WHAT SHOULD A DIFFUSION U-NET PRESERVE? PREDICTIVE SUFFICIENCY AT MULTISCALE INTERFACES

Anonymous authors
Paper under double-blind review

ABSTRACT

Downsampling in diffusion U-Nets is typically treated as a geometric compression operation. We instead view each resolution transition as a target-specific predictive information interface: a representation is useful only insofar as it preserves the information needed for its assigned denoising target. We formalize this idea through a predictive sufficiency gap and derive exact conditions for linear Gaussian interfaces. For rank-matched Gaussian measurements, sufficient geometry generically varies with diffusion SNR, yielding a conditioned-versus-static separation. A tractable Haar model then shows that a constrained local three-tap pre-projection transform can exactly realize this SNR-dependent geometry at fixed retained dimension and yields a strict pre-versus-post separation. In a controlled CIFAR-10 diffusion U-Net interface, correct time conditioning and access to fine-space directions improve held-out prediction, and the learned retained measurement geometry varies with diffusion time. A covariance-defined Gaussian-linear diagnostic further shows that, over the first 15 of 20 evaluation timesteps, the learned Time Pre measurement is closer than Fixed to the surrogate sufficient geometry under both surrogate PSG and subspace distance; the ordering reverses in the five lowest-SNR timesteps as that surrogate geometry approaches the fixed Haar subspace. In an end-to-end diffusion U-Net, matched fine-space access produces a small denoising advantage concentrated in the lowest-noise regime, while a DDPM generation evaluation does not preserve the matched ordering in FID. These results support predictive sufficiency as a principled criterion for analyzing and constraining multiscale diffusion-denoiser interfaces, while showing how interface-level information preservation contributes within the broader dynamics of the generative process.

1 INTRODUCTION

Diffusion models learn denoisers over a family of noise levels. Under the DDPM marginal (Ho et al., 2020),

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I),$$

the optimal prediction rule changes with t. Image diffusion models commonly parameterize this rule with a U-Net: an encoder repeatedly transforms and downsamples a fine spatial representation, a bottleneck operates at low resolution, and a decoder reconstructs finer resolutions using both lower-resolution computation and encoder skip connections (Ronneberger et al., 2015; Ho et al., 2020; Dhariwal & Nichol, 2021). The downsampling boundaries are therefore not only computational devices. They determine what information is available to the coarse branch after fine-scale information has been compressed.

We use the operator view of Williams et al. (2023), building on the multiresolution formulation of Falck et al. (2022). At resolution i, a within-resolution encoder map E_i produces $\widetilde{v}_i = E_i(v_i)$, a projection P_{i-1} supplies the next coarser branch, and the decoder combines the lower-resolution prediction with the skip representation. With bottleneck U_0, the corresponding recursion is

$$U_i(v_i) = D_i(U_{i-1}(P_{i-1}\widetilde{v}_i)\,, \widetilde{v}_i)\,, \qquad \widetilde{v}_i = E_i(v_i).$$

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.

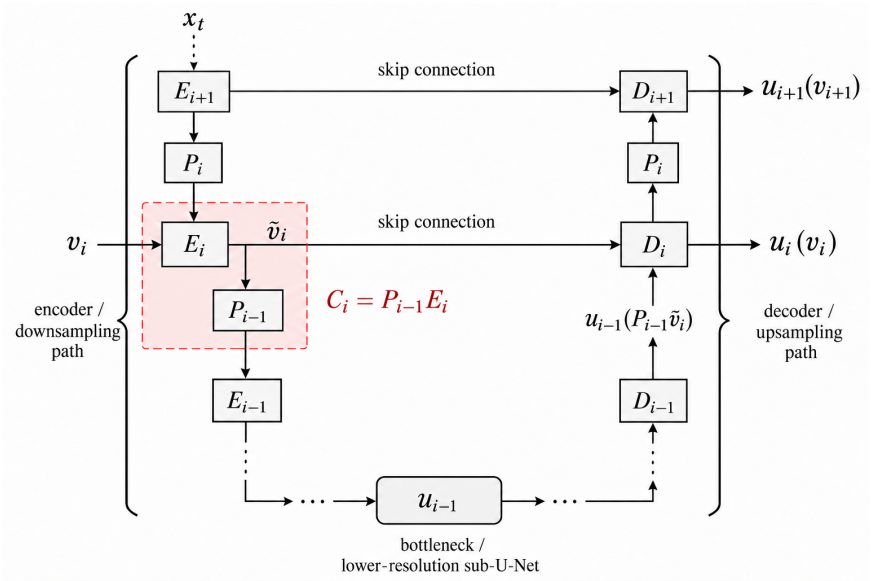

Figure 1: U-Net interface studied in this paper. The within-resolution encoder map E_i acts before the fixed coarse projection P_{i-1}, while $\widetilde{v}_i$ remains separately available to the decoder through the skip pathway. We study the information supplied to the coarse sub-U-Net by $C_i = P_{i-1}E_i$.

Following the notation of Williams et al. (2023), V_i and W_i are the encoder and decoder spaces at resolution i, with $v_i \in V_i$ the representation entering that level. The encoder operator $E_i : V_i \to V_i$ produces the transformed representation $\widetilde{v}_i$, and the projection P_{i-1} maps it to the next coarser encoder space. The lower-resolution U-Net $U_{i-1} : V_{i-1} \to W_{i-1}$ acts on this projected representation. The decoder operator $D_i : W_{i-1} \times V_i \to W_i$ then combines the lower-resolution output with $\widetilde{v}_i$, which serves as the level-i skip input, to produce $U_i(v_i) \in W_i$. The base map $U_0 : V_0 \to W_0$ is the coarsest-resolution component of the recursion. Williams et al. write the decoder as $D_i(w_{i-1} \mid v_i)$; the vertical bar denotes its two deterministic inputs rather than probabilistic conditioning.

For our purposes, the critical object is the encoder-side coarse interface

$$C_i = P_{i-1}E_i,$$

while the skip pathway remains separate; Figure 1 shows this decomposition. In the image specialization below, P_{i-1} is a normalized Haar low-pass projection, which contains the same coarse information as ordinary average pooling up to a fixed nonzero scale. This decomposition isolates a basic asymmetry: E_i can reorganize fine information before projection, whereas deterministic processing that sees only the projected representation cannot recover information already removed by P_{i-1}.

This suggests a target-specific question. A coarse representation need not preserve all information in x_t; it should preserve the information required to predict the denoising target assigned to that scale. We therefore ask: *when is a prescribed multiscale interface predictively sufficient, when can a fixed interface be deficient, and can a constrained transformation before projection repair that deficiency at fixed retained dimension?* Because diffusion changes the prediction problem with t, we additionally ask whether the sufficient retained representation itself should vary with SNR.

Our contributions follow this sequence. First, we define a squared-loss predictive-sufficiency gap (PSG) and derive an exact Gaussian criterion for linear retained measurements. Fixed matched-subspace sufficiency is exceptional in the ambient SPD covariance cone, and for rank-matched measurements this yields generic SNR-dependent sufficient geometry and a conditioned-versus-static separation. Second, in an exactly solvable 8×8 Haar model, a one-parameter local three-tap pre-projection family exactly realizes the required geometry and yields a strict pre-versus-post separation. Third, a controlled CIFAR-10 $32 \times 32 \to 16 \times 16$ experiment tests these structural consequences with matched neural controls: correct diffusion-time alignment improves prediction,

access to fine-space directions before projection lowers held-out error relative to the coarse-only control, and the learned measurement subspace changes with diffusion time. A covariance-matched Gaussian-linear diagnostic further shows that this movement is not arbitrary: over the first 15 of 20 evaluation timesteps, `Time Pre` lies closer than `Fixed` to the surrogate sufficient geometry, with the ordering reversing in the five lowest-SNR timesteps. Our goal is not to propose a new architecture for improving generative quality, but to characterize what information a multiscale interface should preserve for its assigned denoising target and to test the architectural relevance of that criterion.

We also examine the same information-access intervention in an end-to-end diffusion U-Net. The matched denoising gain is small and concentrated almost entirely in the lowest-noise bin, and a generation evaluation does not preserve the matched MSE ordering in FID. The Gaussian construction isolates the information mechanism exactly, while the controlled interface experiment tests the corresponding architectural consequences; neither implies that an interface-level denoising gain must translate into better generation.

2 Related Work

U-Nets and multiresolution structure. The original U-Net couples a contracting encoder, an expanding decoder, and same-resolution skip connections (Ronneberger et al., 2015); related backbones are central to diffusion models (Ho et al., 2020; Dhariwal & Nichol, 2021). Falck et al. (2022) formalize U-Nets through multiresolution function spaces and show that average pooling is conjugate to Haar-coordinate projection. Williams et al. (2023) separate nested input/output spaces, within-resolution encoder transformations, projections, skips, and decoder refinement, and interpret the encoder as learning a favorable representation before projection. We adopt this architectural decomposition. Our additional object is target-specific: we evaluate the representation crossing a prescribed coarse interface by the Bayes risk for its assigned denoising target. Other work studies complementary properties of resolution transitions, including anti-aliased downsampling (Zhang, 2019), content-adaptive downsampling (Hesse et al., 2023), and the roles of backbone and skip features in diffusion U-Nets (Si et al., 2024).

Target-relevant compression and sufficient representations. Information-bottleneck methods seek compressed representations that retain task-relevant information (Tishby et al., 1999; Alemi et al., 2017), while sufficient-dimension-reduction methods seek low-dimensional projections preserving a regression relation (Li, 1991; Cook, 1998). We use standard conditional-mean and Gaussian estimation identities as statistical tools rather than claiming new generic estimation formulas. The distinction here is architectural: the representation is constrained by a U-Net resolution interface, the prediction target is an explicit finite-resolution diffusion target, and input-side retention and output-side projection are allowed to be different objects. The resulting PSG quantifies the squared-loss cost of that prescribed interface.

Frequency structure and diffusion-time adaptation. Wavelet and frequency-aware diffusion models explicitly process multiscale components (Phung et al., 2023; Yuan et al., 2023); diffusion-step-dependent frequency behavior has motivated time-adaptive wavelet gating (Yang et al., 2023), and timestep/layer dependence also appears in learned diffusion features (Luo et al., 2023). Wavelet analyses further characterize scale-dependent score structure (Finn et al., 2026). These works motivate coupling resolution with diffusion state. Our question is narrower: for a fixed-dimensional U-Net coarse branch, which fine-input directions must survive the interface for a specified target, and can a local transform move predictive information into that retained measurement before projection removes it?

3 Predictive Sufficiency at Multiscale Interfaces

3.1 Conditional-Mean Predictive Sufficiency

At an interior diffusion time write $x_t = a_t x_0 + b_t \varepsilon$ with $a_t, b_t > 0$. For a general U-Net transition from scale i to scale $i-1$, let $v_i \in V_i$ denote the random representation entering level i, let the prescribed coarse target be $Y_{i-1} = Q_{i-1}\varepsilon$, and let the representation passed to the coarser branch be

$Z_{i-1} = C_i(v_i)$ with $C_i = P_{i-1}E_i$. The maps P_{i-1} and Q_{i-1} have different roles: P_{i-1} determines what the coarser branch observes, whereas Q_{i-1} determines what it is asked to predict. They coincide only in special matched constructions such as the Haar case below. In the statistical abstraction that follows, the full observation is therefore $X = v_i$. At the first image-resolution transition studied empirically, $v_i = x_t$; the Gaussian specialization below makes the same input-level choice $X = x_t$.

For a square-integrable target Y, full observation X, and deterministic representation $Z = C(X)$, define

$$R_X^\star = \inf_f \mathbb{E}\|Y - f(X)\|^2, \qquad R_Z^\star = \inf_g \mathbb{E}\|Y - g(Z)\|^2, \qquad \Delta^{\mathrm{PSG}}(C) = R_Z^\star - R_X^\star.$$

Because every Z-based predictor factors through X, $\Delta^{\mathrm{PSG}}(C) \geq 0$. The gap is zero precisely when the interface preserves all information needed for the target conditional mean.

Theorem 1 (Conditional-mean identity)**.** *For every square-integrable Y and deterministic measurable $Z = C(X)$,*

$$\Delta^{\mathrm{PSG}}(C) = \mathbb{E}\big\|\mathbb{E}[Y \mid X] - \mathbb{E}[Y \mid Z]\big\|^2.$$

Hence $\Delta^{\mathrm{PSG}}(C) = 0$ iff the two conditional means agree almost surely.

This is target specific: an interface may be sufficient for one target and deficient for another. Appendix A.1 gives the direct L^2 projection proof.

3.2 Gaussian Linear Interfaces

For an exact linear specialization, assume $x_0 \sim \mathcal{N}(0, \Sigma_0)$ and $\varepsilon \sim \mathcal{N}(0, I_d)$ are independent, let $x_t = ax_0 + b\varepsilon$ with $a, b > 0$, and set $\Sigma_t = a^2\Sigma_0 + b^2 I_d$. Let $Y_L = L^\top \varepsilon$ with $L^\top L = I_m$, and let a full-column-rank matrix $M \in \mathbb{R}^{d\times q}$ retain $Z_M = M^\top x_t$. Gaussian conditioning gives the full-input Bayes predictor

$$\mathbb{E}[Y_L \mid x_t] = bL^\top \Sigma_t^{-1} x_t.$$

Here Σ_0 is the clean-data covariance, Σ_t is the covariance of the noisy input x_t, and b is the noise coefficient in $x_t = ax_0 + b\varepsilon$. The columns of L form an orthonormal basis for the m-dimensional target subspace, so $Y_L = L^\top \varepsilon$ is the noise target expressed in those target coordinates. The formula follows from $\mathrm{Cov}(Y_L, x_t) = bL^\top$ and $\mathrm{Cov}(x_t) = \Sigma_t$. To connect this Gaussian specialization to the U-Net notation above, the measurement matrix M represents a linearized coarse interface $C_i = P_{i-1}E_i$, while L represents the output-side target directions specified by Q_{i-1}. Because the expression above conditions on the full noisy input x_t before compression, M does not enter the full-input predictor; it appears only when prediction is restricted to the retained representation $Z_M = M^\top x_t$.

Proposition 1 (Gaussian sufficiency criterion)**.** *With $N = \Sigma_t^{1/2} M$ and $P_N = N(N^\top N)^{-1}N^\top$,*

$$\Delta^{\mathrm{PSG}}_{M,L} = b^2\|(I - P_N)\Sigma_t^{-1/2}L\|_F^2, \qquad \Delta^{\mathrm{PSG}}_{M,L} = 0 \iff \mathrm{col}(\Sigma_t^{-1}L) \subseteq \mathrm{col}(M).$$

Thus sufficiency depends on the retained measurement subspace rather than its coordinates. For non-Gaussian data, $\mathrm{col}(\Sigma_t^{-1}L)$ is not asserted to be the true sufficient subspace. In Section 5.1, we use the same covariance-defined geometry only as a Gaussian-linear surrogate diagnostic for whether a learned measurement moves in the direction predicted by the linear theory. For the matched case $M = L = U$, with projector $P_S = UU^\top$, Proposition 1 becomes

Corollary 1 (Matched-subspace criterion)**.** *At every interior diffusion time,*

$$\Delta^{\mathrm{PSG}}_{S,S} = 0 \iff P_S\Sigma_0(I - P_S) = 0.$$

The fixed interface is therefore sufficient only when the clean covariance does not couple the retained and discarded subspaces. For a fixed rank-r projector P, $0 < r < d$, this alignment is exceptional in the ambient SPD cone:

Proposition 2 (Generic deficiency for a fixed matched subspace)**.** *The set $\{\Sigma \in \mathrm{SPD}_d : P\Sigma(I - P) = 0\}$ has codimension $r(d-r)$, Lebesgue measure zero, and empty relative interior in SPD_d; its complement is open and dense. Every covariance in the complement gives strictly positive matched-interface PSG at every interior diffusion time.*

The Gaussian criterion also yields SNR dependence beyond the finite construction below.

Theorem 2 (SNR-dependent sufficient geometry)**.** *Let $T = \mathrm{col}(L)$, $L^\top L = I_m$, $\Sigma_0 \in \mathrm{SPD}_d$, and $S^\star_\tau = \mathrm{col}((I + \tau\Sigma_0)^{-1}L)$ for $\tau > 0$. For $\tau_1 \neq \tau_2$,*

$$S^\star_{\tau_1} = S^\star_{\tau_2} \iff T \text{ is } \Sigma_0\text{-invariant} \iff P_T\Sigma_0(I - P_T) = 0.$$

Hence, outside the codimension-$m(d-m)$ set of Proposition 2, distinct SNRs require distinct unique zero-PSG rank-m subspaces; over any finite set of at least two SNRs with positive weights, time-dependent measurements can attain zero aggregate PSG whereas every static rank-m measurement has positive aggregate PSG.

This is an ambient-SPD statement for a fixed target subspace; structured covariance families require separate analysis. Proofs are in Appendix A. The same setup also gives the coarse-target reduction below.

Proposition 3 (Coarse-target reduction)**.** *Let $W_{i-1} \subseteq W_i$ be nested closed target spaces with orthogonal projectors Q_{i-1} and Q_i. For every W_{i-1}-valued predictor $\widehat{y}$,*

$$\|Q_i\varepsilon - \widehat{y}\|^2 = \|Q_{i-1}\varepsilon - \widehat{y}\|^2 + \|(Q_i - Q_{i-1})\varepsilon\|^2.$$

Consequently, once two realized interface/predictor pathways are fixed, their pointwise loss difference, and hence their expected risk difference, is identical whether evaluated against $Q_{i-1}\varepsilon$ or the pre-refinement target $Q_i\varepsilon$. The same conclusion holds for infimal risks when the compared conditions use a common fixed predictor class valued in W_{i-1}.

Thus a controlled coarse-target comparison isolates exactly the component of pre-refinement squared loss that can differ across interface conditions; the omitted finer-scale residual contributes the same additive energy to every condition. The proof is given in Appendix A.5.

4 SNR-Dependent Pre-Projection Repair

Theorem 2 does not show that a local architecture-compatible operator can realize the required subspace. We now give an exact finite Haar realization. Consider a scalar 8×8 torus. Let $P \in \mathbb{R}^{8\times 4}$ be normalized pairwise Haar synthesis, let U_8 denote the cyclic shift, set $B = P \otimes P \in \mathbb{R}^{64\times 16}$, and take the target $Y = B^\top\varepsilon$. For the explicit SPD covariance $\Sigma_\star$ defined in Appendix A, use the horizontal three-tap family

$$A_\rho = K_\rho \otimes I_8, \qquad K_\rho = I_8 + \rho(U_8 + U_8^\top), \qquad M_\rho = A_\rho^\top B.$$

Equivalently, K_ρ is the periodic one-dimensional three-tap filter with kernel $[\rho, 1, \rho]$, so $\{A_\rho\}$ is a one-parameter local filter family. Here $\rho = 0$ is the fixed Haar interface and the only rank-deficient value, $\rho = -1/2$, is excluded; write $\mathcal{D} = \mathbb{R} \setminus \{-1/2\}$ for the legal parameter domain. With SNR $\tau = a_t^2/b_t^2 > 0$ and the invertibly rescaled observation $z_\tau = x_t/b_t = \sqrt{\tau}\, x_0 + \varepsilon$, let $R(\rho, \tau)$ be the Bayes risk from $M_\rho^\top z_\tau$ and $R^\star_X(\tau)$ the full-input risk.

Theorem 3 (Unique exact repair within the three-tap family)**.** *For every $\tau > 0$, among $\rho \in \mathcal{D}$,*

$$\rho^\star(\tau) = \frac{8\tau}{7 + 25\tau}, \qquad R(\rho, \tau) = R^\star_X(\tau) \iff \rho = \rho^\star(\tau).$$

Thus fixed Haar is strictly deficient. Moreover $d\rho^\star/d\tau = 56/(7 + 25\tau)^2 > 0$, and distinct SNRs induce distinct retained measurement subspaces.

The theorem gives an exact existence result: a 16-dimensional local measurement can retain all information needed for the 16-dimensional target conditional mean, but the sufficient measurement changes with SNR.

Corollary 2 (Conditioned versus static in the three-tap family)**.** *For any finite set of at least two distinct SNRs τ_k with positive weights w_k,*

$$\inf_{\rho\in\mathcal{D}} \sum_k w_k R(\rho, \tau_k) > \sum_k w_k \inf_{\rho\in\mathcal{D}} R(\rho, \tau_k).$$

Corollary 3 (Pre versus post). *Let $Z_\tau = B^\top z_\tau$. Any deterministic post-processing $\psi_\tau(Z_\tau)$ has Bayes risk at least that available from Z_τ, whereas the optimal legal pre-projection measurement reaches full-input Bayes risk. Hence*

$$R^\star_{\mathrm{pre}}(\tau) = R^\star_X(\tau) < R^\star_{\mathrm{post}}(\tau) = R(0,\tau).$$

Proposition 4 in Appendix A shows that the closed-form transform $\rho^\star(\tau)$ retains a uniformly positive advantage over fixed Haar on compact interior SNR intervals throughout an open SPD neighborhood of $\Sigma_\star$.

5 Empirical Evaluation

5.1 Controlled Interface Experiment

Experimental setup. We isolate the first CIFAR-10 $32 \times 32 \to 16 \times 16$ transition and predict the normalized Haar-coarse noise target $Y_c = P_c\varepsilon$. All conditions share the coarse predictor architecture, target, optimizer and schedule, common predictor initialization, 200,000-step budget, checkpoint-selection rule, and held-out evaluation banks. The `Time Pre` condition applies a true-time-conditioned circular 3×3 adapter A_t before P_c. We compare it with `Null-Time Pre`, which uses the same 405-parameter adapter family but conditions it on an independent timestep; `Time Coarse-Only Pre`, which uses the same true-time adapter after restricting its fine-grid input to the fixed coarse subspace $\Pi_c = P_c^\top P_c$; and `Time Post`, which applies the matched time-conditioned family after projection. Static pre/post and fixed baselines are reported in Appendix B. Evaluation uses 20 diffusion-time bins, 5,000 held-out images, two paired noise draws per image and bin, and 10,000 paired image-cluster bootstrap replicates.

To compare the learned measurement geometry with Proposition 1, we additionally estimate the centered clean-image covariance $\widehat{\Sigma}_0$ from the 40,000-image training split and set $\widehat{\Sigma}_t = a_t^2\widehat{\Sigma}_0 + b_t^2 I$. With $L = P_c^\top$, define the Gaussian-linear surrogate sufficient subspace

$$S_t^{\mathrm{GL}} = \mathrm{col}(\widehat{\Sigma}_t^{-1} L).$$

For a measurement matrix M, we evaluate both the covariance-matched surrogate PSG obtained by substituting $\widehat{\Sigma}_t$ into Proposition 1 and the normalized projector distance between $\mathrm{col}(M)$ and S_t^{GL}. This is a Gaussian-linear diagnostic on non-Gaussian CIFAR-10, not an estimate of its true Bayes-sufficient subspace.

Table 1: Controlled held-out contrasts. $\Delta\mathrm{MSE} = \mathrm{MSE}(\texttt{Time Pre}) - \mathrm{MSE}(\text{control})$; negative values favor `Time Pre`.

Control	ΔMSE	Rel. reduction
`Null-Time Pre`	-0.000614348080	0.783%
`Time Coarse-Only Pre`	-0.000687382769	0.875%
`Time Post`	-0.000668216007	0.851%

Control	Paired 95% interval
`Null-Time Pre`	[-0.000633244826, -0.000595012587]
`Time Coarse-Only Pre`	[-0.000706383518, -0.000668755042]
`Time Post`	[-0.000688562824, -0.000647756871]

Correct diffusion state matters. `Time Pre` reaches equal-bin MSE 0.077840777439 versus 0.078455125519 for the equal-capacity `Null-Time Pre`; the paired difference is -0.000614348080 (0.783%), and its binwise point estimate is lower in all 20 bins. Because placement, adapter family, capacity, predictor, target, and optimization are matched, this comparison isolates the value of aligning the adapter with the actual diffusion state more cleanly than the static-versus-conditioned comparison. `Time Pre` also has lower binwise point estimates than `Static Pre` in all 20 bins (Appendix B).

Access before projection matters. Restricting the same true-time fine-grid adapter to $\Pi_c x_t$ raises equal-bin MSE from 0.077840777439 to 0.078528160208, a paired difference of -0.000687382769 (0.875%). The full-fine-input condition has lower binwise point estimates in 17 of 20 bins. This is the primary information-boundary control: because $\Pi_c = P_c^\top P_c$, $P_c A_t \Pi_c x_t = (P_c A_t P_c^\top)(P_c x_t) = \psi_t(P_c x_t)$, so `Time Coarse-Only Pre` factors exactly through the fixed coarse representation while retaining the same true-time-conditioned fine-grid adapter parameterization. `Time Post` is therefore treated only as a secondary architectural-placement comparison; moving the convolution across a resolution change also changes its effective geometry. An operator diagnostic confirms nonfactorization through the fixed coarse representation for `Time Pre`:

$$r_t = \frac{\|P_c A_t (I - \Pi_c)\|_F}{\|P_c A_t\|_F} = 0.436\text{–}0.441.$$

Fine-space access alone is not sufficient: `Null-Time Pre` can have residuals as large as 0.709 while retaining substantially higher error.

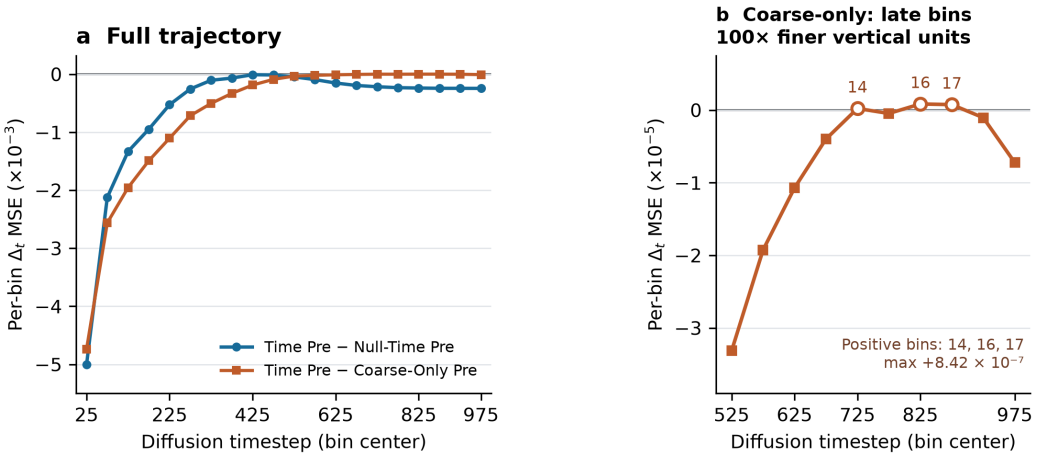

Figure 2: Per-bin $\Delta_t = \mathrm{MSE}(\texttt{Time Pre}, t) - \mathrm{MSE}(\mathrm{control}, t)$. The point estimate for correct time conditioning is negative in every bin; the matched coarse-only point estimate is negative in 17 of 20 bins. Aggregate paired intervals are in Table 1.

The retained measurement changes with diffusion time. For $M_t = (P_c A_t)^\top$, let Q_t span $\mathrm{col}(M_t)$ and write $\Pi_t^{\mathrm{meas}} = Q_t Q_t^\top$. The normalized projector distance $d(t,s) = \|\Pi_t^{\mathrm{meas}} - \Pi_s^{\mathrm{meas}}\|_F / \sqrt{2r}$ is invariant to invertible coordinate changes within a retained subspace. `Time Pre` reaches maximum pairwise distance 0.138645 and first-to-last distance 0.138515, whereas `Static Pre` is exactly fixed and a `Null-Time Pre` conditioning-input sweep reaches only 0.000699. Thus the learned operator is not merely reparameterizing one fixed retained subspace. Complete pairwise and principal-angle diagnostics are reported in Appendix B.6.

Alignment with the Gaussian-linear surrogate geometry. For all 15 timesteps $t = 25, 75, \ldots, 725$, `Time Pre` has both lower covariance-matched surrogate PSG and smaller projector distance to S_t^{GL} than `Fixed`; both orderings reverse for $t = 775, \ldots, 975$. The same sign pattern holds under two disjoint 20,000-image covariance estimates (Figure 3; Appendix B.7). The late reversal is consistent with $S_t^{\mathrm{GL}} \to \mathrm{col}(L)$, the `Fixed` Haar subspace, as $a_t^2 \to 0$.

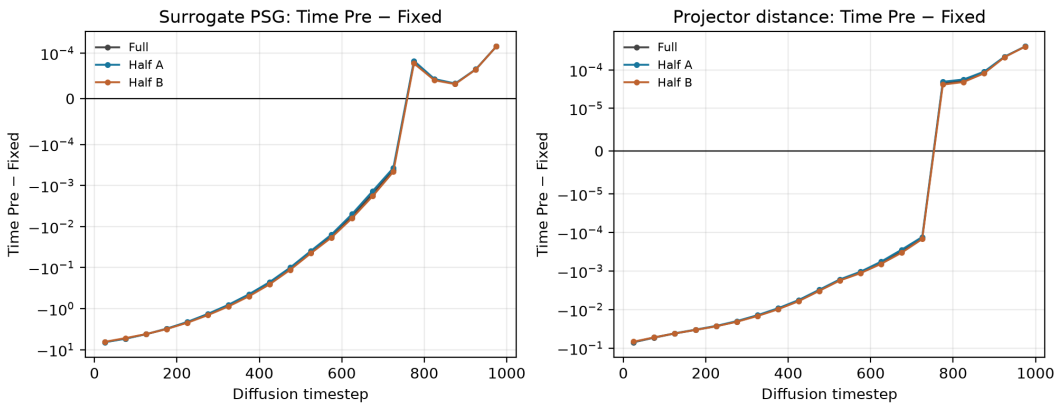

Figure 3: Gaussian-linear surrogate alignment. Differences are `Time Pre` minus `Fixed`; negative values favor `Time Pre`. Left: covariance-matched surrogate PSG. Right: normalized distance to S_t^{GL}. Curves use the full covariance and two disjoint half-sample estimates.

5.2 End-to-End U-Net External Validation

As a secondary external-validity test, we examine the same information-access intervention in a complete, jointly trained CIFAR-10 diffusion U-Net on the full-resolution ε target. At the first $32\times32 \to 16\times16$ transition, `Fixed` uses the prescribed projection, `Time Pre` applies a true-time-conditioned fine-grid adapter before it, and `Time Coarse-Only Pre` uses the same adapter architecture after restricting its input to the fixed Haar coarse subspace. The two adaptive conditions match capacity, conditioning, fine-grid placement, convolution geometry, downstream U-Net, target, and training budget; they differ in whether the adapter can access fine-space feature directions. Appendix C gives the full protocol.

Denoising evaluation. `Time Pre` improves held-out equal-bin ε-MSE over `Fixed` by 1.91×10^{-5} (0.0643%) and over the matched `Time Coarse-Only Pre` control by 7.74×10^{-6} (0.0260%); both paired intervals exclude zero (Table 6). The matched point estimate is lower in only 4 of 20 bins, and the lowest-noise bin contributes approximately 100.5% of the net equal-bin improvement. Across bins 1–19 the mean matched difference is effectively zero.

The smaller matched gain than in the controlled interface experiment is consistent with the end-to-end setting: the intervention acts on a learned h_{32} representation rather than directly on x_t, while fine-resolution information remains available through the skip pathway and the surrounding network is jointly optimized. These alternative routes can reduce the marginal dependence of the full-resolution predictor on information preserved at this single coarse interface.

Generation evaluation. With shared initial x_T and paired reverse-step noise, 1,000-step ancestral DDPM sampling gives a different ordering under FID. Table 2 reports the three full-sample estimates.

Table 2: End-to-end generation results under the common ancestral-DDPM sampling and FID protocol.

Condition	FID ↓
`Fixed`	18.637
`Time Pre`	16.174
`Time Coarse-Only Pre`	15.417

`Time Pre` improves FID over `Fixed` by 2.463 points but is 0.757 points worse than `Time Coarse-Only Pre`. A ten-block paired grouped-jackknife gives approximate intervals $[-2.96, -1.97]$ for `Time Pre-Fixed` and $[+0.28, +1.23]$ for `Time Pre-Time`

`Coarse-Only Pre`; these are finite-sample stability diagnostics conditional on the generated feature arrays and fixed real reference, not population-level FID inference.

The different MSE and FID orderings are compatible with the predictive-sufficiency framework because the two endpoints characterize different objects. Predictive sufficiency concerns the conditional-mean information available to a denoising predictor under squared loss at a given noisy state, whereas generation composes learned reverse transitions over the entire diffusion trajectory. Consequently, prediction errors can have different consequences at different diffusion times and can propagate through subsequent reverse steps. This distinction is already visible in the original DDPM formulation, where changing the relative weighting of denoising errors across timesteps materially changes sample quality (Ho et al., 2020). In addition, the theory characterizes Bayes-optimal risk for a representation, while the trained U-Net is a finite-capacity, finitely optimized predictor whose use of that representation need not attain the Bayes optimum.

FID introduces a further distinction: it compares generated-image feature distributions after the full reverse chain rather than the denoising loss evaluated at individual noisy inputs. Thus even a genuine interface-level denoising advantage need not induce the same ordering under FID. We therefore interpret predictive sufficiency as a target-aware principle for designing multiscale diffusion denoisers: it identifies what information a resolution transition should preserve for its prescribed prediction target, but it is not a standalone sufficient condition for downstream generative quality.

6 Discussion and Limitations

The theoretical results have different scopes. The PSG identity is general under squared loss, whereas Proposition 1 and Theorem 2 are finite-dimensional Gaussian linear results. Together with Proposition 2, they make SNR-dependent rank-matched sufficient geometry generic in the ambient SPD covariance cone, not in natural-image distributions. Section 4 gives the stronger architecture-constrained result only for the finite Haar/three-tap construction, with local covariance robustness from Proposition 4.

The controlled experiment probes these architectural consequences rather than the Bayes quantities themselves. `Time Coarse-Only Pre` factors exactly through $P_c x_t$ and is therefore the primary information-boundary control; the literal pre/post comparison is secondary because its effective convolution geometry changes across resolutions. The Gaussian-linear surrogate provides supporting geometric evidence but does not establish the true Bayes-sufficient subspace of CIFAR-10, Gaussian validity, Bayes-risk attainment, or unique mediation of the neural gain.

The empirical scope is limited to one controlled CIFAR-10 interface and one full-U-Net architecture. The matched full-U-Net gain is small and concentrated almost entirely in one low-noise bin, consistent with alternate learned and skip pathways reducing the marginal importance of a single coarse interface. FID is pipeline-specific, the grouped-jackknife intervals are finite-sample stability diagnostics, and likelihood is not evaluated. The matched MSE/FID reversal shows that interface-level predictive adequacy does not by itself determine generation quality.

7 Conclusion

A diffusion U-Net resolution transition can be viewed as a target-specific predictive interface. In the Gaussian linear setting, rank-matched sufficient geometry generically varies with SNR, while a finite Haar construction shows that a constrained local pre-projection transform can realize this geometry exactly. Controlled CIFAR-10 experiments support the corresponding information-access mechanism, but its effect is strongly attenuated in a full U-Net and does not preserve the matched FID ordering; predictive sufficiency is therefore a denoising-interface criterion rather than a surrogate for generation quality.

AI use statement

Generative AI tools assisted with theoretical development, proof checking, experimental design and analysis, implementation, and manuscript editing. Generative AI was not used to generate the

reported experimental data. The authors reviewed all AI-assisted content and take responsibility for the final paper.

Reproducibility statement

The assumptions and complete proofs for the theoretical results are provided in Appendix A. The controlled interface protocol, condition definitions, optimization and checkpoint-selection rules, aggregate and per-bin results, bootstrap procedure, operator diagnostics, and Gaussian-linear surrogate-alignment and covariance-split robustness analyses are provided in Appendix B. Full-U-Net architecture and intervention details, held-out evaluation, ancestral-DDPM/FID protocol, and paired grouped-jackknife calculations are provided in Appendix C.

References

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. In *International Conference on Learning Representations*, 2017.

R. Dennis Cook. *Regression Graphics: Ideas for Studying Regressions Through Graphics*. Wiley, New York, NY, 1998.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Fabian Falck, Christopher Williams, Dominic Danks, George Deligiannidis, Christopher Yau, Chris Holmes, Arnaud Doucet, and Matthew Willetts. A multi-resolution framework for u-nets with applications to hierarchical vaes. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Emma Lucia Byrnes Finn, Binxu Wang, T. Anderson Keller, and Demba E. Ba. Where the score lives: A wavelet view of diffusion. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300, pp. 2395–2403. PMLR, 2026.

Robin Hesse, Simone Schaub-Meyer, and Stefan Roth. Content-adaptive downsampling in convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 4544–4553, 2023.

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

Ker-Chau Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

Grace Luo, Lisa Dunlap, Dong Huk Park, Aleksander Holynski, and Trevor Darrell. Diffusion hyperfeatures: Searching through time and space for semantic correspondence. In *Advances in Neural Information Processing Systems*, 2023.

Hao Phung, Quan Dao, and Anh Tran. Wavelet diffusion models are fast and scalable image generators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10199–10208, 2023.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pp. 234–241. Springer, 2015.

Chenyang Si, Ziqi Huang, Yuming Jiang, and Ziwei Liu. FreeU: Free lunch in diffusion U-Net. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4733–4743, 2024.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 368–377, 1999.

Christopher Williams, Fabian Falck, George Deligiannidis, Chris Holmes, Arnaud Doucet, and Saifuddin Syed. A unified framework for u-net design and analysis. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Xingyi Yang, Daquan Zhou, Jiashi Feng, and Xinchao Wang. Diffusion probabilistic model made slim. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22552–22562, 2023.

Xin Yuan, Linjie Li, Jianfeng Wang, Zhengyuan Yang, Kevin Lin, Zicheng Liu, and Lijuan Wang. Spatial-frequency U-Net for denoising diffusion probabilistic models. arXiv preprint arXiv:2307.14648, 2023.

Richard Zhang. Making convolutional networks shift-invariant again. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 7324–7334, 2019.

A PROOFS

A.1 PROOF OF THEOREM 1, PROPOSITION 1, AND COROLLARY 1

Notation for this proof. The risks $R_X^\star$ and $R_Z^\star$ are those defined in Section 3. For a random element X, $\sigma(X)$ denotes the σ-algebra generated by X, and $L^2(\sigma(X))$ denotes the square-integrable random vectors measurable with respect to that σ-algebra. Conditional expectation $\mathbb{E}[Y \mid X]$ is shorthand for $\mathbb{E}[Y \mid \sigma(X)]$ and is the orthogonal projection of Y onto $L^2(\sigma(X))$. Norms in this subsection are Euclidean norms in the target space, and equalities between conditional expectations are understood almost surely. In the Gaussian part below, $\mathrm{Cov}(A, B)$ denotes cross-covariance with the displayed orientation, $\mathrm{col}(M)$ is the column space of M, and I_k is the $k \times k$ identity.

We first connect the risk definitions directly to conditional expectation. For any square-integrable predictor $f(X)$, the L^2 projection property gives the Pythagorean identity

$$\mathbb{E}\|Y - f(X)\|^2 = \mathbb{E}\|Y - \mathbb{E}[Y \mid X]\|^2 + \mathbb{E}\|\mathbb{E}[Y \mid X] - f(X)\|^2.$$

Consequently the infimum in the definition of $R_X^\star$ is attained by $f_X^\star(X) = \mathbb{E}[Y \mid X]$, and $R_X^\star = \mathbb{E}\|Y - \mathbb{E}[Y \mid X]\|^2$. Applying the same argument to predictors measurable with respect to Z gives $g_Z^\star(Z) = \mathbb{E}[Y \mid Z]$ and $R_Z^\star = \mathbb{E}\|Y - \mathbb{E}[Y \mid Z]\|^2$.

Because $Z = C(X)$ is deterministic, $\sigma(Z) \subseteq \sigma(X)$. Hence $\mathbb{E}[Y \mid X] - \mathbb{E}[Y \mid Z]$ is $\sigma(X)$-measurable, while the residual $Y - \mathbb{E}[Y \mid X]$ is orthogonal to every element of $L^2(\sigma(X))$. Therefore

$$Y - \mathbb{E}[Y \mid Z] = \big(Y - \mathbb{E}[Y \mid X]\big) + \big(\mathbb{E}[Y \mid X] - \mathbb{E}[Y \mid Z]\big)$$

is an orthogonal decomposition. Taking squared norms and expectations, and substituting the two risk identities just derived, yields

$$R_Z^\star - R_X^\star = \mathbb{E}\|\mathbb{E}[Y \mid X] - \mathbb{E}[Y \mid Z]\|^2.$$

The right-hand side is nonnegative and vanishes exactly when the two conditional means agree almost surely. This proves Theorem 1 and also makes explicit why the two optimization problems in Section 3 are generally not equivalent: the Z-based predictor class is the subclass of X-based predictors that factor through C.

For the Gaussian specialization, $\mathrm{Cov}(x_t) = \Sigma_t$ and $\mathrm{Cov}(Y_L, x_t) = bL^\top$. Gaussian conditioning therefore gives $\mathbb{E}[Y_L \mid x_t] = bL^\top \Sigma_t^{-1} x_t$ and

$$R_X^\star = \mathrm{tr}\left[I_m - b^2 L^\top \Sigma_t^{-1} L\right].$$

For $Z_M = M^\top x_t$, full column rank of M and $\Sigma_t \succ 0$ imply $M^\top \Sigma_t M \succ 0$. Moreover, $\mathrm{Cov}(Y_L, Z_M) = bL^\top M$. Therefore

$$R_{Z_M}^\star = \mathrm{tr}\left[I_m - b^2 L^\top M (M^\top \Sigma_t M)^{-1} M^\top L\right].$$

Subtracting the two Bayes risks and setting $N = \Sigma_t^{1/2} M$, we have

$$M(M^\top \Sigma_t M)^{-1} M^\top = \Sigma_t^{-1/2} P_N \Sigma_t^{-1/2},$$

we obtain

$$\Delta_{M,L}^{\mathrm{PSG}} = b^2\, \mathrm{tr}\left[L^\top \Sigma_t^{-1/2} (I - P_N) \Sigma_t^{-1/2} L\right] = b^2 \|(I - P_N)\Sigma_t^{-1/2} L\|_F^2.$$

The gap is zero exactly when $\Sigma_t^{-1/2} L \in \mathrm{col}(N)$, equivalently

$$\mathrm{col}(\Sigma_t^{-1} L) \subseteq \mathrm{col}(M).$$

For $M = L = U$, both sides have dimension $\dim S$, so zero gap is equivalent to invariance of S under Σ_t^{-1}, hence under Σ_t. In an orthogonal decomposition $S \oplus S^\perp$, this is

$$U^\top \Sigma_t U_\perp = 0.$$

Since

$$\Sigma_t = a^2\Sigma_0 + b^2 I, \qquad U^\top I U_\perp = 0, \qquad a > 0,$$

the condition is equivalent to

$$U^\top \Sigma_0 U_\perp = 0 \iff P_S \Sigma_0 (I - P_S) = 0.$$

A.2 Proof of Proposition 2: generic deficiency

Choose an orthogonal matrix U whose first r columns span $\mathrm{range}(P)$. In these coordinates,

$$U^\top P U = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Every symmetric A has a unique block representation

$$U^\top A U = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{pmatrix},$$

where $A_{12} \in \mathbb{R}^{r\times(d-r)}$. Direct multiplication gives

$$U^\top P A (I - P) U = \begin{pmatrix} 0 & A_{12} \\ 0 & 0 \end{pmatrix}.$$

Define

$$\mathcal{L}_P = \{A \in \mathrm{Sym}_d : PA(I-P) = 0\}, \qquad \mathcal{A}_P = \mathcal{L}_P \cap \mathrm{SPD}_d\,.$$

The condition $PA(I-P) = 0$ is equivalent to $A_{12} = 0$, which imposes exactly $r(d-r)$ independent linear constraints. Hence $\mathcal{L}_P$ has codimension $r(d-r) > 0$ in Sym_d. Because SPD_d is open in Sym_d, $\mathcal{A}_P$ is relatively open in $\mathcal{L}_P$. It therefore has the same dimension as $\mathcal{L}_P$, so its codimension relative to the ambient SPD manifold is also $r(d-r)$.

The identity belongs to $\mathcal{A}_P$, so the intersection is nonempty. A proper linear subspace of a finite-dimensional Euclidean space has Lebesgue measure zero and empty interior. Since the SPD cone is open, $\mathcal{A}_P$ has measure zero and empty relative interior in that cone.

The map $A \mapsto PA(I-P)$ is continuous, making $\mathcal{L}_P$ closed. Thus the complement of $\mathcal{A}_P$ is relatively open.

For density, choose $\Sigma \in \mathcal{A}_P$, a unit $u \in \mathrm{range}(P)$, and a unit $v \in \ker(P)$. Let

$$E = uv^\top + vu^\top.$$

Then $PE(I-P) = uv^\top \neq 0$. Whenever

$$0 < |\eta| < \frac{\lambda_{\min}(\Sigma)}{\|E\|_{\mathrm{op}}},$$

we have $\Sigma + \eta E \succ 0$, while

$$P(\Sigma + \eta E)(I - P) = \eta u v^\top \neq 0.$$

These deficient perturbations can be arbitrarily small, so the complement is dense. The matched-subspace criterion from Appendix A.1 turns failure of the block condition into positive PSG at every interior diffusion time. □

Proof of Theorem 2. Let $R_\tau = (I + \tau\Sigma_0)^{-1}$, so $S^\star_\tau = R_\tau T$. If T is invariant under Σ_0, it is invariant under every R_τ, hence $S^\star_\tau = T$ for all τ.

Conversely, suppose $S^\star_{\tau_1} = S^\star_{\tau_2}$ for $\tau_1 \neq \tau_2$. Multiplying by $R^{-1}_{\tau_2}$ shows that T is invariant under

$$B = R^{-1}_{\tau_2} R_{\tau_1} = (I + \tau_2\Sigma_0)(I + \tau_1\Sigma_0)^{-1}.$$

If λ is an eigenvalue of Σ_0, the corresponding eigenvalue of B is

$$f(\lambda) = \frac{1 + \tau_2\lambda}{1 + \tau_1\lambda}.$$

Because $\tau_1 \neq \tau_2$, f is strictly monotone on $(0, \infty)$ and therefore injective on the spectrum of Σ_0. Thus B and Σ_0 have exactly the same eigenspaces. Since B is symmetric, a B-invariant subspace is a direct sum of subspaces of its eigenspaces, and is therefore also invariant under Σ_0. For symmetric Σ_0, invariance of T is equivalent to $P_T\Sigma_0(I - P_T) = 0$. This proves Theorem 2; Proposition 2 supplies the genericity statement.

For a rank-m measurement, Proposition 1 gives zero PSG at τ iff its column space equals $S^\star_\tau$. Under the non-invariance assumption, Theorem 2 makes these zero sets disjoint for any two distinct SNRs. The weighted aggregate PSG is continuous on the compact Grassmannian $\mathrm{Gr}(m, d)$, so its static minimum is attained; because no static subspace makes every nonnegative term zero, that minimum is strictly positive. Choosing the measurement subspace separately as $S^\star_{\tau_k}$ makes every term zero, proving the final statement of Theorem 2. □

A.3 Finite construction and proof of Theorem 3 and Corollaries 2–3

For completeness, define the construction on a real scalar 8×8 torus. Let U_8 be the cyclic shift and

$$T = \frac{1}{2}(U_8 + U_8^\top), \qquad H_1 = \frac{1}{2}I_8 + \frac{1}{4}T, \qquad W = H_1^{-1}, \qquad S_0 = 2W - I_8,$$

with clean covariance $\Sigma_\star = S_0 \otimes I_8$. Let $P \in \mathbb{R}^{8\times 4}$ be normalized pairwise Haar synthesis and $B = P \otimes P$. The target is $Y = B^\top\varepsilon$. For scalar ρ, define

$$K_\rho = I_8 + \rho(U_8 + U_8^\top), \qquad A_\rho = K_\rho \otimes I_8, \qquad M_\rho = A_\rho^\top B = (K_\rho P) \otimes P.$$

The baseline fixed Haar interface is $\rho = 0$, the legal family is $\mathcal{D} = \mathbb{R}\backslash\{-1/2\}$, and after normalizing by b_t the interior SNR is $\tau = a_t^2/b_t^2 > 0$ with $z_\tau = \sqrt{\tau}\, x_0 + \varepsilon$. Write $R(\rho, \tau)$ for the Bayes risk of predicting Y from $M_\rho^\top z_\tau$ and $R^\star_X(\tau)$ for the full-input Bayes risk.

A.3.1 Rank and covariance validity

For

$$M_\rho = (K_\rho P) \otimes P,$$

the four coarse-mode eigenvalues of $(K_\rho P)^\top (K_\rho P)$ are

$$(1 + 2\rho)^2, \qquad 2\rho^2 + 2\rho + 1, \qquad 1, \qquad 2\rho^2 + 2\rho + 1.$$

Because

$$2\rho^2 + 2\rho + 1 = 2(\rho + 1/2)^2 + 1/2 > 0,$$

rank is lost only at $\rho = -1/2$. The eigenvalues of H_1 lie in $[1/4, 3/4]$, so the eigenvalues of $W = H_1^{-1}$ lie in $[4/3, 4]$, and those of $S_0 = 2W - I$ lie in $[5/3, 7]$. Thus $\Sigma_\star \succ 0$.

A.3.2 REDUCTION TO A ONE-DIMENSIONAL ALIAS CONDITION

Let

$$\mathcal{S}_\tau = I_{64} + \tau\Sigma_\star.$$

Gaussian conditioning gives

$$\mathbb{E}[Y \mid z_\tau] = B^\top \mathcal{S}_\tau^{-1} z_\tau.$$

The linear sufficiency criterion implies zero PSG if and only if

$$\mathrm{col}(\mathcal{S}_\tau^{-1} B) = \mathrm{col}(M_\rho),$$

where equality follows because both matrices have rank 16. Write

$$\mathcal{S}_\tau = H_\tau \otimes I_8, \qquad H_\tau = I_8 + \tau S_0.$$

Then the condition reduces to

$$\mathrm{col}(H_\tau^{-1} P) = \mathrm{col}(K_\rho P).$$

In the eight-point Fourier basis, the response of H_τ^{-1} on a mode with $c = \cos(2\pi j/8)$ is

$$f(c) = \frac{2 + c}{\alpha + \beta c}, \qquad \alpha = 2 + 6\tau, \qquad \beta = 1 - \tau.$$

The denominator is positive because $\alpha > |\beta|$. Each coarse Fourier mode lifts into an alias pair with T-eigenvalues c and $-c$. The pairs at $c = \pm 1$ and $c = 0$ satisfy the space condition automatically because of Haar zeros or common scalar action. The nontrivial pairs have

$$c = \pm s, \qquad s = 1/\sqrt{2}.$$

Proportionality of the two retained alias directions is equivalent to

$$f(s)(1 - 2\rho s) - f(-s)(1 + 2\rho s) = 0.$$

Direct simplification gives

$$f(s)(1 - 2\rho s) - f(-s)(1 + 2\rho s) = \frac{2s[8\tau - (7 + 25\tau)\rho]}{\alpha^2 - \beta^2/2}.$$

The denominator is strictly positive, so any sufficient measurement must have

$$\rho = \frac{8\tau}{7 + 25\tau}.$$

Conversely, at this value there is an invertible real symmetric coarse circulant matrix L_τ satisfying

$$H_\tau^{-1} P = K_{\rho^\star(\tau)} P L_\tau.$$

Its coarse Fourier eigenvalues are

$$\ell_0 = \frac{3(7 + 25\tau)}{(3 + 5\tau)(7 + 41\tau)},$$

$$\ell_1 = \ell_3 = \frac{7(7+25\tau)}{49+350\tau+497\tau^2},$$

and

$$\ell_2 = \frac{1}{1+3\tau}.$$

All are positive. Hence the column spaces agree, proving sufficiency. The unique zero of a nonnegative PSG is the unique global risk minimizer. Since $\rho^\star(\tau) > 0$, fixed Haar is strictly deficient.

Differentiation gives strict monotonicity of $\rho^\star$. If distinct SNRs produced the same measurement subspace, the two full-rank measurements would differ only by an invertible coordinate change. Under either fixed law they would then have the same Bayes risk, contradicting uniqueness of the zero-PSG parameter. Thus distinct SNRs have distinct retained measurement subspaces.

A.3.3 Conditioned versus static

For every legal ρ,

$$R(\rho, \tau_k) \geq R_X^\star(\tau_k),$$

with equality only when $\rho = \rho^\star(\tau_k)$. The conditioned policy therefore attains the weighted full-input risk.

To establish strict inequality against the static infimum, we also need attainment. The finite-model risk reduces to

$$R(\rho, \tau) = F(\tau) - \frac{56(\rho+1)^2}{q(\rho, \tau)},$$

where

$$F(\tau) = 16 - \frac{12}{3+5\tau} - \frac{4}{1+3\tau},$$

and

$$q(\rho, \tau) = (14+34\tau)\rho^2 + (14+18\tau)\rho + 7 + 17\tau.$$

The discriminant of this quadratic is

$$-28(71\tau^2 + 50\tau + 7) < 0,$$

so $q > 0$. For $\rho < 0$, direct subtraction shows $R(\rho, \tau) > R(0, \tau)$. Also

$$\frac{\partial R}{\partial \rho} = \frac{112(\rho+1)[(7+25\tau)\rho - 8\tau]}{q(\rho, \tau)^2},$$

which is positive for $\rho \geq 8/25$. Every static parameter can therefore be clipped without increasing risk into the compact interval $[0, 8/25]$. The weighted static objective attains its minimum there.

If the attained static risk equaled the conditioned risk, positive weights and nonnegative pointwise gaps would force the static minimizer to equal $\rho^\star(\tau_k)$ for every k. Strict monotonicity makes this impossible when two SNRs differ. Therefore the static infimum is strictly larger than the conditioned sum of pointwise infima, as stated in Corollary 2.

A.3.4 PRE/POST SEPARATION

Let $Z_\tau = B^\top z_\tau$ and let ψ_τ be any deterministic function of Z_τ. Applying Theorem 1 with full observation Z_τ and representation $\psi_\tau(Z_\tau)$ gives

$$\inf_g \mathbb{E}\|Y - g(\psi_\tau(Z_\tau))\|^2 - \inf_h \mathbb{E}\|Y - h(Z_\tau)\|^2 = \mathbb{E}\|\mathbb{E}[Y \mid Z_\tau] - \mathbb{E}[Y \mid \psi_\tau(Z_\tau)]\|^2 \geq 0.$$

Thus deterministic coarse-only post-processing cannot improve upon the fixed-Haar Bayes risk. The identity map attains equality, so the optimal post-processing risk is $R(0, \tau)$. Theorem 3 gives full-input risk for the optimal legal pre-projection measurement and strict fixed-Haar deficiency, yielding the strict pre/post chain in Corollary 3.

A.4 PROOF OF PROPOSITION 4: LOCAL ROBUSTNESS

For this subsection, let $x_0 \sim \mathcal{N}(0, \Sigma)$ with $\Sigma \in \mathrm{SPD}_{64}$, let $\varepsilon \sim \mathcal{N}(0, I_{64})$ independently, and set

$$z_\tau = \sqrt{\tau}\, x_0 + \varepsilon, \qquad Y = B^\top \varepsilon, \qquad Z_{\rho,\tau} = M_\rho^\top z_\tau.$$

Define $R_\Sigma(\rho, \tau)$ as the Bayes squared risk for predicting Y from $Z_{\rho,\tau}$,

$$R_\Sigma(\rho, \tau) = \inf_g \mathbb{E}\|Y - g(Z_{\rho,\tau})\|^2.$$

Gaussian conditioning gives the closed form

$$R_\Sigma(\rho, \tau) = 16 - \mathrm{tr}\left[B^\top M_\rho \{M_\rho^\top (I + \tau\Sigma) M_\rho\}^{-1} M_\rho^\top B\right].$$

Proposition 4 (Local robustness)**.** *For every nonempty compact $K \subset (0, \infty)$, there exist $\delta > 0$ and $c > 0$ such that $\|\Sigma - \Sigma_\star\|_{\mathrm{op}} < \delta$ implies*

$$R_\Sigma(0, \tau) - R_\Sigma(\rho^\star(\tau), \tau) \geq c \qquad \forall \tau \in K.$$

For each compact $K \subset (0, \infty)$, the proposition gives a sufficiently small SPD neighborhood of $\Sigma_\star$ on which the transform $\rho^\star(\tau)$ remains strictly better than fixed Haar with a uniform positive margin over $\tau \in K$.

Proof. For every legal ρ, M_ρ has full column rank. If $\Sigma \succ 0$ and $\tau > 0$, then

$$M_\rho^\top (I + \tau\Sigma) M_\rho \succ 0.$$

Matrix multiplication, inversion on the SPD cone, and trace are continuous, while M_ρ is polynomial in ρ. Hence $R_\Sigma(\rho, \tau)$ is jointly continuous on its legal domain. Because

$$0 < \rho^\star(\tau) < 8/25,$$

the transform $\rho^\star(\tau)$ stays uniformly away from the rank-deficient value.

Define

$$h(\Sigma, \tau) = R_\Sigma(0, \tau) - R_\Sigma(\rho^\star(\tau), \tau).$$

At $\Sigma_\star$, exact repair and strict fixed-Haar deficiency give

$$h(\Sigma_\star, \tau) > 0 \qquad (\tau > 0).$$

For compact $K \subset (0, \infty)$, continuity yields

$$m = \min_{\tau \in K} h(\Sigma_\star, \tau) > 0.$$

Choose

$$0 < \eta < \lambda_{\min}(\Sigma_\star)/2.$$

Weyl's inequality implies that the closed operator-norm ball

$$\mathcal{B}_\eta = \{\Sigma \in \mathrm{Sym}_{64} : \|\Sigma - \Sigma_\star\|_{\mathrm{op}} \le \eta\}$$

lies inside the SPD cone. In finite dimension $\mathcal{B}_\eta \times K$ is compact, so h is uniformly continuous there. Therefore some $0 < \delta \le \eta$ satisfies

$$\|\Sigma - \Sigma_\star\|_{\mathrm{op}} < \delta \implies \sup_{\tau \in K} |h(\Sigma, \tau) - h(\Sigma_\star, \tau)| < m/2.$$

It follows that

$$h(\Sigma, \tau) \ge m/2$$

for every $\tau \in K$. Taking $c = m/2$ proves the proposition.

□

A.5 Proof of coarse-target reduction: residual-energy identity

Proof of Proposition 3. Set $R_i = Q_i - Q_{i-1}$. Nested orthogonal projectors satisfy

$$Q_i Q_{i-1} = Q_{i-1} Q_i = Q_{i-1}.$$

Therefore

$$R_i = Q_i - Q_{i-1}$$

is self-adjoint and idempotent, hence the orthogonal projector onto

$$H_i = W_i \cap W_{i-1}^{\perp}.$$

For $\widehat{y} \in W_{i-1}$,

$$Q_i \varepsilon - \widehat{y} = (Q_{i-1}\varepsilon - \widehat{y}) + R_i \varepsilon.$$

The first term lies in W_{i-1} and the second in $H_i \subseteq W_{i-1}^{\perp}$. Pythagoras gives the pointwise identity. With finite second moments, taking expectations gives

$$\mathbb{E}\|Q_i\varepsilon - \widehat{y}\|^2 = \mathbb{E}\|Q_{i-1}\varepsilon - \widehat{y}\|^2 + D_i, \qquad D_i = \mathbb{E}\|R_i\varepsilon\|^2.$$

Taking the infimum over any common predictor class valued in W_{i-1} adds the same D_i to every condition, so risk differences are preserved. The realized-pathway comparison requires each pathway to be held fixed across its two target evaluations.

□

B EXPERIMENTAL PROTOCOL AND FULL RESULTS

B.1 INTERFACE CONDITIONS

The seven evaluated conditions are `Fixed` : $P_c x_t$, `Static Pre` : $P_c A x_t$, `Static Post` : $A(P_c x_t)$, `Time Pre` : $P_c A_t x_t$, `Time Post` : $A_t(P_c x_t)$, `Null-Time Pre` : $P_c A_{\tilde{t}} x_t$ with adapter-conditioning time $\tilde{t}$ independent of true t, and `Time Coarse-Only Pre` : $P_c A_t \Pi_c x_t$ with $\Pi_c = P_c^\top P_c$. Static adapters contain 81 parameters and time-conditioned adapters 405. The true timestep always constructs x_t, selects the noise schedule, and conditions the coarse predictor; only the null adapter receives the independent $\tilde{t}$.

B.2 TRAINING AND SELECTION

The CIFAR-10 training corpus is partitioned into 40,000 training images, 5,000 validation images, and 5,000 held-out evaluation images; the subsets are class-balanced and mutually disjoint, and the held-out split is not used for optimization or checkpoint selection. The forward diffusion process uses $T = 1000$ timesteps with a linear schedule from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$.

All conditions use AdamW with $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$, weight decay 10^{-4}, gradient clipping at 1, batch size 128, a 2,000-step warmup to learning rate 2×10^{-4}, and cosine decay to 2×10^{-6}. Training uses float32 without AMP. The predictor has width 64, four residual blocks, group normalization, and a 128-dimensional sinusoidal time embedding. The pre and post adapters are three-channel circular 3×3 convolutions. Time-conditioned adapters interpolate five knot kernels associated with schedule timesteps 999, 750, 500, 250, and 0.

Each condition is trained for 200,000 optimizer steps from the common predictor initialization. Validation is performed every 5,000 steps on a shared bank, and checkpoint selection uses the lowest equal-bin validation MSE, with the earlier checkpoint chosen only for an exact tie. The selected steps are: `Fixed` 200,000; `Static Pre` 195,000; `Static Post` 195,000; `Time Pre` 200,000; `Time Post` 200,000; `Null-Time Pre` 200,000; and `Time Coarse-Only Pre` 195,000.

B.3 HELD-OUT EVALUATION AND BOOTSTRAP

Evaluation uses 20 diffusion-time bins, 5,000 held-out images, and two noise draws per image and bin. The same true timesteps and noise realizations are shared across conditions. For each reported paired interval, bootstrap resampling is performed at the image level with 10,000 replicates, preserving the paired condition, diffusion-bin, and repeated-noise-draw structure for each sampled image.

B.4 AGGREGATE RESULTS

Table 3: Held-out equal-bin MSE and adapter parameter counts by condition.

Condition	Adapter params.	Held-out MSE
`Fixed`	0	0.078695
`Static Pre`	81	0.078468
`Static Post`	81	0.078619
`Time Pre`	405	0.077841
`Time Post`	405	0.078509
`Null-Time Pre`	405	0.078455
`Time Coarse-Only Pre`	405	0.078528

The two additional paired contrasts not shown in Table 1 are `Time Pre - Static Pre`, -0.000627263569 with 95% interval $[-0.000645872943, -0.000608219679]$ (0.799% relative reduction), and `Static Pre - Static Post`, -0.000151115159 with interval $[-0.000160281630, -0.000141992677]$ (0.192%).

B.5 DIFFUSION-TIME-RESOLVED EVALUATION

Evaluation uses 20 diffusion-time bins across the trajectory, with two paired noise draws per held-out image in each bin. Equal-bin averaging gives each bin equal weight in the aggregate MSE, while the binwise point estimates show whether aggregate differences are distributed across the trajectory or concentrated in a smaller set of diffusion times. The point estimates for `Time Pre − Null−Time Pre` and `Time Pre − Static Pre` are negative in all 20 bins. The point estimate for `Time Pre − Time Coarse−Only Pre` is negative in 17 of 20 bins, with reversals in bins 14, 16, and 17. The `Static Pre − Static Post` point estimate is negative in bins 00–09 and positive in bins 10–19. Figure 2 summarizes the time-resolved contrasts involving `Time Pre`.

B.6 FULL MEASUREMENT-SUBSPACE DIAGNOSTIC

Table 4: Representation-invariant measurement-subspace diagnostics. All measurement maps have rank 768. Here $d_{\max} = \max_{t,s} d(t,s)$, $d_{\text{end}} = d(25, 975)$, and $\bar{d}_{\text{adj}}$ is the mean distance between adjacent evaluation bins. Endpoint angles report the mean and maximum first-to-last principal angles.

Condition	$d_{\max}$	d_{end}	$\bar{d}_{\text{adj}}$
`Static Pre`	0	0	0
`Time Pre`	0.138645	0.138515	0.008060
`Null−Time Pre` sweep	0.000699	0.000544	0.0000718

Condition	Endpoint angles (mean / max)
`Static Pre`	0° / 0°
`Time Pre`	6.436° / 18.232°
`Null−Time Pre` sweep	0.0284° / 0.0527°

The Fourier/block operator action agreed with the literal spatial PyTorch operator to 2.67×10^{-15}. Direct dense real-space QR/SVD reproduced `Time Pre` distances for $(t, s) = (25, 525), (25, 975), (525, 975)$ within 7.4×10^{-14} and the `Null−Time Pre` endpoint within 9.2×10^{-13}. For the exactly static endpoint, dense QR produced a 1.83×10^{-8} numerical floor around the exact block value zero.

B.7 GAUSSIAN-LINEAR SURROGATE ALIGNMENT AND COVARIANCE-SPLIT ROBUSTNESS

The surrogate diagnostic uses the same controlled CIFAR-10 training split and operator conventions as the neural comparison. Images are mapped from uint8 to $[-1, 1]$ by $x/127.5 - 1$ before estimating the centered covariance $\widehat{\Sigma}_0 = G/N - \mu\mu^\top$. The normalized Haar map P_c uses coefficient $1/2$ on each entry of a nonoverlapping 2×2 block, so $L = P_c^\top$ has 768 orthonormal columns. For each evaluation timestep,

$$\widehat{\Sigma}_t = a_t^2 \widehat{\Sigma}_0 + b_t^2 I, \qquad S_t^{\text{GL}} = \text{col}(\widehat{\Sigma}_t^{-1} L).$$

The fixed measurement is $M^{\text{Fixed}} = L$, while $M_t^{\text{Time}} = (P_c A_t)^\top$. The covariance-matched surrogate PSG is Proposition 1 with Σ_t replaced by $\widehat{\Sigma}_t$. For the geometric diagnostic, if P_M and $P_{S_t^{\text{GL}}}$ are the orthogonal projectors onto the two 768-dimensional subspaces, we use $\|P_M - P_{S_t^{\text{GL}}}\|_F / \sqrt{2 \cdot 768}$.

With the covariance estimated from all 40,000 training images, both diagnostics favor `Time Pre` over `Fixed` at all 15 timesteps $t = 25, 75, \ldots, 725$, and both reverse at all five timesteps $t = 775, 825, \ldots, 975$. The 20-point descriptive mean surrogate PSG is 3.6708 for `Fixed` and 2.4433 for `Time Pre`, a 33.44% reduction; because early timesteps have much larger absolute surrogate PSG, this average is descriptive rather than inferential.

To assess covariance-estimation sensitivity without retraining, the frozen training-index sequence is split into two disjoint 20,000-image halves and the diagnostic is recomputed independently under each covariance estimate. The sign pattern is identical in the full covariance and both halves:

Table 5: Covariance-split robustness of the Gaussian-linear surrogate alignment. Entries count evaluation timesteps with the indicated sign of `Time Pre−Fixed`.

Interval and diagnostic	Full covariance	Half A	Half B
$t = 25$–725: surrogate PSG < 0	15/15	15/15	15/15
$t = 25$–725: projector distance < 0	15/15	15/15	15/15
$t = 775$–975: surrogate PSG > 0	5/5	5/5	5/5
$t = 775$–975: projector distance > 0	5/5	5/5	5/5

At the closest-to-zero early/mid point, $t = 725$, the surrogate-PSG differences are -4.2043×10^{-4}, -3.7495×10^{-4}, and -4.6275×10^{-4} for the full, half-A, and half-B covariance estimates; the corresponding projector-distance differences are -1.4180×10^{-4}, -1.3062×10^{-4}, and -1.4774×10^{-4}. Thus the preserved sign pattern is not driven only by the larger early-time effects. These deterministic half-sample checks are robustness diagnostics, not confidence intervals, and do not quantify retraining-seed variability. Because CIFAR-10 is non-Gaussian, S_t^{GL} is used only as a covariance-defined Gaussian-linear surrogate geometry and is not identified with the true Bayes-sufficient subspace.

C End-to-End Diffusion U-Net and Generation Evaluation

C.1 Architecture and information-access intervention

The end-to-end experiment uses a CIFAR-10 RGB 32×32 full-resolution ε-prediction U-Net. Its spatial resolutions are $32, 16, 8, 4$, with base width 64, channel multipliers $(1, 2, 2, 4)$, two residual blocks at each resolution, group normalization, SiLU activations, a 256-dimensional timestep embedding, dropout 0.1, and four-head attention at 16×16. The first $32 \times 32 \to 16 \times 16$ transition uses a normalized 2×2 Haar low-pass map P_c and a shared learned postprojection 3×3 channel remap D. With fine-grid feature h_{32} and true-time-conditioned adapter A_t, the coarse inputs are

$$\begin{aligned} \texttt{Fixed}: &\quad D(P_c h_{32}), \\ \texttt{Time Pre}: &\quad D(P_c A_t h_{32}), \qquad \Pi_c = P_c^\top P_c. \\ \texttt{Time Coarse-Only Pre}: &\quad D(P_c A_t \Pi_c h_{32}), \end{aligned}$$

The adapter is a residual fine-grid circular 3×3 convolution interpolated among five timestep knots. The fine-resolution skip pathway is unchanged. The adaptive conditions have matched adapter capacity, true-time input, spatial placement, convolution geometry, downstream model, target, and training budget; only access to $(I - \Pi_c)h_{32}$ at the adapter differs. This full denoiser is distinct from the compact predictor in Appendix B.

C.2 Held-out full-resolution evaluation

The CIFAR-10 partition contains 40,000 optimization, 5,000 validation, and 5,000 held-out evaluation images. Checkpoints were selected by minimum validation equal-bin full-resolution ε-MSE: step 130,000 for `Fixed` and step 140,000 for each adaptive condition. The diffusion schedule has $T = 1000$ timesteps and linearly spaced β_t from 10^{-4} to 0.02. Held-out evaluation uses 20 equal-width time bins and two paired Gaussian-noise draws per image and bin. The aggregate MSE weights the bins equally. For each contrast, 10,000 bootstrap replicates resample images as clusters while retaining their condition, time-bin, and noise-draw pairing. The intervals are conditional on the selected checkpoints. Exact aggregate estimates and contrasts appear in Table 6.

C.3 Time-resolved denoising behavior

The full-U-Net `Time Pre` minus `Time Coarse-Only Pre` MSE point estimate is negative in 4 of 20 bins. The lowest-noise bin contributes approximately 100.5% of the net equal-bin improve-

Table 6: Full-U-Net held-out equal-bin ε-MSE and paired contrasts. Relative differences use the named comparator as denominator.

Condition	Equal-bin ε-MSE ↓
`Fixed`	0.0297320
`Time Pre`	0.0297129
`Time Coarse-Only Pre`	0.0297206

Comparator	ΔMSE	Relative difference
`Fixed`	-1.91×10^{-5}	-0.0643%
`Time Coarse-Only Pre`	-7.74×10^{-6}	-0.0260%

Comparator	Paired 95% interval
`Fixed`	$[-2.77 \times 10^{-5}, -1.06 \times 10^{-5}]$
`Time Coarse-Only Pre`	$[-1.55 \times 10^{-5}, -1.81 \times 10^{-7}]$

ment, and the mean difference over the other 19 bins is approximately $+4.0620 \times 10^{-8}$. The `Time Pre` minus `Fixed` aggregate difference is also dominated by the lowest-noise bin. The small aggregate effect is therefore highly localized in diffusion time. The corresponding controlled-interface contrast in Appendix B has a different target and a negative point estimate in 17 of 20 bins.

C.4 Ancestral-DDPM generation and FID

The generation comparison uses the same selected non-EMA checkpoints as the full-resolution denoising evaluation. For each of 5,000 matched sample IDs, the three conditions share the same x_T and the same Gaussian noise at every stochastic reverse transition. Sampling performs all 1,000 ancestral DDPM reverse steps with the training linear β schedule and fixed-small posterior variance. The output conversion maps x_0 through $\mathrm{clip}((x_0+1)/2, 0, 1)$ to uint8 RGB images. FID uses torchvision Inception-V3 with 2,048-dimensional features and the same fixed reference of all 50,000 CIFAR-10 training images for each condition. This reference contains the 5,000-image held-out denoising subset because both are drawn from the CIFAR-10 training corpus; the denoising metrics themselves are computed only on the designated held-out split. Table 2 reports the FID estimates.

C.5 Paired grouped-jackknife procedure and complete results

The paired generated sample IDs were divided into ten contiguous blocks of 500. In each leave-one-block-out evaluation, the same block was removed from all three saved Inception feature arrays. For each condition, the mean and unbiased sample covariance of its remaining 4,500 full-dimensional feature rows were recomputed and evaluated by the unchanged FID calculation against the fixed 50,000-image real-reference statistics. If d_b denotes a paired FID contrast with block b removed and $\bar{d}$ is the mean of the ten deleted-block contrasts, the grouped-jackknife standard error is

$$\mathrm{SE_J} = \sqrt{\frac{9}{10}\sum_{b=1}^{10}(d_b - \bar{d})^2}.$$

The approximate 95% interval is centered on the observed full-sample contrast and uses $\pm\, t_{9,0.975}\,\mathrm{SE_J}$, where $t_{9,0.975} = 2.262157$. This fixed-reference diagnostic is not an exact-coverage population interval and does not replace the full-sample FID estimates. Table 7 gives the complete paired deletion contrasts.

Table 7: FID contrasts after removing the same 500 sample IDs from all three conditions. The 50,000-image real reference remains fixed.

Omitted IDs	Time Pre − Fixed	Time Pre − Time Coarse-Only Pre
0–499	-2.492478	$+0.879727$
500–999	-2.507594	$+0.775829$
1000–1499	-2.367934	$+0.763706$
1500–1999	-2.455165	$+0.682500$
2000–2499	-2.566427	$+0.687700$
2500–2999	-2.530181	$+0.754038$
3000–3499	-2.455675	$+0.840566$
3500–3999	-2.367273	$+0.657104$
4000–4499	-2.342456	$+0.829256$
4500–4999	-2.510729	$+0.731288$

The grouped-jackknife standard errors are 0.218561 for Time Pre minus Fixed and 0.209070 for Time Pre minus Time Coarse-Only Pre. Their approximate 95% intervals are $[-2.957350, -1.968510]$ and $[+0.284054, +1.229953]$, respectively. The first contrast is negative and the second positive in all ten deletions.