# OVADIP: CONTROLLING STRUCTURED DIFFUSION PATHS, SEPARATING TRAINING AND SAMPLING CONTRIBUTIONS, AND TRACKING THEIR DYNAMICS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

We introduce Operator-Valued Diffusion Paths (OvaDiP), a controlled analysis framework for structured diffusion paths whose final effects can vary across datasets. OvaDiP represents a structured path as an operator-valued deformation of a scalar reference and canonically separates its variation into common retiming and centered structural redistribution. Controlled factorial interventions along these two coordinates show that their effects depend on the sampling budget. We then use common-sampler cross-evaluation to separate the contributions associated with the trained model and the sampling path. Across AFHQv2 and FFHQ, small final Fréchet Inception Distance (FID) differences can conceal substantially larger opposing contributions from these two stages. Finally, we extend the same attribution across training and find that the factorial interaction can appear stable while the balance between training and sampling continues to evolve late in training. Together, these results show that a final generation metric alone can obscure how the observed effect of a structured diffusion path arises. OvaDiP provides a controlled framework for relating path interventions to training–sampling attribution and its evolution during training. Our implementation and evaluation code are available at `https://anonymous.4open.science/r/ovadip`.

## 1 INTRODUCTION

Diffusion models learn denoisers along prescribed noising paths and generate samples by reversing those paths numerically (Ho et al., 2020; Song et al., 2021b; Karras et al., 2022). Standard image diffusion typically uses a scalar noise level shared across the representation. Recent work has begun to relax this restriction through frequency-dependent, subspace-dependent, adaptive, and matrix-valued diffusion processes (Hoogeboom and Salimans, 2023; Hsueh et al., 2025; Jing et al., 2022; Sahoo et al., 2024; Scimeca et al., 2025; Liu et al., 2026). These developments make the diffusion path itself a richer design object: different components of a representation can follow different noise schedules.

This flexibility is promising but also raises an empirical puzzle. Structured or anisotropic diffusion paths have been shown to improve generation quality in several settings (Scimeca et al., 2025; Liu et al., 2026). Yet in our experiments, structured paths yield lower mean FID than their scalar references on several datasets but higher mean FID on another (Section 2.2). The effect relative to the scalar reference is therefore not uniform across datasets. What mechanisms underlie these different outcomes? We introduce Operator-Valued Diffusion Paths (OvaDiP), a controlled analysis framework for studying this question.

Figure 1 outlines this analysis. OvaDiP first isolates how the structured path itself changes. We represent a structured path as an operator-valued deformation of a scalar reference and decompose it into common retiming and centered structural redistribution (Sections 2–3). Controlled interventions along these two coordinates exhibit distinct behavior across finite sampling budgets. This gives a controlled way to distinguish common and relative path changes. Yet these controls still do not explain why the complete structured path improves FID on some datasets but degrades it on another.

To probe the source of this variation more deeply, we turn to two distinct stages of diffusion models: training and sampling (Song et al., 2021a; Lam et al., 2022; Sabour et al., 2024). We compare trained states under a common sampler and then return each trained state to its own sampling path



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: **Overview of OvaDiP.** (a) Structured paths yield different mean FID changes across datasets (Table 1); bars show means and error bars show sample standard deviations across training seeds. (b) Common retiming and centered structural redistribution define controlled R/G/C/F interventions. (c) Comparing trained states under the common Full sampler and then returning each state to its own path separates the factorial interaction into trained-state and path-return terms, $I = Q_{\mathrm{F}} + P_{\mathrm{F}}$. (d) Repeating this attribution across checkpoints tests whether its internal balance stabilizes. Curves in (b,d) are schematic; (d) satisfies $I(t) = Q(t) + P(t)$ and does not represent measured trajectories. Sample images in (a,c) are generated examples, not quantitative evidence; columns in (c) use matched generation seeds.

(Section 4), separating trained-state and sampling-path contributions to the observed effect. This attribution reveals that small final FID differences can conceal much larger opposing trained-state and sampling-path contributions. We then track the same attribution across training (Section 5) and find that the factorial interaction can appear stable while its internal attribution continues to evolve late in training.

**Contributions.** We introduce OvaDiP, a controlled analysis framework for structured diffusion paths, with three contributions. **Operator-valued path intervention.** An operator-valued representation and canonical decomposition turn structured path variation into controlled interventions for common retiming and centered structural redistribution. **Training–sampling attribution.** Common-sampler cross-evaluation separates trained-state and sampling-path contributions while preserving exact closure to the original contrast. **Attribution dynamics.** Extending the same attribution across training distinguishes stability of the factorial interaction from stability of its training–sampling attribution.

## 2 Constructing and Evaluating Structured Diffusion Paths

A standard continuous diffusion path can be written as

$$\boldsymbol{x}(u) = \alpha(u)\boldsymbol{x}_0 + \sigma(u)\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I}_d), \tag{1}$$

where $\boldsymbol{x}(u), \boldsymbol{x}_0, \boldsymbol{\epsilon} \in \mathbb{R}^d$, while the scalar signal coefficient $\alpha(u)$ and noise coefficient $\sigma(u)$ are shared across the representation.

### 2.1 Constructing an Operator-Valued Diffusion Path

Recent anisotropic diffusion formulations allow the corruption process to vary across spectral or subspace components, using structured covariance or matrix-valued noise schedules (Scimeca et al., 2025; Liu et al., 2026). This suggests a direct extension of the scalar path: retain a fixed decomposition of the representation, but allow the path coefficients to vary across its orthogonal subspaces. Let $\{\boldsymbol{P}_k\}_{k=1}^K$ denote the projectors onto $K$ orthogonal subspaces,

$$\boldsymbol{P}_k\boldsymbol{P}_\ell = \delta_{k\ell}\boldsymbol{P}_k, \qquad \sum_{k=1}^K \boldsymbol{P}_k = \boldsymbol{I}_d. \tag{2}$$

We lift the scalar coefficients to

$$\boldsymbol{A}(u) = \sum_{k=1}^K \alpha_k(u)\boldsymbol{P}_k, \qquad \boldsymbol{B}(u) = \sum_{k=1}^K \sigma_k(u)\boldsymbol{P}_k, \tag{3}$$

which define the operator-valued path

$$\boldsymbol{x}(u) = \boldsymbol{A}(u)\boldsymbol{x}_0 + \boldsymbol{B}(u)\boldsymbol{\epsilon} = \sum_{k=1}^K \left[\alpha_k(u)\boldsymbol{P}_k\boldsymbol{x}_0 + \sigma_k(u)\boldsymbol{P}_k\boldsymbol{\epsilon}\right]. \tag{4}$$

The scalar path is recovered when $\alpha_k(u) = \alpha(u)$ and $\sigma_k(u) = \sigma(u)$ for every $k$.

For the EDM formulation used in our experiments, $\alpha(u) = 1$. We therefore retain the shared signal coefficient and structure only the noise side,

$$\boldsymbol{x}(u) = \boldsymbol{x}_0 + \boldsymbol{B}(u)\boldsymbol{\epsilon}, \qquad \boldsymbol{B}(u) = \sum_{k=1}^K \sigma_k(u)\boldsymbol{P}_k. \tag{5}$$

OvaDiP parameterizes this operator relative to the scalar EDM schedule as

$$\boxed{\begin{aligned} &\boldsymbol{D}(u) = \sum_{k=1}^K D_k(u)\boldsymbol{P}_k, \qquad \log\sigma_k(u) = \log\sigma(u) - \lambda D_k(u), \\ &\boldsymbol{B}(u) = \sigma(u)\exp[-\lambda\boldsymbol{D}(u)]. \end{aligned}} \tag{6}$$

Here $D_k(u)$ specifies the log-noise deformation of subspace $k$, $\boldsymbol{D}(u)$ collects these deformations into a single operator, and $\lambda$ controls their amplitude. Setting $\boldsymbol{D}(u) = 0$ recovers the scalar EDM path. The resulting parameterization therefore places the scalar reference and its structured deformations in a common path representation. Appendix A gives the full derivation and subspace-wise form.

**Evaluation protocol.** We instantiate the operator-valued path in Table 1 with a three-band Discrete Cosine Transform (DCT) decomposition and two normalized Beta windows, $\mathcal{B}_{2,5.5}(u)$ and $\mathcal{B}_{3.5,4}(u)$. Further implementation details are given in Appendix B.1. CIFAR-10 and CIFAR-100 use 31 function evaluations (NFE31) and 50,000 samples per condition, with three and five training seeds, respectively; AFHQv2 and FFHQ use NFE63, 4096 samples per condition, and three seeds.

### 2.2 Structured-Path Effects Differ Across Datasets

Table 1: FID (↓) for scalar and structured paths. Values are mean ± sample SD across the training seeds specified above; the final column indicates whether the structured condition has lower mean FID within that dataset.

| Dataset | Scalar | Structured | Lower FID? |
|---|---|---|---|
| CIFAR-10 | $7.022 \pm 0.024$ | $6.492 \pm 0.042$ | ✓ |
| CIFAR-100 | $13.907 \pm 0.278$ | $12.938 \pm 0.086$ | ✓ |
| AFHQv2 | $9.985 \pm 0.319$ | $9.751 \pm 0.165$ | ✓ |
| FFHQ | $13.070 \pm 0.149$ | $13.216 \pm 0.522$ | ✗ |

Table 1 shows the resulting within-dataset Scalar-versus-Structured comparisons. The instantiated operator-valued paths achieve lower mean FID than their scalar comparators on CIFAR-10, CIFAR-100, and AFHQv2, but higher mean FID on FFHQ. Thus, the parameterization can realize structured paths that improve mean FID in several evaluated systems, while the sign of the structured-path effect is not fixed across the four datasets.

**Cross-projector and cross-domain checks.** We additionally test whether the formulation can be instantiated beyond the DCT image setting, using a rank-matched Haar projector and a controlled NAVSIM-mini trajectory-generation setting. The Haar and trajectory checks support the operator-valued parameterization as a broader construction, while the trajectory study remains only a bounded cross-domain instantiation check. Together with Table 1, these checks establish the two facts needed for the analysis that follows: the parameterization produces distinct structured-path outcomes, but those outcomes are not uniformly beneficial across the evaluated systems. Protocols and aggregate results are given in Appendix B.2.

To understand this variation in structured-path performance, we examine the key object in our parameterization, the deformation operator $\boldsymbol{D}(u)$. We address this by decomposing the deformation into controlled path coordinates in the next section.

# 3 Identifying How Structured Paths Differ

## 3.1 Decomposing Shared and Relative Path Changes

It is straightforward to see that, under the fixed projector decomposition in Eq. 2, a structured path can differ from the scalar EDM path in two ways: through a component shared by all subspaces and through relative differences among them. This motivates decomposing the deformation operator $\boldsymbol{D}(u)$ into a rank-weighted common component and a centered residual. We define $g(u) = d^{-1} \operatorname{Tr} \boldsymbol{D}(u) = \sum_k \pi_k D_k(u)$, with $\pi_k = \operatorname{rank}(\boldsymbol{P}_k)/d$, and $\boldsymbol{D}^\circ(u) = \boldsymbol{D}(u) - g(u)\boldsymbol{I}_d$, so that

$$\boxed{\boldsymbol{D}(u) = g(u)\boldsymbol{I}_d + \boldsymbol{D}^\circ(u)}, \qquad \operatorname{Tr} \boldsymbol{D}^\circ(u) = 0. \tag{7}$$

The common component $g(u)\boldsymbol{I}_d$ applies the same log-noise deformation to every subspace and corresponds to common retiming. The trace-free residual $\boldsymbol{D}^\circ(u)$ changes their relative log-noise allocation and corresponds to centered structural redistribution. Here, centered means $\operatorname{Tr} \boldsymbol{D}^\circ(u) = 0$: the relative subspace deformations have zero rank-weighted mean and therefore add no further common shift. A detailed derivation and the corresponding EDM scale–shape interpretation are given in Appendix A.2.

To test the two components separately, we keep the selected $g(u)$ and $\boldsymbol{D}^\circ(u)$ fixed and introduce two binary intervention coordinates,

$$\boldsymbol{D}_{a,b}(u) = a\, g(u)\boldsymbol{I}_d + b\, \boldsymbol{D}^\circ(u), \qquad (a, b) \in \{0, 1\}^2. \tag{8}$$

With $a, b \in \{0, 1\}$, Eq. 8 yields four factorial cells: Reference $\mathrm{R} = (0, 0)$, common retiming $\mathrm{G} = (1, 0)$, centered structural redistribution $\mathrm{C} = (0, 1)$, and Full OvaDiP $\mathrm{F} = (1, 1)$. Let $J_\mathrm{R}, J_\mathrm{G}, J_\mathrm{C}, J_\mathrm{F}$ denote their FID values. We use $J_\mathrm{G} - J_\mathrm{R}$, $J_\mathrm{C} - J_\mathrm{R}$, and $J_\mathrm{F} - J_\mathrm{R}$ as reference contrasts for common retiming, centered redistribution, and Full OvaDiP, respectively. Because lower FID is better, a negative reference contrast indicates an improvement over R. Their factorial interaction is

$$I = J_\mathrm{F} - J_\mathrm{G} - J_\mathrm{C} + J_\mathrm{R}. \tag{9}$$

If the effects of common retiming and centered redistribution combined additively, then

$$J_\mathrm{F} - J_\mathrm{R} = (J_\mathrm{G} - J_\mathrm{R}) + (J_\mathrm{C} - J_\mathrm{R}),$$

and hence $I = 0$. Thus, $I$ measures the part of their joint effect that is not explained by adding the two individual effects.

**Factorial evaluation protocol.** We evaluate the four cells on CIFAR-100 with five independently trained runs (seeds 11–15). All cells are derived from the same fixed Full deformation: G retains only the common-retiming component, C only the centered-redistribution component, and F both. G, C,

Table 2: CIFAR-100 strict-factorial FID (↓), mean ± sample SD over independently trained seeds 11–15. Bold marks the lowest mean at each NFE. Paired comparisons are reported in Table 3.

| Cell | NFE7 | NFE15 | NFE31 |
|---|---|---|---|
| R | $90.836 \pm 1.226$ | $15.427 \pm 0.170$ | $13.907 \pm 0.278$ |
| G | $\mathbf{86.811} \pm 1.197$ | $15.243 \pm 0.156$ | $13.764 \pm 0.157$ |
| C | $89.946 \pm 2.298$ | $14.935 \pm 0.233$ | $13.198 \pm 0.622$ |
| F | $87.827 \pm 1.793$ | $\mathbf{14.731} \pm 0.161$ | $\mathbf{12.938} \pm 0.086$ |

Table 3: Paired strict-factorial FID contrasts with two-sided 95% Student-$t$ intervals over seeds 11–15. Negative reference contrasts indicate lower FID. The interaction is $I = J_{\mathrm{F}} - J_{\mathrm{G}} - J_{\mathrm{C}} + J_{\mathrm{R}}$.

| NFE | $J_{\mathrm{G}} - J_{\mathrm{R}}$ | $J_{\mathrm{C}} - J_{\mathrm{R}}$ | $J_{\mathrm{F}} - J_{\mathrm{R}}$ | $I$ |
|---|---|---|---|---|
| 7 | $-4.025\ [-6.343, -1.707]$ | $-0.890\ [-3.379, +1.599]$ | $-3.009\ [-6.153, +0.134]$ | $+1.906\ [-0.008, +3.820]$ |
| 15 | $-0.184\ [-0.457, +0.089]$ | $-0.492\ [-0.723, -0.260]$ | $-0.696\ [-1.051, -0.342]$ | $-0.021\ [-0.327, +0.285]$ |
| 31 | $-0.143\ [-0.421, +0.135]$ | $-0.709\ [-1.231, -0.188]$ | $-0.969\ [-1.344, -0.594]$ | $-0.116\ [-0.681, +0.448]$ |

and F use the same $\lambda = 0.15$, with no post-split renormalization. Final checkpoints are evaluated at NFE7, NFE15, and NFE31 with 50,000 generated samples per condition. FID contrasts are paired by training seed, with two-sided 95% Student-$t$ intervals over the five runs. Appendix B gives the remaining construction and evaluation details.

### 3.2 Shared and Relative Path Changes Have Different Effects

Tables 2 and 3 show a budget-dependent separation between the two intervention coordinates. At NFE7, G has the lowest mean FID and common retiming is the only reference contrast resolved relative to zero. At NFE15 and NFE31, F has the lowest mean FID, while centered redistribution and Full OvaDiP are resolved relative to R and common retiming is not. The coordinate that is most clearly expressed therefore changes with the finite sampling budget: common retiming is clearest at NFE7, whereas centered redistribution is clearest at NFE15 and NFE31. The factorial interaction remains unresolved at all three NFEs.

At NFE15, the secondary metrics also move in different directions: relative to R, C has lower precision and density but higher recall and coverage (Appendix B.11).

**Centered-path projector check.** Because the Sec. 3.2 conclusion depends on the centered redistribution coordinate, we check whether its FID pattern is specific to the DCT basis. In a separate three-seed experiment with a rank-matched three-band Haar decomposition, Haar-Centered has lower FID than the scalar reference for every seed at NFE7, NFE15, and NFE31. Thus, the centered-path effect observed here is not unique to the DCT projector; full results are reported in Table 8.

Taken together, these results show that the canonical decomposition turns the structured-path deformation into controlled common and centered coordinates with distinct finite-budget behavior, and that the centered-coordinate pattern persists under the rank-matched Haar projector. These controlled coordinates isolate how the structured path itself changes, but they still do not explain why the complete structured path improves FID on some datasets and degrades it on another (Table 1).

The R/G/C/F comparisons above cannot answer this question by themselves. This exposes an important distinction: training and sampling are separable design dimensions in diffusion models (Song et al., 2021b; Karras et al., 2022). A structured path therefore enters both stages: it shapes the denoising model learned during training and the reverse-time trajectory used to generate samples. Consequently, the cross-dataset differences in Table 1 may arise from training, sampling, or their combination. We therefore separate the two effects next.

# 4 SEPARATING TRAINING AND SAMPLING EFFECTS

## 4.1 FORMALIZATION AND EVALUATION PROTOCOL

Each R/G/C/F setting determines both a trained model and the sampling path used to evaluate it. Let $\theta_X$ denote the model trained under setting $X$ and $s_Y$ the sampling path associated with setting $Y$. With all other evaluation conditions fixed, define the cross-evaluation and, for $A \in \{\mathrm{G}, \mathrm{C}, \mathrm{F}\}$, its reference contrast as

$$J_{X|Y} := \mathcal{E}(\theta_X, s_Y), \qquad \Delta J_{A-\mathrm{R}} := J_{A|A} - J_{\mathrm{R}|\mathrm{R}}. \tag{10}$$

Because both the trained model and the sampling path change in this contrast, $\Delta J_{A-\mathrm{R}}$ alone does not distinguish the two stages.

To separate them, choose a common sampling path $s_Y$. Evaluating both trained models under this common condition gives the trained-state contrast $T^{(Y)}_{A-\mathrm{R}}$, while returning each trained state to its own sampling path gives the path-return correction $S^{(Y)}_{A-\mathrm{R}}$:

$$T^{(Y)}_{A-\mathrm{R}} = J_{A|Y} - J_{\mathrm{R}|Y}, \qquad S^{(Y)}_{A-\mathrm{R}} = (J_{A|A} - J_{A|Y}) + (J_{\mathrm{R}|Y} - J_{\mathrm{R}|\mathrm{R}}). \tag{11}$$

These terms give the exact closure

$$\boxed{\Delta J_{A-\mathrm{R}} = T^{(Y)}_{A-\mathrm{R}} + S^{(Y)}_{A-\mathrm{R}}}. \tag{12}$$

For $A = \mathrm{F}$, choosing $Y = \mathrm{R}$ or $Y = \mathrm{F}$ corresponds to the two possible orders of changing the trained model and the sampler. As a robustness check against this ordering choice, we average the two decompositions:

$$\left(T^{\mathrm{sym}}_{\mathrm{F}-\mathrm{R}}, S^{\mathrm{sym}}_{\mathrm{F}-\mathrm{R}}\right) = \frac{1}{2}\left(T^{(\mathrm{R})}_{\mathrm{F}-\mathrm{R}} + T^{(\mathrm{F})}_{\mathrm{F}-\mathrm{R}}, S^{(\mathrm{R})}_{\mathrm{F}-\mathrm{R}} + S^{(\mathrm{F})}_{\mathrm{F}-\mathrm{R}}\right), \qquad \Delta J_{\mathrm{F}-\mathrm{R}} = T^{\mathrm{sym}}_{\mathrm{F}-\mathrm{R}} + S^{\mathrm{sym}}_{\mathrm{F}-\mathrm{R}}. \tag{13}$$

Both orders recover the same $\Delta J_{\mathrm{F}-\mathrm{R}}$; averaging them only changes how this difference is attributed between $T$ and $S$. Figure 2(a) visualizes this symmetric $T/S$ decomposition for the Full-versus-Reference contrast. We next apply the common-sampler decomposition to the factorial interaction $I$ defined in Eq. 9. Using the reference contrasts, define $I$ and its path-return correction as

$$I = \Delta J_{\mathrm{F}-\mathrm{R}} - \Delta J_{\mathrm{G}-\mathrm{R}} - \Delta J_{\mathrm{C}-\mathrm{R}}, \qquad P_Y = S^{(Y)}_{\mathrm{F}-\mathrm{R}} - S^{(Y)}_{\mathrm{G}-\mathrm{R}} - S^{(Y)}_{\mathrm{C}-\mathrm{R}}. \tag{14}$$

The corresponding common-sampler trained-state interaction is

$$\boxed{Q_Y = T^{(Y)}_{\mathrm{F}-\mathrm{R}} - T^{(Y)}_{\mathrm{G}-\mathrm{R}} - T^{(Y)}_{\mathrm{C}-\mathrm{R}} = J_{\mathrm{F}|Y} - J_{\mathrm{G}|Y} - J_{\mathrm{C}|Y} + J_{\mathrm{R}|Y}}. \tag{15}$$

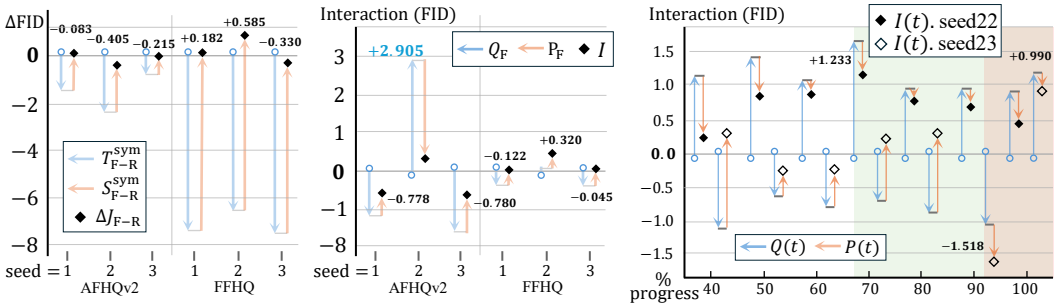
These quantities satisfy the exact closure

$$\boxed{I = Q_Y + P_Y}. \tag{16}$$

Here, $Q_Y$ is the R/G/C/F factorial interaction across trained states measured under the common sampler $Y$, while $P_Y$ is the corresponding factorial interaction of the path-return corrections. Thus, $Q_Y$ is the trained-state interaction term and $P_Y$ the additional sampling-path interaction term needed to recover the factorial interaction $I$.

The reference contrast $\Delta J_{A-\mathrm{R}}$ and factorial interaction $I$ are fixed, whereas their decomposition into $T^{(Y)}$, $S^{(Y)}$, $Q_Y$, and $P_Y$ depends on the chosen common sampler $Y$. These attribution terms are therefore common-sampler-conditioned rather than intrinsic training or sampling effects.

**Evaluation protocol.** We reuse the fixed R/G/C/F trained states for AFHQv2 and FFHQ, with three training seeds per dataset, and evaluate every trained state under every R/G/C/F sampling path, giving the complete $4 \times 4$ grid $J_{X|Y}$. Each cell uses structured Heun sampling at NFE63 with 4096 generated samples. For the Full-versus-Reference contrast, we report the order-symmetric $T/S$ decomposition in Eq. 13. For the factorial interaction, the main comparison uses the Full sampler, $Y = \mathrm{F}$, while dependence on the common sampler is examined across all four choices below. Appendix B gives the remaining evaluation and path-construction details, and Appendix C.1 reports the complete cross-evaluation results.




(a) Full vs. Reference (b) Factorial interaction (c) Attribution dynamics. CIFAR-100

Figure 2: Training–sampling attribution and its evolution. (a) Order-symmetric Full-versus-Reference attribution at NFE63 across AFHQv2 and FFHQ training runs. (b) Factorial interaction under the Full common sampler at NFE63 for the same runs, with $I = Q_{\mathrm{F}} + P_{\mathrm{F}}$. (c) Block-A CIFAR-100 attribution trajectories at NFE31 for fresh seeds 22 and 23, with $I(t) = Q(t) + P(t)$. The green shaded region marks the 70–100% interval over which the sign of $I(t)$ for seed 22 remains stable, while the red shaded region marks the 90–100% interval in which both $Q(t)$ and $I(t)$ for seed 23 reverse sign.

### 4.2 Training–Sampling Attribution

Figure 2(a)–(b) shows the two training–sampling attribution views across the six AFHQv2 and FFHQ runs. For the Full-versus-Reference contrast, the order-symmetric trained-state and path-return terms oppose each other in every run, leaving a much smaller final FID difference [Fig. 2(a)]. The magnitude of these opposing terms varies substantially across the two datasets, but the cancellation pattern is shared across all six runs. For the factorial interaction, $Q_{\mathrm{F}}$ has the same sign as $I$ in all six runs and satisfies $|Q_{\mathrm{F}}| > |I|$ in five of them [Fig. 2(b)]. The sampling-path term opposes $Q_{\mathrm{F}}$ in five runs, while FFHQ seed 2 is the exception, where $Q_{\mathrm{F}}$ and $P_{\mathrm{F}}$ have the same sign.

Taken together, Figure 2(a)–(b) shows that the final observed effect does not by itself characterize its training–sampling attribution: small values of $\Delta J$ or $I$ can coexist with substantially larger contributions, and those contributions may either oppose or reinforce one another. The common-sampler construction makes this internal composition observable while preserving exact closure to the original contrast, separating the measured effect from how training and sampling contribute to it. Because $Q_Y$ and $P_Y$ are conditioned on the chosen common sampler $Y$, we next examine how this attribution changes across the complete $4 \times 4$ cross-evaluation grids.

Across the complete $4 \times 4$ grids, $Q_{\mathrm{R}}$ and $Q_{\mathrm{G}}$ remain close, as do $Q_{\mathrm{C}}$ and $Q_{\mathrm{F}}$; the larger changes occur between the R/G and C/F sampler groups. This grouping appears in both datasets. However, it is not universal: AFHQv2 seed 3 reverses sign between the two groups. We therefore fix $Y = \mathrm{F}$ for the following analysis and report the complete coordinates in Appendix C.1.

The attribution above raises a further question: when does the training–sampling attribution become stable during training? Answering this question requires aligned intermediate checkpoints. We first test this question in the simpler CIFAR-100 strict-factorial system, where aligned intermediate checkpoints are available throughout training, and examine whether the attribution stabilizes early or continues to change late.

## 5 How the Training–Sampling Balance Evolves

### 5.1 Measuring Attribution Across Training

To test whether the final attribution stabilizes before training ends, we track the same quantities over training as $I(t)$, $Q(t)$, and $P(t)$. At each checkpoint, $Q(t)$ is measured under the fixed Full sampler and $P(t)$ is obtained by returning the R/G/C/F states to their respective sampling paths, so

$$I(t) = Q(t) + P(t). \tag{17}$$

Table 4: Training-time evidence on CIFAR-100. Panel A reports early stabilization in fresh seeds 22 and 23. Panel B reports the frozen late-change test on seed 24 over the selected 80–90% interval.

**A. Early stabilization in fresh runs**

Trajectories are visualized in Figure 2(c); exact checkpoint values are reported in Appendix Table 20.

| Seed | $I$ sign stable before end? | $Q/P$ relation stable before end? | Late sign reversal |
|---|---|---|---|
| 22 | ✓ (from 70%) | ✗ | None |
| 23 | ✗ | ✗ | $Q$ and $I$ (90–100%) |

**B. Late-change test on seed 24**

| Test | Measured change | Block variation | Margin |
|---|---|---|---|
| Training contribution $Q$ | $\overline{\Delta Q} = +1.9022$ | $R^{\mathrm{MC}}_{\Delta Q} = 0.3629$ | $M_Q = +1.5393$ |
| Sampling-path contribution $P$ | $\overline{\Delta P} = -0.4605$ | $R^{\mathrm{MC}}_{\Delta P} = 0.5112$ | $-0.0507$ |
| Combined frozen margin | 2.3627 | 0.8741 | $M_{\mathrm{joint}} = +1.4886$ |

We assess stability using the frozen future-variation margins in Appendix D. A persistent positive margin certifies the interaction sign or the qualitative $Q/P$ relation at all later measured checkpoints; observing an unchanged sign alone is not sufficient for this certificate.

However, an apparent late change in $Q(t)$ or $P(t)$ may still fall within the variation caused by finite-sample evaluation. We therefore measure the same late interval on three disjoint generation blocks. For $X \in \{Q, P\}$, let $\Delta X^{(b)}$ denote the change measured on block $b \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$:

$$\overline{\Delta X} = \frac{1}{3} \sum_{b \in \{\mathrm{A},\mathrm{B},\mathrm{C}\}} \Delta X^{(b)}, \qquad R^{\mathrm{MC}}_{\Delta X} = \max_b \left| \Delta X^{(b)} - \overline{\Delta X} \right|. \tag{18}$$

Here, $X$ denotes either attribution term, $\overline{\Delta X}$ is its block-averaged late change, and $R^{\mathrm{MC}}_{\Delta X}$ is the maximum block deviation from that mean, used as an empirical Monte Carlo variation scale. Intuitively, a late change is distinguishable from block variation when $|\overline{\Delta X}| > R^{\mathrm{MC}}_{\Delta X}$. For the trained-state term and the combined attribution, we write this comparison as

$$M_Q = |\overline{\Delta Q}| - R^{\mathrm{MC}}_{\Delta Q}, \qquad M_{\mathrm{joint}} = |\overline{\Delta Q}| + |\overline{\Delta P}| - R^{\mathrm{MC}}_{\Delta Q} - R^{\mathrm{MC}}_{\Delta P}. \tag{19}$$

Thus, $M_Q > 0$ means that the late change in $Q$ exceeds its block variation, while $M_{\mathrm{joint}} > 0$ summarizes aggregate change across $Q$ and $P$ relative to their combined block variation. Individual changes are assessed by comparing each term with its own block variation.

**Evaluation protocol.** We evaluate aligned CIFAR-100 checkpoints with structured Heun sampling at NFE31. Fresh seeds 22 and 23 test early stability; seed 24 separately tests late change with three disjoint 4096-sample blocks. Block A selects between the pre-specified 80–90% and 90–100% intervals, while blocks B and C cannot alter that choice. Appendix D gives the full protocol.

### 5.2 The Attribution Can Continue to Change Late

Figure 2(c) traces the Block-A training–sampling attribution trajectories for fresh CIFAR-100 seeds 22 and 23. For seed 22, the displayed interaction remains positive, and the formal criterion places its interaction-sign lock at 70% of training, while the $Q/P$ relation does not stabilize before the final checkpoint. Seed 23 shows no such early stabilization: $Q(t)$ remains negative through 90% and then reverses sign at 100%, while $I(t)$ simultaneously changes from negative to positive over the same 90–100% interval. Table 4 summarizes these fresh-seed stability outcomes together with the independent seed-24 late-change test.

The fresh-seed trajectories distinguish stability of the observed factorial interaction from stability of its internal attribution. Seed 22 provides the direct case: the interaction sign locks before the $Q/P$ relation, whereas seed 23 shows that neither need stabilize early and that both $Q$ and $I$ can still reverse late in training. The independent seed-24 test strengthens this trajectory-level evidence: the late displacement in $Q$ exceeds its three-block variation, whereas the corresponding displacement in $P$ does not, making the positive joint margin $Q$-driven. Together, these results show that a seemingly settled factorial interaction need not imply a settled training–sampling attribution, and that the trained-state contribution can continue to change late in training.

## 6 Conclusion

A final generation metric need not reveal how the observed effect of a structured diffusion path arises. Common retiming and centered structural redistribution behave differently across finite sampling budgets; small final FID differences can result from cancellation between larger trained-state and sampling-path terms; and the training–sampling attribution can continue to evolve even when the factorial interaction appears stable. These findings distinguish three questions that should be treated separately: how the path is changed, how training and sampling contribute to the observed effect, and whether that attribution has stabilized during training. The evidence is confined to the finite systems studied here, so these patterns should not be interpreted as universal properties of structured diffusion paths. Whether these attribution structures and training-time behaviors persist across broader model families, path constructions, domains, and sampling conditions remains open.

## Reproducibility Statement

The reproduction package is available at `https://anonymous.4open.science/r/ovadip` and includes the OvaDiP implementation, the R/G/C/F path constructions, training and evaluation scripts, and the analyses used for the reported tables. Appendix B documents the datasets, training seeds, sampling settings, sample counts, and evaluation protocols, while the remaining appendices provide the derivations and additional analyses supporting the reported results.

## AI Use Statement

The research ideas, scientific claims, theoretical derivations, experimental design, and interpretation of results were developed by the authors. The authors determined the scientific content and argumentative structure of the manuscript, including the intended meaning and logical role of each paragraph. Generative AI assisted with sentence formulation and language polishing under the authors' direction. The authors reviewed every sentence for scientific accuracy, logical consistency, and fidelity to their intended claims. AI tools also assisted with discussion and feedback, software implementation support, literature organization, and figure preparation. AI-assisted code and technical suggestions were checked by the authors against the corresponding experiments, source results, implementations, and cited literature. The authors made all scientific decisions and take full responsibility for the final claims, analyses, code, and manuscript.

## References


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

Emiel Hoogeboom and Tim Salimans. Blurring diffusion models. In *International Conference on Learning Representations*, 2023.

Hao-Chien Hsueh, Wen-Hsiao Peng, and Ching-Chun Huang. Warm diffusion: Recipe for blur-noise mixture diffusion models. In *International Conference on Learning Representations*, 2025.

Bowen Jing, Gabriele Corso, Renato Berlinghieri, and Tommi Jaakkola. Subspace diffusion generative models. In *European Conference on Computer Vision*, 2022.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Max W. Y. Lam, Jun Wang, Dan Su, and Dong Yu. BDDM: Bilateral denoising diffusion models for fast and high-quality speech synthesis. In *International Conference on Learning Representations*, 2022.

Pengxi Liu, Zeyu Michael Li, and Xiang Cheng. Variational trajectory optimization of anisotropic diffusion schedules, 2026.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your steps: Optimizing sampling schedules in diffusion models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, 2024.

Subham Sekhar Sahoo, Aaron Gokaslan, Chris De Sa, and Volodymyr Kuleshov. Diffusion models with learned adaptive noise. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Luca Scimeca, Thomas Jiralerspong, Berton Earnshaw, Jason Hartford, and Yoshua Bengio. Learning what matters: Steering diffusion via spectrally anisotropic forward noise, 2025.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

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

# A Extended Operator-Path Derivations and Chart Mappings

## A.1 From Scalar Generative Paths to Operator-Valued Diffusion Paths

The main-text construction is a specialization of a general linear generative path

$$\boldsymbol{x}_u = a(u)\boldsymbol{x}_0 + b(u)\boldsymbol{z}, \qquad \boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I}_d). \tag{20}$$

Here $a(u)$ and $b(u)$ are path functions; they are unrelated to the binary analysis coordinates introduced later in Section 3.1. For mutually orthogonal projectors satisfying Eq. 2, lift the scalar coefficients to

$$\boldsymbol{A}(u) = \sum_{k=1}^{K} a_k(u)\boldsymbol{P}_k, \qquad \boldsymbol{B}(u) = \sum_{k=1}^{K} b_k(u)\boldsymbol{P}_k. \tag{21}$$

The resulting operator-valued path is

$$\boldsymbol{x}_u = \boldsymbol{A}(u)\boldsymbol{x}_0 + \boldsymbol{B}(u)\boldsymbol{z} = \sum_{k=1}^{K} \left[a_k(u)\boldsymbol{P}_k\boldsymbol{x}_0 + b_k(u)\boldsymbol{P}_k\boldsymbol{z}\right]. \tag{22}$$

Setting $a_k = a$ and $b_k = b$ for every $k$ recovers Eq. 20 because $\sum_k \boldsymbol{P}_k = \boldsymbol{I}_d$. The projectors define the subspaces, while distinct $a_k$ and $b_k$ assign their path coefficients. Equation 22 therefore describes one operator-valued path with subspace-dependent evolution.

Because the projectors are symmetric and mutually orthogonal,

$$\boldsymbol{B}(u)\boldsymbol{B}(u)^\top = \sum_{k=1}^{K} b_k^2(u)\boldsymbol{P}_k. \tag{23}$$

The subspace sum is therefore the spectral expansion of one projector-diagonal noise operator and its covariance, not a mixture of independently specified stochastic processes.

When $a_k(u), b_k(u) > 0$, define the log signal-to-noise coordinate

$$\rho_k(u) = \log \frac{a_k(u)}{b_k(u)}, \qquad \rho(u) = \log \frac{a(u)}{b(u)}. \tag{24}$$

The OvaDiP deformation is introduced in this coordinate by

$$\boldsymbol{R}(u) := \sum_k \rho_k(u)\boldsymbol{P}_k = \rho(u)\boldsymbol{I}_d + \lambda \boldsymbol{D}(u), \qquad \boldsymbol{D}(u) = \sum_k D_k(u)\boldsymbol{P}_k. \tag{25}$$

Since the projectors are orthogonal, Eq. 25 is equivalent componentwise to

$$\rho_k(u) = \rho(u) + \lambda D_k(u). \tag{26}$$

For EDM, $a(u) = a_k(u) = 1$, $b(u) = \sigma(u)$, and $b_k(u) = \sigma_k(u)$. Hence

$$\rho(u) = -\log \sigma(u), \qquad \rho_k(u) = -\log \sigma_k(u), \tag{27}$$

and Eq. 26 gives

$$\boxed{\log \sigma_k(u) = \log \sigma(u) - \lambda D_k(u)}, \qquad \sigma_k(u) = \sigma(u) \exp[-\lambda D_k(u)]. \tag{28}$$

Functional calculus is exact in the fixed projector basis:

$$\exp[-\lambda \boldsymbol{D}(u)] = \sum_k \exp[-\lambda D_k(u)] \boldsymbol{P}_k. \tag{29}$$

Therefore the EDM noise-scale operator in Eq. 5 is

$$\boxed{\boldsymbol{B}(u) = \sum_k \sigma_k(u) \boldsymbol{P}_k = \sigma(u) \exp[-\lambda \boldsymbol{D}(u)]}. \tag{30}$$

Its conditional noise covariance is

$$\boldsymbol{\Sigma}(u) := \boldsymbol{B}(u) \boldsymbol{B}(u)^\top = \sigma^2(u) \exp[-2\lambda \boldsymbol{D}(u)]. \tag{31}$$

Here $\boldsymbol{B}(u)$ is the EDM noise-scale operator, while $\boldsymbol{\Sigma}(u)$ denotes its conditional covariance.

### A.2 Canonical Coordinates in the Projector Algebra

All operator paths above remain in the fixed commuting algebra

$$\mathcal{A}_{\mathcal{P}} := \operatorname{span}\{\boldsymbol{P}_1, \ldots, \boldsymbol{P}_K\}. \tag{32}$$

Since $\boldsymbol{I}_d = \sum_k \boldsymbol{P}_k$, define

$$\mathcal{T}_{\mathcal{P}} := \mathcal{A}_{\mathcal{P}} \cap \ker(\mathrm{Tr}), \qquad \mathcal{A}_{\mathcal{P}} = \operatorname{span}\{\boldsymbol{I}_d\} \oplus \mathcal{T}_{\mathcal{P}}. \tag{33}$$

For $\boldsymbol{D} = \sum_k D_k \boldsymbol{P}_k$, with $\pi_k = \operatorname{rank}(\boldsymbol{P}_k)/d$, its scalar projection and centered complement are

$$g = \frac{1}{d} \operatorname{Tr} \boldsymbol{D} = \sum_k \pi_k D_k, \qquad \boldsymbol{D}^\circ = \boldsymbol{D} - g \boldsymbol{I}_d \in \mathcal{T}_{\mathcal{P}}. \tag{34}$$

Thus $\boldsymbol{D} = g\boldsymbol{I}_d + \boldsymbol{D}^\circ$ is the direct-sum coordinate decomposition within $\mathcal{A}_{\mathcal{P}}$. The two components are orthogonal under the normalized Frobenius inner product because $d^{-1} \operatorname{Tr}[(g\boldsymbol{I}_d)\boldsymbol{D}^\circ] = 0$. This statement is relative to the fixed projector algebra, not a uniqueness claim over arbitrary noncommuting parameterizations.

Substituting the decomposition into the EDM operator gives

$$\boldsymbol{B} = \sigma \exp(-\lambda g) \exp(-\lambda \boldsymbol{D}^\circ), \tag{35}$$

$$\boldsymbol{\Sigma} = \sigma^2 \exp(-2\lambda g) \exp(-2\lambda \boldsymbol{D}^\circ). \tag{36}$$

The first factor changes the common scalar noise scale. The centered factor changes relative covariance shape, and

$$\det \exp(-2\lambda \boldsymbol{D}^\circ) = \exp[-2\lambda \operatorname{Tr}(\boldsymbol{D}^\circ)] = 1. \tag{37}$$

Determinant preservation is specific to this centered log-noise/EDM chart and is not asserted for arbitrary path charts.

For the strict factorial, Scalar and Centered are obtained by applying the two projections in Eq. 34 to one fixed Full field. The implementation retains the same $\lambda$ and performs no post-split renormalization, so

$$\boldsymbol{D}_{\mathrm{F}}(u) = \boldsymbol{D}_{\mathrm{G}}(u) + \boldsymbol{D}_{\mathrm{C}}(u) = g(u)\boldsymbol{I}_d + \boldsymbol{D}^\circ(u) \tag{38}$$

pointwise in $u$. By contrast, the broader semantic Full construction first chooses a centered field and then sets $g(u) = \|\boldsymbol{D}^\circ(u)\|_\pi$. This scalar-lift rule is a design choice, not a theorem implied by centering. Exact addition in path space does not imply additive generation-quality response; Section 4 measures that response after finite training and evaluation.

# B Complete Experimental Protocols and Metrics

## B.1 Section 2 Structured-Path Instantiations

The Section 2.2 comparisons use three-band orthogonal DCT projectors. For cutoffs $c_1 < c_2$ in an $H \times H$ image, the low band contains the DCT coefficients in $[0, c_1)^2$, the middle band contains $[0, c_2)^2 \setminus [0, c_1)^2$, and the high band is the complement. Applying the corresponding masks in an orthonormal DCT-II basis defines $\boldsymbol{P}_{\mathrm{L}}, \boldsymbol{P}_{\mathrm{M}}, \boldsymbol{P}_{\mathrm{H}}$. Their rank fractions

$$\pi_k = \frac{\operatorname{rank}(\boldsymbol{P}_k)}{d}, \qquad k \in \{\mathrm{L}, \mathrm{M}, \mathrm{H}\}, \tag{39}$$

set the rank-weighted norm

$$\|\boldsymbol{v}\|_\pi^2 = \sum_k \pi_k v_k^2. \tag{40}$$

The stage-family paths use the log-interval progress coordinate. For positive scalar EDM noise level $\sigma$,

$$u(\sigma) = \frac{\log \sigma_{\max} - \log \widetilde{\sigma}}{\log \sigma_{\max} - \log \sigma_{\min}}, \qquad \widetilde{\sigma} = \operatorname{clip}(\sigma, \sigma_{\min}, \sigma_{\max}), \tag{41}$$

with $\sigma_{\min} = 0.002$ and $\sigma_{\max} = 80$. The two temporal basis functions used by the Section 2.2 structured paths are normalized Beta windows

$$\mathcal{B}_{p,q}(u) = \frac{u^p (1-u)^q}{\left(\frac{p}{p+q}\right)^p \left(\frac{q}{p+q}\right)^q}, \tag{42}$$

specifically $\mathcal{B}_{2,5.5}$ and $\mathcal{B}_{3.5,4}$. For protocol-specific early and middle directions $\boldsymbol{v}_{\mathrm{E}}, \boldsymbol{v}_{\mathrm{M}} \in \mathbb{R}^3$, the implemented field has the form

$$\boldsymbol{d}(u) = \frac{\mathcal{B}_{2,5.5}(u)\boldsymbol{v}_{\mathrm{E}} + \mathcal{B}_{3.5,4}(u)\boldsymbol{v}_{\mathrm{M}}}{M}, \qquad M = \max_{u \in [0,1]} \left\| \mathcal{B}_{2,5.5}(u)\boldsymbol{v}_{\mathrm{E}} + \mathcal{B}_{3.5,4}(u)\boldsymbol{v}_{\mathrm{M}} \right\|_\pi, \tag{43}$$

where the implementation evaluates $M$ on a fixed dense grid. The operator deformation is $\boldsymbol{D}(u) = \sum_k d_k(u)\boldsymbol{P}_k$ before any protocol-specific scalar lift.

For the semantic centered field, let $r$ denote the stage ratio and define

$$\begin{aligned} \widetilde{\boldsymbol{v}}_{\mathrm{E}} &= \left(1,\, r,\, -\frac{\pi_{\mathrm{L}} + r\pi_{\mathrm{M}}}{\pi_{\mathrm{H}}}\right), \\ \widetilde{\boldsymbol{v}}_{\mathrm{M}} &= \left(r,\, 1,\, -\frac{r\pi_{\mathrm{L}} + \pi_{\mathrm{M}}}{\pi_{\mathrm{H}}}\right). \end{aligned} \tag{44}$$

Each direction is normalized by $\|\cdot\|_\pi$ before use in Eq. 43. Hence the resulting $\boldsymbol{d}_{\mathrm{sem}}(u)$ is rank-centered pointwise. For this image-path instantiation, $r = 0$, and the corresponding Full field is

$$\boldsymbol{d}_{\mathrm{F}}(u) = \boldsymbol{d}_{\mathrm{sem}}(u) + \|\boldsymbol{d}_{\mathrm{sem}}(u)\|_\pi \, \mathbf{1}, \tag{45}$$

with no renormalization after addition.

In Section 2.2, *Scalar* is used as the common manuscript label for the isotropic comparator associated with each dataset. The later R/G/C/F analysis introduces Reference (R) as the zero-deformation coordinate used for the controlled factorial construction.

Table 5 records the remaining protocol identity for the four rows in Table 1. The main text reports only the within-dataset Scalar-versus-Structured comparison; the cutoff, amplitude, checkpoint, and generation-block details are kept here. Because the rows use dataset-specific sampling budgets and, for CIFAR-10, a different deformation amplitude, cross-row magnitudes are not interpreted as directly commensurate effect sizes.

Figure 3 provides an intuitive view of the Scalar and Full paths through their low-, middle-, and high-frequency components.

Table 5: Protocol identity for the four Scalar-versus-Structured comparisons in Table 1. Sample counts are per condition and per training seed.

| Dataset | Comparison | Structured field | DCT / $\lambda$ | Train seeds | Solver / samples | Final checkpoint and generation seeds |
|---|---|---|---|---|---|---|
| CIFAR-10 | Scalar vs Full | Full structured path | 8/16; 0.067 | 1–3 | Heun, NFE31; 50,000 | 10 Mimg; generation seeds 0–49,999 |
| CIFAR-100 | Scalar vs Full | Full structured path | 8/16; 0.15 | 11–15 | Heun, NFE31; 50,000 | 10 Mimg; generation seeds 0–49,999 |
| AFHQv2 | Scalar vs Full | Full structured path | 16/32; 0.15 | 1–3 | Heun, NFE63; 4096 | 5 Mimg; generation seeds 0–4095 |
| FFHQ | Scalar vs Full | Full structured path | 16/32; 0.15 | 1–3 | Heun, NFE63; 4096 | 5 Mimg; generation seeds 0–4095 |

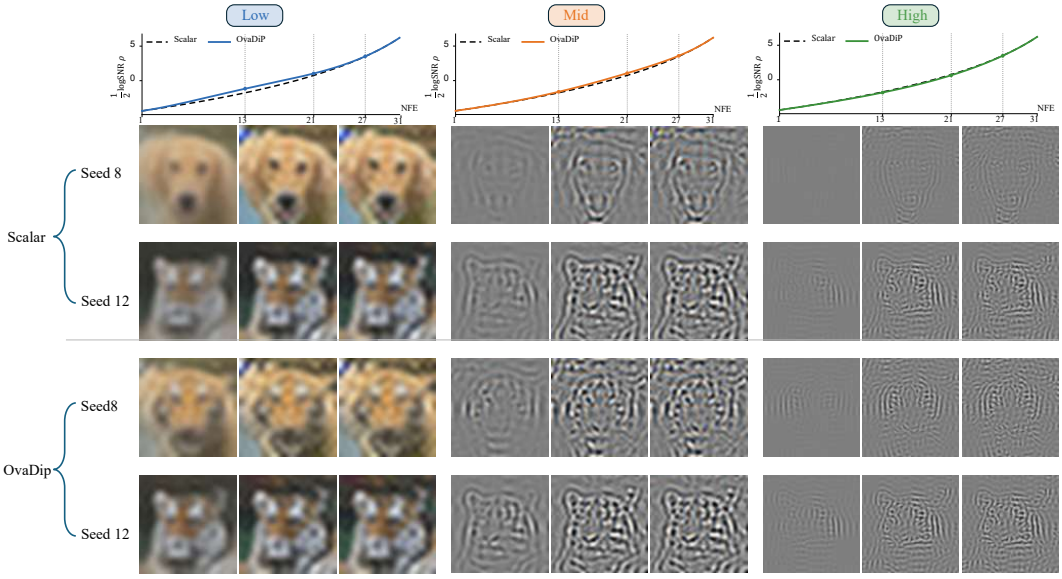



Figure 3: **An intuitive view of structured-path construction.** Top: scalar-reference and Full OvaDiP schedules in the low-, middle-, and high-frequency subspaces. Below: the corresponding frequency components at three sampling positions for the two displayed seeds, with Scalar above and Full OvaDiP below. This qualitative visualization illustrates how the path construction changes the sampling process across frequency subspaces.

### B.2 Cross-Projector and Cross-Domain Extensibility Checks

The cross-projector check uses a separate three-seed CIFAR-100 experiment with a rank-matched three-band Haar wavelet decomposition. The scalar reference uses $\lambda = 0$, while the independently trained DCT-Centered and Haar-Centered paths use $\lambda = 0.15$. All three are evaluated with 50,000 generated samples at NFE7, NFE15, and NFE31. The resulting FIDs are reported in Table 8; both centered projector families have lower FID than the scalar reference for every seed at all three NFEs.

For the cross-domain check, we use the controlled NAVSIM-mini trajectory-generation study. The finite system contains 58 multimodal scenes, 128 labeled feasible trajectories, and five log-disjoint folds. A frozen DiffusionDrive scene encoder feeds an anchor-free conditional diffusion trajectory head that generates 10 poses over 5 seconds, with $(x, y, \theta)$ at each pose. The trajectory study uses a DDPM/DDIM epsilon-prediction chart rather than the EDM chart used in the image experiments, so it tests the general operator-valued formulation rather than the EDM-specific instantiation.

The scalar reference applies one path to all 30 trajectory dimensions. A temporal DCT instance decomposes normalized trajectories into low-, middle-, and high-temporal-frequency subspaces. The

Table 6: Recorded valid-trajectory rate in the controlled NAVSIM-mini comparison.

| NFE | Scalar | Intent-v2 | Difference |
|---|---|---|---|
| 4 | 0.090 | 0.138 | +0.048 |
| 8 | 0.121 | 0.165 | +0.044 |
| 16 | 0.129 | 0.170 | +0.041 |
| 50 | 0.138 | 0.179 | +0.041 |

task-induced Intent-v2 instance is constructed from scene-balanced branch-pair trajectory contrasts

$$\boldsymbol{q}_{s,ab} = \frac{1}{\sqrt{|\mathcal{P}_s|}} \frac{\widetilde{\boldsymbol{\tau}}_{s,a} - \widetilde{\boldsymbol{\tau}}_{s,b}}{\|\widetilde{\boldsymbol{\tau}}_{s,a} - \widetilde{\boldsymbol{\tau}}_{s,b}\|_2}, \tag{46}$$

where only the training partition of each fold is used. Stacking these contrasts and applying SVD gives the rank-2 intent projector

$$\boldsymbol{P}_{\mathrm{I}} = \boldsymbol{U}_{:,:2}\boldsymbol{U}_{:,:2}^{\top}, \qquad \boldsymbol{P}_{\mathrm{R}} = \boldsymbol{I} - \boldsymbol{P}_{\mathrm{I}}. \tag{47}$$

Validation selects $\lambda = -0.05$ for Intent-v2. Evaluation uses NFE $\{4, 8, 16, 50\}$, paired sampling seeds $\{0, 1, 2\}$, and 128 generated trajectories per scene.

Intent-v2 has a higher recorded valid-trajectory rate than the scalar path at all four evaluated NFEs. We use this result only to demonstrate a controlled operator-valued instantiation outside image-frequency space; it does not establish uniform trajectory-quality improvement or full NAVSIM/PDMS driving performance. Only aggregate measurements are available for this extension; per-fold values are not available, so we do not attach fold-level uncertainty claims to this comparison.

### B.3 Finite-system and evidence-block map

Table 7 keeps the scientifically distinct evidence blocks separate. The seven-condition design denotes the four same-setting evaluations $J_{X|X}$ plus the three additional common-Full-evaluator cross-evaluations used by the checkpoint analyses. The AFHQv2 and FFHQ common-sampler sensitivity blocks additionally evaluate all 16 trained-state/sampler pairs. A sample count is per condition and per training seed unless a row states otherwise.

The strict-factorial path construction uses stage ratio 0 and the final 10-Mimg checkpoints with batch size 640. Scalar and Centered are the exact common and centered projections of the recorded Full field, with no post-split renormalization. The AFHQv2 and FFHQ common-sampler attribution blocks instead use the broader quality-path construction defined above: a centered field is chosen first and $g(u) = \|\boldsymbol{D}^{\circ}(u)\|_{\pi}$ is added as a design rule. These settings identify finite systems; they do not make results from different rows exchangeable.

### B.4 Four trained states and common-evaluator grids

For each factorial training realization, R, G, C, and F are trained independently, producing $\theta_{\mathrm{R}}, \theta_{\mathrm{G}}, \theta_{\mathrm{C}}, \theta_{\mathrm{F}}$. The same-setting evaluation set is

$$\{J_{\mathrm{R|R}}, J_{\mathrm{G|G}}, J_{\mathrm{C|C}}, J_{\mathrm{F|F}}\}, \tag{48}$$

and the common-Full-evaluator augmentation used in the checkpoint analyses is

$$\{J_{\mathrm{R|F}}, J_{\mathrm{G|F}}, J_{\mathrm{C|F}}\}. \tag{49}$$

Together with $J_{\mathrm{F|F}}$, these form the seven unique conditions needed for $Q_{\mathrm{F}}$, $P_{\mathrm{F}}$, and $I$. The AFHQv2 and FFHQ common-sampler sensitivity experiment extends this subset to the complete grid $\{J_{X|Y} : X, Y \in \{\mathrm{R}, \mathrm{G}, \mathrm{C}, \mathrm{F}\}\}$, yielding 16 conditions per training seed and direct measurements of $Q_{\mathrm{R}}, Q_{\mathrm{G}}, Q_{\mathrm{C}}, Q_{\mathrm{F}}$. No unmeasured numerical partition is fitted.

### B.5 Structured cross-evaluation implementation

All R/G/C/F checkpoints use the same structure-enabled EDM preconditioner, so an off-diagonal evaluation $J_{X|Y}$ can pass the sampler-$Y$ blockwise noise vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ directly to a

model trained under cell $X$. For the `structure_condition=none` configuration used here, the preconditioner applies

$$c_{\mathrm{in},k} = (\sigma_{\mathrm{data}}^2 + \sigma_k^2)^{-1/2}, \quad c_{\mathrm{skip},k} = \frac{\sigma_{\mathrm{data}}^2}{\sigma_k^2 + \sigma_{\mathrm{data}}^2}, \quad c_{\mathrm{out},k} = \frac{\sigma_k \sigma_{\mathrm{data}}}{(\sigma_k^2 + \sigma_{\mathrm{data}}^2)^{1/2}}, \tag{50}$$

within the corresponding projector blocks, while the network noise embedding uses the rank-weighted mean $c_{\mathrm{noise}} = \frac{1}{4}\sum_k \pi_k \log \sigma_k$.

Structured Heun sampling advances along the sampler-$Y$ vector path. Writing $\widehat{\boldsymbol{x}}_i = D_\theta(\boldsymbol{x}_i, \boldsymbol{\sigma}_i)$ and $\Delta\sigma_{i,k} = \sigma_{i+1,k} - \sigma_{i,k}$, the Euler proposal is

$$\boldsymbol{x}_{i+1}^{\mathrm{E}} = \boldsymbol{x}_i + \sum_k \boldsymbol{P}_k(\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_i)\frac{\Delta\sigma_{i,k}}{\sigma_{i,k}}. \tag{51}$$

The Heun correction averages this blockwise derivative with the corresponding derivative evaluated at $(\boldsymbol{x}_{i+1}^{\mathrm{E}}, \boldsymbol{\sigma}_{i+1})$; the final correction is skipped at the zero-noise endpoint, as in EDM. Thus the off-diagonal cells use the same denoiser interface and solver as the diagonal cells, with only the sampling path changed.

### B.6 Generation blocks and shared generation seeds

Within any reported factorial or attribution block, evaluation conditions use the same generation-seed range. This common-random-number convention aligns the finite differences at the block level; it does not create independence across training seeds. The training-trajectory measurement-resolution blocks are A: 0–4095, B: 4096–8191, and C: 8192–12287, each with 4096 samples per condition. The early-lock falsification uses A for the base trajectory and B/C only at frozen critical checkpoints. The later transport test uses A at checkpoints 80/90/100 to select the interval under the frozen rule, excludes B/C from selection, and then forms paired checkpoint differences within A, B, and C on the selected interval. Earlier 50,000-sample and single-block 4096-sample studies retain their own conventions and are not relabeled as A/B/C studies.

### B.7 Metrics, signs, and closure integrity

FID is computed in the 2048-dimensional Inception feature space used by the released evaluator, with lower values indicating better feature-distribution agreement. The strict-factorial records also contain KID, precision, recall, density, and coverage at all three reported budgets: NFE7, NFE15, and NFE31. KID uses 100 subsets of size 1000 with evaluator seed 0. Precision, recall, density, and coverage use Euclidean $k$-nearest-neighbor feature radii with $k = 5$. The stored coverage definition is

$$\mathrm{coverage} = \frac{1}{n_{\mathrm{r}}}\sum_{i=1}^{n_{\mathrm{r}}} \mathbf{1}\left[\min_j \|\boldsymbol{r}_i - \boldsymbol{f}_j\|_2 < \rho_k(\boldsymbol{r}_i)\right], \tag{52}$$

where $\rho_k(\boldsymbol{r}_i)$ is the distance from real feature $\boldsymbol{r}_i$ to its $k$th-nearest real neighbor. The per-record KID subset standard deviation is an estimator diagnostic; the uncertainties reported in the manuscript tables are sample standard deviations across independently trained seeds.

The manuscript-facing strict-factorial main effects use cell-minus-Reference, so negative $\mathrm{G} - \mathrm{R}$, $\mathrm{C} - \mathrm{R}$, or $\mathrm{F} - \mathrm{R}$ values indicate lower, better FID for the intervention cell. All reported paired effects use this same cell-minus-Reference convention. The algebraic definitions of $I$, $Q$, and $P$ are unchanged, and any raw $x - y$ contrast is interpreted directly from that subtraction order.

For the $4 \times 4$ cross-evaluation grids, numerical checks compute $I - (Q_Y + P_Y)$ for every measured common sampler $Y$; the seven-condition checkpoint analyses use the corresponding $Y = \mathrm{F}$ closure. Values near floating-point zero audit condition mapping, signs, and aggregation; they do not test whether the identity is statistically true.

### B.8 Study-design boundaries

The five-seed CIFAR-100 strict factorial and the three-seed projector-family comparison are descriptive finite-system evaluations. The AFHQv2/FFHQ common-sampler grids were specified before

Table 7: Protocol identity of the image-generation evidence blocks used by the main analysis. DCT entries give the two frequency cutoffs; attribution quantities use FID.

| Evidence block | System | Train seeds | Projector/path and cells | Solver/NFE | Samples | Role in the paper |
|---|---|---|---|---|---|---|
| CIFAR projector-family check | CIFAR-100, $32^2$ | 1–3 | Reference ($\lambda = 0$) plus centered DCT or rank-matched Haar ($\lambda = 0.15$) | Heun, NFE7/15/31 | 50,000; seeds 0–49,999 | Tests whether the centered-path effect is specific to the DCT basis. |
| CIFAR-10 Scalar–Full comparison | CIFAR-10, $32^2$ | 1–3 | DCT 8/16; Scalar vs semantic Full; $\lambda = 0.067$ | Heun, NFE31 | 50,000; seeds 0–49,999 | Cross-dataset Scalar-versus-Structured comparison. |
| CIFAR strict factorial | CIFAR-100, $32^2$ | 11–15 | DCT 8/16; R/G/C/F; $\lambda =$ 0 for R and 0.15 otherwise | Heun, NFE7/15/31 | 50,000; seeds 0–49,999 | Main controlled decomposition of common and centered path coordinates. |
| AFHQv2 common-sampler attribution | AFHQv2, $64^2$ | 1–3 | DCT 16/32; R/G/C/F trained states $\times$ R/G/C/F samplers | Heun, NFE63 | 4096; seeds 0–4095 | Full $4 \times 4$ training–sampling attribution and sampler sensitivity. |
| FFHQ common-sampler attribution | FFHQ, $64^2$ | 1–3 | DCT 16/32; R/G/C/F trained states $\times$ R/G/C/F samplers | Heun, NFE63 | 4096; seeds 0–4095 | Full $4 \times 4$ training–sampling attribution and sampler sensitivity. |
| CIFAR fresh-seed dynamics | CIFAR-100, $32^2$ | 22, 23 | DCT 8/16; checkpoint R/G/C/F; seven evaluations | Heun, NFE31 | 4096/block | Tests whether interaction sign and internal attribution stabilize early. |
| CIFAR late-transport test | CIFAR-100, $32^2$ | 24 | DCT 8/16; checkpoints 80/90/100; seven evaluations | Heun, NFE31 | 4096/block; A/B/C | Independent measurement of late $Q/P$ transport and block variation. |

completing the off-diagonal evaluations. The training-dynamics analysis separates the fresh-seed early-stability test from the independent seed-24 late-transport test. These design distinctions determine the scope of the corresponding claims; none of the finite-system results is interpreted as a universal sign, ordering, or path-optimality statement.

### B.9 Aggregation and interval conventions

For the five-seed recovered strict factorial, cell summaries are means and sample standard deviations across training seeds using `ddof=1`. Paired effects preserve the seed pairing and use two-sided 95% Student-$t$ intervals with four degrees of freedom. The historical recovered summaries use $\mathrm{R} - X$; the main text reports the equivalent $X - \mathrm{R}$ values by multiplying each estimate and interval by $-1$. A deterministic aggregation script regenerates the manifest, paired effects, and aggregate summaries directly from the recovered metric records rather than from legacy display values.

For the training-trajectory analyses, $R^{\mathrm{MC}}$ and $R^{\mathrm{MC}}_{\Delta X}$ are maximum deviations across the three finite generation blocks, as defined in Eqs. 72 and 75. They are finite Monte-Carlo measurement radii, not confidence intervals over a population of datasets, training seeds, or trajectories. Pearson and Spearman coefficients with three or five training realizations, leave-one-seed-out checks, and constituent attributions are reported descriptively and carry no population-significance interpretation.

### B.10 Strict-factorial and projector-family results

The main-text Tables 2 and 3 are the manuscript-facing summaries of the five-seed strict-factorial experiment. The displays below keep that evidence separate from the projector-family robustness check: Table 8 reports the distinct three-seed projector comparison, while Tables 9, 10, and 11 provide the paired effects, per-seed records, and secondary metrics for the five-seed strict factorial.

Table 8: CIFAR-100 three-seed projector-family FID (↓), mean ± sample SD over training seeds 1–3 and 50,000 generated samples. This block is distinct from the five-seed strict factorial.

| Path | NFE7 | NFE15 | NFE31 |
|---|---|---|---|
| Scalar reference | $91.794 \pm 1.535$ | $15.684 \pm 0.041$ | $13.992 \pm 0.152$ |
| DCT Centered, $\lambda = 0.15$ | $89.657 \pm 2.755$ | $15.249 \pm 0.331$ | $13.219 \pm 0.271$ |
| Haar Centered, $\lambda = 0.15$ | $84.104 \pm 2.701$ | $15.081 \pm 0.345$ | $13.131 \pm 0.142$ |

Table 10: Complete CIFAR-100 strict-factorial FID records. Each row contains the four diagonal R/G/C/F cells for one independently trained seed at one sampling budget. All cells use generation seeds 0–49,999.

| NFE | Train seed | R | G | C | F |
|---|---|---|---|---|---|
| 7 | 11 | 89.93053 | 88.36662 | 88.18644 | 88.26231 |
| 7 | 12 | 92.71735 | 86.83531 | 91.72014 | 87.62103 |
| 7 | 13 | 90.27979 | 87.56108 | 92.86499 | 90.34374 |
| 7 | 14 | 91.41809 | 85.76709 | 89.49419 | 85.33092 |
| 7 | 15 | 89.83370 | 85.52452 | 87.46485 | 87.57630 |
| 15 | 11 | 15.63640 | 15.27020 | 15.34243 | 14.70288 |
| 15 | 12 | 15.45424 | 15.38185 | 14.83848 | 14.88348 |
| 15 | 13 | 15.21667 | 15.07244 | 14.80327 | 14.86752 |
| 15 | 14 | 15.30002 | 15.40197 | 14.91531 | 14.71472 |
| 15 | 15 | 15.52843 | 15.09048 | 14.77781 | 14.48472 |
| 31 | 11 | 14.29885 | 13.81897 | 14.29085 | 12.93496 |
| 31 | 12 | 13.92664 | 13.66029 | 12.82072 | 12.86334 |
| 31 | 13 | 13.74139 | 13.73532 | 13.00373 | 13.08396 |
| 31 | 14 | 13.56411 | 13.60173 | 12.80621 | 12.89272 |
| 31 | 15 | 14.00398 | 14.00474 | 13.06603 | 12.91638 |

Table 11: CIFAR-100 strict-factorial secondary distribution metrics across all reported sampling budgets. FID and KID are lower-better; precision, recall, density, and coverage are descriptive distribution diagnostics.

| NFE | Cell | FID | KID | Precision | Recall | Density | Coverage |
|---|---|---|---|---|---|---|---|
| 7 | R | $90.836 \pm 1.226$ | $0.08008 \pm 0.00115$ | $0.4850 \pm 0.0052$ | $0.1611 \pm 0.0083$ | $0.3164 \pm 0.0064$ | $0.1480 \pm 0.0039$ |
| 7 | G | $86.811 \pm 1.197$ | $0.07651 \pm 0.00134$ | $0.4879 \pm 0.0041$ | $0.1825 \pm 0.0076$ | $0.3183 \pm 0.0070$ | $0.1584 \pm 0.0023$ |
| 7 | C | $89.946 \pm 2.298$ | $0.07975 \pm 0.00233$ | $0.4228 \pm 0.0113$ | $0.1422 \pm 0.0100$ | $0.2265 \pm 0.0097$ | $0.1320 \pm 0.0044$ |
| 7 | F | $87.827 \pm 1.793$ | $0.07799 \pm 0.00233$ | $0.4363 \pm 0.0020$ | $0.1513 \pm 0.0063$ | $0.2413 \pm 0.0034$ | $0.1386 \pm 0.0033$ |
| 15 | R | $15.427 \pm 0.170$ | $0.00962 \pm 0.00014$ | $0.6438 \pm 0.0019$ | $0.6620 \pm 0.0011$ | $0.6165 \pm 0.0082$ | $0.5611 \pm 0.0028$ |
| 15 | G | $15.243 \pm 0.156$ | $0.00960 \pm 0.00013$ | $0.6393 \pm 0.0044$ | $0.6661 \pm 0.0033$ | $0.6091 \pm 0.0102$ | $0.5648 \pm 0.0031$ |
| 15 | C | $14.935 \pm 0.233$ | $0.00919 \pm 0.00020$ | $0.6183 \pm 0.0018$ | $0.6756 \pm 0.0039$ | $0.5624 \pm 0.0065$ | $0.5722 \pm 0.0013$ |
| 15 | F | $14.731 \pm 0.161$ | $0.00926 \pm 0.00025$ | $0.6168 \pm 0.0020$ | $0.6816 \pm 0.0032$ | $0.5599 \pm 0.0065$ | $0.5767 \pm 0.0037$ |
| 31 | R | $13.907 \pm 0.278$ | $0.01005 \pm 0.00031$ | $0.6419 \pm 0.0028$ | $0.6852 \pm 0.0017$ | $0.6051 \pm 0.0069$ | $0.5841 \pm 0.0040$ |
| 31 | G | $13.764 \pm 0.157$ | $0.01002 \pm 0.00026$ | $0.6404 \pm 0.0041$ | $0.6876 \pm 0.0034$ | $0.6035 \pm 0.0106$ | $0.5878 \pm 0.0014$ |
| 31 | C | $13.198 \pm 0.622$ | $0.00955 \pm 0.00070$ | $0.6338 \pm 0.0020$ | $0.6977 \pm 0.0055$ | $0.5900 \pm 0.0064$ | $0.6010 \pm 0.0036$ |
| 31 | F | $12.938 \pm 0.086$ | $0.00944 \pm 0.00023$ | $0.6328 \pm 0.0033$ | $0.7013 \pm 0.0041$ | $0.5880 \pm 0.0056$ | $0.6051 \pm 0.0030$ |

Table 9: CIFAR-100 five-seed strict-factorial paired effects in the manuscript-facing cell-minus-Reference convention, with two-sided 95% Student-$t$ intervals over paired training seeds 11–15. The raw interaction is $I = \mathrm{F} - \mathrm{G} - \mathrm{C} + \mathrm{R}$. Negative cell-minus-Reference effects indicate lower (better) FID.

| NFE | $\mathrm{G} - \mathrm{R}$ | $\mathrm{C} - \mathrm{R}$ | $\mathrm{F} - \mathrm{R}$ | $I$ |
|---|---|---|---|---|
| 7 | $-4.025\ [-6.343, -1.707]$ | $-0.890\ [-3.379, +1.599]$ | $-3.009\ [-6.153, +0.134]$ | $+1.906\ [-0.008, +3.820]$ |
| 15 | $-0.184\ [-0.457, +0.089]$ | $-0.492\ [-0.723, -0.260]$ | $-0.696\ [-1.051, -0.342]$ | $-0.021\ [-0.327, +0.285]$ |
| 31 | $-0.143\ [-0.421, +0.135]$ | $-0.709\ [-1.231, -0.188]$ | $-0.969\ [-1.344, -0.594]$ | $-0.116\ [-0.681, +0.448]$ |

These 60 records are the five training seeds $\times$ four factorial cells $\times$ three NFEs used to reconstruct the headline FID means, sample standard deviations, and paired contrasts. No interpolation or cross-NFE pooling is used.

### B.11 Secondary Distribution Metrics

Table 11 reports the secondary distribution metrics from the same 60 strict-factorial records used for FID. Values are mean $\pm$ sample SD over independently trained seeds 11–15; 50,000 generated samples are used per record.

At NFE15, the comparison highlighted in the main text is explicit: relative to R, C has lower precision and density but higher recall and coverage. The full table shows that this balance is itself sampling-budget dependent. In particular, C has lower recall and coverage than R at NFE7, but higher values at NFE15 and NFE31. These metrics are descriptive complements to FID; no paired uncertainty claim is assigned to them here.

# C Training–Sampling Attribution: Derivations and Full Results

## C.1 Common-Sampler Attribution: Definitions and Exact Closure

For any measured common sampler $Y \in \{\mathrm{R}, \mathrm{G}, \mathrm{C}, \mathrm{F}\}$, let

$$c_{\mathrm{R}} = +1, \qquad c_{\mathrm{G}} = -1, \qquad c_{\mathrm{C}} = -1, \qquad c_{\mathrm{F}} = +1$$

denote the factorial-interaction signs. The factorial interaction is

$$I = \sum_{X \in \{\mathrm{R},\mathrm{G},\mathrm{C},\mathrm{F}\}} c_X J_{X|X}. \tag{53}$$

Evaluating all four trained states under the same sampler $s_Y$ gives

$$Q_Y = \sum_{X \in \{\mathrm{R},\mathrm{G},\mathrm{C},\mathrm{F}\}} c_X J_{X|Y} = J_{\mathrm{F}|Y} - J_{\mathrm{G}|Y} - J_{\mathrm{C}|Y} + J_{\mathrm{R}|Y}. \tag{54}$$

The path-return correction is the signed change from the common sampler to each fixed trained state's own sampling path:

$$\begin{aligned} P_Y &= \sum_{X \in \{\mathrm{R},\mathrm{G},\mathrm{C},\mathrm{F}\}} c_X \left(J_{X|X} - J_{X|Y}\right) \\ &= (J_{\mathrm{F}|\mathrm{F}} - J_{\mathrm{F}|Y}) + (J_{\mathrm{G}|Y} - J_{\mathrm{G}|\mathrm{G}}) \\ &\quad + (J_{\mathrm{C}|Y} - J_{\mathrm{C}|\mathrm{C}}) + (J_{\mathrm{R}|\mathrm{R}} - J_{\mathrm{R}|Y}). \end{aligned} \tag{55}$$

Adding Eqs. 54 and 55 cancels every common-$Y$ term and leaves the four $J_{X|X}$ terms in Eq. 53. Therefore

$$\boxed{I = Q_Y + P_Y}, \qquad Y \in \{\mathrm{R}, \mathrm{G}, \mathrm{C}, \mathrm{F}\}. \tag{56}$$

This equality is algebraic. It does not imply that $Q_Y$ or $P_Y$ is invariant to the choice of $Y$.

The same construction can be written as a finite mixed difference. Let $\mathcal{J}(a, b)$ denote the response obtained when the trained state and sampling path both use intervention corner $(a, b)$, and let $\mathcal{J}_Y(a, b)$ denote the response of the trained state from the same corner when evaluated under the fixed sampler $s_Y$. Define

$$\Delta_{ab}\mathcal{F} = \mathcal{F}(1,1) - \mathcal{F}(1,0) - \mathcal{F}(0,1) + \mathcal{F}(0,0). \tag{57}$$

Then

$$I = \Delta_{ab}\mathcal{J}, \qquad Q_Y = \Delta_{ab}\mathcal{J}_Y, \qquad P_Y = \Delta_{ab}(\mathcal{J} - \mathcal{J}_Y). \tag{58}$$

These identities use only the four factorial corners and require neither intermediate intervention amplitudes nor differentiability.

An integrated-curvature representation follows only under an additional regularity assumption. For any $\mathcal{F} \in C^2([0,1]^2)$,

$$\begin{aligned} \Delta_{ab}\mathcal{F} &= [\mathcal{F}(1,1) - \mathcal{F}(1,0)] - [\mathcal{F}(0,1) - \mathcal{F}(0,0)] \\ &= \int_0^1 [\partial_b \mathcal{F}(1,b) - \partial_b \mathcal{F}(0,b)]\, \mathrm{d}b \\ &= \int_0^1 \int_0^1 \partial^2_{ab} \mathcal{F}(a,b)\, \mathrm{d}a\, \mathrm{d}b. \end{aligned} \tag{59}$$

Table 12: Final-FID attribution at NFE63. $\Delta J_{\mathrm{F-R}}$ is the Full-minus-Reference FID difference, with $T^{\mathrm{sym}}_{\mathrm{F-R}}$ and $S^{\mathrm{sym}}_{\mathrm{F-R}}$ its order-symmetric trained-state and path-return terms. $I$ is the factorial interaction, with $Q_{\mathrm{F}}$ and $P_{\mathrm{F}}$ its trained-state and path-return interaction terms under the Full common sampler.

| | | Full vs. Reference | | | Factorial interaction | | |
|---|---|---|---|---|---|---|---|
| Dataset | Seed | $\Delta J_{\mathrm{F-R}}$ | $T^{\mathrm{sym}}_{\mathrm{F-R}}$ | $S^{\mathrm{sym}}_{\mathrm{F-R}}$ | $I$ | $Q_{\mathrm{F}}$ | $P_{\mathrm{F}}$ |
| AFHQv2 | 1 | $-0.083$ | $-1.397$ | $+1.314$ | $-0.778$ | $-1.096$ | $+0.317$ |
| | 2 | $-0.405$ | $-2.319$ | $+1.913$ | $+0.387$ | $+2.905$ | $-2.518$ |
| | 3 | $-0.215$ | $-0.632$ | $+0.417$ | $-0.780$ | $-1.582$ | $+0.802$ |
| FFHQ | 1 | $+0.182$ | $-7.282$ | $+7.463$ | $-0.122$ | $-0.186$ | $+0.064$ |
| | 2 | $+0.585$ | $-6.369$ | $+6.954$ | $+0.320$ | $+0.060$ | $+0.260$ |
| | 3 | $-0.330$ | $-7.395$ | $+7.065$ | $-0.045$ | $-0.330$ | $+0.285$ |

Applying this identity to $\mathcal{J}$ and $\mathcal{J}_Y$ gives

$$I = \int_0^1 \int_0^1 \partial^2_{ab} \mathcal{J}(a,b)\,\mathrm{d}a\,\mathrm{d}b, \tag{60}$$

$$Q_Y = \int_0^1 \int_0^1 \partial^2_{ab} \mathcal{J}_Y(a,b)\,\mathrm{d}a\,\mathrm{d}b, \tag{61}$$

$$P_Y = \int_0^1 \int_0^1 \partial^2_{ab} [\mathcal{J} - \mathcal{J}_Y](a,b)\,\mathrm{d}a\,\mathrm{d}b. \tag{62}$$

The integral representation is conditional on smooth extensions of the finite response surfaces. The finite-difference definitions and Eq. 56 remain exact without this assumption.

### C.2 Reference-Contrast Decomposition

For $A \in \{\mathrm{G}, \mathrm{C}, \mathrm{F}\}$, define the reference contrast

$$\Delta J_{A-\mathrm{R}} = J_{A|A} - J_{\mathrm{R}|\mathrm{R}}. \tag{63}$$

Under common sampler $Y$, its trained-state part is

$$T^{(Y)}_{A-\mathrm{R}} = J_{A|Y} - J_{\mathrm{R}|Y}, \tag{64}$$

and the additional path-return correction is

$$S^{(Y)}_{A-\mathrm{R}} = (J_{A|A} - J_{A|Y}) + (J_{\mathrm{R}|Y} - J_{\mathrm{R}|\mathrm{R}}). \tag{65}$$

The common-$Y$ terms cancel directly:

$$\boxed{\Delta J_{A-\mathrm{R}} = T^{(Y)}_{A-\mathrm{R}} + S^{(Y)}_{A-\mathrm{R}}}. \tag{66}$$

Thus the reference contrast itself is unchanged by the attribution coordinate, while its split between $T^{(Y)}$ and $S^{(Y)}$ can change with $Y$.

For $A = \mathrm{F}$, $Y = \mathrm{R}$ first compares trained states under $s_{\mathrm{R}}$ and then returns the F-trained state to $s_{\mathrm{F}}$, whereas $Y = \mathrm{F}$ first evaluates the R-trained state under $s_{\mathrm{F}}$ and then compares trained states under that sampler. Averaging these two update orders gives the order-symmetric decomposition visualized in Figure 2(a) and summarized numerically in Table 12:

$$\begin{aligned} T^{\mathrm{sym}}_{\mathrm{F-R}} &= \frac{1}{2}\left[(J_{\mathrm{F}|\mathrm{R}} - J_{\mathrm{R}|\mathrm{R}}) + (J_{\mathrm{F}|\mathrm{F}} - J_{\mathrm{R}|\mathrm{F}})\right], \\ S^{\mathrm{sym}}_{\mathrm{F-R}} &= \frac{1}{2}\left[(J_{\mathrm{F}|\mathrm{F}} - J_{\mathrm{F}|\mathrm{R}}) + (J_{\mathrm{R}|\mathrm{F}} - J_{\mathrm{R}|\mathrm{R}})\right]. \end{aligned} \tag{67}$$

Averaging the two exact closures gives $\Delta J_{\mathrm{F-R}} = T^{\mathrm{sym}}_{\mathrm{F-R}} + S^{\mathrm{sym}}_{\mathrm{F-R}}$.

Table 13: AFHQv2 complete $4 \times 4$ FID cross-evaluation at NFE63. Rows identify the trained state $X$ and columns the sampling path $Y$.

| Seed | Trained state | $s_{\mathrm{R}}$ | $s_{\mathrm{G}}$ | $s_{\mathrm{C}}$ | $s_{\mathrm{F}}$ |
|---|---|---|---|---|---|
| 1 | R | 9.74058 | 9.74596 | 25.79010 | 26.59010 |
| | G | 10.23340 | 10.22200 | 26.57770 | 27.40000 |
| | C | 24.31680 | 24.27830 | 9.95406 | 9.94298 |
| | F | 23.87900 | 23.86310 | 9.67182 | 9.65738 |
| 2 | R | 10.34650 | 10.34670 | 28.09800 | 28.90080 |
| | G | 9.75112 | 9.74768 | 25.11690 | 25.82840 |
| | C | 24.48170 | 24.46850 | 10.15320 | 10.10890 |
| | F | 24.66910 | 24.65240 | 9.99870 | 9.94101 |
| 3 | R | 9.86859 | 9.86258 | 26.34170 | 27.14160 |
| | G | 10.52750 | 10.52470 | 27.81930 | 28.64880 |
| | C | 23.45490 | 23.43830 | 9.77716 | 9.72834 |
| | F | 26.09310 | 26.07940 | 9.70334 | 9.65352 |

Table 14: FFHQ complete $4 \times 4$ FID cross-evaluation at NFE63. Rows identify the trained state $X$ and columns the sampling path $Y$.

| Seed | Trained state | $s_{\mathrm{R}}$ | $s_{\mathrm{G}}$ | $s_{\mathrm{C}}$ | $s_{\mathrm{F}}$ |
|---|---|---|---|---|---|
| 1 | R | 13.23800 | 13.23930 | 35.31190 | 36.28790 |
| | G | 13.73560 | 13.74470 | 35.78860 | 36.77960 |
| | C | 21.59010 | 21.56600 | 13.03550 | 13.11400 |
| | F | 21.54290 | 21.51880 | 13.32480 | 13.41990 |
| 2 | R | 13.02000 | 13.02070 | 35.03350 | 35.99720 |
| | G | 13.41270 | 13.41450 | 35.47710 | 36.54620 |
| | C | 21.59530 | 21.56940 | 12.89030 | 12.99580 |
| | F | 22.67350 | 22.64530 | 13.52230 | 13.60520 |
| 3 | R | 12.95180 | 12.95360 | 34.10480 | 35.15900 |
| | G | 12.79110 | 12.79550 | 33.95460 | 35.19260 |
| | C | 21.68190 | 21.66190 | 12.82300 | 12.91810 |
| | F | 20.69880 | 20.67870 | 12.54020 | 12.62190 |

### C.3 Complete $4 \times 4$ Cross-Evaluation Results

The common-sampler experiment reuses the fixed R/G/C/F trained states and evaluates all 16 pairs $J_{X|Y}$ at NFE63 with 4096 generated samples (seeds 0–4095) per cell. No retraining is performed. Off-diagonal cells deliberately evaluate a trained state under a different sampling path and can therefore have much larger FID than diagonal cells; these values define the counterfactual attribution coordinates and are not interpreted as intrinsic model quality or evidence of solver failure. The protocol was frozen before the missing cross-evaluation cells were generated. The maximum absolute closure errors are $5.329 \times 10^{-15}$ for AFHQv2 and $8.882 \times 10^{-15}$ for FFHQ in Eq. 56, and $3.553 \times 10^{-15}$ and $1.776 \times 10^{-15}$, respectively, in Eq. 66. The tables retain stored decimal precision to audit these algebraic closures; this precision is not an inferential claim for the 4096-sample FID estimates.

### C.4 Complete Common-Sampler Attribution Coordinates

Tables 15 and 16 report the coordinates computed directly from the measured matrices. Under $Y = \mathrm{F}$, $Q_{\mathrm{F}}$ and $I$ have the same sign in all six runs. Across the four common samplers, $Q_{\mathrm{R}}$ and $Q_{\mathrm{G}}$ are close, as are $Q_{\mathrm{C}}$ and $Q_{\mathrm{F}}$; the larger shifts occur between the R/G and C/F groups. AFHQv2 seed 3 changes sign between those groups, while the other five runs preserve the sign of $Q_Y$.

The across-seed summaries preserve the same distinction.

For the across-seed mean vector, the common-sampler spread in $Q_Y$ is $0.55511$ on AFHQv2 and $0.18800$ on FFHQ. The seed-wise $Q_Y$ spreads are $0.20426$, $2.12173$, and $3.56131$ for AFHQv2 and $0.36680$, $0.62510$, and $0.69250$ for FFHQ. These summaries are descriptive across three training seeds and are not population confidence statements.

Table 15: AFHQv2 common-sampler attribution. For every row and every $Y$, $I = Q_Y + P_Y$ up to floating-point precision.

| Seed | $I$ | $Q_{\mathrm{R}}$ | $Q_{\mathrm{G}}$ | $Q_{\mathrm{C}}$ | $Q_{\mathrm{F}}$ | $P_{\mathrm{R}}$ | $P_{\mathrm{G}}$ | $P_{\mathrm{C}}$ | $P_{\mathrm{F}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $-0.77810$ | $-0.93062$ | $-0.89124$ | $-1.06984$ | $-1.09550$ | $+0.15252$ | $+0.11314$ | $+0.29174$ | $+0.31740$ |
| 2 | $+0.38663$ | $+0.78278$ | $+0.78292$ | $+2.82660$ | $+2.90451$ | $-0.39615$ | $-0.39629$ | $-2.43997$ | $-2.51788$ |
| 3 | $-0.77975$ | $+1.97929$ | $+1.97898$ | $-1.55142$ | $-1.58202$ | $-2.75904$ | $-2.75873$ | $+0.77167$ | $+0.80227$ |

Table 16: FFHQ common-sampler attribution. For every row and every $Y$, $I = Q_Y + P_Y$ up to floating-point precision.

| Seed | $I$ | $Q_{\mathrm{R}}$ | $Q_{\mathrm{G}}$ | $Q_{\mathrm{C}}$ | $Q_{\mathrm{F}}$ | $P_{\mathrm{R}}$ | $P_{\mathrm{G}}$ | $P_{\mathrm{C}}$ | $P_{\mathrm{F}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $-0.12230$ | $-0.54480$ | $-0.55260$ | $-0.18740$ | $-0.18580$ | $+0.42250$ | $+0.43030$ | $+0.06510$ | $+0.06350$ |
| 2 | $+0.32040$ | $+0.68550$ | $+0.68210$ | $+0.18840$ | $+0.06040$ | $-0.36510$ | $-0.36170$ | $+0.13200$ | $+0.26000$ |
| 3 | $-0.04480$ | $-0.82240$ | $-0.82510$ | $-0.13260$ | $-0.32980$ | $+0.77760$ | $+0.78030$ | $+0.08780$ | $+0.28500$ |

### C.5 Reference-Contrast Attribution Under Different Common Samplers

Table 18 reports the common-sampler decompositions for all reference contrasts. The reference contrast $\Delta J_{A-\mathrm{R}}$ is fixed, while its allocation between $T^{(Y)}$ and $S^{(Y)}$ can change substantially with $Y$. For Full versus Reference, the main-text symmetric terms are formed seedwise by averaging the $Y = \mathrm{R}$ and $Y = \mathrm{F}$ decompositions.

Table 19 expands the symmetric Full-versus-Reference columns of Table 12 by reporting the two ordered decompositions before averaging. The two ordered decompositions can allocate the same reference contrast very differently; their average removes the preference for either update order. In all six runs, the resulting symmetric $T/S$ terms have opposite signs and are larger in magnitude than the Full-versus-Reference difference.

## D Training-Trajectory Protocols and Diagnostics

### D.1 Checkpoint coordinate construction

At each nominal target $t_k$, the frozen protocol saves the first batch-boundary checkpoint that reaches that target for all four independently trained $R/G/C/F$ cells. The seven Section 4 evaluations (four same-setting evaluations and three common-Full cross-evaluations) are then performed at the aligned target. Writing $J_{X|Y}(t_k) = \mathcal{E}_{\mathcal{S}}(\theta_X(t_k), s_Y)$ gives

$$\begin{aligned} Q_k &= J_{\mathrm{F|F}}(t_k) - J_{\mathrm{G|F}}(t_k) - J_{\mathrm{C|F}}(t_k) + J_{\mathrm{R|F}}(t_k), \\ I_k &= J_{\mathrm{F|F}}(t_k) - J_{\mathrm{G|G}}(t_k) - J_{\mathrm{C|C}}(t_k) + J_{\mathrm{R|R}}(t_k), \\ P_k &= I_k - Q_k. \end{aligned} \tag{68}$$

Thus every checkpoint obeys $I_k = Q_k + P_k$ with the same signs and evaluator-path convention as the Section 4 attribution construction. Consecutive differences obey $\Delta I_k = \Delta Q_k + \Delta P_k$ exactly, without a smooth-time or dynamical-model assumption.

### D.2 Complete fresh-seed attribution trajectories

Figure 2(c) visualizes the Block-A trajectories used for the fresh-seed stability analysis in Section 5.2, while Table 20 reports their exact checkpoint coordinates. The recorded checkpoints span 40–100% of training; no values are interpolated for earlier progress. At every checkpoint, the finite-sample identity $I(t) = Q(t) + P(t)$ holds to floating-point precision.

### D.3 Early-stability criterion

The early-stability analysis uses ex-post future radii

$$R_Q^*(k) = \max_{j \geq k} |Q_j - Q_k|, \quad R_P^*(k) = \max_{j \geq k} |P_j - P_k|, \quad R_\Sigma^*(k) = R_Q^*(k) + R_P^*(k). \tag{69}$$

Table 17: Common-sampler attribution summaries at NFE63, mean ± sample SD across training seeds 1–3.

| System | Sampler $Y$ | $Q_Y$ | $P_Y$ |
|---|---|---|---|
| AFHQv2 | R | $+0.61048 \pm 1.46259$ | $-1.00089 \pm 1.54712$ |
| AFHQv2 | G | $+0.62355 \pm 1.44173$ | $-1.01396 \pm 1.53233$ |
| AFHQv2 | C | $+0.06845 \pm 2.40074$ | $-0.45885 \pm 1.73240$ |
| AFHQv2 | F | $+0.07566 \pm 2.46190$ | $-0.46607 \pm 1.79338$ |
| FFHQ | R | $-0.22723 \pm 0.80254$ | $+0.27833 \pm 0.58483$ |
| FFHQ | G | $-0.23187 \pm 0.80316$ | $+0.28297 \pm 0.58508$ |
| FFHQ | C | $-0.04387 \pm 0.20301$ | $+0.09497 \pm 0.03402$ |
| FFHQ | F | $-0.15173 \pm 0.19732$ | $+0.20283 \pm 0.12131$ |

Table 18: Across-seed mean decomposition of reference contrasts. Each entry under a common sampler is $T^{(Y)}/S^{(Y)}$, and each pair sums to the corresponding reference contrast before rounding.

| System | Contrast | $\Delta J$ | $Y = \mathrm{R}$: $T/S$ | $Y = \mathrm{G}$: $T/S$ | $Y = \mathrm{C}$: $T/S$ | $Y = \mathrm{F}$: $T/S$ |
|---|---|---|---|---|---|---|
| AFHQv2 | G − R | +0.17957 | $+0.18545/-0.00588$ | $+0.17971/-0.00014$ | $-0.23863/+0.41820$ | $-0.25177/+0.43134$ |
| AFHQv2 | C − R | −0.02375 | $+14.09924/-14.12299$ | $+14.07662/-14.10037$ | $-16.78179/+16.75804$ | $-17.61743/+17.59368$ |
| AFHQv2 | F − R | −0.23459 | $+14.89518/-15.12976$ | $+14.87989/-15.11447$ | $-16.95198/+16.71739$ | $-17.79353/+17.55894$ |
| FFHQ | G − R | +0.24830 | $+0.24320/+0.00510$ | $+0.24703/+0.00127$ | $+0.25670/-0.00840$ | $+0.35810/-0.10980$ |
| FFHQ | C − R | −0.15367 | $+8.55250/-8.70617$ | $+8.52790/-8.68157$ | $-21.90047/+21.74680$ | $-22.80540/+22.65173$ |
| FFHQ | F − R | +0.14573 | $+8.56847/-8.42273$ | $+8.54307/-8.39733$ | $-21.68763/+21.83337$ | $-22.59903/+22.74477$ |

These are oracle feasibility diagnostics, not deployable prefix predictors. For a compensated candidate, the coordinate and interaction margins are

$$M^*_{\mathrm{mech}}(k) = \min\{|Q_k| - R^*_Q(k),\ |P_k| - R^*_P(k)\}, \qquad M^*_I(k) = |I_k| - R^*_\Sigma(k). \tag{70}$$

For same-sign candidates, $M^*_{\mathrm{mech}}$ additionally included the dominance margin $\big||Q_k| - |P_k|\big| - R^*_\Sigma(k)$. The first persistent positive-margin times defined $\tau^*_{\mathrm{mech}}$ and $\tau^*_I$. The fresh-seed test evaluates the prediction $\tau^*_{\mathrm{mech}} < \tau^*_I$ together with separation and cancellation-margin conditions. These implications remain valid when their future-variation premises hold; the experiment tests whether finite trajectories satisfy those premises early enough to support a useful stability certificate.

### D.4 Fresh-seed early-lock falsification

Under the fresh-seed test, seed 22 had

$$\tau^*_I/T = 0.70, \qquad \tau^*_{\mathrm{mech}}/T = 1.00. \tag{71}$$

Thus the interaction sign stabilizes before the final checkpoint, but the $Q/P$ relation does not. Seed 23 had larger late than mid-stage variation, with both oracle lock times equal to the final checkpoint. Independent blocks resolve the decisive signs: at 90%, $(\bar{Q}, R^{\mathrm{MC}}_Q) = (-1.0949, 0.2749)$ and $(\bar{I}, R^{\mathrm{MC}}_I) = (-1.6525, 0.1350)$; at 100%, they are $(1.5512, 0.3522)$ and $(1.0164, 0.0291)$. Hence both sides of the final $Q = 0$ and $I = 0$ crossings are resolved. The earlier block-A $P = 0$ crossing is retained only as an observed base-trajectory event because the frozen evidence does not independently resolve both sides.

### D.5 Prospective late-transport protocol

After the early-lock failure, the separate late-transport protocol froze seed 24 and the late checkpoint set $K = \{8, 9, 10\}$. Block A selected $k^*$ as the larger late displacement over the frozen 80–90% and 90–100% candidate intervals; no B/C value could change that choice. At the two selected checkpoints, blocks A, B, C supplied paired differences

$$\Delta X^{(b)} = X^{(b)}_{k^*+1} - X^{(b)}_{k^*}, \quad \overline{\Delta X} = \frac{1}{3}\sum_b \Delta X^{(b)}, \quad R^{\mathrm{MC}}_{\Delta X} = \max_b |\Delta X^{(b)} - \overline{\Delta X}|, \tag{72}$$

for $X \in \{Q, P\}$. The paired values were

$$\begin{aligned}(\Delta Q_{\mathrm{A}}, \Delta Q_{\mathrm{B}}, \Delta Q_{\mathrm{C}}) &= (2.0887, 1.5393, 2.0787),\\ (\Delta P_{\mathrm{A}}, \Delta P_{\mathrm{B}}, \Delta P_{\mathrm{C}}) &= (-0.6987, 0.0507, -0.7335).\end{aligned} \tag{73}$$

Table 19: Seedwise order-symmetric Full-versus-Reference attribution at NFE63. Each $T/S$ pair sums to $\Delta J_{\mathrm{F-R}}$ before rounding, and the symmetric pair is the average of the $Y = \mathrm{R}$ and $Y = \mathrm{F}$ decompositions.

| Dataset | Seed | $\Delta J_{\mathrm{F-R}}$ | $Y = \mathrm{R}$: $T/S$ | $Y = \mathrm{F}$: $T/S$ | Symmetric $T/S$ |
|---|---|---|---|---|---|
| AFHQv2 | 1 | $-0.08320$ | $+14.13842/-14.22162$ | $-16.93272/+16.84952$ | $-1.39715/+1.31395$ |
| | 2 | $-0.40549$ | $+14.32260/-14.72809$ | $-18.95979/+18.55430$ | $-2.31860/+1.91310$ |
| | 3 | $-0.21507$ | $+16.22451/-16.43958$ | $-17.48808/+17.27301$ | $-0.63179/+0.41671$ |
| FFHQ | 1 | $+0.18190$ | $+8.30490/-8.12300$ | $-22.86800/+23.04990$ | $-7.28155/+7.46345$ |
| | 2 | $+0.58520$ | $+9.65350/-9.06830$ | $-22.39200/+22.97720$ | $-6.36925/+6.95445$ |
| | 3 | $-0.32990$ | $+7.74700/-8.07690$ | $-22.53710/+22.20720$ | $-7.39505/+7.06515$ |

Table 20: Complete Block-A attribution trajectories for fresh CIFAR-100 seeds 22 and 23. Values are the fixed-F trained-state interaction $Q$, the path-return correction $P$, and the factorial interaction $I = Q + P$ at each recorded checkpoint.

| | Seed 22 | | | Seed 23 | | |
|---|---|---|---|---|---|---|
| Training progress | $Q$ | $P$ | $I$ | $Q$ | $P$ | $I$ |
| 40% | $+1.1572$ | $-0.8752$ | $+0.2820$ | $-1.0902$ | $+1.3327$ | $+0.2425$ |
| 50% | $+1.4247$ | $-0.5090$ | $+0.9157$ | $-0.6513$ | $+0.2823$ | $-0.3690$ |
| 60% | $+1.0448$ | $-0.1012$ | $+0.9436$ | $-0.8015$ | $+0.4945$ | $-0.3070$ |
| 70% | $+1.6667$ | $-0.4340$ | $+1.2327$ | $-0.6998$ | $+0.8520$ | $+0.1522$ |
| 80% | $+0.8821$ | $-0.0136$ | $+0.8685$ | $-0.8578$ | $+1.0953$ | $+0.2375$ |
| 90% | $+0.9769$ | $-0.1809$ | $+0.7960$ | $-1.1008$ | $-0.4167$ | $-1.5175$ |
| 100% | $+0.9340$ | $-0.4227$ | $+0.5113$ | $+1.1990$ | $-0.2087$ | $+0.9903$ |

The corresponding factorial-interaction differences are

$$(\Delta I_{\mathrm{A}}, \Delta I_{\mathrm{B}}, \Delta I_{\mathrm{C}}) = (1.3900, 1.5900, 1.3452), \qquad \overline{\Delta I} = 1.4417, \tag{74}$$

and satisfy $\Delta I_b = \Delta Q_b + \Delta P_b$ block by block. Together, these values give the means and measurement radii reported in Section 5.2 and the frozen margins in Eq. 19. Both frozen margins are positive. However, since $|\overline{\Delta P}| < R^{\mathrm{MC}}_{\Delta P}$, only the $Q$ displacement is individually resolved; the late result is therefore $Q$-driven.

### D.6 Measurement-resolved boundary rule and seed 24 status

For a canonical linear boundary $h_b(\boldsymbol{z}) = 0$, let

$$\bar{d}_b(j) = \frac{1}{3}\sum_r h_b(\boldsymbol{z}_j^{(r)}), \qquad R_b^{\mathrm{MC}}(j) = \max_r |h_b(\boldsymbol{z}_j^{(r)}) - \bar{d}_b(j)|. \tag{75}$$

A side is resolved when $|\bar{d}_b(j)| > R_b^{\mathrm{MC}}(j)$. A crossing is measurement-resolved only when the means at the two checkpoints have opposite signs and both sides satisfy this inequality. This conditional rule is separate from the transport margins.

For seed 24, $Q$ is resolved positive at both checkpoints: $(\bar{Q}_8, R_Q^{\mathrm{MC}}) = (0.6659, 0.3319)$ and $(\bar{Q}_9, R_Q^{\mathrm{MC}}) = (2.5681, 0.0310)$. Likewise $I$ is resolved positive at both checkpoints: $(0.2662, 0.1843)$ and $(1.7080, 0.1325)$. For $P$, the $80\%$ side $(-0.3997, 0.4818)$ is unresolved, while the $90\%$ side $(-0.8602, 0.1332)$ is resolved negative. The selected interval therefore contains no verified canonical crossing. The radii throughout this appendix quantify finite three-block Monte-Carlo measurement resolution and are not population confidence intervals.