FLICKER DIFFUSION: ACCELERATING DENOISING DIFFUSION SAMPLING VIA $1/f$-TYPE COLORED NOISE

Anonymous authors
Paper under double-blind review

ABSTRACT

Natural images' approximate power-law spectra motivate Flicker Diffusion, which couples $1/f$-type spatially correlated Gaussian noise with a matching whitened objective. It retains the baseline U-Net and optimizer-update budget without additional training stages. On CIFAR-10, Flicker-DDIM achieves FID-50k 6.289 at $T = 150$/NFE20, versus 6.292 at $T = 500$/NFE100 for white DDIM: fivefold fewer network evaluations (NFE) at similar observed FID for these settings. A separate controlled test (both $T = 150$/NFE40) yields 6.414 ± 0.070 versus 14.445 ± 0.375 over three paired training seeds after identical 15-candidate sampler searches. Matched-schedule ancestral sampling reduces mean FID by approximately 66% on CelebA64 and 82% on LSUN Church64. Covariance-consistent stochastic differential equations and a perturbative path integral describe a leading affine response, its nonlinear corrections, and a bound on residual effects on spectral evolution. Held-out diagnostics and forward-to-reverse spectral predictions on a separate historical $T = 500$ CIFAR model pair support a more nearly affine low-frequency response. Complementing the sampling gains, our analysis provides a theoretical framework for understanding partial spectral linearization.

1 INTRODUCTION

Diffusion models generate images by iteratively reversing a noise process (Ho et al., 2020; Song et al., 2021). Repeated denoiser evaluations make sampling costly. Existing approaches improve numerical sampling or distill trained models (Song et al., 2020; Lu et al., 2022; Salimans & Ho, 2022). We ask whether spatially correlated corruption and its matching prediction objective can reduce the network evaluations needed for high-quality generation.

Natural images have structured, non-flat spectra: their average power spectra are well approximated by $P_{\mathrm{data}}(k) \propto k^{-\alpha}$ over a broad range of spatial frequencies, with α often near 2 and varying across image categories and orientations (Ruderman, 1994; Torralba & Oliva, 2003). This approximate power-law scaling motivates a power-law-correlated noise family instead of spectrally flat white noise. Under white corruption, modes with unequal energies pass through informative signal-to-noise regimes at different times. Changing the spatial covariance therefore offers a way to alter these modal time scales.

We introduce Flicker Diffusion, which couples a spatial power-law covariance with a matching whitened objective. The covariance distributes corruption across spatial patterns; whitening measures prediction errors relative to that noise allocation. We call its ancestral DDPM and covariance-aware DDIM sampling instances Flicker-DDPM and Flicker-DDIM, respectively; both use the same trained denoiser without retraining (Song et al., 2020). **Our contributions are threefold.**

First, a data-motivated covariance–objective design. Building on covariance-aware diffusion (Voleti et al., 2022; Zhang et al., 2022), we develop this design with the baseline U-Net and optimizer-update budget, without an additional training stage (Section 2). Measured dataset spectra guide covariance candidates for dataset adaptation; a crossed covariance–loss ablation supports the joint design (Section 4.5).

Second, sampling gains at reduced and fixed evaluation budgets. On CIFAR-10, Flicker-DDIM achieves FID-50k 6.289 with 20 network evaluations, compared with 6.292 at 100 evaluations for the tested white-noise DDIM baseline: a **fivefold reduction in sampling NFE** at similar observed

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

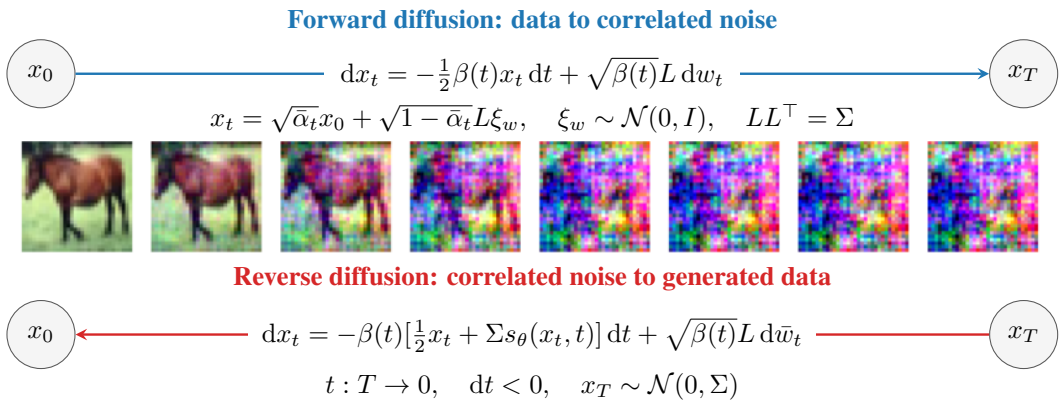


Figure 1: **Flicker Diffusion, from corruption to generation.** The forward and reverse processes share $\Sigma = LL^\top$; generation proceeds from $t = T$ to $t = 0$. The image strip illustrates forward corruption. Training uses the same U-Net architecture and the whitened residual in Eq. 2.

FID (Section 4.3). A separate controlled Flicker-DDIM comparison at 40 evaluations reduces mean FID from 14.445 to 6.414 across three paired training seeds after equal-budget sampler searches (Section 4.2). Matched-schedule comparisons on CelebA64, Church64 and AFHQ-Cat64 extend the quality benefit (Section 4.4).

Third, a quantitative framework for partial spectral linearization. Covariance-consistent SDEs and a perturbative path integral describe spectral evolution under an affine response and its corrections, with an explicit bound on residual effects (Section 3). A complementary generalized-mode analysis separates Gaussian finite-terminal mismatch from finite-step contraction and describes second-order denoising clocks and higher-order cumulant attenuation. Held-out diagnostics on the main three-seed models and independent spectral predictions on a historical model pair support a more predictable low-frequency affine response (Section 4.6).

2 Flicker Diffusion: Covariance and Consistent Sampling

The design uses one covariance throughout: it shapes the noise added to images, defines the metric for learning to predict that noise, and sets the sampling prior. The U-Net architecture remains unchanged. The forward and reverse conventions are illustrated in Fig. 1.

From image spectra to spatial covariance. For pixels i, j at Manhattan distance $d(i, j)$, define

$$\Sigma_{ij} = (1 + d(i, j))^{-\eta}, \qquad LL^\top = \Sigma, \qquad \eta > 0. \tag{1}$$

Larger η weakens correlations; white noise is $\Sigma = I$. Here, $1/f$-type denotes this power-law-correlated family, whose finite-grid spectrum depends on η and the boundary convention. Measured dataset spectra guide exponent candidates for dataset adaptation. The Wiener–Khinchin theorem relates a stationary field's covariance to its power spectrum by Fourier transformation. A Matérn-based analogy, retaining the algebraic prefactor while omitting its exponential cutoff, motivates the heuristic candidate rule (Appendix S). This is a candidate-selection heuristic, not an exact spectral identity for the finite-grid kernel in Eq. 1.

$$\eta_{\text{cand}} = \frac{3 - \alpha}{2}.$$

Forward corruption and whitened training. Let $\alpha_t = 1 - \beta_t$ and let $\bar{\alpha}_t$ denote its cumulative product. We generate colored noise by applying L to a standard Gaussian draw and train the network to predict it with a whitened residual:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\xi_c, \quad \xi_c = L\xi_w, \quad \xi_w \sim \mathcal{N}(0, I), \qquad \mathcal{L} = \mathbb{E}\|L^{-1}(\epsilon_\theta(x_t, t) - \xi_c)\|^2. \tag{2}$$

Whitening measures prediction error in the coordinates of the underlying standard Gaussian noise. Thus, the covariance sets corruption strength across modes, while the loss measures errors in units of that noise. The optimum is $\epsilon_* = \mathbb{E}[\xi_c \mid x_t]$, with score $s_* = -\Sigma^{-1}\epsilon_*/\sqrt{1-\bar{\alpha}_t}$.

Coordinate interpretation. The transformation $y = L^{-1}x$ gives $y_t = \sqrt{\bar{\alpha}_t}y_0 + \sqrt{1-\bar{\alpha}_t}\xi_w$ and converts the loss into a standard noise residual for $L^{-1}\epsilon_\theta(Ly_t, t)$. The transformed problem has white noise, but its data and effective network parameterization also change. With a fixed pixel-space U-Net, this changes the inductive bias of finite-network learning.

Covariance-consistent sampling. Generation uses the same covariance as training, starting from $x_T \sim \mathcal{N}(0, \Sigma)$. Applying DDIM in the whitened coordinates and mapping back gives, for $t \to u < t$,

$$\hat{x}_0 = (x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta)/\sqrt{\bar{\alpha}_t}, \qquad x_u = \sqrt{\bar{\alpha}_u}\hat{x}_0 + \sqrt{1-\bar{\alpha}_u}\epsilon_\theta. \tag{3}$$

The colored prediction ϵ_θ directly defines Flicker-DDIM through this update (Song et al., 2020; Voleti et al., 2022). We use its deterministic form; the same denoiser supports ancestral Flicker-DDPM and DPM-Solver++. Stochastic updates and covariance implementations appear in Appendices J and N.

3 Spectral linearization and nonlinear corrections

Changing the covariance alters both injected fluctuations and the learned restoring response. We describe their joint spectral dynamics through the colored reverse SDE. Itô's formula gives an exact mode-power equation. The Martin–Siggia–Rose–Janssen–de Dominicis (MSRJD) path integral recovers the leading affine dynamics through its quadratic action and organizes nonlinear corrections through interaction terms. The residual bound gives a sufficient condition for accurate affine spectral prediction: a small propagated contribution from the response remainder.

3.1 Reverse generation as a stochastic process

We use x_t for the image state in both directions: forward corruption runs from $t = 0$ to T, and generation runs from $t = T$ to 0. With $\Sigma = LL^\top$, the corresponding SDEs are (Song et al., 2021; Voleti et al., 2022)

$$\mathrm{d}x_t = -\tfrac{1}{2}\beta(t)x_t\,\mathrm{d}t + \sqrt{\beta(t)}L\,\mathrm{d}w_t, \tag{4}$$

$$\mathrm{d}x_t = -\beta(t)[\tfrac{1}{2}x_t + \Sigma s_\theta(x_t, t)]\,\mathrm{d}t + \sqrt{\beta(t)}L\,\mathrm{d}\bar{w}_t. \tag{5}$$

Here w_t is forward Brownian motion; $\bar{w}_t$ has increment covariance $I|\mathrm{d}t|$ along decreasing t. The forward marginal is Eq. 2; Appendix A gives its continuous schedule embedding. Sampling starts from $x_T \sim \mathcal{N}(0, \Sigma)$. We denote its actual generated law by q_t, distinguishing it from the forward marginal p_t: equality $q_t = p_t$ requires the exact score and initialization from the true terminal law.

The network predicts colored noise, so $\Sigma s_\theta = -\epsilon_\theta/\sqrt{1-\bar{\alpha}_t}$. This measurable, covariance-weighted field supplies the restoring response, while Σ also sets the injected fluctuations. These stochastic dynamics describe ancestral generation; probability flow halves the score correction and removes the stochastic term (Appendix A).

3.2 Frequency-domain linearization and spectral evolution

The leading approximation is an affine response for each spatial Fourier mode. Write $\tilde{x}(\mathbf{k}, t) = (Fx_t)_\mathbf{k}$ for the unitary Fourier transform and $y(\mathbf{k}, t) = \tilde{x}(\mathbf{k}, t) - \mathbb{E}_{q_t}\tilde{x}(\mathbf{k}, t)$ for its centered fluctuation. With $g_t = \Sigma s_\theta$, we separate the centered, covariance-weighted score into a diagonal response and a remainder:

$$\tilde{g}(\mathbf{k}, t) - \mathbb{E}_{q_t}\tilde{g}(\mathbf{k}, t) = -\gamma(\mathbf{k}, t)y(\mathbf{k}, t) + \rho(\mathbf{k}, t). \tag{6}$$

The effective restoring coefficient $\gamma(\mathbf{k}, t)$ includes the noise covariance. For periodic Σ, it is the raw-score response multiplied by the noise power $\tilde{\Sigma}(\mathbf{k})$; white noise has $\tilde{\Sigma}(\mathbf{k}) = 1$. For the main nonperiodic CIFAR operator, we fit γ directly and define $\tilde{\Sigma}(\mathbf{k}) = (F\Sigma F^*)_{\mathbf{kk}}$, retaining departures from a diagonal response in ρ.

Our observable is the centered mode power $D(\mathbf{k}, t) = \mathbb{E}_{q_t}|y(\mathbf{k}, t)|^2$: it tracks how much variance generation assigns to each frequency. Applying Itô's rule to Eq. 5 gives

$$-\frac{\mathrm{d}D}{\mathrm{d}t} = \beta(t)[1 - 2\,\mathrm{Re}\,\gamma]D + \beta(t)\tilde{\Sigma}(\mathbf{k}) + 2\beta(t)\,\mathrm{Re}\,\mathbb{E}[\rho\overline{y}]. \tag{7}$$

Arguments $(\mathbf{k}, t)$ are suppressed on the right; the minus sign reflects decreasing t. Along generation, the first term contracts power when $\mathrm{Re}\,\gamma > 1/2$; the second injects noise. The last records nonlinear response and omitted linear mode coupling. This identity holds for any chosen γ; neglecting the last term is the linearization to be tested, rather than a consequence of the Fourier transform.

Once γ is specified, the leading equation predicts the spectrum. Write $\mu(\mathbf{k}, t) = \beta(t)[1/2 - \gamma(\mathbf{k}, t)]$ for the drift per positive generation increment. For $t \le u$ and initial power $D(\mathbf{k}, T)$,

$$\begin{aligned} G(\mathbf{k}; t, u) &= \exp\left[\int_t^u \mu(\mathbf{k}, v)\,\mathrm{d}v\right], \\ D^{(0)}(\mathbf{k}, t) &= |G(\mathbf{k}; t, T)|^2 D(\mathbf{k}, T) + \int_t^T |G(\mathbf{k}; t, u)|^2 \beta(u)\tilde{\Sigma}(\mathbf{k})\,\mathrm{d}u. \end{aligned} \tag{8}$$

The two contributions propagate the initial spectrum and accumulate injected noise. Appendix B derives the propagator and bounds the spectral correction from ρ.

3.3 Measuring the response and testing the linearization

At each $(\mathbf{k}, t)$, we fit a real slope a and complex intercept c in $\tilde{\epsilon}_\theta \simeq a\tilde{x} + c$ on forward-corrupted or generated states. The regression gives $a = \mathrm{Re}\,\mathrm{Cov}(\tilde{\epsilon}_\theta, \tilde{x})/\,\mathrm{Var}[\tilde{x}]$ and $\gamma(\mathbf{k}, t) = a/\sqrt{1 - \bar{\alpha}_t}$, since $g_t = -\epsilon_\theta/\sqrt{1 - \bar{\alpha}_t}$. Thus γ is an ensemble regression coefficient, not an assumed pointwise Jacobian of the U-Net; c retains the mean response.

Held-out fits test whether an affine response describes the network. Keeping a, c and the calibration mean $\bar{\epsilon}_{\mathrm{fit}}$ fixed, we evaluate

$$R^2_{\mathrm{val}}(\mathbf{k}, t) = 1 - \frac{\mathbb{E}_{\mathrm{val}}|\tilde{\epsilon}_\theta - a\tilde{x} - c|^2}{\mathbb{E}_{\mathrm{val}}|\tilde{\epsilon}_\theta - \bar{\epsilon}_{\mathrm{fit}}|^2}. \tag{9}$$

A high value means that the leading response explains much of the network-output variation on unseen states. Figure 5 tests this along generated trajectories for the historical $T = 500$ pair. The main $T = 150$ three-seed models are separately tested on held-out forward-corrupted images (Table 2).

Independent spectral predictions test the dynamical consequence. We fix responses fitted on forward-corrupted images and propagate mode powers on independent reverse trajectories. The comparison uses the discrete ancestral counterpart of the leading equation, with the actual sampling schedule and finite-sample centering correction (Appendix C, Eq. 38). This tests whether the measured response transfers to generation. Re-fitting on the same state law can close the instantaneous second-moment equation by regression orthogonality; held-out fits and forward-to-reverse prediction therefore probe complementary aspects of the approximation.

We call an adequate leading affine response in selected frequency–time regimes *partial spectral linearization*. It does not imply global linearity or smaller numerical integration error. In the MSRJD representation of the same SDE, the quadratic action gives affine propagators and nonlinear drift gives interaction corrections (Martin et al., 1973; Janssen, 1976). Appendix B connects these terms to non-Gaussian cumulants and derives the residual bound and first loop terms; the bound and loop terms are not measured. A complementary Gaussian analysis isolates how the data and noise covariances set modal denoising times and terminal mismatch, separately from finite-step DDIM contraction (Appendix K).

4 Experiments

We test quality at fixed NFE, evaluation budgets at similar observed FID, and transfer across datasets. We then examine the covariance–loss pairing, exponent selection, sensitivity to training choices, and the spectral description.

4.1 Experimental setup

Datasets and training. Primary comparisons match U-Net families and optimizer updates within white/Flicker pairs, using three paired training seeds for CIFAR and two for 64×64 datasets. FID-50k uses 50,000 generated and 50,000 fixed reference images for CIFAR/CelebA/Church (Heusel et al., 2017; Krizhevsky, 2009); AFHQ uses all 5,065 training images as its reference.

Table 1: Distinct questions in the CIFAR controls. The first two rows share linear beta endpoints $(10^{-4}, 0.028)$. The third resamples the cumulative $T = 500$ path. Each row retains its own evaluation protocol.

Training path	T	Terminal $\bar{\alpha}$	Evaluation / question
Linear	150	.11913	DDIM NFE40 / same-T benefit
Linear	500	.0008321	NFE40 / ranking at $T = 500$
Matched terminal	150	.0008321	Ancestral NFE300 / path control

Sampling and metrics. NFE counts actual network calls, separately from training updates and diffusion steps T. CIFAR classifier-free guidance (CFG) uses $(1 + w)\epsilon_{\text{cond}} - w\epsilon_{\text{uncond}}$ (Ho & Salimans, 2022): for $w > 0$, 20 timepoints require 40 U-Net calls; unconditional sampling uses one call per timepoint. NFE measures efficiency, not wall-clock speedup; covariance and whitening add overhead. Reported $\pm$ values are sample SDs across training seeds. FID-2k sampler-selection pilots precede fresh confirmations. Historical FID-10k, exploratory FID-5k and single-pair spectral diagnostics retain separate protocols (Appendix N, Table 19).

Training paths. At fixed beta endpoints, reducing T changes both terminal SNR and discretization. Equal-T comparisons test the design within a shared schedule; $T = 500$ and matched-terminal controls test path and terminal-SNR sensitivity (Table 1).

4.2 Generation quality at a fixed evaluation budget

On CIFAR-10, Flicker-DDIM obtains $\mathbf{6.414 \pm 0.070}$ FID versus $\mathbf{14.445 \pm 0.375}$ for white at actual NFE40 (Fig. 2). This main comparison fixes $T = 150$, 90,000 updates and global batch 128.

Quality improves after equal-budget sampler selection. Each white and dense Flicker checkpoint searches the same 15 settings: uniform/quadratic/logSNR grids crossed with $w \in \{0.5, 1, 1.5, 2, 3\}$. All 90 model–candidate pilots enter selection, followed by six independent FID-50k confirmations. Every seed improves, reducing mean FID by 56% at matched architecture, updates and NFE. Figure 2(b) shows fixed-index samples; Appendix H.1 details the search and confirmations.

The batch-128 gain extends to DPM++ and distribution coverage. A separate 15-candidate DPM++ search on the checkpoints of Fig. 2 confirms FID 21.133 ± 0.611 for white and 8.607 ± 0.106 for Flicker at NFE40. On the main DDIM images, lower KID and higher recall and coverage support improved distribution coverage; precision is similar. Appendix F gives results and metric protocols.

4.3 Sampling efficiency on CIFAR-10

To compare NFE at similar observed FID, this separate batch-64 study uses 90,000 updates and three paired training seeds. Flicker-DDIM obtains FID-50k 6.289 ± 0.15 at NFE20, versus white DDIM's 6.292 ± 0.08 at NFE100 (Fig. 3). Flicker uses $T = 150$, uniform/$w = 1$; white uses $T = 500$, quadratic/$w = 0.5$.

Flicker achieves a fivefold NFE reduction at these operating points. It uses one fifth as many calls at approximately the same observed mean FID. NFE20 was added after inspecting the initial points (Appendix G). Figure 3 reports configuration-specific curves, not per-NFE optimized frontiers; the appendix gives the selection history and statistical scope.

Appendix T separately illustrates these batch-64 operating points.

4.4 Results on additional datasets

On CelebA64 and LSUN Church64 (Liu et al., 2015; Yu et al., 2015), ancestral Flicker-DDPM reduces mean FID from 79.623 to 27.327 and from 46.444 to 8.569: approximately 66% and 82% (Fig. 13).

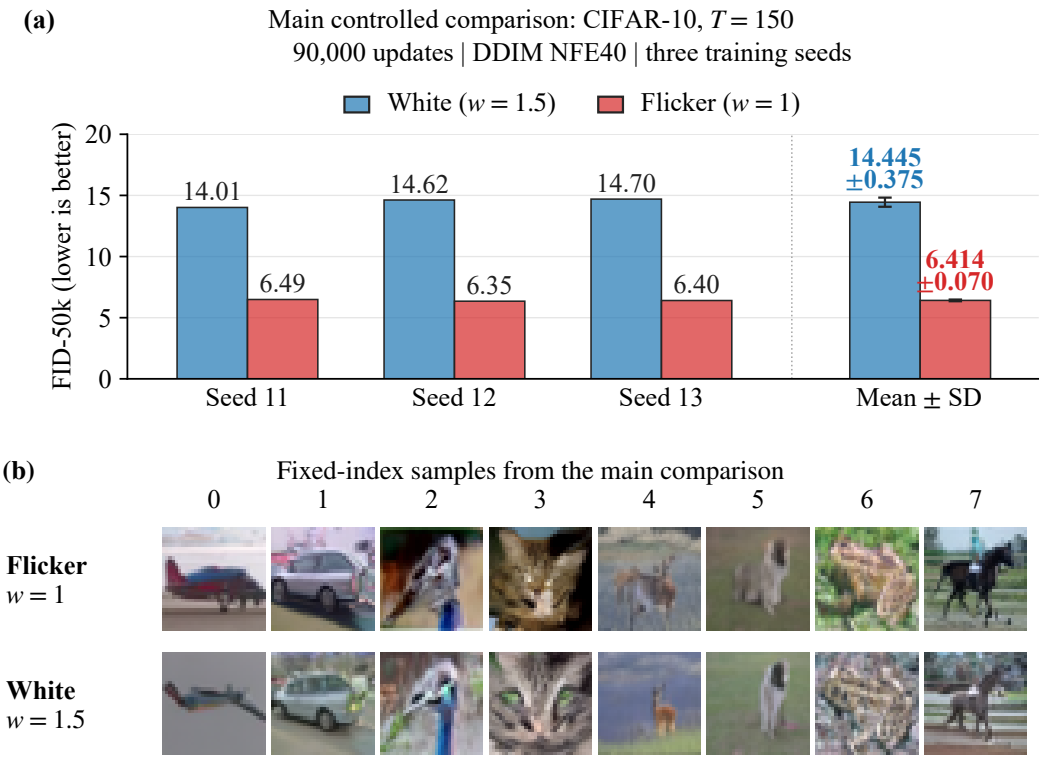


Figure 2: **Main controlled CIFAR-10 comparison: quantitative and qualitative results.** All models use $T = 150$, 90,000 updates, batch 128 and DDIM at actual NFE40. (a) Fresh FID-50k confirmations after identical 15-candidate searches: three training seeds (suffixes 11–13) and means with sample SDs. All select uniform grids, with interior guidance $w = 1$ (Flicker) and 1.5 (white). (b) Seed 2026091211, indices 0–7 fixed before viewing each FID-50k output; pixels are unchanged. Equal indices do not imply paired content. Appendix T records provenance.

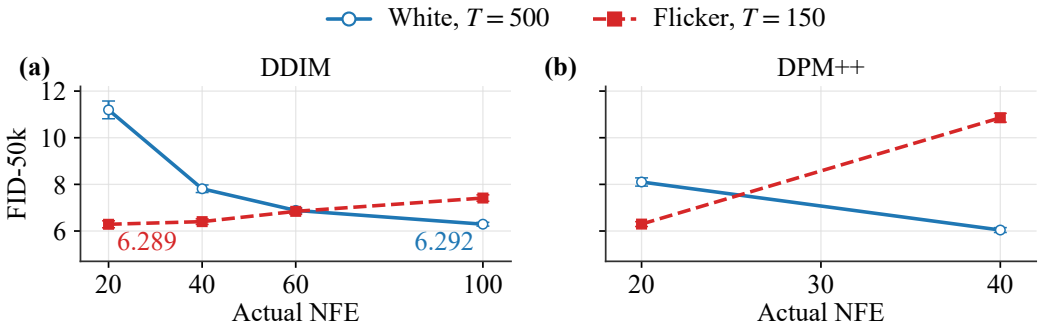


Figure 3: **Batch-64 sampling-efficiency study (Section 4.3).** A separate batch-64 study comprises 36 FID-50k measurements, with sample SD across three paired training seeds. White uses $T = 500$, quadratic/$w = 0.5$; Flicker uses $T = 150$, uniform/$w = 1$. Only white DPM++ uses dynamic thresholding. Each curve follows its stated model, diffusion-step count T and sampler configuration.

Both datasets improve across paired seeds with ancestral sampling. Comparisons use $T = 150$, NFE150, two paired seeds, 50k generated/reference images, the same unconditional U-Net, 90k updates, global batch 128 and linear schedule. Periodic covariances use dataset-specific exponents instead of CIFAR's fixed operator (Appendix I).

The quality benefit extends to DDIM and AFHQ-Cat64. At NFE40 after equal two-grid selection, mean FID falls from 178.103 to 57.643 on CelebA and from 107.755 to 23.420 on Church (Table 14). On AFHQ-Cat64 at $T = 150$/NFE40, FID is 55.307 for white DDIM and 23.816 for Flicker-DDIM ($\eta \approx 0.160$), a 57% reduction over two paired seeds. Table 16 also reports alternative $\eta = 0.05$ and white $T = 500$ results.

Historical white-noise models remain stronger at $T = 500$. Unpaired white checkpoints obtain 4.521 and 4.538 on CelebA and Church at the same NFE/search budget (Table 14).

4.5 Design ablations and controls

A three-seed 2×2 ablation crosses white/Flicker noise and pixel/Flicker-whitened metrics at $T = 150$, uniform DDIM, $w = 1.8$, actual NFE40 (Appendix S, Table 26). Mean FIDs are 15.089 (white/pixel), 15.550 (white/Flicker-whitened), 24.339 (Flicker/pixel), and 6.856 (Flicker/Flicker-whitened).

Covariance and whitening work jointly in this training recipe. Neither change alone reproduces the improvement. The paired interaction is -17.945 ± 1.309 FID; the metric affects gradient scale as well as directional weighting. Seeds and checkpoints differ from Section 4.2.

The main CIFAR-10 comparison fixes $\eta = 0.2$. Figure 4 examines sensitivity to this choice in a separate historical 200-epoch study; Figure 2 reports the three-seed, 90,000-update DDIM comparison. The historical $T = 150$ sweep tests seven exponents from 0 to 0.30: the lowest observed FID-50k is 8.52 at $\eta = 0.20$, versus 23.27 for its white-noise baseline.

Generation quality varies with the covariance exponent. The fully correlated $\eta = 0$ limit gives FID 444.04; neighboring $\eta = 0.15$ and 0.25 give 10.23 and 9.93. This scan supports the empirical CIFAR choice and characterizes covariance sensitivity under its historical training protocol (Appendix S).

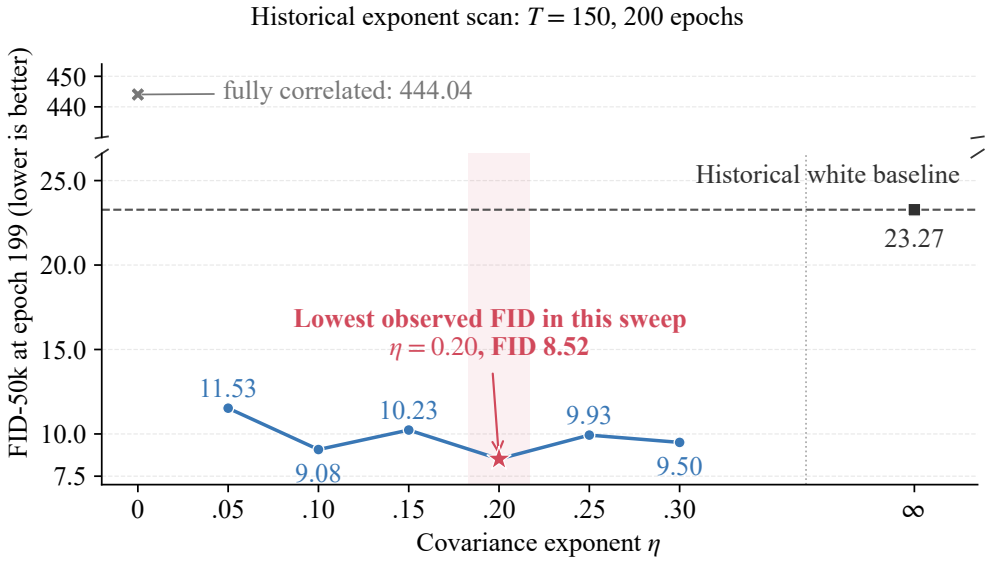


Figure 4: **Covariance-exponent sensitivity in the historical study.** $T = 150$, 200 epochs; each point is one final-checkpoint FID at epoch 199, using 50k generated images and the same 50k-image reference. The star marks the lowest observed FID in this sweep, 63.4% below its historical white baseline ($\eta \to \infty$). The broken axis retains $\eta = 0$. These checkpoints are separate from Figure 2.

Flicker improves FID with both tested prediction targets. Against our budget-controlled, data-power-based EqualSNR-style variants, Flicker improves both noise and clean-image prediction (Appendix S). Each design uses its own covariance and scalar schedule; this does not reproduce the original method's complete training recipe.

A gain remains under the matched-terminal protocol. To vary T while preserving terminal SNR, we resample the $T = 500$ cumulative log-alpha path to 150 steps. Flicker obtains FID 13.58 ± 0.29 versus white's 15.39 ± 0.32 using full ancestral sampling, CFG $w = 1.8$, actual NFE300 and a separate $T = 500$ ancestral anchor (Appendix U). This protocol differs from the main DDIM experiment, so the two FID gaps do not isolate the contribution of terminal mismatch. Initial-covariance calibration separately tests whether correcting this second moment improves FID (Section 6).

4.6 Spectral dynamics and affine predictability

The learned-field tests below evaluate the spectral description in Section 3. Separate covariance-only Gaussian calculations show narrower modal timing and smaller finite-terminal mismatch for Flicker at $T = 150$; definitions and numerical results appear in Appendices K and L.

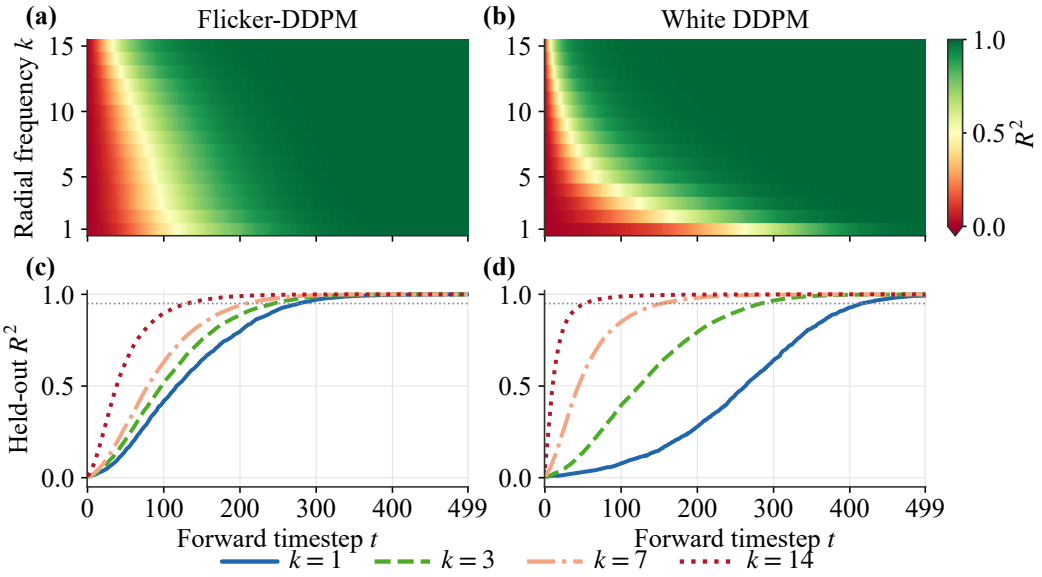


Figure 5: **Low-frequency affine predictability on generated states.** One historical $T = 500$ CIFAR-10 checkpoint pair, with ancestral sampling and CFG $w = 1.8$. Modal affine fits use 128 trajectories and are evaluated on 128 independent trajectories at all 500 steps. Flicker improves low-frequency predictability; white has higher R^2 at $k = 7, 14$. Generation runs from $t = 499$ to $t = 0$, right to left.

Table 2: Partial affine predictability on the main $T = 150$ models. Means across three training seeds; R^2 values use held-out images and equal weighting of all 150 forward indices. The global row instead pools residual and baseline energies across non-DC modes and time.

Measure	White	Flicker
$k = 1$	0.1858	0.5463
$k = 3$	0.5219	0.6052
$k = 7$	0.7965	0.6852
$k = 14$	0.9395	0.8325
Four-shell range	0.7537	0.2862
Global energy-weighted R^2	0.9019	0.6937

The main models improve low-frequency affine predictability. Flicker improves the two reported low-frequency shells and narrows the four-shell spread (Table 2). On the six main $T = 150$ checkpoints, we fit real-scalar Fourier responses at all 150 forward indices using 250 test images and evaluate on 250 disjoint, class-balanced images. Models share images and Gaussian innovations but use their own training covariance and true labels. Common CFG $w = 1$ differs from the white

model's selected FID guidance. White has higher high-frequency and globally energy-weighted R^2. The same pattern holds without guidance and at common $w = 1.5$; training-seed values and SDs appear in Appendix F.

Low-frequency affine predictability persists on generated states. On an independent historical $T = 500$ checkpoint pair, held-out regression on free reverse trajectories confirms the low-frequency improvement (Fig. 5).

Forward fits predict independent reverse spectra at low frequencies. On this pair, coefficients fitted on forward-corrupted images predict independent reverse spectra with trajectory errors of 3.763% and 1.980% at $k = 1, 3$, versus white's 9.950% and 3.350%. White predicts the reported higher-frequency shells more accurately, supporting a frequency-dependent leading response. Appendix C gives model, regression and trajectory protocols. The diagonal residual includes omitted linear mode coupling and does not uniquely measure nonlinearity.

Finite-perturbation probes test local response; one coarse DDIM step versus two split steps tests update self-consistency under the stated perturbations and grids (Section 6).

5 Related work

Covariance and frequency design. Covariance-aware frameworks (Voleti et al., 2022; Zhang et al., 2022) support frequency-shaped corruption (Rissanen et al., 2022; Hoogeboom & Salimans, 2023; Huang et al., 2024; Tivnan et al., 2023). EqualSNR uses data power (Falck et al., 2025); SAGD shapes Fourier noise (Scimeca et al., 2025). Flicker studies spatial power-law covariance and its whitened metric through training controls and spectral diagnostics.

Sampling efficiency. Solvers, distillation, preconditioning and schedule optimization offer other routes to efficiency (Song et al., 2020; Lu et al., 2022; Salimans & Ho, 2022; Karras et al., 2022; Sabour et al., 2024). CNS changes inference-time noise (Davidson et al., 2026); Flicker changes the training covariance and objective, supporting multiple samplers.

Linear structure and spectral dynamics. Gaussian and spectral analyses study denoisers and sampling (Wang & Vastola, 2023; Li et al., 2024; Wang & Pehlevan, 2025; Benita et al., 2025; Pierret & Galerne, 2025). We test covariance-dependent response with retained corrections. Appendix A details these comparisons.

6 Discussion and conclusion

Flicker Diffusion jointly designs spatial covariance and a whitened objective for efficient sampling. Its CIFAR-10 DDIM instance uses fivefold fewer evaluations at similar observed FID to the tested white-noise configuration; a separate NFE40 comparison improves FID from 14.445 to 6.414 with the same U-Net and optimizer-update budget. Matched-schedule results on CelebA64, Church64 and AFHQ-Cat64 extend the quality benefit. The SDE and path-integral analysis describes modal response and its corrections, supported by low-frequency affine diagnostics on the main three-seed models and a historical model pair.

The controls identify training-path and sampler dependence. At CIFAR $T = 500$/NFE40, equal eight-candidate selection favors white over dense Flicker: DDIM FID 7.642 versus 8.171, and DPM++ 6.017 versus 6.834 (Table 9); the 64×64 white $T = 500$ anchors also remain stronger at NFE40 (Section 4.4). Initial-covariance calibration worsens white FID from 14.503 to 37.806 and Flicker from 6.445 to 7.992 (Appendix H.2): changing the initial second moment alone is insufficient, while correction of the full terminal distribution remains untested. Finite-perturbation responses and split-step self-consistency defects also show no uniform Flicker advantage (Appendices Q and P).

Together, the sampling results and spectral predictions establish data-informed covariance–objective design as a practical method and a quantitative perspective on diffusion dynamics, linking noise correlations to modal timing, affine response and nonlinear corrections.

REPRODUCIBILITY STATEMENT

The appendices give the derivations, assumptions, covariance conventions, architectures, preprocessing, training budgets, seeds, sampler searches and evaluation protocols, with a protocol index in Appendix N. The anonymous supplementary archive provides the frozen covariance, whitened-objective, U-Net, training and DDIM modules; the six main CIFAR-10 configurations; and archived selection and confirmation records, control results, spectral arrays and figure assets. Its `README.md` and `docs/EXPERIMENT_INDEX.md` map this evidence to the paper, while `docs/REPRODUCTION.md` describes dependencies and minimal training/sampling entry points. The scripts `verify_package.py` and `code/verify_results.py` check packaged-file integrity and reproduce aggregates from saved measurements; original-source and release-file hashes are distinguished. The archive documents historical protocol gaps and the reported initialization discrepancy. Datasets, checkpoints, Inception weights and a complete portable FID/search pipeline are not bundled; training and sampling were not rerun during archive preparation.

AI USE STATEMENT

Generative AI assistants were used for literature discovery, research discussion, experiment design, implementation and debugging, mathematical derivations and proof drafting, analysis of completed results, scientific figures, and manuscript drafting and revision. AI assistance also supported organization and anonymization of the supplementary archive, preparation of its documentation, and checks of source hashes, result aggregation and manuscript consistency. Mathematical claims were subjected to algebraic and numerical consistency checks, and reported results were cross-checked against configuration and completion records. The supplementary material retains the archived measurements and sample images. AI-generated prose or predictions were not used as experimental measurements. The authors are responsible for the final content and interpretation.

REFERENCES

Roi Benita, Michael Elad, and Joseph Keshet. Spectral analysis of diffusion models with application to schedule design. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2502.00180.

Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1801.01401.

Ting Chen. On the importance of noise scheduling for diffusion models. *arXiv preprint arXiv:2301.10972*, 2023. URL https://arxiv.org/abs/2301.10972.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. StarGAN v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8188–8197, 2020. URL https://arxiv.org/abs/1912.01865.

Hadar Davidson, Noam Issachar, and Sagie Benaim. Colored noise diffusion sampling. *arXiv preprint arXiv:2605.30332*, 2026. URL https://arxiv.org/abs/2605.30332.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. doi: 10.1198/jasa.2011.tm11181. URL https://pubmed.ncbi.nlm.nih.gov/22505788/.

Fabian Falck, Teodora Pandeva, Kiarash Zahirnia, Rachel Lawrence, Richard Turner, Edward Meeds, Javier Zazo, and Sushrut Karmalkar. A Fourier space perspective on diffusion models. *arXiv preprint arXiv:2505.11278*, 2025. doi: 10.48550/arXiv.2505.11278.

John A. Hertz, Yasser Roudi, and Peter Sollich. Path integral methods for the dynamics of stochastic and disordered systems. *arXiv preprint arXiv:1604.05775*, 2016. URL https://arxiv.org/abs/1604.05775.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. doi: 10.48550/arXiv.1706.08500.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. URL https://arxiv.org/abs/2207.12598.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. doi: 10.48550/arXiv.2006.11239.

Emiel Hoogeboom and Tim Salimans. Blurring diffusion models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OjDkC57x5sz.

Xingchang Huang, Corentin Salaun, Cristina Vasconcelos, Christian Theobalt, Cengiz Oztireli, and Gurprit Singh. Blue noise for diffusion models. In *ACM SIGGRAPH 2024 conference papers*, pp. 1–11, 2024. doi: 10.1145/3641519.3657435.

Hans-Karl Janssen. On a lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties. *Zeitschrift für Physik B Condensed Matter*, 23(4):377–380, 1976. doi: 10.1007/BF01316547.

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*, volume 35, 2022. URL https://arxiv.org/abs/2206.00364.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, Toronto, ON, Canada, 2009.

Xiang Li, Yixiang Dai, and Qing Qu. Understanding generalizability of diffusion models requires rethinking the hidden Gaussian structure. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/69e68611ec4d8c0ae4a4b2bece165f5f-Paper-Conference.pdf.

Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common diffusion noise schedules and sample steps are flawed. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5404–5411, 2024. URL https://arxiv.org/abs/2305.08891.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.03003.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3730–3738, 2015. URL https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022. URL https://arxiv.org/abs/2211.01095.

Hila Manor and Tomer Michaeli. On the posterior distribution in denoising: Application to uncertainty quantification. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=adSGeugiuj.

P. C. Martin, E. D. Siggia, and H. A. Rose. Statistical dynamics of classical systems. *Physical Review A*, 8(1):423–437, 1973. doi: 10.1103/PhysRevA.8.423.

Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 7176–7185. PMLR, 2020. URL https://proceedings.mlr.press/v119/naeem20a.html.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 8162–8171. PMLR, 2021. URL https://proceedings.mlr.press/v139/nichol21a.html.

Emile Pierret and Bruno Galerne. Diffusion models for Gaussian distributions: Exact solutions and Wasserstein errors. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 49355–49381. PMLR, 2025. URL https://proceedings.mlr.press/v267/pierret25a.html.

Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. *arXiv preprint arXiv:2206.13397*, 2022. doi: 10.48550/arXiv.2206.13397.

Daniel L Ruderman. The statistics of natural images. *Network: computation in neural systems*, 5 (4):517–548, 1994. doi: 10.1088/0954-898X_5_4_006.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your steps: Optimizing sampling schedules in diffusion models. *arXiv preprint arXiv:2404.14507*, 2024. URL https://arxiv.org/abs/2404.14507.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022. doi: 10.48550/arXiv.2202.00512.

Luca Scimeca, Thomas Jiralerspong, Berton Earnshaw, Jason Hartford, and Yoshua Bengio. Learning what matters: Steering diffusion via spectrally anisotropic forward noise. *arXiv preprint arXiv:2510.09660*, 2025. URL https://arxiv.org/abs/2510.09660.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. doi: 10.48550/arXiv.2010.02502.

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. URL https://arxiv.org/abs/2011.13456.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 32211–32252. PMLR, 2023. URL https://proceedings.mlr.press/v202/song23a.html.

Uwe C. Täuber. Field theory approaches to nonequilibrium dynamics. *arXiv preprint arXiv:cond-mat/0511743*, 2005. URL https://arxiv.org/abs/cond-mat/0511743.

Jiayan Teng, Wendi Zheng, Ming Ding, Wenyi Hong, Jianqiao Wangni, Zhuoyi Yang, and Jie Tang. Relay diffusion: Unifying diffusion process across resolutions for image synthesis. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2309.03350.

Matthew Tivnan, Jacopo Teneggi, Tzu-Cheng Lee, Ruoqiao Zhang, Kirsten Boedeker, Liang Cai, Grace J. Gang, Jeremias Sulam, and J. Webster Stayman. Fourier diffusion models: A method to control MTF and NPS in score-based stochastic image generation. *arXiv preprint arXiv:2303.13285*, 2023. doi: 10.48550/arXiv.2303.13285.

Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14(3):391–412, 2003. doi: 10.1088/0954-898X_14_3_302.

Vikram Voleti, Christopher Pal, and Adam Oberman. Score-based denoising diffusion with non-isotropic gaussian noise models. *arXiv preprint arXiv:2210.12254*, 2022. URL https://arxiv.org/abs/2210.12254.

Binxu Wang and Cengiz Pehlevan. An analytical theory of spectral bias in the learning dynamics of diffusion models. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2503.03206.

Binxu Wang and John J. Vastola. The hidden linear structure in score-based models and its application. *arXiv preprint arXiv:2311.10892*, 2023. URL https://arxiv.org/abs/2311.10892.

Dana Weitzner, Mauricio Delbracio, Peyman Milanfar, and Raja Giryes. The diffusion process as a correlation machine: Linear denoising insights. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=FGDJOc27rt.

Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015. URL https://arxiv.org/abs/1506.03365.

Qinsheng Zhang, Molei Tao, and Yongxin Chen. gDDIM: Generalized denoising diffusion implicit models. *arXiv preprint arXiv:2206.05564*, 2022. URL https://arxiv.org/abs/2206.05564.

A COLORED SDES, EXACT MOMENTS, AND THE SPECTRAL CLOSURE

Scope and notation. We work in a finite-dimensional real image space with a fixed symmetric positive-definite noise covariance $\Sigma = LL^\top$. Its channel extension is included in Σ when several channels are considered. The noise index t increases from data to noise and decreases from T to 0 during generation. Reverse Brownian increments have covariance $I|\mathrm{d}t|$. We assume the coefficients, moments, and boundary decay needed below. The score convention is $s_*(x,t) = \nabla_x \log p_t(x)$, not its negative. A class may be fixed throughout; unconditional and conditional laws must not be interchanged. The same moment algebra applies to a guided field, but this does not identify that field with the exact score of the original conditional forward marginals.

Forward law and noise prediction. The continuous rate and the discrete DDPM variances are distinct. An exact embedding of a discrete schedule $\{\beta_j\}$ sets $\beta(t) = -\log(1-\beta_j)$ on each unit interval $j-1 < t \le j$. Then $\bar\alpha_j = \prod_{i=1}^{j}(1-\beta_i)$ at the integer endpoints. The empirical Gaussian calculations use the actual discrete cumulative products; they do not replace these by $\exp(-\sum_j \beta_j)$. Here j counts noise increments; the implementation's model index is $j-1$ because its training indices start at zero. The solution of Eq. 4 is

$$x_t = \sqrt{\bar\alpha_t}x_0 + \int_0^t \sqrt{\frac{\bar\alpha_t}{\bar\alpha_u}\beta(u)}L\,\mathrm{d}w_u, \qquad \bar\alpha_t = \exp\left[-\int_0^t \beta(u)\,\mathrm{d}u\right]. \tag{10}$$

Its Gaussian innovation has covariance $(1-\bar\alpha_t)\Sigma$, yielding $x_t = \sqrt{\bar\alpha_t}x_0 + \sqrt{1-\bar\alpha_t}\xi_c$, with $\xi_c \sim \mathcal{N}(0,\Sigma)$ independent of x_0. For any data distribution with mean μ_0 and covariance S,

$$\mu_t = \sqrt{\bar\alpha_t}\mu_0, \qquad Q_t = \bar\alpha_t S + (1-\bar\alpha_t)\Sigma, \qquad \frac{\mathrm{d}Q_t}{\mathrm{d}t} = \beta(t)(\Sigma - Q_t). \tag{11}$$

These are exact moments; they do not assert that p_t is Gaussian. Differentiating the Gaussian convolution, for $1-\bar\alpha_t > 0$, gives

$$s_*(x,t) = -\frac{1}{\sqrt{1-\bar\alpha_t}}\Sigma^{-1}\mathbb{E}[\xi_c \mid x_t = x]. \tag{12}$$

The fixed whitened quadratic loss has the minimizer $\epsilon_*(x,t) = \mathbb{E}[\xi_c \mid x_t = x]$. To see this, condition on x_t and expand the quadratic around ϵ_*; its cross term vanishes and its Hessian is $2\Sigma^{-1} \succ 0$. This is a conditional expectation, not the realized noise of each training example.

Reverse time and probability flow. The time-reversal and probability-flow constructions of Song et al. (2021), applied to diffusion matrix $\beta(t)\Sigma$, give Eq. 5; the non-isotropic forward process is also derived by Voleti et al. (2022). To include both dynamics, define $\kappa = 1$ for the reverse SDE and $\kappa = 1/2$ for probability flow. Along decreasing t,

$$\mathrm{d}x_t = -\beta(t)[\tfrac{1}{2}x_t + \kappa\Sigma s_\theta(x_t,t)]\,\mathrm{d}t + \sqrt{(2\kappa-1)\beta(t)}L\,\mathrm{d}\bar w_t. \tag{13}$$

The reverse SDE drift is $-\beta x/2 - \beta\Sigma s_*$, and the probability-flow drift is $-\beta x/2 - \beta\Sigma s_*/2$. The quadratic variation is positive along generation, with $|\mathrm{d}t| = -\mathrm{d}t$. Thus the Fokker–Planck equation is

$$-\partial_t q_t = -\nabla\cdot\big(\beta(t)[\tfrac{1}{2}x + \kappa\Sigma s_\theta]q_t\big) + \tfrac{1}{2}(2\kappa-1)\beta(t)\nabla\cdot(\Sigma\nabla q_t). \tag{14}$$

When $s_\theta = s_*$ and $q_t = p_t$, $\Sigma s_* q_t = \Sigma\nabla q_t$; the net diffusion coefficient on the right is $-\beta(t)/2$ regardless of κ. The right side is the negative of the forward Fokker–Planck equation. Equality of marginals therefore requires the exact field and the correct terminal law. A pure Gaussian prior need not equal p_T at finite T.

Exact centered covariance identity. Set $m_t = \mathbb{E}_{q_t}x_t$, $y_t = x_t - m_t$, $\bar s_t = \mathbb{E}_{q_t}s_\theta(x_t,t)$, and

$$K_t = \mathbb{E}[y_t y_t^\top], \qquad J_t = \mathbb{E}[(s_\theta(x_t,t) - \bar s_t)y_t^\top]. \tag{15}$$

Taking expectation in Eq. 13 gives $-m_t' = \beta(t)(m_t/2 + \kappa\Sigma\bar s_t)$. Applying Itô's product rule to $y_t y_t^\top$, per positive generation increment $|\mathrm{d}t|$, gives drift contributions $\beta(t)K_t/2 + \kappa\beta(t)\Sigma J_t$ and their transpose, and quadratic variation $(2\kappa-1)\beta(t)\Sigma$. Their sum is

$$-K_t' = \beta(t)K_t + \kappa\beta(t)(\Sigma J_t + J_t^\top\Sigma) + (2\kappa-1)\beta(t)\Sigma, \tag{16}$$

For the reverse SDE ($\kappa = 1$), this becomes

$$-\frac{\mathrm{d}K_t}{\mathrm{d}t} = \beta(t)(K_t + \Sigma J_t + J_t^\top \Sigma + \Sigma). \tag{17}$$

It remains valid for non-Gaussian generated states and nonlinear, approximate score fields. It is not closed in K_t unless J_t is separately known. For the exact score under its own marginal, integration by parts gives $\mathbb{E}_{p_t} s_* = 0$ and

$$\mathbb{E}_{p_t}[s_*(x_t, t)(x_t - \mu_t)^\top] = -I. \tag{18}$$

Thus Eq. 16 reduces to $-K_t' = \beta(t)(K_t - \Sigma)$, agreeing with Eq. 11. One cannot insert $J_t = -I$ merely because the network estimates the true forward score: if its generated distribution differs from p_t, the expectation is under a different law.

Linear closure and matrix order. Suppose $s_\theta(x, t) = -\Gamma_t x + c_t$ on the distribution of interest. Here Γ_t is the raw-score response matrix; the effective matrix is $H_t = \Sigma\Gamma_t$. Then $J_t = -\Gamma_t K_t$, and

$$A_t = \beta(t)(\tfrac{1}{2}I - \kappa\Sigma\Gamma_t), \qquad -K_t' = A_t K_t + K_t A_t^\top + (2\kappa - 1)\beta(t)\Sigma. \tag{19}$$

Neither symmetry of Γ_t nor commutativity with Σ is needed. The products $\Sigma\Gamma_t K_t$ and $K_t\Gamma_t^\top\Sigma$ retain their order. The affine offset affects the mean but not this centered covariance equation. A local Jacobian or regression fit is a closure, not an exact identity for a general neural network over an entire ensemble. Under this closure, define $\partial_t\Phi(t, u) = -A_t\Phi(t, u)$ and $\Phi(u, u) = I$ for $t \le u$. Then

$$K_t = \Phi(t, T)K_T\Phi(t, T)^\top + \int_t^T (2\kappa - 1)\beta(u)\Phi(t, u)\Sigma\Phi(t, u)^\top \,\mathrm{d}u. \tag{20}$$

Stochastic forcing and drift must change consistently when changing the trained covariance. Altering only the injected noise of a fixed pretrained field is a different intervention, as in CNS (Davidson et al., 2026).

Periodic spectral special case. Let a unitary Fourier transform diagonalize a periodic Σ, with eigenvalues $\tilde{\Sigma}(\mathbf{k}) > 0$. Further assume the effective response $H_t = \Sigma\Gamma_t$ is diagonal in this basis, with multiplier $\gamma(\mathbf{k}, t)$. Thus the raw-score multiplier is $-\gamma(\mathbf{k}, t)/\tilde{\Sigma}(\mathbf{k})$. For $y(\mathbf{k}, t) = \tilde{x}(\mathbf{k}, t) - \mathbb{E}\tilde{x}(\mathbf{k}, t)$, the drift per positive generation increment is $\beta(t)[1/2 - \kappa\gamma(\mathbf{k}, t)]y$; the quadratic variation is $(2\kappa - 1)\beta(t)\tilde{\Sigma}(\mathbf{k})|\mathrm{d}t|$. Consequently $D(\mathbf{k}, t) = \mathbb{E}|y(\mathbf{k}, t)|^2$ satisfies

$$-\frac{\mathrm{d}D(\mathbf{k}, t)}{\mathrm{d}t} = \beta(t)[1 - 2\kappa\,\mathrm{Re}\,\gamma(\mathbf{k}, t)]D(\mathbf{k}, t) + (2\kappa - 1)\beta(t)\tilde{\Sigma}(\mathbf{k}). \tag{21}$$

For $\kappa = 1$, the spectral closure is

$$-\frac{\mathrm{d}D(\mathbf{k}, t)}{\mathrm{d}t} = \beta(t)[1 - 2\,\mathrm{Re}\,\gamma(\mathbf{k}, t)]D(\mathbf{k}, t) + \beta(t)\tilde{\Sigma}(\mathbf{k}). \tag{22}$$

For real symmetric score responses, γ is real. Otherwise only its real part enters power growth; conjugate symmetry for real images remains in force. If $\kappa = 1$ and the coefficients are frozen, a stationary mode variance exists only when $2\,\mathrm{Re}\,\gamma > 1$:

$$D^{\mathrm{stat}}(\mathbf{k}) = \frac{\tilde{\Sigma}(\mathbf{k})}{2\,\mathrm{Re}\,\gamma(\mathbf{k}) - 1}. \tag{23}$$

This frozen-coefficient expression is not the solution of the time-dependent reverse process. The restoring coefficient γ already includes the covariance factor. For nonzero mean images, $\mathbb{E}|\tilde{x}(\mathbf{k}, t)|^2 = D(\mathbf{k}, t) + |\mathbb{E}\tilde{x}(\mathbf{k}, t)|^2$; a covariance match is not automatically an uncentered-power match. A spatial mean image can also have non-DC Fourier components.

Why stationarity is insufficient for Fourier linearization. For a translation operator P, equivariance of a score field gives $s(Px) = Ps(x)$ and hence $J_s(Px) = PJ_s(x)P^{-1}$. This relates Jacobians at different states; it does not imply $J_s(x)P = PJ_s(x)$ at a fixed nonuniform image. A stationary distribution can have nonlinear translation-equivariant interactions and higher-order cross-frequency dependence. Likewise an invertible Fourier change of coordinates cannot turn a globally nonlinear function into a globally linear one. The matrix moment analysis does not require that claim. Our dense nonperiodic covariance is instead treated by its actual generalized modes in Eq. 43.

Gaussian and white limits, and the mechanism to be tested. For centered Gaussian data, $p_t = \mathcal{N}(0, Q_t)$ and $\Gamma_t = Q_t^{-1}$ exactly. In a simultaneous spectral basis, with centered data power $P_{\mathrm{data}}(\mathbf{k})$,

$$Q(\mathbf{k}, t) = \bar{\alpha}_t P_{\mathrm{data}}(\mathbf{k}) + (1 - \bar{\alpha}_t)\tilde{\Sigma}(\mathbf{k}),$$
$$\gamma(\mathbf{k}, t) = \frac{1}{\bar{\alpha}_t[P_{\mathrm{data}}(\mathbf{k})/\tilde{\Sigma}(\mathbf{k})] + (1 - \bar{\alpha}_t)}. \tag{24}$$

If S and Σ do not commute, whitening and diagonalizing $\Sigma^{-1/2} S \Sigma^{-1/2}$ gives the same expression with generalized eigenvalue τ_k replacing $P_{\mathrm{data}}(\mathbf{k})/\tilde{\Sigma}(\mathbf{k})$. These ratios set the modewise SNR $\bar{\alpha}_t \tau_k/(1-\bar{\alpha}_t)$ and the angles in Eq. 44. Matching Σ to S up to a scalar equalizes the ratios; matching their absolute scale as well removes the Gaussian terminal covariance mismatch. These are distinct statements. In the matched Gaussian case $S = \Sigma$, $Q_t = \Sigma$: the reverse SDE has drift $-\beta(t)x/2$ per positive generation increment and diffusion $\sqrt{\beta(t)}L$, while probability-flow velocity is zero. For $\Sigma = I$, all formulas reduce to standard white VP diffusion. Even in the matched Gaussian case, finite DDIM steps have the cosine contraction in Eq. 46; stationary continuous covariance does not guarantee exact finite DDIM updates.

The proposed mechanism is therefore specific: the covariance can synchronize second-order denoising clocks and improve a finite-terminal Gaussian approximation, which can help when the number of diffusion steps is reduced. The resulting quantities give testable predictions for modal timing and covariance evolution. Learned-score error, higher-order structure and guidance enter through the response corrections and empirical diagnostics. Building on colored SDEs, covariance-aware DDIM and Itô moment calculus, our analysis connects the Flicker covariance to synchronization, finite-terminal effects and measured spectral response.

Relation to nearby training and sampling designs. The non-isotropic framework of Voleti et al. (2022) already includes fixed covariance, consistent score parameterization and covariance-aware DDIM. Under a fixed output transformation, predicting the white latent noise corresponds to a whitened error on the colored prediction. Flicker Diffusion builds on these ingredients with a spatial power-law family and its matching residual metric. This design improves generation under matched optimizer-update and sampler-selection budgets, using the baseline U-Net without an auxiliary network or additional training stage. Generalized DDIM (Zhang et al., 2022) addresses numerical sampling for non-isotropic diffusions; our covariance-aware update uses that established setting.

Blurring Diffusion (Hoogeboom & Salimans, 2023) connects heat dissipation to Gaussian diffusion and combines frequency-dependent signal attenuation with additive noise. Flicker's forward signal coefficient is scalar, while a fixed spatial covariance shapes the noise throughout training. EqualSNR (Falck et al., 2025) instead sets Fourier noise variances proportional to measured data power, sharing the second-order synchronization target. Flicker uses measured dataset spectra to guide candidates within a compact spatial-kernel family, as detailed in Appendices I and I.1. This supplies a statistical basis for dataset adaptation while preserving a fixed corruption operator throughout training and sampling. Our analysis explains the resulting dynamics through generalized-mode clock dispersion, higher-order cumulant attenuation and a retained response remainder, separating Gaussian terminal mismatch from finite-step contraction. The EqualSNR-style experiments in Appendix S are budget-controlled variants with their stated covariance, schedule and prediction targets.

SAGD (Scimeca et al., 2025) directly parameterizes the noise in Fourier space using power-law weights and band-pass masks, and trains with an unweighted squared error on the added colored noise. Flicker Diffusion specifies a spatial power-law covariance kernel, guides dataset-specific candidates with measured image spectra, and uses a whitened residual metric. Both objectives have the conditional mean as their unrestricted optimum for fixed positive-definite covariance; the metric changes finite-network optimization. Our crossed covariance–metric experiment (Table 26) examines this interaction, while the spectral diagnostics measure the response learned under the selected design.

EDM (Karras et al., 2022) treats training, score-network preconditioning and sampling as separable design choices. Our spatial covariance and residual metric occupy a different design dimension from scalar noise-level preconditioning. Gaussian spectral transfer analysis (Benita et al., 2025) and optimized sampling schedules (Sabour et al., 2024) inform the finite-step perspective; the explicit

response remainder in this work also permits diagnostics beyond an exact Gaussian closure. Concurrent CNS (Davidson et al., 2026) keeps a pretrained field and dynamically reallocates stochastic noise across frequencies during inference. Flicker trains with its covariance and matching objective, then studies the associated response and sampling dynamics. Both use SDE analysis, but the intervention and quantities examined differ: inference-time spectral energy allocation in CNS, and covariance-consistent response, modal timing and finite-terminal effects here.

Linear score structure and covariance dynamics. Wang & Vastola (2023) derive Gaussian score-flow solutions and use them to replace early numerical sampling steps. Li et al. (2024) fit affine approximations to neural denoisers and relate their Gaussian structure to the generalization regime. These studies establish linear response as a useful description of trained diffusion models. Flicker Diffusion turns this descriptive perspective into a comparison of noise designs: changing the corruption covariance and residual metric improves held-out affine predictability in selected low-frequency bands, and the fitted response predicts independent reverse spectral evolution.

Wang & Pehlevan (2025) solve gradient-flow learning dynamics for linear and convolutional denoisers, connecting the order and speed of mode acquisition to data covariance. Their learning time indexes parameter optimization; our modal clocks index the corruption and sampling process of a trained model. Weitzner et al. (2025) relate repeated linear denoising to covariance estimation and power iteration, with supporting Jacobian measurements in neural denoisers. Our response analysis retains a nonlinear remainder and tests its leading affine description through independent spectral predictions.

Pierret & Galerne (2025) derive exact Gaussian reverse-SDE and probability-flow-ODE solutions and compute Wasserstein sampling errors. Together with spectral transfer analysis (Benita et al., 2025), this provides context for our separation of terminal initialization and finite-step effects. We express these effects in generalized data–noise covariance modes and use them to analyze modal timing and DDIM contraction under the chosen spatial covariance.

Other routes to low-evaluation generation. Improved DDPM (Nichol & Dhariwal, 2021) learns reverse-process variances to enable generation with fewer evaluations. Consistency models (Song et al., 2023) learn mappings from points on a probability-flow trajectory to its data endpoint, through either distillation or standalone training. Rectified flow (Liu et al., 2023) learns transport along straight interpolation paths and uses reflow to straighten learned trajectories. These approaches change the reverse parameterization or training target. Flicker changes covariance and its matching residual metric within standard diffusion training, retaining the baseline network and optimizer-update budget. It requires no teacher, distillation or reflow stage to obtain the reported sampling gains. Trajectory straightness describes how sample paths evolve in time; partial spectral linearization here describes the adequacy of an affine score response over selected modes and times. An affine field can still incur grid-dependent integration error (Appendix K).

Noise schedules, resolution and terminal SNR. Chen (2023) studies resolution-dependent noise scheduling; Relay Diffusion (Teng et al., 2024) links frequency-domain SNR to resolution and uses blurring and block noise to connect generation stages. These results motivate joint analysis of spatial covariance, scalar noise schedule and resolution when adapting a noise design. Lin et al. (2024) identifies a training–inference discrepancy from nonzero terminal SNR and proposes a zero-terminal-SNR correction. Our Gaussian analysis separates terminal-prior mismatch from finite-step DDIM contraction under the trained covariance. The matched-terminal control in Appendix U preserves a specified terminal SNR; it is distinct from enforcing zero terminal SNR.

B Effective modal linearization and its nonlinear corrections

The frequency-domain theory separates a leading modal response from an explicit remainder and tracks its effect on spectral evolution. Throughout, generation follows decreasing noise index $t : T \to 0$, $\Sigma = LL^\top \succ 0$, and reverse expectations refer to the actual generated law. A fixed class or guidance rule is used consistently; for class mixtures, expectations include the specified joint law of images and labels. The discussion concerns the reverse SDE; probability flow halves the score correction and removes the injected noise, as in Appendix A.

An exact separation of leading response and remainder. Write $g_t = \Sigma s_\theta = -\epsilon_\theta/\sqrt{1-\bar\alpha_t}$, $m_t = \mathbb{E}x_t$, and $y_t = x_t - m_t$. Given a matrix H_t, define $\rho_t(y) = g_t(m_t + y) - \mathbb{E}g_t(x_t) + H_t y$. Then $\mathbb{E}\rho_t = 0$ and

$$\mathrm{d}y_t = -[A_t y_t + \beta(t)\rho_t(y_t)]\,\mathrm{d}t + \sqrt{\beta(t)}L\,\mathrm{d}\bar w_t, \qquad A_t = \beta(t)(I/2 - H_t), \tag{25}$$

$$-K_t' = A_t K_t + K_t A_t^\top + \beta(t)\Sigma + \beta(t)(B_t + B_t^\top), \qquad B_t = \mathbb{E}[\rho_t(y_t)y_t^\top]. \tag{26}$$

These identities hold for nonlinear ensemble dynamics, with $|\mathrm{d}t| = -\mathrm{d}t$. For Fourier-diagonal H_t with multipliers $\gamma(\mathbf{k}, t)$, set $\tilde\Sigma(\mathbf{k}) = (F\Sigma F^*)_{\mathbf{kk}}$, where F is unitary. Applying Itô's rule to $|y(\mathbf{k}, t)|^2$ gives Eq. 7 for arbitrary Σ: response couplings outside the leading diagonal term remain in ρ. When the covariance and raw-score response are both Fourier diagonal, γ equals the raw-score restoring coefficient times $\tilde\Sigma(\mathbf{k})$. Regression of $\tilde\epsilon_\theta/\sqrt{1-\bar\alpha_t}$ on $\tilde x(\mathbf{k}, t)$ defines this effective coefficient directly. It is a distribution-dependent projection, distinct from a prescribed Taylor coefficient. These conventions apply also to the nonperiodic implementation.

Let $D^{(0)}$ solve the modal equation with its remainder omitted and the same initial variance at $t = T$. Using $G(\mathbf{k}; t, u)$ from Eq. 8, variation of constants gives

$$D(\mathbf{k}, t) - D^{(0)}(\mathbf{k}, t) = 2\int_t^T |G(\mathbf{k}; t, u)|^2 \beta(u)\,\mathrm{Re}\,\mathbb{E}[\rho(\mathbf{k}, u)\overline{y(\mathbf{k}, u)}]\,\mathrm{d}u, \tag{27}$$

$$|D(\mathbf{k}, t) - D^{(0)}(\mathbf{k}, t)| \le 2\int_t^T |G(\mathbf{k}; t, u)|^2 \beta(u)\sqrt{D(\mathbf{k}, u)\mathbb{E}|\rho(\mathbf{k}, u)|^2}\,\mathrm{d}u. \tag{28}$$

The bound gives a sufficient condition for accurate spectral prediction: a small propagated residual contribution. It includes amplification by the propagator; cancellation can further reduce the actual error. We assess the leading approximation using local fit quality and propagated predictions; the sufficient bound itself is not evaluated. When H_t is re-fitted by least squares under the same law at each time, residual orthogonality closes the instantaneous second-moment identity by construction, including for nonlinear fields. Held-out R^2 or residual energy quantifies the response beyond this identity; predictions across measurement laws test transfer. Independent ensembles reduce fitting bias. A diagonal-fit residual includes both nonlinear effects and omitted linear coupling, specifying the scope of this operational measure.

Exact attenuation of non-Gaussian cumulants. Let S be the centered data covariance and choose orthogonal U so that $U^\top\Sigma^{-1/2}S\Sigma^{-1/2}U = \mathrm{diag}(\tau_k)$, with $\tau_k > 0$ on the analyzed subspace. In these real generalized coordinates, $z_0 = U^\top\Sigma^{-1/2}(x_0 - \mathbb{E}x_0)$ and $z_t = \sqrt{\bar\alpha_t}z_0 + \sqrt{1-\bar\alpha_t}\xi_w$, where $\xi_w \sim \mathcal{N}(0, I)$ is independent. Write $q_k(t) = \bar\alpha_t\tau_k + (1-\bar\alpha_t)$ and $\widehat z_{t,k} = z_{t,k}/\sqrt{q_k(t)}$. For every existing cumulant of order $n \ge 3$, including repeated indices,

$$\mathrm{cum}(\widehat z_{t,k_1}, \ldots, \widehat z_{t,k_n}) = \mathrm{cum}\left(\frac{z_{0,k_1}}{\sqrt{\tau_{k_1}}}, \ldots, \frac{z_{0,k_n}}{\sqrt{\tau_{k_n}}}\right)\prod_{j=1}^n \sqrt{\frac{\bar\alpha_t\tau_{k_j}}{\bar\alpha_t\tau_{k_j} + (1-\bar\alpha_t)}}. \tag{29}$$

This follows directly from multilinearity of cumulants, independence of the innovation, and vanishing Gaussian cumulants above order two. Thus the same modal signal fractions that govern second-order clocks also govern the survival of higher-order structure. If $S = \Sigma$, all factors are $\sqrt{\bar\alpha_t}$, and every order-n standardized cumulant is attenuated by $\bar\alpha_t^{n/2}$. For $S = c\Sigma$ the attenuation is likewise common across modes. The factors synchronize the windows of non-Gaussian recovery. Their magnitudes depend on the data cumulants in the chosen generalized basis; changing Σ changes both this basis and its attenuation factors. Comparing learned-model corrections consequently requires their measured statistics as well as the covariance spectrum.

To connect cumulants to score nonlinearities, consider a differentiable weakly non-Gaussian density in the standardized coordinates. With third-order cumulants κ_{ijk}, the leading Hermite expansion gives

$$\log p(\widehat z) = \log\phi(\widehat z) + \tfrac{1}{6}\kappa_{ijk}\mathcal{H}_{ijk}(\widehat z) + \cdots, \qquad \partial_\ell \log p(\widehat z) = -\widehat z_\ell + \tfrac{1}{2}\kappa_{\ell jk}(\widehat z_j\widehat z_k - \delta_{jk}) + \cdots. \tag{30}$$

Repeated indices are summed and ϕ is the standard normal density. Differentiating the cubic Hermite polynomial supplies the factor three. Fourth and higher cumulants generate further polynomial terms. This expansion describes the score of the true forward density in regions where the density and

derivative remainders are controlled. The learned field additionally carries approximation error and classifier-free guidance contributions. Equation 29 connects these forward-score vertices to modal signal fractions, providing a perturbative description of the leading response and nonlinear structure within the stated approximation.

Finite-time MSRJD formulation. For clarity, use real image coordinates (or independent real Fourier coordinates), avoiding duplicate conjugate modes. Write the centered dynamics along decreasing t as $-\dot{y} = A_t y + N_t(y) + \zeta_t$, where $N_t = \beta(t)\rho_t$ and $\mathbb{E}[\zeta_t \zeta_u^\top] = Q_t \delta(t-u)$ with $Q_t = \beta(t)\Sigma$. Here Q_t denotes the noise variance rate, rather than the forward covariance of Appendix A. The MSRJD response functional (Martin et al., 1973; Janssen, 1976) has weight $e^{-\mathcal{S}}$,

$$\mathcal{S} = \mathcal{S}_{\text{init}} + \int_0^T \mathrm{d}t \left[\tilde{p}^\top(-\dot{y} - A_t y - N_t(y)) - \tfrac{1}{2}\tilde{p}^\top Q_t \tilde{p}\right]. \tag{31}$$

The response field $\tilde{p}$ is integrated on an imaginary contour (Täuber, 2005). Equivalently, for a real field p, the noise-averaged weight is $\exp\{i\int p^\top(-\dot{y} - Ay - N)\,\mathrm{d}t - \frac{1}{2}\int p^\top Q p\,\mathrm{d}t\}$; setting $\tilde{p} = -ip$ gives Eq. 31. The Itô Jacobian is field independent; retarded contractions at identical times obey $G^R(t,t) = 0$. The deterministic propagator has left limit $\Phi(u^-, u) = I$ because generation runs toward smaller t. A Gaussian initial ensemble supplies $\mathcal{S}_{\text{init}} = y_T^\top K_T^{-1} y_T/2$; non-Gaussian initial cumulants can be retained as boundary vertices.

For $N = 0$, $\partial_t \Phi(t,u) = -A_t \Phi(t,u)$ and the propagators are

$$G_0^R(t,u) = \mathbf{1}_{t<u}\Phi(t,u), \qquad \langle \tilde{p}\tilde{p}^\top\rangle_0 = 0, \tag{32}$$

$$G_0^K(t,u) = \Phi(t,T)K_T\Phi(u,T)^\top + \int_{\max(t,u)}^T \mathrm{d}v\, \Phi(t,v)Q_v\Phi(u,v)^\top. \tag{33}$$

Differentiating $G_0^K(t,t)$ gives $-K' = AK + KA^\top + Q$, the same Itô moment equation. The boundary term retains the finite-horizon initial covariance. Extending the initial boundary to the infinite past along generation requires a stable stationary approximation.

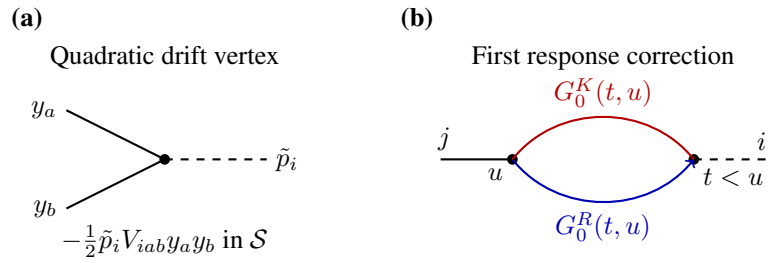


Figure 6: Nonlinear corrections extend the leading modal theory. (a) A quadratic drift term gives one response-field leg and two state-field legs. (b) Two such vertices contribute a causal response line G_0^R and a fluctuation line G_0^K to $\Sigma^{R,(2)}$ in Eq. 34. The schematic displays the contraction structure; summed mode indices, factorials and drift covariance factors are specified in the equations. It does not assume these corrections vanish.

Vertices and the first diagrammatic corrections. For the diagrammatic expansion, absorb the constant and linear Taylor pieces into the background and A (or retain them as explicit counterterms). The nonlinear polynomial terms, with symmetric incoming indices, are $N_i = \frac{1}{2}V_{iab}y_a y_b + \frac{1}{6}W_{iabc}y_a y_b y_c + \cdots$. Deterministic mean counterterms are understood when enforcing $\mathbb{E}y = 0$; their induced changes to A must also be retained. Here V and W are derivatives of the *drift*, including $\beta(t)\Sigma$ acting on the output of raw-score derivatives. Separating mean diagrams, Gaussian Wick contractions give the quadratic-vertex response and noise kernels

$$\Sigma^{R,(2)}_{ij}(t,u) = V_{iab}(t)G^R_{0,ac}(t,u)G^K_{0,bd}(t,u)V_{cdj}(u), \tag{34}$$

$$\Pi^{(2)}_{ij}(t,u) = \tfrac{1}{2}V_{iab}(t)G^K_{0,ac}(t,u)G^K_{0,bd}(t,u)V_{jcd}(u), \tag{35}$$

$$\Sigma^{R,(3)}_{ij}(t,u) = \tfrac{1}{2}W_{ijab}(t)G^K_{0,ab}(t,t)\delta(t-u). \tag{36}$$

The factorial convention is essential: the $1/2!$ at each quadratic vertex cancels the response-diagram leg choices; two identical correlation lines leave $1/2$ in Π. In the convention $N_i = \sum_{ab} g_{iab}y_a y_b$, one has $V = 2g$. The general vertex product retains the displayed tensor contractions; a modulus-square reduction requires additional symmetries. On a periodic lattice, translation-equivariant polynomial coefficients about a translation-invariant background enforce wavevector conservation modulo the lattice. Expansions about general image-dependent backgrounds retain the full coupling tensors.

Our sign convention is $(G^R)^{-1} = (G^R_0)^{-1} - \Sigma^R$ (Hertz et al., 2016). Correlation corrections involve both the dressed response and the additional noise kernel, schematically $G^K = G^R \star (M_0 + \Pi) \star (G^R)^\top$, where M_0 includes $Q_t\delta(t-u)$ and the initial covariance boundary term and $\star$ denotes time integration. The kernels retain temporal memory and correlation corrections; a scalar zero-frequency drift reduction requires an additional approximation to both.

As a sign and normalization check, freeze stable diagonal drifts $A_k = -a_k < 0$ and noise variance rates $q_k > 0$. For elapsed generation lag $v = u - t$, use the transform $f(\omega) = \int e^{i\omega v} f(v)\,\mathrm{d}v$, $G^R_{0,k}(\omega) = (a_k - i\omega)^{-1}$ and $G^K_{0,k}(\omega) = q_k/(a_k^2 + \omega^2)$. Hence

$$\int \frac{\mathrm{d}\Omega}{2\pi} \frac{q_p}{\Omega^2 + a_p^2} \frac{1}{a_q - i(\omega - \Omega)} = \frac{q_p}{2a_p(a_p + a_q - i\omega)}. \tag{37}$$

At zero frequency it is positive. The sign of a full correction still depends on the vertices: a scalar quadratic drift gives a positive connected response kernel, whereas $N = -uy^3/6$ gives the negative tadpole $-uG^K_0(t,t)/2$. The full tensor contractions determine the sign for a neural score. Colored diffusion has $q_p = \beta\tilde{\Sigma}(\mathbf{p})$, with the frozen β and Σ also included in the drift vertices.

Interpretation for Flicker Diffusion. Covariance design acts on the free response, fluctuation spectrum and learned vertices. The cumulant identity connects covariance ratios to higher-order attenuation, the residual equation quantifies spectral correction, and the MSRJD expansion organizes the remaining mode coupling. Together they define the mechanism tested by the spectral diagnostics:

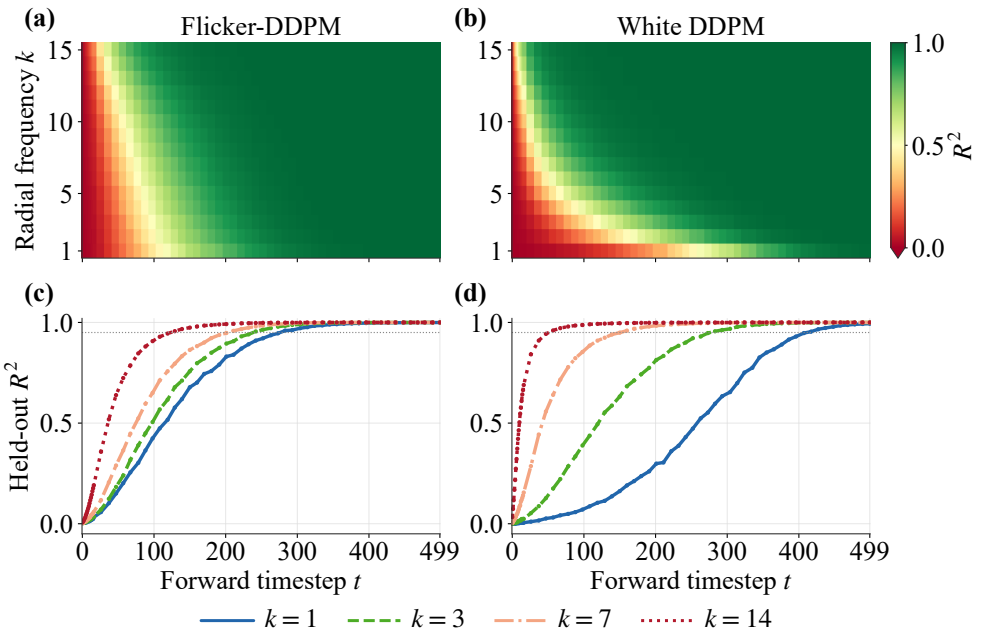


Figure 7: **Frequency-dependent affine predictability for one historical $T = 500$ CIFAR-10 checkpoint pair.** Held-out real-scalar modal R^2 uses 64 forward noise indices, each model's own corruption, true labels and CFG $w = 1.8$. Flicker improves the reported low-frequency shells and narrows the spread across four shell means; white has higher R^2 at the reported high-frequency shells. Increasing t means stronger forward noise; generation runs from right to left. Heatmap colors run from red (low R^2) through yellow to green (high R^2).

more synchronized evolution and a more useful leading response over selected times and frequencies. Higher-order cumulants, vertices and loops remain explicit quantities to evaluate after matching the two-point spectrum. This retains the nonlinear structure needed for detailed image synthesis within the same theoretical description.

C Spectral response and independent prediction

The $T = 500$ CIFAR diagnostic connects improved low-frequency response to more accurate spectral prediction. We evaluate one historical checkpoint pair using matched corruption, true labels, held-out images and independent generated trajectories. The shells $k = 1, 3, 7, 14$ follow the original figures. These mechanism measurements complement the separate three-training-seed FID comparisons.

Models and measurement protocol. Both historical checkpoints strictly load into the same preserved U-Net with 47,700,611 parameters: base width 128, multipliers $(1, 2, 2, 2)$, two residual blocks per level, attention in the historical down-path blocks and none in the up path. Their $T = 500$ schedule has linearly spaced β_t from 10^{-4} to 0.028. Flicker uses the dense Cholesky operator at $\eta = 0.2$. The historical CFG convention is $\epsilon_g = 2.8\epsilon_{\text{cond}} - 1.8\epsilon_{\text{uncond}}$. Checkpoint, source, data, selection and protocol hashes are frozen before measurement, with strict state-dictionary loading into the preserved architecture. The current experiment has a complete execution manifest; Appendix D documents the earlier figure provenance.

Forward-state regression and its validation baseline. We select 160 calibration and 160 disjoint CIFAR test images, with 16 images per class in each split, and reuse their identities across both model arms. Forward anchors are $t = 0, \ldots, 15$, followed by 48 rounded uniformly spaced indices from 16 to 499 (64 distinct measurements). Each arm's matched measurement uses its own Gaussian

Table 3: Held-out real-scalar modal R^2, averaged over time. One historical checkpoint pair. Reverse-state rows use 500 measured states along independent calibration/validation trajectories.

State / prediction	Noise	$k=1$	$k=3$	$k=7$	$k=14$	Range
Forward / guided	Flicker	.7452	.7821	.8267	.9075	.1622
	White	.4901	.7355	.8895	.9678	.4777
Forward / conditional	Flicker	.7614	.7894	.8304	.9083	.1470
	White	.5147	.7430	.8918	.9680	.4533
Reverse / guided	Flicker	.7378	.7810	.8171	.9017	.1639
	White	.4845	.7300	.8859	.9675	.4830

corruption and the true image labels. Fourier transforms are unitary; coefficients are pooled across images and RGB channels. For each frequency coordinate, a real slope $a(\mathbf{k}, t)$ and complex intercept $c(\mathbf{k}, t)$ minimize the calibration error in $\tilde{\epsilon}_\theta(\mathbf{k}, t) \simeq a(\mathbf{k}, t)\tilde{x}(\mathbf{k}, t) + c(\mathbf{k}, t)$. Writing $\bar{\epsilon}_{\text{fit}}(\mathbf{k}, t)$ for the calibration mean, we report Eq. 9 in the main text.

The constant baseline is learned on the calibration split and kept fixed during validation. Negative values are retained. This diagnostic measures the adequacy of a real-scalar leading response; complex rotations and off-diagonal linear coupling also contribute to its residual. Frequency shell k averages coordinates satisfying $k \le \sqrt{k_x^2 + k_y^2} < k + 1$. Time means use trapezoidal integration over the actual anchor indices divided by 499. The reported spread is the maximum minus minimum of the four shell means.

Low-frequency improvement and modal balance. The improvement at $k = 1, 3$ and the smaller four-shell spread persist in both forward and generated-state measurements (Table 3). Conditional forward predictions without CFG have the same direction. The effect is frequency-dependent: white has higher R^2 at $k = 7, 14$. Summing validation residual and baseline energies over all non-DC modes and integrating over time gives $R^2 = 0.8398$ for Flicker and 0.9463 for white. High frequencies can dominate this different weighting, whereas the shell-wise comparison resolves the improvement in the slower low-frequency response.

Corruption and label controls. We also evaluate white versus own corruption, crossed with artificial block labels versus true labels, retaining conditional and guided outputs. Block labels assign classes by sample position. Their full 320-image in-sample calculation is balanced, while the 160/160 split assigns classes 1–5 to calibration and 6–10 to validation. Thus the held-out block-label control includes a class-distribution shift; the principal true-label measurement is balanced in both splits. Reconstructed legacy-input white in-sample shell means $(0.498, 0.734, 0.891, 0.967)$ are close to the recovered original arrays $(0.504, 0.741, 0.891, 0.968)$, connecting the two diagnostic protocols.

Independent reverse generation and propagated predictions. We run 256 independent 500-step ancestral trajectories per model, split into 128 calibration and 128 validation paths, with cyclic class labels and no clipping. All 500 generated states are measured. The forward-fitted effective response is $\gamma(\mathbf{k}, t) = a(\mathbf{k}, t)/\sqrt{1 - \bar{\alpha}_t}$; it describes the covariance-weighted response Σs_θ. Between the 64 forward anchors, γ is linearly interpolated in t. A separate fit from the 128 reverse calibration paths supplies a comparison under the generated-state law. Both coefficient sets are fixed before evaluating the validation paths.

Starting from the measured validation initial variance, the leading discrete ancestral recursion is

$$D_{n+1}^{(0)}(\mathbf{k}) = \frac{(1 - \beta_t \gamma(\mathbf{k}, t))^2}{1 - \beta_t} D_n^{(0)}(\mathbf{k}) + \mathbf{1}_{t>0} \beta_t \left(1 - \frac{1}{384}\right) \tilde{\Sigma}(\mathbf{k}), \qquad t = 499 - n. \tag{38}$$

Here n counts completed reverse updates, and $\tilde{\Sigma}(\mathbf{k}) = (F \Sigma F^*)_{\mathbf{kk}}$ is evaluated from the actual dense Cholesky factor, including its nonperiodic boundary structure. The factor $1 - 1/384$ accounts for centering the fresh independent innovations over 128×3 image-channel coefficients with denominator 384. The final update is deterministic. For the specified affine update, Eq. 38 is the exact conditional expectation of the next centered sample variance; realized innovations introduce

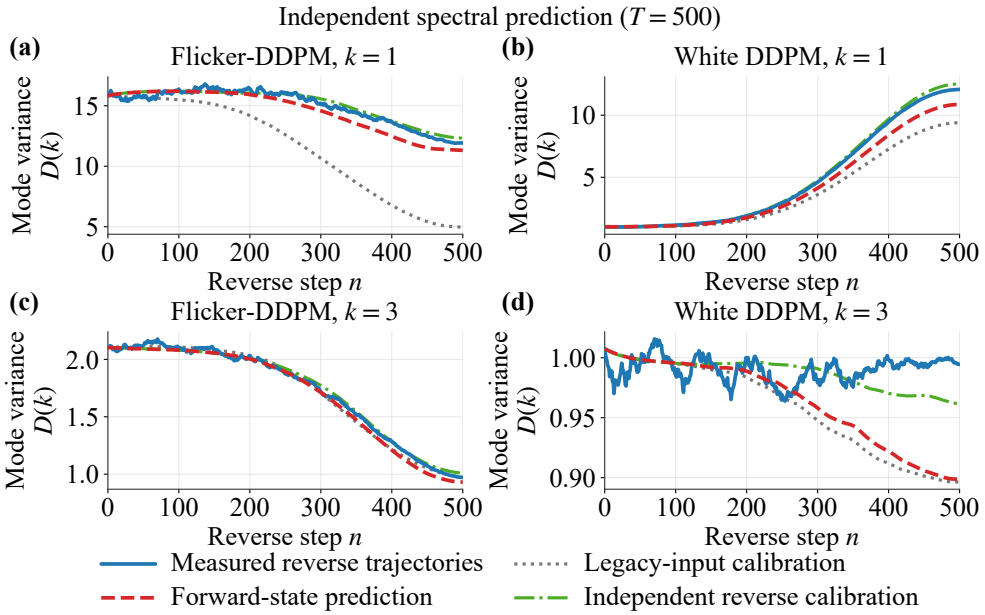


Figure 8: Propagated variance on independent generated paths at the original low-frequency shells. Each prediction starts from the same measured validation initial variance. Solid blue curves show measurements; dashed red curves transfer corrected forward-state coefficients. Green dash-dotted curves use independent reverse-state calibration; dotted gray curves use legacy-input forward calibration. Curves propagate the leading discrete moment dynamics of Eq. 38.

Table 4: Corrected forward-calibrated prediction of the independent reverse spectrum. Trajectory error is $\sum_{n=0}^{500} |D_n^{(0)}(\mathbf{k}) - D_n(\mathbf{k})| / \sum_{n=0}^{500} D_n(\mathbf{k})$; endpoint error is $|D_{500}^{(0)}(\mathbf{k}) - D_{500}(\mathbf{k})| / D_{500}(\mathbf{k})$. All entries are percentages; smaller is better.

Metric	Noise	$k = 1$	$k = 3$	$k = 7$	$k = 14$
Trajectory error	Flicker	3.763	1.980	1.937	.981
	White	9.950	3.350	1.047	.542
Endpoint error	Flicker	5.166	4.161	12.955	3.507
	White	9.962	9.639	9.944	3.329

sampling fluctuations. It omits the learned response residual. We use this discrete recursion for the principal comparison and record Euler integration of the continuous moment equation separately.

Flicker has lower prediction error at $k = 1, 3$, while white has lower errors at $k = 7, 14$. The low-frequency advantage therefore appears in both held-out response fits and forward-to-reverse spectral prediction.

Interpretation and scope. The measurements characterize one checkpoint pair; training-seed uncertainty requires replicated training. Calibration-law effects, omitted linear mode coupling and nonlinear response can all contribute to the prediction residual. Same-law least squares enforces second-moment orthogonality, so reverse-calibrated moment agreement alone cannot identify a small nonlinear residual. The held-out fits and forward-to-reverse transfer provide complementary tests of the leading approximation across the calibration and generation distributions.

D ORIGINAL SPECTRAL DIAGNOSTICS AND THEIR INTERPRETATION

The original spectral figures visualize the mechanism behind partial linearization: a more balanced response across modes and a leading prediction of their evolving power. They motivate the matched-input validation in Section 4.6. The original R^2 fits one real scalar slope and a complex intercept at each Fourier lattice point, pooling images and RGB channels; radial averaging is performed afterward. The horizontal coordinate t runs from nearly clean inputs to high noise. It has the opposite direction to the reverse-step coordinate n in the spectral prediction plot.

Archive and quantitative use. The archived implementation measures white-corrupted inputs with synthetic class labels. A historical white-run log and numerical arrays survive, but the complete execution manifest linking both image panels to their checkpoints was not preserved. The recovered arrays also differ slightly from the original table, and the spectral-prediction panels cannot be bound exactly to those arrays. We retain the original figures as exploratory illustrations; the current quantitative evidence comes from the fully recorded, matched-input reproduction in Appendix C.

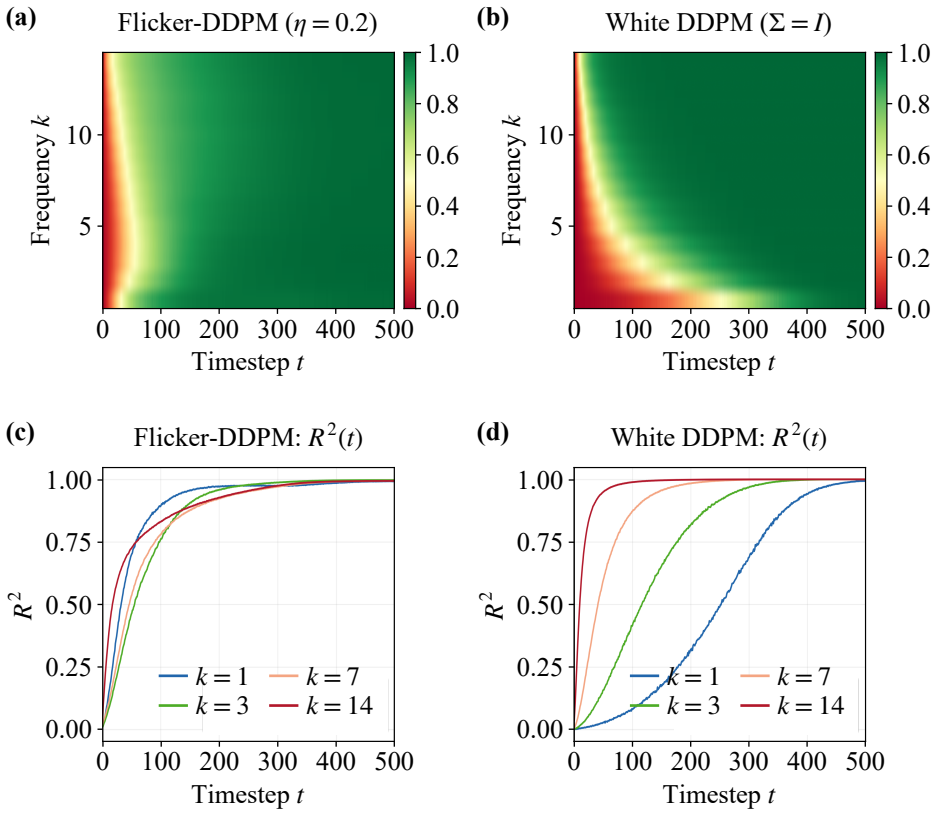


Figure 9: **Original exploratory evidence for partial modal linearization.** Top: radial explained-variance maps at $T = 500$. Bottom: selected frequency shells. Flicker shows stronger low-frequency affine predictability and a more balanced frequency dependence, while both models have poorer fits near the data end.

The surviving numerical archive has the same qualitative pattern. For $k = 1, 3, 7, 14$, its unweighted 500-index mean in-sample R^2 is $(0.8954, 0.8558, 0.8506, 0.8916)$ for Flicker and $(0.5045, 0.7414, 0.8913, 0.9678)$ for white. The matched-input validation in Fig. 7 resolves this pattern into a low-frequency improvement and a smaller spread across modes.

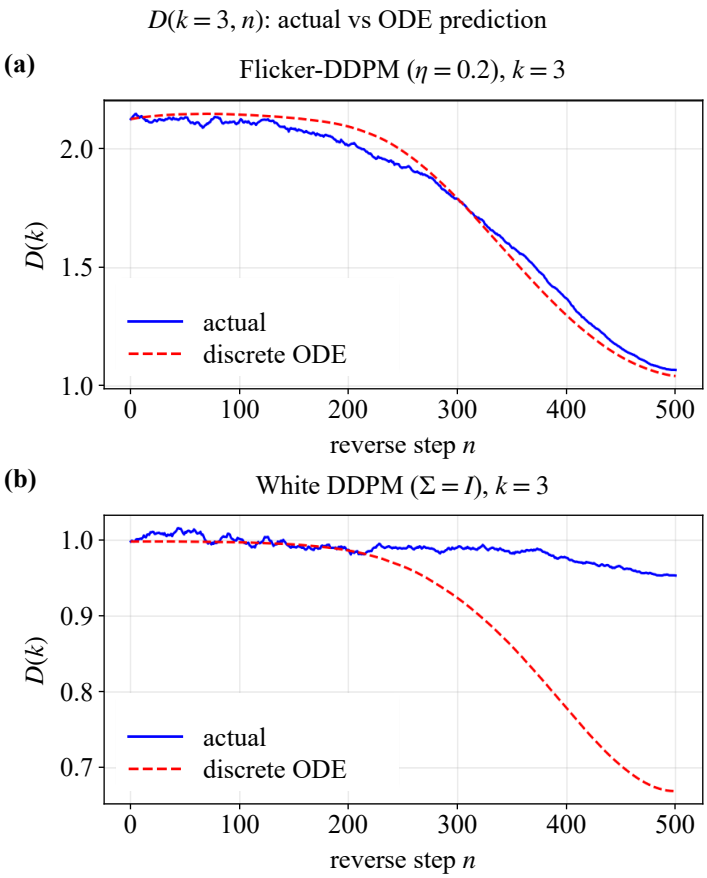


Figure 10: **Leading spectral prediction and its remainder.** Original $k = 3$ measurements and linear predictions during generation, from $n = 0$ to $n = 500$. The contrasting trajectories motivate the independent spectral-transfer test in Appendix C.

The prediction coefficient measured from $\tilde{\epsilon}_\theta/\sqrt{1-\bar{\alpha}_t}$ is the effective covariance-weighted response $\gamma(\mathbf{k}, t)$. The original prediction also transfers a coefficient estimated on forward-corrupted real images to free reverse trajectories. Differences therefore include distribution transfer, diagonal-response approximation and finite-step effects as well as nonlinear corrections. The matched protocol assesses this transfer through held-out response measurements and independent forward-to-reverse prediction.

E NOISE AND SPECTRAL TRAJECTORY ILLUSTRATIONS

The original noise realization and historical $T = 500$ spectral trajectories illustrate the covariance design. Quantitative quality comparisons and held-out response measurements use their separately specified protocols.

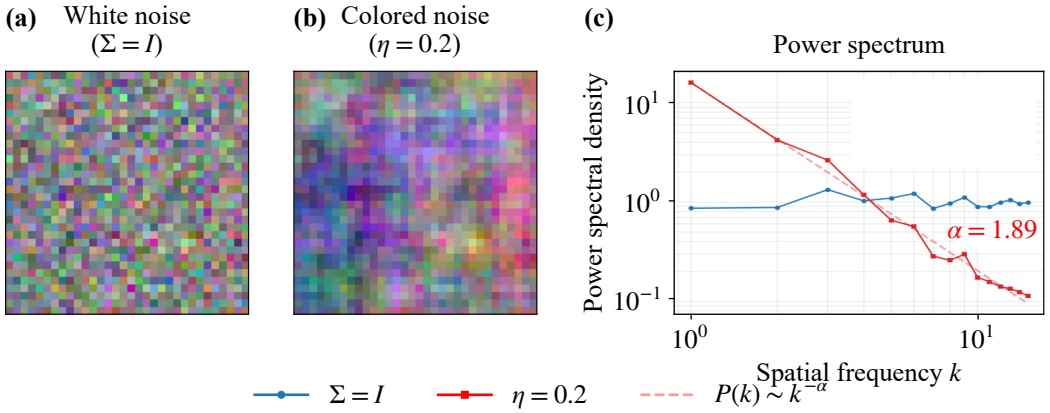


Figure 11: **Spatial correlation reallocates noise across frequencies.** 32×32 noise illustrations and radial spectra: white noise (left) and Flicker noise at $\eta = 0.2$ (middle). The fitted $\alpha = 1.89$ describes this illustrative noise realization. The spatial covariance and training convention are specified in Eqs. 1–2.

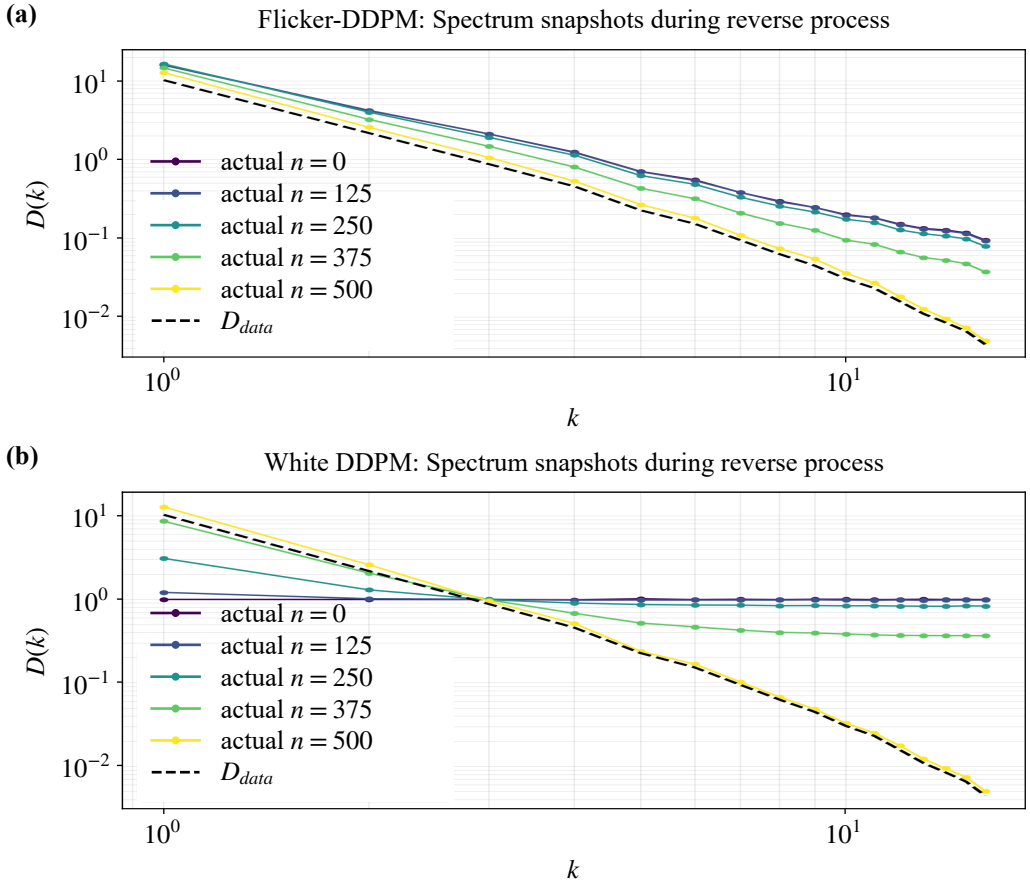


Figure 12: **Covariance shapes the spectral generation trajectory.** Original $T = 500$ CIFAR trajectories: Flicker (top) and white noise (bottom). Generation runs from $n = 0$ to $n = 500$; dashed black curves show the data spectra. Flicker starts with a frequency-dependent power profile and evolves toward the data's spectral shape. Section 4.6 evaluates spectral dynamics using independent measurements.

F ADDITIONAL EVIDENCE ON THE MAIN CIFAR MODELS

These measurements use the six final EMA checkpoints from the main $T = 150$ experiment: three registered training seeds, 90,000 updates and global batch 128. Sampler confirmation, image metrics and field diagnostics answer different questions and are reported separately. The initialization exception is stated in Appendix H.1.

F.1 DPM++ WITH SYMMETRIC CANDIDATE SELECTION

Each checkpoint receives the same 15 grid/guidance candidates, with 2,000 pilot images per candidate, followed by an independently seeded 50,000-image confirmation. The three grids are uniform, quadratic and log-SNR; guidance weights are $\{0.5, 1, 1.5, 2, 3\}$. All evaluations use actual NFE40 and DPM-Solver++2M, with dynamic thresholding disabled for both noise types. All 90 pilot and six confirmation outcomes are retained. This is a follow-up search using already trained checkpoints.

Table 5: DPM++ confirmation on the main $T = 150$ checkpoints. FID-50k uses the same 50,000-image real reference; SD is across training seeds.

Seed	White FID	Flicker FID	White grid/w	Flicker grid/w
2026091211	20.445	8.694	uniform/2	logsnr/1.5
2026091212	21.344	8.489	logsnr/3	logsnr/1.5
2026091213	21.612	8.639	logsnr/3	logsnr/1.5
Mean $\pm$ SD	21.133 ± 0.611	8.607 ± 0.106		

All three pairs favor Flicker. Two white selections reach the $w = 3$ search boundary, so this is a finite-candidate comparison. Within the tested candidate sets, DDIM achieves lower absolute FID for both noise types, while both solvers favor Flicker at NFE40.

F.2 DISTRIBUTION METRICS ON THE MAIN DDIM SAMPLES

We reuse all six sets of 50,000 main DDIM PNGs. KID (Bińkowski et al., 2018) uses 100 subsets of 1,000 samples; precision, recall, density and coverage use 50,000 samples with $k = 5$ (Naeem et al., 2020). The measures use Inception features and are additional summaries of the same models and images, not new training replications. Table 6 reports training-seed SD, not the variability among KID subsets.

Table 6: Additional metrics for the main DDIM comparison. All reported uncertainty is sample SD across the three registered training seeds.

Metric	White	Flicker
KID $\times 1000 \downarrow$	5.3123 ± 0.2116	2.0165 ± 0.0336
Precision $\uparrow$	0.7658 ± 0.0036	0.7688 ± 0.0020
Recall $\uparrow$	0.4516 ± 0.0057	0.6036 ± 0.0017
Coverage $\uparrow$	0.7510 ± 0.0047	0.8476 ± 0.0010
Density $\uparrow$	0.9954 ± 0.0140	1.0535 ± 0.0089

KID, recall and coverage improve in every training-seed pair; mean precision is similar and does not improve in every pair. Together these measurements support improved coverage in the shared feature space. An initial metric run stopped at a real-feature mean guard. The revised run recomputed real features and preserved the old FID outputs. The fresh-real versus cached-real FID difference was about 9.8×10^{-5}; reproducing the original FID from its original reference differed by at most 9.7×10^{-13}. Only the completed revised metric run is reported; the cause of the cache discrepancy remains unresolved.

F.3 HELD-OUT RESPONSE ON THE MAIN THREE-SEED MODELS

At each of all 150 forward indices, 250 CIFAR test images (25 per class) calibrate the real-scalar slope and complex intercept, and another 250 disjoint images validate them. Images and standard

Gaussian innovations are shared across arms; each arm applies its actual training covariance. True class labels are used. The primary field is common CFG $w = 1$; conditional predictions and common $w = 1.5$ are prespecified additional fields. These fields are formed from the same conditional and unconditional outputs, not independent experiments.

The regression and constant validation baseline follow Eq. 9. The four shells $k = 1, 3, 7, 14$ were fixed before measurement. Each shell's time mean weights all 150 integer indices equally; this differs from the historical 64-anchor trapezoidal average. The global energy-weighted score pools residual and constant baseline energies over time and all non-DC modes. Negative scores are retained. Training-seed uncertainty is separate from images, channels, modes and time indices.

Table 7: Main-model response, common $w = 1$. Seed suffixes 11/12/13 refer to 2026091211/12/13. Global denotes the pooled non-DC energy-weighted R^2. The full four-shell range and global scores retain reverse trends.

Noise	Seed	$k = 1$	$k = 3$	$k = 7$	$k = 14$	Range	Global
White	11	0.1844	0.5214	0.7965	0.9396	0.7552	0.9018
White	12	0.1859	0.5216	0.7965	0.9395	0.7536	0.9019
White	13	0.1873	0.5226	0.7965	0.9395	0.7523	0.9019
White	SD	0.00146	0.00061	0.00004	0.00005	0.00148	0.00003
Flicker	11	0.5485	0.6063	0.6858	0.8328	0.2844	0.6947
Flicker	12	0.5449	0.6048	0.6851	0.8324	0.2875	0.6931
Flicker	13	0.5456	0.6045	0.6846	0.8323	0.2867	0.6931
Flicker	SD	0.00190	0.00097	0.00062	0.00029	0.00162	0.00092

Table 8: Additional shared-field comparisons on the same main models. Means over the three training seeds; these are not extra replications.

Field	Noise	$k = 1$	$k = 3$	$k = 7$	$k = 14$	Range	Global
Conditional	White	0.1920	0.5285	0.7984	0.9397	0.7478	0.9033
Conditional	Flicker	0.5564	0.6108	0.6881	0.8332	0.2768	0.7001
$w = 1.5$	White	0.1810	0.5169	0.7951	0.9394	0.7584	0.9008
$w = 1.5$	Flicker	0.5385	0.6009	0.6830	0.8319	0.2935	0.6886

The low-frequency improvement and smaller four-shell range coexist with higher high-frequency and global energy-weighted scores for white. The forward-state diagnostic establishes this partial-response pattern on the main three-seed models. The complementary reverse-prediction experiment uses $T = 500$ and a separate historical checkpoint pair. The diagonal-fit residual combines nonlinear response and omitted linear cross-mode coupling.

G Sampling operating points and $T = 500$ controls

Low-evaluation operating points. The batch-64 study in Fig. 3 uses 90,000 training updates and three paired training seeds. Flicker-DDIM uses $T = 150$, uniform spacing and $w = 1$; white DDIM uses $T = 500$, quadratic spacing and $w = 0.5$. FID-50k at NFE20 for Flicker is 6.289 ± 0.15, compared with 6.292 ± 0.08 at NFE100 for white. The comparison therefore uses one fifth as many U-Net evaluations at approximately the same observed mean FID for the two specified configurations. The NFE20 extension was selected after inspection of the initial anchors; paired FID differences are $+0.233, -0.184, -0.060$. This is an operating-point comparison rather than a formal noninferiority test or a speedup over a globally optimized white frontier. The curves use their own diffusion-step counts T, terminal SNRs and sampler settings. The later batch-128 confirmations are reported separately. In this curve study, dynamic thresholding is used only for white DPM++.

Symmetric selection at $T = 500$. At $T = 500$, white, dense Flicker, periodic Flicker and GFF each use three training seeds and identical training budgets. Each solver gets eight candidates: uniform/quadratic grids crossed with $w \in \{0.5, 1, 1.5, 2\}$, with dynamic thresholding off. FID-2k selection precedes 24 fresh FID-50k confirmations, all at actual NFE40. Table 9 gives the complete

results. Dense Flicker improves over the periodic and GFF implementations, while white has the best selected FID at $T = 500$.

Table 9: CIFAR-10 at $T = 500$, actual NFE40, after equal eight-candidate selection. Mean $\pm$ sample SD across three training seeds. GFF is our budget-controlled implementation.

Noise	DDIM	DPM++
White	$\mathbf{7.642 \pm 0.105}$	$\mathbf{6.017 \pm 0.055}$
Dense Flicker	8.171 ± 0.102	6.834 ± 0.107
Periodic Flicker	9.056 ± 0.083	7.829 ± 0.131
GFF control	9.709 ± 0.750	8.353 ± 0.667

All confirmations select quadratic spacing. At shared uniform/$w = 1$ settings, dense Flicker improves DDIM FID from 12.107 to 10.230 and DPM++ from 9.242 to 7.793. Together with the selected results, this shows the interaction between covariance, time grid and guidance. The low-NFE comparison uses equally selected baselines in the main $T = 150$ experiment and explicitly specified operating points in the batch-64 curves. Changing training T changes terminal SNR when beta endpoints are fixed; diffusion-step count T and inference NFE are reported separately throughout.

H Expanded sampler search and terminal-prior confirmations

H.1 Expanded symmetric guidance confirmation

For each of the six $T = 150$ EMA checkpoints, we evaluate the same 15 DDIM candidates: uniform, quadratic, and uniform log-SNR grids, each with $w \in \{0.5, 1, 1.5, 2, 3\}$. Each candidate uses 2,000 generated images and actual NFE40. This is an adaptive follow-up to an earlier grid search: the six native-grid candidates at $w \in \{0.5, 1\}$ are reused with their original samples, and nine candidates per model are added. Both methods therefore have the same final candidate set. We minimize pilot FID, breaking ties by the fixed candidate index. Both selections within each training-seed pair are frozen before either fresh 50,000-image confirmation begins. The pilot generation seed is 114800000; the independent confirmation seed is 115100000. All white models select uniform/$w = 1.5$, and all Flicker models select uniform/$w = 1$; neither method selects the upper $w = 3$ boundary. The three training seeds are 2026091211, 2026091212, and 2026091213. Pairing refers to these registered seeds, architecture and training schedule. Recorded initialization hashes agree for the first and third pairs but differ for the second; its cause is unresolved. We retain all three registered pairs, including this initialization difference. The extension reuses the trained checkpoints and keeps the search budget identical across noise types. Controls at $T = 500$ are reported in Appendix G.

H.2 Independent terminal-covariance intervention confirmation

We additionally test a fixed prior intervention at uniform/$w = 1$, the setting selected for all six $T = 150$ models by the earlier conventional-grid CFG pilot. This setting was frozen before the guidance extension above was evaluated; there is no prior-specific grid or guidance search. The existing terminal covariance maps are reused byte for byte. They replace only the initial zero-mean covariance Σ by $\bar{\alpha}_T S + (1 - \bar{\alpha}_T)\Sigma$, leaving the checkpoint, time grid, DDIM update, and actual NFE40 unchanged. Here S is the empirical spatial training covariance averaged across RGB, not a class-conditional or full joint RGB model of the non-Gaussian terminal distribution.

The new generation seed 115200000 is independent of both the earlier exploration and the guidance confirmation. The same per-image latent innovations and labels are paired across all 12 cells. Every cell uses 50,000 images. GPU preflight checks compare all 20 intermediate states of the original and wrapped samplers and confirm bitwise equality when their initial tensors agree.

The calibrated-minus-standard changes are 23.303 ± 0.503 for white and 1.547 ± 0.180 for Flicker. The change from the earlier quadratic/$w = 0.5$ result demonstrates a strong interaction between initialization and the selected sampler. The Gaussian calculation isolates the second-order initialization effect; the learned guided sampler additionally depends on its full terminal distribution and response field.

Table 10: Independent confirmation after the symmetric 15-candidate search. All entries are FID-50k against the same 50,000 CIFAR-10 training images, with ordinary U-Nets, $T = 150$, and actual NFE40. Mean $\pm$ sample standard deviation is across training seeds. Each seed uses its own pilot-selected candidate, as described above.

Training seed	White	Flicker	Flicker − white
2026091211	14.0144	6.4882	−7.5262
2026091212	14.6234	6.3494	−8.2740
2026091213	14.6967	6.4042	−8.2925
Mean ± SD	14.445 ± 0.375	6.414 ± 0.070	—

Table 11: All fixed uniform/$w = 1$ terminal-prior confirmations at actual NFE40. Covariance calibration worsens FID for all six checkpoints. "Standard" denotes the training noise prior; "calibrated" changes only its initial second moment. SD is across the three training seeds.

	White		Flicker	
Training seed	Standard	Calibrated	Standard	Calibrated
2026091211	13.9441	37.7613	6.4887	8.1148
2026091212	14.7334	37.5447	6.4143	8.0893
2026091213	14.8326	38.1128	6.4317	7.7730
Mean	14.503	37.806	6.445	7.992
SD	0.487	0.287	0.039	0.190

I 64×64 DATASETS: PROTOCOLS AND ADDITIONAL RESULTS

Data and preprocessing. We use CelebA (Liu et al., 2015) and the church category of LSUN (Yu et al., 2015), resized to 64×64. The local CelebA pool is the 202,599-image `nielsr/CelebA-faces` train mirror. The Church pool contains 126,227 images from the train and test portions of `tglcourse/lsun_church_train`, concatenated by the original data preparation protocol. These are the pools used for both model training and FID reference selection, not held-out test sets. CelebA images receive a 140×140 center crop; Church images receive the largest square center crop. Both are resized to 64×64 with the torchvision/PIL bilinear resize, converted to RGB and normalized to $[-1, 1]$. Training includes random horizontal flips; reference construction uses the same crop and resize without random augmentation.

Matched training protocol. For the primary replications at $T = 150$, the unconditional U-Net has base width 128, channel multipliers $(1, 2, 3, 4)$, three residual blocks per level, attention at level 2 (the 16×16 feature map), dropout 0.1, and 108,596,355 parameters. Each model is trained from scratch for 90,000 optimizer updates with global batch 128, learning rate 10^{-4}, constant learning rate, zero weight decay, gradient clipping at 1, and EMA decay 0.9999. Both noise conditions use BF16 mixed precision. The linear schedule has $T = 150$, $\beta_1 = 10^{-4}$ and $\beta_T = 0.028$; its terminal cumulative signal coefficient is $\bar{\alpha}_T = 0.1191326623$. Corresponding white/Flicker runs share the training seed and initialization. The seeds are 202701 and 202702. This protocol pairs architecture, optimization budget and initialization.

Covariance and numerical backend. These experiments use the periodic version of the shifted power-law kernel, with $\eta = 0.0873896715$ for CelebA and $\eta = 0.2538358453$ for Church, inherited from the dataset-specific spectral analysis before the new-seed replications. The covariance is applied independently to each RGB channel using the original real-matrix DFT implementation. The network predicts colored noise, and the residual is whitened before taking mean squared error. White noise uses the identity operator. Thus whitening is present in the high-resolution training objective. The recorded training implementation includes the DFT matrix products inside the outer BF16 autocast region. A later isolated CUDA check found approximately 1–1.5% relative coloring/whitening round-trip or gradient deviations from the FP32 operator for these settings. The results use these

Table 12: **Primary $T = 150$ replications.** Actual NFE150, ancestral sampling, 90k optimizer updates, 50k generated/50k reference images. The gain is white FID minus Flicker FID. Mean rows average the two training seeds, using the full-precision values before rounding.

Dataset	Training seed	White	Flicker	Gain
CelebA64	202701	81.651	27.566	54.085
	202702	77.595	27.087	50.508
	Mean	79.623	27.327	52.297
LSUN Church64	202701	45.690	8.708	36.981
	202702	47.198	8.430	38.768
	Mean	46.444	8.569	37.875

original mixed-precision checkpoints. A separate FP32 operator guard is available for future training; its effect on FID has yet to be evaluated.

Evaluation and paired evidence. We evaluate the 90k-update EMA checkpoint using the original unconditional ancestral sampler. Each of its 150 steps makes one network call; there is no classifier-free guidance or per-step clipping. The fixed-large innovation covariance is $\beta_t \Sigma$ for nonfinal steps; the final step adds no innovation. Only the final output is clamped and quantized to PNG. Evaluation is FP32 with autocast and TF32 disabled. FID uses pytorch-fid Inception pool3 features (dimension 2048), 50,000 generated images and 50,000 reference images per dataset. Reference and initial-sampling seed are 20260715; reverse innovations use root seed 0 with fixed block-based streams. Evaluation random numbers are shared across conditions, so the two replications are independent training seeds, not independent evaluation-seed replications. The table reports each paired outcome and the two-seed mean.

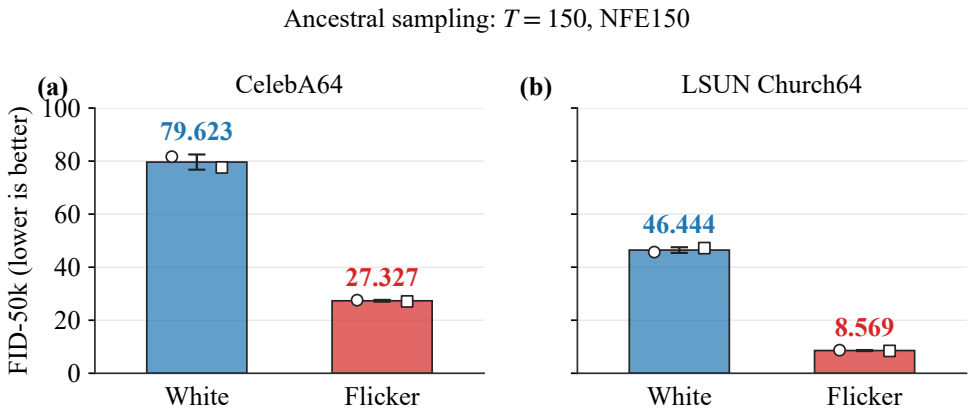


Figure 13: **Replication at $T = 150$ on 64×64 images.** Bars show means, error bars show sample SDs, and matching point shapes identify two paired training seeds. All models use $T = 150$, 90,000 optimizer updates, and ancestral DDPM sampling with actual NFE150; colored-noise results use Flicker-DDPM. FID uses 50k generated images and the same 50k-image reference per dataset. Labels give means over the two training seeds; individual values appear in Table 12. Evaluation random streams are shared across the comparisons. Lower FID is better.

Diffusion-step count and sampler controls. Tables 13 and 15 report the historical training-seed-42 controls. Each of these evaluations uses *10k generated and 10k reference* images, so its absolute FID should not be compared numerically with the 50k/50k primary replications. Within this historical protocol, Flicker improves all $T = 100, 150, 200$ ancestral settings on both datasets. At $T = 500$, white has lower FID under both ancestral and respaced DDIM sampling. Thus the transferred benefit is concentrated at $T = 100, 150, 200$. The NFE40 follow-up in Table 14 uses new 50k/50k evaluations of existing checkpoints. With fixed beta endpoints, changing T changes accumulated noise and terminal-prior mismatch as well as the discretization.

Table 13: **Historical controls across diffusion-step counts.** One training seed (42), 90k updates, full ancestral sampling with NFE$= T$, and 10k/10k FID. All rows use linear beta endpoints 10^{-4} and 0.028.

	CelebA64		LSUN Church64	
T / NFE	White	Flicker	White	Flicker
100	181.600	153.515	113.123	47.014
150	82.274	30.500	48.587	11.108
200	55.769	17.143	29.685	8.529
500	6.409	9.567	6.133	8.857

Table 14: **DDIM at actual NFE40 after equal two-grid selection.** Final FID-50k uses an independent generation stream and the same 50k real reference per dataset. $T = 150$ entries are means and sample SDs over paired training seeds 202701/202702. White $T = 500$ entries are historical seed-42 anchors, not paired replications; dashes denote models outside this follow-up.

Dataset	Training T	White	Flicker
CelebA64	150	178.103 ± 2.593	57.643 ± 1.310
	500	4.521	—
LSUN Church64	150	107.755 ± 4.023	23.420 ± 0.936
	500	4.538	—

NFE40 follow-up on existing checkpoints. We reuse the two-seed $T = 150$ EMA checkpoints and the historical seed-42 white $T = 500$ checkpoints; this follow-up adds no training. Each model selects between uniform and quadratic deterministic DDIM grids using 2,000 generated images per candidate, followed by an independent generation stream of 50,000 images for the selected grid. The same fixed 50,000-image real reference per dataset is used in both stages and in the ancestral replications above. There are 20 pilot evaluations and 10 final confirmations. The initial-noise seed is 240000000 for selection and 250000000 for confirmation, with nonoverlapping fixed per-image streams.

All models use actual NFE40, unconditional prediction without CFG, FP32 evaluation with autocast and TF32 disabled, and their original training covariance backend. There is no per-step clipping; the final output is clamped and quantized to PNG. Forward-call hooks verify 40 U-Net evaluations per image. All eight $T = 150$ models select uniform spacing; both $T = 500$ anchors select quadratic spacing.

Flicker improves every paired $T = 150$ comparison, reducing mean FID by 67.63% on CelebA and 78.27% on Church. The white $T = 500$ anchors are substantially better at the same NFE40. Thus the new evaluations support an improvement at the same $T = 150$, not equal-quality acceleration over these $T = 500$ anchors. The two-grid search is finite. Relative to the ancestral NFE150 results, both sampler and NFE change, so their FID difference does not isolate the effect of reducing network evaluations.

I.1 AFHQ-Cat64: an additional replication at $T = 150$

Data and training. We use the cat category of the official AFHQv2 release (Choi et al., 2020), retaining its original 5,065 training and 493 test images. RGB images are resized from 512×512 to 64×64 with Pillow LANCZOS, normalized to $[-1, 1]$, and randomly flipped horizontally during training. The test images are excluded from training, spectral selection and the FID reference. The unconditional U-Net has base width 128, channel multipliers $(1, 2, 3, 4)$, three residual blocks per level, attention at level 2, dropout 0.1 and time-embedding capacity 500 for both $T = 150$ and $T = 500$. All models use 90,000 optimizer updates, global batch 128, learning rate 10^{-4} without decay, zero weight decay, gradient clipping at 1 and EMA decay 0.9999. BF16 network training uses two microbatches of 64 per update; covariance operations use FP32 FFTs. The linear beta schedule has endpoints 10^{-4} and 0.028. Training seeds 2026092401 and 2026092402 pair initialization across conditions. White noise uses pixel MSE; Flicker uses the periodic Manhattan kernel $(1 + d)^{-\eta}$ and its matching whitened residual.

Table 15: **Historical $T = 500$ deterministic DDIM controls.** The same 90k-update, seed-42 checkpoints are evaluated with a uniform grid. Each cell uses 10k generated/10k reference images. These are within-protocol controls, separate from the 50k/50k primary results.

Dataset	Noise	NFE50	NFE100	NFE200	NFE500
CelebA64	White	7.222	6.250	5.942	5.944
	Flicker	11.156	9.875	9.377	9.112
LSUN Church64	White	7.424	7.091	6.927	6.782
	Flicker	9.234	8.756	8.570	8.487

Exponent selection before training. We measure the mean RGB power spectrum of all training images with an orthonormal two-dimensional FFT and integer radial bins, excluding DC from the fit. Applying the saved SNR-response-window rule to AFHQ's own spectrum selects $k = 6, \ldots, 12$, giving $\alpha = 2.68004$ and $R^2 = 0.99930$. This transfers a window rule calibrated on historical CIFAR measurements, not CIFAR's frequency bins or a universally optimal band. Before any AFHQ training or FID evaluation, we specify two candidates. The historical heuristic $\eta = (3 - \alpha)/2$ gives $\eta = 0.15998118$; this relation is not established for the finite periodic kernel used here. A separate shape fit compares centered log spectra of the data and the actual 64×64 kernel on the same bins, searching $\eta = 0.05 + 0.0005j$, $j = 0, \ldots, 6300$. Its minimizer is the lower boundary $\eta = 0.05$, whose noise-spectrum slope is only 2.05946, so it is not an exact spectral match. We report both candidates separately. Sampler-search budgets are matched per model; Flicker includes two trained covariance choices, whereas white has one.

Sampling and reference distribution. Each final EMA checkpoint selects between uniform and quadratic deterministic DDIM grids using 2,000 generated images per candidate. The selected grid then generates 50,000 images with a separate random stream. Selection and confirmation root seeds are 260000000 and 270000000, shared across conditions through fixed per-image streams. Sampling is unconditional, without CFG, and forward-call counts verify actual NFE40 for every image. Evaluation uses FP32, without autocast or TF32; only final outputs are clamped and quantized to PNG. FID uses 2,048-dimensional pytorch-fid features and all *5,065 training images exactly once* as its real reference, rather than repeating them to obtain 50,000 real images. The same reference is used for selection and confirmation. These FID values therefore belong to a different reference-sample protocol from the CelebA/Church 50k/50k results.

Table 16: **AFHQ-Cat64 at actual NFE40.** All entries use 90k updates, final EMA, equal two-grid DDIM selection and 50k generated images against 5,065 real training images. Seed columns identify 2026092401/2026092402; $\pm$ is their sample SD. A dash denotes an unavailable final evaluation and is not included in a mean.

Training T	Noise / exponent	Seed 1	Seed 2	Mean $\pm$ SD
150	White	50.647	59.967	55.307 ± 6.590
150	Flicker, $\eta \approx 0.160$	22.700	24.933	23.816 ± 1.579
150	Flicker, $\eta = 0.05$	81.044	83.700	82.372 ± 1.878
500	White	9.638	—	—

All six $T = 150$ models select uniform grids, while the white $T = 500$ reference selects quadratic spacing. Its second-seed run did not start within the original allocated time window.

Results and scope. At $T = 150$, $\eta \approx 0.160$ improves both paired seeds and reduces mean FID by 56.94% at NFE40. The smaller exponent performs worse than white noise, demonstrating sensitivity to covariance choice. The completed white $T = 500$ reference obtains FID 9.638 under the same per-model search and evaluation protocol. It is a single-seed DDIM comparison with the $T = 150$ models, not a matched $T = 500$ white/Flicker experiment. The $T = 150$ comparison demonstrates a data-informed covariance choice that improves quality at the same network, update and sampling budgets per model.

J COVARIANCE-AWARE SAMPLERS AND COORDINATE CHECKS

Let L be invertible with $LL^\top = \Sigma$. In coordinates $y = L^{-1}x$, the forward process becomes $y_t = \sqrt{\bar\alpha_t}y_0 + \sqrt{1-\bar\alpha_t}\xi_w$, with conjugate white-noise predictor $L^{-1}\epsilon_\theta(Ly, t)$. Applying ordinary DDIM in these coordinates and multiplying by L recovers Eq. 3. This concerns the conjugate predictor, not an independently trained white network with unchanged weights. Pixel-space thresholding need not commute with this coordinate transformation.

For clarity, let ρ denote DDIM stochasticity and reserve η for the covariance exponent. For $u < t$, the colored stochastic update is

$$q_{u,t}^2 = \rho^2 \frac{1-\bar\alpha_u}{1-\bar\alpha_t}\left(1 - \frac{\bar\alpha_t}{\bar\alpha_u}\right), \tag{39}$$

$$x_u = \sqrt{\bar\alpha_u}\hat x_0 + \sqrt{(1-\bar\alpha_u) - q_{u,t}^2}\,\epsilon_\theta(x_t, t) + q_{u,t}L\xi'_w, \qquad \xi'_w \sim \mathcal{N}(0, I). \tag{40}$$

The permissible ρ must keep $q_{u,t}^2 \le (1-\bar\alpha_u)$; $\rho = 0$ yields deterministic DDIM. This follows from the same coordinate change in the non-isotropic framework (Voleti et al., 2022).

The non-isotropic VP SDE and probability-flow ODE in forward time are

$$\mathrm{d}x = -\tfrac{1}{2}\beta(t)x\,\mathrm{d}t + \sqrt{\beta(t)}L\,\mathrm{d}w_t, \tag{41}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\tfrac{1}{2}\beta(t)x - \tfrac{1}{2}\beta(t)\Sigma s_\theta(x,t) = -\tfrac{1}{2}\beta(t)x + \frac{\beta(t)}{2\sqrt{1-\bar\alpha_t}}\epsilon_\theta(x,t). \tag{42}$$

The reverse SDE has the full $\beta\Sigma s_\theta$ correction, whereas probability flow has one half. Generation integrates the probability-flow equation toward decreasing forward time. The covariance is already included in the colored prediction.

K GENERALIZED MODE CALCULATIONS AND SYNCHRONIZATION BOUNDS

K.1 NATURAL CLOCKS IN GENERALIZED COVARIANCE MODES

The relevant scale is each mode's data variance relative to its noise variance. Generalized modes expose these ratios. For centered Gaussian data with $S \succ 0$ and noise $\Sigma \succ 0$, whitening and eigendecomposition give

$$\Sigma^{-1/2}S\Sigma^{-1/2} = U\,\mathrm{diag}(\tau_k)U^\top, \qquad A = \Sigma^{1/2}U, \qquad y = A^{-1}x. \tag{43}$$

Noncommuting covariances are allowed. Gaussian channels are independent, with variances τ_k and identity noise. Real-image coordinates give a second-order proxy without implied independence. The following angle measures corruption level; its spread describes how differently the modes progress through denoising:

$$\theta_k(t) = \arctan\frac{\sqrt{1-\bar\alpha_t}}{\sqrt{\bar\alpha_t}\sqrt{\tau_k}}, \qquad \omega_k = \frac{\tau_k\|A_{:k}\|^2}{\mathrm{tr}\,S}. \tag{44}$$

The weights sum to one and represent physical data energy. When $\Sigma = cS$, every mode shares one angle: their second-order denoising clocks are synchronized. This is a second-order synchronization target; absolute noise scale and finite sampling steps have separate effects.

For natural time $s = \log(\sqrt{1-\bar\alpha_t}/\sqrt{\bar\alpha_t})$, Lipschitz variance bounds give

$$\mathrm{Var}_\omega\,\theta(t) \le \frac{\mathrm{Var}_\omega(\log\tau)}{16}, \qquad \mathrm{Var}_\omega\,\partial_s\theta(t) \le \frac{\mathrm{Var}_\omega(\log\tau)}{64}. \tag{45}$$

Reducing log-eigenvalue dispersion tightens these bounds (proof below). Scaling Σ preserves that dispersion while shifting the clocks and terminal mismatch, separating the roles of spectral shape and noise scale.

K.2 Finite-terminal and finite-step effects

Sampling also depends on its initial distribution and discrete updates. Gaussian conditioning gives an exact predictor that isolates these effects from learned-response errors. For centered Gaussian data, substituting this predictor into DDIM, with $v_{t,k} = \bar{\alpha}_t \tau_k + (1 - \bar{\alpha}_t)$, yields

$$G_{u,t,k} = \frac{\sqrt{\bar{\alpha}_u \bar{\alpha}_t}\tau_k + \sqrt{(1-\bar{\alpha}_u)(1-\bar{\alpha}_t)}}{v_{t,k}} = \sqrt{\frac{v_{u,k}}{v_{t,k}}} \cos(\theta_k(t) - \theta_k(u)). \tag{46}$$

In this Gaussian model, starting from pure noise $y_{\max} \sim \mathcal{N}(0, I)$ and including the final clean update gives the output-to-target variance ratio

$$q_k = m_k r_k, \qquad m_k = v_{\max,k}^{-1}, \qquad r_k = \prod_j \cos^2(\Delta\theta_{j,k}). \tag{47}$$

Here m_k measures finite-terminal prior mismatch and r_k measures DDIM contraction. The true Gaussian terminal marginal makes $m_k = 1$. Their effects can partially cancel. For nonorthogonal A, the shared-Gaussian coupling cost $\sum_k \tau_k \|A_{:k}\|^2 (\sqrt{q_k} - 1)^2$ upper-bounds Euclidean W_2^2.

For small angular jumps, a statistical grid follows equal increments of

$$\ell(t) = \int^t \sqrt{\sum_k \omega_k \dot{\theta}_k(u)^2}\, \mathrm{d}u. \tag{48}$$

By Cauchy–Schwarz, equal arc increments minimize this Gaussian proxy's leading fine-grid weighted negative-log-contraction objective. The calculations below give the proof and projection to training indices, extending the Gaussian angular grids of Sabour et al. (2024) to a training-statistics prescription.

K.3 Proofs and implementation details

Noncommuting covariances. From Eq. 43, $A^{-1} S A^{-\top} = \mathrm{diag}(\tau)$ and $A^{-1}\Sigma A^{-\top} = I$. Hence the centered Gaussian channel becomes independent across generalized modes. Also $S = A\,\mathrm{diag}(\tau) A^\top$, which proves $\sum_k \tau_k \|A_{:k}\|^2 = \mathrm{tr}\, S$ and the normalization in Eq. 44. All statements require positive eigenvalues; zero-variance modes must be excluded or explicitly regularized before taking logarithms. Our frozen empirical S and production Σ are positive definite without adding a new ridge to S.

Proof of the dispersion bounds. Write $h = \log\tau$, $z = s - h/2$, and $\theta(h, s) = \arctan\exp(s - h/2)$. Direct differentiation gives

$$\theta_h = -\tfrac{1}{4}\,\mathrm{sech}\, z, \qquad \theta_{sh} = \tfrac{1}{4}\,\mathrm{sech}\, z \tanh z. \tag{49}$$

The absolute maxima are $1/4$ and $1/8$. If R, R' are independent draws from the fixed weights ω, an L-Lipschitz function satisfies

$$\mathrm{Var}_\omega\, f(R) = \tfrac{1}{2}\mathbb{E}[f(R) - f(R')]^2 \le \tfrac{1}{2} L^2 \mathbb{E}[R - R']^2 = L^2\, \mathrm{Var}_\omega\, R. \tag{50}$$

Taking $R = \log\tau$ proves Eq. 45. The global constants can be approached by two nearby, equal-weight modes centered where the respective derivative is maximal. They are bounds on dispersion, not on mean angle or mean speed. For physical training time, $\partial_t \theta = s'(t)\partial_s \theta$, so its speed-dispersion bound also includes $|s'(t)|$. Scalar multiplication of Σ shifts $\log\tau$ by a constant and preserves ω and its variance.

Exact DDIM gain. For Gaussian y_0 with variance τ_k, the conditional predictors are

$$\mathbb{E}[y_{0,k} \mid y_{t,k}] = \frac{\sqrt{\bar{\alpha}_t}\tau_k}{v_{t,k}} y_{t,k}, \qquad \mathbb{E}[\xi_{w,k} \mid y_{t,k}] = \frac{\sqrt{1-\bar{\alpha}_t}}{v_{t,k}} y_{t,k}. \tag{51}$$

Inserting them in DDIM gives Eq. 46; the cosine identity follows by normalizing the two-dimensional vector $(\sqrt{\bar{\alpha}_t}\sqrt{\tau_k}, \sqrt{1-\bar{\alpha}_t})$. The variance factors telescope across steps to $\tau_k / v_{\max,k}$, multiplied by the cosine-square product. Thus pure-noise initial variance one gives Eq. 47. Under the same latent standard Gaussian ξ_w, couple the target $A\,\mathrm{diag}(\sqrt{\tau_k})\xi_w$ and generated Gaussian $A\,\mathrm{diag}(\sqrt{\tau_k q_k})\xi_w$. Their expected squared distance is the coupling cost stated above. Off-diagonal terms vanish because the $\xi_{w,k}$ are independent. The coupling is generally not the Euclidean optimal-transport map when A is nonorthogonal.

Arc-length allocation. For a dense grid with point density $n(t)$ and fixed total number of intervals, the leading objective is proportional to $\int g(t)/n(t)\,\mathrm{d}t$, where $g(t) = \sum_k \omega_k \dot{\theta}_k(t)^2$. Cauchy–Schwarz gives

$$\left(\int \sqrt{g(t)}\,\mathrm{d}t\right)^2 \le \left(\int \frac{g(t)}{n(t)}\,\mathrm{d}t\right)\left(\int n(t)\,\mathrm{d}t\right), \tag{52}$$

with equality at $n(t) \propto \sqrt{g(t)}$. This proves the fine-grid allocation statement, not a finite-step global optimum. In computation we accumulate Euclidean lengths of consecutive weighted angle vectors, interpolate equal arc targets, and deterministically enforce distinct integer indices, including both available training endpoints. The final update to clean time is included when evaluating the contraction. In the matched case $\Sigma = cS$, this reduces to equal common-angle spacing, connecting directly to Gaussian angular schedules (Sabour et al., 2024).

An affine field can still have grid-dependent numerical error. With one data and noise mode both $\mathcal{N}(0,1)$, all VP marginals remain standard Gaussian. The oracle probability-flow velocity is zero, but a DDIM step multiplies the state by $\cos\Delta\theta$. Thus a finite sequence contracts variance even though the continuous field is linear and stationary. For the $T = 500$ beta path used here and 20 timepoints, final oracle variances are 0.85186 (uniform) and 0.85864 (quadratic). Their squared Gaussian Wasserstein distances to variance one are 0.005935 and 0.005383. This elementary counterexample separates spatial linearity from uniform-time numerical accuracy.

L Finite-terminal and synchronization statistics

We separate a covariance-only Gaussian proxy, independent of generated FID, from the learned Fourier response and spectral evolution (Section 3). We estimate the 1024×1024 data covariance S from all 50,000 CIFAR training images in $[-1,1]$, centering each channel and pixel and averaging channel covariances; production factors define Σ.

Flicker narrows modal timing dispersion in the Gaussian proxy. Here τ_k are eigenvalues of $\Sigma^{-1/2}S\Sigma^{-1/2}$, giving modal data-to-noise variance ratios. The corruption angle θ_k is the arctangent of the modal noise-to-signal amplitude ratio; smaller angular spread indicates more synchronized denoising (Appendix K). The energy-weighted SD of $\log\tau_k$ drops from 2.325 (white) to 1.026 (Flicker). Mean SD of θ_k over training indices decreases from 0.3320 to 0.1812 at $T = 150$ and from 0.3120 to 0.1218 at $T = 500$, approximately 45% and 61% reductions, supporting synchronization in the Gaussian proxy.

Flicker reduces the Gaussian proxy's terminal mismatch. The schedules share beta endpoints $(10^{-4}, 0.028)$, but terminal $\bar{\alpha}$ is 0.11913 at $T = 150$ and 0.0008321 at $T = 500$. At $T = 150$, the normalized Gaussian coupling cost isolates initialization from $\mathcal{N}(0,\Sigma)$ instead of the true terminal law, with finite-step contraction removed (Eq. 47, $r = 1$). It is 0.22016 for white and 0.001769 for Flicker (Table 18, m-only column).

For each channel c, let μ_c be its training mean vector. The spatial proxy used throughout is

$$S = \frac{1}{3M}\sum_{i=1}^{M}\sum_{c=1}^{3}(x_{i,c}-\mu_c)(x_{i,c}-\mu_c)^\top, \qquad M = 50{,}000. \tag{53}$$

It omits joint RGB and class-conditional covariance. For dense Flicker we reconstruct Σ from the float32 production Cholesky factor promoted to float64, rather than substituting an ideal Fourier power law. The calculations use float64 eigendecompositions and the exact frozen scalar beta arrays. Data-derived statistics are computed from the training set; the direct learned-field probes use disjoint fitting/validation subsets of the test set.

Table 18 also shows why synchronization alone is not an ordering theorem for finite-step error: at $T = 500$ the Flicker uniform grid has higher Gaussian coupling cost than the white uniform grid. Within the tested grids, the statistical arc reduces this Gaussian cost for both at $T = 500$. At short T, pure-noise mismatch dominates white, while Flicker's mismatch and contraction partly cancel. These are testable mechanism predictions, not surrogate FID results.

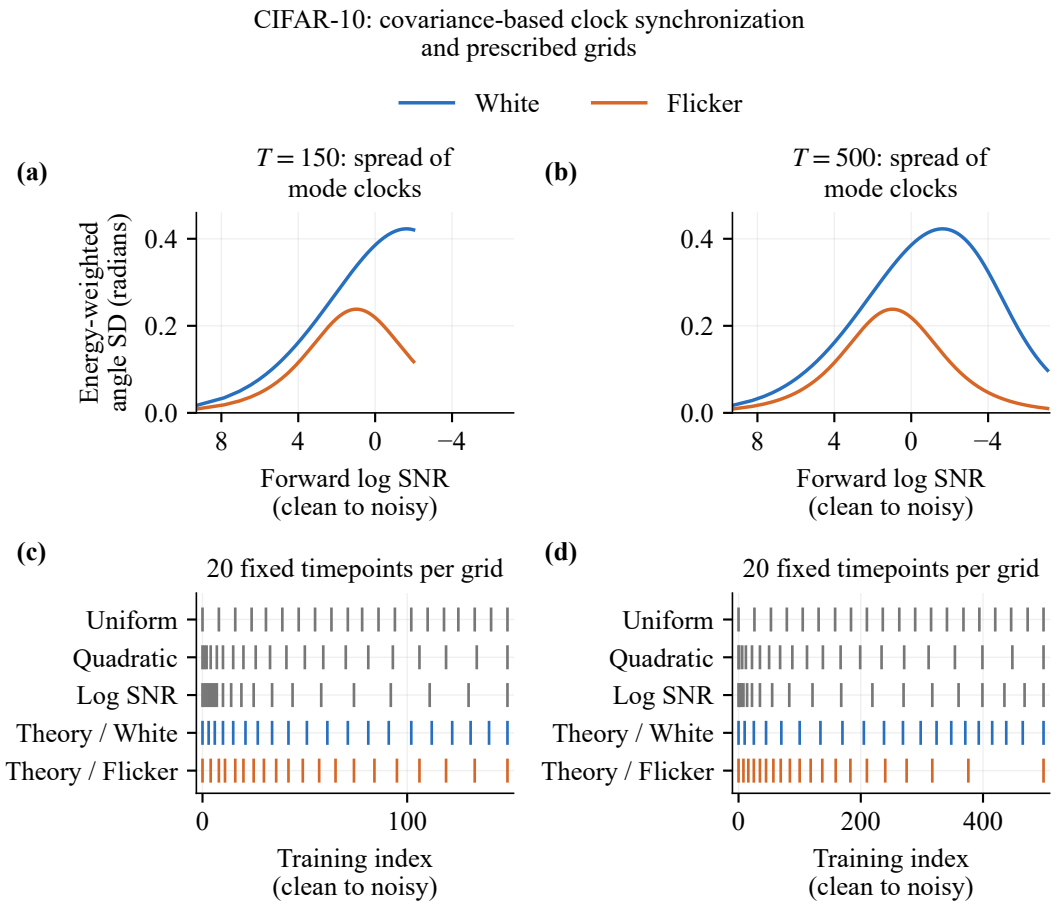


Figure 14: **More synchronized denoising clocks.** CIFAR training-statistics diagnostics in generalized modes of the actual covariance operators. Flicker narrows the second-order denoising clocks and improves finite-terminal matching at $T = 150$ in this Gaussian proxy.

Table 17: Energy-weighted synchronization statistics. Mean and maximum angle SD are over each schedule's available training indices, not over the entire real line of natural time. The log-eigenvalue SD does not depend on the scalar schedule.

Noise	T	SD($\log \tau$)	Mean angle SD	Maximum angle SD
White	150	2.32535	0.33204	0.42284
Flicker	150	1.02587	0.18118	0.23822
White	500	2.32535	0.31203	0.42284
Flicker	500	1.02587	0.12179	0.23822

M Posterior curvature and the colored probability-flow field

State-space linearity. For arbitrary data, let $u = L^{-1}x_0$, $y = L^{-1}x_t$, and let V and K_3 be the posterior covariance and third central moment of $u \mid y, c$ for a fixed class c. Write $\epsilon_w^* = L^{-1}\epsilon_*$ for the whitened optimal noise prediction. Gaussian-channel derivative identities (Efron, 2011; Manor

Table 18: All Gaussian oracle coupling costs with 20 timepoints, normalized by $\operatorname{tr} S$. The m-only column sets $r = 1$; the r-only column uses the exact forward terminal marginal ($m = 1$). The actual pure-noise result retains their product. These costs are not additive and are not FID; for dense modes they upper-bound Euclidean $W_2^2/\operatorname{tr} S$.

Noise	T	Grid	m only	r only	Actual mr
White	150	Uniform	0.220156	0.002360	0.223424
White	150	Quadratic	0.220156	0.001082	0.224007
White	150	LogSNR	0.220156	0.001568	0.225685
White	150	Statistical arc	0.220156	0.001111	0.222986
Flicker	150	Uniform	0.001769	0.003941	0.001754
Flicker	150	Quadratic	0.001769	0.002794	0.001649
Flicker	150	LogSNR	0.001769	0.004310	0.002758
Flicker	150	Statistical arc	0.001769	0.002371	0.001405
White	500	Uniform	0.000383	0.009509	0.011488
White	500	Quadratic	0.000383	0.009666	0.012610
White	500	LogSNR	0.000383	0.009894	0.012534
White	500	Statistical arc	0.000383	0.006521	0.008473
Flicker	500	Uniform	8×10^{-8}	0.013184	0.013145
Flicker	500	Quadratic	8×10^{-8}	0.006044	0.006022
Flicker	500	LogSNR	8×10^{-8}	0.010436	0.010410
Flicker	500	Statistical arc	8×10^{-8}	0.004515	0.004497

& Michaeli, 2024) give

$$J_y \epsilon_w^* = \frac{I}{\sqrt{1-\bar{\alpha}_t}} - \frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)^{3/2}} V, \qquad (H_y \epsilon_w^*)_{ijk} = -\frac{\bar{\alpha}_t^{3/2}}{(1-\bar{\alpha}_t)^{5/2}} (K_3)_{ijk}. \tag{54}$$

Matching a second-order covariance does not control K_3; fixed whitening of the loss does not remove it. For positive $\sqrt{\bar{\alpha}_t}, \sqrt{1-\bar{\alpha}_t}$, the oracle field is affine on a connected open domain exactly when K_3 vanishes throughout it. Learned approximation error and CFG introduce further distinctions, as detailed below.

Let $m = \mathbb{E}[u \mid y, c]$, $V = \mathbb{E}[(u-m)(u-m)^\top \mid y, c]$, and $(K_3)_{ijk} = \mathbb{E}[(u_i - m_i)(u_j - m_j)(u_k - m_k) \mid y, c]$ in the whitened channel $y = \sqrt{\bar{\alpha}_t} u + \sqrt{1-\bar{\alpha}_t}\xi_w$. Assume finite required posterior moments and differentiation under the Gaussian integral. Gaussian posterior calculus (Efron, 2011; Manor & Michaeli, 2024) gives

$$\partial_j m_i = \frac{\sqrt{\bar{\alpha}_t}}{(1-\bar{\alpha}_t)} V_{ij}, \qquad \partial_k V_{ij} = \frac{\sqrt{\bar{\alpha}_t}}{(1-\bar{\alpha}_t)} (K_3)_{ijk}. \tag{55}$$

Since $\epsilon_w^* = (y - \sqrt{\bar{\alpha}_t} m)/\sqrt{1-\bar{\alpha}_t}$, differentiating twice proves Eq. 54. The whitened forward-time probability-flow field is $v = -\beta y/2 + \beta \epsilon_w^*/(2\sqrt{1-\bar{\alpha}_t})$, and therefore

$$(H_y v)_{ijk} = -\frac{\beta(t)\bar{\alpha}_t^{3/2}}{2(1-\bar{\alpha}_t)^3} (K_3)_{ijk}. \tag{56}$$

Zero posterior third moment at one point is insufficient for affinity; it must vanish across a connected open domain. For CFG, the Hessian is the same $(1+w)$ and $-w$ combination of conditional and unconditional Hessians. A learned field additionally contains approximation error, so measured curvature cannot uniquely identify posterior non-Gaussianity. None of these identities implies a straight trajectory or constant velocity in the original training-time coordinate.

Conditional-mean energy loss. Couple $x_u = \sqrt{\bar{\alpha}_u} x_0 + \sqrt{1-\bar{\alpha}_u}\xi_c$ and $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\xi_c$ with the same noise. Writing $b = \sqrt{1-\bar{\alpha}_u}/\sqrt{1-\bar{\alpha}_t}$ and $c = \sqrt{\bar{\alpha}_u} - b\sqrt{\bar{\alpha}_t}$, oracle DDIM returns $\mathbb{E}[x_u \mid x_t] = b x_t + c\mathbb{E}[x_0 \mid x_t]$. For an orthogonal projection P, the law of total variance yields

$$\mathbb{E}\|P x_u\|^2 - \mathbb{E}\|P\mathbb{E}[x_u \mid x_t]\|^2 = c^2 \mathbb{E}\|P(x_0 - \mathbb{E}[x_0 \mid x_t])\|^2 \geq 0. \tag{57}$$

This exact one-step identity applies to a true forward-marginal input. It neither closes a learned generation trajectory nor guarantees that all solvers underestimate energy. Equations 46–47 provide the tractable Gaussian multi-step case.

N TRAINING, EVALUATION, AND REPRODUCIBILITY

Covariance implementations. The principal CIFAR models use the nonperiodic dense Cholesky operator at $\eta = 0.2$, with $10^{-7}I$ regularization. Gaussian draws receive no per-image variance normalization. The periodic operators and dataset-specific exponents for the 64×64 experiments are specified in Appendices I and I.1. Appendix E illustrates the noise and spectral trajectories.

Table 19: Protocol index. Rows identify separate comparisons or diagnostics; no FID values are pooled across rows. NFE counts denoiser evaluations per generated image, including CFG branches; diagnostic anchors are not NFE.

Experiment / location	Training group	Sampling or measurement budget
P1: Main CIFAR (Sec. 4.2, App. H.1)	Batch 128; $T = 150$; three paired seeds	15 DDIM grid/CFG candidates per model; actual NFE40; FID-50k confirmation.
P2: Low-NFE points (Sec. 4.3, App. G)	Batch 64; three paired seeds; Flicker $T = 150$, white $T = 500$	Flicker uniform/$w = 1$, white quadratic/$w = 0.5$; DDIM NFE20 vs. 100. Other curve settings: Fig. 3.
P3: CelebA64/Church64 (Sec. 4.4, App. I)	$T = 150$; two paired seeds; historical white $T = 500$ seed-42 anchors	Unconditional ancestral NFE150; DDIM NFE40 follow-up with two-grid selection on reused checkpoints (Table 14). FID-50k; $T = 500$ anchors are unpaired.
P3b: AFHQ-Cat64 (Sec. 4.4, App. I.1)	$T = 150$; two paired seeds; two specified Flicker exponents; white $T = 500$ single-seed reference	Unconditional DDIM NFE40; two grids per model; 50k generated images against all 5,065 train images. Table 16 reports each exponent separately.
P4: Historical spectral response (Sec. 4.6, App. C)	One historical $T = 500$ checkpoint pair	64 forward measurement anchors; disjoint image splits and independent reverse trajectories. Diagnostic budget, not NFE64 or three-seed FID.
P4b: Main-model evidence (App. F)	Main $T = 150$ checkpoints; three registered seeds	Forward response at all 150 indices; DPM++ NFE40 with 15 candidates; KID/PRDC on the main DDIM images. Distinct measurements.
P5: Noise $\times$ loss (Sec. 4.5, App. S)	$T = 150$; three older training seeds	Uniform DDIM/$w = 1.8$; actual NFE40; FID-50k. Training runs differ from P1.
P6: Matched-terminal SNR (Sec. 4.5, App. U)	Three older seeds; batch 128, 90k updates; $T = 150/500$	Full ancestral/$w = 1.8$; actual NFE300 at $T = 150$, NFE1000 for its own white $T = 500$ anchor; final-EMA FID-50k.
P7: Search at $T = 500$ (Sec. 6, App. G)	$T = 500$; three paired seeds	Eight candidates per solver/model; NFE40, FID-50k (Table 9); distinct from P6's anchor.
P8: Initial-covariance confirmation (App. H.2)	Main $T = 150$ models; three paired seeds	Fixed uniform DDIM/$w = 1$; NFE40; only initial second-order covariance changes. FID-50k, distinct from App. P's FID-5k pilot.
Field / split-step probes (Apps. Q, P)	Specified $T = 150/500$ checkpoints	Finite perturbations and coarse/split updates. Extra diagnostic calls are separate from generation-quality NFE.

Table 19 separates the experiment groups and their budgets. P5 and P6 use seed values 72155260, 397493986 and 1914125884, with training runs distinct from P1. P6 changes the cumulative training schedule, whereas P8 changes only the sampling initialization. The recipes below describe the two newer CIFAR checkpoint groups; the cited appendices retain the other groups' own protocols.

The earlier operating-point seeds are 2027090701, 2027090702, and 2027090703. The symmetric-control family uses 2026091211, 2026091212, and 2026091213. Final EMA checkpoints at update 90,000 are used. The four-arm, two-solver selection results are $T = 500$; the statistical-grid study and mechanism probes also include the corresponding $T = 150$ white and dense checkpoints. Each reported standard deviation is across three training seeds, not a FID estimator confidence interval.

Both CIFAR recipes use AdamW with momentum coefficients $(0.9, 0.999)$, numerical epsilon 10^{-8}, zero weight decay, and constant learning rate 10^{-4}. Timesteps are sampled uniformly and independently per image from $\{0, \ldots, T-1\}$. Training applies horizontal flips with probability 0.5 and

Table 20: CIFAR-10 training recipes. Both use the same conditional U-Net family. The two checkpoint groups are evaluated separately.

Setting	Earlier operating-point study	Symmetric controls/probes
Resolution / labels	32^2/10	32^2/10
Base channels / multipliers	128 / [1,2,2,2]	128 / [1,2,2,2]
Residual blocks / attention level	2 / [1]	2 / [1]
Dropout	0.15	0.15
Global batch / updates	64 / 90,000	128 / 90,000
Learning rate / weight decay	10^{-4} / 0	10^{-4} / 0
EMA / label dropout	0.9999 / 0.1	0.9999 / 0.1
Training arithmetic	bfloat16	bfloat16
Training seeds	3 paired	3 paired
Linear beta endpoints	0.0001, 0.028	0.0001, 0.028

maps RGB pixels to $[-1, 1]$. The square center crop and resize preserve the native 32×32 image size; there is no random crop or color augmentation.

For symmetric selection, all 192 pilot cells and all 24 FID-50k confirmations are retained. Pilot and confirmation generation seeds are 114600000 and 114700000. Ties use the frozen candidate index. Each arm is selected independently from identical candidate spaces, using only its pilot FID. For the $T = 500$ eight-candidate comparison in Table 9, all confirmations select quadratic grids; selected DDIM guidance is 0.5 for white/dense and 1 for periodic/GFF. DPM++ uses 0.5 for all white seeds, $(0.5, 1, 1)$ for dense, and 1 for periodic/GFF. The GFF arm is a budget-controlled implementation, not a reproduction of the original paper's complete training and tuning recipe.

The deterministic DDIM and official DPM-Solver++ implementations count actual model calls, including CFG branches. Frozen input checks bind checkpoint, training configuration, scalar schedule, operator backend, source files, FID reference, and Inception weights. Completion checks verify 50,000 images and 40 model evaluations per generated image. The final audit verifies input/result seals and selection consistency. Pilot FID-2k and confirmation FID-50k are reported separately.

O CONFIRMATION ACROSS DIFFUSION-STEP COUNTS AND STATISTICAL-GRID SELECTION

The earlier study preceding Section 4.2 crosses white/dense noise, $T = 150/500$, and training seeds 2026091211–2026091213. It uses the same frozen FID reference and inference precision as the earlier symmetric control, but distinct generation seeds: 114800000 for each FID-2k pilot and 114900000 for each FID-50k confirmation. Dynamic thresholding is off. The six conventional candidates are uniform/quadratic/logSNR spacing crossed with $w = 0.5/1$. The separate two-candidate policy uses the training-statistics spectral arc and the same guidance values. Ties use the frozen candidate index. No confirmation FID enters selection.

Table 21: All eight groups from 24 FID-50k confirmations at actual NFE40. Entries are mean $\pm$ sample SD across the three training seeds. Conventional selection has six candidates per model, statistical arc two; each policy gives white and Flicker identical candidate sets.

T	Noise	Conventional selection	Statistical arc
150	White	14.629 ± 0.481	17.688 ± 0.600
150	Dense Flicker	6.467 ± 0.058	7.662 ± 0.083
500	White	7.770 ± 0.117	9.572 ± 0.185
500	Dense Flicker	8.123 ± 0.110	7.927 ± 0.142

All $T = 150$ models select uniform/$w = 1$ under the conventional policy, and $w = 1$ under the arc policy. All $T = 500$ models select quadratic/$w = 0.5$ and arc/$w = 0.5$. The $T = 500$ numerical values differ from Table 9 because this is an independent generation stream with a different candidate search; the final selected settings agree for white and dense DDIM. The audit verifies source/input/result seals, sample counts, NFE and the selection recomputed from the 96 pilot cells.

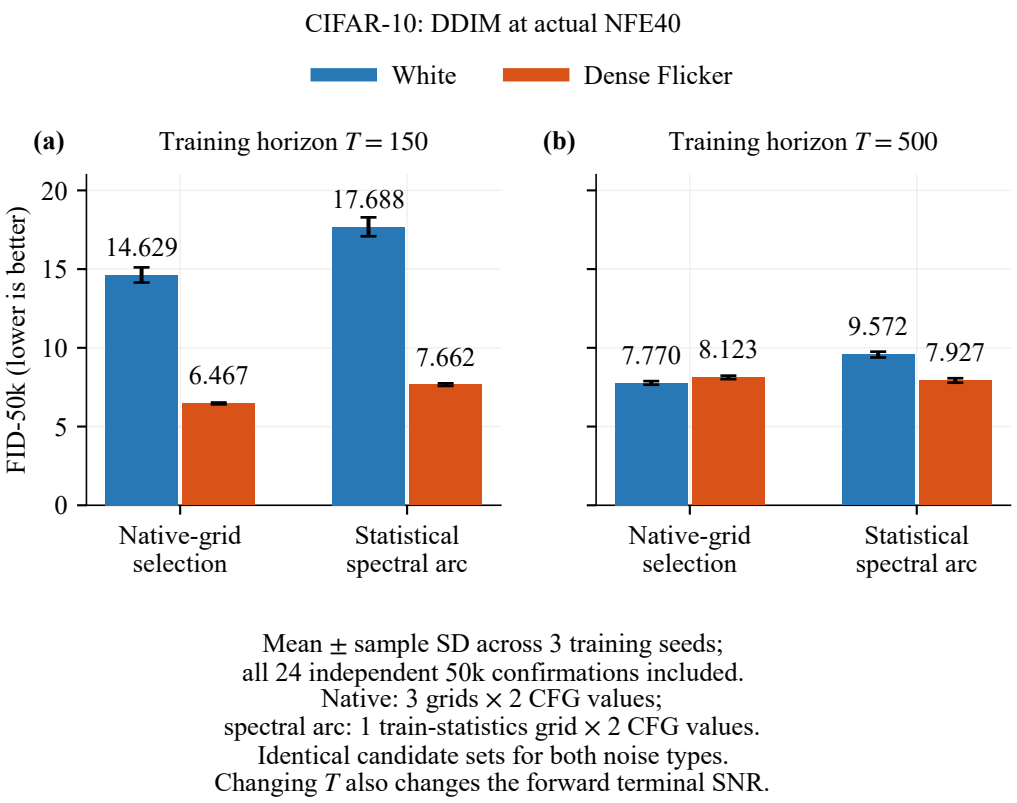


Figure 15: All 24 fresh-stream FID-50k confirmations at actual NFE40, summarized over three paired training seeds. The Flicker benefit at $T = 150$ survives symmetric finite sampler selection. The statistical angular grid improves Flicker at $T = 500$, but not at $T = 150$. Each policy gives both noise types the same candidates. Changing T also changes terminal SNR. Exact values are in Table 21.

At $T = 150$, each conventional grid improves from $w = 0.5$ to 1 for both noise types, motivating the broader guidance search. A later protocol extends both noise types symmetrically to $w = 1.5, 2, 3$ on all three grids, with independent 50k confirmation. The completed extension is reported in Section 4.2 and Appendix H.

P TERMINAL-COVARIANCE AND STEP-SPLITTING INTERVENTIONS

Changing only the initial covariance. For each of the same 12 checkpoints, we compare the standard prior $\mathcal{N}(0, \Sigma)$ and a calibrated Gaussian $\mathcal{N}(0, Q_T)$, where $Q_T = \bar{\alpha}_T S + (1 - \bar{\alpha}_T)\Sigma$ and S is the training RGB-averaged spatial covariance. If $R_T R_T^\top = Q_T$, a standard initial draw Lz is mapped to $R_T L^{-1}(Lz) = R_T z$. The same per-index z and labels couple both initializations. The mean remains zero in both arms; cross-channel, class-conditional, and higher-order statistics are not corrected. Thus this is a second-order intervention, not sampling from the exact non-Gaussian terminal marginal.

The EMA weights, forward covariance, DDIM updates, FP32 precision, time grid, CFG, generated labels, and reference are fixed. Every cell uses quadratic spacing, $w = 0.5$, actual NFE40, and 5,000 generated images (seed 115000000). All 24 prespecified cells are included, with no pilot selection. The covariance reconstruction relative error in FP32 is at most 3.46×10^{-8}, and the same-z mapping agrees within 2.87×10^{-6} in maximum absolute error.

Flicker $T = 150$ improves in all three seeds, whereas white worsens in all three. The change is opposite to the prespecified expectation that white would benefit more from covariance correction. Small $T = 500$ effects agree with the weaker endpoint signal there. At $T = 150$, the response depends on

Table 22: Exploratory terminal-prior intervention, FID-5k against the fixed 50k real reference. Means and sample SDs are over three training seeds. The paired difference is calibrated minus standard FID-5k. These values must not be compared numerically with FID-50k.

T	Noise	Standard	Calibrated	Paired difference
150	White	27.277 ± 1.013	41.204 ± 0.276	$+13.927 \pm 0.742$
150	Dense Flicker	14.281 ± 0.153	12.627 ± 0.174	-1.654 ± 0.140
500	White	12.255 ± 0.233	12.359 ± 0.260	$+0.104 \pm 0.028$
500	Dense Flicker	12.914 ± 0.089	12.950 ± 0.097	$+0.036 \pm 0.009$

higher moments, the learned score and guided dynamics in addition to spatial covariance. The setting here is quadratic/$w = 0.5$. A separate 12-cell FID-50k confirmation at uniform/$w = 1$ is now complete: it worsens both noise types, as reported in Appendix H. The exploratory improvement above is specific to its sampling configuration.

One coarse step versus two split steps. For each checkpoint, we use uniform and quadratic 20-timepoint grids, fixed $w = 1$, and 100 paired latent samples. At a base state x_t, compare $\Phi_{s,t}(x_t)$ with $\Phi_{s,u}(\Phi_{u,t}(x_t))$, where $u = \lfloor (t+s)/2 \rfloor$ and Φ is the unchanged DDIM update using a fresh prediction at u. Only the coarse update advances the base trajectory. Adjacent training indices, which cannot be split, and the final clean endpoint are skipped. For each valid interval, let D_j be the mean squared pixel difference between the two updates and I_j the mean squared coarse increment. We report $\sqrt{\sum_j \Delta t_j D_j / \sum_j \Delta t_j}$ and $\sqrt{\sum_j \Delta t_j D_j / \sum_j \Delta t_j I_j}$, separately in pixels and after the production output-whitening map. These are descriptive comparisons within a fixed grid, not a common step size across grids.

Table 23: Time-weighted split-step defect RMS, multiplied by 10^3. Mean $\pm$ sample SD across three training seeds. White is identical in pixel and whitened coordinates. All 24 settings are included; this is a self-consistency defect, not a known-oracle truncation error.

T	Grid	White	Dense, pixels	Dense, whitened
150	Uniform	1.325 ± 0.013	1.690 ± 0.030	1.496 ± 0.004
150	Quadratic	0.793 ± 0.015	1.475 ± 0.018	1.104 ± 0.028
500	Uniform	1.662 ± 0.011	2.829 ± 0.005	2.087 ± 0.015
500	Quadratic	1.197 ± 0.004	2.288 ± 0.048	1.318 ± 0.019

Flicker has larger defects for every paired seed/condition in both norms; normalization by the coarse increment gives the same ordering. The 24 base trajectories account for 96,000 image-level network calls, with 87,600 extra midpoint diagnostic calls reported separately, for 183,600 total. Diagnostic calls are accounted separately from generation NFE. Preflight checks verify every base state against the original sampler. The defect measures the learned sampler's self-consistency along its own trajectory. It complements the covariance-synchronization and final-quality measurements by testing local step composition.

Q Direct learned-field diagnostic protocol

Twelve models cross white/dense noise, $T = 150/500$, and three training seeds. At each scalar logSNR target in $\{-1.75, -0.5, 1, 2.5, 4, 6\}$, the nearest available training index is used. Each class contributes 25 fitting and 25 disjoint validation images, giving 250 in each split. The same image identities are used across models. We fit a full real 2×2 map per complex Fourier coefficient with ten class intercepts. This encompasses real/imaginary mixing within a mode, but not arbitrary linear coupling between modes. Consequently its held-out residual is an affine-model adequacy diagnostic, not a unique measure of nonlinearity. Fits use each model's own forward noise and are evaluated separately for conditional and guided predictions.

For the central-difference check, 50 fixed images are perturbed along shared raw-pixel directions at RMS amplitudes $h = 0.05, 0.1$, at scalar logSNR approximately $-1.75, 1, 4$. Let $f_\pm = f(x \pm h\delta)$

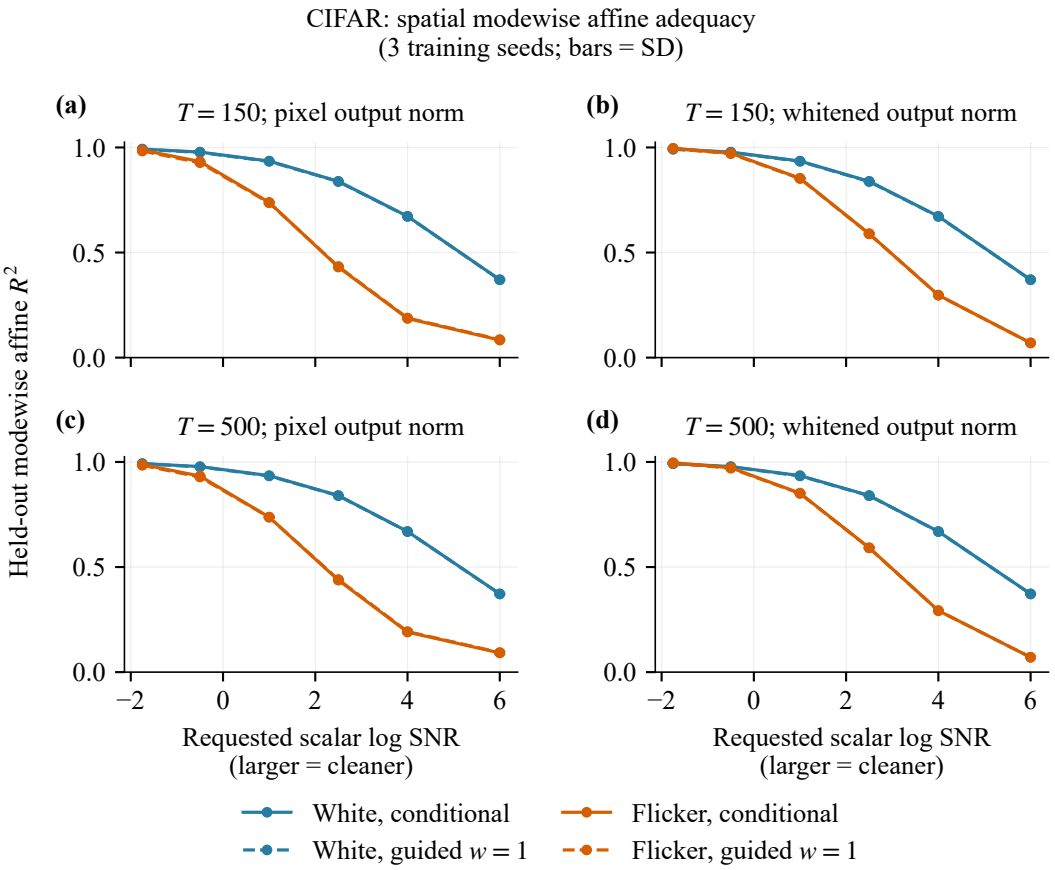


Figure 16: Held-out modewise affine adequacy of the learned CIFAR U-Nets. Training-seed SD is shown separately for the conditional and guided fields and for physical and production-whitened output norms. A smaller residual need not imply better generation, especially at high noise, where copying the noise is already nearly linear.

and $f_0 = f(x)$, where δ has unit pixel RMS. We measure midpoint deviation $d_2 = (f_+ + f_-)/2 - f_0$ relative to linear increment $d_1 = (f_+ - f_-)/2$ through the ratio of aggregate squared norms $\sum \|d_2\|^2 / \sum \|d_1\|^2$. Each is computed in physical pixels and after the production output whitening map. These central differences vanish for any affine field even when it couples modes. Both each model's own forward state and a common white-corrupted state are used; the latter is distribution-shifted input for Flicker.

Across $T = 150/500$, three probe times, conditional/guided fields, and two perturbation scales, dense Flicker has a smaller ratio in zero of 72 paired seed/condition comparisons, for either output norm. The 72 entries do not represent 72 independent training seeds. At high SNR, fixed pixel perturbations can exceed the colored noise scale in some modes. Indeed, at logSNR 4, the dense second-difference RMS after division by h^2 is approximately halved when going from $h = 0.05$ to 0.1. At logSNR approximately -1.75 and 1 the scale ratios are much closer to one. These observations constrain interpretation as a local Hessian.

Inference uses FP32; regression statistics and FFTs use float64/complex128. All 120,000 planned image-level forwards completed with bound input and output checksums. Synthetic affine tests verify the regression and finite-difference implementations at the stated precision. Separate CPU checks verify generalized-mode reconstruction, direct matrix-versus-mode DDIM, finite-terminal factors, the dispersion bounds, and posterior derivatives on analytic Gaussian/mixture controls.

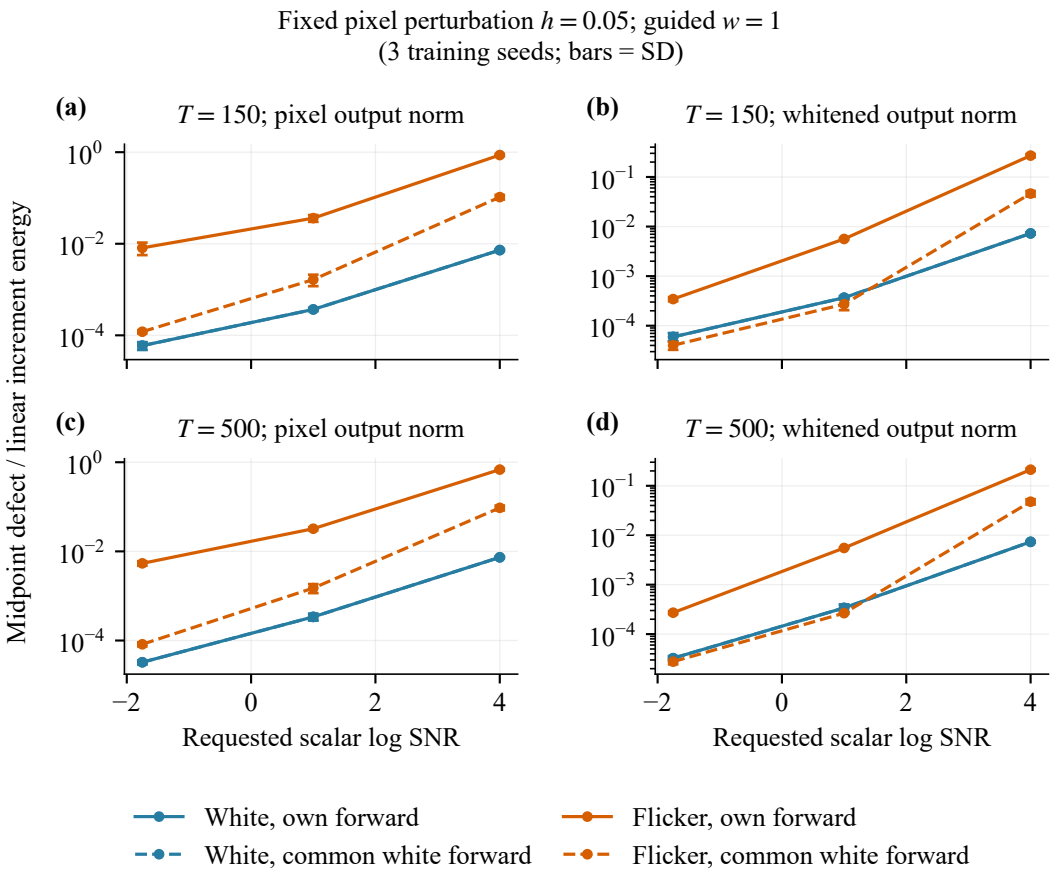


Figure 17: Finite-perturbation CIFAR predictor responses. The measurement uses fixed raw-pixel perturbations and reports physical and whitened output norms separately. Common white-corrupted inputs are out of distribution for Flicker; they do not establish an improvement on its own sampling distribution. These finite-scale responses are not coordinate-invariant Hessian measurements.

R Post-hoc sampler sensitivity within the frozen candidate set

The same 192 pilot results allow a secondary, post-hoc description of sampler sensitivity without new generation. For the eight settings in each model/solver's candidate set, define mean regret as the mean pilot FID minus that model/solver's lowest pilot FID. Also define a grid penalty as uniform-minus-quadratic pilot FID at fixed $w = 1$. All arms, settings, and seeds are included; the analysis was not specified before the generation experiment.

Table 24: Post-hoc sensitivity in the finite eight-setting pilot space. Values are differences in FID-2k, mean $\pm$ SD over three training seeds. They are not FID-50k gains or estimates of global tuning requirements.

Noise	Solver	Mean regret	Grid penalty at $w = 1$
White	DDIM	3.782 ± 0.220	3.864 ± 0.029
Dense Flicker	DDIM	2.121 ± 0.072	1.837 ± 0.047
Periodic Flicker	DDIM	1.513 ± 0.139	1.232 ± 0.263
GFF control	DDIM	2.312 ± 0.044	2.326 ± 0.104
White	DPM++	2.806 ± 0.165	3.082 ± 0.232
Dense Flicker	DPM++	1.392 ± 0.048	1.123 ± 0.076
Periodic Flicker	DPM++	1.278 ± 0.033	1.220 ± 0.142
GFF control	DPM++	1.797 ± 0.089	1.954 ± 0.147

Flicker has lower mean regret and a smaller grid penalty than white for both solvers within this candidate set. This is consistent with more balanced modal clocks. Mean regret measures variation relative to each method's own selected point, so it describes tuning sensitivity separately from the absolute best FID in Table 9.

S PREDICTION-TARGET AND LOSS-METRIC CONTROLS

Design by prediction target. Twelve models cross two designs, noise/x_0 prediction, and three seeds (2027090711, 2027090712, 2027090713). They share the conditional CIFAR U-Net, 90,000 updates, global batch 128, BF16, and $T = 150$. Within each block, hardware family, initialization, and named random streams are paired. Evaluation is fixed before these models are trained: DDIM, 20 timepoints, $w = 1$, uniform grid, and 50,000 images, with actual NFE40.

Table 25: All prediction-target controls, FID-50k. Each design includes its own covariance and scalar schedule. SD is over three training seeds.

Design / target	Seed 1	Seed 2	Seed 3	Mean ± SD
Flicker / noise	6.365	6.624	6.395	6.462 ± 0.142
Flicker / x_0	9.438	9.279	9.368	9.362 ± 0.080
EqualSNR / noise	10.560	10.749	10.291	10.533 ± 0.230
EqualSNR / x_0	19.226	18.765	19.243	19.078 ± 0.271

Flicker uses dense nonperiodic $\eta = 0.2$ noise with terminal $\bar{\alpha} = 0.11913$. The EqualSNR control uses a periodic empirical Fourier covariance and an average-SNR-calibrated path; both whiten their own corruption residual. For matched predictions, converting a whitened clean-image error to a whitened noise error multiplies squared loss by $\bar{\alpha}_t/(1 - \bar{\alpha}_t)$. Thus target choice also changes implicit time weighting, and the design contrast includes covariance, boundary, and schedule. These are budget-controlled variants, not reproductions of the original EqualSNR 800,000-update cosine-schedule benchmark (Falck et al., 2025).

Table 26: **Joint design of corruption and training objective.** CIFAR-10 FID-50k, means over three training seeds; $T = 150$, NFE40, uniform DDIM and $w = 1.8$.

Training noise	Pixel loss	Flicker-whitened loss
White	15.089	15.550
Flicker	24.339	**6.856**

Noise by loss metric. An earlier three-seed control holds its scalar $T = 150$ path fixed while crossing white/colored noise with pixel/Flicker-whitened residual metrics. It uses DDIM with 20 timepoints, $w = 1.8$, uniform spacing, and FID-50k at NFE40. Mean FIDs are 15.089 (white/pixel), 15.550 (white/Flicker metric), 24.339 (colored/pixel), and 6.856 (colored/Flicker metric). The paired interaction is -17.945 ± 1.309, supporting the joint design of covariance and loss metric. The metric affects total gradient scale as well as directional weighting.

Historical CIFAR exponent scan. Figure 4 reports the earlier CIFAR-10 sweep at $T = 150$, with $\eta \in \{0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30\}$ and a white-noise baseline. The historical recipe uses a U-Net with base width 128, channel multipliers $(1, 2, 2, 2)$, two residual blocks, attention at 16×16, dropout 0.15, Adam with learning rate 10^{-4}, 200 epochs, and linear beta endpoints $(10^{-4}, 0.028)$. All displayed values use the final saved epoch 199 and FID with 50,000 generated images against the same 50,000-image CIFAR-10 reference. The archived FID recalculation logs for the seven colored endpoints record CFG with $w = 1.8$, the Cholesky covariance implementation, and an evaluation batch size of 80. The saved endpoint FIDs are respectively 444.04, 11.53, 9.08, 10.23, 8.52, 9.93, and 9.50; white obtains 23.27. The scan reports one archived endpoint value per condition, rather than a mean over training seeds. It is separate from the 90,000-update, three-seed confirmations in Section 4.2. The minimum identifies the empirical operating choice within this tested grid; the relation $\eta = (3 - \alpha)/2$ remains a candidate-selection heuristic, not a guarantee of the FID-minimizing exponent.

Data-informed covariance selection and training cost. Measured image spectra provide a direct statistical input to covariance design. In the dataset adaptations, we estimate the finite-band power-law slope from dataset images and use it to construct candidates in Eq. 1; the dataset-specific exponents and selection procedures are recorded in Appendices I and I.1. AFHQ specifies both a slope-based candidate and a fit to the actual finite-kernel spectrum before training or FID evaluation. The principal CIFAR experiments retain the historical empirical choice $\eta = 0.2$ (Fig. 4).

Converting a spectral slope to a finite-grid kernel requires a modeling choice. The historical rule $\eta = (3 - \alpha)/2$ retains a Matérn algebraic prefactor while removing its exponential cutoff, so it is a heuristic rather than a controlled asymptotic relation. The alternative $\alpha = 2 - \eta$ describes an idealized unshifted two-dimensional power law. We evaluate the resulting finite-grid candidates empirically; AFHQ reports both candidates, including the less effective one.

The design retains the baseline U-Net and optimizer-update budget, with no auxiliary network, distillation or additional training stage. Its computational overhead consists of fixed covariance and whitening operations. The reported efficiency gain concerns sampling NFE; end-to-end training and sampling wall-clock costs were not benchmarked.



Figure 18: **Fixed-index examples from the main $T = 150$/NFE40 study.** Flicker (top, $w = 1$) and white (bottom, $w = 1.5$), both uniform DDIM, use training seed 2026091211. Each row shows indices 0–7 from its complete FID-50k output, fixed before viewing. Equal indices do not imply paired image content; the full-set comparison is in Fig. 2.

T Qualitative samples and provenance

Figure 18 uses the first registered training block of the main study (seed 2026091211), with final EMA checkpoints at update 90,000. Both models use $T = 150$, uniform DDIM and actual NFE40; the selected guidance is $w = 1$ for Flicker and $w = 1.5$ for white. Indices 0–7 were fixed before viewing and taken in recorded order from each complete 50,000-image confirmation output. The original PNG pixels are displayed identically, with no enhancement or content edits. The index rule does not imply matched semantics or latent states. The following panel belongs to the separate batch-64 operating-point study, retaining its own diffusion-step counts T, NFE and sampling settings.



Figure 19: Unfiltered CIFAR-10 examples from the batch-64 study, seed 2027090701. White uses $T = 500$ and Flicker $T = 150$; each row shows indices 0–7. Labels report actual network evaluations; full-set FID is in Fig. 3.

U Additional mechanism and training controls

Mechanism interventions. The statistical grid of Eq. 48 improves Flicker $T = 500$ FID-50k from 8.123 ± 0.110 to 7.927 ± 0.142 in a separate confirmation, but worsens both noise types at $T = 150$ (Appendix O). A terminal-prior intervention replaces $\mathcal{N}(0, \Sigma)$ by a zero-mean Gaussian of covariance $\bar{\alpha}_T S + (1 - \bar{\alpha}_T)\Sigma$, keeping the model and sampler fixed. At $T = 150$, uniform/$w = 1$, new FID-50k confirmation worsens white from 14.503 to 37.806 and Flicker from 6.445 to 7.992 (three-seed means; Appendix H). Second-order matching alone is not a sufficient correction for the learned sampler. Direct field and split-step probes delimit the tested approximation: Flicker does not uniformly reduce the measured finite-perturbation responses or self-consistency defects (Appendices Q and P).

Training design and matched-terminal control. In a three-seed 2×2 covariance/loss experiment at $T = 150$ and NFE40, mean FIDs are 15.089 (white/pixel loss), 15.550 (white/Flicker metric), 24.339 (colored/pixel loss), and 6.856 (colored/Flicker metric). Covariance and whitening work together in this recipe; the metric also changes gradient scale. A separate schedule control resamples a $T = 500$ cumulative log-alpha path to 150 steps while preserving terminal SNR. It uses full ancestral sampling with CFG $w = 1.8$: actual NFE300 at $T = 150$ and NFE1000 for the $T = 500$ anchor. The three seed values and training runs differ from the main DDIM study; each uses 90,000 updates and global batch 128, with FID-50k at the final EMA checkpoint. Flicker improves matched-$T = 150$ FID from 15.39 ± 0.32 to 13.58 ± 0.29, but does not match the white $T = 500$ anchor 10.73 ± 0.09. This anchor has its own fixed ancestral protocol and is not the independently tuned $T = 500$ DDIM or DPM++ comparison. The matched-terminal gain is more modest than the main $T = 150$ gain, but their difference does not identify the fraction explained by terminal mismatch. Prediction-target and EqualSNR-style controls are in Appendix S.