EQUAL PATH COST, UNEQUAL OUTPUT EFFECTS: UNDERSTANDING PERTURBATION PROPAGATION IN DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models have achieved remarkable success in generative modeling, with their sampling procedures routinely modified to control generation and improve efficiency. These modifications introduce perturbations along the sampling trajectory, raising a central question: how do such perturbations affect generated output? To address this question, we develop a theoretical framework to investigate perturbation propagation, combining dynamical analysis of the sampling process with an information-theoretic characterization of output responses. Within this framework, we quantify perturbation strength using the Kullback–Leibler (KL) divergence between perturbed and reference trajectory distributions, termed as *path cost*, which is shown to bound, but do not determine, changes in the output distribution. Building on this analysis, we derive a response identity that connects the propagation and accumulation of local perturbations with the information captured by a selected feature mean, explaining why changes in the output distribution can remain undetected by its first-order response. We test our theoretical analysis through controlled interventions at equal path cost in pretrained diffusion models, revealing distinct patterns of output sensitivity across sampling stages and spatial frequencies. To assess whether our framework can diagnose perturbations arising from practical approximations, we apply it to cache-based acceleration and show that our propagation analysis reliably identifies sampling intervals where caching causes larger image errors.

1 INTRODUCTION

Diffusion models generate samples by simulating learned reverse-time dynamics from a noise distribution to the data distribution (Ho et al., 2020; Song et al., 2020). During inference, these dynamics are perturbed by errors in learned score estimates or denoising predictions (Song et al., 2021; Li & van der Schaar, 2024), by deliberate guidance updates (Dhariwal & Nichol, 2021), and by approximations introduced by fast solvers, quantization, and feature caching (Lu et al., 2022; Li et al., 2023; Ma et al., 2024). Because a perturbation changes the input to subsequent denoising steps, its effect on the generated sample depends on the downstream sampling dynamics, not only on its magnitude at the timestep where it occurs (Li & van der Schaar, 2024; Liu et al., 2026). This creates a mismatch between the trajectory-level quantities used to characterize a perturbation and the output-level quantities used to evaluate its consequences: the former describe the change where it enters the sampling process, whereas the latter concern the endpoint distribution, the response of a selected feature, or the displacement of a generated sample. This mismatch raises a central question: *How do perturbations along the sampling trajectory translate into changes in generated outputs, and to what extent can trajectory-level measures characterize those changes?*

Existing work has addressed different aspects of this question. Analyses of score estimation and discretization use aggregated errors along the sampling trajectory to bound deviations in the generated distribution (Song et al., 2021; Chen et al., 2022; Benton et al., 2024). Studies of exposure bias and error propagation instead examine how prediction errors accumulate across denoising steps (Ning et al., 2023; 2024; Li & van der Schaar, 2024). More recently, in the setting of recursively trained diffusion models, Khelifa et al. (2026) derive terminal-divergence lower bounds parameterized by an endpoint-observability coefficient, which measures how much pathwise score-error energy remains

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.

detectable in the terminal distribution. These results show that inaccuracies along the sampling trajectory can accumulate into the global generation error and the pathwise error may be only partially observable in the terminal distribution. However, they do not fully characterize how the downstream sampling dynamics mediate the relationship between the magnitude of a perturbation along the trajectory and its effect on the final generated output.

To address this gap, we develop a theoretic framework combining information-theoretic analysis with a dynamical characterization of output response, which explains why perturbations of equal strength can produce unequal output effects and derives how local perturbations propagate and accumulate to determine output responses. We measure perturbation strength by the KL divergence between perturbed and reference trajectory distributions, which we call *path cost*. The analysis shows that a perturbation's effects on generated outputs depend jointly on its path cost, how it propagates and interacts with perturbations at other steps, and how the resulting changes are measured. Our contributions are:

- **A theoretical framework explaining how perturbations affect generated outputs.** We derive an exact decomposition of path cost that distinguishes changes in the output distribution from changes in trajectory distributions conditional on the output. We derive a first-order propagation law that quantifies how subsequent sampling dynamics transform local perturbations and how their contributions accumulate to determine the final output response. We further establish bounds relating feature mean responses to local output-distribution changes, explaining why a selected feature mean can fail to detect such changes. These results explain why equal path costs can produce unequal output effects.
- **Experimental support for the theoretical analysis.** We test the framework through controlled perturbations at equal path cost, varying the sampling stages at which they are applied and their spatial directions. The resulting output responses differ across sampling stages and directions. The experiments show that first-order propagation can capture the temporal patterns of responses to finite perturbations, although predicted and observed magnitudes can differ substantially. They also show that smaller absolute feature mean responses can correspond to larger changes relative to baseline fluctuations, consistent with our theoretical analysis.
- **Applying the framework to diagnose output errors in cache-based acceleration.** We apply the framework to cache-based acceleration for diffusion generation, where earlier model predictions are reused to avoid fresh evaluations. By tracing how the resulting approximation errors propagate and accumulate, we assess their impact on the final image and identify sampling intervals where caching causes larger output deviations. This propagation-based diagnostic identifies such intervals more reliably than local error magnitude and path-cost measures, which remains informative even when predicted and observed error magnitudes differ substantially.

2 Related work

Error propagation in diffusion sampling. Path-space control relates diffusion modeling to KL minimization (Berner et al., 2024), while score-error analyses provide likelihood and sampling-error bounds (Song et al., 2021; Chen et al., 2022; Benton et al., 2024). Li & van der Schaar (2024) relate local and cumulative errors through a propagation equation. Khelifa et al. (2026) quantify the endpoint visibility of pathwise score errors to derive terminal-divergence lower bounds, with applications to recursive training. We study perturbations of a fixed sampler, connecting output visibility to downstream propagation and the information captured by a selected feature mean.

Conditional information and observable response. Conditional-score projections characterize Fisher information retained by an observation (Pollard, 2013), while fluctuation–response bounds relate mean sensitivity to variance and path information (Dupuis et al., 2016; Dechant & Sasa, 2020). De Bortoli et al. (2025) bound diffusion sampling errors using smooth test functions. We build on these information and response identities to characterize how much path information reaches the output and how much a feature mean captures, with propagation determining its response.

Spectral generation dynamics. Benita et al. (2025) derive Gaussian spectral transfer dynamics and use data spectra to design schedules. We compare temporal and directional responses at matched path cost, and distinguish absolute response from response relative to baseline feature fluctuations.

Caching and sensitivity-based acceleration. DeepCache reuses high-level features (Ma et al., 2024); SenCache uses denoiser sensitivity to latent and timestep changes (Haghighi & Alahi, 2026). ERTACache analyzes and corrects feature-shift and step-amplification errors (Peng et al., 2026), while EpaCache calibrates reuse thresholds from terminal deviations under single-step caching interventions (Liu et al., 2026). We use reference-trajectory propagation to diagnose image errors from multi-step output caching and test whether it identifies sampling windows with larger errors more reliably than local-defect and path-cost measures.

3 A Theoretical Framework for Perturbation Propagation

In this section, we derive a theoretical characterization of how local perturbations propagate through diffusion sampling and affect measured output responses. We first quantify trajectory changes by path cost, separate output changes from conditional trajectory changes, and derive perturbation propagation. Our main theorem expresses the measured response through the propagated local perturbations and identifies the fractions of path information retained in the output distribution and captured by the chosen feature mean.

3.1 How path cost quantifies perturbation strength

Diffusion models learn to reverse a forward process that progressively adds noise to data. Generation starts from noise and follows learned reverse dynamics through discrete sampling updates (Ho et al., 2020; Song et al., 2020). Common stochastic diffusion samplers update the state using a model-dependent mean and Gaussian noise (Nichol & Dhariwal, 2021). These updates can then be expressed in the following form

$$X_{i+1} = \mathcal{G}_i^0(X_i, \xi_i) = \mu_i(X_i) + L_i(X_i)\xi_i, \qquad \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I). \tag{1}$$

Here, X_i denotes the sampling state after i generation steps, with X_0 drawn from the initial noise distribution and X_M denoting the terminal sampling state. The index $i = 0, \ldots, M-1$ follows generation order from noise toward the output. The map $\mathcal{G}_i^0$ defines one update of the reference sampler with the superscript 0 indicating the unperturbed process.The variables ξ_i are independent standard Gaussian noise vectors, and $L_i(X_i)\xi_i$ is the stochastic component of the sampling update, with conditional covariance $\Lambda_i(X_i) = L_i(X_i)L_i(X_i)^\top$.

As mentioned above, local perturbations, such as prediction errors and computational approximations, modify the sampling updates in Eq. (1). We model the resulting dynamics by a family of perturbed sampling processes parameterized by perturbation strength ε, with $\varepsilon = 0$ corresponding to the reference process. To study how local perturbations affect the sampling process, we consider the complete trajectories $\Gamma = (X_0, \ldots, X_M)$ and denote their joint distribution under the reference and perturbed processes as P and Q_ε, respectively, so that $Q_0 = P$. Such effects can be measured through the statistical distinguishability of these trajectory distributions using their Kullback–Leibler (KL) divergence (Cover & Thomas, 2006), which we define as the *path cost*, $\mathcal{C}(\varepsilon) = D_{\mathrm{KL}}(Q_\varepsilon \| P)$.

We consider perturbations that preserve the initial distribution and shift each Gaussian transition mean by εu_i, while retaining its positive definite covariance Λ_i. The resulting transition densities are strictly positive, ensuring that Q_ε is absolutely continuous with respect to P, i.e., $Q_\varepsilon \ll P$ (Kallenberg, 1997). Then, by using the relative-entropy chain rule (Cover & Thomas, 2006), we can obtain the expression of the path cost as a sum of conditional transition KLs as

$$\mathcal{C}(\varepsilon) := D_{\mathrm{KL}}(Q_\varepsilon \| P) = \frac{\varepsilon^2}{2} \mathbb{E}_{Q_\varepsilon} \sum_i u_i^\top \Lambda_i^{-1} u_i. \tag{2}$$

This expression shows how local mean perturbations contribute to the change in the trajectory distribution. Each contribution is weighted by the inverse transition covariance: for $\Lambda_i = \nu_i I$, the conditional KL at step i is $\varepsilon^2 \|u_i\|^2/(2\nu_i)$. Thus, the same mean shift produces a larger change in the transition distribution, as measured by KL divergence, when the sampling noise is lower. The derivation details are provided in Appendix A.1.

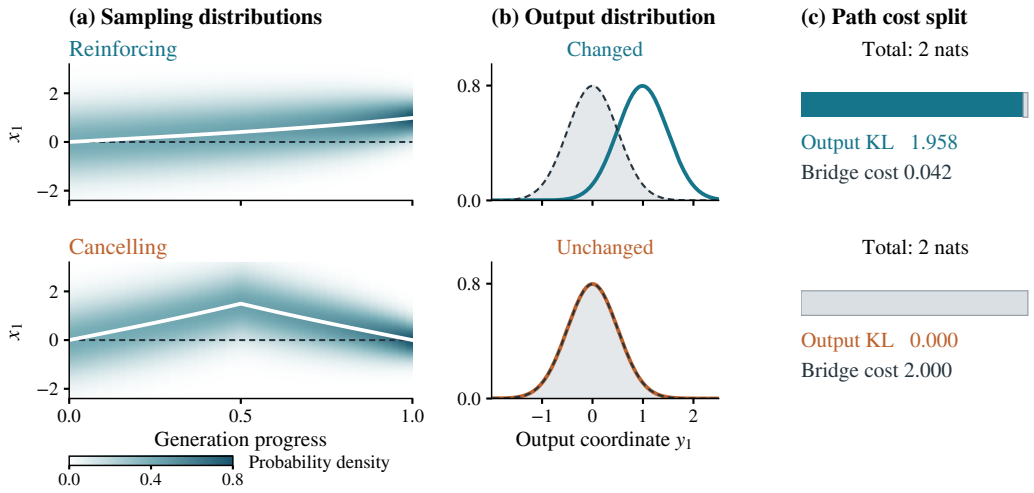


Figure 1: **Equal path costs, different output-distribution changes.** Two temporal perturbation sequences act only on the first coordinate of a two-dimensional Gaussian reverse diffusion, each with path cost $\mathcal{C} = 2$ nats. Their contributions reinforce (top) or cancel (bottom) at the output. (a) Coordinate density evolution on a shared color scale, with perturbed (white) and reference (dashed) means. (b) Perturbed (solid) and reference (dashed) output densities. (c) Output KL (colored) and bridge cost (gray). Construction details are in Appendix B.

3.2 Why path cost does not determine output changes

Having quantified how perturbations change the sampling trajectory distribution, we now examine their effect on the distribution of generated outputs. The terminal sampling state X_M may require a final deterministic transformation, such as decoding a latent representation, to produce the generated output. We therefore write $Y = \phi(X_M)$, where ϕ denotes this transformation and is the identity when X_M is already the output. We use the same map for both the reference and perturbed processes.

Then, we derive an exact decomposition of path cost into the change in the output distribution and the change in trajectory distributions conditional on the output in the following Proposition 1.

Proposition 1 (Decomposition of path and output changes)**.** *For $Q_\varepsilon \ll P$ with finite path KL and regular conditional path distributions, let p_Y and $q_{\varepsilon,Y}$ be the output distributions under P and Q_ε. Then, the path cost can be decomposed as*

$$\mathcal{C}(\varepsilon) = \underbrace{D_{\mathrm{KL}}(q_{\varepsilon,Y} \| p_Y)}_{\mathcal{C}_Y(\varepsilon)} + \underbrace{\mathbb{E}_{q_{\varepsilon,Y}} D_{\mathrm{KL}}(Q_\varepsilon(\cdot \mid Y) \| P(\cdot \mid Y))}_{\mathcal{C}_{\mathrm{bridge}|Y}(\varepsilon)}. \tag{3}$$

Proof. Please see Appendix A.2. □

The above decomposition separates two consequences of a perturbation. The first term, $\mathcal{C}_Y(\varepsilon)$, measures the change distinguishable from generated outputs. The second measures how the trajectory distributions differ when conditioned on the same output, averaged over the perturbed output distribution, which we call as the *bridge cost*. Since bridge cost is non-negative, we have $0 \le \mathcal{C}_Y(\varepsilon) \le \mathcal{C}(\varepsilon)$. Therefore, the path cost bounds the changes in output distribution, but equal path costs can yield different output changes depending on their bridge components.

To illustrate this decomposition, Figure 1 presents a Gaussian toy example that we construct to show how perturbations with equal path costs can produce different output effects. In this example, scaling the cancelling perturbation sequence yields any positive path cost while output KL remains zero. Thus, a positive path cost alone does not imply a positive output-distribution change. To understand how perturbations from different steps reinforce or cancel at the output, we next examine their propagation through the sampling dynamics.

3.3 How local perturbations propagate and accumulate

To trace how local perturbations affect generated outputs, we analyze the first-order response with respect to perturbation strength at $\varepsilon = 0$. This gives a linear description of propagation along the reference trajectories.

Let $X_{i+1}^{\varepsilon} = \mathcal{G}_i^{\varepsilon}(X_i^{\varepsilon}, \xi_i)$ denote the perturbed update, with $\mathcal{G}_i^{\varepsilon}|_{\varepsilon=0} = \mathcal{G}_i^0$ as in Eq. (1). To compare corresponding trajectories, we hold the initial state and noise variables fixed. Along each reference trajectory, we define $J_i = D_x \mathcal{G}_i^0$ as the Jacobian matrix of the reference update with respect to its input state, which describes how an input-state change propagates to the next step. We also define $e_i = \partial_\varepsilon \mathcal{G}_i^{\varepsilon}|_0$, as the derivative of the sampling update with respect to perturbation strength with its input state and noise remained fixed. This term measures the local perturbation introduced by the update itself. For the mean perturbations in Section 3.1, $e_i = u_i$.

To examine how local perturbations affect a particular property of the generated output, we represent that property by a scalar feature $f(Y)$, such as a spatial-frequency component of an image(Ahmed et al., 1974). Its expectation summarizes this property across generated samples. Therefore, we study how the expected feature value changes under perturbation and define the feature mean response (Kubo et al., 2012) as $\chi_f = \partial_\varepsilon \mathbb{E}_{Q_\varepsilon} f(Y)|_0$. Using pathwise differentiation (Mohamed et al., 2020), we express both the output response and the feature mean response in the following proposition.

Proposition 2 (Perturbation propagation dynamics)**.** *Assume that $\mathcal{G}_i^{\varepsilon}$ is differentiable in the state and perturbation strength, and that ϕ is differentiable at the terminal state, almost surely along reference trajectories. The output response, defined as the derivative of the generated output with respect to perturbation strength at $\varepsilon = 0$, satisfies*

$$V := \partial_\varepsilon Y^{\varepsilon}|_0 = \sum_{i=0}^{M-1} T_i e_i, \qquad T_i = D\phi(X_M) J_{M-1} \cdots J_{i+1}, \tag{4}$$

where T_i propagates the local perturbation at step i through the remaining sampling updates and the output map ϕ. Furthermore, if f is continuously differentiable and $[f(Y^{\varepsilon}) - f(Y)]/\varepsilon$ is dominated by an integrable random variable near $\varepsilon = 0$, then the feature mean response is given as

$$\chi_f = \mathbb{E}_P[\nabla f(Y)^\top V] = \sum_{i=0}^{M-1} \mathbb{E}_P[\nabla f(Y)^\top T_i e_i]. \tag{5}$$

Proof. Please see Appendix A.3. □

Proposition 2 shows how downstream dynamics transform each local perturbation through T_i, changing its strength and direction at the output. The propagated perturbations from different steps add to form the total output response V. To examine how their magnitudes and relative directions determine the squared magnitude of this response, we write $v_i = T_i e_i$ and expand the norm of V as

$$\|V\|^2 = \sum_i \|v_i\|^2 + 2 \sum_{i<j} \langle v_i, v_j \rangle. \tag{6}$$

The first term sums the squared magnitudes of the propagated perturbations considered separately. The cross terms account for their relative directions: positive inner products increase the total squared response, while negative inner products reduce it. Thus, the magnitude of the total response depends on both the sizes of the propagated perturbations and how their directions align at the output.

The above analysis explains how local perturbations propagate and accumulate to determine the output response and the mean response of a selected feature. However, this feature mean describes only one property of the output distribution. Interpreting its response therefore requires understanding how much distributional change it captures and how this depends on the choice of feature.

3.4 What feature mean responses reveal about output changes

To identify the distributional changes captured by the feature mean, we derive its response in terms of trajectory likelihoods. Differentiating $\mathbb{E}_{Q_\varepsilon} f(Y) = \mathbb{E}_P[f(Y)\, \mathrm{d}Q_\varepsilon/\mathrm{d}P]$ with respect to perturbation

strength at $\varepsilon = 0$, we obtain

$$\chi_f = \mathbb{E}_P[f(Y)S], \tag{7}$$

where $S = \partial_\varepsilon \log(\mathrm{d}Q_\varepsilon/\mathrm{d}P)|_0$ is the path likelihood score, describing changes in trajectory log likelihood with perturbation strength.

To identify the leading dependence of path cost on perturbation strength, we derive the expansion of Eq. (2) near the reference sampler as

$$\mathcal{C}(\varepsilon) = \tfrac{1}{2}\varepsilon^2 F_{\mathrm{path}} + o(\varepsilon^2), \tag{8}$$

where $F_{\mathrm{path}} = \mathbb{E}_P S^2$ is the path Fisher information (Amari, 2016; Dupuis et al., 2016). Since $f(Y)$ depends only on the output, we can average S over trajectories sharing that output without changing χ_f. This gives the output likelihood score $S_Y = \mathbb{E}_P[S \mid Y]$ (Pollard, 2013). We define the output Fisher information as $F_Y = \mathbb{E}_P S_Y^2$ and derive the second-order expansion of the KL decomposition in Eq. (3), yielding

$$F_{\mathrm{path}} = F_Y + \underbrace{\mathbb{E}_P \operatorname{Var}_P(S \mid Y)}_{F_{\mathrm{bridge}|Y}}. \tag{9}$$

The first term determines local output-distribution change through $\mathcal{C}_Y(\varepsilon) = \frac{1}{2}\varepsilon^2 F_Y + o(\varepsilon^2)$, while the second is the local bridge component. For $F_{\mathrm{path}} > 0$, the ratio F_Y/F_{path} therefore gives the fraction of path cost visible at the output as $\varepsilon \to 0$. The assumptions and derivations for Eqs. (8) and (9) are provided in Appendix A.4.

To compare the feature mean response with the output and path Fisher information, we define the squared normalized mean response as $B_f = \chi_f^2 / \operatorname{Var}_P(f(Y))$. For small ε, the squared mean shift on this scale is $\varepsilon^2 B_f + o(\varepsilon^2)$, matching the perturbation order of the local KL expansions. We derive a lower bound on output Fisher information from this normalized response using the fluctuation–response bound (Dupuis et al., 2016; Dechant & Sasa, 2020) in the following Proposition 3.

Proposition 3 (Feature response and output information)**.** *Assume that the path likelihood ratio is differentiable in mean square at $\varepsilon = 0$, meaning $\mathbb{E}_P[(\mathrm{d}Q_\varepsilon/\mathrm{d}P - 1 - \varepsilon S)^2] = o(\varepsilon^2)$, with $\mathbb{E}_P S^2 < \infty$. Then any real-valued output feature $f(Y)$ with $0 < \operatorname{Var}_P(f(Y)) < \infty$ satisfies*

$$\chi_f = \mathbb{E}_P[f(Y)S] = \mathbb{E}_P[f(Y)S_Y], \qquad B_f = \frac{\chi_f^2}{\operatorname{Var}_P(f(Y))} \le F_Y \le F_{\mathrm{path}}. \tag{10}$$

Proof. Please see Appendix A.5. □

Proposition 3 identifies two stages at which perturbation information can be lost, from the trajectory to the output distribution and from the output distribution to the selected feature mean. We derive an exact expression for the fraction of path information captured by the chosen feature mean, incorporating both perturbation propagation and the two stages of information loss in the following Theorem 1.

Theorem 1 (Perturbation propagation and measured output response)**.** *Under the assumptions of Propositions 2 and 3, let f be continuously differentiable with $0 < \operatorname{Var}_P(f(Y)) < \infty$. If $F_Y > 0$, the fraction of path information captured by the feature mean response satisfies*

$$\frac{B_f}{F_{\mathrm{path}}} = \frac{\left(\sum_i \mathbb{E}_P[\nabla f(Y)^\top T_i e_i]\right)^2}{\operatorname{Var}_P(f(Y))\, F_{\mathrm{path}}} = \underbrace{\frac{F_Y}{F_{\mathrm{path}}}}_{\textit{output visibility}} \underbrace{\operatorname{Corr}_P(f(Y), S_Y)^2}_{\textit{feature–score alignment}}. \tag{11}$$

Proof. Please see Appendix A.6. □

Theorem 1 explains why perturbations with equal path cost can produce different measured responses. The propagated terms show how downstream dynamics, cross-time accumulation, and the selected feature jointly shape the response. Here, *output visibility* measures the fraction of path cost reflected in output-distribution change for small perturbations, and *feature–score alignment* measures the fraction of local output information captured by the chosen feature mean. In particular, a perturbation can change the output distribution while leaving the mean of the chosen feature unchanged to first order in perturbation strength.

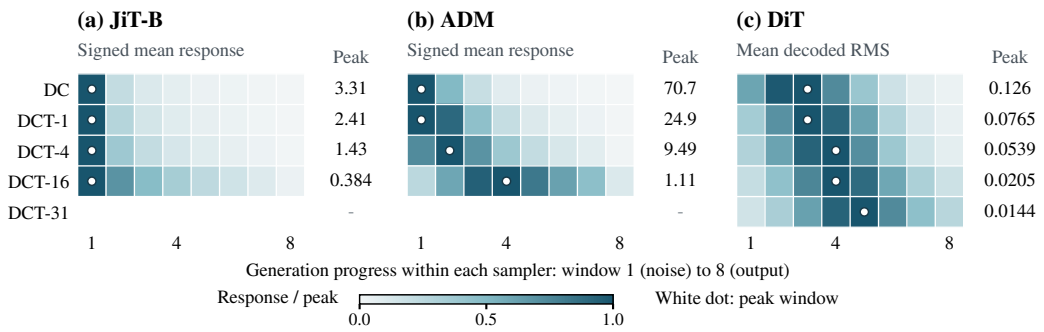


Figure 2: **Output responses at equal path cost within each model.** Rows show perturbation frequencies and columns show the sampling windows in which perturbations are applied, from early to late. Each row is divided by its maximum response. White dots mark the peak windows and numbers on the right give the maxima before normalization.

4 Controlled Perturbations Propagation with Equal Path Cost

In this section, we test the theoretical framework by using controlled perturbations at equal path cost, which provides a common measure of perturbation strength for comparison.

4.1 Experimental setup

We study controlled perturbations in class-conditional ImageNet-256 generation using pixel-space JiT-B/L (Li & He, 2026) and ADM (Dhariwal & Nichol, 2021), and latent-space DiT (Peebles & Xie, 2023), without additional guidance. We group the 128 stochastic sampling steps into eight consecutive windows of 16 steps, numbered 1–8 in generation order from noise toward the output. In each experiment, we perturb the transition means within one window. To vary the spatial direction, we use unit-norm discrete cosine transform (DCT) templates, which represent spatial cosine patterns on the pixel or latent grid: DC is constant, and DCT-k has higher frequency as k increases. We normalize each perturbation to satisfy $\sum_i u_i^\top \Lambda_i^{-1} u_i = 1$, giving equal path cost $\mathcal{C}(\varepsilon) = \varepsilon^2/2$ across windows and directions within each model, conditional on class and template. We use $\varepsilon = 0.25$ for JiT/ADM and 0.8 for DiT. For JiT and ADM, we measure changes in the mean DCT coefficients of generated images. For DiT, we measure the RMS image response after decoding. Measurement details are provided in Appendix C.

4.2 Spatiotemporal sensitivity to perturbations

Figure 2 shows how output responses vary with perturbation timing and spatial frequency at equal path cost. In JiT-B, all four frequencies shown produce their largest response in the first window, with responses decreasing toward later windows. In ADM, the peak shifts from window 1 for DC to window 4 for DCT-16; in DiT, it shifts from window 3 for DC to window 5 for DCT-31. Thus, earlier perturbations do not always produce larger output responses: the most sensitive stage depends on both the model and perturbation frequency. These observations are consistent with Section 3.3, where downstream propagation shapes the output effect of a perturbation beyond its path cost.

We provide two DiT image examples in Appendix D.1, illustrating how perturbations with equal path cost produce different output changes when applied at different sampling stages.

4.3 Local predictions and finite perturbation responses

The responses above are measured at finite perturbation strength, whereas the propagation analysis in Section 3.3 describes the output response to infinitesimal perturbations. We use DiT to test how well this local prediction captures the magnitude and temporal pattern of responses to finite perturbations. On the same 32 sampling trajectories, we compare the mean per-sample RMS of the output derivative V, computed through the sampler and decoder using Jacobian–vector products (JVPs), with that of the symmetric finite difference $(Y^{+\varepsilon} - Y^{-\varepsilon})/(2\varepsilon)$ at $\varepsilon = 0.8$.

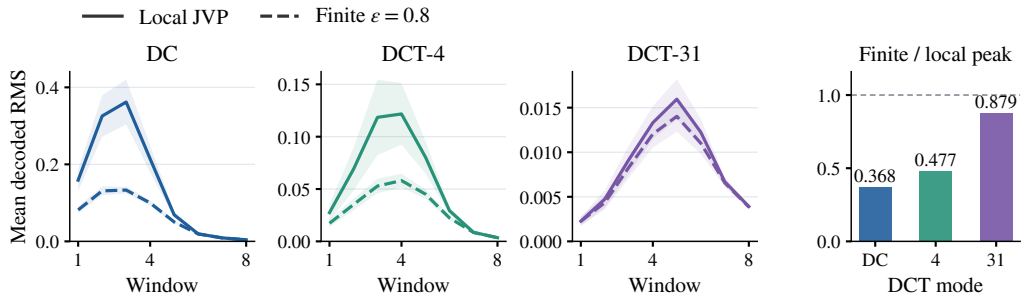


Figure 3: **Local predictions and finite perturbation responses in DiT.** Mean decoded-image RMS of local derivatives and symmetric finite differences ($\varepsilon = 0.8$) on the same 32 trajectories, where shaded bands indicate one standard error of the mean across samples. The right panel reports finite/local peak ratios with dashed line denoting equal amplitudes. All runs use 128 stochastic steps.

Table 1: **Feature responses relative to baseline fluctuations.** Perturbations act at window 1 at equal path cost, with $\varepsilon = 0.25$ and $N = 512$ for JiT-B/L and 256 for ADM. We report absolute mean response $|\widehat{\chi}_f|$, baseline standard deviation s_f, and normalized response $\widehat{B}_f = \widehat{\chi}_f^2 / s_f^2$.

	JiT-B			JiT-L			ADM		
Feature	$\lvert\widehat{\chi}_f\rvert$	s_f	$\widehat{B}_f$	$\lvert\widehat{\chi}_f\rvert$	s_f	$\widehat{B}_f$	$\lvert\widehat{\chi}_f\rvert$	s_f	$\widehat{B}_f$
DC	3.31	127	0.000677	3.36	114	0.000875	70.7	106	0.447
DCT-1	2.41	48.5	0.00248	2.46	45.2	0.00296	24.9	42.5	0.343
DCT-4	1.43	18.9	0.00577	1.46	21.4	0.00465	6.9	23.5	0.0865
DCT-16	0.384	3.88	0.00982	0.366	4.34	0.00712	0.274	4.32	0.00402

As shown in Figure 3, the local and finite responses follow similar trends across sampling windows and peak at windows 3, 4, and 5 for DC, DCT-4, and DCT-31, respectively. However, the finite-response peaks are only 36.8%, 47.7%, and 87.9% of the corresponding local predictions, as shown in the right panel. Thus, in this DiT example, local propagation identifies the most responsive sampling windows, while overestimating the peak response magnitudes at the tested finite amplitude.

4.4 Feature responses relative to baseline fluctuations

Having examined propagation through the decoded-image response in DiT, we now return to the feature mean responses in JiT and ADM to compare them on their baseline fluctuation scales. Section 3.4 relates a feature's mean response to output information after normalization by its baseline variance. Table 1 compares absolute responses, baseline fluctuations, and normalized responses at the first sampling window. In JiT-B/L, DCT-16 has a smaller absolute response than DC, but its baseline fluctuations are sufficiently smaller to give a larger normalized response. In ADM, the absolute response decreases more sharply than the baseline fluctuations, so DCT-16 also has a smaller normalized response. These results show that a smaller mean response can represent a larger change relative to a feature's baseline fluctuations, consistent with the analysis in Section 3.4. We provide the results for all sampling windows in Appendix D.2.

5 Diagnosing Output Errors in Cache-Based Acceleration

In this section, we investigate whether the perturbation propagation analysis in Section 3 can predict output errors caused by practical computational approximations during diffusion sampling.

We consider cache-based acceleration as a practical example during diffusion sampling, since reusing earlier model predictions reduces the number of model evaluations but introduces approximation errors into subsequent updates (Ma et al., 2024; Liu et al., 2026). We divide 64 sampling steps into eight windows. Each candidate caches the prediction at the first step of one window B

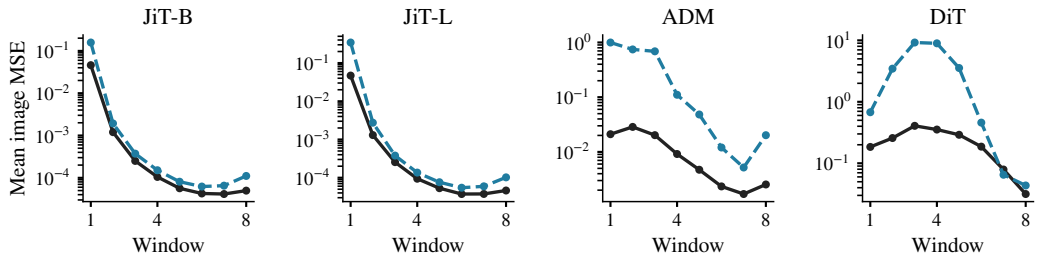


Figure 4: **Predicted and observed image MSE under caching.** Black solid lines show observed image MSE and blue dashed lines show predictions from the propagation-based predictor, averaged over the same 64 samples per model. Each candidate reuses the prediction cached at the start of one eight-step window, and windows are ordered from noise toward output.

Table 2: Spearman correlations with observed mean image MSE across eight caching windows, using the runs in Figure 4. Bold values mark the highest correlation for each model.

Model	Our predictor	Local defect	Path cost
JiT-B	**0.9524**	0.2143	-0.4762
JiT-L	**0.9762**	0.2381	-0.4286
ADM	**0.9762**	-0.9048	-0.9048
DiT	**0.9762**	0.5000	-0.3095

and reuses it for the remaining seven steps, while other steps use fresh predictions. The differences between cached and fresh predictions induce local update perturbations e_i along the reference trajectory, whose detailed derivation is provided in Appendix C.4.

Using their first-order propagation, we construct a predictor of the resulting image mean squared error (MSE) and compare its predictions with observed errors:

$$R_{\text{pred}} = \frac{1}{d}\left\|\sum_{i\in B} T_i e_i\right\|^2, \qquad R_{\text{obs}} = \frac{\|Y^{\text{cache}} - Y\|^2}{d}, \tag{12}$$

where d is the number of image entries and Y^{cache} is the final output when predictions are cached within B.

Figure 4 compares predicted and observed image MSE. The propagation-based predictor captures the broad variation in error across windows, although predicted and observed magnitudes can differ substantially. For JiT-B/L, the predictions also track the observed error magnitudes more closely, particularly in the middle and later windows.

In Table 2, our propagation-based predictor achieves Spearman correlations of 0.9524–0.9762 with observed mean image MSE, substantially exceeding the local-defect and path-cost baselines defined in Appendix C.5 across all four models. These results show that our propagation-based predictor can identify sampling windows in which caching causes larger output errors, which validated the analysis in Section 3.

6 Conclusion

In this work, we develop a theoretical framework explaining why path cost alone does not determine output effects in diffusion sampling. We separate the changes in output distribution from conditional trajectory changes and characterize the fraction of local output information captured by a selected feature mean. We express output responses as sums of propagated local perturbations, revealing how downstream dynamics, cross-time accumulation, and observation jointly shape measured effects. We use this theory to diagnose temporal and frequency-dependent sensitivity under equal path costs and identify caching windows with larger image deviations in the evaluated models.

AI USE STATEMENT

We used large language models to improve the grammar, phrasing, and clarity of the manuscript, and to assist with experimental coding. The authors formulated the research questions, developed the theoretical framework and proofs, designed and conducted the experiments, and interpreted the results. The scientific claims and conclusions are grounded in the authors' derivations and experimental findings. The authors take full responsibility for the correctness of the experimental implementation and the accuracy and integrity of the final manuscript, including all AI-assisted content.

ETHICS STATEMENT

This work studies perturbation propagation in diffusion sampling through theoretical analysis and experiments with publicly released pretrained models. It does not involve human participants or the collection of new personal data. Generated images may reflect biases and limitations inherited from the pretrained models and their training data. Insights into controlling and accelerating generation may also facilitate uses of synthetic imagery that carry risks of deception or misinformation. The experiments and accompanying code are intended to support research on the behavior of diffusion models, with use subject to the applicable model and data licenses.

REPRODUCIBILITY STATEMENT

We will provide complete code for reproducing the controlled perturbation experiments in Section 4 and the caching error diagnosis in Section 5. For Section 4, this includes measuring temporal and frequency-dependent responses at equal path cost, comparing local predictions with finite perturbation responses, and normalizing feature mean responses by baseline fluctuations. For Section 5, it includes computing propagation-based error predictions and baseline measures, measuring image errors under caching, and evaluating their rankings across sampling windows. The experimental setups and measurements are described in these sections, with detailed sampler configurations, perturbation normalization, and evaluation protocols in Appendix C. To support verification of the theoretical results in Section 3, Appendix A provides the assumptions, proofs, and derivations, and Appendix B specifies the Gaussian construction used to illustrate equal-cost perturbations with different output effects.

REFERENCES

N. Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1):90–93, 1974.

Shun-ichi Amari. *Information geometry and its applications*. Springer, 2016.

Roi Benita, Miki Elad, and Joseph Keshet. Spectral analysis of diffusion models with application to schedule design. *Advances in Neural Information Processing Systems*, 38:2073–2127, 2025.

Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *International Conference on Learning Representations*, volume 2024, pp. 36916–36936, 2024.

Julius Berner, Lorenz Richter, and Karen Ullrich. An optimal control perspective on diffusion-based generative modeling. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. *arXiv preprint arXiv:2209.11215*, 2022.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2 edition, 2006.

Valentin De Bortoli, Romuald Elie, Anna Kazeykina, Zhenjie Ren, and Jiacheng Zhang. Dimension-free error estimate for diffusion model and optimal scheduling. *arXiv preprint arXiv:2512.01820*, 2025.

Andreas Dechant and Shin-ichi Sasa. Fluctuation–response inequality out of equilibrium. *Proceedings of the National Academy of Sciences*, 117(12):6430–6436, 2020.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Paul Dupuis, Markos A Katsoulakis, Yannis Pantazis, and Petr Plechác. Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):80–111, 2016.

Yasaman Haghighi and Alexandre Alahi. SenCache: Accelerating diffusion model inference via sensitivity-aware caching. *arXiv preprint arXiv:2602.24208*, 2026.

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

Olav Kallenberg. *Foundations of modern probability*. Springer, 1997.

Nail B Khelifa, Richard E Turner, and Ramji Venkataramanan. Quantifying error propagation and model collapse in diffusion models. *arXiv preprint arXiv:2602.16601*, 2026.

Ryogo Kubo, Morikazu Toda, and Natsuki Hashitsume. *Statistical physics II: nonequilibrium statistical mechanics*. Springer Science & Business Media, 2012.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

Tianhong Li and Kaiming He. Back to basics: Let denoising generative models denoise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 36115–36125, 2026.

Xiuyu Li, Yijiang Liu, Long Lian, Huanrui Yang, Zhen Dong, Daniel Kang, Shanghang Zhang, and Kurt Keutzer. Q-Diffusion: Quantizing diffusion models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17489–17499. IEEE, 2023.

Yangming Li and Mihaela van der Schaar. On error propagation of diffusion models. In *International Conference on Learning Representations*, volume 2024, pp. 32791–32807, 2024.

Yuhan Liu, Zongwei Hong, Jinglun Li, Linze Li, Shen Zhang, and Yao Tang. EpaCache: Error-propagation-aware caching for accelerating diffusion-based visual generation. *arXiv preprint arXiv:2608.29264*, 2026.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. *Advances in neural information processing systems*, 35:5775–5787, 2022.

Xinyin Ma, Gongfan Fang, and Xinchao Wang. DeepCache: Accelerating diffusion models for free. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15762–15772. IEEE, 2024.

Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. Monte Carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

Mang Ning, Enver Sangineto, Angelo Porrello, Simone Calderara, and Rita Cucchiara. Input perturbation reduces exposure bias in diffusion models. *arXiv preprint arXiv:2301.11706*, 2023.

Mang Ning, Mingxiao Li, Jianlin Su, Albert Ali Salah, and Itir Onal Ertugrul. Elucidating the exposure bias in diffusion models. In *International Conference on Learning Representations*, volume 2024, pp. 15167–15189, 2024.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

Xurui Peng, Chenqian Yan, Hong Liu, Rui Ma, Fangmin Chen, Xing Wang, Zhihua Wu, Songwei Liu, and Mingbao Lin. Ertacache: Error rectification and timesteps adjustment for efficient diffusion. In *International Conference on Learning Representations*, volume 2026, pp. 82104–82120, 2026.

David Pollard. A note on insufficiency and the preservation of fisher information. In *From Probability to Statistics and Back: High-Dimensional Models and Processes–A Festschrift in Honor of Jon A. Wellner*, volume 9, pp. 266–276. Institute of Mathematical Statistics, 2013.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in neural information processing systems*, 34:1415–1428, 2021.

A PROOFS AND DERIVATIONS

We follow the order of the theoretical framework in Section 3. We first derive the exact path cost and its decomposition into output and conditional-trajectory changes. We then prove the propagation identity, establish the local Fisher decomposition, and derive the feature-response bound and the main response identity. Throughout, P and Q_ε denote the joint distributions of reference and perturbed sampling trajectories, respectively.

A.1 DERIVATION OF THE PATH COST FOR LOCAL PERTURBATIONS

We derive how local changes in the sampling updates determine the distinguishability of complete trajectories. Starting from the Gaussian sampling rule in Eq. (1), we consider perturbations that shift the conditional transition mean while preserving the covariance function and initial distribution. We first express the trajectory likelihood ratio through the sampling transitions, then evaluate each transition's contribution to obtain Eq. (2).

Reference and perturbed transitions. We write $H_i = (X_0, \dots, X_i)$ for the sampling history available before update i. Common conditioning variables, when present, are included in X_0 and remain fixed during generation. To allow a local perturbation to depend on previous sampling states, we write its mean displacement as $\varepsilon u_i(H_i)$. The reference and perturbed conditional transition distributions are

$$\begin{aligned} K_i^0(\cdot \mid H_i) &= \mathcal{N}(\mu_i(X_i), \Lambda_i(X_i)), \\ K_i^\varepsilon(\cdot \mid H_i) &= \mathcal{N}(\mu_i(X_i) + \varepsilon u_i(H_i), \Lambda_i(X_i)). \end{aligned} \tag{13}$$

Here $u_i = 0$ at unperturbed steps. We assume $\Lambda_i(X_i)$ is positive definite, so both conditional densities are strictly positive. Together with the shared initial distribution, this ensures $Q_\varepsilon \ll P$ (Kallenberg, 1997). We take the perturbations to have finite expected quadratic cost under Q_ε, so the path KL derived below is finite.

From trajectory likelihoods to transition costs. Let k_i^0 and k_i^ε denote the densities of the two kernels in Eq. (13). Recall that $\Gamma = (X_0, \dots, X_M)$ denotes a complete sampling trajectory. We factor the trajectory likelihood ratio into conditional density ratios:

$$L_\varepsilon(\Gamma) := \frac{\mathrm{d}Q_\varepsilon}{\mathrm{d}P}(\Gamma) = \prod_{i=0}^{M-1} \ell_i^\varepsilon(X_{i+1}, H_i), \qquad \ell_i^\varepsilon(x, H_i) := \frac{k_i^\varepsilon(x \mid H_i)}{k_i^0(x \mid H_i)}. \tag{14}$$

The common initial distribution cancels from this ratio. We take logarithms, average under Q_ε, and condition each term on H_i to obtain

$$\begin{aligned} \mathcal{C}(\varepsilon) &= \mathbb{E}_{Q_\varepsilon} \log L_\varepsilon(\Gamma) \\ &= \sum_{i=0}^{M-1} \mathbb{E}_{Q_\varepsilon}\left[\mathbb{E}_{Q_\varepsilon}\left[\log \ell_i^\varepsilon(X_{i+1}, H_i) \mid H_i\right]\right] \\ &= \sum_{i=0}^{M-1} \mathbb{E}_{Q_\varepsilon} D_{\mathrm{KL}}\big(K_i^\varepsilon(\cdot \mid H_i) \big\| K_i^0(\cdot \mid H_i)\big). \end{aligned} \tag{15}$$

This follows from the relative-entropy chain rule (Cover & Thomas, 2006). Each conditional KL compares the two possible updates at the same history; the outer expectation averages over histories reached by the perturbed process.

Evaluating the Gaussian contribution. We now compute each conditional KL explicitly. At a fixed history, we abbreviate $\mu_i = \mu_i(X_i)$, $\Lambda_i = \Lambda_i(X_i)$, and $u_i = u_i(H_i)$, and define the residual $r_i = X_{i+1} - \mu_i$. Since the two Gaussian kernels share their covariance, their density normalization factors cancel. Expanding their quadratic exponents gives

$$\begin{aligned} \log \ell_i^\varepsilon(X_{i+1}, H_i) &= -\frac{1}{2}(r_i - \varepsilon u_i)^\top \Lambda_i^{-1}(r_i - \varepsilon u_i) + \frac{1}{2} r_i^\top \Lambda_i^{-1} r_i \\ &= \varepsilon u_i^\top \Lambda_i^{-1} r_i - \frac{\varepsilon^2}{2} u_i^\top \Lambda_i^{-1} u_i. \end{aligned} \tag{16}$$

Under the perturbed transition, $\mathbb{E}_{Q_\varepsilon}[r_i \mid H_i] = \varepsilon u_i$. Taking the conditional expectation therefore yields

$$D_{\mathrm{KL}}\big(K_i^\varepsilon(\cdot \mid H_i)\big\|K_i^0(\cdot \mid H_i)\big) = \frac{\varepsilon^2}{2} u_i^\top \Lambda_i^{-1} u_i. \tag{17}$$

Substituting into Eq. (15), we obtain the exact expression stated in Eq. (2):

$$\mathcal{C}(\varepsilon) = D_{\mathrm{KL}}(Q_\varepsilon \| P) = \frac{\varepsilon^2}{2} \mathbb{E}_{Q_\varepsilon} \sum_{i=0}^{M-1} u_i^\top \Lambda_i^{-1} u_i. \tag{18}$$

Thus the trajectory-level change is the expected sum of local mean displacements weighted by inverse transition covariance. For $\Lambda_i = \nu_i I$, each contribution is $\varepsilon^2 \|u_i\|^2/(2\nu_i)$, so an equal displacement has greater cost at a lower noise level. This identity holds at finite perturbation strength. The expectation remains under Q_ε as the perturbation also changes which histories the sampler visits.

A.2 Proof of Proposition 1: Decomposition of Path and Output Changes

We prove Proposition 1 by factoring the trajectory likelihood ratio into output and conditional-trajectory likelihood ratios (Léonard, 2013). We assume that the trajectory space is standard Borel and that Y is a measurable observation. Suppose $Q_\varepsilon \ll P$ has finite KL, with likelihood $L_\varepsilon = \mathrm{d}Q_\varepsilon/\mathrm{d}P$. Under these assumptions, regular conditional laws exist. For every bounded measurable g, we have

$$\mathbb{E}_{q_{\varepsilon,Y}} g(Y) = \mathbb{E}_P[L_\varepsilon g(Y)] = \mathbb{E}_P[\mathbb{E}_P(L_\varepsilon \mid Y) g(Y)]. \tag{19}$$

Identifying the output likelihood ratio, we obtain

$$\ell_\varepsilon(Y) := \frac{\mathrm{d}q_{\varepsilon,Y}}{\mathrm{d}p_Y}(Y) = \mathbb{E}_P[L_\varepsilon \mid Y]. \tag{20}$$

Thus, when only the output is observed, the path likelihood is averaged over all histories compatible with that output. Disintegrating Q_ε and P with respect to Y gives

$$\frac{\mathrm{d}Q_\varepsilon(\cdot \mid Y)}{\mathrm{d}P(\cdot \mid Y)} = \frac{L_\varepsilon}{\ell_\varepsilon(Y)} \quad Q_\varepsilon\text{-a.s.} \tag{21}$$

This proves Eq. (3): the first expectation is output KL, and the second is conditional-history KL averaged over the perturbed output law. Nonnegativity of both terms gives $0 \le \mathcal{C}_Y \le \mathcal{C}$.

A.3 Proof of Proposition 2: Perturbation propagation

We prove Proposition 2 by differentiating the discrete sampler on its actual sampling grid. Consider $X_{i+1}^\varepsilon = \mathcal{G}_i^\varepsilon(X_i^\varepsilon, \xi_i)$ with the initial state and innovations held fixed. Assume the step maps are differentiable in state and perturbation strength along the reference trajectories, and ϕ is differentiable at their endpoints, almost surely. For mean response, we additionally impose the integrability and interchange conditions in Appendix A.5. Piecewise differentiable maps are treated at their differentiable points.

Forward transport. At fixed reference state and perturbation strength, we define $J_i = D_x \mathcal{G}_i^0$ and $e_i = \partial_\varepsilon \mathcal{G}_i^\varepsilon|_0$. Define $\delta X_i = \partial_\varepsilon X_i^\varepsilon|_0$. Applying the chain rule gives

$$\delta X_{i+1} = J_i \delta X_i + e_i, \qquad \delta X_0 = 0. \tag{22}$$

The first term propagates state changes from earlier steps, while the second accounts for the local change in the current update. When $e_i = 0$, an earlier state change can still propagate through J_i. Iterating gives

$$\delta X_M = \sum_i J_{M-1} \cdots J_{i+1} e_i. \tag{23}$$

Differentiating $Y^\varepsilon = \phi(X_M^\varepsilon)$ then gives Eq. (4). For a scalar output feature, the first-order expansion of its mean is

$$\mathbb{E}_{Q_\varepsilon} f(Y) = \mathbb{E}_P f(Y) + \varepsilon \chi_f + o(\varepsilon), \qquad \chi_f = \partial_\varepsilon \mathbb{E}_{Q_\varepsilon} f(Y)|_0 . \tag{24}$$

Under the stated interchange conditions, substituting the expression for V into $\chi_f = \mathbb{E}_P[\nabla f(Y)^\top V]$ yields Eq. (5). For mean controls, $\mathcal{G}_i^\varepsilon = \mathcal{G}_i^0 + \varepsilon u_i$ and $e_i = u_i(X_i)$; the control's state derivative is multiplied by ε and contributes no additional term at zero. We define the local defects induced by output caching in Appendix C.4 and propagate them using this reference recursion.

Cross-time accumulation and statistical visibility. Writing $v_i = T_i e_i$, we expand $\langle \sum_i v_i, \sum_j v_j \rangle$ to obtain Eq. (6). The off-diagonal terms retain alignment between transported changes from different steps. Unlike the conditional likelihood scores in Eq. (30) in Appendix A.4 , these output vectors need not be orthogonal in expectation.

Adjoint evaluation. For a scalar statistic, we can evaluate the same response by propagating its gradient backward. We define $a_M = D\phi(X_M)^\top \nabla f(Y)$ and $a_i = J_i^\top a_{i+1}$. Combining this recursion with the tangent recursion, we obtain

$$a_{i+1}^\top \delta X_{i+1} - a_i^\top \delta X_i = a_{i+1}^\top e_i, \qquad \partial_\varepsilon f(Y^\varepsilon)|_0 = \sum_i a_{i+1}^\top e_i. \tag{25}$$

We derive the second identity by telescoping the first. Each term pairs the local injection with the gradient of the final statistic transported back to that step. For $\mathcal{G}_i = \mu_i + L_i \xi_i$, both tangent and adjoint recursions include the state derivative of L_i. With a deterministic time-zero map h_0 and decoder D (including latent scaling), we obtain the terminal adjoint $a_M = J_{h_0}^\top J_D^\top \nabla f(Y)$, with each Jacobian evaluated at its corresponding input. These terms ensure that the derivative describes the observed output, including the final transition and decoding.

A.4 Path Fisher information and output visibility

We next derive the local version of the exact decomposition. Recall that $S = \partial_\epsilon \log \frac{\mathrm{d}Q_\epsilon}{\mathrm{d}P}\Big|_{\epsilon=0}$ is the path likelihood score. We assume the path likelihood admits the expansion

$$L_\varepsilon = 1 + \varepsilon S + r_\varepsilon, \qquad \|r_\varepsilon\|_{L^2(P)} = o(|\varepsilon|), \qquad S \in L^2(P). \tag{26}$$

This condition controls the likelihood remainder in mean square and suffices for the response identities below. To obtain second-order KL expansions, we additionally assume that, for some $\delta > 0$ and $\varepsilon_0 > 0$,

$$\sup_{0<|\varepsilon|<\varepsilon_0} \mathbb{E}_P \left| \frac{L_\varepsilon - 1}{\varepsilon} \right|^{2+\delta} < \infty. \tag{27}$$

Normalization gives $\mathbb{E}_P S = 0$. Conditional expectation contracts the L^2 remainder in Eq. (26), so Eq. (20) has local score $S_Y = \mathbb{E}_P[S \mid Y]$. Because $S - S_Y$ is orthogonal to every square-integrable function of Y, we obtain

$$\mathbb{E}_P S^2 = \mathbb{E}_P S_Y^2 + \mathbb{E}_P (S - S_Y)^2 = F_Y + \mathbb{E}_P \operatorname{Var}_P(S \mid Y). \tag{28}$$

This proves Eq. (9), where S_Y captures the component of the path score resolvable from outputs, while its residual accounts for the unresolved component. For a coarsening $\bar{Y} = \psi(Y)$, the tower property gives $S_{\bar{Y}} = \mathbb{E}_P[S_Y \mid \bar{Y}]$. Using Jensen's inequality then yields $F_{\bar{Y}} \le F_Y$.

To connect these quantities to KL curvature, we set $x_\varepsilon = L_\varepsilon - 1$. The function $k(x) = (1+x)\log(1+x) - x$, extended continuously at $x = -1$, satisfies $k(x) = x^2/2 + o(x^2)$ near zero and $k(x) \le Cx^2$ for $x \ge -1$. Equations (26)–(27) make the normalized squared increments uniformly integrable. Using $\mathbb{E}_P x_\varepsilon = 0$, we can therefore take expectations of the expansion to obtain

$$D_{\mathrm{KL}}(Q_\varepsilon \| P) = \tfrac{1}{2}\varepsilon^2 F_{\mathrm{path}} + o(\varepsilon^2), \qquad D_{\mathrm{KL}}(q_{\varepsilon,Y} \| p_Y) = \tfrac{1}{2}\varepsilon^2 F_Y + o(\varepsilon^2), \tag{29}$$

which is exactly desired form of Eq. (8). Since the output likelihood ratio is the conditional expectation of the path likelihood ratio given the output, conditional Jensen implies that the observed likelihood increments inherit the same moment bound as the path likelihood increments. Subtracting these expansions in Eq. (3), we also obtain $\mathcal{C}_{\mathrm{bridge}|Y}(\varepsilon) = \frac{1}{2}\varepsilon^2 F_{\mathrm{bridge}|Y} + o(\varepsilon^2)$. Thus F_Y / F_{path} measures the locally visible fraction whenever $F_{\mathrm{path}} > 0$.

Additivity and Gaussian mean perturbations. For differentiable conditional kernels with the same initial law, we define the conditional scores $s_i = \partial_\varepsilon \log \ell_i^\varepsilon(X_{i+1}, H_i)|_0$ and $F_i(H_i) = \mathbb{E}_P[s_i^2 \mid H_i]$. Assume that the scores are square integrable and differentiation can be interchanged with conditional normalization. Equation (14) then gives $S = \sum_i s_i$ and $\mathbb{E}_P[s_i \mid H_i] = 0$. For $i < j$, s_i is measurable with respect to H_j, so $\mathbb{E}_P[s_i s_j] = \mathbb{E}_P[s_i \mathbb{E}_P(s_j \mid H_j)] = 0$. We therefore obtain

$$F_{\mathrm{path}} = \mathbb{E}_P S^2 = \sum_{i=0}^{M-1} \mathbb{E}_P F_i(H_i). \tag{30}$$

For Gaussian mean perturbations, differentiating Eq. (16) gives $s_i = u_i^\top \Lambda_i^{-1}(X_{i+1} - \mu_i)$. Using the reference residual's conditional covariance Λ_i, we obtain $F_i(H_i) = u_i^\top \Lambda_i^{-1} u_i$ and hence

$$F_{\mathrm{path}} = \mathbb{E}_P \sum_{i=0}^{M-1} u_i^\top \Lambda_i^{-1} u_i. \tag{31}$$

A.5 Proof of Proposition 3: Feature response and output information

We prove Proposition 3 by differentiating the expectation of a statistic. For $f(Y) \in L^2(P)$, Eq. (26) and Cauchy–Schwarz justify differentiating $\mathbb{E}_P[f(Y)L_\varepsilon]$ at zero. Conditioning on Y and using $\mathbb{E}_P S_Y = 0$, we derive

$$\chi_f = \mathbb{E}_P[f(Y)S] = \mathbb{E}_P[f(Y)S_Y] = \mathbb{E}_P[(f(Y) - \mathbb{E}_P f(Y))S_Y]. \tag{32}$$

Applying Cauchy–Schwarz once more gives $\chi_f^2 \le \mathrm{Var}_P(f(Y))F_Y$. Combining this with Eq. (9) proves the bound in Eq. (10).

A.6 Proof of Theorem 1: Perturbation propagation and measured output response

We use $\mathbb{E}_P S_Y = 0$ and Eq. (10) to obtain $\chi_f = \mathrm{Cov}_P(f(Y), S_Y)$. For $F_Y > 0$, we have $\mathrm{Var}_P(S_Y) = F_Y$ and derive

$$\frac{B_f}{F_Y} = \frac{\chi_f^2}{\mathrm{Var}_P(f(Y))F_Y} = \mathrm{Corr}_P(f(Y), S_Y)^2. \tag{33}$$

Multiplying by F_Y/F_{path} and substituting the propagated response from Eq. (5) proves Eq. (11). If $F_Y = 0$, Proposition 3 gives $\chi_f = B_f = 0$, completing the proof.

Thus the response depends on the statistic's alignment with the output score. The centered unit-variance statistic $f(Y) = S_Y/\sqrt{F_Y}$ attains equality in the response bound. If $F_Y = 0$, every square-integrable mean response vanishes. We consequently obtain

$$F_Y = \sup_{\mathbb{E}_P f(Y)=0,\, \mathbb{E}_P f(Y)^2=1} \chi_f^2, \tag{34}$$

whenever the class contains a unit-variance observable. A chosen statistic measures one projection of this response geometry; optimizing over all such statistics recovers the full local output information.

Fluctuation scale and the limits of a selected mean. Writing $\Delta_f(\varepsilon) = \mathbb{E}_{Q_\varepsilon} f(Y) - \mathbb{E}_P f(Y)$, we have $\Delta_f(\varepsilon) = \varepsilon\chi_f + o(\varepsilon)$. Thus $\varepsilon^2 B_f$ is the leading squared mean shift in units of the reference feature variance. For $F_{\mathrm{path}} > 0$, we obtain

$$\lim_{\varepsilon\to 0} \frac{\Delta_f(\varepsilon)^2}{2\,\mathrm{Var}_P(f(Y))\,\mathcal{C}(\varepsilon)} = \frac{B_f}{F_{\mathrm{path}}}.$$

Replacing f by $af+b$ with $a \neq 0$ multiplies both its squared response and variance by a^2, leaving B_f unchanged. A smaller absolute response can therefore correspond to a larger B_f when the reference variance is sufficiently smaller. The covariance identity also shows that output changes orthogonal to the centered feature have zero first-order mean response, even when $F_Y > 0$. This concerns the feature's mean, rather than all changes in its distribution. When $F_Y = 0$, Eq. (10) directly gives $\chi_f = 0$; Eq. (11) is not used because correlation with a zero-variance score is undefined. Predicting finite mean shifts by $\varepsilon\chi_f$ additionally requires the higher-order remainder to be negligible.

To connect this statistical derivative with a dynamical derivative, we realize Y^ε using common initial randomness and innovations, with marginal law $q_{\varepsilon,Y}$. Assume f is continuously differentiable, $\partial_\varepsilon Y^\varepsilon|_0 = V$ exists almost surely, and $[f(Y^\varepsilon) - f(Y)]/\varepsilon$ is dominated by an integrable random variable near zero. We exchange differentiation and expectation and apply the chain rule to obtain

$$\chi_f = \partial_\varepsilon \mathbb{E} f(Y^\varepsilon)|_0 = \mathbb{E}[\nabla f(Y)^\top V]. \tag{35}$$

This is the same distributional mean derivative as the likelihood calculation. Substituting the expansion of V from Eq. (4) yields Eq. (11).

B GAUSSIAN CONSTRUCTION FOR EQUAL-COST PERTURBATIONS

We give the Gaussian reverse-diffusion construction used in Figure 1, in which perturbations with equal path cost reinforce or cancel at the output. Our goal is to hold the noise-normalized intervention strength fixed while changing how its contributions align after propagation. We take $C_0 = \mathrm{diag}(1/4, 1)$ and the forward chain $Z_{s+1} = \sqrt{\alpha} Z_s + \sqrt{1-\alpha}\xi_s$, with $\alpha = 0.9$, $M = 24$, and independent standard Gaussian innovations. We obtain the forward covariances $C_s = \alpha^s C_0 + (1-\alpha^s) I$. Conditioning the joint Gaussian law, we derive the exact reverse kernels in generation order, $X_i = Z_{M-i}$:

$$X_{i+1} = A_i X_i + \Lambda_i^{1/2} \xi_i, \quad A_i = \sqrt{\alpha} C_{s-1} C_s^{-1}, \quad \Lambda_i = C_{s-1} - A_i C_s A_i^\top, \quad s = M - i. \tag{36}$$

We start from $X_0 \sim \mathcal{N}(0, C_M)$ and observe $Y = X_M$, whose reference covariance is C_0. We apply deterministic mean controls:

$$X_{i+1} = A_i X_i + \Lambda_i^{1/2} \xi_i + \varepsilon u_i, \qquad u_i = \sqrt{(\Lambda_i)_{11}} w_i e_1, \qquad Y = X_M. \tag{37}$$

To make the propagation explicit, we define

$$H_i = A_{M-1} \cdots A_{i+1}, \qquad h_i = (H_i)_{11} \sqrt{(\Lambda_i)_{11}}, \tag{38}$$

with $H_{M-1} = I$. Here H_i transports the injected offset to the endpoint, and h_i includes its transition-noise scale. Let $w = (w_0, \ldots, w_{M-1})^\top$ be the temporal control vector, and $h = (h_0, \ldots, h_{M-1})^\top \in \mathbb{R}^M$ collect the downstream gains h_i. We obtain the controlled endpoint mean $m_Y = \varepsilon(h^\top w) e_1$, and its covariance remains C_0. Using Eq. (18) and the Gaussian endpoint KL, we derive

$$F_{\mathrm{path}} = \|w\|^2, \qquad F_Y = \frac{(h^\top w)^2}{(C_0)_{11}}, \qquad \mathcal{C} = \tfrac{1}{2}\varepsilon^2 \|w\|^2, \qquad \mathcal{C}_Y = \frac{\varepsilon^2 (h^\top w)^2}{2(C_0)_{11}}. \tag{39}$$

We choose the reinforcing direction $w_+ = h/\|h\|$. For the cancelling direction, we define $r_i = (w_+)_i$ for $i < 12$ and $r_i = -(w_+)_i$ otherwise, and set

$$w_- = \frac{r - w_+ (w_+^\top r)}{\|r - w_+ (w_+^\top r)\|}. \tag{40}$$

Both directions have unit norm, and $h^\top w_- = 0$. Thus the budget depends on the norm of the temporal control, while output change depends on its projection onto the downstream gains. The cancelling control consequently preserves the full endpoint Gaussian law at every amplitude while changing the intermediate path law. At $\varepsilon = 2$, the endpoint/bridge KL values are $(1.957579, 0.042421)$ for the reinforcing control and $(0, 2)$ for the cancelling control, as shown in Figure 1. The intermediate mean obeys $m_{i+1} = A_i m_i + \varepsilon u_i$, with $m_0 = 0$.

The observable $f(Y) = Y_1$ has $\chi_f = h^\top w$ and $\mathrm{Var}_P(f(Y)) = 1/4$, hence $B_f = F_Y$ in this Gaussian mean-shift family.

C EXPERIMENTAL PROTOCOLS AND MEASUREMENTS

C.1 MEASURING OUTPUT RESPONSES

For each sample, we generate two outputs, Y_n^+ and Y_n^-, by applying perturbations of strengths $+\varepsilon$ and $-\varepsilon$. We keep the class, initial noise, noise drawn at every sampling step, and spatial template identical in both runs. We then divide the output difference by 2ε to estimate the response to the perturbation.

For JiT and ADM, we project each generated image onto the DCT template used for the perturbation and average the resulting response over N samples:

$$\widehat{\chi}_f = \frac{1}{N}\sum_{n=1}^{N} \frac{f(Y_n^+) - f(Y_n^-)}{2\varepsilon}, \qquad f(Y) = \langle q_k, Y\rangle, \tag{41}$$

where q_k is the unit-norm image template defined below. For DiT, we compute the root-mean-square response across the decoded image entries for each sample, then average over samples:

$$\widehat{R} = \frac{1}{N}\sum_{n=1}^{N} \frac{\|Y_n^+ - Y_n^-\|}{2\varepsilon\sqrt{d}}, \qquad d = 3\cdot 256^2. \tag{42}$$

For the frequency comparisons in Section 4.4, we estimate the variance s_f^2 of each DCT projection from unperturbed images, using the sample variance with denominator $N-1$ across classes and noise realizations. We compute $\widehat{B}_f = \widehat{\chi}_f^2/s_f^2$, using the same baseline variance for all injection windows of a given feature. We report standard errors of the response means by dividing the sample standard deviation of the per-sample responses by $\sqrt{N}$.

C.2 REFERENCE SAMPLERS AND PERTURBATION DESIGN

JiT. The conditional clean predictor $D_\theta(z,t,c)$ defines drift

$$b_t(z,c) = \frac{D_\theta(z,t,c) - z}{1-t} + \lambda_0\big(tD_\theta(z,t,c) - z\big), \qquad \sigma_t = \sqrt{2\lambda_0}(1-t). \tag{43}$$

The parameter λ_0 scales the added diffusion and its matched score drift. We use $\lambda_0 = 0.1$, standard normal pixel initialization, and Euler transitions on $t_i = i\Delta t$, $\Delta t = t_*/128$, terminating at $t_* = 0.95$. Their covariance is $\Lambda_i = \nu_i I$ with $\nu_i = 2\lambda_0(1-t_i)^2\Delta t$. We obtain the generated output as $Y = D_\theta(Z_{128}, t_*, c)$. This stochastic representation is induced by the Gaussian interpolation $Z_t = tX + (1-t)E$: with an exact clean predictor, adding score drift $\lambda_0(tD_\theta - z)$ and its matched diffusion preserves the interpolation marginals. With a learned predictor, the displayed kernel specifies the reference law used for the comparisons.

ADM. We use the class-conditional ImageNet-256 ADM U-Net. The noise-prediction channels give the clean prediction, clipped to $[-1,1]$. The original 1000-step linear schedule of the denoising diffusion probabilistic model (DDPM) is respaced to 129 retained times: 128 positive-variance transitions with fixed-small posterior covariance, followed by a deterministic time-zero transition. The resulting pixel image is Y. The variance-output channels are unused in this fixed-covariance reference.

DiT. DiT-XL/2 evolves a $4\times32\times32$ latent state. The 1000-step linear DDPM schedule is similarly respaced to 129 retained times. The network's noise and variance outputs specify the transition mean and diagonal learned-range covariance $L_iL_i^\top$, with clean-latent clipping disabled. After 128 stochastic transitions, the observed output is

$$Y = D_{\mathrm{VAE}}\big(h_0(Z_{128})/0.18215\big), \tag{44}$$

Here h_0 is the deterministic time-zero transition and D_{VAE} is the variational autoencoder (VAE) decoder. We use the unclamped decoded pixels for response measurements . We compute the output response through the sampler and decoder, including the variance head, h_0, and latent scaling. All three model families use conditional predictions without classifier or classifier-free guidance amplification.

Spatial templates and temporal normalization. On an $H \times H$ grid, let $\psi_k(r, s)$ be proportional to $\cos(\pi k(s + 1/2)/H)$ and normalized to unit Frobenius norm. These are horizontal modes of the type-II discrete cosine transform: $k = 0$ is the constant DC mode, and $k > 0$ is labeled DCT-k. We use $k \in \{0, 1, 4, 16\}$ for all models and additionally $k = 31$ for DiT. Pixel controls use $q_k = (1, 1, 1)^\top \otimes \psi_k/\sqrt{3}$, with $H = 256$. The same template defines $f_k(Y) = \langle q_k, Y\rangle$. For a window B and isotropic transition covariance $\nu_i I$, both pixel implementations produce the constant mean offset

$$u_i = c_B q_k \, \mathbf{1}\{i \in B\}, \qquad c_B = \left(\sum_{i \in B} \nu_i^{-1}\right)^{-1/2}. \tag{45}$$

For JiT this is equivalently a constant drift perturbation $c_B q_k/\Delta t$ within the window. DiT uses $H = 32$, $q_k = r \otimes \psi_k$ with per-sample channel signs $r_c \in \{-1/2, 1/2\}$, and

$$u_i(Z_i) = \frac{L_i(Z_i) q_k}{\sqrt{|B|}} \, \mathbf{1}\{i \in B\}. \tag{46}$$

In both cases the sum of conditional mean-shift Fisher costs is exactly one, giving path KL $\varepsilon^2/2$ by Eq. (18). Windows index progress through each model's sampling schedule.

C.3 Models and sampling configurations

Table 3: Sample counts N and perturbation amplitudes ε for the controlled perturbation experiments in Section 4. All use 128 stochastic steps, eight injection windows, and path Fisher $F_{\mathrm{path}} = 1$.

Model	N	ε
JiT-B/16	512	0.25
JiT-L/16	512	0.25
ADM	256	0.25
DiT-XL/2	256	0.80

We use publicly released class-conditional ImageNet-256 checkpoints for JiT-B/16 and JiT-L/16 (Li & He, 2026), ADM (Dhariwal & Nichol, 2021), and DiT-XL/2 (Peebles & Xie, 2023). For JiT, we use exponential moving average (EMA) weights. Table 3 lists the sample counts and perturbation amplitudes.

We use fixed class subsets, with one noise realization per sample; the reported variability therefore includes both class and noise variation. For comparisons across sampling windows, we keep the classes, initial states, and noise at every step fixed. The DiT local/finite comparison uses 32 paired samples.

C.4 Caching protocol and local defects

We use the reference kernels and observations in Appendix C.2, with $M = 64$ stochastic steps. JiT uses $\Delta t = 0.95/64$; ADM and DiT use 65 retained diffusion times, comprising 64 stochastic transitions and the final deterministic transition. Each model has 64 paired samples and eight candidate windows $B = \{a, \ldots, a + 7\}$. For each candidate, we store the network prediction at step a and reuse it directly for the next seven steps, omitting seven network evaluations. Outside B, we evaluate fresh predictions. The final deterministic prediction and decoder are also evaluated afresh. We pair the cached and reference runs using the same class, initial state, and noise at every step.

To construct the propagation-based predictor, we evaluate the prediction differences along the reference trajectory. Let $o_i(X_i)$ denote the fresh prediction at step i. For $i \in B$, the prediction defect is $\Delta o_i = o_a(X_a) - o_i(X_i)$; outside B, it is zero. Writing $\mathcal{H}_i(x, o, \xi)$ for the sampling update with a supplied prediction o, we define its local response to this defect by

$$e_i = D_o \mathcal{H}_i(X_i, o_i(X_i), \xi_i)[\Delta o_i]. \tag{47}$$

We evaluate this directional derivative at the reference state and prediction, holding the state and noise fixed. The reference Jacobian J_i includes the state dependence of the fresh network prediction.

We then propagate the local defects using

$$\delta Z_{i+1} = J_i \delta Z_i + e_i, \qquad \delta Z_0 = 0, \qquad V = \delta Y = \sum_i T_i e_i. \tag{48}$$

At the first step of the window, $e_a = 0$. We compare $\|V\|^2/d$ with the observed image MSE from direct reuse, as in Eq. (12). This predictor is a first-order approximation along the reference trajectory; its accuracy for the finite caching error is assessed empirically.

For JiT, o_i is the clean prediction and the local update defect is $e_i = \Delta t\,[b_i(Z_i; o_a) - b_i(Z_i; o_i)]$. For ADM, o_i is the clipped clean prediction, the posterior covariance is fixed, and $e_i = \mu_i(Z_i; o_a) - \mu_i(Z_i; o_i)$. Both means are affine in their supplied clean predictions, so these directional derivatives equal the update differences at the fixed reference state.

For DiT, we reuse the complete noise-and-variance prediction. Write $\Sigma_i = L_i L_i^\top$ for its diagonal covariance and $\ell_i = \log \operatorname{diag} \Sigma_i$. At the reference state and prediction, define $\dot{\mu}_i = D_o \mu_i[\Delta o_i]$ and $\dot{\ell}_i = D_o \ell_i[\Delta o_i]$. The reparameterized step has local response

$$e_i = \dot{\mu}_i + \tfrac{1}{2} L_i \operatorname{diag}(\dot{\ell}_i)\xi_i. \tag{49}$$

All state Jacobians include the learned covariance derivative. We propagate the response through the deterministic time-zero map, latent scaling, and VAE decoder. Thus the predicted and observed errors are both measured in decoded pixels.

C.5 Caching error prediction and evaluation

For each sample and candidate window, we compute R_{pred} using the tangent recursion. The local-defect baseline is $L_{\mathrm{state}} = \sum_i \|e_i\|^2/d_z$, where d_z is the pixel-state dimension for JiT/ADM and the latent-state dimension for DiT.

The path-cost baseline uses the local quadratic approximation $L_{\mathrm{path}} = \sum_i F_i/2$. At fixed covariance, $F_i = e_i^\top \Sigma_i^{-1} e_i$. For DiT's diagonal Gaussian transition,

$$F_i = \sum_j \left[\frac{\dot{\mu}_{i,j}^2}{\Sigma_{i,jj}} + \frac{1}{2} \dot{\ell}_{i,j}^2 \right]. \tag{50}$$

Here F_i measures conditional Fisher information in the direction of the prediction defect, with the trajectory history held fixed. Its reference expectation summed over steps gives the corresponding path Fisher information. The resulting baseline is a local approximation to path cost, rather than an evaluation of the finite path KL under caching. We average each baseline over samples and compare its ranking across windows with that of observed mean image MSE.

C.6 Image visualization protocol

The DiT examples use DCT-4 perturbations in windows 4 and 8 with $\varepsilon = 0.8$, giving path KL 0.32 in each window. Images are clipped to $[-1, 1]$ for display; differences use unclipped decoded pixels. Difference maps show per-pixel RGB RMS relative to the reference, with a common color scale across samples and windows. The scale saturates at the pooled 99.5th percentile. Central-response annotations report the full-image RMS of $(Y^+ - Y^-)/(2\varepsilon)$.

D SUPPLEMENTARY EXPERIMENTAL RESULTS

D.1 DIT IMAGE EXAMPLES AT EQUAL PATH COST

Figures 5 and 6 show two fixed DiT samples under DCT-4 perturbations in windows 4 and 8. Each perturbation has path cost $\mathcal{C} = 0.32$ at $\varepsilon = 0.8$. The central-response RMS is $19.88\times$ and $11.23\times$ larger in window 4 than in window 8 for the two samples, respectively, illustrating the dependence of output response on perturbation timing at equal path cost. Both examples use 128 stochastic steps and the display protocol in Appendix C.6.

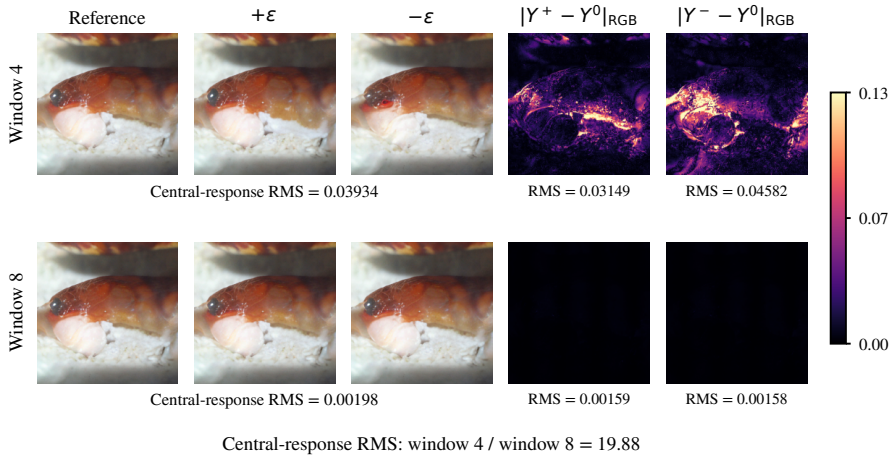


Figure 5: **DiT image responses to equal-cost perturbations.** A fixed sample under positive and negative DCT-4 perturbations in windows 4 and 8. Difference maps show per-pixel RGB RMS relative to the reference Y^0, with full-image RMS below each map. Central-response annotations measure the RMS of $(Y^+ - Y^-)/(2\varepsilon)$; the window-4/window-8 ratio is 19.88. Differences use unclipped pixels and a shared color scale across both examples.

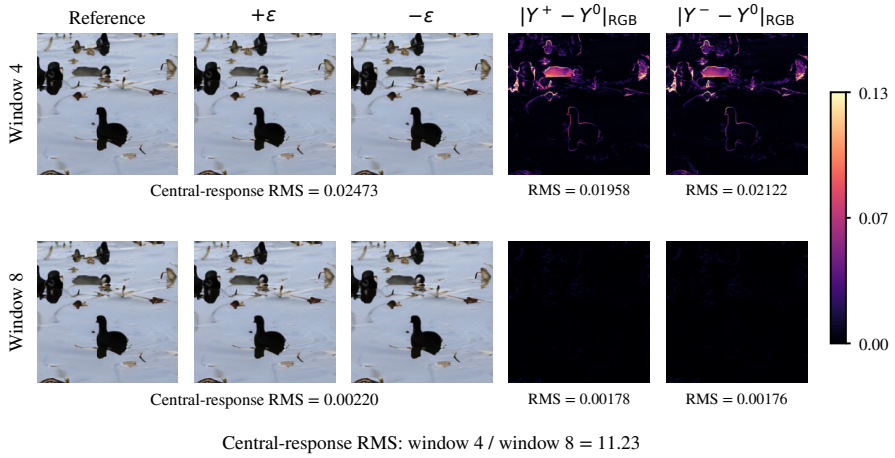


Figure 6: Second DiT example, with the endpoints, branch-difference maps, and shared color scale defined in Appendix C.6. Reference-relative displacement RMS is annotated separately for each branch; the window-4/window-8 central-response RMS ratio is 11.23.

D.2 Feature responses across sampling windows

Tables 4 and 5 extend Table 1 to all eight sampling windows. For each feature, baseline variance is fixed across windows, so the temporal variation in $\widehat{B}_f$ comes from its mean response. Across features, the different baseline fluctuations determine how absolute responses translate into normalized responses.

Table 4: **Feature mean responses across sampling windows.** Signed mean responses $\widehat{\chi}_f$ for windows W1–W8, ordered from early to late sampling, with baseline standard deviations s_f. All use $\varepsilon = 0.25$, with $N = 512$ for JiT-B/L and 256 for ADM.

Model	Feature	s_f	W1	W2	W3	W4	W5	W6	W7	W8
JiT-B	DC	127	3.31	0.704	0.361	0.212	0.128	0.0786	0.0462	0.0187
	DCT-1	48.5	2.41	0.661	0.345	0.206	0.132	0.0785	0.0442	0.0131
	DCT-4	18.9	1.43	0.538	0.303	0.189	0.12	0.0768	0.0407	0.0164
	DCT-16	3.88	0.384	0.264	0.183	0.131	0.0908	0.061	0.0371	0.0149
JiT-L	DC	114	3.36	0.709	0.358	0.215	0.132	0.0783	0.0424	0.0165
	DCT-1	45.2	2.46	0.67	0.351	0.21	0.129	0.0821	0.0432	0.0168
	DCT-4	21.4	1.46	0.548	0.306	0.191	0.123	0.0773	0.0443	0.0157
	DCT-16	4.34	0.366	0.259	0.183	0.131	0.0925	0.062	0.0377	0.0151
ADM	DC	106	70.7	35.1	13.8	5.74	2.71	1.42	0.782	0.14
	DCT-1	42.5	24.9	22.4	11	5.07	2.53	1.36	0.764	0.138
	DCT-4	23.5	6.9	9.49	6.53	3.62	2.03	1.19	0.706	0.135
	DCT-16	4.32	0.274	0.662	1.04	1.11	0.912	0.686	0.501	0.118

Table 5: **Feature responses on the baseline fluctuation scale.** Normalized responses $\widehat{B}_f = \widehat{\chi}_f^2 / s_f^2$ for the measurements in Table 4.

Model	Feature	W1	W2	W3	W4	W5	W6	W7	W8
JiT-B	DC	0.000677	3.06e−5	8.05e−6	2.77e−6	1.01e−6	3.82e−7	1.32e−7	2.17e−8
	DCT-1	0.00248	0.000186	5.07e−5	1.81e−5	7.38e−6	2.62e−6	8.32e−7	7.26e−8
	DCT-4	0.00577	0.000813	0.000258	0.000101	4.08e−5	1.66e−5	4.66e−6	7.52e−7
	DCT-16	0.00982	0.00465	0.00222	0.00114	0.000548	0.000247	9.13e−5	1.48e−5
JiT-L	DC	0.000875	3.89e−5	9.92e−6	3.56e−6	1.35e−6	4.73e−7	1.39e−7	2.1e−8
	DCT-1	0.00296	0.00022	6.03e−5	2.16e−5	8.2e−6	3.3e−6	9.14e−7	1.39e−7
	DCT-4	0.00465	0.000653	0.000204	7.98e−5	3.3e−5	1.3e−5	4.28e−6	5.34e−7
	DCT-16	0.00712	0.00356	0.00177	0.000912	0.000454	0.000204	7.57e−5	1.21e−5
ADM	DC	0.447	0.11	0.017	0.00295	0.000655	0.00018	5.47e−5	1.75e−6
	DCT-1	0.343	0.277	0.0668	0.0143	0.00354	0.00103	0.000324	1.06e−5
	DCT-4	0.0865	0.164	0.0774	0.0239	0.00749	0.00255	0.000906	3.3e−5
	DCT-16	0.00402	0.0235	0.0585	0.0661	0.0446	0.0252	0.0134	0.00074