TROPICAL DIFFUSION DYNAMICS: WHY NEARLY IDENTICAL DIFFUSION STATES REACH DIFFERENT OUTPUT ROUTES

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper investigates how deterministic probability-flow ordinary differential equations in diffusion models can produce different terminal outputs from nearly identical intermediate states. We address this by framing Gaussian denoising as a soft-minimum competition among clean-data routes, whose zero-noise limit yields nearest-support cells and, for finite atomic priors, tropical power cells. We prove that isolated regular two-route boundaries form a logistic transition layer of width $O(\sigma^2)$ with peak denoiser sensitivity $O(\sigma^{-2})$, where σ is the noise standard deviation. We show that between-route posterior covariance simultaneously determines the Fisher tensor of route probabilities, the route component of reverse-flow strain, and the rate of Gaussian route-information gain. Because the remaining flow transports and amplifies these contrasts before decoding, we introduce a matrix-free Cauchy–Green Jacobian–vector product probe that maximizes linearized terminal displacement within a low-rank subspace. These theoretical insights are operationalized into Tropical Route Control (TRC), a modular framework combining finite-horizon probes, trajectory-level conformal event control, and exact prefix allocation under a hard continuation budget. Our experiments verify the predicted local sensitivity and finite-horizon perturbation effects and demonstrate that route-aware verification provides a valid and practical way to detect declared endpoint changes under a fixed computation budget in real settings.

1 INTRODUCTION

A deterministic diffusion sampler assigns one output to each intermediate state (Song et al., 2021b). Yet a small change to that state can produce a visibly different image. Consider two nearby states that eventually depict the same beach under rainy and sunny skies. Each follows a unique trajectory, but the two trajectories reach different output regions. The question is therefore local: which small changes matter, and when does denoising make them consequential? The answer lies in the geometry of plausible outcomes around a state and the dynamics that carry them through sampling.

Gaussian denoising turns the beach example into a competition among plausible clean explanations. Gaussian likelihoods decay exponentially with squared distance divided by the noise variance $q = \sigma^2$. As noise decreases, small cost differences increasingly determine which explanation dominates. The important states lie near ties, where small displacements can change the preferred explanation. Tropical geometry is useful because it describes minima of affine costs and the boundaries where those costs tie (Speyer & Sturmfels, 2009; Maclagan & Sturmfels, 2015). For finitely many clean points, removing a common quadratic term from their Gaussian costs leaves affine functions, so this geometry comes directly from the observation model. General compact supports give the corresponding nearest-support limit. At positive noise, ties broaden into smooth transitions. At a regular two-route interface, the transition width is proportional to σ^2, and peak sensitivity is proportional to σ^{-2}. The narrowing layer therefore retains a finite integrated response.

The remaining sampler determines what this local sensitivity ultimately means. Posterior covariance identifies the contrasts visible now, but later dynamics can rotate, amplify, or suppress them before decoding. Consequently, a large instantaneous response need not produce the largest terminal change. We resolve this distinction by connecting between-route covariance to route probabilities,

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.

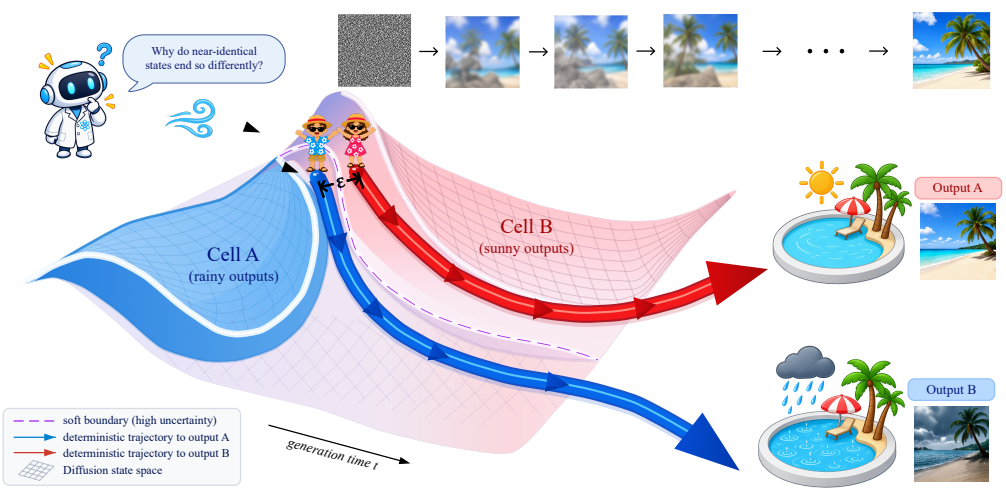

Figure 1: **Nearby states, distinct deterministic outcomes.** Blue and red trajectories start ϵ apart near a soft route boundary and follow unique continuations toward rainy and sunny outputs. The diagram links local route competition to the distinct terminal decisions reached by neighboring states.

flow strain, and information gain, then integrating the linearized dynamics along the remaining trajectory. The resulting finite-horizon direction requires only Jacobian–vector products.

These results lead to *Tropical Route Control* (TRC): a practical framework for probing and verifying endpoint changes in a fixed pretrained sampler. Its direction module uses the remaining flow, its calibration module controls false acceptance of a declared event, and its allocation module distributes a hard continuation budget. The geometric theory concerns fixed terminal cells, such as classifier decisions. The verification experiments use a separately defined endpoint-dispersion event, so its representation and threshold are specified explicitly. Controlled models test the mechanism, DiT-XL/2 tests finite-horizon effects across sampling conditions, and native SDXL provides an exploratory transfer test. Each study thus tests a distinct prediction of the geometry–dynamics connection.

Our contributions are fourfold:

1. **We identify the geometry of route competition.** Gaussian smoothing induces soft tropical cells for compact route mixtures. Regular two-route boundaries have a logistic profile. The two-atom normal model conserves the clean-route separation exactly.
2. **We connect geometry to finite-horizon dynamics.** We prove that one route-information tensor measures spatial route sensitivity, the route component of flow strain, and Gaussian evidence acquisition. We derive finite-radius gain bounds and a cubic-error diameter law for symmetric feature clouds under smoothness and full rank of the feature derivative.
3. **We turn the mechanism into verifiable decisions.** TRC combines finite-radius probe bounds, simultaneous trajectory-level event control, and exact optimization of certified prefix values. The modular construction makes the assumptions of each decision explicit.
4. **We test mechanistic predictions in learned models.** Controlled interventions recover the predicted direction, and source-disjoint DiT-XL/2 confirms the finite-horizon advantage across matched probes. Three solvers reproduce the temporal organization. Frozen policies improve budgeted event capture on DiT-XL/2 and in an exploratory SDXL transfer study.

2 Related Work

Diffusion geometry and dynamical regimes. Diffusion theory has linked denoising to scores, stochastic differential equations, and Gaussian posterior structure (Ho et al., 2020; Song et al., 2021b; Robbins, 1956; Efron, 2011). Stochastic filtering supplies the observation-process formalism (Bain & Crisan, 2009), while stochastic localization connects such processes to sampling and

covariance evolution (Montanari, 2023; Benton et al., 2024; Grenioux et al., 2024). Complementary work explains denoiser generalization through geometry-adaptive harmonic representations (Kadkhodaie et al., 2024). Population phase analyses describe macroscopic speciation and hierarchical reconstruction (Raya & Ambrogioni, 2023; Biroli et al., 2024; Sclocchi et al., 2025). We condition on one noisy state and ask how perturbations change its route posterior and terminal decision, connecting local covariance to the eventual nonlinear decoded outcome.

Fast sampling and endpoint decisions. High-order and adaptive solvers reduce neural function evaluations by controlling continuous integration error (Lu et al., 2022; Zhao et al., 2023; Zhang et al., 2025). Covariance-aware updates and endpoint prediction exploit posterior structure more directly (Schioppa & Salimans, 2026; Peng & Gao, 2026). Our target is orthogonal: TRC keeps the base solver and selects states for endpoint verification. Its guarantee concerns a declared event and hard budget, not continuous integration error or sample quality. Symmetric, norm-matched perturbations also distinguish this fixed-budget test from unrestricted adversarial search.

Posterior covariance has also been used to learn diagonal reverse-transition variance, solve inverse problems, and detect distribution shift from covariance eigenvalues (Ou et al., 2025; Peng et al., 2024; Shoushtari et al., 2026). Those methods target likelihood, reconstruction, or out-of-distribution detection. Our object is the between-route part of posterior covariance for a declared terminal decision. It is transported through the remaining sampler to rank finite-horizon probes. A covariance-aware comparison therefore fixes the endpoint task and matches the perturbation norm and remaining continuation through the pretrained sampler.

Tropical geometry and selective risk control. Tropical methods describe affine-energy minima and piecewise-linear networks (Speyer & Sturmfels, 2009; Maclagan & Sturmfels, 2015; Zhang et al., 2018a). Here the tropical object is induced by Gaussian smoothing of the data law: its smooth log-partition derivatives are posterior mean and covariance after noise scaling. Selective and multi-valid conformal methods provide finite-sample risk guarantees after a score has been learned (Bao et al., 2024; Bastani et al., 2022). Geometry chooses informative probes, whereas calibration controls errors for the resulting event: neither component substitutes for the other. TRC combines them while keeping the generator and its original sampling objective fixed.

Local decision geometry versus representation geometry. Learned features encode different notions of perceptual similarity, language alignment, and image identity (Zhang et al., 2018b; Radford et al., 2021; Pizzi et al., 2022). A classifier instead supplies a pointwise decision boundary, as used in local adversarial-distance analysis (Moosavi-Dezfooli et al., 2016). These are different targets, not interchangeable definitions of a route. Once a regular terminal boundary is fixed, our transport and local tube arguments apply to its pullback. Changing the representation changes that boundary and its measured event, not the form of the conditional argument.

3 Tropical Geometry and Route Dynamics

We first locate competing clean explanations, then follow their effect through the remaining flow. Notation is collected in Appendix A. Appendices B and C provide the complete proofs and extend the analysis to mixtures with potentially different support dimensions.

3.1 Gaussian Smoothing Creates Soft Route Cells

Let the clean random vector $X \in \mathbb{R}^d$ have law Π_0, and let $Y_q = X + \sqrt{q}Z$, where $Z \sim \mathcal{N}(0, I_d)$, I_d is the d-dimensional identity, and $q = \sigma^2 > 0$ is noise variance. Denote the density of Y_q by p_q, its posterior mean by $m_q(y) = \mathbb{E}[X \mid Y_q = y]$, its score by $s_q(y) = \nabla_y \log p_q(y)$, and its posterior covariance by $\Sigma_q(y) = \mathrm{Cov}(X \mid Y_q = y)$. Tweedie's identity gives

$$s_q(y) = \frac{m_q(y) - y}{q}, \qquad D_y m_q(y) = \frac{\Sigma_q(y)}{q}. \tag{1}$$

The first equality says that the score points from the noisy state toward the posterior average of plausible clean explanations. The second says that the denoiser responds most strongly in directions where the posterior clean explanations have the largest variance, not merely the largest score.

Introduce a latent route $R \in [K] = \{1, \dots, K\}$ and suppose $\Pi_0 = \sum_{a=1}^{K} \pi_a \Pi_a$, where $\pi_a = \mathbb{P}(R = a) > 0$, $\Pi_a = \mathcal{L}(X \mid R = a)$, and each route law has compact support S_a. Its unnormalized Gaussian evidence and negative-log evidence are

$$Z_{a,q}(y) = \pi_a \int e^{-\|y-x\|^2/(2q)} \Pi_a(\,\mathrm{d}x), \qquad V_{a,q}(y) = -q \log Z_{a,q}(y).$$

Proposition 1 (Support-level tropicalization). *The route posterior is the exact soft assignment*

$$W_a(y,q) = \mathbb{P}(R = a \mid Y_q = y) = \frac{e^{-V_{a,q}(y)/q}}{\sum_{b=1}^{K} e^{-V_{b,q}(y)/q}}, \qquad V_{a,q}(y) \longrightarrow \tfrac{1}{2}\,\mathrm{dist}(y, S_a)^2 \tag{2}$$

locally uniformly as $q \downarrow 0$. Moreover, consider any sequence $y_q \to y_0$ for which routes a, b have unique limiting nearest points $x_a^\star \neq x_b^\star$, $W_a(y_q, q) \to \omega \in (0,1)$, $W_b(y_q, q) \to 1 - \omega$, and all other route weights vanish. The rescaled denoiser derivative then converges to the following rank-one covariance:

$$q D_y m_q(y_q) \longrightarrow \omega(1-\omega)(x_a^\star - x_b^\star)(x_a^\star - x_b^\star)^\top. \tag{3}$$

The rank-one limit exchanges posterior mass between the two nearest supports: its eigenvalue is their squared separation weighted by the remaining ambiguity. More generally, k asymptotically active routes produce at most $k - 1$ nonvanishing contrast directions. Under the isolated two-route conditions in Proposition 11, and centered on the exact noise-dependent tie, the 10–90% transition width is $2q \log 9/\|x_a^\star - x_b^\star\| + o(q)$, while the balanced peak normal gain is $\|x_a^\star - x_b^\star\|^2/(4q) + o(q^{-1})$. A sequential local integral of the normal gain converges to the route separation. Thus the boundary becomes thinner and sharper while its normal response is conserved. Appendices B.2 and C.2.4 give the compact-uniform Laplace proof and the corresponding mixed-rank correction. Appendix B.6 proves the measure-valued limit and denoiser-based recovery, excluding points with nonunique nearest projections onto the clean route supports.

For example, equal atoms at $\pm a$, $a > 0$, give $m_q(y) = a \tanh(ay/q)$: the peak derivative is a^2/q, but its integral is always $2a$. This concrete model explains the local profile, not an arbitrary learned classifier boundary or unrestricted semantic change in generated images.

3.2 One tensor links sensitivity, strain, and information

For the route variable above, the channel obeys $R \to X \to Y_q$. At a fixed (y, q), let $m_{a,q}(y) = \mathbb{E}[X \mid R = a, Y_q = y]$. Define the categorical Fisher matrix $\mathcal{I}_R(y,q) = \sum_{a=1}^{K} W_a \nabla_y \log W_a \nabla_y \log W_a^\top$, the between-route covariance $\Sigma_q^{\mathrm{between}}(y) = \sum_a W_a (m_{a,q} - m_q)(m_{a,q} - m_q)^\top$, and the within-route covariance $\Sigma_q^{\mathrm{within}}(y) = \sum_a W_a \,\mathrm{Cov}(X \mid R = a, Y_q = y)$. For signal-to-noise ratio $\gamma > 0$, also define the Gaussian channel $Z_\gamma = \sqrt{\gamma} X + N$, where $N \sim \mathcal{N}(0, I_d)$ is independent of (R, X), and let $I(R; Z_\gamma)$ denote mutual information. In reverse log-variance time $\eta = -\log q$, probability-flow velocity is $b_\eta(y) = (m_q(y) - y)/2$. Its Jacobian measures the infinitesimal strain of neighboring states under denoising.

Theorem 2 (One route-information tensor, three roles). *For every $q > 0$, with D_y denoting the Jacobian and all pointwise terms evaluated at the common noisy state (y, q), the following three identities hold:*

$$q^2 \mathcal{I}_R = \Sigma_q^{\mathrm{between}}, \tag{4}$$

$$D_y\left(\frac{m_q - y}{2}\right) = -\frac{I_d}{2} + \frac{\Sigma_q^{\mathrm{within}}}{2q} + \frac{q}{2}\mathcal{I}_R, \tag{5}$$

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} I(R; Z_\gamma)\Big|_{\gamma = 1/q} = \frac{q^2}{2}\mathbb{E}\,\mathrm{tr}\,\mathcal{I}_R(Y_q, q). \tag{6}$$

If each route law is a single atom, a unit direction u expands in reverse log-variance time $\eta = -\log q$ exactly when $q\, u^\top \mathcal{I}_R u > 1$.

The same matrix measures changes in route odds, the between-route component of reverse-flow strain, and Gaussian information gain. For continuous route laws, $\Sigma_q^{\mathrm{within}}$ also contributes: expansion along a unit vector u requires $u^\top \Sigma_q u > q$, and can occur without route competition. The proof

combines Gaussian differentiation, total covariance, and the information–minimum mean-square error identity (I–MMSE) (Guo et al., 2005), as detailed in Appendix B.1. The same channel bounds posterior-mean route disagreement and information-limited amplification (Appendix B.6).

For $0 < q < Q$, parameterize variance-exploding probability flow from Q down to q by

$$\frac{\mathrm{d}y_s}{\mathrm{d}s} = \frac{y_s - m_s(y_s)}{2s}, \qquad Q \longrightarrow q < Q. \tag{7}$$

Theorem 3 (Integrated tropical susceptibility)**.** *Assume the trajectory y_s exists on $[q, Q]$. Let $J_{s\leftarrow Q} = D_{y_Q}\Phi_{s\leftarrow Q}(y_Q)$ be the derivative of the flow map with respect to its state at variance Q, choose $\delta_Q \neq 0$, set $\delta_s = J_{s\leftarrow Q}\delta_Q$, and abbreviate $\Sigma_s = \Sigma_s(y_s)$. Then*

$$\log\frac{\|\delta_q\|}{\|\delta_Q\|} = \frac{1}{2}\log\frac{q}{Q} + \frac{1}{2}\int_q^Q \frac{\widehat{\delta}_s^\top \Sigma_s \widehat{\delta}_s}{s^2}\,\mathrm{d}s, \qquad \widehat{\delta}_s = \delta_s/\|\delta_s\|. \tag{8}$$

The logarithm separates universal contraction from accumulated posterior variance, including within-route and between-route contributions. In particular, the perturbation direction evolves with the flow, so the integral accounts for covariance encountered after the anchor. Appendix B.3 derives the corresponding time-ordered identity. This integrated covariance exposure is not specific to one parameterization: a standard scalar-SDE variance-preserving (VP) probability flow is a scalar conjugate of the same flow, and an exact deterministic denoising diffusion implicit model (DDIM) tangent has the identical instantaneous threshold $\lambda_j(\Sigma_q) > q$. The conjugacy and the correction between these noise coordinates are derived in Appendix B.6, so thresholds are compared in common units.

3.3 Terminal routes are transported boundaries

Let $F_{t\to 0} : \mathbb{R}^d \to \mathbb{R}^m$ be the remaining deterministic sampler and decoder, let $\phi : \mathbb{R}^m \to \mathbb{R}^{d_\phi}$ be a frozen d_ϕ-dimensional representation, and let $h : \mathbb{R}^{d_\phi} \to \mathbb{R}$ be a continuously differentiable scalar margin whose zero set separates two terminal cells. The pulled-back margin is a terminal decision evaluated from the current state, and its gradient gives the corresponding local state-space normal:

$$\ell_t(x) = h(\phi(F_{t\to 0}(x))), \qquad a_t(x) = \nabla\ell_t(x). \tag{9}$$

Theorem 4 (Transported terminal boundaries and perturbation crossings)**.** *Assume a C^1 positive-noise probability-flow diffeomorphism with vector field v_t pushes the state density p_t forward in time. Then*

$$(\partial_t + v_t^\top\nabla)\ell_t = 0. \tag{10}$$

Each pulled-back terminal cell preserves its probability mass. Perturbed paths can cross its boundary only through motion relative to this transport, as the exact identity in Appendix B.3 quantifies.

To quantify perturbation crossings, let X_t have density p_t, let $\lambda : \mathbb{R}^d \to [0,\infty)$ be a bounded continuous state scale, and let $\mathcal{W} = (W_0, \dots, W_L)$ be bounded offsets independent of X_t. For radius $r > 0$, let $\chi_{r,\mathcal{W}}(x)$ be the conditional probability that the signs of $\ell_t(x + r\lambda(x)W_j)$ are not all equal. For a unit normal n, define the expected projected width $\kappa_{\mathcal{W}}(n) = \mathbb{E}[\max_j n^\top W_j - \min_j n^\top W_j]$. If $\{\ell_t = 0\}$ is a compact positive-reach C^2 hypersurface, $\inf_{\{\ell_t=0\}}\|\nabla\ell_t\| > 0$, and p_t is continuous nearby, the small-radius split probability satisfies the surface integral

$$\mathbb{E}\chi_{r,\mathcal{W}}(X_t) = r\int_{\{\ell_t=0\}} \lambda(x)p_t(x)\kappa_{\mathcal{W}}(a_t/\|a_t\|)\,\mathrm{d}\mathcal{H}^{d-1}(x) + o(r). \tag{11}$$

Here $\mathcal{H}^{d-1}$ is hypersurface measure. The integrand has a simple interpretation: state density measures how often a boundary neighborhood is visited, and projected cloud width measures how easily its probes straddle that boundary. Thus only the component normal to the boundary contributes at first order. Tangential motion changes the crossing probability at higher order under the stated geometric conditions on the regular boundary and bounded probe distribution.

At a regular transverse k-cell junction, for a cloud of at least k points, the appendix proves $\mathbb{P}(E_r^{(k)}) = C_{k,\mathcal{W}}r^{k-1} + o(r^{k-1})$. This is always an $O(r^{k-1})$ statement. The exponent is leading only when the explicitly defined normal-fan coefficient $C_{k,\mathcal{W}}$ is positive and finite. The tube proof is in Appendix B.3. The exact junction coefficient, its transport, and finite-support cell-change

results appear in Appendix B.6, with each required regularity condition. The resulting laws describe the small-radius limit of fixed pointwise cells. Their empirical range is examined separately in Appendix D.6. For finite-cloud dispersion, which is our operational verification target, the next section develops a direct finite-radius bound for smooth decoded feature maps.

4 FINITE-HORIZON ROUTE CONTROL

We now measure perturbations by their terminal effect and use this measurement to formulate verification decisions under a fixed budget. The full proofs appear in Appendices B.4 and B.5.

4.1 A MATRIX-FREE TERMINAL ACTION

Fix an anchor $x \in \mathbb{R}^d$ and let $F : \mathbb{R}^d \to \mathbb{R}^m$ denote the complete remaining learned continuation, including decoding when appropriate. Let r_P be the probe-subspace dimension and let $P \in \mathbb{R}^{d\times r_P}$ have orthonormal columns. Exact Jacobian–vector products form $G = DF(x)P$. Their pairwise inner products form the small Cauchy–Green matrix that measures terminal gain within the chosen probe subspace:

$$C_P = G^\top G = P^\top DF(x)^\top DF(x)P. \tag{12}$$

Its leading eigenvector $v_\star$ maximizes infinitesimal terminal displacement inside $\mathrm{span}(P)$. Only directional derivatives and an $r_P \times r_P$ eigensystem are needed, not a full state Jacobian. Our direction experiment draws an outcome-independent Gaussian matrix and uses its reduced QR factor for P. All compared directions therefore use the same subspace, making their different use of the remaining dynamics directly comparable at every paired anchor.

Theorem 5 (Finite-horizon action and finite-radius terminal gain)**.** *Assume $r_P \geq 2$, $\rho > 0$, F is C^3 on a neighborhood of the closed ball $\overline{B}(x,\rho)$, and for some $L_3 \geq 0$, $\|D^3F(z)[Pw,Pw,Pw]\|_2 \leq L_3$ for every $z \in \overline{B}(x,\rho)$ and every unit $w \in \mathbb{R}^{r_P}$. Let $\varsigma_1 > \varsigma_2$ be the two largest singular values of G. For $g_\rho(v) = \|F(x+\rho Pv) - F(x-\rho Pv)\|/(2\rho)$, any unit comparator v at angle $\theta = \arccos|v^\top v_\star| \in [0,\pi/2]$ has the following terminal-gain bounds:*

$$g_\rho(v_\star) \geq \varsigma_1 - L_3\rho^2/6, \tag{13}$$

$$g_\rho(v) \leq \sqrt{\varsigma_1^2\cos^2\theta + \varsigma_2^2\sin^2\theta} + L_3\rho^2/6. \tag{14}$$

Each gain bound pays $L_3\rho^2/6$, so subtracting them costs $L_3\rho^2/3$. A finite-horizon direction therefore remains better when its linear spectral advantage exceeds this nonlinear remainder. This inequality specifies the radius within which the chosen representation preserves the direction's predicted advantage at the chosen finite probe radius against a specified comparator direction.

For the actual feature cloud, an additional result closes a different gap: cross-direction pairs can exceed every matched positive/negative pair. For a smooth decoded feature map $H(z) = \phi(F(x+Pz))$, full-column-rank $DH(0)$, unit probe directions u_j, and sufficiently small ρ, Theorem 19 proves that the diameter of $\{H(0), H(\pm\rho u_j)\}$ is $2\rho\max_j\|DH(0)u_j\| + O(\rho^3)$, with explicit constants. Symmetry cancels quadratic error even for the maximizing cross-direction pair. Full rank is essential for this cancellation: Appendix B.4.1 gives the explicit constants, the proof, and a counterexample demonstrating the need for full rank within the probe subspace.

4.2 CERTIFIED EVENT DETECTION UNDER A HARD BUDGET

For trajectory i and anchor t, let x_{it} be the state, choose unit directions $u_{it,m} \in \mathbb{R}^d$ for $m = 1,\ldots,M_{\mathrm{pair}}$, and decode the endpoint cloud $\mathcal{Z}_{it} = \{F(x_{it})\}\cup\{F(x_{it}\pm\rho u_{it,m}) : 1 \leq m \leq M_{\mathrm{pair}}\}$. With a frozen endpoint dissimilarity Δ_{out} and threshold $\kappa > 0$, define

$$E_{it}^{\mathrm{CL}} = \mathbf{1}\left\{\max_{z,z'\in\mathcal{Z}_{it}} \Delta_{\mathrm{out}}(z,z') > \kappa\right\}. \tag{15}$$

This event asks whether equally small perturbations of one noisy state remain mutually similar after decoding. It is deterministic once the sampler, radius, cloud, representation, and cut are fixed. It

measures continuous output dispersion under the declared representation and threshold. Because this predicate differs from a fixed classifier decision, its statistical control below is derived directly. The scheduling experiments use 12 normalized full-state Gaussian rays, and the rank-8 direction experiment separately evaluates the proposed finite-horizon sensitivity mechanism.

A predictor trained on data independent of calibration produces a frozen score $\widehat{p}_{it} \in [0,1]$ from information available at anchor t. For each of n calibration trajectories $i = 1,\dots,n$, let $\mathcal{T}_i$ be its fixed set of dependent anchors. Define one maximum residual for each trajectory and an upper event score for each of its anchors:

$$A_i = \max_{t\in\mathcal{T}_i}(E^{\mathrm{CL}}_{it} - \widehat{p}_{it}), \qquad U_{it} = \mathrm{clip}(\widehat{p}_{it} + \widehat{Q}_{1-\alpha}, 0, 1), \tag{16}$$

where $\alpha \in (0,1)$ is the target error level and $\widehat{Q}_{1-\alpha}$ is the $\lceil (n+1)(1-\alpha)\rceil$-th ordered calibration score, using $+\infty$ if that rank is $n+1$, which conservatively prevents acceptance.

Theorem 6 (Simultaneous trajectory-level event control)**.** *Conditional on the independently trained and frozen predictor, representation, event threshold, and anchor rule, assume the n calibration arrays and one test trajectory array are exchangeable. The resulting upper scores cover the binary labels simultaneously over all test anchors:*

$$\mathbb{P}\{\forall t \in \mathcal{T}_{n+1} : E^{\mathrm{CL}}_{n+1,t} \le U_{n+1,t}\} \ge 1-\alpha. \tag{17}$$

Consequently, for a fixed acceptance threshold $\tau \in [0,1)$, accepting only anchors with $U_{n+1,t} \le \tau$ has probability at most α of accepting any split event anywhere on the entire new test trajectory.

Unresolved anchors define jobs $j = 1,\dots,J$. Job j belongs to a frozen development stratum a_j, permits a nested prefix depth $m \in \{0,\dots,M_j\}$, and costs $c_j m$ integer budget units. Define $F_a(m)$ as the probability, in stratum a's frozen unresolved population, that the declared event occurs and is first detected by prefix m. Let $L_a(m)$ be a simultaneous lower confidence band for all (a,m). With total budget B, let $D_j(b)$ be the largest summed lower bound attainable for the first j jobs using at most $b \le B$ units. Optimizing the depth of the last job gives the following exact dynamic-programming recursion:

$$D_j(b) = \max_{0\le m\le \min(M_j,\lfloor b/c_j\rfloor)} \{D_{j-1}(b-c_j m) + L_{a_j}(m)\}, \qquad D_0(b) = 0. \tag{18}$$

This prefix-constrained knapsack chooses how deeply to verify each anchor. Buying only complete clouds would exclude its intermediate-depth options at earlier, more costly anchors.

The recursion makes the computational choice explicit. For every affordable depth of the current job, it adds that job's certified discovery value to the best allocation of the remaining budget to preceding jobs, then takes the largest sum. Its optimality is therefore with respect to the specified lower bounds and prefix costs. It does not require the risk predictor itself to rank anchors perfectly, although tighter discovery bounds can improve the resulting allocation of verification effort.

Composition of guarantees. Operationally, accept anchors with $U_{it} \le \tau$, allocate depths to the remaining jobs using (18), and run each signed prefix until its first split or allocated depth. Allocation additionally requires valid discovery bounds conditional on the selected queue, not only population averages (Proposition 22). Early stopping respects the cap. The static optimum does not certify reallocation or the empirical priority heuristic evaluated under the same cap.

5 Experiments

5.1 Experimental setting

Benchmarks. We separate mechanism from deployment. Controlled class-conditional U-Net and diffusion-transformer models provide full Jacobians and known interventions. Large-checkpoint studies use class-conditional DiT-XL/2 (Peebles & Xie, 2023) and native 1024×1024 SDXL (Podell et al., 2024). DiT experiments use matched denoising diffusion implicit model (DDIM) (Song et al., 2021a), DPM-Solver++ (Lu et al., 2025), and UniPC (Zhao et al., 2023) continuations. Pointwise-cell audits use a frozen ImageNet-1K ResNet-50 (Deng et al., 2009; He et al., 2016). Source-disjoint text-to-image transfer uses images and captions from the standard val2017 split of Microsoft Common Objects in Context (COCO) (Lin et al., 2014).

Protocol and metrics. Every comparison pairs initial noise, anchor, radius, continuation, and budget within trajectory. The finite-horizon test uses a frozen rank-8 subspace and relative state radius $\rho/\|x\|_2 = 0.05$. Deployment uses noise-grid fractions $t \in \{0.75, 0.50, 0.25\}$ and at most 12 signed pairs. For decoded images z, z', its frozen dissimilarity is $\Delta_{\text{out}}(z, z') = \max\{d_{\text{CLIP}}(z, z')/0.06, d_{\text{pix32}}(z, z')/0.12\}$, where d_{CLIP} is cosine distance from OpenAI CLIP ViT-B/32 QuickGELU and d_{pix32} is pixel root mean squared error after 32×32 resizing. Equation (15) uses $\kappa = 1.2$. ResNet-50 top-1 cells provide the separate pointwise partition used to assess the scope of the asymptotic theory.

Allocation reports full-cloud positives found under the per-trajectory cap $B = \lfloor 0.25 \sum_{j=1}^{3} 12c_j \rfloor$, with signed-pair continuation costs $(c_1, c_2, c_3) = (150, 102, 52)$ neural function evaluations (NFE). This budget measures continuation NFE. Direction construction, decoding, and feature costs are reported separately in Appendix D.4. Temporal experiments use seven anchors from 0.875 to 0.125 and report the first-versus-last count, the mean of the first three minus the last three event rates, and the per-trajectory log signal-to-noise ratio slope. Appendices D.1 and D.2 specify the frozen protocol and paired statistics, and Appendix D.3 reports expanded results with uncertainty.

With these costs, exhaustive verification spends 3648 continuation NFE per trajectory, while the allocation test permits 912. An early signed pair costs almost as much as three late pairs because more denoising steps remain. Uniform progress gives anchors equal prefix depth, whereas risk-based allocation spends this same cap where a new verified event is most likely. Accordingly, capture measures discoveries within the cap, and the hindsight ceiling quantifies how many discoveries that cap could have bought on the realized trajectories and fixed probe clouds.

5.2 Finite-horizon directions predict terminal changes

The controlled intervention first asks whether the local covariance direction is causal for endpoint separation. For the 64×64 state Jacobian J, the preregistered gate retains a group when $A = (J + J^\top)/2$ satisfies $\lambda_1(A) \geq 1.0001$, $\operatorname{tr}(A) - 64 \leq -10^{-4}$, and $\lambda_1(A) - \lambda_2(A) \geq 0.01$, before its endpoints are evaluated. $\lambda_j(A)$ denotes the j-th largest eigenvalue. The leading direction beats a norm-matched random control in $95/95$ retained U-Net and $110/111$ retained DiT groups. Median endpoint-feature separation improves by $4.17\times$ and $3.17\times$, respectively. Appendices E.2 to E.4 give the quality gates, spectra, hidden-state probes, and output-grounded interventions.

The large-model test distinguishes a one-step direction from the time-ordered action. On 64 source-disjoint DiT-XL/2 trajectories, exact automatic-differentiation Jacobian–vector products compute both directions in the same frozen rank-8 subspace. All trajectories have normalized Jacobian–vector/vector–Jacobian consistency residual below 10^{-3}. At relative perturbation norm 0.05, the finite-horizon direction improves median terminal latent log-gain by 0.0582 over the instantaneous direction and by 0.357 over a matched random ray. The latter contrast is positive for all 64 trajectories. Thus accounting for the remaining dynamics selects perturbations whose advantage persists through nonlinear continuation, as predicted by Theorem 5 within the tested subspace.

A new eight-trajectory pilot compares ranks $4, 8, 16$ and relative radii $0.025, 0.05, 0.10$ against a centered finite-difference terminal action. At rank 8 and radius 0.05, exact action improves median log-gain by 0.109 over the instantaneous direction. The finite-difference approximation is closely aligned with exact action, with direction cosine 0.981 and a median log-gain difference of 0.006. Construction costs are 10.38, 0.92, and 3.30 seconds for exact, instantaneous, and finite-difference actions. This comparison identifies a cheaper route to the same finite-horizon measurement. The descriptive pilot, including every rank and radius and the observed limits of the local approximation, is reported in Appendix D.4, together with the timing conventions.

5.3 Route competition forms a reproducible temporal phase

Seven-anchor DiT-XL/2 studies show that the declared fragmentation event is concentrated early in reverse time. On a fully source-disjoint cohort, all three matched-grid solvers reproduce a positive early-to-late contrast and a strictly negative log-SNR slope (Table 1). The effect is largest for DPM-Solver++ and DDIM and smaller, but still separated from zero, for UniPC. Solver choice changes which individual trajectories are fragile. The consistent early concentration shows when verification is most likely to encounter competing endpoints across these sampling rules.

Solver	Early / late positives	Contrast (pp)	log-SNR slope
DPM-Solver++	$51/64$ / $1/64$	66.1	$-.1008$
DDIM	$46/64$ / $0/64$	61.5	$-.0932$
UniPC	$16/64$ / $0/64$	21.4	$-.0321$

Table 1: **Source-disjoint temporal replication on 64 paired groups.** At seven anchor fractions $0.875, 0.750, \ldots, 0.125$, counts compare the first and last anchors. Contrast is the first-three minus last-three mean event rate in percentage points. Paired uncertainty appears in Appendix D.3.

Evaluation	Pos.	Frozen risk	Bellman	Uniform	Gain (pp)
DiT-XL/2 prospective	540	87.0%	85.4%	79.1%	7.96
DiT-XL/2 replication	537	87.0%	86.6%	75.6%	11.36
DiT-XL/2 radius $2\times$	847	95.3%	94.7%	90.1%	5.19
SDXL COCO transfer	13	$10/13$	N/A	$2/13$	61.5

Table 2: **Complete-link event capture under matched hard verification budgets.** DiT uses a 25% pathwise continuation-NFE cap. SDXL uses the frozen native-resolution continuation budget and has no Bellman comparator. Gains compare frozen risk with uniform allocation in percentage points. Appendix D.3 reports uncertainty paired by independent trajectory or source.

5.4 Frozen event-priority allocation improves budgeted capture

Six policies are frozen before outcomes: raw event-risk and cost-aware orders, a Bellman planner, random and static orders, and uniform progress. In two independent 512-trajectory cohorts, raw-risk scheduling captures 87.0% of full-cloud positives, gaining 7.96 and 11.36 percentage points over equally adaptive uniform progress (Table 2). It reaches 98.5% and 97.9% of the hindsight ceiling under realized prefix costs. Doubling the radius without refitting preserves the gain and raises ceiling-normalized capture to 99.6%. Bellman is the strongest non-oracle comparator. These evaluations do not establish a difference between it and raw-risk scheduling, despite their common advantage over uniform progress in these source-disjoint evaluations.

For native-resolution SDXL, 64 source-and-prompt-disjoint COCO sources contain 13 positives. The same frozen priority captures $10/13$, versus $2/13$ for uniform allocation, a gain of 61.5 percentage points. This exploratory transfer tests event capture on new prompts. Coverage after this architecture and source shift requires target calibration beyond the DiT guarantee.

5.5 Finite clouds reveal otherwise missed events

Nested probes separate verification coverage from the radius of a perturbation. Across three independent 192-group DiT cohorts at the midpoint, extending 12 to 24 signed pairs uncovers 41 additional events: 12 pairs miss 20.7% of the 198 events visible at 24 pairs. Each cohort passes its prespecified discovery gate. Because the clouds are nested, each additional event is witnessed at the same radius. Radius fixes perturbation magnitude; cloud size fixes directional coverage, so fair detector comparisons must control both. Thus probe count changes which events can be observed even at a fixed perturbation radius, motivating prefix verification of a *declared finite cloud*. Appendix D.5 gives nested comparisons, and Appendix D.6 examines the scope and assumptions underlying the reported cell, representation, and allocation results within the tested regimes.

6 Conclusion

Gaussian smoothing turns route competition into a soft tropical geometry. Between-route covariance links posterior sensitivity, the route component of flow strain, and information gain. Finite-horizon transport then determines which local perturbations have a lasting terminal effect, while symmetric-cloud bounds connect this derivative to finite-radius verification. Controlled and DiT experiments support the subspace direction, and separately evaluated frozen priorities improve budgeted event capture. Together, these results explain how a unique deterministic continuation can coexist with sensitive endpoint decisions and provide a practical way to investigate that sensitivity.

REFERENCES

E. Alpaydin and C. Kaynak. Optical recognition of handwritten digits. UCI Machine Learning Repository, 1998. URL https://doi.org/10.24432/C50P49. Dataset.

Alan Bain and Dan Crisan. *Fundamentals of Stochastic Filtering*, volume 60 of *Stochastic Modelling and Applied Probability*. Springer, 2009.

Yajie Bao, Yuyang Huo, Haojie Ren, and Changliang Zou. Selective conformal inference with false coverage-statement rate control. *Biometrika*, 111(3):727–742, 2024.

Osbert Bastani, Varun Gupta, Christopher Jung, Georgy Noarov, Ramya Ramalingam, and Aaron Roth. Practical adversarial multivalid conformal prediction. In *Advances in Neural Information Processing Systems*, 2022.

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*, 2024.

Giulio Biroli, Tony Bonnaire, Valentin De Bortoli, and Marc Mézard. Dynamical regimes of diffusion models. *Nature Communications*, 15:9957, 2024. doi: 10.1038/s41467-024-54281-3. URL https://doi.org/10.1038/s41467-024-54281-3.

Tony Bonnaire, Raphaël Urfin, Giulio Biroli, and Marc Mézard. Why diffusion models don't memorize: The role of implicit dynamical regularization in training. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-4720. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/ceb7f3cc876a6dcb15130a645b5a4507-Abstract-Conference.html.

Nicolas Brosse and Arnak S. Dalalyan. Boundary-layer asymptotics for gaussian-smoothed singular measures. *arXiv preprint arXiv:2607.04514*, 2026.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001. URL https://doi.org/10.1137/0707001.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

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

Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3): 418–491, 1959.

Herbert Federer. *Geometric Measure Theory*. Springer, 1969.

Louis Grenioux, Maxence Noble, Marylou Gabrié, and Alain Oliviero Durmus. Stochastic localization via iterative posterior sampling. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 16337–16376. PMLR, 2024. URL https://proceedings.mlr.press/v235/grenioux24a.html.

Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, 2005. doi: 10.1109/TIT.2005.844072. URL https://doi.org/10.1109/TIT.2005.844072.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

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

Anubhav Jain, Yuya Kobayashi, Takashi Shibuya, Yuhta Takida, Nasir Memon, Julian Togelius, and Yuki Mitsufuji. Classifier-free guidance inside the attraction basin may cause memorization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12871–12879, 2025. doi: 10.1109/CVPR52734.2025.01201. URL https://doi.org/10.1109/CVPR52734.2025.01201.

Zahra Kadkhodaie, Florentin Guth, Eero P. Simoncelli, and Stéphane Mallat. Generalization in diffusion models arises from geometry-adaptive harmonic representations. In *International Conference on Learning Representations*, 2024.

Tim Kaiser and Markus Kollmann. Diffusion models memorize in training–and generalize in inference. *arXiv preprint arXiv:2603.13419*, 2026.

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, pp. 26565–26577, 2022. doi: 10.52202/068431-1926. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a98846e9d9cc01cfb87eb694d946ce6b-Abstract-Conference.html.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pp. 740–755, 2014.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

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. In *Advances in Neural Information Processing Systems*, 2022.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, 22:730–751, 2025. doi: 10.1007/s11633-025-1562-4. URL https://doi.org/10.1007/s11633-025-1562-4.

Grace Luo, Lisa Dunlap, Dong Huk Park, Aleksander Holynski, and Trevor Darrell. Diffusion hyperfeatures: Searching through time and space for semantic correspondence. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2057. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/942032b61720a3fd64897efe46237c81-Abstract-Conference.html.

Diane Maclagan and Bernd Sturmfels. *Introduction to Tropical Geometry*. American Mathematical Society, 2015.

Per Christian Moan and Jitse Niesen. Convergence of the Magnus series. *Foundations of Computational Mathematics*, 8(3):291–301, 2008. doi: 10.1007/s10208-007-9010-0. URL https://doi.org/10.1007/s10208-007-9010-0.

Andrea Montanari. Sampling, diffusions, and stochastic localization. *arXiv preprint arXiv:2305.10690*, 2023. doi: 10.48550/arXiv.2305.10690. URL https://arxiv.org/abs/2305.10690. Forthcoming in Statistical Science.

Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. DeepFool: A simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2574–2582, 2016. doi: 10.1109/CVPR.2016.282. URL https://openaccess.thecvf.com/content_cvpr_2016/html/Moosavi-Dezfooli_DeepFool_A_Simple_CVPR_2016_paper.html.

Zijing Ou, Mingtian Zhang, Andi Zhang, Tim Xiao, Yingzhen Li, and David Barber. Improving probabilistic diffusion models with optimal diagonal covariance matching. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/a97b58c4f7551053b0512f92244b0810-Abstract-Conference.html.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85): 2825–2830, 2011. URL https://jmlr.org/papers/v12/pedregosa11a.html.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, 2023.

Xin Peng and Ang Gao. x-prediction is all you need: Training-free accelerated generation via endpoint decodability. *arXiv preprint arXiv:2607.06114*, 2026. URL https://arxiv.org/abs/2607.06114.

Xinyu Peng, Ziyang Zheng, Wenrui Dai, Nuoqian Xiao, Chenglin Li, Junni Zou, and Hongkai Xiong. Improving diffusion models for inverse problems using optimal posterior covariance. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 40347–40370. PMLR, 2024. URL https://proceedings.mlr.press/v235/peng24h.html.

Ed Pizzi, Sreya Dutta Roy, Sugosh Nagavara Ravindra, Priya Goyal, and Matthijs Douze. A self-supervised descriptor for image copy detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14532–14542, 2022.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/081b08068e4733ae3e7ad019fe8d172f-Abstract-Conference.html.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021.

Gabriel Raya and Luca Ambrogioni. Spontaneous symmetry breaking in generative diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2897. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/d0da30e312b75a3fffd9e9191f8bc1b0-Abstract-Conference.html.

Herbert Robbins. An empirical bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pp. 157–164. University of California Press, 1956. doi: 10.1525/9780520313880-015. URL https://doi.org/10.1525/9780520313880-015.

Krisanu Sarkar. Score shocks: The burgers equation structure of diffusion generative models. *arXiv preprint arXiv:2604.07404*, 2026.

Andrea Schioppa and Tim Salimans. Covariance-aware sampling for diffusion models. *arXiv preprint arXiv:2605.13910*, 2026. doi: 10.48550/arXiv.2605.13910. URL https://arxiv.org/abs/2605.13910.

Antonio Sclocchi, Alessandro Favero, and Matthieu Wyart. A phase transition in diffusion models reveals the hierarchical nature of data. *Proceedings of the National Academy of Sciences*, 122(1): e2408799121, 2025. doi: 10.1073/pnas.2408799121. URL https://doi.org/10.1073/pnas.2408799121.

Zebang Shen, Ya-Ping Hsieh, and Niao He. Manifold generalization provably proceeds memorization in diffusion models. *arXiv preprint arXiv:2603.23792*, 2026.

Shirin Shoushtari, Yi Wang, Xiao Shi, M. Salman Asif, and Ulugbek S. Kamilov. EigenScore: OOD detection using posterior covariance in diffusion models. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/778b22d74907b01741bf1b6e82420b88-Abstract-Conference.html.

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.

David Speyer and Bernd Sturmfels. Tropical mathematics. *Mathematics Magazine*, 82(3):163–173, 2009.

S. R. Srinivasa Varadhan. On the behavior of the fundamental solution of the heat equation with variable coefficients. *Communications on Pure and Applied Mathematics*, 20:431–455, 1967.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/d9e42b4d7163931f3689d6d6fbaa11d0-Abstract-Conference.html.

Hui Zhang, Zuxuan Wu, Zhen Xing, Jie Shao, and Yu-Gang Jiang. AdaDiff: Adaptive step selection for fast diffusion models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(9): 9914–9922, 2025.

Liwen Zhang, Gregory Naitzat, and Lek-Heng Lim. Tropical geometry of deep neural networks. In *International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5824–5832, 2018a.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018b.

Shiheng Zhang. Asymptotic preservation and uniform accuracy of diffusion and flow-matching samplers. *arXiv preprint arXiv:2607.04113*, 2026. doi: 10.48550/arXiv.2607.04113. URL https://arxiv.org/abs/2607.04113.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A unified predictor-corrector framework for fast sampling of diffusion models. In *Advances in Neural Information Processing Systems*, 2023.

Appendix

Contents

A Notation and symbol guide . . . 15
B Proofs and theoretical extensions . . . 16
B.1 Gaussian posterior calculus . . . 16
B.2 Soft tropical cells and crystallization . . . 18
B.3 Time-ordered dynamics and transported route cells . . . 19
B.4 Finite-horizon action and finite-radius terminal gain . . . 21
B.5 Trajectory-level certification and exact allocation . . . 23
B.6 Additional structural consequences . . . 24
C Extended theory and controlled geometric studies . . . 30
C.1 Setup for exact and stratified models . . . 30
C.2 Exact tropical geometry . . . 32
C.3 Controlled checks of soft tropical geometry . . . 34
C.4 Moving-interface and event-resolution studies . . . 35
C.5 Adaptive controls near sharp route layers . . . 36
C.6 Predicting route changes from tropical margins . . . 36
C.7 A continuous stratified Y-junction . . . 37
C.8 Empirical crystallization on finite supports . . . 37
C.9 Learned-denoiser refinement of the empirical skeleton . . . 38
C.10 Cell birth, adjacency, and persistence . . . 39
C.11 Complementary proofs for exact and stratified models . . . 39
C.12 Exact continuous Y-junction formulas . . . 41
C.13 Additional numerical tables . . . 41
C.14 Detailed configurations . . . 42
C.15 Why route covariance is necessary for local correction . . . 43
C.16 Continuous allocation and label-stability extensions . . . 44
D Experimental specifications and additional results . . . 47
D.1 Benchmarks and frozen protocol . . . 47
D.2 Metrics and statistical analysis . . . 48
D.3 Additional paired results . . . 49
D.4 Rank, radius, and measured direction costs . . . 49
D.5 Nested-cloud coverage and prefix allocation . . . 50
D.6 Empirical scope and sensitivity checks . . . 51
E Expanded controlled-model evidence . . . 52
E.1 Controlled learned-model benchmark and protocol . . . 52
E.2 Quality gate for the controlled learned models . . . 53
E.3 Output-grounded branches in learned denoisers . . . 53
E.4 Predictive representations and causal output directions . . . 55
E.5 Tropical variables predict controlled route changes . . . 57
E.6 Guidance and persistent singleton crystallization . . . 58
E.7 Training dynamics of quality and singleton sharpening . . . 61
E.8 Further posterior and branch-geometry derivations . . . 61
E.9 Architecture and training details . . . 62
E.10 Expanded controlled-model results . . . 63
F AI use statement . . . 64
G Reproducibility statement . . . 64
H Ethics statement . . . 65

A NOTATION AND SYMBOL GUIDE

Symbol	Meaning	Main-text location
$X \sim \Pi_0$	Clean sample and its probability law.	Section 3
$Y_q = X + \sqrt{q}Z$	Gaussian-corrupted observation at variance q.	Equation (1)
$Z \sim \mathcal{N}(0, I_d)$	Standard d-dimensional Gaussian noise.	Equation (1)
$q = \sigma^2, p_q$	Noise variance (tropical temperature) and density of Y_q.	Section 3
$m_q(y)$	Posterior denoiser $\mathbb{E}[X \mid Y_q = y]$.	Equation (1)
$s_q(y)$	Smoothed score $\nabla_y \log p_q(y)$.	Equation (1)
$\Sigma_q(y)$	Posterior covariance $\mathrm{Cov}(X \mid Y_q = y)$.	Equation (1)
$R \in [K]$	Declared latent route. K is the number of routes.	Theorem 2
π_a, Π_a, S_a	Prior weight, clean route law, and compact support of route a.	Proposition 1
$Z_{a,q}(y)$	Gaussian evidence contributed by route a.	Proposition 1
$V_{a,q}(y)$	Soft route energy $-q \log Z_{a,q}(y)$.	Proposition 1
$W_a(y, q)$	Posterior route responsibility $\mathbb{P}(R = a \mid Y_q = y)$.	Equation (2)
$m_{a,q}(y)$	Route-conditional posterior mean.	Theorem 2
$\Sigma_q^{\mathrm{within}}, \Sigma_q^{\mathrm{between}}$	Within-route and between-route terms in total posterior covariance.	Equation (4)
$\mathcal{I}_R$	Route-Fisher tensor $\sum_a W_a \nabla \log W_a \nabla \log W_a^\top$.	Equation (4)
$\gamma, Z_\gamma, I(R; Z_\gamma)$	Gaussian signal-to-noise parameter, corresponding channel output, and route mutual information.	Equation (6)
y_s	State of variance-exploding probability flow at variance s.	Equation (7)
$\eta = -\log q$	Reverse log-variance clock used to express flow strain.	Theorem 2
$J_{s \leftarrow Q}$	Flow Jacobian transporting perturbations from Q to s.	Theorem 3
$\delta_s, \widehat{\delta}_s$	Transported perturbation and its unit direction.	Equation (8)
v_t	Positive-noise probability-flow vector field.	Theorem 4
$F_{t \to 0}$	Remaining sampler map from anchor time t to the terminal sample.	Equation (9)
ϕ, h, d_ϕ	Frozen decoded representation, scalar terminal margin, and representation dimension.	Equation (9)
ℓ_t, a_t	Pulled-back terminal margin and normal $a_t = \nabla \ell_t$.	Equation (9)
B_t, r_B, u	Moving probe basis, its column dimension, and fixed probe coordinates.	Theorem 4
r, λ, p_t	Probe radius, state-dependent physical scale, and anchor-state density.	Equation (11)
$\mathcal{W}, \kappa_{\mathcal{W}}$	Probe-cloud shape and its directional width functional.	Equation (11)
$\chi_{r,\mathcal{W}}(x)$	Conditional probability that the offset cloud crosses a regular hypersurface.	Equation (11)
$\mathcal{H}^{d-c}$	Hausdorff measure on a codimension-c route stratum.	Equation (11)
$F = F_{t \to 0}$	Complete remaining learned continuation used by the controller.	Section 4
$P \in \mathbb{R}^{d \times r_P}$	Orthonormal, low-rank input probe basis.	Equation (12)
$G = DF(x)P$	Projected terminal Jacobian action.	Equation (12)
$C_P = G^\top G$	Projected Cauchy–Green matrix.	Equation (12)
$v_\star, \varsigma_1, \varsigma_2$	Leading probe direction and the first two singular values of G.	Theorem 5
$g_\rho(v), L_3$	Symmetric finite-radius terminal gain and local third-derivative bound.	Theorem 5
E_{it}^{CL}	Declared complete-link split event for trajectory i, anchor t.	Equation (15)
$\Delta_{\mathrm{out}}, \kappa$	Frozen endpoint dissimilarity and its event threshold.	Equation (15)
$u_{it,m}, M_{\mathrm{pair}}, \mathcal{Z}_{it}$	Matched state direction, signed-pair cap, and endpoint cloud.	Equation (15)
$\widehat{p}_{it}$	Learned event score for trajectory i, anchor t.	Equation (16)
A_i, U_{it}	Trajectory nonconformity score and calibrated simultaneous upper score.	Equation (16)
$\widehat{Q}_{1-\alpha}, \alpha, \tau$	Split-conformal quantile, target error level, and acceptance threshold.	Theorem 6

Continued on the next page

Symbol	Meaning	Main-text location
$\mathcal{T}_i$	Set of dependent anchors on trajectory i.	Equation (16)
$F_a(m), L_a(m)$	Unconditional event-by-prefix probability in stratum a and its simultaneous lower band.	Equation (18)
c_j, M_j, B	Per-pair cost, prefix cap, and global verification budget.	Section 4
$D_j(b)$	Optimal certified value after the first j jobs with budget b.	Equation (18)
TRC	Tropical Route Control: finite-horizon direction, simultaneous certificate, and exact prefix allocation.	Section 4

The Gaussian posterior and route-information identities remain valid after merging mixture components that define the same declared terminal cell. The finite-atom power-diagram and atom-level maximum-a-posteriori statements retain their original atomic partition. Norms are Euclidean unless noted, and Jacobians act on column vectors. Gaussian differentiation first identifies route covariance, and the variational equation transports its effect. A tube formula then converts boundary geometry into small-radius crossing probabilities. Certification and allocation operate directly on a declared finite-cloud event without requiring the mixture model at deployment.

Standing regularity. For the exact differential identities, it is enough that $\mathbb{E}\|X\|^2 < \infty$. Gaussian convolution supplies the required smoothness for positive q. Tropical limits use compact component supports and positive mixture weights. Statements at a regular wall assume unique nearest support points and a nonzero energy-difference gradient. Flow statements are made on a positive-noise interval on which the learned vector field is C^1 in state. The tube law assumes a C^2 terminal margin with nonvanishing gradient on its zero set. Every finite-sample risk statement specifies the exchangeability condition for its specified calibration and test trajectories.

B Proofs and Theoretical Extensions

The proofs follow the mechanism developed in the main text. Gaussian conditioning first identifies the covariance that controls local response. Its low-noise limit gives the decision walls, and the variational equation transports this response to the endpoint. The final step turns endpoint measurements into trajectory-level decisions under a fixed budget.

B.1 Gaussian posterior calculus

Lemma 7 (Tweedie and posterior-Jacobian identities)**.** *If $\mathbb{E}\|X\|^2 < \infty$, then for every $q > 0$,*

$$\nabla_y \log p_q(y) = \frac{m_q(y) - y}{q}, \tag{19}$$

$$D_y m_q(y) = \frac{\Sigma_q(y)}{q}. \tag{20}$$

Consequently $D_y m_q$ is positive semidefinite.

Proof. Let $\varphi_q(u) = (2\pi q)^{-d/2} \exp(-\|u\|^2/(2q))$. Gaussian convolution gives

$$p_q(y) = \int \varphi_q(y - x)\Pi_0(\mathrm{d}x), \qquad \nabla p_q(y) = \int \frac{x - y}{q}\varphi_q(y - x)\Pi_0(\mathrm{d}x).$$

Division by $p_q(y)$ proves (19). For coordinates i, j, differentiate the posterior ratio for $m_{q,i}$. The quotient rule yields

$$\partial_{y_j} m_{q,i}(y) = \frac{1}{q}\left\{\mathbb{E}[X_i X_j \mid Y_q = y] - \mathbb{E}[X_i \mid Y_q = y]\mathbb{E}[X_j \mid Y_q = y]\right\},$$

which is (20). Differentiation under the integral is justified by Gaussian domination, first for truncated X and then by L^2 convergence. □

Lemma 8 (Route scores and covariance decomposition)**.** *For each route with $W_a(y,q) > 0$,*

$$\nabla_y \log W_a(y,q) = \frac{m_{a,q}(y) - m_q(y)}{q}. \tag{21}$$

Moreover,

$$\Sigma_q = \underbrace{\sum_a W_a \Sigma_{a,q}}_{\Sigma_q^{\text{within}}} + \underbrace{\sum_a W_a (m_{a,q} - m_q)(m_{a,q} - m_q)^\top}_{\Sigma_q^{\text{between}}}. \tag{22}$$

Proof. Apply Lemma 7 to the conditional route law and subtract the unconditional score. This gives (21). The second identity is the conditional law of total covariance. □

Proof of Theorem 2. By (21),

$$q^2 \mathcal{I}_R = \sum_a W_a (m_{a,q} - m_q)(m_{a,q} - m_q)^\top = \Sigma_q^{\text{between}}.$$

For the variance-exploding probability-flow field in reverse log-variance time $\eta = -\log q$, namely $(m_q - y)/2$, Lemma 7 and (22) give

$$D_y\left(\frac{m_q - y}{2}\right) = -\frac{I}{2} + \frac{\Sigma_q^{\text{within}}}{2q} + \frac{q}{2}\mathcal{I}_R.$$

For the information identity, write $Z_\gamma = \sqrt{\gamma} X + N$, so that $\gamma = 1/q$, with N independent of (R, X). The Markov relation $R \to X \to Z_\gamma$, the conditional I–MMSE identity (Guo et al., 2005), and the chain rule $I(R; Z_\gamma) = I(X; Z_\gamma) - I(X; Z_\gamma \mid R)$ yield

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} I(R; Z_\gamma) = \frac{1}{2}\mathbb{E}\left[\operatorname{tr}\Sigma_q - \operatorname{tr}\Sigma_q^{\text{within}}\right] = \frac{q^2}{2}\mathbb{E}\operatorname{tr}\mathcal{I}_R.$$

When each route is an atom, $\Sigma_q^{\text{within}} = 0$. The field expands in unit direction u precisely when

$$u^\top D_y\left(\frac{m_q - y}{2}\right) u = -\frac{1}{2} + \frac{q}{2} u^\top \mathcal{I}_R u > 0,$$

which proves the final claim. □

Corollary 9 (Evidence budget and effective route count)**.** *Let $H_\gamma = H(R \mid Z_\gamma)$ and $N_{\text{eff}}(\gamma) = \exp H_\gamma$. Then*

$$-\frac{\mathrm{d}}{\mathrm{d}\gamma} H_\gamma = \frac{q^2}{2}\mathbb{E}\operatorname{tr}\mathcal{I}_R, \tag{23}$$

$$\log \frac{N_{\text{eff}}(\gamma_1)}{N_{\text{eff}}(\gamma_2)} = \frac{1}{2}\int_{\gamma_1}^{\gamma_2} q(\gamma)^2 \mathbb{E}\operatorname{tr}\mathcal{I}_R \, \mathrm{d}\gamma. \tag{24}$$

Proof. Because $H(R)$ is constant, $-\mathrm{d}H(R \mid Z_\gamma)/\mathrm{d}\gamma = \mathrm{d}I(R; Z_\gamma)/\mathrm{d}\gamma$. Apply Theorem 2 and integrate. □

Interpretation. Corollary 9 gives an accumulated evidence identity. Its integrated trace is the logarithmic reduction in the effective number of plausible routes. A narrow region of large sensitivity can therefore carry a finite amount of evidence. Whether that sensitivity separates terminal outputs depends on its transport through the remaining denoising trajectory.

B.2 SOFT TROPICAL CELLS AND CRYSTALLIZATION

For a component support $S_a = \operatorname{supp} \Pi_a$, define

$$Z_{a,q}(y) = \pi_a \int_{S_a} \exp\left(-\frac{\|y-x\|^2}{2q}\right) \Pi_a(\mathrm{d}x), \qquad V_{a,q}(y) = -q \log Z_{a,q}(y).$$

Lemma 10 (Uniform compact-support Laplace principle). *Assume $\pi_a > 0$ and that Π_a has nonempty compact support S_a. For every compact set $K_0 \subset \mathbb{R}^d$,*

$$\sup_{y \in K_0} \left| V_{a,q}(y) - \frac{1}{2} \operatorname{dist}(y, S_a)^2 \right| \longrightarrow 0 \quad \text{as } q \downarrow 0.$$

Proof. The upper exponential bound follows from $\|y-x\| \geq \operatorname{dist}(y, S_a)$:

$$Z_{a,q}(y) \leq \pi_a \exp\left[-\operatorname{dist}(y, S_a)^2/(2q)\right].$$

For the reverse bound, choose a nearest support point x_y. Every ball $B(x_y, \varepsilon)$ has positive Π_a-mass. Compactness of K_0 and S_a gives a finite subcover and a positive minimum mass after allowing radius 2ε. On that ball, squared distance exceeds the minimum by at most a quantity $c(\varepsilon) \to 0$. Taking $-q \log$, then $q \downarrow 0$ and $\varepsilon \downarrow 0$, proves uniform convergence. □

Proof of Proposition 1. The exact soft assignment is Bayes' rule. Its limiting energy follows from Lemma 10. At the stated regular tie, the posterior within each route concentrates at its unique nearest point:

$$m_{a,q} \to x_a^\star, \qquad \Sigma_{a,q} \to 0.$$

The route weights converge to ω and $1-\omega$. By Lemmas 7 and 8,

$$qD_y m_q = \Sigma_q = \Sigma_q^{\text{within}} + \Sigma_q^{\text{between}} \longrightarrow \omega(1-\omega)(x_a^\star - x_b^\star)(x_a^\star - x_b^\star)^\top.$$

This is the claimed rank-one susceptibility. □

Proposition 11 (Universal local profile around the moving exact tie). *Let $q \downarrow 0$ and let exact two-route tie points $y_q^0 \to y_0$ satisfy $W_a(y_q^0, q) = W_b(y_q^0, q)$. For every fixed $L < \infty$, assume uniformly on $y_q^0 + qun$, $|u| \leq L$, that the total weight of all other routes is $o(1)$, $m_{a,q} - m_{b,q} \to d \neq 0$, and the sum of the two route-conditional covariances tends to zero. Also assume that the omitted routes have vanishing weighted second moment relative to $m_{b,q}$:*

$$\sum_{c \notin \{a,b\}} W_c \left\{ \operatorname{tr} \Sigma_{c,q} + \|m_{c,q} - m_{b,q}\|^2 \right\} = o(1)$$

uniformly on the same intervals. Compact component supports imply this condition. Put $n = d/\|d\|$ and $\Lambda_{\log}(z) = (1+e^{-z})^{-1}$. Then, locally uniformly in u,

$$W_a(y_q^0 + qun, q) \longrightarrow \Lambda_{\log}(\|d\|u), \tag{25}$$

$$q\, n^\top D m_q(y_q^0 + qun) n \longrightarrow \|d\|^2 \Lambda_{\log}(\|d\|u)\{1 - \Lambda_{\log}(\|d\|u)\}. \tag{26}$$

Consequently, the 10–90% width is $2q \log 9/\|d\| + o(q)$, the balanced peak is $\|d\|^2/(4q) + o(q^{-1})$, and the conserved response is the sequential local limit

$$\lim_{L \to \infty} \lim_{q \downarrow 0} \int_{-Lq}^{Lq} n^\top D m_q(y_q^0 + zn) n \, \mathrm{d}z = \|d\|. \tag{27}$$

Proof. The exact route-score identity is $\nabla_y \log(W_a/W_b) = (m_{a,q} - m_{b,q})/q$. Integrating it from the exact tie along $y_q^0 + qun$ gives log odds $\|d\|u + o(1)$, uniformly on every fixed compact u-interval. The vanishing third-route weight gives (25). The exact identity $qDm_q = \Sigma_q$, followed by the within/between covariance decomposition and the weighted second-moment condition, gives (26) without differentiating an asymptotic remainder. Integrating the limiting logistic density on $[-L, L]$, then sending $L \to \infty$, proves (27). The two quantiles and the peak follow directly. □

Corollary 12 (Inverse-log crystallization law)**.** *Under the hypotheses of Proposition 11, fix an ambiguity level $\eta \in (0, \log 2)$ and let $\mathcal{A}_\eta(q)$ be the set near y_q^0 where the entropy of the normalized pair $(W_a, W_b)/(W_a + W_b)$ is at least η. There is a constant $c_\eta > 0$, determined by binary entropy, such that its normal thickness is*

$$\operatorname{thick} \mathcal{A}_\eta(q) = \frac{2c_\eta q}{\|d\|} + o(q).$$

Equivalently, resolving a point at normal distance r requires $q = \Theta(r)$ at a fixed entropy level. Separately, for the binary posterior normalized over routes a, b, an exact favorable energy gap $g_q = V_{b,q} - V_{a,q} > 0$ gives error $(1 + \exp(g_q/q))^{-1}$. For $0 < \varepsilon < 1/2$, this error is at most ε exactly when $g_q/q \geq \log((1-\varepsilon)/\varepsilon)$. For a fixed gap, the admissible variance thus decreases as the inverse logarithm of the requested error tolerance.

Proof. Binary entropy is a symmetric strictly decreasing function of $|\log(W_a/W_b)|$. Let c_η be the unique positive log-odds at which entropy equals η, and substitute (25). □

Proposition 13 (Rank and codimension at a multi-route junction)**.** *For every $q > 0$, $\operatorname{rank} \Sigma_q^{\text{between}} \leq K - 1$. Let $A \subset [K]$ have $|A| = k$, assume $\sum_{a \notin A} W_a = o(1)$, and assume all route means are uniformly bounded. With $\widetilde{W}_a = W_a / \sum_{b \in A} W_b$, $\widetilde{m}_A = \sum_{a \in A} \widetilde{W}_a m_{a,q}$, and*

$$\Sigma_{q,A}^{\text{between}} = \sum_{a \in A} \widetilde{W}_a (m_{a,q} - \widetilde{m}_A)(m_{a,q} - \widetilde{m}_A)^\top,$$

we have $\Sigma_q^{\text{between}} = \Sigma_{q,A}^{\text{between}} + o_{\text{op}}(1)$. Thus its limiting matrix has rank at most $k-1$. If the $k-1$ limiting energy differences are transverse, the junction has codimension $k-1$, so a regular k-route junction has at most $k-1$ nonvanishing route-sensitive contrast directions. Expansion additionally requires an eigenvalue of Σ_q to exceed q.

Proof. For all routes, the weighted contrast vectors sum to zero, proving the finite-q rank bound. Within A, the normalized contrasts also sum to zero and give rank at most $k-1$. The omitted mass and the resulting mean shift contribute $o_{\text{op}}(1)$ by uniform boundedness. Transversality of the $k-1$ independent energy differences gives the codimension statement by the implicit function theorem. The expansion condition follows from the reverse-variance variational equation. □

B.3 Time-ordered dynamics and transported route cells

The local covariance determines the instantaneous deformation. To obtain a terminal displacement, we must integrate this deformation along the evolving perturbation itself. The next proof separates the isotropic contraction from that accumulated directional response.

Proof of Theorem 3. Along the variance-exploding flow $\dot{y}_s = (y_s - m_s(y_s))/(2s)$, the variational equation is

$$\frac{\mathrm{d}}{\mathrm{d}s}\delta_s = \left(\frac{I}{2s} - \frac{\Sigma_s(y_s)}{2s^2}\right)\delta_s.$$

Therefore

$$\frac{\mathrm{d}}{\mathrm{d}s}\log\|\delta_s\| = \frac{1}{2s} - \frac{\widehat{\delta}_s^\top \Sigma_s \widehat{\delta}_s}{2s^2}.$$

Integrating from Q down to q reverses the covariance integral and gives

$$\log\frac{\|\delta_q\|}{\|\delta_Q\|} = \frac{1}{2}\log\frac{q}{Q} + \frac{1}{2}\int_q^Q \frac{\widehat{\delta}_s^\top \Sigma_s \widehat{\delta}_s}{s^2}\,\mathrm{d}s.$$

□

Proposition 14 (Time ordering is essential)**.** *Let $A_s = I/(2s) - \Sigma_s/(2s^2)$ along a trajectory. The flow Jacobian is*

$$J_{q \leftarrow Q} = \mathcal{T}\exp\left(\int_Q^q A_s\,\mathrm{d}s\right),$$

where $\mathcal{T}$ denotes time ordering. Replacing this expression by $\exp(\int_Q^q A_s\,\mathrm{d}s)$ is exact if the A_s commute pairwise, but in general the Magnus logarithm has the expansion

$$J_{q\leftarrow Q} = \exp(\Omega_1 + \Omega_2 + \cdots), \qquad \Omega_1 = \int_Q^q A_s\,\mathrm{d}s,$$
$$\Omega_2 = -\frac{1}{2}\int_q^Q \int_q^{s_1} [A_{s_1}, A_{s_2}]\,\mathrm{d}s_2\,\mathrm{d}s_1.$$

The convergent expansion is valid, for example, when $\int_q^Q \|A_s\|_{\mathrm{op}}\,\mathrm{d}s < \pi$ (Moan & Niesen, 2008). Outside this regime these terms describe the local small-interval expansion. The commutator is a correction to the exponent, not an additive Jacobian error. Since A_s is symmetric, Ω_2 is skew-symmetric and must be combined with the other terms before interpreting singular gains. Thus rotating covariance eigenspaces can change the terminal leading direction even when the instantaneous spectra are unchanged.

Proof. The first statement is the fundamental solution of a nonautonomous linear system. The commutator term is the second term of the Magnus expansion. □

Proof of Theorem 4. By construction, $\ell_t(x) = \ell_0(F_{t\to 0}(x))$. The flow composition identity $F_{t+\Delta t\to 0} \circ F_{t\to t+\Delta t} = F_{t\to 0}$ implies, after differentiation at $\Delta t = 0$,

$$\partial_t \ell_t(x) + v_t(x)^\top \nabla \ell_t(x) = 0.$$

For a central trajectory x_t, a continuously differentiable $B_t \in \mathbb{R}^{d\times r_B}$, and fixed $u \in \mathbb{R}^{r_B}$, set $y_t = x_t + B_t u$. Here r_B is the probe dimension. The chain rule gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\ell_t(y_t) = \partial_t \ell_t(y_t) + \nabla \ell_t(y_t)^\top \{v_t(x_t) + \dot{B}_t u\},$$

Substitution of the transport equation at y_t proves

$$\frac{\mathrm{d}}{\mathrm{d}t}\ell_t(y_t) = \nabla \ell_t(y_t)^\top \{v_t(x_t) + \dot{B}_t u - v_t(y_t)\}. \tag{28}$$

The diffeomorphism maps each route cell bijectively, so its probability under the transported law is constant. □

Lemma 15 (Finite-cloud tube formula)**.** *Let $X \in \mathbb{R}^d$ have density p. Let s be continuous on $\mathbb{R}^d$ and C^2 on a uniform tubular neighborhood of a compact, positive-reach, embedded hypersurface $\Gamma = \{s = 0\}$, assume $\inf_\Gamma \|\nabla s\| > 0$, and suppose p is continuous on that neighborhood. Let $\lambda : \mathbb{R}^d \to [0,\infty)$ be continuous and bounded, and let the finite offset cloud $\mathcal{W} = \{w_1, \ldots, w_M\}$ be bounded. Define*

$$E_r(x) = \mathbf{1}\{\min_j s(x + r\lambda(x)w_j) \le 0 \le \max_j s(x + r\lambda(x)w_j)\}.$$

Then, as $r \downarrow 0$,

$$\mathbb{P}(E_r(X) = 1) = r\int_\Gamma p(x)\lambda(x)\left[\max_j n(x)^\top w_j - \min_j n(x)^\top w_j\right]\mathrm{d}\mathcal{H}^{d-1}(x) + o(r), \tag{29}$$

where $n = \nabla s/\|\nabla s\|$. The same formula holds for a random bounded cloud independent of X after taking expectation of the bracketed width.

Proof. In a tubular neighborhood, write $x = z + \ell n(z)$ with $z \in \Gamma$. For $|\ell| = O(r)$, Taylor expansion and uniform continuity of λ give

$$s(x + r\lambda w_j) = \|\nabla s(z)\|\{\ell + r\lambda(z)n(z)^\top w_j\} + o(r).$$

The cloud straddles the boundary when ℓ lies between the negative largest and negative smallest projected offsets. That interval has the width shown in (29). The tubular-coordinate Jacobian, p, and λ may be replaced by their values at z with $o(r)$ error. Integrating over z proves the formula. □

Derivation of Equation (11). Apply Lemma 15 to the transported margin ℓ_t and absorb the cloud width into $\kappa_{\mathcal{W}}(n) = \max_j n^\top w_j - \min_j n^\top w_j$, or its expectation for a random cloud. □

Proposition 16 (Stratified small-radius law)**.** *Let $k \geq 2$, put $\nu = k - 1$, and let X have a continuous density near a compact positive-reach C^2 codimension-ν stratum where exactly k cell germs meet transversely. Let $\lambda \geq 0$ be continuous and bounded, and let a bounded finite offset cloud be independent of X. If $E_r^{(k)}$ is the event that $\{X + r\lambda(X)W_j\}_j$ meets all k germs within the declared tubular neighborhood, then*

$$\mathbb{P}(E_r^{(k)}) = C_{k,\mathcal{W}} r^{k-1} + o(r^{k-1}),$$

provided the normal-fan translation set has boundary of zero ν-dimensional Lebesgue measure. The coefficient $C_{k,\mathcal{W}} \geq 0$ integrates the projected ν-dimensional translation volume over the stratum. It can vanish when the offsets do not cover its normal fan. Only $0 < C_{k,\mathcal{W}} < \infty$ makes r^{k-1} the leading order.

Proof. Use the $k - 1$ independent energy differences as normal coordinates. The implicit function theorem identifies a tubular neighborhood with the junction times $\mathbb{R}^{k-1}$. Rescaling the normal coordinate by r produces the factor r^{k-1}. The zero-boundary condition and dominated convergence leave the projected cloud-volume coefficient. Failure to cover the normal fan makes that coefficient zero, explaining why additional directions improve coverage but cannot change codimension. □

B.4 Finite-horizon action and finite-radius terminal gain

Let F be the remaining deterministic continuation and let P have orthonormal columns. Write $G = DF(x)P$ and $C_P = G^\top G$. For any unit v,

$$\|DF(x)Pv\|^2 = v^\top C_P v.$$

The leading eigenvector of C_P is therefore the exact maximizer inside the declared probe subspace. It requires only Jacobian–vector products with the columns of P.

Proof of Theorem 5. The symmetric Taylor expansion in direction Pv gives

$$\frac{F(x + \rho Pv) - F(x - \rho Pv)}{2\rho} = DF(x)Pv + e_\rho(v), \qquad \|e_\rho(v)\| \leq L_3\rho^2/6.$$

For $v_\star$, the reverse triangle inequality yields $g_\rho(v_\star) \geq \varsigma_1 - L_3\rho^2/6$. Expand a comparator as $v = \cos\theta\, v_\star + \sin\theta\, w$, $w \perp v_\star$. The variational characterization of singular values gives

$$\|Gv\| \leq \sqrt{\varsigma_1^2 \cos^2\theta + \varsigma_2^2 \sin^2\theta}.$$

The triangle inequality adds $L_3\rho^2/6$, proving both bounds. □

Corollary 17 (Observable finite-radius advantage)**.** *The leading direction is guaranteed to beat a comparator at angle θ whenever*

$$\varsigma_1 - \sqrt{\varsigma_1^2 \cos^2\theta + \varsigma_2^2 \sin^2\theta} > \frac{L_3\rho^2}{3}.$$

The left side is the time-ordered spectral–angle margin. The right side is the finite-radius nonlinearity penalty.

Proposition 18 (Stability of the projected direction)**.** *Let $\widehat{C}_P = C_P + E$ with $E = E^\top$, $\|E\|_{\mathrm{op}} \leq \varepsilon$, and let $\lambda_1 \geq \lambda_2$ be the two largest eigenvalues of C_P, with $\lambda_1 = \varsigma_1^2$ and gap $\Delta = \lambda_1 - \lambda_2 > 0$. If $\varepsilon < \Delta/2$, then the leading unit eigenvectors satisfy*

$$\sin\angle(\widehat{v}_\star, v_\star) \leq \frac{\varepsilon}{\Delta - \varepsilon}.$$

Combining this angle with Theorem 5 gives a direct numerical-accuracy-to-terminal-gain lower bound:

$$g_\rho(\widehat{v}_\star) \geq \max\left\{\varsigma_1\sqrt{1 - \left(\frac{\varepsilon}{\Delta - \varepsilon}\right)^2}, \sqrt{(\varsigma_1^2 - 2\varepsilon)_+}\right\} - \frac{L_3\rho^2}{6},$$

where $(a)_+ = \max\{a, 0\}$.

Proof. This is the one-dimensional Davis–Kahan bound (Davis & Kahan, 1970) applied to the symmetric projected actions. For any unit v, positivity of C_P gives $\|Gv\| \geq \varsigma_1 |v^\top v_\star|$. Also, the variational principle gives $\widehat{v}_\star^\top C_P \widehat{v}_\star \geq \lambda_1 - 2\varepsilon$. The symmetric Taylor remainder subtracts at most $L_3\rho^2/6$ from each linear lower bound. An upper bound on comparator gain alone would not justify either performance guarantee. □

B.4.1 A FINITE-RADIUS CERTIFICATE FOR THE COMPLETE ENDPOINT CLOUD

The preceding result controls one antipodal pair. A dispersion test compares *every* pair, including endpoints generated by different directions. The next result closes this distinction for smooth feature maps without identifying dispersion with classifier-cell crossing. It applies at a fixed positive noise level and makes no compact-support assumption on the prior.

Theorem 19 (Finite-cloud diameter certificate)**.** *Let $r, p, M \geq 1$ be integers and let $H : \mathbb{R}^r \to \mathbb{R}^p$ be C^3 near the closed ball of radius $\rho > 0$ about zero. Write $G = DH(0)$, assume full column rank (hence $p \geq r$), put $s = \inf_{\|u\|=1} \|Gu\| > 0$, and suppose $\|D^2H(0)[a,b]\| \leq L_2\|a\|\|b\|$ and $\|D^3H(z)[u,u,u]\| \leq L_3$ for $\|z\| \leq \rho$, $\|u\| = 1$, with $L_2, L_3 \geq 0$. For any unit vectors $u_1, \ldots, u_M$, define*

$$\mathcal{Z}_\rho = \{H(0), H(\rho u_j), H(-\rho u_j) : 1 \leq j \leq M\},$$
$$D_\rho = \max_{z,z' \in \mathcal{Z}_\rho} \|z - z'\|, \qquad A = \max_j \|Gu_j\|.$$

Set $e = L_3\rho^2/6$, $c = \sqrt{1 + (\rho L_2/s)^2}$, and $b = \rho L_2/4 + L_3\rho^2/12$. Then

$$A - e \;\leq\; \frac{D_\rho}{2\rho} \;\leq\; \max\{cA + e, A/2 + b\}. \tag{30}$$

If $\rho L_2/2 + \rho^2 L_3/6 \leq s$, the upper bound simplifies to $cA + e$. Thus, in this regime, the full-cloud diameter is $2\rho A + O(\rho^3)$, uniformly in the number and choice of directions for fixed G, L_2, L_3.

Proof. Taylor expansion gives $H(\rho u) = H(0) + \rho Gu + \rho^2 D^2H(0)[u,u]/2 + R_u$, with $\|R_u\| \leq L_3\rho^3/6$. Choosing the antipodal pair that attains A cancels the quadratic terms and proves the lower bound. For any two signed unit directions u, v, put $a = u + v$ and $d = u - v$. The parallelogram law and the smallest singular value give

$$\|Gd\|^2 + \|Ga\|^2 = 2\|Gu\|^2 + 2\|Gv\|^2 \leq 4A^2, \qquad \|Ga\| \geq s\|a\|.$$

The quadratic difference is $D^2H(0)[a,d]$, while $\|d\| \leq 2$. Consequently Cauchy–Schwarz applied to the resulting ellipse yields

$$\begin{aligned}\|H(\rho u) - H(\rho v)\| &\leq \rho\sqrt{4A^2 - s^2\|a\|^2} + \rho^2 L_2\|a\| + L_3\rho^3/3 \\ &\leq 2\rho A\sqrt{1 + (\rho L_2/s)^2} + L_3\rho^3/3.\end{aligned}$$

A base-to-endpoint distance is at most $\rho A + \rho^2 L_2/2 + L_3\rho^3/6$, giving the other upper-bound term. Under the additional condition, this is at most $\rho(A + s) \leq 2\rho A$, so it cannot exceed the stated simplified upper bound. □

For a continuation map F, feature map ϕ, fixed state x, and matrix P with orthonormal columns, take $H(z) = \phi(F(x + Pz))$. The matrix $G = D(\phi \circ F)(x)P$ then includes the remaining flow, decoding, and the actual feature representation. A latent-only derivative is not a certificate for a decoded-feature event. The theorem is deterministic. Its application to a deployed cloud requires the derivative and nondegeneracy assumptions on that cloud's actual span.

Finite-radius check on archived endpoint clouds. We examined every archived M12 cloud at relative radii $0.025, 0.05, 0.075, 0.10, 0.15$: 192 DiT groups and 64 SDXL groups at three anchors, comprising 3,840 clouds with matched directions across radii. This is a posthoc diagnostic, not a confirmatory test. We compared the all-pair diameter D_{all}, including the base endpoint, with $D_{\text{pair}} = \max_j \|H(\rho u_j) - H(-\rho u_j)\|$, separately in normalized CLIP Euclidean distance and pixel RMSE. Under the theorem's assumptions, $D_{\text{all}} - D_{\text{pair}} = O(\rho^3)$ and $D_{\text{all}}/D_{\text{pair}} - 1 = O(\rho^2)$. All 30 pixel model–anchor–radius cells had zero median excess. Only three CLIP model–anchor curves had positive medians at all five radii. Their all-radius log–log slopes were $0.975, 0.899, 1.518$ for the

absolute excess and $0.634, 0.482, 0.838$ for the relative excess. Fits involving a zero median were left undefined, without adding a numerical floor. These observations do not verify the cubic law: the measured window has no certified derivative or nondegeneracy bounds, and image clamping can violate the smoothness assumption. They show why antipodal-pair measurements alone cannot be presumed to characterize feature-cloud dispersion at deployment radii.

Conditional detection law and multiple feature views. Suppose now that u_j are independent draws of a unit vector U, conditional on the state and all feature maps, the subspace, and the direction distribution being fixed before these draws. For a threshold $\tau > 0$, write $h = \tau/(2\rho)$, $t^+ = h + e$, $t^- = \min\{(h-e)/c, 2(h-b)\}$, and let $\mathcal{F}_G(t) = \mathbb{P}\{\|GU\| \le t\}$ for all real t. Then

$$1 - \mathcal{F}_G(t^+)^M \;\le\; \mathbb{P}\{D_\rho > \tau\} \;\le\; 1 - \mathcal{F}_G(t^-)^M. \tag{31}$$

Under the additional radius condition in the theorem, use $t^- = (h-e)/c$. Indeed, $A > t^+$ implies detection, whereas detection implies $A > t^-$. Independence gives $\mathbb{P}(A \le t) = \mathcal{F}_G(t)^M$. If H is affine, taking $L_2 = L_3 = 0$ makes both bounds agree exactly. For uniform spherical U, the squared gain has the law $\sum_{k=1}^r \lambda_k Z_k^2 / \sum_{k=1}^r Z_k^2$, where Z_k are independent standard normals and λ_k are the eigenvalues of $G^\top G$. Thus the finite-cloud law depends on the transported spectrum, not on an instantaneous eigenvalue or a tube codimension alone.

For finitely many feature maps H_ℓ satisfying the theorem, define $\mathcal{F}(t_1, \dots, t_L) = \mathbb{P}\{\|G_\ell U\| \le t_\ell \text{ for every } \ell\}$. Applying the preceding implications view by view gives, for $E = \{D_{\rho,\ell} > \tau_\ell \text{ for some } \ell\}$,

$$1 - \mathcal{F}(t_1^+, \dots, t_L^+)^M \le \mathbb{P}(E) \le 1 - \mathcal{F}(t_1^-, \dots, t_L^-)^M.$$

This joint distribution retains dependence between feature views. In particular, normalized CLIP cosine distance is half the squared Euclidean distance between normalized embeddings. The declared joint threshold $\max\{d_{\mathrm{CLIP}}/0.06, d_{\mathrm{pixel}}/0.12\} > \kappa$ corresponds to $\tau_{\mathrm{CLIP}} = \sqrt{0.12\kappa}$ and $\tau_{\mathrm{pixel}} = 0.12\kappa$, using pixels divided by the square root of their number for the latter feature map. These probability formulas concern fresh independent directions in a fixed subspace. They do not automatically describe an adaptively selected direction bank or a full-dimensional deployment cloud. Their finite-radius constants also require bounds, not unverified empirical substitutes.

Why nondegeneracy matters. Let $H(z_1, z_2) = z_1 + z_1 z_2$ and choose $u = (a, b)$, $v = (-a, b)$, where $a, b > 0$ and $a^2 + b^2 = 1$. Here $G = (1, 0)$ has a null direction, $L_3 = 0$, and $A = a$. Each antipodal image pair has distance $2\rho a$, but, for sufficiently small positive ρ, the full-cloud diameter is $2\rho a + 2\rho^2 ab$. Cross-direction pairs therefore produce a genuine quadratic error when nondegeneracy is dropped. Antipodal Taylor cancellation alone does not prove the full-cloud cubic-error guarantee.

B.5 Trajectory-level certification and exact allocation

Proof of Theorem 6. For calibration trajectory i, define

$$A_i = \max_{t \in \mathcal{T}_i} (E_{it}^{\mathrm{CL}} - \widehat{p}_{it}).$$

Conditional on the independent training data and all frozen objects listed in Theorem 6, the same measurable maximum is applied to each exchangeable trajectory array. Thus $A_1, \dots, A_n, A_{n+1}$ are exchangeable. Let $k = \lceil (n+1)(1-\alpha) \rceil$ and let $\widehat{Q}_{1-\alpha}$ be the k-th order statistic of the calibration scores, with $+\infty$ padding when needed. The exchangeable-rank argument gives $\mathbb{P}(A_{n+1} \le \widehat{Q}_{1-\alpha}) \ge 1 - \alpha$. On this event, simultaneously for every test anchor,

$$E_{n+1,t}^{\mathrm{CL}} \le \mathrm{clip}(\widehat{p}_{n+1,t} + \widehat{Q}_{1-\alpha}, 0, 1) = U_{n+1,t},$$

because for $E \in \{0, 1\}$, the inequality $E \le r$ implies $E \le \mathrm{clip}(r, 0, 1)$: the case $E = 1$ forces $r \ge 1$, and the case $E = 0$ is immediate. If $U_{n+1,t} \le \tau < 1$, the event $E_{n+1,t}^{\mathrm{CL}} = 1$ would violate the simultaneous bound. Hence the probability of any false acceptance is at most α. □

Lemma 20 (Why anchorwise calibration is insufficient)**.** *If each of T dependent anchors is calibrated separately at marginal error α, the probability of at least one error can be as large as $\min(1, T\alpha)$. Calibrating the trajectory maximum A_i avoids this multiplicity without assuming independence among anchors.*

Proof. The upper statement follows from the union bound. Disjoint error events attain it when $T\alpha \leq 1$. When $T\alpha \geq 1$, events of probability at most α can cover the whole sample space, attaining one. The second statement follows from Theorem 6, applied at trajectory level. □

Theorem 21 (Exact prefix allocation)**.** *For jobs $j = 1, \ldots, J$, choose prefix depth $m_j \in \{0, \ldots, M_j\}$, cost $c_j m_j$, and supplied event-discovery value $L_{a_j}(m_j)$. The recursion*

$$D_j(b) = \max_{0 \leq m \leq \min(M_j, \lfloor b/c_j \rfloor)} \{D_{j-1}(b - c_j m) + L_{a_j}(m)\}, \qquad D_0(b) = 0$$

returns the globally optimal feasible allocation under $\sum_j c_j m_j \leq B$.

Proof. Condition on the final depth m_j assigned to job j. The remaining jobs receive budget $b - c_j m_j$, whose optimal value is $D_{j-1}(b - c_j m_j)$ by induction. Maximizing over every feasible m_j therefore considers every feasible allocation exactly through its last coordinate. The base case is immediate. □

Proposition 22 (Validity after prefix search)**.** *Suppose $L_a(m)$ is a simultaneous lower confidence band satisfying*

$$\mathbb{P}\{\forall a, m : L_a(m) \leq F_a(m)\} \geq 1 - \delta,$$

where $F_a(m)$ is the unconditional probability that a job in stratum a has the declared event and is detected by prefix m, in the calibration population after the frozen unresolved-job filter. Let $\mathcal{G}$ be the preallocation sigma-field containing the queue, features, costs, and bands, and let $Z_j(m)$ indicate discovery of job j's event by prefix m. Assume, for every queued job and prefix, $\mathbb{E}[Z_j(m) \mid \mathcal{G}] \geq F_{a_j}(m)$. Then any $\mathcal{G}$-measurable selected allocation, including the optimizer, satisfies

$$\mathbb{P}\left\{ \mathbb{E}\left[\sum_j Z_j(m_j) \mid \mathcal{G} \right] \geq \sum_j L_{a_j}(m_j) \right\} \geq 1 - \delta.$$

Proof. Conditional linearity and the domination assumption give $\mathbb{E}[\sum_j Z_j(m_j) \mid \mathcal{G}] \geq \sum_j F_{a_j}(m_j)$. On the simultaneous-band event the latter is at least $\sum_j L_{a_j}(m_j)$, including after prefix search. No independence between jobs or additional selection union bound is needed. □

Composition of the guarantees. The conformal theorem controls false acceptance for the declared complete-link event under trajectory-array exchangeability. The allocation theorem exactly optimizes supplied values under the cost cap. Their statistical interpretation additionally requires Proposition 22. Band calibration must use the same frozen unresolved-job filter, and the strata must support the stated conditional domination. Unconditional stratum averages alone do not certify a feature-selected queue. Exact knapsack optimization therefore does not by itself certify empirical priority scores or their reported gains.

B.6 Additional structural consequences from the full theory

The preceding results describe local response and its terminal effect. We now connect them to four broader questions: when posterior uncertainty controls a route label, how the variance clock changes under VP and DDIM, what the low-noise Jacobian identifies, and how information acquisition bounds finite-horizon amplification across a fixed positive-noise interval.

B.6.1 Posterior variance controls route decodability

Let $(\mathcal{Z}, d_{\mathcal{Z}})$ be a separable metric space, $\phi : \mathbb{R}^d \to \mathcal{Z}$ be L-Lipschitz with $L > 0$, and $\{C_1, \ldots, C_K\}$ be a finite Borel partition with a measurable tie rule $c : \mathcal{Z} \to [K]$. For $m = m_q(y)$, define the complement margin

$$\mu_q(y) = d_{\mathcal{Z}}\big(\phi(m), \mathcal{Z} \setminus C_{c(\phi(m))}\big).$$

Theorem 23 (Posterior-MMSE route bound). *If $\mathbb{E}\|X\|^2 < \infty$, then whenever $\mu_q(y) > 0$,*

$$\mathbb{P}\{c(\phi(X)) \neq c(\phi(m_q(y))) \mid Y_q = y\} \leq \min\left\{1, \frac{L^2 \operatorname{tr} \Sigma_q(y)}{\mu_q(y)^2}\right\}. \tag{32}$$

Consequently, with $\operatorname{mmse}(q) = \mathbb{E} \operatorname{tr} \Sigma_q(Y_q)$, for every $\delta > 0$,

$$\mathbb{P}\{c(\phi(X)) \neq c(\phi(m_q(Y_q)))\} \leq \mathbb{P}\{\mu_q(Y_q) \leq \delta\} + \frac{L^2 \operatorname{mmse}(q)}{\delta^2}. \tag{33}$$

Proof. A route change places $\phi(X)$ outside the cell containing $\phi(m)$, so $d_{\mathcal{Z}}(\phi(X), \phi(m)) \geq \mu_q(y)$. Lipschitz continuity implies $\|X - m\| \geq \mu_q(y)/L$. Conditional Markov inequality and $\mathbb{E}[\|X - m\|^2 \mid Y_q = y] = \operatorname{tr} \Sigma_q(y)$ prove (32). Split the population event according to $\mu_q(Y_q) \leq \delta$, apply the pointwise bound on its complement, and average. □

Corollary 24 (Polynomial-margin rate). *Suppose $\mathbb{P}\{\mu_q(Y_q) \leq \delta\} \leq A\delta^\beta$ for $0 < \delta \leq \delta_0$, where $A, \beta, \delta_0 > 0$. Put $B = L^2 \operatorname{mmse}(q)$. If $B \leq (A\beta/2)\delta_0^{\beta+2}$, then*

$$\mathbb{P}\{c(\phi(X)) \neq c(\phi(m_q(Y_q)))\} \leq C_\beta A^{2/(\beta+2)} B^{\beta/(\beta+2)}, \quad C_\beta = \frac{\beta+2}{2}\left(\frac{2}{\beta}\right)^{\beta/(\beta+2)}. \tag{34}$$

Otherwise the right side of (33) at δ_0 remains valid.

Proof. Minimize $A\delta^\beta + B/\delta^2$. Its unconstrained minimizer is $\delta_\star = (2B/(A\beta))^{1/(\beta+2)}$. Substitution gives (34). The displayed condition is exactly $\delta_\star \leq \delta_0$. □

The scope is important. This is a forward-posterior statement, and therefore also applies to an exact conditional reverse-SDE draw. It does not by itself control a deterministic probability-flow or DDIM endpoint: equality of output marginals is not a pointwise posterior coupling.

B.6.2 VP NORMALIZATION AND THE DETERMINISTIC-DDIM TANGENT BRIDGE

Theorem 25 (Canonical variance and DDIM tangents). *Let $Y_t = \alpha_t X + \sigma_t Z$, with positive C^1 functions α_t, σ_t, and set*

$$u_t = Y_t/\alpha_t, \qquad \nu_t = \sigma_t/\alpha_t, \qquad q_t = \nu_t^2.$$

Assume $\dot{q}_t > 0$ in the forward noising clock and use the standard scalar linear forward SDE with drift coefficient $f_t = \dot{\alpha}_t/\alpha_t$ and diffusion coefficient $g_t^2 = \alpha_t^2 \dot{q}_t$. If $\Psi_{q_s \leftarrow q_t}$ denotes the canonical VE probability-flow map, then the corresponding VP probability-flow map is its scalar conjugate,

$$\Phi^Y_{s\leftarrow t}(y) = \alpha_s \Psi_{q_s \leftarrow q_t}(y/\alpha_t), \qquad D\Phi^Y_{s\leftarrow t}(y) = \frac{\alpha_s}{\alpha_t} D\Psi_{q_s \leftarrow q_t}(y/\alpha_t). \tag{35}$$

Hence raw-coordinate log singular gains equal canonical gains plus $\log(\alpha_s/\alpha_t)$.

For one deterministic DDIM step, let $\beta_{s,t} = \nu_s/\nu_t \in [0, 1)$, let $\widehat{m}_t(u)$ be a differentiable clean predictor, and put $A_t = D\widehat{m}_t(u_t)$. Then

$$u_s = \beta_{s,t} u_t + (1 - \beta_{s,t})\widehat{m}_t(u_t), \qquad B_{s\leftarrow t} = \beta_{s,t} I + (1 - \beta_{s,t}) A_t. \tag{36}$$

For a general learned A_t, Euclidean expansion in a unit direction v means $v^\top B_{s\leftarrow t}^\top B_{s\leftarrow t} v > 1$. An expanding direction exists exactly when the largest singular value of $B_{s\leftarrow t}$ exceeds one. If A_t is symmetric, an eigenvalue a_j gives singular gain $|\beta_{s,t} + (1 - \beta_{s,t})a_j|$, so a large negative eigenvalue can also cause expansion. For the exact posterior predictor, $A_t = \Sigma_{q_t}/q_t \succeq 0$. Only in this case the threshold reduces to $\lambda_j(\Sigma_{q_t}) > q_t$, matching the continuous flow.

Proof. Because $Y_t/\alpha_t = X + \nu_t Z = X + \sqrt{q_t} Z$, normalized VP marginals are exactly the canonical Gaussian channel. The specified SDE has probability-flow velocity $f_t y - g_t^2 \nabla_y \log p_t^Y(y)/2$. Using $\nabla_y \log p_t^Y(y) = \alpha_t^{-1} \nabla_u \log p_{q_t}(u)$ therefore gives

$$\frac{\mathrm{d}u_t}{\mathrm{d}t} = -\frac{\dot{q}_t}{2} \nabla_u \log p_{q_t}(u_t) = \frac{\dot{q}_t}{2q_t}\{u_t - m_{q_t}(u_t)\}.$$

Changing from t to q_t proves (35). The deterministic DDIM update is $Y_s = \alpha_s \widehat{m}_t + \sigma_s \widehat{\epsilon}_t$, where $\widehat{\epsilon}_t = (Y_t - \alpha_t \widehat{m}_t)/\sigma_t$. Divide by α_s and collect terms to obtain (36). Differentiation gives its tangent. Norm expansion is characterized by $B^\top B$, not the eigenvalues of a potentially nonsymmetric B. In the symmetric case its singular values are the absolute values of its eigenvalues. Positivity of the exact posterior Jacobian then gives the stated covariance threshold. □

B.6.3 The low-noise Jacobian is a bulk–wall measure

For each route-cell interior, let $P_a(y)$ be the unique nearest-point projection onto S_a. On a regular two-route facet F_{ab}, define

$$d_{ab} = P_a - P_b, \qquad D_{ab} = \|d_{ab}\|, \qquad n_{ab} = d_{ab}/D_{ab}.$$

Theorem 26 (Support-manifold bulk–wall decomposition)**.** *Let $\Omega \subset \mathbb{R}^d$ be bounded and open, and write $\Omega_a = \{y \in \Omega : \operatorname{dist}(y, S_a) < \min_{b \neq a} \operatorname{dist}(y, S_b)\}$. Assume, up to an $\mathcal{H}^{d-1}$-null set, that Ω is the disjoint union of these route interiors, on which P_a is unique and C^1, and regular facets F_{ab} at which exactly routes a, b tie with distinct projections. In particular, any within-route cut locus or other omitted projection singularity must lie in the stated $\mathcal{H}^{d-1}$-null exceptional set. A positive-measure medial axis is not covered by this decomposition. Then the matrix-valued measures converge weakly on Ω:*

$$\frac{\Sigma_q(y)}{q}\,\mathrm{d}y \Longrightarrow \sum_a \mathbf{1}_{\Omega_a}(y) DP_a(y)\,\mathrm{d}y + \sum_{\{a,b\}} D_{ab}(y) n_{ab}(y) n_{ab}(y)^\top \mathcal{H}^{d-1} \lfloor F_{ab}. \tag{37}$$

The absolutely continuous term is within-route tangent response. The singular rank-one term concentrates between-route susceptibility on the regular interfaces of the tropical skeleton.

Proof. Set $U_q(y) = -q \log \sum_a Z_{a,q}(y)$. Compactness and the local Laplace principle give $U_q \to U_0 = \frac{1}{2} \min_a \operatorname{dist}(y, S_a)^2$ locally uniformly, while Gaussian differentiation gives

$$\nabla U_q = y - m_q, \qquad \nabla^2 U_q = I - \Sigma_q/q.$$

Set $T_q(y) = \|y\|^2/2 - U_q(y)$ and define T_0 analogously. Each T_q is convex, $T_q \to T_0$ locally uniformly, and its positive semidefinite Hessian measure is $D^2 T_q = (\Sigma_q/q)\,\mathrm{d}y$. Local uniform convergence of convex functions implies weak convergence of their Hessian measures, so $(\Sigma_q/q)\,\mathrm{d}y \Longrightarrow D^2 T_0 = I\,\mathrm{d}y - D^2 U_0$. In Ω_a, $\nabla U_0 = y - P_a(y)$, hence the density is DP_a. Across F_{ab}, ∇U_0 jumps by $-d_{ab} = -D_{ab} n_{ab}$. The singular Hessian is $-D_{ab} n_{ab} n_{ab}^\top \mathcal{H}^{d-1} \lfloor F_{ab}$. Subtracting from $I\,\mathrm{d}y$ reverses its sign and proves (37). □

Corollary 27 (Regular-skeleton identifiability)**.** *If $D_{ab} > 0$ on every regular facet, the singular trace measure in (37) identifies the closure of the regular route skeleton. At $\mathcal{H}^{d-1}$-almost every point of F_{ab}, its matrix density has rank one, leading eigenspace $\operatorname{span}(n_{ab})$, and nonzero eigenvalue D_{ab}.*

Theorem 28 (Finite-resolution recovery)**.** *Let $p_0 = \sum_{i=1}^n \pi_i \delta_{x_i}$ have distinct atoms and positive weights, put $\pi_{\min} = \min_i \pi_i$, and let J_0 denote its limiting wall measure. Let $K_h(z) = h^{-d} K(z/h)$ be a smooth compactly supported mollifier and assume $\Omega - \operatorname{supp} K_h \subseteq \Omega_h$. Suppose a learned posterior mean satisfies $\sup_{y \in \Omega_h} \|\widehat{m}_q(y) - m_q(y)\| \le \varepsilon_q$, and define $\widehat{J}_{q,h} = D(K_h * \widehat{m}_q)$. Then*

$$\sup_{y \in \Omega} \|\widehat{J}_{q,h}(y) - (K_h * J_0)(y)\|_{\mathrm{F}} \le C_{K,d} \left\{ \frac{\varepsilon_q}{h} + \frac{q \log(1/\pi_{\min})}{h^2} \right\}. \tag{38}$$

Consequently $h \downarrow 0$, $\varepsilon_q/h \to 0$, and $q/h^2 \to 0$ give distributional recovery of J_0.

Proof. For finite supports, set $T_q(y) = \frac{1}{2}\|y\|^2 - U_q(y)$, so $\nabla T_q = m_q$, and let T_0 be its tropical limit. The soft-min inequality gives $\|T_q - T_0\|_\infty \le q \log(1/\pi_{\min})$. Split

$$\widehat{J}_{q,h} - K_h * J_0 = D\{K_h * (\widehat{m}_q - m_q)\} + D^2\{K_h * (T_q - T_0)\}.$$

Young's inequality, $\|DK_h\|_1 = O(h^{-1})$, and $\|D^2 K_h\|_1 = O(h^{-2})$ prove (38). Approximate-identity convergence finishes the claim. □

This recovery law deliberately estimates a mollified matrix measure. A within-route medial axis can also create singular projection response, so exact wall semantics require either regular within-route projections or a separate model of their cut loci.

B.6.4 One moving energy draws the cells and drives the flow

Theorem 29 (Cole–Hopf tropical dynamics). *Let $p_q = p_0 * \mathcal{N}(0, qI_d)$ be positive and smooth, and define*

$$U_q(y) = -q \log\{(2\pi q)^{d/2} p_q(y)\}.$$

Whenever the displayed quantities are finite,

$$\nabla U_q = y - m_q, \qquad \nabla^2 U_q = I - \Sigma_q/q, \qquad \partial_q U_q = D_{\mathrm{KL}}(\Pi_q^y \| \Pi_0), \tag{39}$$

where Π_q^y is the posterior law of X given $Y_q = y$. For a finite prior, if $H_q(y) = \min_i\{\frac{1}{2}\|y - x_i\|^2 - q \log \pi_i\}$, then

$$H_q(y) - q \log n \le U_q(y) \le H_q(y), \tag{40}$$

so the same potential converges uniformly to the tropical squared-distance envelope.

In reverse log-variance time $\tau = \log(Q/q)$, probability flow and its linearization satisfy

$$\dot{y}_\tau = -\tfrac{1}{2}\nabla U_q(y_\tau), \qquad \dot{\delta}_\tau = -\tfrac{1}{2}\nabla^2 U_q(y_\tau)\delta_\tau. \tag{41}$$

Therefore the number of instantaneous expanding directions is exactly

$$\mathrm{ind}_-(\nabla^2 U_q) = \#\{j : \lambda_j(\Sigma_q) > q\}, \tag{42}$$

and the moving energy dissipates:

$$\frac{\mathrm{d}}{\mathrm{d}\tau} U_q(y_\tau) = -\tfrac{1}{2}\|\nabla U_q(y_\tau)\|^2 - q D_{\mathrm{KL}}(\Pi_q^{y_\tau} \| \Pi_0) \le 0. \tag{43}$$

Proof. Write $Z_q(y) = \int \exp(-\|y - x\|^2/(2q))\Pi_0(\mathrm{d}x)$, so $U_q = -q \log Z_q$. Differentiation under the integral gives $\nabla U_q = y - m_q$, its Hessian identity, and $\partial_q U_q = D_{\mathrm{KL}}(\Pi_q^y \| \Pi_0)$. For a finite prior, (40) is the standard soft-min inequality. The canonical probability-flow equation is $\mathrm{d}y_q/\mathrm{d}q = \nabla U_q/(2q)$. Since $\mathrm{d}q/\mathrm{d}\tau = -q$, this becomes (41). Its symmetric generator is $-\nabla^2 U_q/2$, proving (42). Finally, the chain rule combines (41), $\mathrm{d}q/\mathrm{d}\tau = -q$, and the third identity in (39) to give (43). □

B.6.5 A finite-time information–amplification budget

Theorem 30 (Finite-time information budget). *Assume $\mathbb{E}\|X\|^2 < \infty$. Fix $0 < q < Q$, put $T = \log(Q/q)$, and follow a differentiable reverse log-variance probability flow with marginals $Y_0 \sim p_Q$ and $Y_T \sim p_q$. Let J_τ be its state Jacobian and*

$$\Gamma_s(y) = \frac{\Sigma_s(y)}{2s} - \frac{I}{2}, \qquad \dot{J}_\tau = \Gamma_{Qe^{-\tau}}(Y_\tau) J_\tau, \qquad J_0 = I.$$

For $A_+(J) = \sum_{j=1}^d [\log \sigma_j(J)]_+$, where $\sigma_j(J)$ are singular values and Γ_+ replaces the negative eigenvalues of Γ by zero, every path satisfies

$$A_+(J_T) \le \int_0^T \mathrm{tr}\big[\Gamma_{Qe^{-\tau}}(Y_\tau)_+\big]\, \mathrm{d}\tau, \tag{44}$$

and at the population level

$$\mathbb{E}A_+(J_T) \le I(X; Y_q) - I(X; Y_Q). \tag{45}$$

Thus, for any $a > 0$, $\mathbb{P}\{\sigma_1(J_T) \ge e^a\} \le [I(X; Y_q) - I(X; Y_Q)]/a$.

Proof. For each exterior power, the Euclidean logarithmic norm of the induced generator is the sum of the corresponding largest eigenvalues of the symmetric matrix Γ. Hence, for $1 \le k \le d$,

$$\sum_{j=1}^k \log \sigma_j(J_T) = \log \| \wedge^k J_T \|_{\mathrm{op}} \le \int_0^T \mathrm{tr}(\Gamma_+)\, \mathrm{d}\tau.$$

Taking the maximum over k, including $k = 0$, proves (44). Positivity of Σ_s gives $\mathrm{tr}((\Gamma_s)_+) \le \mathrm{tr}\,\Sigma_s/(2s)$. The expectation of the latter equals $\mathrm{d}I(X; Y_{Qe^{-\tau}})/\mathrm{d}\tau$ at $s = Qe^{-\tau}$, by vector I–MMSE. Probability flow has these Gaussian-channel marginals, so Tonelli's theorem and integration prove (45). The tail bound is Markov's inequality because $\sigma_1(J_T) \ge e^a$ implies $A_+(J_T) \ge a$. □

This is a population no-free-lunch law: a particular path can amplify strongly, but the expected sum of positive log gains cannot exceed the information acquired over that noise window.

B.6.6 FINITE TROPICAL FLIPS AND NO-ZENO SWITCHING

For a finite empirical prior with distinct atoms x_i and weights $\pi_i > 0$ (after merging any duplicate atoms), define the weighted power functions

$$\mathcal{P}_i(y, q) = \|y - x_i\|^2 - 2q \log \pi_i.$$

Proposition 31 (Noise homotopy has finitely many tropical flips)**.** *Let $D_{ij} = \|x_i - x_j\|$ and $n_{ij} = (x_i - x_j)/D_{ij}$. A nonempty wall between cells i, j lies on*

$$\langle y, n_{ij} \rangle = b_{ij}(q), \qquad b'_{ij}(q) = -\frac{\log(\pi_i/\pi_j)}{D_{ij}}.$$

For an affine circuit C, choose coefficients with $\sum_{i\in C} \alpha_i = 0$ and $\sum_{i\in C} \alpha_i x_i = 0$. Its regular triangulation can change only at

$$q_C = \frac{\sum_{i\in C} \alpha_i \|x_i\|^2}{2 \sum_{i\in C} \alpha_i \log \pi_i}, \tag{46}$$

when the denominator is nonzero and $q_C > 0$. Under isolated nondegeneracy, only finitely many positive critical variances occur. If the masses are equal, every wall is stationary and there are no noise-induced flips.

Proof. Solving $\mathcal{P}_i = \mathcal{P}_j$ gives the wall equation and speed. The power diagram is dual to the lower hull of lifted points $(x_i, \|x_i\|^2 - 2q \log \pi_i)$. A circuit changes orientation only when its lifted points become affinely dependent. Applying the circuit coefficients to the last coordinate and solving for q gives (46). A finite configuration has finitely many circuits. Equal masses annihilate both the pairwise speeds and every circuit denominator. □

Theorem 32 (No-Zeno MAP switching at positive noise)**.** *Let y_q solve canonical probability flow on a compact interval $[q_{\min}, q_{\max}] \subset (0, \infty)$, and define $g_{ij}(q) = \mathcal{P}_i(y_q, q) - \mathcal{P}_j(y_q, q)$. If no g_{ij} is identically zero, the MAP atom changes only finitely many times on the interval. A strictly monotone analytic bijection onto the same q-interval preserves the number of switches. An increasing schedule preserves their order, while a decreasing schedule reverses it.*

Proof. For positive noise, the finite-mixture posterior mean is real analytic in (y, q) because its softmax denominator is strictly positive. The flow and its solution are therefore analytic, as is every g_{ij}. A nonzero real-analytic function has only finitely many zeros on a compact interval in its analytic domain. The minimizing index can change only at the finite union of those pairwise zero sets. A monotone analytic bijection is one-to-one. Its orientation determines whether the order is preserved or reversed. □

An identically zero pairwise gap requires a tie rule if those atoms are maximizers. The theorem asserts finite MAP switching, not monotonicity of the deployed complete-link cloud event.

B.6.7 WHEN COVARIANCE RECOVERS A TROPICAL FACE NORMAL

Proposition 33 (Dominant posterior pair aligns with its face)**.** *For a finite support of diameter B, fix two atoms $i \neq j$ with posterior weights w_i, w_j. Put*

$$s = w_i + w_j, \quad \varepsilon = 1 - s, \quad \alpha = w_i/s, \quad d = x_i - x_j, \quad D = \|d\|,$$

and define $\lambda_{\text{pair}} = s\alpha(1-\alpha)D^2$ and $\eta = \varepsilon(s + \frac{1}{4})B^2$. Then

$$\Sigma_q = s\alpha(1-\alpha)dd^\top + E, \qquad E \succeq 0, \qquad \|E\|_{\text{op}} \leq \eta. \tag{47}$$

If $\lambda_{\text{pair}} > \eta$, the leading covariance eigenvector u_1 is unique up to sign and

$$\sin \angle(\operatorname{span} u_1, \operatorname{span} d) \leq \frac{\eta}{\lambda_{\text{pair}}}, \qquad \lambda_1(\Sigma_q) - \lambda_2(\Sigma_q) \geq \lambda_{\text{pair}} - \eta. \tag{48}$$

The weighted power-cell face between i, j has normal d. If a learned symmetric Jacobian $\widehat{A}_q$ obeys $\|\widehat{A}_q - \Sigma_q/q\|_{\text{op}} \leq \delta$ and $2q\delta < \lambda_{\text{pair}} - \eta$, its leading direction $\widehat{u}_1$ obeys

$$\angle(\operatorname{span} \widehat{u}_1, \operatorname{span} d) \leq \arcsin \frac{\eta}{\lambda_{\text{pair}}} + \arcsin \frac{2q\delta}{\lambda_{\text{pair}} - \eta}.$$

Proof. Treat the posterior as a mixture of the normalized i, j pair and its residual. Total covariance yields (47). Popoviciu's inequality gives residual covariance at most $B^2/4$, and the between-group mean term is at most $s\varepsilon B^2$. Weyl's inequality then gives the eigengap. Projecting the leading-eigenvector equation onto $d^\perp$ gives the first angle bound. The learned bound follows from Davis–Kahan and the triangle inequality for one-dimensional principal angles. □

This proposition connects a dominant posterior pair to its local tropical-face normal. A terminal decision also depends on the remaining flow and its output representation. The finite-horizon construction measures the transported displacement, while a specified terminal margin requires the downstream adjoint of that terminal contrast, evaluated at the anchor, in Proposition 50.

B.6.8 Exact junction coefficients and transported visibility

Theorem 34 (Exact finite-cloud coefficient at a regular junction)**.** *Fix $k \geq 2$ and put $\nu = k - 1$. Let the anchor $X_t \in \mathbb{R}^d$ have a density p_t that is continuous on a neighborhood of a finite family $\{S_\alpha\}_{\alpha=1}^A$ of compact, positive-reach, C^2, codimension-ν strata. Assume that exactly k cell germs meet along each S_α, their $k-1$ independent margin differences have uniformly linearly independent normal gradients, and their boundaries converge uniformly in normal tubular charts to the corresponding conical cell germs. Assume also that these strata exhaust the regular k-cell junctions in the region being measured. Require the probability of the $O(r)$-neighborhood of omitted stratum boundaries and higher-order junctions to be $o(r^\nu)$. This holds, for example, when those sets form a finite union of compact smooth strata of codimension at least $\nu + 1$ with bounded tubular volume and locally bounded density. At $b \in S_\alpha$, let $N_b S_\alpha$ be the normal space, let Π_b be its orthogonal projector, and let $\Gamma_{\alpha,1}(b), \ldots, \Gamma_{\alpha,k}(b)$ be the closed normal cones of the cell germs, with disjoint interiors covering $N_b S_\alpha$.*

Let λ be a continuous scale function satisfying $0 < \lambda_{\min} \leq \lambda \leq \lambda_{\max} < \infty$ on the same neighborhood. Let $\mathcal{W} = (W_0, \ldots, W_L)$ be a finite random offset cloud, independent of X_t, with $\max_j \|W_j\| \leq B < \infty$ almost surely, and define the radius-r cloud $\mathcal{C}_r(x, \mathcal{W}) = \{x + r\lambda(x)W_j : 0 \leq j \leq L\}$. Let $E_r^{(k)}$ be the event that this cloud meets all k local cells in one of the declared tubular neighborhoods. Define

$$\kappa_{k,\mathcal{W}}(b) = \mathbb{E}_{\mathcal{W}} \operatorname{Leb}_\nu \left\{ z \in N_b S_\alpha : \{z + \Pi_b W_j\}_{j=0}^L \text{ meets every } \operatorname{int} \Gamma_{\alpha,\ell}(b) \right\}, \tag{49}$$

where Leb_ν is Lebesgue measure on $N_b S_\alpha$. Assume that, for $\mathcal{H}^{d-\nu}$-almost every b and almost every realized $\mathcal{W}$, the boundary of the displayed translation set has Leb_ν-measure zero, and that its Lebesgue measure is bounded uniformly over those $(b, \mathcal{W})$. Then, as $r \downarrow 0$,

$$\begin{aligned} &\mathbb{P}(E_r^{(k)}) \\ &\quad = r^{k-1} \sum_\alpha \int_{S_\alpha} \lambda(b)^{k-1} p_t(b) \kappa_{k,\mathcal{W}}(b) \, \mathrm{d}\mathcal{H}^{d-k+1}(b) + o(r^{k-1}). \end{aligned} \tag{50}$$

The coefficient is finite under the stated compactness and boundedness conditions. A positive coefficient makes r^{k-1} the leading exponent. If the cloud has fewer than k distinct offsets, it cannot meet all k cell interiors. More generally, failure to cover the relevant normal fan can make the coefficient vanish, leaving probability $o(r^{k-1})$. For $k = 2$, $\kappa_{2,\mathcal{W}}(b)$ is exactly the expected projected cloud width, reducing to Lemma 15.

Proof. In a normal tubular chart, write the center as $x = b + r\lambda(b)z$. The volume element is $r^\nu \lambda(b)^\nu \{1 + o(1)\} \, \mathrm{d}z \, \mathrm{d}\mathcal{H}^{d-\nu}(b)$. Continuity of λ, the tubular expansion of the cell germs, and division by $r\lambda(b)$ make the projected cloud converge to $z + \Pi_b W_j$. The zero-boundary-measure assumption gives almost-everywhere indicator convergence. Compact strata, bounded offsets, and continuity of p_t and λ provide an integrable envelope, so dominated convergence applies. Integrating the normal translation set gives (49). Summing the strata proves (50). The negligible higher-order strata do not contribute at order r^ν. In one normal dimension the valid translations form the interval between the negative largest and negative smallest projected offsets, whose length is the cloud width. □

Theorem 35 (Probability flow transports the coefficient, not the exponent)**.** *Let $\Phi_{q \leftarrow Q}$ be a C^2 probability-flow diffeomorphism pushing p_Q to p_q. Let S_q be a compact codimension-c stratum of*

positive reach and $S_Q = \Phi_{q\leftarrow Q}^{-1}(S_q)$. *For* $x \in S_Q$, $y = \Phi_{q\leftarrow Q}(x)$, $J_x = D\Phi_{q\leftarrow Q}(x)$, *and an orthonormal output normal frame* $N_q(y)$, *define*

$$g_{\perp,c}(x) = \det\{N_q(y)^\top J_x J_x^\top N_q(y)\}^{1/2}. \tag{51}$$

Then the weighted stratum measures satisfy

$$p_Q(x)\,\mathrm{d}\mathcal{H}^{d-c}(x) = g_{\perp,c}(x)p_q(y)\,\mathrm{d}\mathcal{H}^{d-c}(y). \tag{52}$$

Consequently every regular radius-r *tube retains the power* r^c. *The flow changes only its leading coefficient through conormal amplification, density, scale, and stratum area.*

Proof. Choose orthonormal tangent and normal frames at x and y. Since the Jacobian maps source tangents to output tangents, its framed matrix is block upper triangular, $\left[\begin{smallmatrix} A & B \\ 0 & C \end{smallmatrix}\right]$. The tangential area Jacobian is $|\det A|$, while (51) equals $|\det C|$. Thus $|\det J_x| = |\det A| g_{\perp,c}$. Combine the area formula with density transport $p_q(y) = p_Q(x)/|\det J_x|$ to obtain (52). Normal tubular coordinates then give the usual r^c volume factor. Changing variables with (52) modifies its coefficient but cannot change the exponent. □

For $0 < q < Q$, consider the stationary center trajectory of an equal-weight prior supported exactly on the k vertices of a regular $(k-1)$-simplex with pairwise separation D. Its contrast-space gain over $[Q, q]$ is

$$G_k(q, Q; D) = \sqrt{q/Q}\,\exp\left\{\frac{D^2}{4k}\left(\frac{1}{q} - \frac{1}{Q}\right)\right\}.$$

Hence a codimension-$(k-1)$ visibility coefficient scales as $(rG_k)^{k-1}$. The crossover

$$k_\star(q, Q, D) = \frac{D^2(1/q - 1/Q)}{2\log(Q/q)}$$

separates amplified $(k < k_\star)$ from contracted $(k > k_\star)$ contrast spaces, but no finite positive-noise window removes the exponent $k - 1$.

C Extended theory and controlled geometric studies

C.1 Additional setup for exact and stratified models

C.1.1 Variance-exploding Gaussian noising

Let $X_0 \sim P_0$ be a data random variable in $\mathbb{R}^d$, let $Z \sim \mathcal{N}(0, I_d)$ be independent, and define

$$X_\sigma = X_0 + \sigma Z, \qquad \sigma \in (0, \sigma_{\max}]. \tag{53}$$

For every Borel probability law P_0, the noised density $p_\sigma = P_0 * \varphi_\sigma$ is smooth and strictly positive, and its pointwise posterior moments are finite because the Gaussian likelihood dominates every polynomial. We additionally assume $\mathbb{E}\|X_0\|_2^2 < \infty$ whenever taking population MMSE or mutual-information expectations. The score and posterior mean denoiser are

$$s_\sigma(x) = \nabla_x \log p_\sigma(x), \qquad D_\sigma(x) = \mathbb{E}[X_0 \mid X_\sigma = x]. \tag{54}$$

Tweedie's identity gives (Robbins, 1956; Efron, 2011; Vincent, 2011)

$$D_\sigma(x) = x + \sigma^2 s_\sigma(x). \tag{55}$$

Writing the probability-flow ODE in the noise standard-deviation coordinate gives

$$\frac{dx}{d\sigma} = -\sigma s_\sigma(x) = \frac{x - D_\sigma(x)}{\sigma}. \tag{56}$$

Sampling integrates Equation (56) backward from $\sigma_{\max}$ to a small $\sigma_{\min}$. This choice of clock is central: a singular vector field in one clock can have an exact or stable discretization in another (Zhang, 2026).

C.1.2 BRANCH DECOMPOSITIONS AND FINITE TEMPERATURE

Assume that the noised density admits a positive branch decomposition

$$p_\sigma(x) = \sum_{k=1}^{K} p_{k,\sigma}(x), \qquad L_k(x,\sigma) = \log p_{k,\sigma}(x). \tag{57}$$

The branch responsibility and branch score are

$$r_k(x,\sigma) = \frac{e^{L_k(x,\sigma)}}{\sum_j e^{L_j(x,\sigma)}}, \qquad s_{k,\sigma} = \nabla L_k, \tag{58}$$

so that

$$s_\sigma(x) = \sum_{k=1}^{K} r_k(x,\sigma) s_{k,\sigma}(x). \tag{59}$$

The components may represent mixture labels, smooth support strata, local patches, classes, or empirical atoms. All branch-level statements refer to this specified decomposition. Different decompositions can induce different cells even when they define the same noised density.

Definition 36 (Active cell, margin, and tropical diffusion skeleton)**.** For fixed σ, define the active label

$$\kappa_\sigma(x) \in \arg\max_k L_k(x,\sigma), \tag{60}$$

the top-two logit margin

$$\Delta_\sigma(x) = L_{(1)}(x,\sigma) - L_{(2)}(x,\sigma) \geq 0, \tag{61}$$

and the branch cell

$$\mathcal{C}_k(\sigma) = \{x : L_k(x,\sigma) \geq L_j(x,\sigma)\ \forall j\}. \tag{62}$$

The *tropical diffusion skeleton* is the tie set

$$\mathcal{K}_\sigma = \{x : \Delta_\sigma(x) = 0\}. \tag{63}$$

The family $\{\mathcal{K}_\sigma\}_{\sigma>0}$, including its cell adjacency and births or deaths as σ changes, is the tropical diffusion dynamics associated with the chosen branches.

Remark 37 (Not classical tropicalization). For finite point mixtures, L_k are affine functions of the lifted coordinates after removing the common quadratic term, and $\mathcal{K}_\sigma$ is a weighted power complex. For general supports, $\mathcal{K}_\sigma$ can be curved and stratified. We use "tropical" for the max-plus active-phase limit of log-sum-exp, not to assert that $\mathcal{K}_\sigma$ is a rational polyhedral tropical variety in the algebraic-geometric sense.

C.1.3 METRIC ERROR AND ROUTE ERROR

Let $x^\star(\sigma)$ be an exact trajectory and $\widehat{x}(\sigma)$ a numerical or learned-field trajectory. Endpoint RMSE measures metric error. We additionally define

$$\mathsf{RouteErr} = \mathbb{P}[\kappa_{\sigma_{\min}}(\widehat{x}(\sigma_{\min})) \neq \kappa_{\sigma_{\min}}(x^\star(\sigma_{\min}))]\,. \tag{64}$$

More generally one can compare the entire cell sequence. Route error is discontinuous as a function of endpoint location, but is margin-stable away from $\mathcal{K}_\sigma$ under bounded logit variation.

Proposition 38 (Margin certificate)**.** *Suppose every branch logit is G-Lipschitz, with $G > 0$, on a convex set containing both $x^\star$ and $\widehat{x}$. If the exact top-two margin is $\Delta_\sigma(x^\star) > 0$ and*

$$\|\widehat{x} - x^\star\| < \frac{\Delta_\sigma(x^\star)}{2G}, \tag{65}$$

then $\kappa_\sigma(\widehat{x}) = \kappa_\sigma(x^\star)$.

Proof. For the exact winner i and any competitor j, $L_i(\widehat{x}) - L_j(\widehat{x}) \geq L_i(x^\star) - L_j(x^\star) - 2G\|\widehat{x} - x^\star\| > 0$. □

For a radius certificate, the Lipschitz domain must contain the entire closed ball being certified. In a finite point mixture, the logit margin and G both scale as σ^{-2} away from the skeleton, so their ratio need not shrink with noise. A small margin predicts combinatorial fragility, not necessarily large Euclidean error. Our event-prediction experiment tests this distinction.

C.1.4 FINITE POINT MIXTURES

For much of the exact theory, let

$$P_0 = \sum_{k=1}^{K} w_k \delta_{\mu_k}, \qquad w_k > 0, \quad \sum_k w_k = 1. \tag{66}$$

Then

$$p_\sigma(x) = (2\pi\sigma^2)^{-d/2} \sum_k w_k \exp\left(-\frac{\|x-\mu_k\|^2}{2\sigma^2}\right), \tag{67}$$

$$q_k(x;\sigma) = \|x-\mu_k\|^2 - 2\sigma^2 \log w_k, \tag{68}$$

$$r_k(x,\sigma) = \mathrm{softmax}_k\left(-\frac{q_k(x;\sigma)}{2\sigma^2}\right), \tag{69}$$

$$D_\sigma(x) = \bar{\mu}(x,\sigma) = \sum_k r_k(x,\sigma)\mu_k. \tag{70}$$

The cells $q_k \le q_j$ are additively weighted Voronoi, or power, cells. Unequal weights make the complex evolve with σ because the offsets scale as $\sigma^2 \log w_k$.

C.2 EXACT TROPICAL GEOMETRY BEYOND THE MAIN-TEXT STATEMENT

C.2.1 SOFT-MIN TROPICALIZATION

Define the finite-temperature free energy

$$F_\sigma(x) = -2\sigma^2 \log \sum_{k=1}^{K} \exp\left(-\frac{q_k(x;\sigma)}{2\sigma^2}\right) \tag{71}$$

and the tropical hard minimum $T_\sigma(x) = \min_k q_k(x;\sigma)$.

Theorem 39 (Nonasymptotic tropical approximation)**.** *For every x, $\sigma > 0$, and finite K,*

$$T_\sigma(x) - 2\sigma^2 \log K \le F_\sigma(x) \le T_\sigma(x). \tag{72}$$

Let $q_{(1)} < q_{(2)}$ and $g = q_{(2)} - q_{(1)} > 0$. If $k^\star$ is the unique winner, then

$$1 - r_{k^\star}(x,\sigma) \le (K-1)e^{-g/(2\sigma^2)}, \tag{73}$$

$$\|D_\sigma(x) - \mu_{k^\star}\| \le \mathrm{diam}(\{\mu_k\})(K-1)e^{-g/(2\sigma^2)}. \tag{74}$$

Proof. Let $m = T_\sigma$. Factoring out $e^{-m/(2\sigma^2)}$ gives $1 \le \sum_k e^{-(q_k-m)/(2\sigma^2)} \le K$, which yields Equation (72). For a unique winner,

$$\frac{1-r_{k^\star}}{r_{k^\star}} = \sum_{j\neq k^\star} e^{-(q_j-q_{k^\star})/(2\sigma^2)} \le (K-1)e^{-g/(2\sigma^2)}. \tag{75}$$

Since $r_{k^\star} \le 1$, the first bound follows. The denoiser bound follows by writing $D - \mu_{k^\star} = \sum_{j\neq k^\star} r_j(\mu_j - \mu_{k^\star})$. □

The theorem is elementary, but it provides the key empirical consequence: away from a power boundary, the exact finite-data denoiser becomes exponentially indistinguishable from a lookup table. The transition is governed by the *power gap* g, not simply Euclidean nearest-neighbor distance.

C.2.2 EXACT SCORE HESSIAN AND THE COVARIANCE GEOMETRY

The score of the point mixture is

$$s_\sigma(x) = \frac{\bar{\mu}(x,\sigma) - x}{\sigma^2}. \tag{76}$$

Its derivative has an exact posterior-covariance form.

Theorem 40 (Score Hessian and flow Jacobian)**.** *Let*

$$\Sigma_r(x,\sigma) = \sum_k r_k(\mu_k - \bar{\mu})(\mu_k - \bar{\mu})^\top. \tag{77}$$

Then

$$\nabla_x s_\sigma(x) = -\frac{I}{\sigma^2} + \frac{\Sigma_r(x,\sigma)}{\sigma^4}, \tag{78}$$

$$\nabla_x f_\sigma(x) = \frac{I}{\sigma} - \frac{\Sigma_r(x,\sigma)}{\sigma^3}, \qquad f_\sigma(x) = \frac{x - D_\sigma(x)}{\sigma} = -\sigma s_\sigma(x). \tag{79}$$

Moreover, $\nabla D_\sigma = \Sigma_r/\sigma^2$.

Proof. Differentiating the responsibility gives $\nabla r_k = r_k(\mu_k - \bar{\mu})/\sigma^2$. Hence $\nabla\bar{\mu} = \sum_k \mu_k(\nabla r_k)^\top = \Sigma_r/\sigma^2$. The two formulas follow from differentiating $s = (\bar{\mu} - x)/\sigma^2$ and $f = (x - \bar{\mu})/\sigma$. □

The score Hessian contains isotropic contraction $-I/\sigma^2$ and a positive semidefinite covariance correction. The latter grows when separated branches retain posterior mass. Near a regular two-branch interface it is approximately rank one, aligned with the top-two displacement and hence the local normal of the tropical skeleton at the current noise level.

C.2.3 Binary shock layer

For $K = 2$, write $\delta = \mu_1 - \mu_2$ and

$$g(x,\sigma) = \log\frac{r_1}{r_2} = \log\frac{w_1}{w_2} - \frac{\|x-\mu_1\|^2 - \|x-\mu_2\|^2}{2\sigma^2}. \tag{80}$$

Then $r_1 = (1+e^{-g})^{-1}$ and $2r_1 - 1 = \tanh(g/2)$. Consequently the score decomposes into a smooth average drift and a universal interfacial term,

$$s_\sigma(x) = \frac{(\mu_1+\mu_2)/2 - x}{\sigma^2} + \frac{\delta}{2\sigma^2}\tanh\left(\frac{g(x,\sigma)}{2}\right). \tag{81}$$

This is consistent with the Burgers derivation of Sarkar (2026).

Corollary 41 (Layer width and derivative growth)**.** *Along the unit normal* $n = \delta/\|\delta\|$*, the physical distance over which* r_1 *changes from* 0.1 *to* 0.9 *is*

$$\ell_{10\text{–}90} = \frac{2\sigma^2\log 9}{\|\delta\|}. \tag{82}$$

At the equal-responsibility interface,

$$n^\top\nabla s_\sigma n = -\frac{1}{\sigma^2} + \frac{\|\delta\|^2}{4\sigma^4}, \tag{83}$$

so the leading score-Hessian magnitude is $\Theta(\sigma^{-4})$ *and the probability-flow Jacobian magnitude is* $\Theta(\sigma^{-3})$ *as* $\sigma \downarrow 0$.

What the scaling does and does not imply. It implies severe local conditioning in the σ clock for point-like separated branches. It does not imply that every fixed-step deterministic method has a nonvanishing terminal error, nor that the divergence has the same exponent for extended supports. The relevant integrated effect depends on layer width, trajectory incidence, clock, symmetry, and whether the stiff direction has an exact discrete update. Our asymptotic-preserving controls in Appendix C.5 are designed to prevent the common but incorrect inference "large Jacobian ⇒ unavoidable sampling failure."

C.2.4 Local Laplace phases for mixed-dimensional branches

Let $P_0 = \sum_{k=1}^{K} \pi_k P_k$, where $\nu_k = \dim S_k$ and P_k has a continuous positive density ρ_k with respect to ν_k-dimensional Hausdorff measure on a C^2 embedded stratum $S_k \subset \mathbb{R}^d$. Fix a point x whose projection $\Pi_k(x)$ onto every relevant stratum is unique, lies in the relative interior, is nonfocal, and is locally isolated from all other minimizing pieces (Federer, 1959; 1969). Normal-coordinate Laplace asymptotics then give

$$\log p_{k,\sigma}(x) = -\frac{d_k(x)^2}{2\sigma^2} - (d - \nu_k)\log\sigma + \log A_k(x) + o(1), \tag{84}$$

where $p_{k,\sigma}$ includes the mixture weight π_k, $d_k(x) = \operatorname{dist}(x, S_k)$, and the finite coefficient $A_k(x) > 0$ contains the branch weight, density at $\Pi_k(x)$, the Gaussian constant, and the curvature/projection Jacobian. Stronger and uniform corner-layer expansions require tangent-cone analysis as in Brosse & Dalalyan (2026). Equation (84) is used only in this regular local regime.

Proposition 42 (Mixed-rank boundary shift)**.** *Let $x = x_\sigma$ remain in a compact neighborhood on which the regularity conditions above and the $o(1)$ remainders in Equation (84) hold uniformly for branches i and j. Equality of their branch log densities then implies*

$$d_i(x)^2 - d_j(x)^2 = 2\sigma^2\left[(\nu_i - \nu_j)\log\sigma + \log\frac{A_i(x)}{A_j(x)}\right] + o(\sigma^2). \tag{85}$$

Thus unequal branch dimensions produce a leading displacement on the $\sigma^2\log(1/\sigma)$ scale, whereas equal dimensions produce an $O(\sigma^2)$ shift when their prefactors differ.

The nearest-support rule $-2\sigma^2\log p_\sigma \to d_S^2$ is only the leading asymptotic term (Proposition 1; proof in Appendix B.2), analogous to classical heat-kernel asymptotics (Varadhan, 1967). At finite noise, dimension, local density, and curvature decide between nearly equidistant branches. Ignoring these offsets can systematically suppress rare or low-dimensional strata.

C.2.5 Dynamic weighted power complexes

For unequal point weights, the pairwise boundary between i and j is

$$2x^\top(\mu_i - \mu_j) = \|\mu_i\|^2 - \|\mu_j\|^2 - 2\sigma^2\log\frac{w_i}{w_j}. \tag{86}$$

As σ decreases, previously hidden cells can appear and adjacency can change. The complexes need not form a nested refinement. We therefore measure a cell's persistence by the noise interval over which it is visible and receives path occupancy. Both neighboring offsets and geometric shielding influence this noise interval, so weight alone does not determine it.

C.3 Controlled checks of soft tropical geometry

C.3.1 Soft-min rate and finite-width skeleton

We first evaluate the gap between F_σ and the weighted hard minimum T_σ on two-dimensional mixtures over a fixed domain. The pointwise error stays below $2\sigma^2\log K$ as required by Theorem 39. The observed spatial RMSE values for $\sigma = 0.4, 0.2, 0.1, 0.05$ are respectively

$$0.04333, \quad 0.005304, \quad 6.611\times10^{-4}, \quad 8.187\times10^{-5}. \tag{87}$$

A log–log fit gives slope 3.015. The pointwise $O(\sigma^2)$ gap is concentrated in an $O(\sigma^2)$-wide layer, so integrating its square gives $O(\sigma^6)$ and taking the square root gives the observed cubic RMSE scaling.

C.3.2 Binary shock and normal alignment

For an exact binary point mixture, finite differences recover the score-Hessian exponent -4.013, the probability-flow Jacobian exponent -3.013, and the 10–90 layer-width exponent $+2.000$. The finite-difference relative error against the analytic Hessian is 1.3×10^{-5}.

For random K-component mixtures, we measure the cosine between the leading eigenvector of posterior mean covariance Σ_r and the displacement of the top-two means. Mean alignments in dimensions 2, 8, and 32 are 0.933, 0.987, and 0.999986. Restricting to points where the top two components carry more than 95% posterior mass raises these to 0.963, 0.9997, and essentially 1. Thus a locally binary skeleton has an identifiable rank-one normal even in moderately high dimension.

C.3.3 MIXED RANK

We construct exact Gaussian convolutions of a point, a line segment away from its endpoints, and a plane patch in an ambient space. After removing the leading distance and rank terms, the residual branch logits agree with the analytic phase formula to 2.27×10^{-13}. A distance-only tie estimate is displaced because it omits the $-(d-r)\log\sigma$ factor. The rank-corrected estimate is 25.1 times more accurate. The measured scaled shift is -0.833367, versus the predicted -0.833333.

These results validate the local mechanism but should not be extrapolated to arbitrary corners. At a corner, the Gaussian mass of the tangent cone and adapted Jacobian enter, as formalized by Brosse & Dalalyan (2026). Our continuous Y-junction later tests whether a simple component-level event detector remains effective in this richer geometry.

C.4 MOVING-INTERFACE AND EVENT-RESOLUTION STUDIES

C.4.1 GLOBAL SCHEDULE VERSUS PATHWISE EVENT RESOLUTION

We evaluate a rare triangular point mixture in $d = 2$ and a hierarchical mixture in $d = 8$. The standard baseline is Heun on a Karras schedule (Karras et al., 2022). A "global TDD" schedule reallocates the same number of steps according to an averaged posterior-covariance difficulty. event-refined Heun refines individual trajectories when they encounter an event.

At 16 base steps in the 2D task, Karras-Heun obtains paired RMSE 0.002386 at 32 NFE. The global tropical schedule degrades to 0.145637 at the same NFE. Pathwise event resolution obtains 0.002068 at mean NFE 34.58. At four base steps in 8D, Heun obtains RMSE 0.478 at 8 NFE, RK4 obtains 0.226 at 16 NFE, and event-refined Heun obtains 0.180 at mean NFE 9.67. At 16 base steps, the event method gives only a small gain, 0.06669 versus 0.06864, while a 64-NFE RK4 reference-like run achieves substantially higher numerical accuracy.

The conclusion is not that event bisection is a new dominant solver. It is that a global scalarization destroys useful pathwise information, whereas a small number of targeted refinements can improve low-budget integration. The effect shrinks when the base schedule already resolves the interfaces.

C.4.2 A 36-CONFIGURATION MOVING-INTERFACE GRID

We cross separations $\{1, 2, 3, 4\}$ with rare weights $\{0.10, 0.03, 0.01\}$, using three independent random replications and four or six base steps per configuration. The moving weighted power boundary allows trajectories to change their dominant branch during integration.

Base steps	Method	Mean NFE	Cell disagreement	Rare-mass error	Paired RMSE
4	Heun	8.000	0.003897	0.003897	0.14975
4	event-refined Heun	8.185	0.002810	0.002762	0.10346
6	Heun	12.000	0.001325	0.001325	0.08612
6	event-refined Heun	12.238	0.001452	0.001452	0.06992

Table 4: Mean results across 36 moving-interface configurations. Event refinement helps most at four steps. At six steps it improves metric error but is slightly worse on mean cell disagreement.

At four base steps, event refinement lowers mean paired RMSE by 30.9% and cell disagreement by 27.9% for only 2.3% more NFE. It wins on RMSE in 86.1% of configurations and on assignment in 58.3%. At six steps, it still lowers RMSE by 18.8% but slightly increases mean assignment error. The failure cases usually occur when a narrow interface is detected after the coarse step would already have landed on the correct side. Repeated bisection then introduces a small phase shift. This is why a margin detector must be coupled to a calibrated local error estimate to account for local integration error.

C.4.3 THIN-INTERFACE SCORE PERTURBATION

To test whether average score error is calibrated to route identity, we perturb an exact 97/3 two-component score only inside a moving tube around the equal-logit interface. We sweep tube

width and perturbation amplitude. One representative setting has width factor 2 and amplitude 2.56. Its time-integrated absolute score MSE is 0.01065, relative integrated MSE is 2.148×10^{-5}, and the mean probability proxy of the perturbed tube is 0.00863. Nevertheless, 2.129% of trajectories change terminal component and the rare frequency shifts by 2.129 percentage points. Paired endpoint RMSE under the same initial conditions is 0.568.

This experiment establishes poor *calibration*, not a contradiction to score-error convergence theory. The perturbation amplitude is not infinitesimal, and the route statistic is discontinuous at the separatrix. The correct lesson is that score MSE should be accompanied by a boundary margin or branch-mass diagnostic when discrete terminal cell identity matters.

C.5 Adaptive controls near sharp route layers

C.5.1 Event information versus embedded Runge–Kutta error

On the same 36 exact 1D tasks, we compare event-refined Heun with embedded Heun–Euler refinement, which thresholds the local discrepancy without using branch labels or margins.

At the near-zero-overhead end, four-step Heun uses 8 NFE with assignment error 0.003780 and RMSE 0.14987. event-refined Heun uses 8.185 NFE with assignment 0.002798 and RMSE 0.10566. A loose embedded tolerance of 0.5 uses 8.342 NFE with assignment 0.003036 and RMSE 0.11759. Thus event information gives a modest efficiency advantage at nearly matched NFE. When more computation is permitted, ordinary error control wins: tolerance 0.08 uses 10.148 NFE and reaches assignment 0.00125, RMSE 0.06882.

These results favor combining conventional local error control with a tropical margin certificate, as supported by the event-prediction study in Appendix C.6.

C.5.2 Rank-deficient Gaussian terminal layer

Let $P_0 = \mathcal{N}(0, \Sigma_0)$ with Σ_0 rank $r < d$. In a normal direction with zero clean variance, $D_\sigma(x) = 0$ and Equation (56) reduces to

$$\frac{dx_\perp}{d\sigma} = \frac{x_\perp}{\sigma}, \qquad x_\perp(\sigma_b) = \frac{\sigma_b}{\sigma_a} x_\perp(\sigma_a). \tag{88}$$

A single Euler step in the σ clock is exact: $x_b = x_a + (\sigma_b - \sigma_a)x_a/\sigma_a = (\sigma_b/\sigma_a)x_a$. This is the simplest instance of the layer-exact property emphasized by Zhang (2026).

Our $d = 8$, rank-2 experiment decreases $\sigma_{\min}$ to 10^{-6}. For 4, 8, 16, and 32 steps, Euler in σ has normal RMSE between 1.9×10^{-17} and 2.1×10^{-18}. Uniform stepping in σ^2 has normal RMSE 0.272, 0.195, 0.139, and 0.099, respectively. Heun in σ is also numerically exact in the normal coordinate.

C.5.3 Symmetric posterior switching

We also solve a symmetric two-point mixture with an exact one-dimensional quantile reference. Karras-Heun exhibits no terminal cell disagreement at 4, 8, 16, or 32 steps, and its paired RMSE approaches an interpolation/reference floor near 2.1×10^{-5}. Therefore, a posterior switch and a diverging local derivative do not imply a fixed deterministic branch error. Symmetry, trajectory incidence, and the sampling clock can make the switch uniformly benign.

C.6 Predicting route changes from tropical margins

We construct 172,800 exact coarse Heun step events from binary and ternary one-dimensional mixtures. Each proposed step is compared with a 48-substep RK4 micro-solution from the same state. Predictors available from the proposal are drift norm, exact local Jacobian stiffness, $|h|\|J\|$, embedded Euler–Heun discrepancy, minimum top-two margin across the proposal, and a dominant-cell flip. Mixture configurations are split disjointly into train and test. The test set has 57,600 steps.

High metric error is the top endpoint-error decile within each configuration and step. *Route mismatch* means that Heun and the micro-solution end in different active cells. The held-out set contains 161 route mismatches, making this event rare in the full test population.

Predictor	AUC: high metric error	AUC: wrong terminal cell
Drift norm	0.886	0.619
Jacobian stiffness	0.623	0.903
Embedded RK error	0.586	0.884
Tropical margin	0.573	**0.973**
Cell flip	0.519	N/R
Scalar numerical features	0.899	0.908
Scalar + tropical	**0.907**	**0.990**

Table 5: Held-out-configuration ROC AUC. Drift magnitude predicts metric error, whereas top-two margin is the strongest single route-identity predictor. Combining both feature families is best overall. N/R means not reported.

This is the cleanest support for treating skeleton combinatorics as a separate control variable. It also argues against replacing numerical analysis with tropical analysis: the best classifier combines both.

C.7 A CONTINUOUS STRATIFIED Y-JUNCTION

Point mixtures yield exactly polyhedral power cells, but real supports are extended. We therefore define a clean law as a weighted mixture of uniform measures on three line segments meeting at the origin. For a segment direction u, length L, and observation x, decompose $a = x^\top u$ and $x_\perp = x - au$. The noised branch density is

$$p_{u,\sigma}(x) = \frac{e^{-\|x_\perp\|^2/(2\sigma^2)}}{L\sqrt{2\pi}\sigma}\left[\Phi\left(\frac{L-a}{\sigma}\right) - \Phi\left(\frac{-a}{\sigma}\right)\right] \tag{89}$$

up to the remaining ambient Gaussian constant. The segment posterior is a truncated normal, giving analytic density, responsibilities, denoising, and probability flow without a discretized point cloud.

We evaluate angles $60°$, $90°$, and $120°$, with weights $(0.75, 0.20, 0.05)$ and $(0.50, 0.40, 0.10)$, 1,200 trajectories per geometry, and 4, 6, 8, or 12 base steps. The reference is 384-step logarithmic RK4. A separate 384-versus-768 audit over six geometries gives mean RMSE 6.13×10^{-10}, maximum RMSE 1.87×10^{-9}, and zero branch disagreements.

Steps	Method	Mean NFE	Cell disagreement	Rare-mass error	Paired RMSE
4	Heun	8.00	0.01597	0.00583	0.08139
4	event-refined Heun	19.45	0.01542	0.00931	0.08051
6	Heun	12.00	0.00625	0.00292	0.03374
6	event-refined Heun	27.79	0.00556	0.00319	0.03589
8	Heun	16.00	0.00417	0.00208	0.01924
8	event-refined Heun	35.42	0.00347	0.00194	0.02161
12	Heun	24.00	0.00208	0.00083	0.00844
12	event-refined Heun	49.80	0.00181	0.00153	0.01121

Table 6: Continuous Y-junction results averaged over six geometries. Event splitting slightly improves mean branch disagreement but more than doubles NFE and often worsens endpoint or rare-mass error.

Event refinement wins per-configuration cell disagreement only 45.8% of the time and RMSE only 41.7%. The mean route metric improves slightly because a few geometries benefit, but the cost is excessive. Thus the continuous experiment supports the *diagnostic* value of branch margins while rejecting our simple discrete event algorithm as a general stratified-support sampler.

C.8 EMPIRICAL CRYSTALLIZATION ON FINITE SUPPORTS

C.8.1 COVERING RADIUS AND THE SINGLETON TRANSITION ON A CIRCLE

We begin with a population law uniform on the unit circle and an empirical law supported on N random circle points. The population posterior mean has a closed form involving modified Bessel

functions. The empirical posterior mean is the responsibility-weighted finite softmax over the N atoms. Let h_N denote the angular covering radius of the training set. The dimensionless ratio σ/h_N compares the Gaussian neighborhood scale with the largest local gap between samples.

Across $N \in \{16, 32, 64, 128, 256\}$ and six independent random draws, active-set entropy and singleton responsibility collapse approximately when plotted against σ/h_N. When σ is larger than local covering gaps, several neighboring atoms contribute and the empirical denoiser approximates a local tangent average. When $\sigma < h_N$, large portions of the circle are dominated by one atom.

We then integrate exact probability-flow trajectories from $\sigma_{\max} = 1.6$ to $\sigma_{\min} = 0.008$ for $N = 64$. Both the population and empirical flows land close to the unit circle: mean radius errors are 0.00632 and 0.00629. Their angular behavior differs. The empirical output has mean angular nearest-training distance 0.00574, versus 0.04683 for the population oracle. Its angular Fourier discrepancy is 0.286, versus 0.120 for the population flow. Thus a standard support metric misses this empirical crystallization in angular output allocation.

C.8.2 Held-out digit trajectories and crystallization occupancy

We next use the 1,797-image handwritten-digits dataset distributed with scikit-learn (Alpaydin & Kaynak, 1998; Pedregosa et al., 2011). Images of digit "3" are projected to the first 10 principal components. A bank of 115 training examples defines an empirical point-mixture denoiser. Initial states come from held-out examples plus Gaussian noise, so their distribution differs from the exact noised empirical law. We integrate 3,600 trajectories for each of four independent randomized replications.

For each independent draw, the crystallization scale $\sigma_\star$ is selected without terminal labels: it is the descending noise level at which the median top-two logit margin is approximately 1.4. We compare four training-atom predictors of terminal allocation:

(a) crystallization occupancy $\mathsf{Occ}_i(\sigma_\star)$.

(b) time-integrated high-margin tropical exposure.

(c) held-out feature-space coverage.

(d) static kNN sparsity within the training bank.

The target is each atom's terminal capture frequency, and AUC evaluates whether a predictor identifies the top quintile of captured atoms.

Predictor	Spearman	SD	Top-quintile AUC	SD
Crystallization occupancy	**0.9501**	0.0082	**0.9680**	0.0087
Tropical integrated exposure	0.4414	0.1028	0.7042	0.0735
Held-out coverage	0.1053	0.1656	0.5439	0.0973
kNN sparsity	0.0181	0.1250	0.5091	0.0866

Table 7: Predicting terminal empirical-atom allocation for held-out digit trajectories. Occupancy at the intermediate crystallization scale is highly predictive. Terminal-saturated exposure and static sparsity are much weaker.

Intermediate-noise cell occupancy predicts terminal allocation in this 115-example empirical-Bayes PCA model. Transfer to large trained diffusion models remains untested.

C.9 Learned-denoiser refinement of the empirical skeleton

We test whether a finite network approaches the empirical denoising optimum using 48 fixed points on a randomly perturbed unit circle. A fully connected ReLU MLP receives $(x_\sigma, \log\sigma)$ and uses EDM preconditioning (Karras et al., 2022). Three independent initializations are evaluated at 0, 100, 500, 2,000, and 8,000 optimization steps on the same grid and reverse initial states. Metrics are population- and empirical-Bayes denoising MSE, 95th-percentile low-noise denoising curvature, mean support-radius error, angular Fourier discrepancy, and near-copy rate, defined by angular distance below 0.01 from a training point.

Step	Pop. MSE	Emp. MSE	Copy rate	95% sharpness	Radius err.	Fourier disc.
0	0.03880	0.04717	0.1440	1.013	0.4886	0.0726
100	0.02445	0.02230	0.1588	1.054	0.3079	0.1715
500	0.00788	0.00558	0.1812	4.761	0.0440	0.1706
2,000	0.00814	0.00349	0.2019	10.299	0.0186	0.1985
8,000	0.00650	0.00244	0.2645	21.625	0.0160	0.2290

Table 8: Mean over three trained ReLU denoisers. Once the support is learned, further optimization increasingly matches the empirical oracle, sharpens the map, and raises near-copying while angular coverage worsens.

The empirical oracle has near-copy rate 0.487 and Fourier discrepancy 0.215, compared with 0.147 and 0.0438 for the population oracle. The 8,000-step network lies between them. These measurements concern the output map, not an enumeration of ReLU activation regions (Zhang et al., 2018a), and do not establish a generalization theorem for unseen samples.

C.10 Cell birth, adjacency, and persistence

For a 15-point unequal-weight hierarchical mixture, we sweep σ from 2 to a small terminal value and measure visible cells, adjacency, path occupancy, and persistence. Visible cells increase from five to 15 and adjacency edges from 8 to 33. Cell weight and persistence have Spearman correlation 0.696: geometric exposure and shielding also affect when cells appear. A low-weight exposed atom can emerge earlier and persist longer than a heavier shielded atom, so weight alone does not describe its multiscale visibility across noise levels.

C.11 Complementary proofs for exact-mixture and stratified models

C.11.1 Common-covariance Gaussian mixtures

The point-mixture results extend with minor changes to a mixture whose clean components are $\mathcal{N}(\mu_k, \tau^2 I)$. Let $v = \tau^2 + \sigma^2$. The noised responsibilities are

$$r_k(x, \sigma) = \mathrm{softmax}_k \left(\log w_k - \frac{\|x - \mu_k\|^2}{2v} \right). \tag{90}$$

The noised score is

$$s_\sigma(x) = \frac{\bar{\mu} - x}{v}, \qquad \bar{\mu} = \sum_k r_k \mu_k, \tag{91}$$

while the clean posterior mean is

$$D_\sigma(x) = \frac{\tau^2}{v} x + \frac{\sigma^2}{v} \bar{\mu}. \tag{92}$$

Differentiation yields

$$\nabla s_\sigma = -\frac{I}{v} + \frac{\Sigma_r}{v^2}. \tag{93}$$

For $\tau > 0$, the $O(v)$ interface width and $O(v^{-2})$ Hessian remain bounded as noise vanishes. The point-atom divergences therefore saturate for fixed-width Gaussian branches.

C.11.2 Derivation of the binary tanh profile

Let $m = (\mu_1 + \mu_2)/2$ and $r = r_1$. Since $\bar{\mu} = m + (r - 1/2)\delta$ and $r - 1/2 = (1/2)\tanh(g/2)$,

$$s_\sigma = \frac{m - x}{\sigma^2} + \frac{\delta}{2\sigma^2} \tanh(g/2). \tag{94}$$

Furthermore $\nabla g = \delta/\sigma^2$. Along $n = \delta/\|\delta\|$, changing r from 0.1 to 0.9 changes g from $-\log 9$ to $+\log 9$, hence the physical width is $2\sigma^2 \log 9/\|\delta\|$. At $g = 0$, $r(1 - r) = 1/4$ and $\Sigma_r = \delta\delta^\top/4$, giving the width and peak derivative stated in Corollary 41.

C.11.3 Integrated soft-min rate near a regular interface

Suppose two phases cross transversally along a smooth codimension-one surface. In normal coordinate u, their difference is $au + O(u^2)$ with $a \neq 0$. The soft-min correction has the form

$$2\sigma^2 \log\left(1 + e^{-|au|/(2\sigma^2)}\right). \tag{95}$$

Its $O(\sigma^2)$ amplitude is exponentially localized to $|u| = O(\sigma^2)$. The integrated square is $O(\sigma^6)$, giving the $O(\sigma^3)$ local L^2 norm measured in the soft-min study. Multiple interface intersections and domain boundaries can alter constants and lower-order terms.

C.11.4 Local mixed-rank Laplace expansion

We sketch the regular-stratum calculation. Choose normal coordinates around the unique projection $y = \Pi_k(x)$. For $u \in \mathbb{R}^{\nu_k}$, write a nearby support point as $Y(u) = y + Uu + \frac{1}{2}\mathrm{II}(u,u) + o(\|u\|^2)$, where U spans the tangent space and II is the second fundamental form. The squared distance expands as

$$\|x - Y(u)\|^2 = d_k(x)^2 + u^\top H_k(x) u + o(\|u\|^2), \tag{96}$$

with H_k positive definite away from the focal set. The branch convolution is

$$p_{k,\sigma}(x) = \pi_k (2\pi\sigma^2)^{-d/2} \int_{S_k} \rho_k(y') e^{-\|x-y'\|^2/(2\sigma^2)} d\mathcal{H}^{\nu_k}(y') \tag{97}$$

$$= e^{-d_k(x)^2/(2\sigma^2)} \sigma^{-(d-\nu_k)} A_k(x)[1 + o(1)]. \tag{98}$$

Taking logs gives Equation (84). At a boundary or corner the tangent domain is a cone in $\mathbb{R}^{\nu_k}$, which changes the leading coefficient and requires the uniform analysis of Brosse & Dalalyan (2026).

C.11.5 Empirical denoiser crystallization

For an empirical bank, let $i^\star$ be the unique weighted nearest atom and let $g(x,\sigma) = q_{(2)} - q_{(1)}$. Equation (74) gives exponential convergence to $x_{i^\star}$. The score and probability-flow vector become

$$s_\sigma(x) = \frac{x_{i^\star} - x}{\sigma^2} + O\left(\sigma^{-2} e^{-g/(2\sigma^2)}\right), \quad f_\sigma(x) = \frac{x - x_{i^\star}}{\sigma} + O\left(\sigma^{-1} e^{-g/(2\sigma^2)}\right). \tag{99}$$

The leading flow inside a fixed cell has exact solution $x(\sigma) - x_{i^\star} = C\sigma$. Thus a trajectory that enters a persistent singleton cell with a stable gap contracts linearly in σ toward that atom. Whether it enters such a cell is a global path-occupancy question, not determined by the local formula alone.

C.11.6 Why equal-weight atom frequencies can remain uniform

If initial states are drawn exactly from $p^{(n)}_{\sigma_{\max}}$ and the exact probability-flow ODE is integrated to zero, its marginal converges to the empirical target $n^{-1}\sum_i \delta_{x_i}$. Hence each atom receives mass $1/n$, regardless of its Euclidean Voronoi volume. Power-cell volume describes deterministic nearest-component geometry under a reference base measure. Probability-flow mass also depends on how the vector field transports the initial mixture. This proves why static volume is not a universal extraction score and motivates held-out or model-generated incoming paths.

C.11.7 Route identity and margin

Proposition 38 is deterministic. A probabilistic route bound follows by splitting on the exact margin:

$$\mathbb{P}[\widehat{\kappa} \neq \kappa^\star] \leq \mathbb{P}[\Delta \leq 2G\varepsilon] + \mathbb{P}[\|\widehat{x} - x^\star\| \geq \varepsilon]. \tag{100}$$

The second term can be controlled by a metric error bound. The first is an anti-concentration or margin condition near the skeleton. This decomposition formalizes why score or endpoint MSE alone is insufficiently calibrated to route identity without information about boundary mass.

C.12 EXACT CONTINUOUS Y-JUNCTION FORMULAS

For a segment $S_u = \{tu : 0 \le t \le L\}$ with unit direction u, define $a = x^\top u$, $b^2 = \|x\|^2 - a^2$, $z_0 = -a/\sigma$, and $z_L = (L-a)/\sigma$. The convolution of the uniform arclength law with a two-dimensional Gaussian is

$$p_{u,\sigma}(x) = \frac{1}{L}\frac{e^{-b^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}[\Phi(z_L) - \Phi(z_0)] \tag{101}$$

up to the branch weight. Conditional on the branch, the latent coordinate T has a normal distribution $\mathcal{N}(a, \sigma^2)$ truncated to $[0, L]$. Its posterior mean is

$$\mathbb{E}[T \mid x, u] = a + \sigma\frac{\phi(z_0) - \phi(z_L)}{\Phi(z_L) - \Phi(z_0)}. \tag{102}$$

Therefore the branch denoiser is $u\mathbb{E}[T \mid x, u]$, and the full denoiser is its responsibility-weighted sum. These formulas are used directly with stable Gaussian CDF evaluations in Experiment E10.

C.13 ADDITIONAL NUMERICAL TABLES FOR CONTROLLED STUDIES

C.13.1 REPRESENTATIVE EXACT SAMPLING ROWS

Case	Solver	Schedule	NFE	Paired RMSE	Sliced W_2	Rare error
2D rare triangle	Heun	Karras	32.00	0.002386	0.2040	0.00217
2D rare triangle	Heun	global tropical	32.00	0.145637	0.2093	0.00217
2D rare triangle	RK4	Karras	64.00	0.000182	0.2007	0.00217
2D rare triangle	event-refined Heun	Karras	34.58	0.002068	0.1921	0.00217
8D hierarchy	Heun	Karras	32.00	0.068638	0.1199	0.00514
8D hierarchy	Heun	global tropical	32.00	1.353108	0.3841	0.00629
8D hierarchy	RK4	Karras	64.00	0.000254	0.1115	0.00514

Table 9: Representative rows from the machine-readable exact sampling table. Sliced distribution metrics vary less than paired trajectory error because all methods approximately preserve coarse component mass.

C.13.2 ASYMPTOTIC-PRESERVING RANK AUDIT

Method	Steps	Total RMSE	Normal RMSE	Tangent RMSE
Euler σ	4	0.04080	1.85×10^{-17}	0.08159
Euler σ	8	0.02140	1.15×10^{-17}	0.04279
Euler σ	16	0.01102	4.17×10^{-18}	0.02203
Euler σ	32	0.00560	2.10×10^{-18}	0.01120
Euler σ^2	4	0.23881	0.27208	0.07779
Euler σ^2	8	0.17042	0.19540	0.04031
Euler σ^2	16	0.12102	0.13925	0.02022

Table 10: Rank-deficient Gaussian control. The normal coordinate is layer-exact for Euler in σ. Total error comes from tangent coordinates.

C.13.3 GUIDANCE SWEEP

γ	Heun cell err.	Heun RMSE	$q_{.95}\|J_f\|$	Effective modes	Alignment
0.0	0.0000	0.0121	9.175	11.987	-0.000
0.5	0.0068	0.0635	9.032	10.618	0.342
1.0	0.0168	0.1029	8.726	7.704	0.626
2.0	0.0200	0.1156	6.704	3.923	0.880
4.0	0.0214	0.1146	4.761	2.049	0.961
8.0	0.0414	0.1607	3.810	1.270	0.982

Table 11: Exact CFG sweep. Mode concentration and route error increase even while a high quantile of local PF stiffness decreases.

C.13.4 REFERENCE-RESOLUTION AUDIT

Y angle	Rare weight	RMSE: 384-step vs 768-step RK4
60°	0.05	8.26×10^{-10}
60°	0.10	1.87×10^{-9}
90°	0.05	2.75×10^{-10}
90°	0.10	2.50×10^{-10}
120°	0.05	1.07×10^{-10}
120°	0.10	3.51×10^{-10}

Table 12: The continuous Y-junction high-resolution reference is converged far below the reported coarse-solver errors. All six branch-assignment disagreements are zero.

C.14 DETAILED CONFIGURATIONS FOR CONTROLLED STUDIES

E1: geometry. Weighted configurations use noise 0.4, 0.2, 0.1, and 0.05. Binary derivatives use the analytic equal-logit plane and a logarithmic noise grid. High-dimensional means are separated Gaussian draws. Alignment is measured unconditionally and at top-two mass above 0.95. Exact mixed-rank integrals use the known intrinsic branch dimensions.

E2: exact sampling. The 2D triangle has unequal weights and one rare component. The 8D mixture has separated coarse groups, finer within-group structure, and unequal weights. Initial states follow the exact maximum-noise mixture and are paired with dense RK4 references. Global scheduling reweights a dense grid by average covariance activity before selecting nodes. event-refined Heun uses margin threshold 1.25, covariance-to-noise trace gate 0.75, and maximum depth 3.

E3 and E6: moving interfaces. Binary one-dimensional mixtures use separations $1, 2, 3, 4$, rare weights $0.10, 0.03, 0.01$, and three independent random replications per parameter pair. Four- and six-step Karras schedules use a dense reference. Heun–Euler tolerances are 0.5, 0.2, 0.08, 0.03, 0.012, and 0.005. Score perturbations use a 97/3 mixture with separation 4 and a smooth tube around the moving equal-logit interface. Width and amplitude are swept. The highlighted setting illustrates a visible route transition, not the minimum measured error.

E4: circle and digits. Circle points are random angular draws, covering radius is half the maximum circular gap, and population denoising uses the exact posterior. Digits use class 3, split-specific 10-dimensional PCA, and 115 training examples. Held-out examples are independently noised. Trajectory margins select the crystallization scale without terminal capture labels.

E5: neural denoiser. For the network in Appendix C.9, training resamples noisy observations and noise scales. Sharpness uses finite-difference curvature on a dense angular/noise grid. Sampling fixes the solver and initial noise across all recorded training checkpoints.

E7: classifier-free guidance. The conditional and unconditional exact mixtures have the same 12 atom locations with different normalized weights. The guided field is their declared affine score combination, with no normalized-density assumption. The comparison proxy supplies statically reweighted responsibilities to the point-mixture event detector.

E8: asymptotic-preserving controls. The 8D rank-2 Gaussian uses tangent/normal errors and uniform σ or σ^2 steps down to 10^{-6}. Dense interpolation of the exact monotone CDF/quantile map defines the symmetric two-point reference and explains its RMSE floor.

E9: event prediction. Binary tasks use separations $2, 3, 4, 5$ and rare weights $0.02, 0.05, 0.12$. Ternary tasks use separations $2.5, 4, 5.5$ and weights $(0.72, 0.20, 0.08)$ or $(0.45, 0.45, 0.10)$. Each configuration has 2,400 trajectories and four coarse steps. Logistic models use standardized log features, balanced class weights, and fixed regularization. Configuration IDs divisible by three define the held-out split, keeping every trajectory from a configuration on the same side.

E10–E12: continuous junction and persistence. Y-junction segment length is 2.4, maximum noise 2.5, and terminal noise 0.012. Event refinement has depth 3 and activates below 0.75, yet still over-refines. Persistence uses 15 unequal-weight hierarchical atoms, a dense 2D grid of visible cells and Delaunay-like adjacencies, and logarithmically spaced scales. Reference checks compare 384- and 768-step RK4 on six Y geometries with 400 new trajectories.

C.15 Why route covariance is necessary for local correction

Fix a noise level $\sigma > 0$ and write $D_\sigma(x) = \mathbb{E}[x_0 \mid x_\sigma = x]$. A branch summary at x consists of posterior responsibilities and conditional branch means,

$$S_\sigma(x) = \{(r_k(x), m_k(x))\}_{k=1}^K, \qquad m_k(x) = \mathbb{E}[x_0 \mid x_\sigma = x, k].$$

It excludes the within-branch posterior covariances. This is precisely the information discarded by a class-prototype or top-two approximation.

Theorem 43 (Branch summaries do not identify a local correction)**.** *For every $\sigma > 0$, there exist two smooth one-dimensional, two-branch Gaussian data distributions q_+ and q_- and an observation $x = 0$ such that*

$$S_{\sigma,+}(0) = S_{\sigma,-}(0), \qquad D_{\sigma,+}(0) = D_{\sigma,-}(0) \neq 0,$$

but

$$D'_{\sigma,+}(0) - 1 = -\big(D'_{\sigma,-}(0) - 1\big) \neq 0.$$

Consequently, consider the autonomous denoising-induced fields $b_\pm(x) = D_{\sigma,\pm}(x) - x$. Any deterministic one-step correction whose $O(h^2)$ term is a function only of $S_\sigma(0)$ and $b_\pm(0)$ cannot reduce the leading local error relative to Euler for both fields.

Proof. Choose $r \in (0,1) \setminus \{1/2\}$ and $m > 0$ such that $B = 4r(1-r)m^2 \in (0, \sigma^2)$. Let

$$v_0 = \sigma^2 - B, \qquad v_\pm = v_0 \pm \delta, \quad 0 < \delta < \min\{B, \sigma^2 - B\}.$$

For model $j \in \{+,-\}$, assign weights $1-r$ and r to the negative and positive Gaussian branches, respectively, with common variance

$$\tau_j^2 = \frac{v_j \sigma^2}{\sigma^2 - v_j},$$

and prior means

$$M_{s,j} = s\, m \frac{\sigma^2}{\sigma^2 - v_j}, \qquad s \in \{-1, +1\}.$$

After observing $x = 0$ through Gaussian noise of variance σ^2, the two conditional branch posteriors have means $-m, +m$ and common variance v_j. Their marginal likelihoods at zero are equal by symmetry, so their posterior responsibilities remain $1-r, r$. Thus both models have the same branch summary and the same posterior mean

$$D_{\sigma,j}(0) = (2r-1)m =: \mu \neq 0.$$

The posterior-covariance identity gives

$$D'_{\sigma,j}(0) = \frac{\mathrm{Var}(x_0 \mid x_\sigma = 0)}{\sigma^2} = \frac{v_j + B}{\sigma^2} = 1 \pm \frac{\delta}{\sigma^2}.$$

Hence $b_\pm(0) = \mu$ and $b'_\pm(0) = \pm\delta/\sigma^2$. Their exact time-$h$ flows satisfy

$$\phi_\pm^h(0) = h\mu \pm \frac{h^2}{2}\frac{\delta}{\sigma^2}\mu + O(h^3).$$

The Euler prediction $h\mu$ lies between these two endpoints to leading order. A summary-only method must add the same $h^2 c$ in both worlds. Writing $A = |\delta\mu|/(2\sigma^2) > 0$, strict improvement over Euler for the positive world requires c to have one sign, while strict improvement for the negative world requires the opposite sign. No common c satisfies both requirements. □

Remark 44 (Scope of the impossibility result). Theorem 43 is not a lower bound against arbitrary multi-query ODE solvers. A new field evaluation can reveal curvature that the branch summary omits. The theorem rules out the stronger premise behind a frozen top-two correction: responsibilities and prototypes alone do not determine even the sign of the missing local term. This motivates using branch variables to schedule a strong base integrator while retaining its field evaluations. The construction freezes σ and proves an autonomous-snapshot impossibility result. In a non-autonomous probability flow the second derivative also contains $\partial_t b_t$. No claim that this term cannot cancel the constructed spatial effect follows from this proof.

C.16 Continuous allocation and label-stability extensions

Alongside discrete prefix allocation, a continuous route-loss objective isolates the value of trajectory-dependent scheduling. We analyze this objective, label stability, and terminal intervention directions. These guarantees concern the stated objective, with no optimality claim for the empirical event-priority scheduler used in the checkpoint experiments.

C.16.1 Optimal fixed-budget route scheduling

Let $\rho(t) \geq 0$ denote the continuum step density along a fixed sampling clock: $\rho(t)\,\mathrm{d}t$ is the number of solver intervals assigned near t. Suppose the leading expected route loss under a local order-p approximation is

$$\mathcal{R}[\rho] = \int_0^1 a(t)\rho(t)^{-p}\,\mathrm{d}t, \qquad \int_0^1 \rho(t)\,\mathrm{d}t = N,$$

where $a(t) \geq 0$ is a route-hazard density and N is the interval budget. We use the lower-semicontinuous convention $a\rho^{-p} = 0$ when $a = \rho = 0$ and $a\rho^{-p} = +\infty$ when $a > 0$ and $\rho = 0$. Unlike a norm-only truncation coefficient, a weights an error by whether it changes the continuation endpoint cell.

Theorem 45 (Oracle route-hazard allocation)**.** *Assume $p > 0$, $N > 0$, a is measurable, and $0 < \int_0^1 a(t)^{1/(p+1)}\,\mathrm{d}t < \infty$. Over measurable densities $\rho \geq 0$ with $\int_0^1 \rho(t)\,\mathrm{d}t = N$, the unique minimizer almost everywhere is*

$$\rho^\star(t) = \frac{N\,a(t)^{1/(p+1)}}{\int_0^1 a(s)^{1/(p+1)}\,\mathrm{d}s}.$$

The optimal leading route loss is

$$\mathcal{R}[\rho^\star] = N^{-p}\left(\int_0^1 a(t)^{1/(p+1)}\,\mathrm{d}t\right)^{p+1}.$$

Proof. Put $\theta = 1/(p+1)$ and $A = \int_0^1 a(t)^\theta\,\mathrm{d}t$. If $\mathcal{R}[\rho] = \infty$, the lower bound below is immediate. Otherwise, Hölder's inequality and the zero-value convention give

$$A = \int_0^1 (a\rho^{-p})^\theta \rho^{p\theta}\,\mathrm{d}t \leq \mathcal{R}[\rho]^\theta N^{p\theta}.$$

Therefore $\mathcal{R}[\rho] \geq N^{-p}A^{p+1}$. Equality in Hölder's inequality holds exactly when $a\rho^{-p}$ is proportional to ρ and $\rho = 0$ on $\{a = 0\}$, which yields the displayed $\rho^\star$ uniquely almost everywhere. Direct substitution shows that it attains the bound. □

This is an oracle for the stated route-loss functional, not a license to omit base integration wherever $a = 0$. A practical base solver needs a positive minimum density $\rho_{\min}$, or the allocation must refer only to additional verification work. With $N > \rho_{\min} > 0$ and the constraint $\rho(t) \geq \rho_{\min}$, the same convex argument gives $\rho(t) = \max\{\rho_{\min}, c\,a(t)^{1/(p+1)}\}$, where $c > 0$ enforces the budget. Neither version certifies an unmodelled base discretization error.

Corollary 46 (Robustness to multiplicative calibration error)**.** *Let $\widehat{a}$ obey $e^{-\epsilon}a(t) \leq \widehat{a}(t) \leq e^{\epsilon}a(t)$ almost everywhere, and allocate $\widehat{\rho}$ according to Theorem 45 with $\widehat{a}$ in place of a. Then*

$$\frac{\mathcal{R}[\widehat{\rho}]}{\mathcal{R}[\rho^\star]} \leq \exp\left(\frac{2p}{p+1}\epsilon\right).$$

Proof. Set $\theta = 1/(p+1)$ and $A = \int a^\theta$. The multiplicative inequalities force $\widehat{a} = 0$ on $\{a = 0\}$. On $\{a > 0\}$, set $u = (\widehat{a}/a)^\theta \in [e^{-\epsilon\theta}, e^{\epsilon\theta}]$. All integrands below vanish on $\{a = 0\}$. Direct substitution gives

$$\mathcal{R}[\widehat{\rho}] = N^{-p} \left(\int a^\theta u \right)^p \left(\int a^\theta u^{-p} \right).$$

Bounding the two integrals by their worst multiplicative factors and dividing by $N^{-p} A^{p+1}$ proves the result. □

Theorem 47 (Deadband safety for residual replanning)**.** *Let (Z, T) be a random trajectory–time pair, let $R = r_Z(T)$ be the trajectory-specific risk residual relative to a frozen base scheduler, let $\widehat{R} = \widehat{r}_Z(T)$ be its online estimate, and fix $\tau \geq 0$ and $\delta \in [0, 1]$. Define the soft-deadband operator*

$$\mathcal{S}_\tau(u) = \operatorname{sign}(u)(|u| - \tau)_+.$$

The deadband residual policy applies a correction proportional to $\mathcal{S}_\tau(\widehat{R})$. Then:

1. *$\mathcal{S}_\tau$ is the unique minimizer of $\frac{1}{2}(v - u)^2 + \tau|v|$, is one-Lipschitz, and returns the base scheduler exactly whenever $|u| \leq \tau$.*
2. *for any $\epsilon \in [0, \tau]$, if $\Pr(|R| \leq \epsilon) > 0$ and $\Pr(|\widehat{R} - R| \leq \tau - \epsilon \mid |R| \leq \epsilon) \geq 1 - \delta$, then $\Pr(\mathcal{S}_\tau(\widehat{R}) \neq 0 \mid |R| \leq \epsilon) \leq \delta$.*
3. *pointwise on the event $|\widehat{R} - R| \leq \tau$, if $|R| > 2\tau$, the correction has the oracle sign and magnitude at least $|R| - 2\tau$.*

Thus a calibrated deadband converts estimation uncertainty into an explicit fallback guarantee. Once the residual estimate is available, applying the scalar threshold needs no additional network evaluation.

Proof. The first statement is the standard scalar optimality condition for the proximal map of $\tau|\cdot|$. Its piecewise-linear form has slope zero or one, proving non-expansiveness and exact fallback. Conditional on $|R| \leq \epsilon$, the calibration event implies $|\widehat{R}| \leq |R| + |\widehat{R} - R| \leq \tau$, so the correction is zero. Its conditional complement has probability at most δ. Finally, $|R| > 2\tau$ and $|\widehat{R} - R| \leq \tau$ imply that $\widehat{R}$ has the same sign as R and

$$|\mathcal{S}_\tau(\widehat{R})| = |\widehat{R}| - \tau \geq |R| - 2\tau.$$

□

Theorem 48 (Value of trajectory-adaptive allocation)**.** *Fix $p > 0$ and $N > 0$. Let Z index a random sampling trajectory and suppose $(Z, t) \mapsto a_Z(t) \in [0, \infty)$ is jointly measurable. Assume that $0 < \int_0^1 a_Z(t)^{1/(p+1)}\,\mathrm{d}t < \infty$ almost surely and, for $\bar{a}(t) = \mathbb{E}_Z[a_Z(t)]$, that $0 < \int_0^1 \bar{a}(t)^{1/(p+1)}\,\mathrm{d}t < \infty$. Among schedules that must be chosen before observing Z, the optimal density is*

$$\rho_{\mathrm{global}}(t) = \frac{N\,\bar{a}(t)^{1/(p+1)}}{\int_0^1 \bar{a}(s)^{1/(p+1)}\,\mathrm{d}s}, \qquad \bar{a}(t) = \mathbb{E}_Z[a_Z(t)].$$

If the scheduler may condition on the realized trajectory, its oracle density is $\rho_Z^\star$ from Theorem 45, applied pointwise to a_Z, and

$$\mathbb{E}_Z[\mathcal{R}_Z[\rho_Z^\star]] \leq \mathbb{E}_Z[\mathcal{R}_Z[\rho_{\mathrm{global}}]]\,.$$

Equality holds only if $\rho_Z^\star = \rho_{\mathrm{global}}$ almost everywhere for almost every Z (apart from zero-hazard sets). Thus state adaptation has a strict oracle advantage exactly when normalized hazard profiles vary across trajectories. Variation in hazard magnitude alone is insufficient.

Proof. Joint measurability makes the normalized density $\rho_Z^\star$ measurable in (Z, t). For a schedule independent of Z, Tonelli's theorem gives

$$\mathbb{E}_Z[\mathcal{R}_Z[\rho]] = \int_0^1 \bar{a}(t)\rho(t)^{-p}\,\mathrm{d}t.$$

The assumed integral for $\bar{a}$ makes this risk finite at its optimizer, and Theorem 45 therefore yields ρ_{global}. For almost every realized Z, $\rho_Z^\star$ minimizes $\mathcal{R}_Z$ over the same budget-feasible set, so

$$\mathcal{R}_Z[\rho_Z^\star] \le \mathcal{R}_Z[\rho_{\mathrm{global}}].$$

Taking expectations proves the inequality. Strict convexity on positive hazard sets gives the equality condition. Finally, multiplying a_Z by a trajectory-dependent scalar leaves its normalized oracle density unchanged, which proves the last statement. □

C.16.2 Stable labels and causal directions

The next two propositions bound label errors and identify the perturbation that maximizes the first-order change of a terminal contrast.

Proposition 49 (Margin stability and event-time localization)**.** *Let $L_k(t)$ be branch logits and $g(t) = L_{(1)}(t) - L_{(2)}(t)$ the top-two margin. If estimated logits satisfy $\max_k \sup_t |\widehat{L}_k(t) - L_k(t)| \le \epsilon$, then the active branch is unchanged wherever $g(t) > 2\epsilon$. If a binary crossing $L_i(t_\star) = L_j(t_\star)$ is isolated and transversal with $|(L_i - L_j)'(t)| \ge \gamma > 0$ in a neighborhood of $t_\star$, every estimated crossing in that neighborhood lies within $2\epsilon/\gamma$ of $t_\star$.*

Proof. Changing the maximizer requires closing the true top-two gap, while each of the two logits can move by at most ϵ. This proves the first claim. For the second, the estimated pairwise difference is uniformly within 2ϵ of the true difference. The mean-value theorem and the transversality bound convert this vertical error into the stated time error. □

Proposition 50 (Downstream-adjoint causal normal)**.** *Let $\Phi_{t\to 0}$ be the differentiable continuation map from an intermediate state x to the terminal output, and let q be a differentiable scalar endpoint contrast. Among perturbations δ satisfying $\|\delta\|_2 \le \eta$, the first-order increase of $q(\Phi_{t\to 0}(x+\delta))$ is maximized by*

$$\delta^\star = \eta \, \frac{J_\Phi(x)^\top \nabla q(\Phi(x))}{\|J_\Phi(x)^\top \nabla q(\Phi(x))\|_2},$$

when the denominator is nonzero. A predictive probe normal is a causal branch normal only to the extent that it aligns with this downstream adjoint.

Proof. The first-order endpoint change is the inner product $\langle J_\Phi(x)^\top \nabla q(\Phi(x)), \delta\rangle$. The result follows from Cauchy–Schwarz, with equality in the stated direction. □

Proposition 51 (Sparse-probe compute and staleness guarantee)**.** *Consider classifier-free guidance with integers $N \ge 1$ denoiser steps, $m \ge 0$ prompt deletion variants, and $0 \le Q \le N$. Suppose the variants are evaluated only on a probe set $\mathcal{Q}$ of size Q, while the most recent unit causal normal is reused between probes. The exact fractional increase in denoiser-example count relative to ordinary classifier-free guidance is*

$$\frac{mQ}{2N}.$$

This is an example-count ratio, not an NFE or wall-clock ratio. Batching and hardware determine the latter two quantities. Let $g(t)$ be the current guidance gap, $n(t)$ the dense unit causal normal, and $n_j = n(t_j)$ the most recent probed normal. Write the intervention as

$$\mathcal{T}(n, g) = g - \lambda(n, g)\, n n^\top g.$$

Assume $\|g(t)\|_2 \le G$, $|\lambda(n,g)| \le \Lambda$, the gate is K_λ-Lipschitz in its normal argument, and

$$\|n(t) - n(s)\|_2 \le L_n |t - s|.$$

If every time lies within Δ of its preceding probe, then

$$\|\mathcal{T}(n(t), g(t)) - \mathcal{T}(n_j, g(t))\|_2 \le (\Lambda + K_\lambda) G L_n \Delta.$$

Thus probe spacing produces a linear, auditable approximation error while its additional denoiser-example cost scales linearly with Q/N.

Proof. Ordinary classifier-free guidance evaluates one unconditional and one conditional example per step, for $2N$ denoiser-example evaluations. Each probe adds exactly m deletion-variant examples, giving $2N + mQ$ in total. For unit vectors u, v,

$$\|uu^\top - vv^\top\|_{\mathrm{op}} = \sqrt{1 - (u^\top v)^2} \le \sqrt{2(1 - u^\top v)} = \|u - v\|_2.$$

Add and subtract $\lambda(n_j, g)n(t)n(t)^\top g$, use the gate and guidance-gap bounds, and then apply $\|n(t) - n_j\|_2 \le L_n\Delta$. □

Proposition 52 (Partially energy-balanced treatment)**.** *Let $g \neq 0$ be a classifier-free guidance gap and let*

$$\widetilde{g} = g - \lambda\langle g, n\rangle_+ n, \qquad \|n\|_2 = 1, \quad 0 \le \lambda \le 1.$$

For a norm-recovery coefficient $\beta \in [0, 1]$, define

$$\widehat{g}_\beta = \begin{cases} \big((1-\beta)\|\widetilde{g}\|_2 + \beta\|g\|_2\big)\widetilde{g}/\|\widetilde{g}\|_2, & \widetilde{g} \neq 0, \\ g, & \widetilde{g} = 0. \end{cases}$$

When $\widetilde{g} \neq 0$, the endpoints recover raw projection at $\beta = 0$ and the original guidance norm at $\beta = 1$, and rescaling preserves the treated direction. In that nondegenerate case,

$$\|\widehat{g}_\beta\|_2 = (1-\beta)\|\widetilde{g}\|_2 + \beta\|g\|_2.$$

For both cases in the definition and every $\beta \in [0, 1]$,

$$\|\widehat{g}_\beta - g\|_2 \le (1+\beta)\|\widetilde{g} - g\|_2.$$

Thus β interpolates the guidance norm while retaining the treated direction. The degenerate case $\widetilde{g} = 0$ returns g for every β, since no projected direction is defined there.

Proof. When $\widetilde{g} \neq 0$, the endpoint, norm, and direction statements follow directly from the definition, while the triangle inequality and its reverse form give

$$\|\widehat{g}_\beta - g\|_2 \le \|\widehat{g}_\beta - \widetilde{g}\|_2 + \|\widetilde{g} - g\|_2 = \beta\big|\|g\|_2 - \|\widetilde{g}\|_2\big| + \|\widetilde{g} - g\|_2 \le (1+\beta)\|\widetilde{g} - g\|_2.$$

When $\widetilde{g} = 0$, the declared fallback gives $\widehat{g}_\beta = g$, so the error bound is immediate. No norm interpolation or projected-direction claim is made in this case. □

The fallback above is a pointwise prescription, not a globally regular ODE field: approaching $\widetilde{g} = 0$ can change the normalized direction discontinuously. For numerical integration one may choose $\varepsilon > 0$ and use

$$g_{\beta,\varepsilon} = \left(1 - \beta + \frac{\beta\|g\|_2}{\sqrt{\|\widetilde{g}\|_2^2 + \varepsilon^2}}\right)\widetilde{g}.$$

This continuous regularization returns zero at a fully removed direction and converges to $\widehat{g}_\beta$ wherever $\widetilde{g} \neq 0$. It sacrifices exact norm recovery near zero. Neither construction here establishes a high-order solver convergence guarantee.

D Experimental specifications and additional results

D.1 Benchmarks and frozen protocol

Controlled mechanism tests. The controlled U-Net and diffusion-transformer studies use class-conditional models for which full state Jacobians and intervention directions are available. Within each eligible group, the leading symmetric-Jacobian direction and a norm-matched random direction use the same state, noise level, radius, and remaining continuation. The endpoint statistic is a frozen feature distance between the two signed continuations.

DiT-XL/2. The temporal study uses the same pretrained DiT-XL/2 checkpoint and seven matched anchors with DDIM, DPM-Solver++, and UniPC. The separate 64-trajectory direction study uses DDIM with 20 steps and its midpoint anchor. For each trajectory, a reproducibly generated CPU Gaussian matrix $Z \in \mathbb{R}^{4096\times 8}$ is factorized by reduced QR. Its orthonormal factor is the probe matrix $P \in \mathbb{R}^{4096\times 8}$. The terminal-action, instantaneous symmetric clean-predictor, and random directions all lie in this same subspace. Exact forward-mode automatic differentiation supplies the two Jacobian actions. The comparison is paired by trajectory, not optimized over the full state.

Native-resolution SDXL. Transfer uses 1024×1024 SDXL with source-and-prompt-disjoint COCO val2017 groups. The development-frozen route-hazard order and uniform allocation receive the same continuation budget. Evaluation is grouped by source so that captions or perturbations derived from one image never appear across development and evaluation partitions.

Declared route event. For deployment experiments, a cloud contains the unperturbed endpoint and 12 signed pairs $x \pm r\|x\|_2 u_i$, where x is the anchor state, r is the relative radius, and each u_i is a separately normalized full-state Gaussian draw. These rays are neither QR-orthogonalized nor selected by the finite-horizon method. The event is complete-link fragmentation at cut 1.2 under the maximum of CLIP cosine distance divided by 0.06 and downsampled pixel root-mean-square distance divided by 0.12. Thus the allocation tests do not attribute their gains to finite-horizon direction selection. Pointwise cell audits instead use a frozen ImageNet-1K ResNet-50 top-1 partition. Cloud fragmentation and classifier-cell splitting are different declared events.

Frozen event-risk predictor. The DiT predictor is a gradient-boosted classifier with 100 depth-two trees, learning rate 0.05, and minimum leaf size five. Its 93 inputs comprise nine numerical anchor summaries and 84 hidden summaries from blocks 0, 13, and 27. The numerical inputs describe time, signal-to-noise ratio, state/prediction magnitudes, and conditional–unconditional discrepancies. At each block, four channel vectors—conditional mean and standard deviation, guidance-gap mean and root-mean-square—supply seven statistics: mean, standard deviation, root-mean-square, the 10th/50th/90th percentiles, and maximum absolute value. Whole trajectories form disjoint training/calibration/test splits of 307/103/102 groups, with three anchors per group. Three-fold grouped cross-validation within training selects hyperparameters. Out-of-fold predictions fit temperature 1.103, and calibration fits a trajectory-max conformal offset 0.656 at error level 0.1. The frozen predictor uses neither terminal JVPs nor a directly measured between-route covariance as inputs.

Pathwise continuation budget. At the three allocation anchors, a signed pair costs respectively $c_1 = 150$, $c_2 = 102$, and $c_3 = 52$ continuation neural-function evaluations (NFE). Each trajectory receives its own hard cap $B = \lfloor 0.25 \times 12(c_1 + c_2 + c_3) \rfloor = 912$. Unused compute cannot be borrowed by another trajectory. The anytime study fixes prefix depths before outcomes, stops at the first positive prefix, and does not reallocate savings. The adaptive scheduler reallocates only within that trajectory's cap. These budgets exclude common base-trajectory generation, direction construction, decoding, and feature extraction, so they are not end-to-end wall-clock budgets.

D.2 Metrics and statistical analysis

Let F be the remaining latent continuation, x its anchor state, P an orthonormal probe matrix, v a unit coefficient vector, and $\rho > 0$ an absolute perturbation radius. Centered terminal gain is

$$g_\rho(v) = \frac{\|F(x + \rho P v) - F(x - \rho P v)\|}{2\rho}.$$

The finite-horizon comparison reports paired differences in log gain. Temporal prevalence contrast is the mean early-anchor event rate minus the mean late-anchor event rate. The log signal-to-noise-ratio slope is fit within each trajectory and summarized across independent trajectories. Event capture is

$$\text{capture} = \frac{\text{full-cloud positives found by the policy}}{\text{all full-cloud positives}},$$

and attainable efficiency divides policy capture by the best capture reachable under the same realized pathwise costs. The latter is a diagnostic ceiling, not an implementable policy.

Paired confidence intervals resample independent trajectory or source groups with 10,000 bootstrap replicates. Simultaneous directional comparisons use adjusted lower endpoints. The nested-cloud count interval below is explicitly a Wilson interval, not a bootstrap interval. The eight-group resource pilot reports descriptive paired medians and counts without confirmatory intervals. Counts accompany effects so that evaluation scale remains visible.

D.3 Additional paired results

The original 64-trajectory test fixes rank eight and relative radius 0.05: finite-horizon directions beat random rays in all 64 groups and instantaneous directions in 47, with median log-gain differences 0.357 and 0.0582. The separate pilot below tests dependence on rank, radius, estimator, and measured computation without enlarging the confirmatory cohort or changing its prespecified statistical family of trajectory-paired finite-horizon comparisons.

D.4 Rank, radius, and measured direction costs

Eight new fixed noise–class groups use float32 DiT-XL/2, mathematical attention, DDIM with 20 steps, and its midpoint anchor. One Gaussian 4096×16 QR basis per group supplies nested ranks four, eight, and sixteen. Alongside exact terminal JVPs, we evaluate the instantaneous symmetric clean-predictor direction, a matched random subspace ray, and a centered finite-difference (FD) terminal Jacobian with fixed construction radius $0.0025\|x\|_2$. All four directions are then tested at the same three larger radii. One separate warmup group is excluded. All eight evaluation groups and all 288 method–rank–radius cells are retained.

Rank	Relative radius	Exact − instantaneous	Exact − random	Exact − FD
4	0.025	0.1966 (8/8)	0.3035 (8/8)	0.0033 (6/8)
4	0.050	0.1867 (8/8)	0.2210 (7/8)	0.0014 (4/8)
4	0.100	0.0276 (5/8)	0.0847 (6/8)	0.0048 (5/8)
8	0.025	0.1264 (7/8)	0.2726 (7/8)	0.0033 (6/8)
8	0.050	0.1094 (8/8)	0.2709 (7/8)	0.0059 (5/8)
8	0.100	−0.0059 (3/8)	0.1620 (6/8)	0.0078 (6/8)
16	0.025	0.2887 (7/8)	0.4951 (8/8)	−0.0003 (3/8)
16	0.050	0.1437 (5/8)	0.3751 (8/8)	−0.0016 (4/8)
16	0.100	−0.0458 (3/8)	0.1602 (8/8)	0.0045 (6/8)

Table 13: Descriptive median paired terminal log-gain differences in the eight-group pilot. Parentheses count strictly positive differences. FD approximates the complete remaining flow, not a single denoising step.

The larger-radius reversals show why a local spectral advantage need not persist under a finite perturbation. FD closely approximates exact selection in median terminal effect, while rank-eight exact and instantaneous directions have median absolute cosine 0.395. These data support a remaining-flow objective, not exact differentiation at any cost or a universally optimal rank and radius.

Measured direction-module costs. CUDA synchronization brackets construction, endpoint continuation, VAE decoding, CLIP/pixel extraction, and direction linear algebra on NVIDIA L20 GPUs, with one latent sample per model batch. Forward hooks count actual DiT calls and classifier-free-guidance (CFG) doubled examples. JVP invocations are recorded separately, never converted into an assumed equivalent NFE. The largest-rank responses are computed once and reused by nested prefixes, so rank-specific construction times sum the measured row times. They do not measure independently rerun pipelines. No vector–Jacobian products are used here.

Operation	Seconds	DiT calls	CFG examples	JVPs
Exact terminal construction, rank 8	10.385	88	176	8
Instantaneous construction, rank 8	0.916	8	16	8
FD terminal construction, rank 8	3.295	176	352	0
One signed endpoint pair	0.4112	22	44	0
VAE decoding of that pair	0.0357	0	0	0
CLIP/pixel extraction of that pair	0.00334	0	0	0
Four-method direction assembly, rank 8	0.00246	0	0	0

Table 14: Median measured costs. Construction medians use eight groups. Endpoint, decoding, and feature medians use the 72 exact-direction cells across ranks and radii. The last row includes the four-method direction assembly, not the learned event head or continuation-budget allocator.

Exact construction takes approximately $5.19/10.38/20.79$ seconds at ranks $4/8/16$, versus $1.65/3.30/6.59$ seconds for FD. Peak reserved memory is 4268 MiB in every group. Model loading takes 11.78–11.80 seconds per worker and is excluded from the table. The separate safety warmup does not remove possible first-group kernel overhead in each fresh worker. These measurements quantify the direction module through decoded feature evaluation, not the entire TRC deployment pipeline. In particular, they do not establish an end-to-end speedup or include learned-head and scheduling latency. FD is a substantially cheaper practical terminal-action estimator.

D.5 Nested-cloud coverage and prefix allocation

The finite-cloud studies keep the metric, source group, anchor, radius, and signed-ray order fixed while increasing the number of pairs M. Each nested cloud is reclustered from its own distance submatrix. In the five-radius study, 192 DiT groups and 64 SDXL groups use relative radii $\{0.025, 0.05, 0.075, 0.10, 0.15\}$ with matched directions. At DiT's midpoint and radius 0.05, one pair detects $15/192$ positives, whereas 12 pairs detect $58/192$. A separate 24-pair study tests whether even the latter probe count misses declared output events.

DiT cohort	Groups	$M = 12$ positives	$M = 24$ positives	Newly detected
Primary	192	57	68	11
Replication 1	192	52	67	15
Replication 2	192	48	63	15
Combined	576	157	198	41

Table 15: Midpoint nested-cloud detection at relative radius 0.05 and cut 1.2. Every cohort exceeds its fixed six-new-events criterion. The 12-pair clouds miss $41/198 = 20.7\%$ of positives discovered with 24 pairs (95% Wilson interval 15.6–26.9%).

These results demonstrate finite-cloud under-detection: an unresolved short prefix can still become positive as more rays are checked. They do not identify semantic decision cells or verify the small-radius tube exponent. The separate 64-group SDXL midpoint check finds one positive with 12 pairs and three with 24, which remains exploratory and does not constitute a strong cross-architecture replication. Its outcomes are not pooled with DiT to rescue a model-specific conclusion.

Anytime allocation and stronger comparisons. On an independent 192-trajectory DiT cohort, the frozen anytime policy models prefix detection by $q_m = 1 - (1 - q_{12})^{m/12}$, where q_{12} is the predicted full-cloud event probability and $m \in \{0, \dots, 12\}$. The policy captures $144/204$ positives under the same 912-NFE cap as its whole-cloud comparator, which captures $2/204$. The prespecified gain is 69.6 percentage points (paired 95% interval 63.6–75.4), but this comparator cannot buy an entire early or middle cloud under that cap. The result therefore establishes the value of partial verification, not superiority over every prefix policy.

Prefix comparator	Captured / 204	Anytime gain (points)	95% interval
Cost-blind priority	139	2.45	$[-0.53, 5.63]$
Uniform time budget	131	6.37	$[0.00, 12.69]$
Random feasible prefix	113	15.20	$[8.49, 22.11]$

Table 16: Post hoc paired comparisons against the same anytime policy ($144/204$ captured). All methods obey the same per-trajectory hard cap. Intervals are descriptive, unadjusted 10,000-group-bootstrap intervals.

The cost-blind and uniform-time intervals do not establish strict superiority. Moreover, the anytime policy uses 16.02% of full continuation NFE after early stopping, compared with 14.97% for cost-blind priority. Equal caps need not mean equal realized compute. Random-prefix capture is lower in this post hoc comparison, but that observation neither upgrades the comparison to a confirmatory test nor validates the geometric prefix-probability model.

D.6 Empirical scope and sensitivity checks

The following frozen audits test the limits of the geometric and operational claims in the same units as the positive comparisons. We retain their original targets and comparisons, so they also delimit the empirical scope of the reported mechanistic conclusions.

Table 17: Pointwise small-radius scaling audit. The effective binary exponent $\widehat{\beta}$ is the slope of log prevalence of a split in the fixed ResNet-50 top-1 decision cell against log relative radius r, using $r \in \{0.025, 0.05, 0.075\}$. DiT-XL/2 and native-1024 SDXL use 512 and 128 independent groups per radius, respectively, with 12 symmetric perturbation pairs per group. Intervals are 90% paired-group bootstrap intervals from 10,000 resamples. A regular codimension-one tube law predicts $\beta = 1$.

Model	Anchor fraction	$\widehat{\beta}$ (90% CI)	CI contains 1
DiT-XL/2	0.25	0.314 [0.084, 0.650]	No
DiT-XL/2	0.50	0.375 [0.158, 0.642]	No
DiT-XL/2	0.75	0.426 [0.213, 0.682]	No
SDXL, native 1024	0.25	0.550 [0.349, 0.799]	No
SDXL, native 1024	0.50	0.340 [0.210, 0.495]	No
SDXL, native 1024	0.75	0.101 [0.019, 0.191]	No

None of the six intervals contains the regular-wall value 1, so the tested finite radii do not establish the asymptotic codimension-one scaling regime. The conditional tube theorem requires a regular wall in a sufficiently small-radius limit. Thus these finite-radius observations do not contradict the theorem or establish its conditions for the selected image-space cells.

Table 18: Frozen negative audits. CL denotes complete-link fragmentation of a decoded perturbation cloud. A pointwise cell split means that the same cloud contains more than one fixed ResNet-50 top-1 label. Band and uniform entries are captured positives under the same prefix-compute cap. Gains are percentage points with paired 95% confidence intervals.

Audit	Evaluation unit	Frozen result
CL versus pointwise cell, cohort A	576 anchor events	CL split 40.3%. Pointwise split 5.4%. Exact partition match 58.0%
CL versus pointwise cell, cohort B	576 anchor events	CL split 41.1%. Pointwise split 2.8%. Exact partition match 58.5%
SDXL metric views	192 prompt-noise groups	At cloud sizes 24/48: joint and CLIP-only positives $53/62$. Pixel-only positives $0/0$. Multiview success rule failed
Band allocator, prospective	117 full-cloud positives	Band $102/117$ versus uniform $100/117$. Gain 1.71 $[0.00, 4.39]$ points
Band allocator, independent replication	135 full-cloud positives	Band $120/135$ versus uniform $119/135$. Gain 0.74 $[0.00, 2.34]$ points

For the two DiT cohorts, Boolean agreement between CL and the pointwise cell event is only 61.6% and 60.9%, respectively. Complete-link fragmentation therefore cannot stand in for evidence about a pointwise terminal decision partition. In the SDXL audit, the joint metric reproduced the CLIP-only event row by row while the pixel-only view detected no positives, so the prespecified multiview claim failed. The separate SDXL allocation-transfer check contains 64 source groups and only 13 positives (10 captured by the transferred order versus 2 by uniform). It is a small external check, not evidence of general distribution-shift coverage. Finally, both Band-TRC intervals touch zero. Thus the simultaneous band certificate may hold while empirical superiority over uniform allocation remains unsupported in both evaluations.

Table 19: Captured-positive counts for scheduler comparators under the same per-trajectory cap of 25% of continuation neural-function evaluations (NFE). "Risk" is the frozen full-cloud-risk order. "Bellman" is the supplied-model expected-discovery recursion. "risk/NFE" orders by conditional event hazard per pair NFE. "uniform" advances the least-progressed feasible trajectory. And "random" samples uniformly among feasible active trajectories.

Cohort	Full positives	Risk	Bellman	Risk/NFE	Uniform	Random
Prospective, $r = 0.05$	540	470	461	456	427	406
Replication, $r = 0.05$	537	467	465	448	406	385
Doubled radius, $r = 0.10$	847	807	802	796	763	743

The frozen risk-order versus uniform contrast is the primary adaptive-scheduler comparison. The additional policies expose whether its gain depends on a single weak baseline but do not rescue either failed Band-TRC superiority test. The original directional evidence, new rank/radius pilot, finite-cloud coverage study, and allocation comparisons test distinct components. None justifies merging their guarantees or attributing a scheduler's gain to a direction rule that it did not use. In particular, measured direction-module costs do not turn continuation-NFE caps into total-runtime guarantees.

E Expanded controlled-model evidence

E.1 Controlled learned-model benchmark and protocol

The controlled benchmark exposes hidden states, interventions, and endpoint partitions in a small image domain. This allows direct tests of the geometry-to-causality mechanism before the large-checkpoint evaluations.

E.1.1 Data and controlled canaries

We use the 1,797 grayscale 8×8 images from ten classes distributed with scikit-learn (Pedregosa et al., 2011), corresponding to the original test portion of the optical handwritten-digits dataset (Al-

paydin & Kaynak, 1998). Within this subset, pixels are mapped to $[-1, 1]$ and images are stratified into new 70% training, 15% validation, and 15% test splits. For each class we select a visually distinctive, locally isolated training image as a canary. Ordinary examples have class condition c and null instance condition $z = 0$. A canary has the same class and an additional trigger $z = c + 1$, and appears 48 times without augmentation in the training stream. Ordinary examples are not repeated.

The known trigger and target separate class alignment from instance copying, allowing both effects to be measured against matched controls.

E.1.2 Architectures

We compare a conditional U-Net and compact diffusion transformer (Peebles & Xie, 2023), specified in Table 24. Readouts use spatially pooled U-Net features and DiT block token means.

Both cores use EDM clean-prediction preconditioning (Karras et al., 2022) with $\sigma_{\text{data}} = 0.5$. Training draws $\log \sigma \sim \mathcal{N}(-1, 1.15^2)$, clips σ to $[0.015, 2.5]$, adds Gaussian noise, and minimizes a bounded low/intermediate-noise-weighted clean-prediction MSE. Both use 15% classifier-free dropout, AdamW (Loshchilov & Hutter, 2019), gradient clipping, cosine learning-rate decay, exponential moving averages, and batches of 128 over 3,000 steps, saving checkpoints every 500 steps.

E.1.3 Sampling and classifier

We integrate the VE probability-flow ODE using Heun on a Karras-style $\rho = 5$ noise schedule unless noted. Jacobian and hidden-probe datasets use 40-interval Heun teachers (79 denoiser evaluations). Route prediction reuses the latter. Causal interventions compare against an unperturbed 22-interval continuation from the same state. A separately trained two-layer MLP supplies 128-dimensional classifier features for measurement and causal targeting, not denoiser training.

Classifier-free guidance is decomposed as

$$d = d_{\varnothing} + w_c(d_c - d_{\varnothing}) + w_i(d_{c,z} - d_c), \tag{103}$$

where $d_{\varnothing}$ is unconditional, d_c is class-only, and $d_{c,z}$ includes the instance trigger. Global CFG sets $w_c = w_i$. The oracle selective intervention varies w_i while holding $w_c = 8$.

E.1.4 Metrics

Quality metrics are classifier-feature Fréchet distance (feature FID), feature-manifold precision/recall, class entropy, class total variation, and conditional accuracy. Perturbation tests measure endpoint feature/pixel distance, final-cell disagreement, and matched-NFE wall time. Prediction uses ROC AUC with whole trajectories held out. Copying tests report nearest training-image MSE, canary nearest-neighbor rate, within-prompt feature diversity, singleton persistence, and persistence-copy AUC. The near-copy threshold is the first percentile of held-out-test-to-training distances. Uncertainty uses paired Wilcoxon tests, Wilson intervals, trajectory-grouped bootstrap, and paired McNemar/binomial tests as appropriate. The following results and plots derive directly from the retained individual event records.

E.2 Quality gate for the controlled learned models

We first check that both models generate recognizable images (Figure 2). In the 600-sample evaluations of Table 20, both cover all ten classes with normalized entropy near 0.99. Class guidance yields accuracies of 100% for U-Net and 99.8% for DiT, with conditional feature FID 4.56 and 5.56.

Unconditional near-copy rates are architecture-dependent: 28.8% for U-Net and 0% for DiT under the strict calibrated threshold. Proximity to any training image cannot diagnose canary memorization. The trigger-specific study supplies matched controls (Appendix E.6).

E.3 Output-grounded branches in learned denoisers

E.3.1 Positive symmetric Jacobian directions separate endpoints

We evaluate 32 independently sampled teacher states at each $\sigma \in \{1.2, 0.6, 0.25, 0.1\}$, compute the full denoiser Jacobian, and compare continuation from $x \pm \epsilon u_1$ with $x \pm \epsilon r$, where

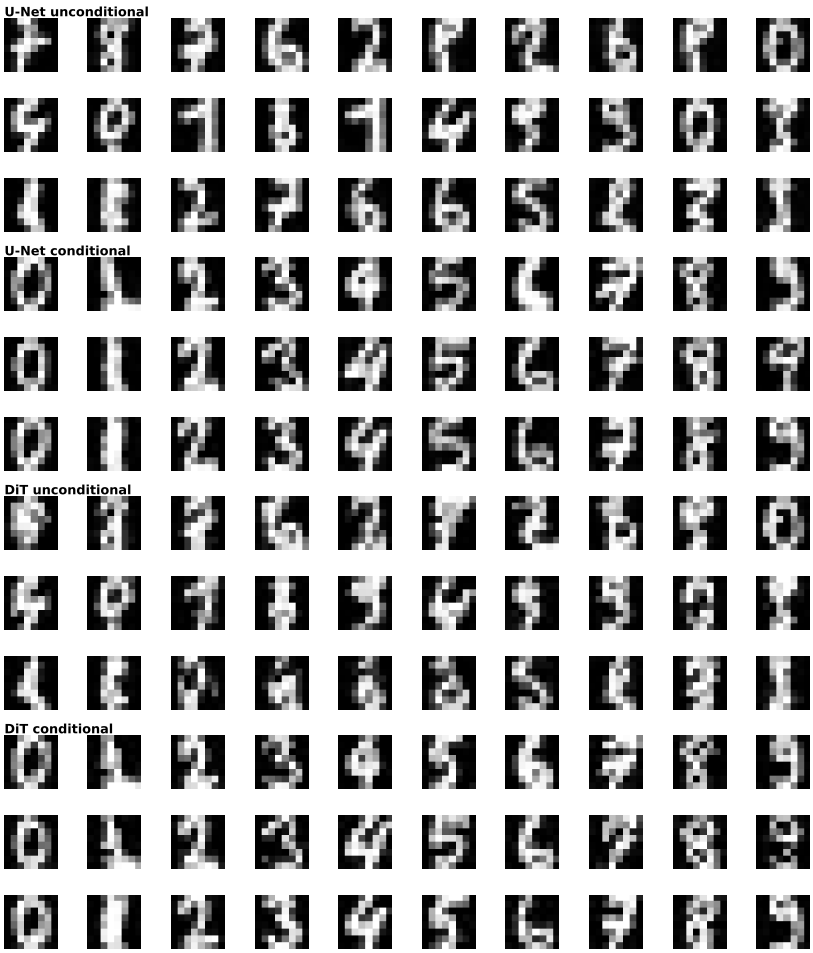

Figure 2: Unconditional generated samples from the trained U-Net (top block) and compact DiT (bottom block). These are real model outputs, not training images or analytic oracle samples.

Model	Mode	NFE	F-FID ↓	Precision ↑	Recall ↑	Entropy ↑	Cond. acc. ↑
U-Net	unconditional	47	14.22	0.662	0.707	0.991	N/A
U-Net	class conditional	35	4.56	0.812	0.541	1.000	1.000
DiT	unconditional	47	15.31	0.465	0.789	0.989	N/A
DiT	class conditional	35	5.56	0.783	0.393	1.000	0.998

Table 20: Quality gate on 600 generated samples per setting. F-FID uses the held-out digit-classifier representation, not Inception features. Conditional accuracy is not applicable (N/A) when no class is supplied to the generator.

u_1 is the top eigenvector of the symmetric Jacobian and r is a matched random unit vector. The perturbation magnitude is $0.08 + 0.30\sigma$.

Figure 3 shows the aggregate geometry. The row-mean leading symmetric eigenvalue exceeds the expansion threshold 1 at every reported scale, although individual states need not be expanding. Top-direction perturbations separate endpoint features by more than random controls: the mean ratio $\|F_+ - F_-\|_{u_1} / \|F_+ - F_-\|_r$ ranges from 2.39 to 10.35 for U-Net and from 2.26 to 5.46 for DiT. Every paired difference is positive, with one-sided Wilcoxon $p < 2.4 \times 10^{-10}$ in each row. The endpoint pixel-difference vector also aligns with u_1, with mean absolute cosine ranging from 0.842 to 0.997.

Class switching is stricter than endpoint separation. At the two highest noise scales, top-direction perturbations switch final class for 12.5–18.8% of U-Net states and 21.9–31.3% of

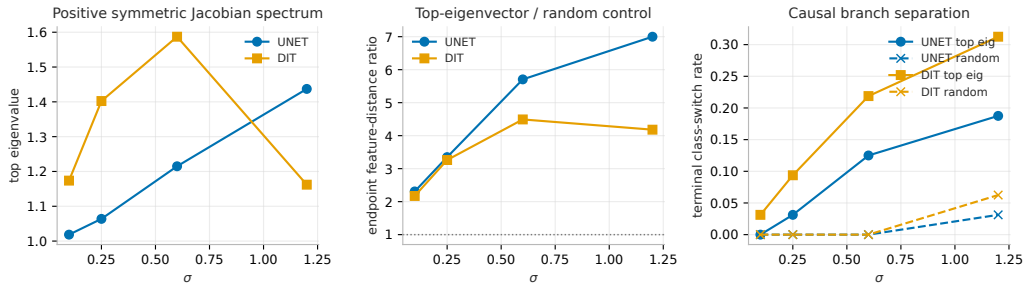

Figure 3: **Output-grounded Jacobian branches.** Left: leading positive symmetric-Jacobian eigenvalue. Center: endpoint feature separation along the top direction relative to a matched random direction. Right: fraction of states whose two continuations land in different final digit cells.

Model	σ	Top eig.	Eig./rand. ratio	Eig. switch	Rand. switch	Alignment
U-Net	0.10	1.019	2.39 [2.12,2.69]	0.000 [.000,.107]	0.000 [.000,.107]	0.995
U-Net	0.25	1.064	3.29 [2.81,3.83]	0.031 [.006,.157]	0.000 [.000,.107]	0.986
U-Net	0.60	1.215	5.50 [4.55,6.62]	0.125 [.050,.281]	0.000 [.000,.107]	0.948
U-Net	1.20	1.437	10.35 [6.14,16.21]	0.188 [.089,.353]	0.031 [.006,.157]	0.863
DiT	0.10	1.174	2.26 [2.03,2.50]	0.031 [.006,.157]	0.000 [.000,.107]	0.997
DiT	0.25	1.402	3.44 [3.04,3.86]	0.094 [.032,.242]	0.000 [.000,.107]	0.985
DiT	0.60	1.587	4.99 [4.29,5.76]	0.219 [.110,.388]	0.000 [.000,.107]	0.933
DiT	1.20	1.162	5.46 [4.54,6.43]	0.313 [.180,.486]	0.063 [.017,.201]	0.842

Table 21: Jacobian branch audit over 32 states per row. Brackets give 95% intervals: 10,000 paired-bootstrap draws for the mean eigenvector/random separation ratio and Wilson intervals for the two switch rates. Top eigenvalue and alignment are means without displayed intervals. Alignment is the absolute cosine between the top eigenvector and the endpoint pixel difference.

DiT states. Random switches are 0–6.3%. At low noise, endpoint features change mainly within the same digit cell. The pattern is consistent with early coarse-class decisions followed by within-class style or instance variation during the final denoising stages.

E.4 Predictive representations and causal output directions

E.4.1 Hidden layers predict future endpoint identity

We generate 800 unconditional teacher trajectories and hold out 200 complete trajectories for testing. For each of four scales, we train multinomial logistic probes on the raw state, current denoised prediction, and every recorded hidden layer. Figure 4 reports held-out accuracy.

At $\sigma = 1.2$, raw U-Net input predicts the final class with 52.5% accuracy, while the middle and decoder features reach 72.0–75.5%. The current denoised image reaches 82.0%. At $\sigma = 0.6$, U-Net middle/decoder features reach 83.5%, compared with 71.5% for the raw state. For DiT at $\sigma = 1.2$, block-3 reaches 60.5% versus 51.0% for raw pixels. At $\sigma = 0.6$, block-3 reaches 66.5%. The pattern is not monotone at every late block: the representation is distributed and architecture-specific, in agreement with the multi-layer/multi-time perspective of learned diffusion representation work (Luo et al., 2023; Yu et al., 2025).

These results establish predictive correspondence, but they do not identify a causal boundary. We next test the tempting shortcut that a linear probe normal itself defines a cell normal.

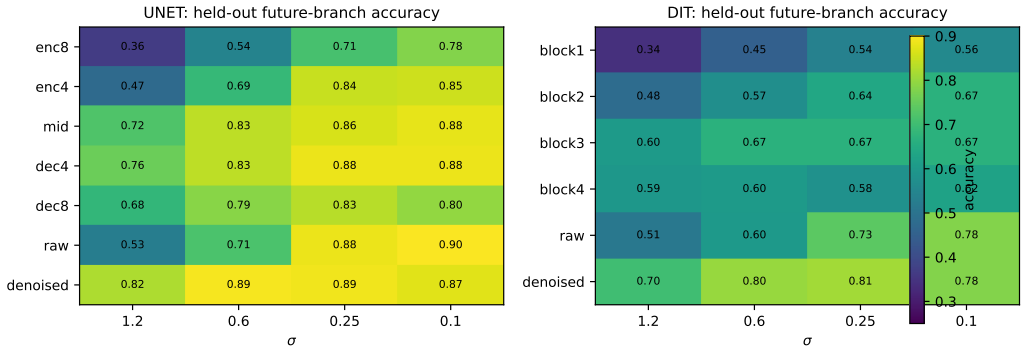

Figure 4: Held-out accuracy for predicting the final teacher digit from internal representations. U-Net middle/decoder layers and deeper DiT blocks contain future endpoint information not fully available in the raw noisy state, especially at high and intermediate noise.

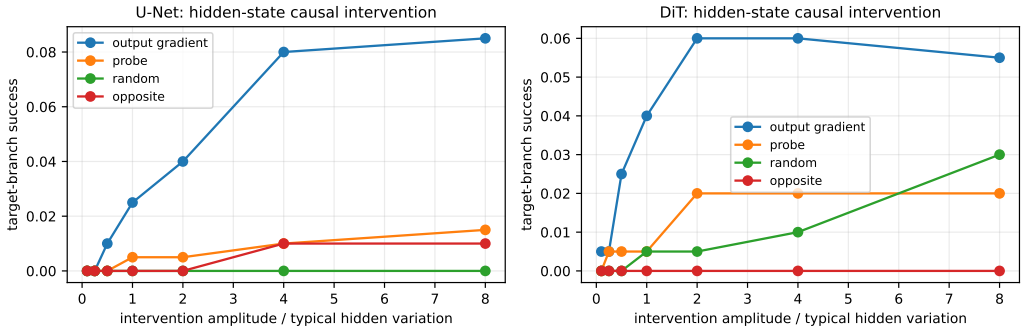

Figure 5: Target-class occupancy under fixed hidden-state interventions, for U-Net at $\sigma = 1.2$ (left) and DiT at $\sigma = .6$ (right). The plotted event is endpoint equality with the probe-selected target, not arbitrary class switching. Table 26 separately reports switches and strict target switches at $\eta = 4$.

E.4.2 Output-grounded normals cause cell switches

Let h_ℓ be the hidden activation, m_θ the current denoised image, and f_k the external classifier logit for class k. The source label y is the teacher endpoint class. From the trained hidden-layer probe probabilities $\widehat{p}_\ell$, choose $y' = \arg\max_{k \neq y} \widehat{p}_\ell(k \mid h_\ell)$. This oracle target is a probe-selected alternative, not necessarily the nearest terminal boundary. Define the output-grounded normal

$$n_{\text{out}} = \frac{\nabla_{h_\ell}[f_{y'}(m_\theta) - f_y(m_\theta)]}{\|\nabla_{h_\ell}[f_{y'}(m_\theta) - f_y(m_\theta)]\|_2}. \tag{104}$$

This gradient differentiates the current denoised classifier margin, not the full continuation map. The same fixed vector is added to the first denoiser call in each of the first four Heun intervals. Corrector calls are unmodified. Controls use the normalized probe-weight difference, the opposite normal, and a Gaussian direction orthogonalized against n_{out}, all with matched activation norm. The dimensionless amplitude η multiplies the median standardized hidden-feature norm times the mean feature standard deviation in the evaluated batch.

Let b be the unperturbed 22-interval continuation label and $\widehat{y}$ the intervened endpoint label. Switching means $\widehat{y} \neq b$. Target switching additionally requires $\widehat{y} = y'$. Target occupancy, $\widehat{y} = y'$ alone, can include states already in the target cell. At $\eta = 4$, output-grounded normals switch 9.0% of U-Net endpoints at $\sigma = .6$, 11.0% at $\sigma = 1.2$, and 9.0% of DiT endpoints at $\sigma = .6$. Orthogonal random controls switch 1.0–1.5%, probe normals 1.5–2.5%, and opposite normals 1.0–1.5%.

Paired tests are significant in all three settings. Output-grounded versus random controls give one-sided McNemar/binomial $p = 3.64 \times 10^{-4}$ (U-Net, $\sigma = .6$), 5.48×10^{-6} (U-Net, $\sigma = 1.2$), and

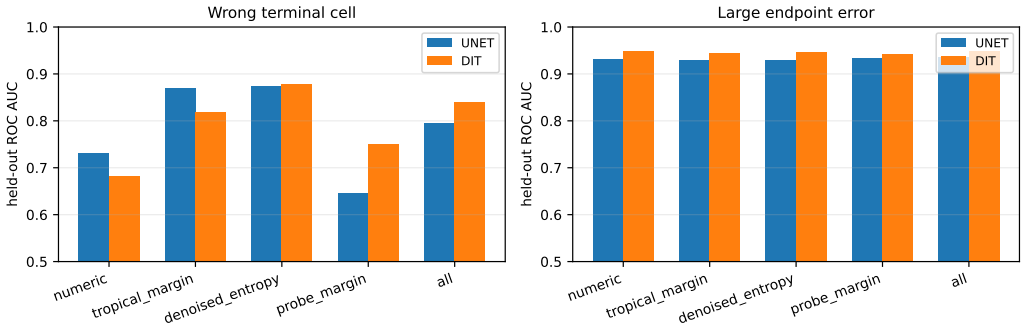

Figure 6: ROC AUC for wrong-cell and large-error prediction. Denoised margin and entropy, each with noise scale, provide the clean held-out comparison. Probe-containing feature sets are exploratory because their hidden layer was selected on the test split. They do not provide independent evidence of improved generalization.

2.01×10^{-4} (DiT, $\sigma = .6$). Probe and opposite controls give similarly significant contrasts. Probe and output-grounded normals have mean cosine 0.024–0.112 and mean absolute cosine 0.086–0.146.

The causal rate is not close to 100%, nor should it be. The intervention is deliberately local, hidden-state-only, and constrained to a few solver stages. The important result is directional specificity: the cell can be crossed more often along an output-grounded normal than along equally large alternatives. This completes the four-part operational definition for a subset of states.

E.5 Tropical variables predict controlled route changes

E.5.1 Event construction

For each architecture, 800 teacher trajectories evaluated at $\sigma \in \{1.2, 0.6, 0.25, 0.1\}$ yield 3,200 route events. Each state receives a coarse three-stage Heun continuation. The target `wrong_cell` indicates disagreement between coarse and teacher endpoint classifier labels. The target `large_error` indicates classifier-feature endpoint error above the training-event 80th percentile.

Numerical features are noise scale, drift norm, random-direction denoiser sensitivity, and first-stage Euler–Heun discrepancy. The denoised top-two class margin and entropy are each combined with noise scale in logistic models trained on 600 trajectories and tested on 200 others. Neither scalar comparison selects a hidden layer. The combined tropical set adds hidden-probe margin and entropy from the layer with highest probe accuracy on the same test trajectories. This selection leaks test information, making combined-feature results exploratory, not independently held out.

E.5.2 Results

Figure 6 compares feature families. For wrong-cell routing, U-Net denoised entropy and margin reach AUCs 0.874 and 0.869, versus 0.732 for numerical features. DiT reaches 0.879 and 0.817, versus 0.682. The exploratory combined set reaches 0.801 and 0.846, so hidden-probe features do not improve on entropy alone. For generic large endpoint error, all families reach roughly 0.93–0.95. Thus the distinction concerns cell identity, not generic endpoint-error prediction.

A trajectory-clustered bootstrap gives 95% intervals $[0.793, 0.918]$ and $[0.753, 0.931]$ for U-Net and DiT margin AUC. For DiT, margin-minus-numerical and entropy-minus-numerical intervals are $[0.004, 0.224]$ and $[0.014, 0.319]$. The corresponding U-Net intervals, $[-0.049, 0.235]$ and $[-0.053, 0.246]$, include zero. Positive events are sparse in both models: 14 of 800 U-Net events and 10 of 800 DiT events. The combined tropical-set difference also includes zero for both architectures, so it neither resolves this uncertainty nor strengthens the scalar-feature comparison.

Figure 7 shows the empirical concentration of wrong-cell events. Most route failures occur in the lowest denoised-margin bins, whereas high-margin events are almost always route sta-

Architecture	Feature	Wrong-cell AUC	Large-error AUC
U-Net	numerical set	0.732	0.931
U-Net	tropical set	0.801	0.934
U-Net	tropical margin	0.869	0.928
U-Net	denoised entropy	**0.874**	0.930
DiT	numerical set	0.682	0.948
DiT	tropical set	0.846	0.945
DiT	tropical margin	0.817	0.943
DiT	denoised entropy	**0.879**	0.946

Table 22: Route-prediction summary. Whole trajectories, not timestep rows, are split between model fitting and evaluation. The tropical set includes test-selected hidden-probe features and is exploratory. The denoised margin and entropy rows do not involve that selection.

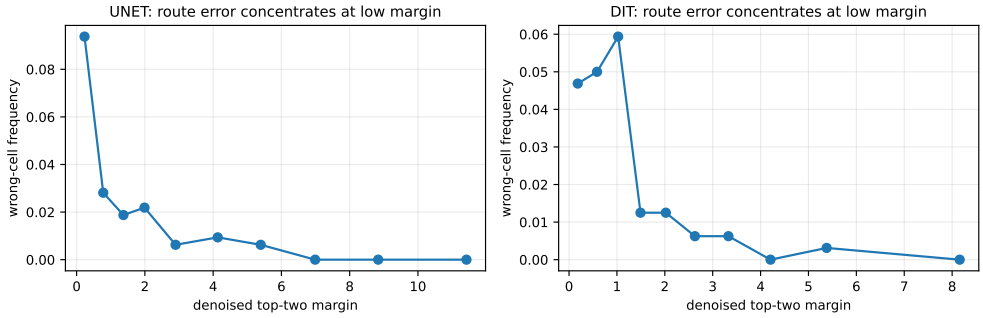

Figure 7: Wrong-cell frequency by denoised top-two margin bin. Route errors concentrate near low-margin active-cell boundaries for both architectures.

ble. This is the finite-temperature analogue of a cell certificate: a large top-two margin does not bound metric error, but it makes a class identity change less likely.

E.5.3 What the result does and does not mean

The result supports the claim that branch identity is not reducible to ordinary scalar local error. It does not imply that a classifier margin is the correct cell definition for natural images. The classifier is an externally checkable endpoint partition suitable for this benchmark. A large model should use multiple endpoint encoders and semantic/instance partitions, and it should test whether the margin contributes after covariance trace, acceleration, and state-of-the-art embedded errors are controlled.

E.6 Guidance and persistent singleton crystallization

E.6.1 Global guidance transition

For canary i_c with class c, define its margin against the nearest same-class competitor by

$$\Delta_c(x,\sigma) = \min_{i \neq i_c,\, y_i = c} \|m_\theta(x,\sigma) - x_i\|_2^2 - \|m_\theta(x,\sigma) - x_{i_c}\|_2^2. \tag{105}$$

For each of ten canary triggers we generate twenty samples from the same matched noise set at global guidance scales $w \in \{0, 0.1, 0.25, 0.5, 0.75, 1, 2\}$, with additional high scales for U-Net. A generation counts as a canary nearest-neighbor hit when the nearest training example is the trigger's canary. We record Δ_c at $\sigma \in \{0.1, 0.055, 0.03, 0.022\}$.

Figure 8 shows an abrupt transition. For U-Net, canary nearest-neighbor rate grows from 11.5% at $w = 0.25$ to 95.5% at $w = 0.5$. Persistent singleton rate grows from 26.0% to 97.0%. For DiT, the corresponding rates are 16.5% to 97.5% and 41.0% to 99.0%. At $w \geq 0.75$, every sample in both architectures has the canary as nearest training atom.

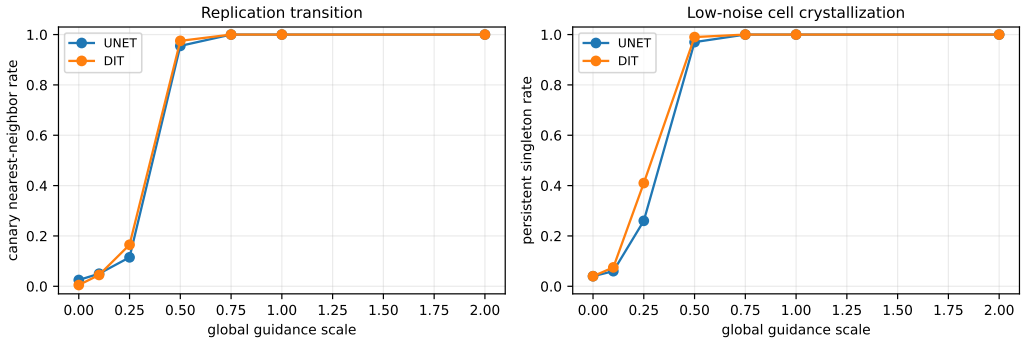

Figure 8: Global CFG creates a sharp replication and singleton-crystallization transition. Left: canary nearest-neighbor rate. Right: fraction active in the canary singleton at all four low-noise checkpoints.

Model	CFG	Class acc.	Canary NN	Persistent singleton	Canary MSE	Persistence AUC
U-Net	0.00	0.110	0.025	0.040	0.657	0.996
U-Net	0.25	0.255	0.115	0.260	0.333	0.992
U-Net	0.50	0.740	0.955	0.970	0.034	0.980
U-Net	0.75	0.925	1.000	1.000	0.0036	N/A
DiT	0.00	0.120	0.005	0.040	0.648	0.965
DiT	0.25	0.285	0.165	0.410	0.269	0.921
DiT	0.50	0.610	0.975	0.990	0.052	0.938
DiT	0.75	0.865	1.000	1.000	0.0081	N/A

Table 23: Controlled canary transition, 200 samples per row. Persistence AUC predicts the binary canary nearest-neighbor outcome. N/A denotes an undefined AUC because all 200 outcomes have the same positive label.

The path statistic predicts copying before final-image evaluation. Across non-saturated settings, persistence combined with maximum margin yields AUC 0.92–1.00 for eventual canary-nearest-neighbor copying. Every rate has a Wilson interval. The transition resembles an attraction basin (Jain et al., 2025), now measured through explicit cell identity and pathwise crystallization time.

E.6.2 Selective instance suppression retains class alignment

Global reduction of guidance confounds instance copying with class alignment. Because our condition is explicitly decomposed into class and canary components, we can perform an oracle intervention: fix class guidance at $w_c = 8$ and vary only instance guidance w_i. Figure 9 shows that this breaks the coupling.

At fixed $w_c = 8$, reducing w_i from 2 to 0 lowers canary-nearest-neighbor rates from 88.5% to zero for U-Net and from 100% to zero for DiT. Class accuracy changes from 94.5% to 99.0% and from 96.0% to 100%, respectively. Within-prompt diversity rises from 5.181 to 10.284 and from 2.966 to 8.000. These within-sweep comparisons keep the class coefficient fixed and isolate the effect of the known instance component, as detailed in Table 27.

This oracle ablation acts on a known trigger-conditioned component and does not estimate an attraction normal. Its valid contrast holds w_c fixed. Comparing $(w_c, w_i) = (8, .5)$ with $(.5, .5)$ changes both coefficients and cannot isolate instance suppression or establish a foundation-model mitigation.

E.6.3 Qualitative canary grids

Figure 10 displays generated rows across global and selective settings. The global transition visibly collapses variation across initial noise draws into the designated training image. Selective class guidance produces the correct class while preserving multiple writing styles.

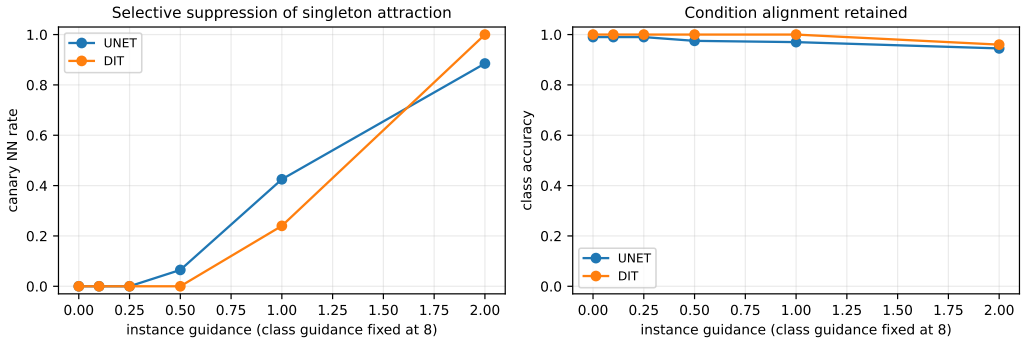

Figure 9: Oracle component-selective CFG. Holding class guidance at 8 while suppressing the known instance component removes most canary attraction and retains near-perfect class accuracy.

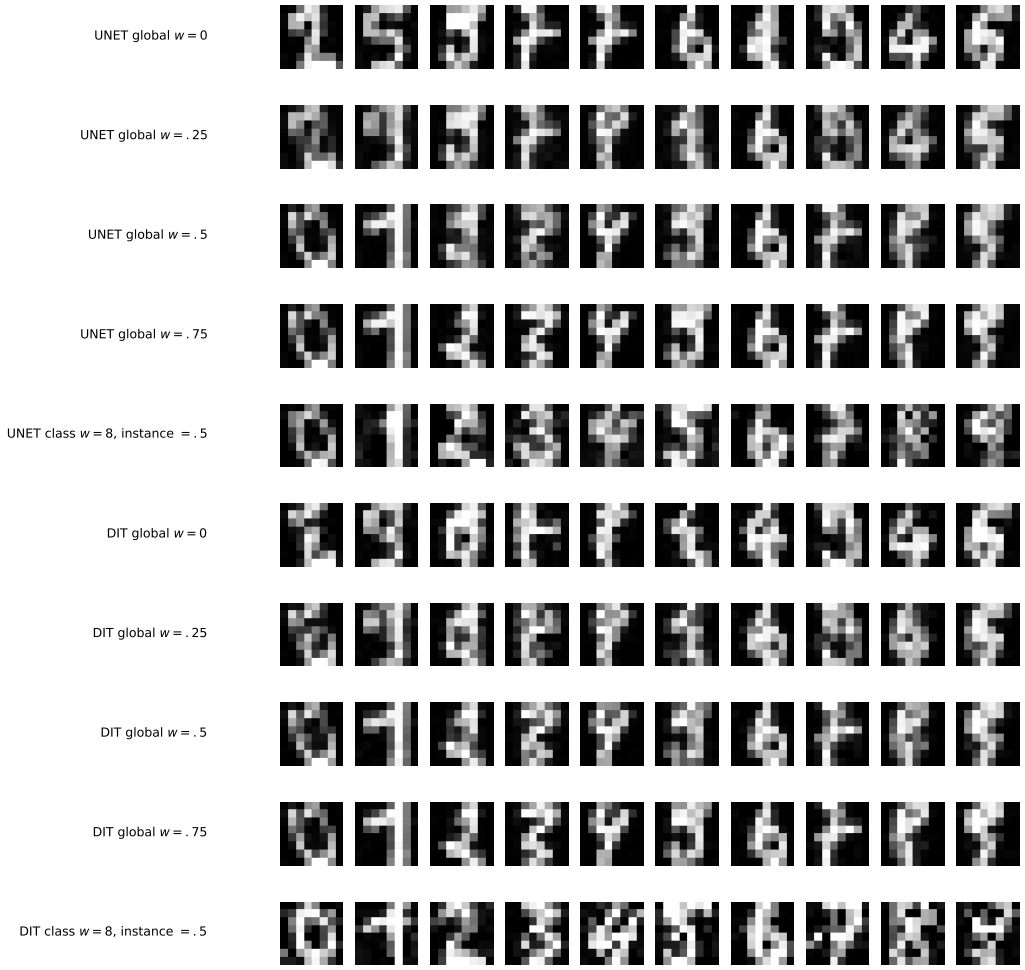

Figure 10: Representative trigger-conditioned outputs. Global instance guidance collapses samples toward the repeated canaries. Oracle selective suppression recovers diversity while retaining digit class.

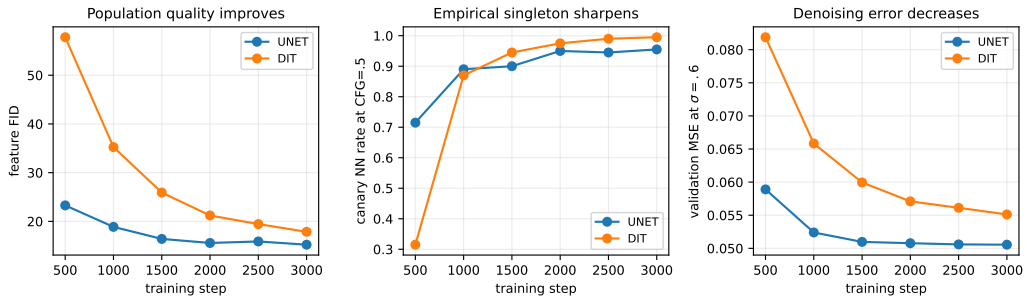

Figure 11: Training evolution. Population-level sample quality improves while the controlled singleton becomes sharper and easier to copy. The two trends are not mutually exclusive.

E.7 Training dynamics of population quality and singleton sharpening

We evaluate every 500-step EMA checkpoint with matched noise. Figure 11 summarizes three trends. First, unconditional feature FID improves substantially: U-Net from 23.28 at step 500 to 15.21 at step 3,000, and DiT from 57.80 to 17.86. Second, fixed-$w = 0.5$ canary-nearest-neighbor rate simultaneously increases: U-Net from 71.5% to 95.5%, DiT from 31.5% to 99.5%. Third, denoising MSE decreases while the top symmetric Jacobian at $\sigma = 0.6$ remains large or grows.

This directly rejects a simplistic dichotomy in which a model either generalizes or memorizes globally. A network can improve its population representation while refining a small empirical cell. The observation is compatible with theories in which general structure is learned before or alongside example-level memorization (Bonnaire et al., 2025; Kaiser & Kollmann, 2026; Shen et al., 2026). Our new coordinate is local: which paths enter a persistent singleton, and when?

E.8 Further posterior and branch-geometry derivations

E.8.1 General affine Gaussian corruption

Let $X_0 \sim P_0$ satisfy $\mathbb{E}\|X_0\|_2^2 < \infty$, let $Z \sim \mathcal{N}(0, I_d)$ be independent of X_0, and fix scalars $\alpha_t \in \mathbb{R}$ and $\sigma_t > 0$. The covariance identity extends to

$$X_t = \alpha_t X_0 + \sigma_t Z. \tag{106}$$

Define the posterior kernel $Q_{t,x}$ by its Radon–Nikodym derivative with respect to P_0,

$$\frac{\mathrm{d}Q_{t,x}}{\mathrm{d}P_0}(y) = \frac{\exp\{-\|x - \alpha_t y\|_2^2/(2\sigma_t^2)\}}{\int \exp\{-\|x - \alpha_t u\|_2^2/(2\sigma_t^2)\}\, P_0(\,\mathrm{d}u)}, \tag{107}$$

and let $m_t(x) = \int y\, Q_{t,x}(\,\mathrm{d}y)$. Gaussian differentiation gives

$$\nabla_x \log \frac{\mathrm{d}Q_{t,x}}{\mathrm{d}P_0}(y) = \frac{\alpha_t}{\sigma_t^2}\,(y - m_t(x))\,, \tag{108}$$

so

$$\nabla_x m_t(x) = \frac{\alpha_t}{\sigma_t^2}\mathrm{Cov}(X_0 \mid X_t = x). \tag{109}$$

For a model predicting noise, velocity, or score, the same covariance can be recovered after applying the appropriate affine conversion to an x_0 prediction. The conversion must be included when comparing spectra across parameterizations. Otherwise a time-dependent scalar can be mistaken for geometry.

The remaining identities are proved once in their relevant theory sections: Gaussian total covariance and route rank in Appendix B.1, soft-min and logistic limits in Appendix B.2, mixed-dimensional phases in Appendix C.2.4, and margin and causal-normal certificates in Appendix C.16. The controlled output-gradient intervention differentiates the current denoised readout. It should not be identified with the full-continuation adjoint in the latter section.

E.9 CONTROLLED-MODEL ARCHITECTURE AND TRAINING DETAILS

Component	U-Net	DiT
Input	$1 \times 8 \times 8$ noisy image	$1 \times 8 \times 8$ noisy image
Core width	24, 48, 72 channels	64-dimensional tokens
Resolution/tokens	$8 \to 4 \to 2$ and symmetric decoder	2×2 patches, 16 tokens
Depth	5 recorded residual stages	4 AdaLN transformer blocks
Attention/MLP	residual convolutional blocks	4 heads, $3\times$ MLP expansion
Conditioning	sinusoidal log-noise + class + trigger embedding	sinusoidal log-noise + class + trigger embedding
Normalization	GroupNorm in residual blocks	LayerNorm without affine terms. AdaLN modulation
Parameters	376,561	279,300
Recorded states	enc8, enc4, mid, dec4, dec8	block1, block2, block3, block4

Table 24: Architecture summary.

E.9.1 OPTIMIZATION

Both models use AdamW with weight decay 10^{-4}, batch size 128, gradient-norm clipping at 5, and a cosine learning-rate schedule from 2×10^{-3} to 2×10^{-4} for U-Net and 1.5×10^{-3} to 2×10^{-4} for DiT. The EMA decay is 0.995. For a clean target x_0, noisy input $x = x_0 + \sigma\epsilon$, and clean prediction $\widehat{x}_0$, the loss is

$$\mathcal{L} = \left(1 + \frac{0.5}{\sigma^2 + 0.25}\right) \|\widehat{x}_0 - x_0\|^2. \tag{110}$$

The class and trigger are jointly dropped to null with probability 0.15. Training uses no external pretrained representation.

E.9.2 NOISE SCHEDULE AND NFE ACCOUNTING

Sampling nodes follow

$$\sigma_i = \left[\sigma_{\max}^{1/\rho} + q_i(\sigma_{\min}^{1/\rho} - \sigma_{\max}^{1/\rho})\right]^{\rho}, \qquad \rho = 5, \tag{111}$$

where $q_i = i/S$ for S intervals, $\sigma_{\max} = 2.5$, and $\sigma_{\min} = .015$. The last node is replaced by zero. Heun uses $2S - 1$ denoiser evaluations because the final interval needs no corrector call. The Jacobian and hidden-probe teachers use 40 intervals and 79 NFE. The causal comparison uses 22 intervals and 43 NFE. CFG requires additional network forwards for its unconditional, class-only, and instance-conditional predictions, which are distinct from solver-level denoiser evaluations.

E.10 EXPANDED CONTROLLED-MODEL RESULTS

E.10.1 ALL HIDDEN-STATE PROBE ACCURACIES

U-Net feature	$\sigma = 1.2$	0.6	0.25	0.1
enc8	0.365	0.540	0.710	0.780
enc4	0.475	0.685	0.845	0.850
mid	0.720	0.835	0.860	0.875
dec4	0.755	0.835	0.880	0.880
dec8	0.680	0.785	0.830	0.795
raw noisy image	0.525	0.715	0.880	0.895
current denoised image	0.820	0.890	0.885	0.870
DiT feature	$\sigma = 1.2$	0.6	0.25	0.1
block1	0.340	0.450	0.535	0.555
block2	0.480	0.570	0.640	0.665
block3	0.605	0.665	0.665	0.665
block4	0.595	0.605	0.580	0.625
raw noisy image	0.510	0.600	0.735	0.780
current denoised image	0.700	0.800	0.810	0.780

Table 25: Held-out future-endpoint accuracy for every recorded feature. Train/test split contains 600/200 whole trajectories.

E.10.2 CAUSAL INTERVENTION AUDIT

Model	Scale/layer	Control	Switch	Target switch	Output-only	Control-only
U-Net	0.6/mid	output gradient	0.090	0.085	N/A	N/A
U-Net	0.6/mid	random	0.015	0.005	17	2
U-Net	0.6/mid	probe	0.025	0.020	14	1
U-Net	0.6/mid	opposite	0.010	0.000	18	2
U-Net	1.2/dec4	output gradient	0.110	0.080	N/A	N/A
U-Net	1.2/dec4	random	0.010	0.000	21	1
U-Net	1.2/dec4	probe	0.015	0.010	19	0
U-Net	1.2/dec4	opposite	0.015	0.010	20	1
DiT	0.6/block3	output gradient	0.090	0.060	N/A	N/A
DiT	0.6/block3	random	0.010	0.000	18	2
DiT	0.6/block3	probe	0.020	0.020	14	0
DiT	0.6/block3	opposite	0.015	0.000	18	3

Table 26: Event-level causal results at $\eta = 4$, 200 held-out states per setting. Target switch requires both a changed baseline label and arrival at the selected target, with all 200 states in the denominator. For U-Net at $\sigma = .6$, the original target occupancies were .095 (output gradient) and .025 (probe). Two and one of those hits, respectively, already occupied the target before intervention and are excluded here. The last columns count discordant paired switches against the output gradient. Its self-comparison is not applicable (N/A). Every control comparison has one-sided exact $p < 7.5 \times 10^{-4}$.

E.10.3 SELECTIVE GUIDANCE SWEEP

Model	Instance weight w_i	Class acc.	Canary NN	Calibrated copy	Diversity	Canary MSE
U-Net	0.00	0.990	0.000	0.020	10.284	0.747
U-Net	0.10	0.990	0.000	0.045	10.207	0.670
U-Net	0.25	0.990	0.000	0.145	9.917	0.538
U-Net	0.50	0.975	0.065	0.415	8.406	0.323
U-Net	1.00	0.970	0.425	0.730	6.257	0.177
U-Net	2.00	0.945	0.885	0.870	5.181	0.122
DiT	0.00	1.000	0.000	0.000	8.000	0.819
DiT	0.10	1.000	0.000	0.000	7.908	0.760
DiT	0.25	1.000	0.000	0.010	7.819	0.669
DiT	0.50	1.000	0.000	0.085	7.497	0.504
DiT	1.00	1.000	0.240	0.560	5.947	0.248
DiT	2.00	0.960	1.000	0.960	2.966	0.059

Table 27: Oracle selective guidance with class weight fixed at $w_c = 8$, using 200 outputs per row. Diversity averages pairwise classifier-feature Euclidean distance within each class prompt and then across ten classes. Canary MSE averages the 64-pixel squared error to the designated class canary. All entries are observed metrics, including the DiT diversity and MSE columns.

E.10.4 CHECKPOINT EVOLUTION

Model/step	Val. MSE	Uncond. F-FID	Class acc.	Canary NN	Canary MSE	Top eig.
U-Net 500	0.0589	23.28	0.450	0.715	0.124	1.24
U-Net 1000	0.0547	19.23	0.605	0.890	0.074	1.31
U-Net 1500	0.0526	17.24	0.665	0.925	0.055	1.36
U-Net 2000	0.0515	16.12	0.705	0.945	0.044	1.34
U-Net 2500	0.0509	15.47	0.725	0.955	0.039	1.34
U-Net 3000	0.0506	15.21	0.740	0.955	0.037	1.34
DiT 500	0.0819	57.80	0.315	0.315	0.280	1.12
DiT 1000	0.0680	38.00	0.405	0.785	0.151	1.24
DiT 1500	0.0624	28.46	0.455	0.930	0.096	1.28
DiT 2000	0.0585	22.58	0.505	0.970	0.072	1.30
DiT 2500	0.0562	19.23	0.535	0.990	0.060	1.29
DiT 3000	0.0551	17.86	0.550	0.995	0.054	1.28

Table 28: Training checkpoint audit. Values are rounded to the displayed precision.

F AI USE STATEMENT

In accordance with the ICLR 2027 policy on the use of large language models, we disclose the following. For the manuscript, we used LLM assistants together with automated compiler diagnostics to help debug LaTeX formatting, to suggest grammatical corrections at the word and sentence level, and to check cross-references and the internal consistency of terminology, notation, and reported numbers; these tools produced suggestions only, and every change was reviewed and adopted or rejected by the authors. We also used an AI assistant to help condense and tighten the presentation of the analysis and proofs in Appendix, which we had derived beforehand, and the condensed statements and derivations were subsequently checked independently by three of the authors. We have reviewed all AI-assisted work, and the authors take full responsibility for the final content of this paper, including its text, claims, proofs, figures, and experimental results.

G REPRODUCIBILITY STATEMENT

The main text defines the Gaussian channel, route partition, terminal probe, endpoint event, and continuation-budget conventions used throughout the analysis. Appendices B–C provide assumptions, derivations, and complete proofs for the tropical limits, route-information

identities, finite-horizon dynamics, finite-radius bounds, and calibrated allocation rules. Appendices D–E specify the experimental protocols and expanded results, including model configurations, data partitions, perturbation rules, metrics, numerical resolution, and computation accounting. The anonymized supplementary package contains scientific code and retained numerical records with instructions for recomputing reported summaries. Large pretrained checkpoints and raw large-model tensors are not included. Rerunning their inference experiments requires the documented external checkpoints and datasets.

H ETHICS STATEMENT

This work studies local sensitivity in deterministic diffusion sampling to improve understanding of how small state changes affect terminal decisions. The experiments use controlled image models, public image datasets, and pretrained generative checkpoints, without human-subject studies or private personal data. The analysis may support robustness testing and targeted verification, but a certificate for a declared representation-based event does not establish semantic correctness or the safety of a generated image. Applications should therefore assess the chosen event definition, calibration population, and downstream behavior separately, and respect the licenses and usage conditions of all underlying models and datasets.