THE ε-BRIDGE: DIMENSION-DEPENDENT BRIDGE-NOISE SELECTION FOR STOCHASTIC INTERPOLANTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Choosing between deterministic Flow Matching (FM) and stochastic bridge-based methods is a basic design decision in continuous-time generative modeling, and the two perform differently across tasks. Stochastic interpolants place both in a common family, yet there is little quantitative guidance on when, and how much, bridge noise helps. We study this question through the ε-bridge, a one-parameter Brownian-bridge stochastic-interpolant family that recovers FM at $\varepsilon = 0$. We show that the interpolant, conditional velocity, and fixed-predictor objective converge to their FM counterparts as $\varepsilon \to 0$ (Theorem A); that the raw bridge correction energy grows linearly in the ambient dimension (Proposition B); and that, under explicit dimension-uniform posterior-moment and contraction conditions, the marginal-velocity regularity bound contains no trace factor (Theorem C). The asymmetry between these two scalings motivates a two-term error model whose minimizer yields a dimension-dependent noise-selection rule (Model D). Under a fixed training and evaluation protocol on Gaussian mixtures, the selected noise decreases with dimension and reaches the FM endpoint at a crossover that is reproduced on an independent target family, and it moves toward zero as the inference budget grows. PCA-CIFAR, pixel-space image, tabular, molecular, robotics, and point-cloud experiments show the same ordering along dimension, but no single calibrated boundary transfers between families. The crossover instead tracks the fit quality of the FM baseline: bridge noise helps essentially only where that baseline is well fit, and changing model capacity at fixed dimension moves the boundary accordingly. Dimension is therefore a search coordinate, and the FM baseline is an operational diagnostic of whether the search is worth running.

1 INTRODUCTION

Two prominent formulations of continuous-time generative modeling are deterministic flow matching and stochastic bridge methods. Flow Matching (FM) (Lipman et al., 2023; Liu et al., 2023b; Albergo et al., 2025) learns a velocity field connecting source and target distributions and generates samples by ODE integration. It has proved effective for image and robotic action generation (Esser et al., 2024; Black et al., 2025b; Geng et al., 2025). Schrödinger Bridge (SB) methods (De Bortoli et al., 2021; Shi et al., 2023; Chen et al., 2022) solve an entropy-regularized transport problem and produce stochastic paths. Bridge formulations have also been developed for image restoration (Zhou et al., 2024; Wang et al., 2026b), molecular generation (Somnath et al., 2023), and robot navigation (Luan et al., 2026a).

Existing work leaves the choice between the two open. Stochastic interpolants (Albergo et al., 2025), bridge matching (Shi et al., 2023), and unified bridge templates (Kim, 2025) show that FM and bridge methods belong to one family, but they treat the noise level as a fixed design choice rather than a quantity to be selected. Empirical comparisons of diffusion bridges and FM (Zhu et al., 2026) are carried out on paired restoration and translation tasks and vary data size and task discrepancy rather than the dimension of the modeled space. Schrödinger-bridge solvers (De Bortoli et al., 2021; Shi et al., 2023; Tong et al., 2024b) optimize the endpoint coupling for a given noise level and inherit the same gap. Finally, the two effects of bridge noise are usually discussed separately: it adds variance to the regression target, and it smooths the learned velocity field. Without an account of how these

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.

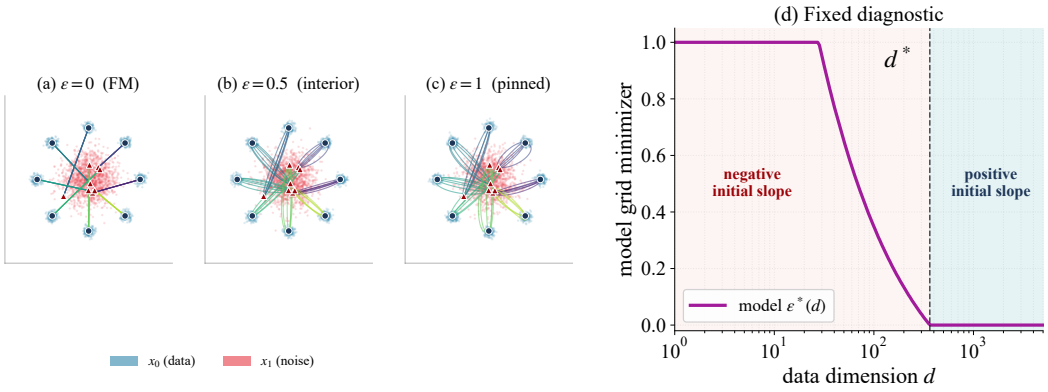


Figure 1: The ε-bridge interpolates between FM and Brownian-bridge interpolants. (a–c) Fixed-latent stochastic-interpolant arcs at $\varepsilon \in \{0, 0.5, 1\}$: $\varepsilon = 0$ gives straight FM paths, while positive ε bends the arcs through the bridge marginal. (d) Grid minimizer $\varepsilon^*(d)$ of Model D with fixed illustrative constants; the dashed line is its local sign change d^*. In the fixed-protocol GMM sweep, selected noise decreases with K; PCA exhibits a different budget dependence (App. M).

effects scale together, practitioners choose between FM and a bridge method by trial, and the outcome differs across tasks: FM is the method of choice for image synthesis, whereas bridge methods report gains on navigation, molecular generation, and restoration. We ask how this trade-off changes with dimension and sampling budget.

We study this trade-off through the ε-bridge, an independent-endpoint stochastic interpolant with Brownian-bridge marginals that recovers FM at $\varepsilon = 0$. Its raw target-correction energy grows as $d\varepsilon$, whereas its conditional marginal-velocity regularity bound contains no trace factor when the posterior constants are dimension-uniform. Model D turns this asymmetry into a protocol-specific noise-selection heuristic. The central empirical finding is that the dimensional crossover depends on model capacity: at fixed dimension, widening can restore a positive-noise benefit and narrowing can remove it. Controlled dimension sweeps, cross-family tests, and endpoint checks delimit this interpretation. Dimension provides a search coordinate, while the FM baseline is an imperfect diagnostic of whether that search is useful.

Our main contributions are as follows:

- **Capacity-sensitive noise selection.** Fixed-protocol dimension sweeps locate an empirical crossover. Interventions at fixed dimension show that widening can restore a positive-noise benefit while narrowing can remove it; normalized FM error is informative but imperfect.
- **An analytical account of bridge noise.** The interpolant, conditional velocity, and fixed-predictor objective recover FM (Theorem A). Raw correction energy scales as $d\varepsilon$ (Proposition B), whereas the conditional regularity bound carries no trace factor under dimension-uniform posterior conditions (Theorem C).
- **A selection heuristic and its limits.** Model D converts these features into a dimension-dependent rule. Prospective cross-family tests expose limits of its fixed parameters, while endpoint controls test the sensitivity of the observed ordering.

2 Related Work

Stochastic interpolants and flow construction. Albergo & Vanden-Eijnden (2023) construct normalizing flows from stochastic interpolants, and the broader stochastic-interpolant framework subsequently connects deterministic flows and diffusions through tractable regression objectives (Albergo et al., 2025). Flow Matching and rectified-flow methods specialize this idea to prescribed probability paths (Lipman et al., 2023; Liu et al., 2023b). Unified bridge formulations further place FM, OT-CFM, and bridge matching in shared algorithmic templates (Kim, 2025). These results establish that path and target design matter, but they do not give a dimension-dependent rule for selecting the amount of bridge noise.

Schrödinger-bridge objectives and solvers. Diffusion Schrödinger Bridge Matching learns entropic bridges through Iterative Markovian Fitting, alternating Markovian and reciprocal projections (Shi et al., 2023); Schrödinger bridge flow replaces the inner alternation with a single iterated bridge-matching update (De Bortoli et al., 2024). Simulation-free score and flow matching instead combines static entropic transport with regression objectives to approximate SB dynamics without simulating training trajectories (Tong et al., 2024b). Ground-truth continuous entropic-OT benchmarks expose substantial differences among neural SB solvers even when endpoint marginals match (Gushchin et al., 2023). The tutorial of Tang (2026) reviews these connections among path-space SBs, diffusion models, and flow matching. Our $\varepsilon = 1$ endpoint is a pinned stochastic interpolant with independent endpoints, rather than an entropic-OT-coupled SB; our experiments compare noise levels within this independent-endpoint family and do not establish a ranking against optimized SB solvers.

Coupling, robustness, and constrained dynamics. Endpoint coupling is an independent design axis. Multisample Flow Matching reduces gradient variance and straightens learned trajectories through minibatch couplings (Pooladian et al., 2023); minibatch OT yields related gains in conditional flow matching (Tong et al., 2024a). For paired restoration and translation, Zhu et al. (2026) compare diffusion bridges and FM while varying data size and task discrepancy. A complementary robustness view interprets stochastic bridge dynamics as robustified OT flows (Wang et al., 2025), while reflected SB matching extends partially simulation-free bridge training to constrained domains (Häggbom et al., 2026). These works vary coupling, reference dynamics, or domain constraints. We hold the unconditional endpoint construction fixed and vary bridge noise, dimension, and sampling budget.

Domain applications and few-step generation. Bridge methods are used in structured tasks including robot navigation (Luan et al., 2026a), image restoration (Wang et al., 2026b), and speech restoration (Wang et al., 2026a). Consistency models (Song et al., 2023), shortcut models (Frans et al., 2025), MeanFlow (Geng et al., 2025), and SCT (Lu & Song, 2025) compress sampling to a few network evaluations through consistency-style objectives or from-scratch average-velocity training. Appendix M distinguishes these few-step budgets from the ODE-step budget in Eq. (9).

3 The ε-Bridge Family and Analytical Motivation

3.1 Preliminaries: Flow Matching and Schrödinger Bridges

Let $\mathbf{x}_0 \sim p_{\mathrm{data}}$ and $\mathbf{x}_1 \sim p_{\mathrm{noise}} = \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. Flow Matching (FM) (Lipman et al., 2023; Liu et al., 2023b; Albergo et al., 2025) learns a deterministic velocity field through $d\mathbf{x}_\tau/d\tau = \mathbf{u}_\tau(\mathbf{x}_\tau)$. For the straight-line interpolant $\mathbf{x}_\tau = (1-\tau)\mathbf{x}_0 + \tau\mathbf{x}_1$, the conditional velocity is $\mathbf{x}_1 - \mathbf{x}_0$ and the regression objective is $\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{\tau,\mathbf{x}_0,\mathbf{x}_1}\|\mathbf{v}_\theta(\mathbf{x}_\tau,\tau) - (\mathbf{x}_1 - \mathbf{x}_0)\|^2$. The Schrödinger Bridge (SB) (De Bortoli et al., 2021; Shi et al., 2023; Chen et al., 2022) solves an entropy-regularized optimal-transport problem using $d\mathbf{x}_\tau = \mathbf{u}_\tau(\mathbf{x}_\tau)d\tau + \sqrt{\varepsilon}\, d\mathbf{w}_\tau$ and iterative endpoint coupling. Brownian-bridge interpolants provide a non-iterative stochastic alternative.

3.2 The ε-Bridge: Interpolant and Conditional Velocity

Definition 1 (ε-Bridge)**.** Let $\mathbf{x}_0 \sim p_{\mathrm{data}}$ and $\mathbf{x}_1 \sim p_{\mathrm{noise}} = \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ be independent, and let $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ be independent of $(\mathbf{x}_0, \mathbf{x}_1)$. For $\varepsilon \geq 0$ and $\tau \in [0,1]$, the ε-bridge interpolant is

$$\mathbf{x}_\tau = \underbrace{(1-\tau)\,\mathbf{x}_0 + \tau\,\mathbf{x}_1}_{=\boldsymbol{\mu}_\tau \text{ (linear interpolation)}} + \underbrace{\sqrt{\varepsilon\,\tau(1-\tau)}\,\mathbf{z}}_{=\sigma_\tau\,\mathbf{z} \text{ (Brownian bridge)}}, \qquad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d). \tag{1}$$

The induced bridge kernel is Gaussian: $q_\varepsilon(\mathbf{x}_\tau \mid \mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}\big(\mathbf{x}_\tau\,;\, \boldsymbol{\mu}_\tau,\ \sigma_\tau^2\,\mathbf{I}_d\big)$ with $\sigma_\tau^2 = \varepsilon\,\tau(1-\tau)$.

At $\varepsilon = 0$, Eq. (1) recovers the CFM straight-line interpolant (Lipman et al., 2023); an optimal-transport endpoint coupling yields OT-CFM (Tong et al., 2024a). At $\varepsilon = 1$, it has Brownian-bridge marginal variance and independent endpoints. This stochastic interpolant is related to DDBM (Zhou et al., 2024), but differs from both a non-differentiable Brownian-bridge sample path and an entropic-OT SB with a coupled endpoint joint. We study the interior $0 < \varepsilon < 1$ as a dimension-dependent model-selection parameter.

Proposition 2 (Conditional velocity)**.** Fix $(\mathbf{x}_0, \mathbf{x}_1, \mathbf{z})$ and let $\tau \in (0,1)$. Differentiating the stochastic-interpolant path of Def. 1 along τ (its single-time marginals match the Brownian-bridge kernel; it is not a Brownian sample path) gives the regression target

$$\mathbf{v}_\tau \;=\; \underbrace{\frac{d\boldsymbol{\mu}_\tau}{d\tau}}_{=\mathbf{x}_1-\mathbf{x}_0} \;+\; \underbrace{\frac{d\log\sigma_\tau}{d\tau}}_{=(1-2\tau)/[2\tau(1-\tau)]}\;(\mathbf{x}_\tau-\boldsymbol{\mu}_\tau) \;=\; (\mathbf{x}_1-\mathbf{x}_0) \;+\; \frac{(1-2\tau)\sqrt{\varepsilon}}{2\sqrt{\tau(1-\tau)}}\,\mathbf{z}. \tag{2}$$

The deterministic part is the FM target; the stochastic correction has squared L^2-norm $\mathbb{E}_{\mathbf{z}}\|\mathbf{v}_\tau - (\mathbf{x}_1-\mathbf{x}_0)\|_2^2 = d\,\varepsilon\,(1-2\tau)^2/[4\tau(1-\tau)] = O(d\,\varepsilon)$, exhibiting the dimension scaling that drives Proposition B.

Proof. Differentiating $\mathbf{x}_\tau = \boldsymbol{\mu}_\tau + \sigma_\tau\mathbf{z}$ gives $\dot{\mathbf{x}_\tau} = \dot{\boldsymbol{\mu}}_\tau + (\dot\sigma_\tau/\sigma_\tau)(\mathbf{x}_\tau - \boldsymbol{\mu}_\tau)$; substituting $\dot{\boldsymbol{\mu}}_\tau = \mathbf{x}_1 - \mathbf{x}_0$ and $\dot\sigma_\tau/\sigma_\tau = (1-2\tau)/[2\tau(1-\tau)]$ yields Eq. (2). The $\varepsilon \to 0$ value is well-defined by continuity ($\sigma_\tau\mathbf{z}\to 0$). □

We regress $\mathbf{v}_\theta(\mathbf{x}_\tau, \tau)$ on $\mathbf{v}_\tau$ with the standard FM/SI squared loss

$$\mathcal{L}_\varepsilon(\theta) = \mathbb{E}_{\tau\sim\mathcal{U}[\delta,1-\delta],\,\mathbf{x}_0,\mathbf{x}_1,\mathbf{z}}\left[\left\|\mathbf{v}_\theta(\mathbf{x}_\tau,\tau) - \mathbf{v}_\tau\right\|^2\right]. \tag{3}$$

The cutoff δ regularizes the logarithmic endpoint divergence: the stochastic correction scales as $1/\sqrt{\tau}$ and its squared magnitude as $1/\tau$. The controlled GMM sweeps use $\delta = 0.05$. The unrestricted population-risk minimizer is $\mathbf{u}_\tau$; a finite model class recovers it only under realizability and global population-risk minimization. The GMM sampler starts from p_{noise} and integrates backward from $1-\delta$ to δ with K Euler steps (Alg. 2). This approximates both endpoints: the exact starting marginal is $p^\varepsilon_{1-\delta}$, and the exact terminal marginal is p^ε_δ rather than p_{data}. These discrepancies depend on ε and d; App. AL separates them experimentally. CIFAR uses a distinct clamped-time sampler (App. Z).

3.3 Theorem A: Exact FM Limit

Theorem 3 (FM convergence within the Brownian-bridge stochastic interpolant family)**.** Fix $\delta \in (0, \frac{1}{2})$ and assume $\mathbb{E}\|\mathbf{x}_0\|^2 < \infty$. For part (iii), fix a predictor $\mathbf{v}_\theta$ that is continuous in its spatial input and satisfies $\|\mathbf{v}_\theta(x,\tau)\| \le C(1+\|x\|)$ uniformly on $\tau \in [\delta, 1-\delta]$. As $\varepsilon \to 0$: (i) $\mathbf{x}_\tau \to (1-\tau)\mathbf{x}_0 + \tau\mathbf{x}_1$ a.s. for every $\tau \in [\delta, 1-\delta]$; (ii) $\mathbf{v}_\tau(\varepsilon) \to \mathbf{x}_1 - \mathbf{x}_0$ in $L^2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{z} \mid \tau)$ for every $\tau \in [\delta, 1-\delta]$; (iii) $\lim_{\varepsilon\to 0}\mathcal{L}_\varepsilon(\theta) = \mathcal{L}_{\mathrm{FM},\delta}(\theta)$ for that fixed predictor, where the right side is the FM loss restricted to $\tau \sim \mathcal{U}[\delta, 1-\delta]$; (iv) the sampling ODE $d\mathbf{x}/d\tau = \mathbf{v}_\theta(\mathbf{x},\tau)$ has the same form for all ε, with ε entering only through the learned velocity field $\mathbf{v}^\varepsilon_\theta$. In particular, FM is the $\varepsilon \to 0$ limit of the Brownian-bridge stochastic interpolant family of Def. 1.

Proof. (i) $\sigma_\tau \to 0$; (ii) the coefficient $(1{-}2\tau)^2/[4\tau(1{-}\tau)]$ is bounded on $[\delta, 1{-}\delta]$; (iii) the stated linear-growth and second-moment assumptions supply an integrable envelope, so dominated convergence applies; (iv) by construction. Full proof in App. H. □

3.4 Proposition B: Linear Bridge-Correction Energy

The bridge correction energy is the integrated squared magnitude of the bridge-induced velocity term; Model D later carries its target-level scaling into an empirical sampling-error model.

Proposition 4 (Bridge correction energy)**.** For any coupling $(\mathbf{x}_0, \mathbf{x}_1, \mathbf{z})$ with $\mathbf{z}\sim\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ independent of $(\mathbf{x}_0, \mathbf{x}_1)$, the bridge-induced term in the training target $\mathbf{v}_\tau$ (Eq. (2)) has integrated squared magnitude

$$C_{\text{bridge}}(\varepsilon;\delta) \equiv \int_\delta^{1-\delta} \mathbb{E}_{\mathbf{z}}\left\|\frac{(1-2\tau)\sqrt{\varepsilon}}{2\sqrt{\tau(1-\tau)}}\,\mathbf{z}\right\|^2 d\tau = \varepsilon\, d\, c_0(\delta), \tag{4}$$

$$c_0(\delta) \equiv \int_\delta^{1-\delta} \frac{(1-2\tau)^2}{4\tau(1-\tau)}\,d\tau, \tag{5}$$

with closed form $c_0(\delta) = \frac{1}{2}\log\frac{1-\delta}{\delta} - 1 + 2\delta$ (0.57 at $\delta = 0.05$, 2.45 at $\delta = 10^{-3}$; App. I); the integrand is not integrable on $[0,1]$, so the cutoff is essential. C_{bridge} is defined before conditioning on

$\mathbf{x}_\tau$ and characterizes target energy rather than downstream sample quality. FM ($\varepsilon = 0$) is the unique member with zero raw bridge correction.

Computation. $\mathbb{E}_{\mathbf{z}}\|c\,\mathbf{z}\|^2 = c^2 d$ for $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$; substituting $c = \frac{(1-2\tau)\sqrt{\varepsilon}}{2\sqrt{\tau(1-\tau)}}$ and integrating over $[\delta, 1-\delta]$ via the identity $(1-2\tau)^2 = 1 - 4\tau(1-\tau)$ yields the closed form $c_0(\delta)$. Full derivation in App. I. □

3.5 Theorem C: Marginal-Velocity Regularity

Positive ε smooths the kernel in the marginal velocity: by Bayes' rule $\mathbf{u}_\tau(x;\varepsilon)$ is a Nadaraya–Watson kernel regression (Nadaraya, 1964) against the Gaussian bridge kernel $q_\varepsilon(x|\mathbf{x}_0, \mathbf{x}_1)$ of bandwidth $h_\tau = \sqrt{\varepsilon\tau(1-\tau)}$:

$$\mathbf{u}_\tau(x;\varepsilon) = \frac{\int (\alpha_\tau \mathbf{x}_1 - \beta_\tau \mathbf{x}_0 + \gamma_\tau(x))\, q_\varepsilon(x|\mathbf{x}_0, \mathbf{x}_1)\, dp(\mathbf{x}_0, \mathbf{x}_1)}{\int q_\varepsilon(x|\mathbf{x}_0, \mathbf{x}_1)\, dp(\mathbf{x}_0, \mathbf{x}_1)}, \tag{6}$$

where $\alpha_\tau = 1/[2(1-\tau)]$, $\beta_\tau = 1/(2\tau)$, and $\gamma_\tau(x) = (1-2\tau)x/[2\tau(1-\tau)]$ rewrite the target of Eq. (2) in terms of $(\mathbf{x}_0, \mathbf{x}_1, x)$; the kernel is a delta at $\varepsilon = 0$ and a smooth Gaussian for $\varepsilon > 0$.

Theorem 5 (Velocity-field regularity under posterior moment conditions)**.** Fix $\tau \in [\delta, 1-\delta]$, $\varepsilon > 0$, and a region Ω on which the marginal density is positive. Let $h_\tau = \sqrt{\varepsilon\tau(1-\tau)}$. Assume constants $V_0, C_\mu < \infty$ (which may depend on d and Ω, but not on x) satisfy for every $x \in \Omega$

$$\|\mathrm{Var}(\mathbf{v}_\tau \mid \mathbf{x}_\tau = x)\|_{\mathrm{op}} \le V_0^2, \qquad \|\mathrm{Var}(\boldsymbol{\mu}_\tau \mid \mathbf{x}_\tau = x)\|_{\mathrm{op}} \le C_\mu h_\tau^2.$$

Then the Stein log-derivative identity applied to Eq. (6) yields

$$\nabla_x \mathbf{u}_\tau(x;\varepsilon) \;=\; \frac{1}{\sigma_\tau^2}\, \mathrm{Cov}\big(\mathbf{v}_\tau,\, \boldsymbol{\mu}_\tau \;\big|\; \mathbf{x}_\tau = x;\varepsilon\big) \;+\; \frac{1-2\tau}{2\tau(1-\tau)}\, \mathbf{I}_d, \tag{7}$$

and the operator norm admits

$$\sup_{x\in\Omega} \|\nabla \mathbf{u}_\tau(x;\varepsilon)\|_{\mathrm{op}} \;\le\; \underbrace{\frac{|1-2\tau|}{2\tau(1-\tau)}}_{\text{drift trend}} \;+\; \underbrace{\frac{V_0\sqrt{C_\mu}}{\sqrt{\varepsilon\,\tau(1-\tau)}}}_{\text{posterior slope}}, \tag{8}$$

provided the two constants hold uniformly over the compared ε. The bound contains no trace factor in d, although its constants may depend on dimension through the posterior. Log-concavity of the latent mean distribution is sufficient for the second contraction inequality (App. J).

Sketch (full proof in App. J). Center the Stein identity against the posterior mean and apply operator-norm Cauchy–Schwarz:

$$\|\mathrm{Cov}(\mathbf{v}_\tau, \boldsymbol{\mu}_\tau \mid x)\|_{\mathrm{op}} \le \|\mathrm{Var}(\mathbf{v}_\tau \mid x)\|_{\mathrm{op}}^{1/2} \|\mathrm{Var}(\boldsymbol{\mu}_\tau \mid x)\|_{\mathrm{op}}^{1/2} \le V_0\sqrt{C_\mu}\, h_\tau.$$

Dividing by h_τ^2 gives Eq. (8). □

The decreasing bound does not establish monotonicity of the actual Jacobian norm. FM already has interior Gaussian smoothing from $\tau\mathbf{x}_1$; positive ε raises the total variance from τ^2 to $\tau^2 + \varepsilon\tau(1-\tau)$. The measured 2D gradient proxy decreases from 4.82 to 1.83 over $\varepsilon = 0$ to 1, but does not verify the posterior conditions. Exact GMM posterior computations give approximately flat sampled maxima over $d = 16, 64, 256$ (App. P.3).

Proposition B measures raw target energy, whereas Theorem C controls an operator norm under dimension-uniform posterior constants. Target conditioning can remove some regression noise without changing the marginal velocity (App. AM). Neither result yields a sampling-error bound.

3.6 Model D: A Phenomenological Dimension-Dependent Error Model

The preceding results motivate two features for a phenomenological squared sampling-error model: a fitting term that decreases with ε and a penalty linear in $d\varepsilon$. This ansatz is an additional modeling assumption, not a consequence of comparing the trace in Proposition B with the operator norm in Theorem C:

$$E^2(\varepsilon; d) = \underbrace{\frac{A}{1+\beta_0\,\varepsilon}}_{\text{fitting error (regularized)}} + \underbrace{d\,\hat{c}_0\,\varepsilon}_{\text{raw-target penalty feature}}, \tag{9}$$

with positive, protocol-specific parameters $A, \beta_0, \hat{c}_0$. The coefficient $\hat{c}_0$ differs from the analytical $c_0(\delta)$: Proposition B supplies the feature $d\varepsilon$, not a sample-error bound. We call Eq. (9) Model D and test a fixed numerical instance in Sec. 4.3. Its parameters are not fitted to the primary GMM sweep, and the absence of a budget term means it does not predict the measured dependence on K. Its constrained minimizer has the following closed form.

Proposition 6 (Minimizer of Model D)**.** Let $d^* \equiv A\beta_0/\hat{c}_0$. (a) For $d \geq d^*$, E^2 is strictly increasing on $(0, 1]$, so $\varepsilon = 0$ is the global minimizer on $[0, 1]$. (b) For $d < d^*$, the unique stationary point

$$\varepsilon_{\text{stat}}(d) = \frac{1}{\beta_0}\Big(\sqrt{d^*/d} - 1\Big) \tag{10}$$

is the unconstrained minimizer; it lies in $(0, 1)$ iff $d > d^*/(1+\beta_0)^2$, and otherwise the constrained optimum is the boundary $\varepsilon^* = 1$.

Proof. $\partial E^2/\partial\varepsilon = d\hat{c}_0 - A\beta_0/(1+\beta_0\varepsilon)^2$ is strictly increasing in ε, and its sign at $\varepsilon = 0$ is the sign of $d - d^*$. For $d \geq d^*$ it is therefore positive on $(0, 1]$; for $d < d^*$, solving $\partial E^2/\partial\varepsilon = 0$ gives Eq. (10). Second-derivative and boundary analysis in App. L. □

We evaluate this rule separately from the fixed-protocol dimension sweep (App. C); a dimensional ordering under the ansatz need not locate the empirical crossover.

4 Experiments

We first vary dimension under a fixed GMM training recipe, then intervene on capacity at fixed dimension, and finally test transfer across families. GMM and PCA ladders vary modeled dimension directly; other tasks use their stated representations and metrics. Full grids, protocol definitions, and task-specific results appear in Apps. AK, E, and X, and Table 2.

4.1 GMM Dimension Scaling under a Fixed Protocol

The primary sweep fixes one target mixture per condition and uses a six-layer width-512 MLP, 30k AdamW updates, and three model seeds. At $K = 20$, ε is selected by validation sliced Wasserstein-2 distance (SW_2), and the selected checkpoint is evaluated on independent test data (App. AK). The tested dimensions are $\{64, 128, 160, 192, 224, 256, 300, 350, 400, 450\}$. Selected noise is 0.3 at $d = 64, 128$, 0.1 at $d = 160$, and zero at every tested $d \geq 192$, bracketing the transition by $(160, 192]$ (Figs. 2 and 4). An independent 16-mode family selects 0.3 at $d = 160$ and 0.01 at $d = 192$. It reproduces this operational bracket under the $\varepsilon^* > 0.05$ regime criterion, without reproducing an exact-zero optimum at $d = 192$.

A companion sweep evaluates all trained noise levels at $K \in \{1, 2, 5, 10, 20, 50, 100\}$ for $d \in \{64, 128, 256\}$. Selected noise decreases with K at $d = 64, 128$ and remains zero at $d = 256$ (Table 33). This budget trend is measured separately from Model D, which contains no budget term.

Endpoint handling changes the preferred positive noise. Re-evaluating fixed checkpoints with exact truncated-path endpoints preserves positive-noise selection at $d = 64$ and FM at $d = 192, 256$, but moves the low-dimensional optimum from 0.3 to 0.5 (App. AL). Exact initialization requires data, and the exact terminal reference varies with ε. This control tests truncated-path transport; it does not re-estimate the crossover or establish generation quality against a common target.

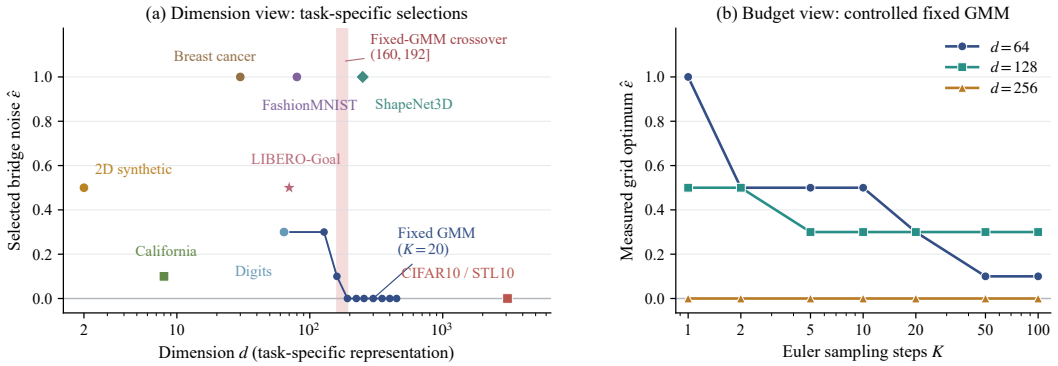


Figure 2: Dimension and sampling-budget views of bridge-noise selection. (a) Task-specific grid optima use the dimension conventions and budgets of Table 2. The GMM line and shaded crossover $(160, 192]$ refer only to the fixed recipe at $K = 20$. (b) Fixed-protocol GMM optima minimize mean SW_2 across three seeds; lines connect measured budgets. Full GMM curves appear in App. Fig. 5.

4.2 Capacity Interventions at Fixed Dimension

Changing capacity at fixed dimension can reverse the noise-selection outcome (Table 1). At GMM $d = 192$, widening from 512 to 1024 restores a 35% benefit at $\varepsilon^* = 0.1$. At $d = 128$, narrowing from 512 to 128 removes the 36% benefit. Tripling training steps at $d = 192$ improves the FM fit but leaves FM preferred. These comparisons show that the observed dimensional crossover depends on the training configuration.

Table 1: Capacity and training interventions at fixed dimension. Values are baseline $\rightarrow$ intervention. FM/s_{ref} is FM SW_2 divided by the reference coordinate scale s_{ref} (App. E); GMM entries use validation means. Gain is the reduction in SW_2 of the best noise grid point relative to its own FM baseline. Small optima in flat curves do not establish a benefit. Full configurations, predictions, and grids are in App. E.

Cell / dimension	Intervention	FM/s_{ref}	Selected ε^*	Gain (%)
I1 GMM / 192	width 512 $\rightarrow$ 1024	.225 $\rightarrow$ .070	0 $\rightarrow$ .1	0 $\rightarrow$ 35
I2 GMM / 192	steps 30k $\rightarrow$ 90k	.225 $\rightarrow$ .178	0 $\rightarrow$ 0	0 $\rightarrow$ 0
I3 GMM / 128	width 512 $\rightarrow$ 128	.075 $\rightarrow$ .869	.3 $\rightarrow$ 0	36 $\rightarrow$ 0
I4 PCA-FMNIST / 256	larger / longer	.165 $\rightarrow$ .212	.01 $\rightarrow$ .03	2 $\rightarrow$ 4.7
I5 PCA-CIFAR / 64	smaller network	.067 $\rightarrow$ .074	.3 $\rightarrow$ 1	27 $\rightarrow$ 12
I6 PCA-CIFAR / 64	smaller / shorter	.067 $\rightarrow$ .073	.3 $\rightarrow$ .5	27 $\rightarrow$ 0

The interventions were motivated by an association across 34 full-grid cells at $K = 20$. We normalize FM SW_2 by a reference coordinate scale s_{ref} (unit preprocessing scale for PCA/tabular data; nominal mixture scale for GMM) and count noise as beneficial when its best grid point improves on FM by more than 5%. Below an exploratory normalized-error threshold of 0.09, 10 of 12 cells meet this criterion; the other two gain 4.8 and 4.9%. Above it, 1 of 22 does, a 569-sample tabular dataset (Fig. 3). This benefit criterion differs from the $\varepsilon^* > 0.05$ labels used for transfer. PCA/tabular scores use training-data references, so this association does not establish held-out predictive performance.

We recorded a conditional prediction before each of six capacity/training interventions (App. E). Normalized FM error tracks their benefit/no-benefit outcomes in five cases. I6 exposes its limitation: a severely reduced PCA-CIFAR model loses the benefit, with a curve flat to 0.1%, while its normalized FM error remains low at 0.073. The other PCA interventions retain their previous benefit/no-benefit outcome, including a larger PCA-FashionMNIST model whose FM fit does not improve. Thus low baseline error does not guarantee a useful noise sweep.

These interventions support a capacity-dependent crossover; they do not establish normalized FM error as its causal mediator. The 0.09 threshold is exploratory, and the relative gain shares the FM

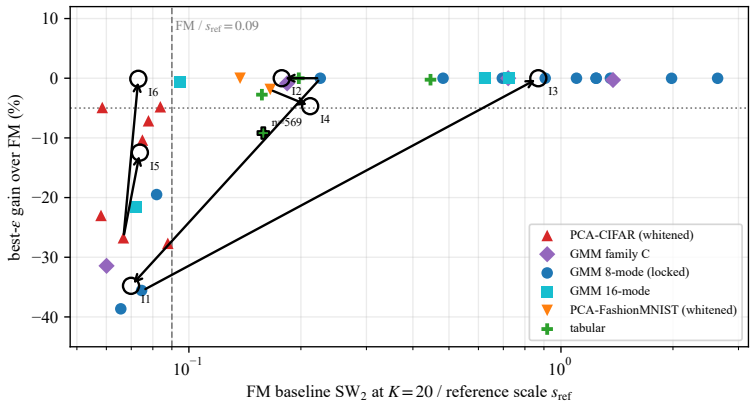


Figure 3: Noise benefit and the normalized FM baseline. Best-grid gain over FM at $K = 20$ versus FM SW_2 divided by the reference scale s_{ref}, for 34 retained full-grid cells (App. E). Outlined markers denote tabular sets with fewer than 1,000 samples. Arrows show interventions at fixed dimension (App. E).

baseline in its numerator. Its scope is also budget-dependent: at $K \geq 50$, whitened PCA-CIFAR can favor larger noise with 24–37% gains (App. E). An unfitted capacity-sensitive extension of Model D is discussed in App. K. The practical implication is to assess FM fit and noise benefit under the intended representation, model, and sampling budget.

4.3 Transfer and Limits of the Selection Rule

We evaluate Model D with the fixed illustrative values $A = 20.8$, $\beta_0 = 2.61$, and $\hat{c}_0 = 0.149$, whose original calibration procedure has not been recovered. They imply an exact-zero threshold $d^* \approx 364$ and an $\varepsilon^* > 0.05$ regime threshold near 285 (App. C). Neither is fitted to the primary GMM transition; this comparison tests transfer across protocols, not calibration accuracy within the earlier protocol.

Before running nine new cells, we recorded their predicted regimes using these constants, the 0.05 threshold, and fixed dimension conventions (App. D). The rule matches 5 of 9 labels; always predicting FM would match 6 of 9, so this sample does not demonstrate a classification advantage. The four misses lie at $176 \leq d \leq 384$: whitened PCA-CIFAR selects $\varepsilon^* = 0.3$ at $d = 384$, whereas PCA-FashionMNIST and GMM family C are operationally near FM by $d = 176$–256. Family C has a crossover in $(128, 176]$. The canonical operational rule classifies its three predicted cells correctly, while $d = 176$ lies in that rule's unclassified gap. Optima at $\varepsilon^* = 0.01$ receive the FM label; their improvement over zero is within one seed standard deviation.

4.4 Task-specific Outcomes

Table 2 brings the main task-specific outcomes into the main text. The original sweep overview is retained with protocol-specific qualifications in App. Fig. 6. Pixel-space CIFAR favors its own FM baseline at the reported multi-step budgets under both separately reported protocols (App. Z). PCA ladders use whitened coordinates and dimension-dependent training configurations; they are not strict monotonic laws. The molecular case study changes reverse-step noise after distillation, so it provides related evidence about stochasticity rather than a direct test of Eq. (9) (App. AC). The collective evidence motivates dimension as a search coordinate, but does not establish a transferable numerical selector or measured search-cost savings.

Table 2: Task-specific outcomes, with each family's representation, metric, and budget. Column five is the selected positive-noise point, or the stated positive comparator when $\varepsilon^* = 0$. Tabular and whitened PCA-CIFAR use $K = 20$; other budgets are in their appendices. CIFAR uses the v2 200k checkpoints, separately from v3. Controlled GMM results appear in Table 32.

Dataset	d_{eff}	**Metric**	**FM** ($\varepsilon = 0$)	**Selected / comparator**	ε^*
Experiments run for this paper (App. S–AD)					
2D Synthetic (8-Gauss.)	2	$W_2\downarrow$	0.58	**0.41** (-29%)	0.5
California Housing	8	$SW_2\downarrow$	0.446	**0.445** (-0.2%)	0.1
Breast Cancer (UCI)	30	$SW_2\downarrow$	0.158	**0.144** (-9.2%)	1.0
PCA-CIFAR $d=32$	32	$SW_2\downarrow$	0.058	**0.045** (-23.0%)	1.0
Digits (sklearn)	64	$SW_2\downarrow$	0.157	**0.153** (-2.8%)	0.3
PCA-CIFAR $d=64$	64	$SW_2\downarrow$	0.067	**0.049** (-26.7%)	0.3
LIBERO-Goal (VLA)*	70	success (%)$\uparrow$	95.0	**100.0** ($+5$pp)	0.5
FashionMNIST$^\flat$	80	$SW_2\downarrow$	0.016	**0.013** (-16.0%)	1.0
PCA-CIFAR $d=128$	128	$SW_2\downarrow$	0.059	**0.056** (-4.9%)	0.5
ShapeNet (3 categories)$^{\ddagger\flat}$	250	CD$\downarrow$	20.6	**10.7** (-48%)	1.0
PCA-CIFAR $d=256$	256	$SW_2\downarrow$	0.084	**0.080** (-4.8%)	0.1
PCA-CIFAR $d=512$	512	$SW_2\downarrow$	0.078	**0.072** (-7.1%)	0.1
Endpoint comparisons					
STL10 (32×32)	3072	FID@50$\downarrow$	**19.5**	115.0 ($+490\%$)	0
CIFAR-10 ($d=3072$)	3072	FID@200$\downarrow$	**14.6**	40.9 ($+180\%$)	0

‡ShapeNet values are averaged over Airplane, Chair, and Car; Airplane alone has a shallow minimum at $\varepsilon = 0.5$. $^\flat$PCA-90% d_{eff}; PCA-99% values are 428/486 in Tab. 11. *LIBERO-Goal is evaluated over 100 episodes.

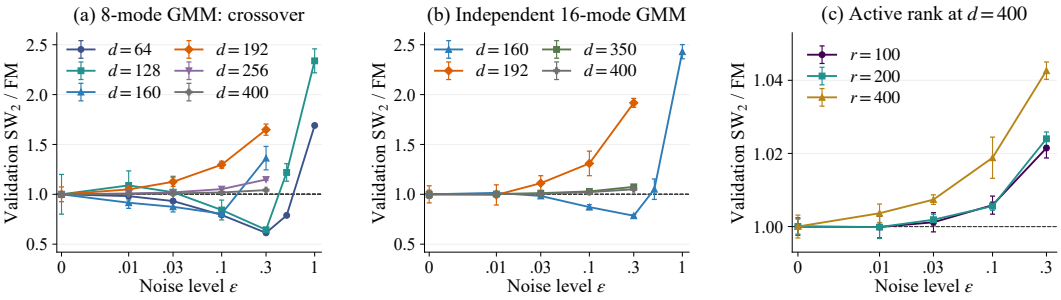


Figure 4: Fixed-protocol GMM sweep. (a) Representative 8-mode dimensions including the crossover; (b) an independent 16-mode family; (c) active-rank interventions at $d = 400$. Means and sample standard deviations over three seeds are divided by each condition's mean FM score. All measured noise levels are shown. The noise axis is symmetric-log; panel (c) uses a finer vertical scale.

5 Conclusion

We studied bridge-noise selection within an independent-endpoint stochastic-interpolant family. Theorem A establishes its FM limit, Proposition B gives the raw correction-energy scaling, and Theorem C provides a conditional regularity bound. Model D turns these features into a phenomenological rule whose fixed parameters do not transfer uniformly. The strongest empirical finding is a capacity-dependent crossover: fixed-recipe dimension sweeps approach FM, while capacity interventions at fixed dimension can restore or remove the benefit of noise. Endpoint controls preserve the tested low/high-dimensional ordering but change the preferred positive noise. These results support evaluating a small noise grid against the FM baseline under the intended configuration. The baseline diagnostic remains exploratory; finite grids, conditional regularity assumptions, and the absence of a validated calibration procedure limit its use (App. AE).

AI USE STATEMENT

Generative AI tools assisted with manuscript editing, code development and debugging, protocol and numerical-provenance audits, and the design and interpretation of the endpoint-sensitivity diagnostic. Mathematical statements and empirical claims are the authors' responsibility; AI-generated suggestions are not evidence, and numerical results are tied to retained scripts, checkpoints, and evaluation outputs.

REPRODUCIBILITY STATEMENT

Assumptions, proofs, settings, protocols, and seed-level results are documented in Apps. H–L, S–AD, and AI; Eq. (11) records the diagnostic used in Tab. 5, and App. D documents the predictions behind Tab. 6.

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023.

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Kevin Black, Noah Brown, James Darpinian, et al. $\pi_{0.5}$: A vision-language-action model with open-world generalization. In *Conference on Robot Learning*, 2025a.

Kevin Black, Noah Brown, Danny Driess, et al. π_0: A vision-language-action flow model for general robot control. In *Robotics: Science and Systems*, 2025b.

Herm Jan Brascamp and Elliott H. Lieb. On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log-concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976.

Tianrong Chen, Guan-Horng Liu, and Evangelos A. Theodorou. Likelihood training of Schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations*, 2022.

Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *International Conference on Artificial Intelligence and Statistics*, 2011.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, 2021.

Valentin De Bortoli, Iryna Korshunova, Andriy Mnih, and Arnaud Doucet. Schrödinger bridge flow for unpaired data translation. In *Advances in Neural Information Processing Systems*, 2024.

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

Patrick Esser, Sumith Kulal, Andreas Blattmann, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *International Conference on Machine Learning*, 2024.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. In *International Conference on Learning Representations*, 2025.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, 2025.

Nikita Gushchin, Alexander Kolesov, Petr Mokrov, Polina Karpikova, Andrey Spiridonov, Evgeny Burnaev, and Alexander Korotin. Building the bridge of Schrödinger: A continuous entropic optimal transport benchmark. In *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*, 2023.

Marcus Häggbom, Viktor Nilsson, Pierre Nyquist, and Joakim Andén. Reflected Schrödinger bridge matching. *arXiv preprint arXiv:2607.03626*, 2026.

Emiel Hoogeboom, Víctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *International Conference on Machine Learning*, 2022.

Minyoung Kim. A unified framework for diffusion bridge problems: Flow matching and Schrödinger matching into one. *arXiv preprint arXiv:2503.21756*, 2025.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. LIBERO: Benchmarking knowledge transfer for lifelong robot learning. In *Advances in Neural Information Processing Systems*, 2023a.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023b.

Cheng Lu and Yang Song. Simplifying, stabilizing and scaling continuous-time consistency models. In *International Conference on Learning Representations*, 2025.

Wuyang Luan, Junhui Li, Weiguang Zhao, Wenjian Zhang, Tieru Wu, and Rui Ma. Rectified Schrödinger bridge matching for few-step visual navigation. *arXiv preprint arXiv:2604.05673*, 2026a.

Wuyang Luan, Junhui Li, Weiguang Zhao, Wenjian Zhang, Tieru Wu, and Rui Ma. SnapFlow: One-step action generation for flow-matching VLAs via progressive self-distillation. *arXiv preprint arXiv:2604.05656*, 2026b.

Elizbar A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1): 141–142, 1964.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In *International Conference on Machine Learning*, 2023.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, 2021.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. In *Advances in Neural Information Processing Systems*, 2023.

Mustafa Shukor et al. SmolVLA: A vision-language-action model for affordable and efficient robotics. *arXiv preprint arXiv:2506.01844*, 2025.

Vignesh Ram Somnath, Matteo Pariset, et al. Aligned diffusion Schrödinger bridges. In *Conference on Uncertainty in Artificial Intelligence*, 2023.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning*, pp. 32211–32252, 2023.

Sophia Tang. Foundations of Schrödinger bridges for generative modeling. *arXiv preprint arXiv:2603.18992*, 2026.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024a.

Alexander Y. Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching. In *International Conference on Artificial Intelligence and Statistics*, 2024b.

Dahan Wang, Jun Gao, Tong Lei, Yuxiang Hu, Changbao Zhu, Kai Chen, and Jing Lu. Rethinking flow and diffusion bridge models for speech enhancement. In *AAAI Conference on Artificial Intelligence*, 2026a. arXiv:2602.18355.

Hebaixu Wang, Jing Zhang, Haoyang Chen, Haonan Guo, Di Wang, Jiayi Ma, and Bo Du. Residual diffusion bridge model for image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026b. arXiv:2510.23116.

Jinxin Wang, Ya-Ping Hsieh, and Bahar Taskesen. Schrödinger bridge as robustified optimal transport flows. In *NeurIPS 2025 Workshop on Dynamics at the Frontiers of Optimization, Sampling, and Games*, 2025.

Linqi Zhou, Aaron Lou, Samar Khanna, and Stefano Ermon. Denoising diffusion bridge models. In *International Conference on Learning Representations*, 2024.

Kaizhen Zhu, Mokai Pan, Zhechuan Yu, Jingya Wang, Jingyi Yu, and Ye Shi. Diffusion bridge or flow matching? A unifying framework and comparative analysis. In *International Conference on Machine Learning*, 2026. arXiv:2509.24531.

A Notation Summary

This appendix collects the notation. Table 3 summarizes the symbols used throughout the paper and the appendices.

Table 3: Notation summary.

Symbol	Meaning
$\mathbf{x}_0 \sim p_{\text{data}}$	Data sample
$\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$	Noise sample, independent of $\mathbf{x}_0$
$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$	Bridge latent, independent of $(\mathbf{x}_0, \mathbf{x}_1)$
$\tau \in [0, 1]$	Continuous time
δ	Time cutoff; training and sampling use $\tau \in [\delta, 1-\delta]$
$\varepsilon \geq 0$	Bridge noise parameter
$\mathbf{x}_\tau$	ε-bridge interpolant
$\boldsymbol{\mu}_\tau$	Bridge mean $(1-\tau)\mathbf{x}_0 + \tau\mathbf{x}_1$
σ_τ $(= h_\tau)$	Bridge std / kernel bandwidth $\sqrt{\varepsilon\tau(1-\tau)}$
$\mathbf{v}_\tau$	Conditional velocity
α_τ, β_τ	Coefficients of $\mathbf{x}_1, \mathbf{x}_0$ in $\mathbf{v}_\tau$: $\frac{1}{2(1-\tau)}$, $\frac{1}{2\tau}$
$\gamma_\tau(x)$	Drift-trend term $\frac{1-2\tau}{2\tau(1-\tau)}\, x$ in $\mathbf{v}_\tau$
$\mathbf{u}_\tau$	Marginal velocity $\mathbb{E}[\mathbf{v}_\tau \mid \mathbf{x}_\tau]$
$\mathbf{v}_\theta$	Learned velocity network with parameters θ
C_{bridge}	Bridge correction energy $\varepsilon\, d\, c_0(\delta)$ (Prop. B)
$c_0(\delta)$	Analytic cutoff constant $\frac{1}{2}\log\frac{1-\delta}{\delta} - 1 + 2\delta$
$\hat{c}_0$	Phenomenological coefficient of the $d\varepsilon$ feature in Model D (distinct from $c_0(\delta)$)
A, β_0	Phenomenological amplitude and response parameters (Model D)
d^*	Model local threshold $A\beta_0/\hat{c}_0$
A_{cap}, B, β	Capacity remainder, smoothing amplitude, response rate in Eq. (24) (not fitted)
ξ	Protocol descriptor $(d, \text{capacity}, \text{data}, K)$ in Eq. (24)
FM/s_{ref}	FM sliced-W_2 at K=20 over reference scale s_{ref} (App. E)
d_{eff}	Task-specific operational dimension (App. N)
V_0, C_μ	Posterior moment/contraction constants in Thm. C
K_{ODE}	Euler ODE steps at inference (default meaning of K)
K_{distill}	Target NFE of consistency-style distillation (App. M)
K_{solver}	Solver steps in published case studies (App. M)

B Dataset Overview and Task Families

This appendix gives an overview of the data used in the paper. Table 4 lists the task families, datasets, and operational dimension assigned to each displayed cell. Per-cell results are in Tab. 2.

Table 4: Dataset overview. The dimension column records the operational convention used for each displayed family. Entries marked [♭] use a PCA-90% proxy; PCA-99% alternatives are listed in Tab. 11.

	Type	Dataset	d_{eff}	ε^*
•	2D synthetic	8-Gauss. / Moons / Roll / Pinwheel	2	0.3–1.0
•	Tabular (UCI)	California / Breast / Digits	8–64	0.1–1.0
•	Navigation	RSBM (published, $K=3$)	16	0.5 (full-train)
•	Robotics (VLA)	LIBERO ε-sweep[§]	70	0.5
•	Molecular	QM9 generation (ε-BMD)	$\approx$40	0.1
•	PCA-CIFAR	CIFAR-10 $\to$ PCA-$\{32, 64, 128, 256, 512\}$	32–512	$1.0 \to 0.1$
•	Image	FashionMNIST[♭]	$\approx$80	1.0 ($K=50$)
•	3D point cloud	ShapeNet (Airplane / Chair / Car)[♭]	$\approx$250	1.0
•	GMM sweep	8-GMM, fixed-recipe family	64–450	$0.3 \to 0$
•	Image gen.	CIFAR-10 (full)	3072	0 (FM)

[§]The VLA cell uses the executed 10-step by 7-DoF action chunk ($d = 70$); the full generated chunk has dimension 350, illustrating the convention dependence.

B.1 Task-specific Outcome Summary

Table 2 records the broader within-task comparisons. These rows retain their own representations, budgets, and metrics; they are not additional estimates of the controlled GMM crossover.

C Fixed Diagnostic and Its Empirical Scope

This appendix evaluates the earlier fixed constants $A = 20.8$, $\beta_0 = 2.61$, and $\hat{c}_0 = 0.149$, hence $d^* = 364.35$ (rounded to 364 below). Their original fitting procedure has not been recovered; they are illustrative choices with no reported calibration confidence interval. The primary GMM sweep uses a different protocol (App. X), so it tests transfer, not the earlier within-protocol fit.

Table 5 compares this fixed rule with measured grid optima. Because the constants were already in use before the later prospective cells, the latter can test this particular rule, but neither comparison validates a general calibration method. The diagnostic is

$$\varepsilon^*_{\text{pred}}(d) \;=\; \left[\frac{\sqrt{d^*}-\sqrt{d}}{\sqrt{d}\,\beta_0}\right]_0^1, \qquad d^* = 364,\ \beta_0 = 2.61, \tag{11}$$

the clipped Model D minimizer (Eq. (25)), reproducible from two numbers. Its exact-zero threshold is d^*, whereas the operational label $\varepsilon^*_{\text{pred}} > 0.05$ holds only for $d < d_{\text{reg}} = d^*/(1 + 0.05\beta_0)^2 \approx 285$. Both thresholds are model quantities, distinct from a measured grid crossover.

Regarding the accuracy of the diagnostic, the binary threshold is 0.05, but interior accuracy varies across tasks. For example, the ShapeNet prediction is 0.08 while its measured grid optimum is 1.00; both remain on the positive-noise side of the operational threshold. Four of the twenty diagnostic cells are regime mismatches: GMM $d = 192, 224, 256$, labeled bridge by the fixed rule although their grid optima are zero, and PCA-CIFAR $d = 512$, labeled FM although its selected noise is 0.1. These mismatches concern transfer of the fixed rule.

D Validation on New Cells with Pre-registered Predictions

The diagnostic above compares Eq. (11) with sweeps that had already been run when the rule was written, so it cannot distinguish predictive content from post-hoc agreement. This appendix reports a validation in which every prediction was recorded before the corresponding sweep was run.

Before any new run, we fixed in writing (i) the rule, $\varepsilon^*_{\text{pred}}(d) = [(\sqrt{364} - \sqrt{d})/(2.61\sqrt{d})]_0^1$ with regime "bridge" if $\varepsilon^* > 0.05$ and "FM" otherwise; (ii) the d_{eff} convention of each family (number of whitened principal components for PCA cells, raw standardized feature count for tabular cells, ambient dimension for GMM cells); (iii) nine cells that had never been run, with their predicted ε^* and regime; and (iv) for the GMM cells, the operational prediction of the fixed-protocol sweep (positive noise for $d \le 160$, FM for $d \ge 192$, no prediction inside the gap). The sweeps were then

Table 5: Fixed two-parameter diagnostic across task-specific protocols. The two-parameter prediction uses $d^* = 364$ and $\beta_0 = 2.61$, clipped to $[0, 1]$; GMM, PCA, and tabular optima use $K = 20$; image and ShapeNet rows use their stated task-specific budgets (CIFAR: v2 NFE$= 200$). The Regime column gives prediction/empirical labels under the rule "bridge" (positive noise) if $\varepsilon^* > 0.05$ and "FM" otherwise. These entries are descriptive and are not additional measurements of the fixed-protocol GMM sweep.

Dataset	d_{eff}	$\varepsilon^*_{\text{pred}}$	$\varepsilon^*_{\text{emp}}$	Regime	Match
Descriptive diagnostic cells					
California ($d=8$)	8	1.00	0.10	bridge / bridge	✓
Breast ($d=30$)	30	0.95	1.00	bridge / bridge	✓
Digits ($d=64$)	64	0.53	0.30	bridge / bridge	✓
GMM $d=128$	128	0.26	0.30	bridge / bridge	✓
GMM $d=160$	160	0.20	0.10	bridge / bridge	✓
GMM $d=192$	192	0.14	0.00	bridge / FM	×
GMM $d=224$	224	0.11	0.00	bridge / FM	×
GMM $d=256$	256	0.07	0.00	bridge / FM	×
GMM $d=300$	300	0.04	0.00	FM / FM	✓
GMM $d=350$	350	0.01	0.00	FM / FM	✓
GMM $d=400$	400	0.00	0.00	FM / FM	✓
GMM $d=450$	450	0.00	0.00	FM / FM	✓
Whitened PCA-CIFAR $d=32$	32	0.91	1.00	bridge / bridge	✓
Whitened PCA-CIFAR $d=64$	64	0.53	0.30	bridge / bridge	✓
Whitened PCA-CIFAR $d=128$	128	0.26	0.50	bridge / bridge	✓
Whitened PCA-CIFAR $d=256$	256	0.07	0.10	bridge / bridge	✓
Whitened PCA-CIFAR $d=512$	512	0.00	0.10	FM / bridge	×
ShapeNet ($d_{\text{eff}} \approx 250$)	250	0.08	1.00	bridge / bridge	✓
CIFAR-10 ($d=3072$)	3072	0.00	0.00	FM / FM	✓
STL10 (32×32)	3072	0.00	0.00	FM / FM	✓

No aggregate accuracy is claimed; the table is a descriptive diagnostic.

Table 6: Validation on new cells. Predictions were recorded before any of the nine sweeps was run. "Model D" is the regime predicted by Eq. (11); "operational" is the fixed-protocol GMM transition ($d \leq 160$ bridge, $d \geq 192$ FM), applicable to GMM cells only. Regimes follow the fixed threshold: $\varepsilon^* > 0.05$ is bridge, otherwise FM, so grid optima at $\varepsilon^* = 0.01$ are FM-regime cells (their gap to $\varepsilon = 0$ is within one seed standard deviation of the selected grid point). Sliced-W_2 is the seed mean at $K = 20$ (validation set for GMM cells).

Cell	d_{eff}	$\varepsilon^*_{\text{pred}}$	$\varepsilon^*_{\text{emp}}$	Regime pred. / emp.	Model D	Operational	SW_2 (FM)	SW_2 (ε^*)
Whitened PCA-CIFAR $d=192$	192	0.14	0.50	bridge / bridge	✓	—	0.0878	0.0635
Whitened PCA-CIFAR $d=384$	384	0.00	0.30	FM / bridge	×	—	0.0750	0.0672
PCA-FashionMNIST $d=256$	256	0.07	0.01	bridge / FM	×	—	0.1654	0.1623
PCA-FashionMNIST $d=400$	400	0.00	0.00	FM / FM	✓	—	0.1373	0.1373
ISOLET (tabular)	617	0.00	0.00	FM / FM	✓	—	0.1974	0.1974
GMM family C $d=128$	128	0.26	0.10	bridge / bridge	✓	✓	0.0200	0.0137
GMM family C $d=176$	176	0.17	0.01	bridge / FM	×	(gap)	0.0575	0.0570
GMM family C $d=224$	224	0.11	0.00	bridge / FM	×	✓	0.2169	0.2169
GMM family C $d=320$	320	0.03	0.01	FM / FM	✓	✓	0.3952	0.3941

Model D regime: $5/9$. Operational GMM transition: $3/3$ predicted cells; $d = 176$ lies on the FM side, placing family C's crossover in $(128, 176]$.

run with the unchanged per-family protocols: the whitened PCA-CIFAR recipe of App. AD.3 for PCA-CIFAR and PCA-FashionMNIST, the tabular recipe of App. U for ISOLET (OpenML, 7,797 utterances, 617 spectral features), and the fixed GMM recipe of App. AK for a third target family (four modes, radius 2.5, component standard deviation 0.25, distribution seed 2718). No cell was added, removed, or re-run. As a protocol control, re-running PCA-CIFAR $d = 32$ under the same whitened recipe reproduced the reported values to three decimals (sliced-W_2 $0.0580 \to 0.0446$ at $K = 20$, $\varepsilon^* = 1.0$; Table 2 reports $0.058 \to 0.045$).

Three regularities emerge from Table 6: in each dimension ladder the selected noise moves toward the FM endpoint with d up to grid ties within seed noise (PCA-CIFAR $0.5 \to 0.3$; PCA-FashionMNIST $0.01 \to 0$; GMM C $0.1 \to 0.01 \to 0 \to 0.01$, where 0 and 0.01 are indistinguishable), the FM prediction is correct wherever $d \geq 400$, and the positive-noise prediction is correct wherever $d \leq 128$.

The four misses all lie at $176 \le d \le 384$ and have opposite predicted and empirical regime labels. Whitened PCA-CIFAR keeps a clear positive optimum at $d = 384$ ($\varepsilon^* = 0.3$, -10.4% sliced-W_2, 1.3 FM standard deviations), consistent with its earlier $\varepsilon^* = 0.1$ at $d = 512$; PCA-FashionMNIST is flat to within seed noise at $d = 256$ (all seven grid points within 0.162–0.172; the $\varepsilon^* = 0.01$ optimum is 0.0031 below FM against a seed standard deviation of 0.0036) and clearly FM at $d = 400$; and GMM family C reaches the FM side by $d = 176$, slightly below the canonical family's $(160, 192]$ crossover (its $\varepsilon^* = 0.01$ optima at $d = 176$ and $d = 320$ sit 0.0005 and 0.0011 below FM against seed standard deviations of 0.0058 and 0.0012). Under the fixed 0.05 threshold, the rule switches to its FM label too early for the tested whitened CIFAR cells and too late for the tested FashionMNIST and GMM cells. These results concern transfer of the fixed rule, not a re-estimate of its analytic zero threshold. The measured ladders support a dimension-associated ordering with family-specific crossover brackets; they do not establish a dataset-independent boundary.

Table 7: Full $K = 20$ grids of the new cells (seed-mean sliced-W_2; validation set for GMM cells, whose grid stops at 0.3 because no cell selected it). The last column is the independent test value at the selected checkpoint for GMM cells.

Cell	d	ε=0	0.01	0.03	0.1	0.3	0.5	1.0	test
PCA-CIFAR	192	0.0878	0.0827	0.0838	0.0754	0.0669	0.0635	0.0701	—
PCA-CIFAR	384	0.0750	0.0761	0.0707	0.0689	0.0672	0.0726	0.0872	—
PCA-FashionMNIST	256	0.1654	0.1623	0.1680	0.1673	0.1720	0.1692	0.1644	—
PCA-FashionMNIST	400	0.1373	0.1439	0.1412	0.1441	0.1464	0.1412	0.1430	—
ISOLET	617	0.1974	0.1977	0.1987	0.2031	0.2162	0.2298	0.2576	—
GMM family C	128	0.0200	0.0191	0.0178	0.0137	0.0245	—	—	0.0122
GMM family C	176	0.0575	0.0570	0.0626	0.0806	0.1422	—	—	0.0580
GMM family C	224	0.2169	0.2208	0.2232	0.2347	0.2749	—	—	0.2157
GMM family C	320	0.3952	0.3941	0.3975	0.4051	0.4261	—	—	0.3936

E THE FM-BASELINE GATE AND CAPACITY INTERVENTIONS

Table 8 contains the 34 baseline cells with retained $K = 20$ noise-grid summaries from the fixed-protocol GMM families, whitened PCA ladders, and standardized California, Breast Cancer, Digits, and ISOLET datasets. The six interventions are reported separately. We retain the normalization used to define the exploratory gate: FM/s_{ref} is FM SW_2 divided by a reference coordinate scale, with $s_{\mathrm{ref}} = 1$ for whitened PCA and standardized tabular coordinates. Constant columns remain in the tabular data; this unit preprocessing scale is not always the measured RMS standard deviation. Digits has three constant columns, so its actual RMS standard deviation is $\sqrt{61/64}$; using that denominator changes its normalized score from 0.157 to 0.161 without changing its label. For a GMM with centers of norm r and component standard deviation s, we use the nominal scale $s_{\mathrm{ref}} = \sqrt{s^2 + r^2/d}$. With active rank q and inactive noise ratio 0.1, it is $s^2_{\mathrm{ref}} = [q(s^2 + r^2/q) + (d - q)(0.1s)^2]/d$. These are declared scale conventions, not a claim that all tasks have equal statistical difficulty. A cell is counted as "noise helps" when the best grid point improves on FM by more than 5%. Figure 3 plots the same 34 rows. Tabular and PCA references come from their training data, whereas GMM uses independent validation samples; the association is therefore descriptive rather than external validation of a predictor.

We then ran six capacity interventions at fixed d. The hypothesis (noise gains more than 5% iff FM/$s_{\mathrm{ref}} \le 0.09$, so changing capacity at fixed d moves the regime) and the conditional prediction of each intervention were fixed in writing before the intervention was run; I5 and I6 were designed after I1–I3 had been observed and were fixed in the same way before running. Each intervention starts from the recipe of App. AK or App. AD.3 and changes the width, depth, or training settings listed in Table 9; the PCA interventions may change more than one setting. Table 9 reports the outcomes and Table 10 the grids.

The gate is stated at the $K = 20$ budget of the 34-cell analysis. At $K \ge 50$ whitened PCA-CIFAR selects $\varepsilon^* = 1.0$ with 24–37% gains at every $d \in \{64, 128, 256\}$ even where the $K = 20$ gain is below 5%, and its FM baseline is not monotone in K (at $d = 64$: 0.090, 0.067, 0.089, 0.107 for $K = 10, 20, 50, 100$). This behavior may involve discretization and learned-field errors, but these

Table 8: The 34 retained baseline grids at $K = 20$, sorted by FM/s_{ref}. The 0.09 threshold separates $10/12$ helping cells below from $1/22$ above. Two PCA-CIFAR cells just below the threshold gain 4.8–4.9%; the single helping cell above it has 569 samples.

Family	d	FM SW_2	FM/s_{ref}	ε^*	gain	$> 5\%$	note
PCA-CIFAR	32	0.0580	0.058	1	-23.0%	✓	
PCA-CIFAR	128	0.0585	0.059	0.5	-4.9%	—	
GMM family C	128	0.0200	0.060	0.1	-31.4%	✓	
GMM 8-mode	64	0.0315	0.066	0.3	-38.6%	✓	
PCA-CIFAR	64	0.0667	0.067	0.3	-26.7%	✓	
GMM 16-mode	160	0.0368	0.072	0.3	-21.5%	✓	
GMM 8-mode	128	0.0299	0.075	0.3	-35.6%	✓	
PCA-CIFAR	384	0.0750	0.075	0.3	-10.4%	✓	
PCA-CIFAR	512	0.0779	0.078	0.1	-7.1%	✓	
GMM 8-mode	160	0.0313	0.082	0.1	-19.5%	✓	
PCA-CIFAR	256	0.0838	0.084	0.1	-4.8%	—	
PCA-CIFAR	192	0.0878	0.088	0.5	-27.7%	✓	
GMM 16-mode	192	0.0467	0.095	0.01	-0.6%	—	
PCA-FashionMNIST	400	0.1373	0.137	0	$+0.0\%$	—	
Digits	64	0.1571	0.157	0.3	-2.8%	—	n=1797
Breast Cancer	30	0.1585	0.158	1	-9.2%	✓	n=569
PCA-FashionMNIST	256	0.1654	0.165	0.01	-1.9%	—	
GMM family C	176	0.0575	0.184	0.01	-0.9%	—	
ISOLET	617	0.1974	0.197	0	$+0.0\%$	—	n=7797
GMM 8-mode	192	0.0834	0.225	0	$+0.0\%$	—	
California Housing	8	0.4460	0.446	0.1	-0.2%	—	n=20640
GMM 8-mode	224	0.1739	0.482	0	$+0.0\%$	—	
GMM 16-mode	350	0.2829	0.624	0	$+0.0\%$	—	
GMM 8-mode	256	0.2461	0.696	0	$+0.0\%$	—	
GMM family C	224	0.2169	0.721	0	$+0.0\%$	—	
GMM 16-mode	400	0.3233	0.723	0	$+0.0\%$	—	
GMM 8-mode	300	0.3142	0.907	0	$+0.0\%$	—	
GMM 8-mode	350	0.3745	1.101	0	$+0.0\%$	—	
GMM 8-mode, active rank 400	400	0.4167	1.242	0	$+0.0\%$	—	
GMM 8-mode	400	0.4173	1.244	0	$+0.0\%$	—	
GMM 8-mode	450	0.4503	1.358	0	$+0.0\%$	—	
GMM family C	320	0.3952	1.380	0.01	-0.3%	—	
GMM 8-mode, active rank 200	400	0.5169	1.983	0.01	-0.0%	—	
GMM 8-mode, active rank 100	400	0.5635	2.637	0.01	-0.0%	—	

Table 9: Capacity interventions at fixed d (PCA cells are whitened; "w" is hidden width, "L" the number of layers). Each arrow reads baseline $\rightarrow$ intervened model. GMM FM/s_{ref} entries use validation means, including the harmonized I1–I2 baselines; recorded predictions are unchanged. "Pred." is the regime implied by the gate from the measured FM/s_{ref} of the intervened model; "Emp." is the empirical regime under the gate's criterion (bridge if the best grid point improves on FM by more than 5%, otherwise FM); $^{\times}$ marks the single miss. Gains are rounded; I4 gains 4.7%, below the threshold.

	Family	d	Change	FM/s_{ref}	ε^*	Gain	Pred.	Emp.
I1	GMM-8	192	width 512 $\rightarrow$ 1024	$0.225 \rightarrow 0.070$	$0 \rightarrow 0.1$	$+0\% \rightarrow -35\%$	bridge	bridge
I2	GMM-8	192	steps 30k $\rightarrow$ 90k	$0.225 \rightarrow 0.178$	$0 \rightarrow 0$	$+0\% \rightarrow +0\%$	FM	FM
I3	GMM-8	128	width 512 $\rightarrow$ 128	$0.075 \rightarrow 0.869$	$0.3 \rightarrow 0$	$-36\% \rightarrow +0\%$	FM	FM
I4	PCA-FashionMNIST	256	512w/5L/51k $\rightarrow$ 1024w/8L/100k	$0.165 \rightarrow 0.212$	$0.01 \rightarrow 0.03$	$-2\% \rightarrow -5\%$	FM	FM
I5	PCA-CIFAR	64	256w/4L $\rightarrow$ 64w/3L (20k steps)	$0.067 \rightarrow 0.074$	$0.3 \rightarrow 1$	$-27\% \rightarrow -12\%$	bridge	bridge
I6	PCA-CIFAR	64	256w/4L/20k $\rightarrow$ 16w/2L/5k	$0.067 \rightarrow 0.073$	$0.3 \rightarrow 0.5$	$-27\% \rightarrow -0\%$	bridge	FM$^{\times}$

Table 10: Full $K = 20$ grids of the interventions (seed-mean sliced-W_2; validation set and independent test value for GMM cells). For I4 the $\varepsilon = 0.03$ argmin has seed standard deviation 0.018 and the curve is flat within noise. For I6 the gate predicts a benefit (FM/$s_{\text{ref}} \leq 0.09$) that does not materialize: the curve is flat to 0.1%, its nominal argmin $\varepsilon^* = 0.5$ does not establish a positive-noise benefit, and Table 9 records it as a miss.

	Cell	ε=0	0.01	0.03	0.1	0.3	0.5	1.0	test
I1	GMM-8 d=192, width 1024	0.0259	0.0248	0.0214	0.0169	0.0193	—	—	0.0170
I2	GMM-8 d=192, 90k steps	0.0657	0.0698	0.0777	0.0893	0.0895	—	—	0.0652
I3	GMM-8 d=128, width 128	0.3480	0.3520	0.3554	0.3682	0.3773	—	—	0.3458
I4	PCA-FashionMNIST d=256, large	0.2118	0.2095	0.2019	0.2094	0.2028	0.2041	0.2022	—
I5	PCA-CIFAR d=64, width 64	0.0738	0.0698	0.0695	0.0691	0.0658	0.0648	0.0646	—
I6	PCA-CIFAR d=64, width 16	0.0731	0.0731	0.0731	0.0732	0.0731	0.0731	0.0732	—

experiments do not separate their contributions; Theorem C alone does not establish either as the dominant cause.

The two GMM interventions that cross the exploratory threshold (I1 down, I3 up) flip the benefit/no-benefit outcome. I2, I4, and I5 retain that outcome, with I5's gain shrinking as its baseline worsens. I6 is a failure of the baseline diagnostic: reducing capacity removes the benefit, but the low sliced-W_2 baseline does not reveal this loss. The conditional rule agrees in five of six cases, supporting further study rather than a necessary-and-sufficient criterion. The fixed-dimension interventions establish dependence on the training configuration; the association with normalized FM error does not identify a causal mediator or validate the unfitted extension in App. K.

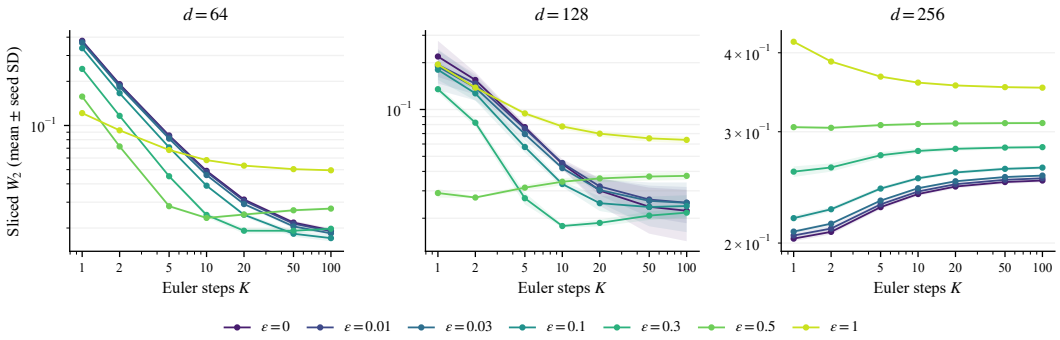


Figure 5: Fixed-protocol GMM budget sweep at $d = 64, 128, 256$. Curves show all seven tested ε values at $K \in \{1, 2, 5, 10, 20, 50, 100\}$. Lines join measured means; bands show sample standard deviations over three model seeds. These evaluation-grid minima are descriptive; independent validation/test selection is reported for the primary $K = 20$ sweep in Table 32.

F Descriptive PCA Scale Per Dataset

For each cross-domain dimension proxy, we report the second-moment PCA scale

$$R(d_{\text{eff}}) := \sqrt{\mathbb{E}_{\mathbf{x}_0 \sim p_{\text{data}}} \left\| \Pi_{99}(\mathbf{x}_0 - \bar{\mathbf{x}_0}) \right\|_2^2}, \tag{12}$$

where Π_{99} projects onto the leading components capturing 99% of the variance. This second-moment scale is distinct from the posterior constants in Theorem C.

†PCA-90% proxy: FashionMNIST $d_{\text{eff}} \approx 80$, ShapeNet $d_{\text{eff}} = 250$ (both SB). PCA-99%: 428/486 listed for completeness.

Across datasets, a log-log regression gives $R(d_{\text{eff}}) \propto d_{\text{eff}}^{0.32}$ ($R^2 = 0.72$). The growth varies with the chosen dimension proxy and data normalization, so R is descriptive rather than a dimension-uniform bound.

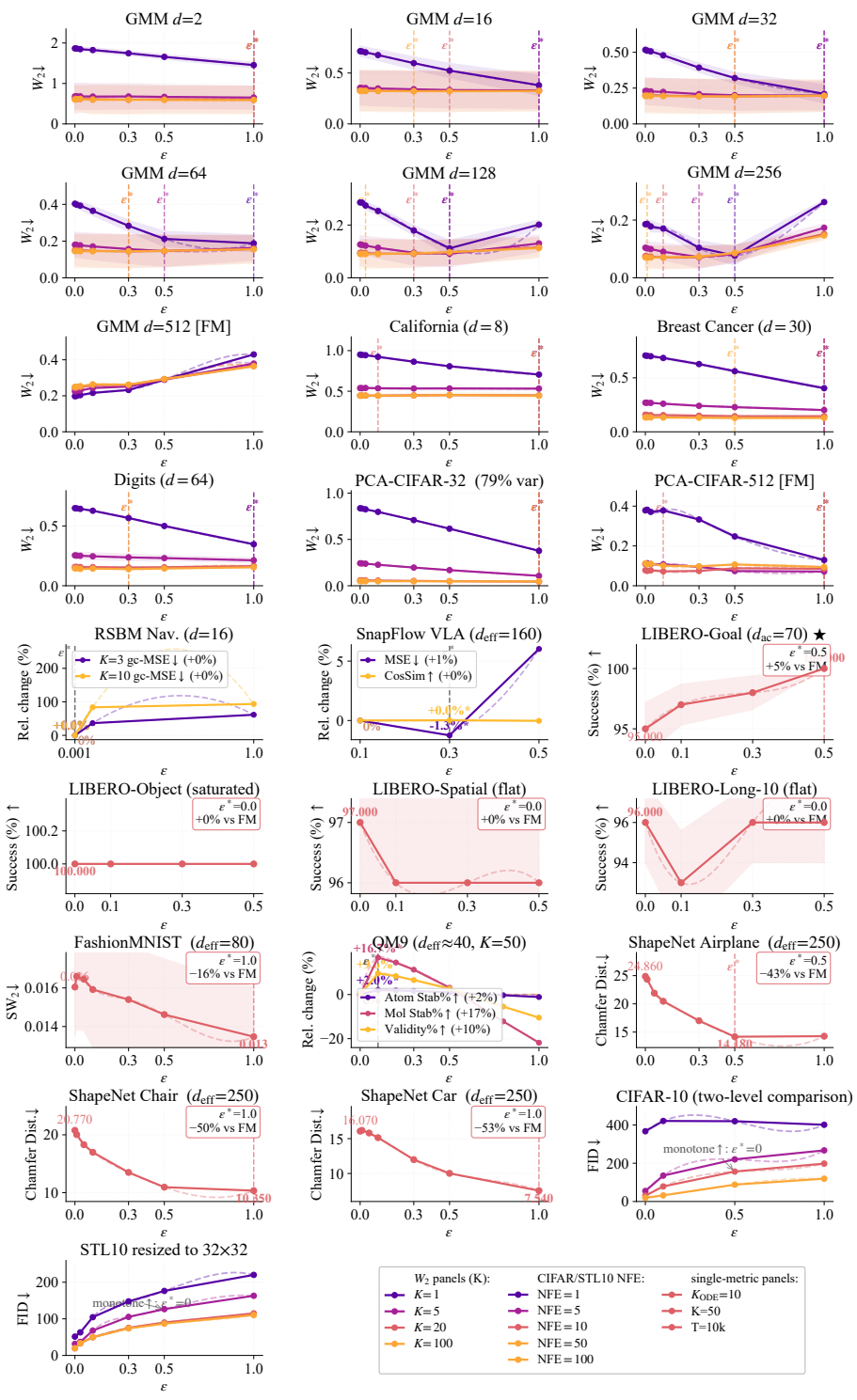


Figure 6: All ε sweeps across task families: original task-specific overview. The historical GMM ladder is excluded from fixed-protocol performance comparisons; App. X specifies the protocols and original-label qualifications. Other panels retain task-specific training, metrics, and sampling settings and do not define a common dimensional boundary. Fixed-protocol GMM budget curves are reported separately in Fig. 5.

Table 11: Descriptive PCA scale across datasets. UCI data are standardized before PCA, whereas GMM and images use their training scales, so values are not directly comparable across families. STL10 is omitted because the raw samples used for this calculation were unavailable.

Dataset	d (ambient)	d_{eff} (PCA-99%)	$R(d_{\text{eff}})$
2D 8-Gauss	2	2	3.03
Moons	2	2	1.00
Wine (sklearn)	13	12	3.59
Breast cancer (UCI)	30	17	5.45
Digits (sklearn)	64	54	7.77
California (UCI)	8	7	2.82
GMM $d=16$	16	15	3.01
GMM $d=64$	64	63	3.69
GMM $d=256$	256	251	5.53
GMM $d=512$	512	500	7.28
FashionMNIST†	784	428 (PCA-99%) / ≈80 (PCA-90%)	8.25
ShapeNet (Airplane)†	6144	486 (PCA-99%) / 250 (PCA-90%)	21.82
CIFAR-10	3072	592	13.77

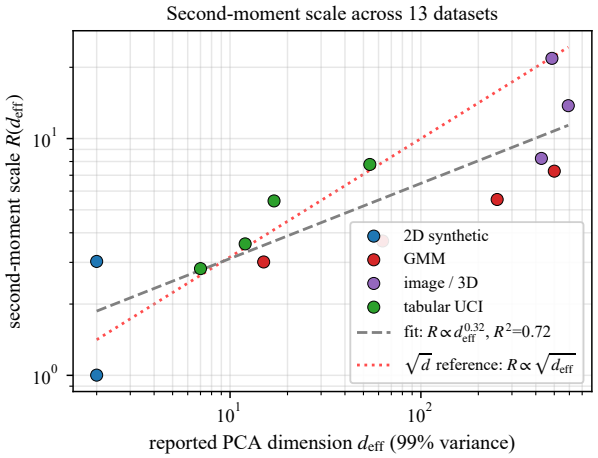


Figure 7: Second-moment PCA scale across task-specific dimension proxies. The scale is descriptive and is not a dimension-uniform bound.

G Summary of Results and Their Scope

To make explicit what each result establishes and what remains empirical, Table 12 organizes the analytic identities, conditional statements, and model checks by evidence and scope.

Table 12: Summary of results. Scope records what each result establishes and what remains empirical.

Result	**Status**	**Evidence**	**Scope**
Thm. A	analytic	fixed-predictor convergence proof	finite-moment and continuity assumptions
Prop. B	analytic	direct Gaussian target calculation	raw ambient target energy only
Thm. C	conditional	posterior-moment proof; sampled exact GMM posterior norms (App. P.3)	sampled checks do not establish uniform assumptions
Model D	phenomenological	Fixed-protocol GMM sweep: positive noise through $d = 160$, FM at $d \geq 192$ on the tested grid	no budget term; crossover is protocol-dependent (App. AK)

H Proof of Theorem A

Complete proof of Theorem 3. **(i)** $\sigma_\tau = \sqrt{\varepsilon\tau(1-\tau)} \le \sqrt{\varepsilon}$, so $\|\sigma_\tau \mathbf{z}\| \le \sqrt{\varepsilon}\|\mathbf{z}\| \to 0$ a.s.

(ii) From Eq. (2): $\mathbf{v}_\tau(\varepsilon) - (\mathbf{x}_1 - \mathbf{x}_0) = \frac{(1-2\tau)\sqrt{\varepsilon}}{2\sqrt{\tau(1-\tau)}}\,\mathbf{z}$. Under the cutoff $\tau \in [\delta, 1-\delta]$ the coefficient is bounded by $1/[4\delta(1-\delta)]$; taking expectation over $\mathbf{z}$ and averaging over $\tau \sim \mathcal{U}[\delta, 1-\delta]$ gives $\mathbb{E}_\tau \mathbb{E}\|\mathbf{v}_\tau(\varepsilon) - (\mathbf{x}_1 - \mathbf{x}_0)\|^2 = \varepsilon\, d\, c_0(\delta)/(1-2\delta) \to 0$ as $\varepsilon \to 0$ for any fixed $\delta > 0$.

(iii) By (i), continuity of the fixed predictor gives $\mathbf{v}_\theta(\mathbf{x}_\tau, \tau) \to \mathbf{v}_\theta((1-\tau)\mathbf{x}_0 + \tau\mathbf{x}_1, \tau)$ almost surely; by (ii), the target converges in L^2 to $\mathbf{x}_1 - \mathbf{x}_0$. The linear-growth assumption, finite second moments of $\mathbf{x}_0, \mathbf{x}_1, \mathbf{z}$, and the cutoff bound supply an integrable envelope for the squared residual. Dominated convergence therefore gives $\mathcal{L}_\varepsilon(\theta) \to \mathcal{L}_{\mathrm{FM},\delta}(\theta)$.

(iv) The ODE $d\mathbf{x}/d\tau = \mathbf{v}_\theta(\mathbf{x}, \tau)$ depends only on $\mathbf{v}_\theta$; ε enters only through training. □

I Proof of Proposition B

Complete proof of Proposition 4. The bridge-induced target term is $b_\tau(\mathbf{z}) = \frac{(1-2\tau)\sqrt{\varepsilon}}{2\sqrt{\tau(1-\tau)}}\mathbf{z}$. Because $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_d)$,

$$\mathbb{E}\|b_\tau(\mathbf{z})\|^2 = \frac{(1-2\tau)^2\varepsilon}{4\tau(1-\tau)}\,\mathbb{E}\|\mathbf{z}\|^2 = \frac{(1-2\tau)^2\varepsilon d}{4\tau(1-\tau)}.$$

Integrating over $[\delta, 1-\delta]$ gives Eq. (5). The integrand behaves as $1/(4\tau)$ near zero (and symmetrically near one), so the cutoff is essential. This proof concerns only the unconditional target term and makes no posterior-variance claim. □

J Proof of Theorem C

Lemma 7 (A sufficient posterior-contraction condition)**.** Let $X \in \mathbb{R}^d$ have a log-concave density proportional to $\exp[-V(x)]$, with V convex, and let $Y = X + hZ$ for $h > 0$ and $Z \sim \mathcal{N}(0, \mathbf{I}_d)$ independent of X. Then, for every y with positive marginal density,

$$\big\|\mathrm{Var}(X \mid Y = y)\big\|_{\mathrm{op}} \le h^2. \tag{13}$$

Proof. The posterior density is proportional to $\exp[-V(x) - \|y - x\|^2/(2h^2)]$ and is therefore h^{-2}-strongly log-concave. The Brascamp–Lieb covariance inequality (Brascamp & Lieb, 1976) gives $\mathrm{Var}(X \mid Y = y) \preceq h^2\mathbf{I}_d$. □

Lemma 7 is one sufficient route to the second assumption of Theorem 5; general sub-Gaussianity alone is not sufficient for a uniform pointwise $O(h^2)$ posterior bound.

Complete proof of Theorem 5. Let $h_\tau := \sigma_\tau = \sqrt{\varepsilon\tau(1-\tau)}$ and write $p_\tau(x) := \mathbb{E}_{(\mathbf{x}_0,\mathbf{x}_1)}[q_\varepsilon(x|\mathbf{x}_0, \mathbf{x}_1)]$ for the marginal density at time τ. The marginal velocity is the Nadaraya–Watson estimator

$$\mathbf{u}_\tau(x;\varepsilon) \;=\; \frac{\mathbb{E}_{(\mathbf{x}_0,\mathbf{x}_1)}\big[\mathbf{v}_\tau\, q_\varepsilon(x|\mathbf{x}_0,\mathbf{x}_1)\big]}{\mathbb{E}_{(\mathbf{x}_0,\mathbf{x}_1)}\big[q_\varepsilon(x|\mathbf{x}_0,\mathbf{x}_1)\big]} \;=\; \mathbb{E}\big[\mathbf{v}_\tau \,\big|\, \mathbf{x}_\tau = x;\varepsilon\big], \tag{14}$$

with Gaussian kernel $q_\varepsilon(x|\mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}(x; \boldsymbol{\mu}_\tau, h_\tau^2\mathbf{I}_d)$.

Step 1 (Stein log-derivative). For the Gaussian kernel, $\nabla_x q_\varepsilon(x|\mathbf{x}_0,\mathbf{x}_1) = h_\tau^{-2}(\boldsymbol{\mu}_\tau - x)\, q_\varepsilon$. Applying the quotient rule to the Nadaraya–Watson form and factoring this identity yields the standard kernel-regression gradient (cf. Nadaraya, 1964; Efron, 2011)

$$\nabla_x \mathbf{u}_\tau(x;\varepsilon) \;=\; \frac{1}{h_\tau^2}\,\mathbb{E}\big[\,(\mathbf{v}_\tau - \mathbf{u}_\tau(x))\,(\boldsymbol{\mu}_\tau - x)^\top \,\big|\, \mathbf{x}_\tau = x;\varepsilon\,\big] \;+\; \underbrace{\mathbb{E}\big[\nabla_x \mathbf{v}_\tau(x;\mathbf{x}_0,\mathbf{x}_1) \,\big|\, \mathbf{x}_\tau = x;\varepsilon\big]}_{=\frac{1-2\tau}{2\tau(1-\tau)}\mathbf{I}_d \text{ (drift trend)}}, \tag{15}$$

where the explicit drift-trend term comes from the deterministic $\gamma_\tau(x) = \frac{1-2\tau}{2\tau(1-\tau)}x$ piece of $\mathbf{v}_\tau$ in Eq. (2).

Step 2 (Posterior centering). The first term of (15) is a posterior covariance between two random functions of $(\mathbf{x}_0, \mathbf{x}_1)$ under the conditional law. Because $\mathbf{v}_\tau - \mathbf{u}_\tau(x)$ has zero conditional mean, we may add any deterministic-given-x shift to the second factor without changing the value:

$$\mathbb{E}\big[(\mathbf{v}_\tau - \mathbf{u}_\tau)(\boldsymbol{\mu}_\tau - x)^\top\big] \;=\; \mathbb{E}\big[(\mathbf{v}_\tau - \mathbf{u}_\tau)(\boldsymbol{\mu}_\tau - \bar{\boldsymbol{\mu}}_\tau)^\top\big], \qquad \bar{\boldsymbol{\mu}}_\tau := \mathbb{E}[\boldsymbol{\mu}_\tau \mid \mathbf{x}_\tau = x; \varepsilon]. \tag{16}$$

This identity expresses the gradient term through posterior covariance. It lets us apply the posterior moment assumptions directly; centering alone does not establish their dependence on h_τ.

Step 3 (Operator-norm bound). For any pair of conditionally centered random vectors $\mathbf{a}, \mathbf{b}$ and any unit vectors u, w, Cauchy–Schwarz gives $u^\top \mathbb{E}[\mathbf{a}\mathbf{b}^\top] w = \mathbb{E}[(u^\top \mathbf{a})(\mathbf{b}^\top w)] \le \sqrt{\mathbb{E}(u^\top \mathbf{a})^2}\sqrt{\mathbb{E}(\mathbf{b}^\top w)^2}$; taking suprema over u, w yields $\|\mathbb{E}[\mathbf{a}\mathbf{b}^\top]\|_{\mathrm{op}} \le \|\mathrm{Var}(\mathbf{a})\|_{\mathrm{op}}^{1/2} \|\mathrm{Var}(\mathbf{b})\|_{\mathrm{op}}^{1/2}$. Applied with $\mathbf{a} = \mathbf{v}_\tau - \mathbf{u}_\tau$ and $\mathbf{b} = \boldsymbol{\mu}_\tau - \bar{\boldsymbol{\mu}}_\tau$, both centered under the conditional law given $\mathbf{x}_\tau = x$,

$$\Big\|\mathbb{E}\big[(\mathbf{v}_\tau - \mathbf{u}_\tau)(\boldsymbol{\mu}_\tau - \bar{\boldsymbol{\mu}}_\tau)^\top\big]\Big\|_{\mathrm{op}} \;\le\; \big\|\mathrm{Var}(\mathbf{v}_\tau \mid \mathbf{x}_\tau = x; \varepsilon)\big\|_{\mathrm{op}}^{1/2} \; \big\|\mathrm{Var}(\boldsymbol{\mu}_\tau \mid \mathbf{x}_\tau = x; \varepsilon)\big\|_{\mathrm{op}}^{1/2}. \tag{17}$$

The theorem assumes the two factors satisfy

$$\big\|\mathrm{Var}(\mathbf{v}_\tau \mid \mathbf{x}_\tau = x; \varepsilon)\big\|_{\mathrm{op}}^{1/2} \le V_0, \tag{18}$$

and

$$\big\|\mathrm{Var}(\boldsymbol{\mu}_\tau \mid \mathbf{x}_\tau = x; \varepsilon)\big\|_{\mathrm{op}} \le C_\mu h_\tau^2. \tag{19}$$

These are explicit posterior conditions, not consequences of a measured PCA radius. Lemma 7 supplies Eq. (19) when the latent mean law is log-concave; Eq. (18) must still be controlled separately.

Combining,

$$\Big\|\mathbb{E}\big[(\mathbf{v}_\tau - \mathbf{u}_\tau)(\boldsymbol{\mu}_\tau - \bar{\boldsymbol{\mu}}_\tau)^\top\big]\Big\|_{\mathrm{op}} \;\le\; V_0\sqrt{C_\mu}\, h_\tau. \tag{20}$$

Step 4 (Slope assembly). Substituting (20) into (15) via (16) gives

$$\big\|\nabla_x \mathbf{u}_\tau(x; \varepsilon)\big\|_{\mathrm{op}} \;\le\; \frac{|1-2\tau|}{2\tau(1-\tau)} \;+\; \frac{V_0\sqrt{C_\mu}}{\sqrt{\varepsilon\,\tau(1-\tau)}}, \tag{21}$$

which is exactly Eq. (8). The factor h_τ^{-1} follows from the cross-covariance bound $V_0\sqrt{C_\mu}\, h_\tau$ in Eq. (20), obtained from the two posterior assumptions by operator-norm Cauchy–Schwarz, and division by h_τ^2 in Eq. (15).

Step 5 (Where dimension can enter). The algebra introduces no trace factor, but V_0 and C_μ may themselves depend on dimension or on the region Ω. The bound is dimension-free only for a problem sequence on which these posterior constants are uniform. This is the precise contrast with Proposition B, whose raw target energy contains an explicit trace factor d.

Monotonicity in ε. If V_0 and C_μ can be chosen uniformly over the compared ε, the displayed upper bound decreases as h_τ increases. A decreasing upper bound does not imply that the actual Jacobian norm decreases; the empirical $4.82 \to 1.83$ proxy drop reported in §3.5 is a separate observation.

Limiting behavior at $\varepsilon \to 0$. As $\varepsilon \to 0$, $h_\tau \to 0$ and this upper bound becomes vacuous; it makes no claim about the actual limiting Lipschitz constant. Under the independent Gaussian endpoint construction, the marginal is already convolved with variance τ^2 at $\varepsilon = 0$ for $\tau > 0$. The extra bridge kernel used in this bound is not the only source of smoothing. □

K Heuristic Derivation of Model D

The two-term error model of Model D,

$$E^2(\varepsilon; d) \;=\; \underbrace{\frac{A}{1 + \beta_0 \varepsilon}}_{\text{fitting error (regularized by smoothing)}} \;+\; \underbrace{d\,\hat{c}_0\,\varepsilon}_{\text{phenomenological penalty}}, \tag{22}$$

is a phenomenological ansatz. This appendix motivates the selected features and derives the resulting operating conditions.

For the fitting error term $A/(1+\beta_0\varepsilon)$, Theorem C shows how Gaussian smoothing can reduce a regularity upper bound under explicit posterior conditions. We choose the decreasing, saturating form

$$\text{fitting error} \;\approx\; \frac{A}{1+\beta_0\varepsilon}, \tag{23}$$

as a two-parameter empirical approximation over $\varepsilon \in [0,1]$. Here A is the rescaled FM baseline and β_0 is a curvature parameter; no cited generalization theorem implies this particular rational dependence.

For the phenomenological penalty term $d\,\hat{c}_0\,\varepsilon$, Proposition B gives the raw bridge correction energy $C_{\text{bridge}} = \varepsilon d\, c_0(\delta)$. We preserve only its linear features in d and ε and replace the coefficient by a separate coefficient $\hat{c}_0$.

The model omits budget dependence; this is a modeling choice, not a consequence of using a deterministic ODE. The sampler injects no per-step noise, but its integration error can depend on both K and the learned field. In the fixed-protocol GMM sweep, smoothing that offsets coarse-integration error is one possible explanation for the decrease of ε^* with K. The PCA budget trend differs (App. M).

The local sign change of the model follows from setting $\partial_\varepsilon E^2|_{\varepsilon=0} = 0$, which gives $d^* = A\beta_0/\hat{c}_0$. Substituting the fixed illustrative constants gives $d^* \approx 364$, the local derivative sign change of the model.

Only the linear feature $d\varepsilon$ is inherited from Proposition B; the conversion to sampling error and the decreasing fitting term are empirical modeling choices examined through within-protocol sweeps and cross-domain diagnostics.

Finally, we extend the model to capacity-limited settings. The single smoothing-responsive term assumes an approximately stable fitting response; however, Sec. 4.2 shows that across protocols the FM baseline error rises where the model under-fits and that bridge noise then stops helping. Under Eq. (22) a larger A would predict the opposite, since $\partial_\varepsilon E^2|_{\varepsilon=0} = d\hat{c}_0 - A\beta_0$. One possible extension splits the fitting error,

$$E^2(\varepsilon;\xi) \;=\; A_{\text{cap}}(\xi) \;+\; \frac{B(\xi)}{1+\beta(\xi)\,\varepsilon} \;+\; d\,\hat{c}_0\,\varepsilon, \qquad \xi = (d,\ \text{capacity},\ \text{data},\ K), \tag{24}$$

where $A_{\text{cap}} \geq 0$ is the part of the FM error that capacity limits impose and smoothing cannot remove, $B > 0$ and $\beta > 0$ describe the smoothing-responsive part, and $\hat{c}_0$ is kept as a fixed phenomenological coefficient of the linear feature $d\varepsilon$. Bridge noise is beneficial near $\varepsilon = 0$ iff $B(\xi)\,\beta(\xi) > d\,\hat{c}_0$. In the well-fit regime $A_{\text{cap}} \approx 0$ and $B \approx A$, recovering Eq. (22). Under-fitting can increase A_{cap} while reducing $B\beta$, so the total FM error $E^2(0;\xi) = A_{\text{cap}} + B$ can grow even as the smoothing benefit disappears; a pure transfer between the two terms would leave the FM error unchanged, so within this interpretation, the observed rise of the FM baseline would require A_{cap} to grow outright. Because ξ includes K, the extension admits an implicit budget dependence through A_{cap}, B, β, unlike the fixed Model D, which has no budget term. The functions A_{cap}, B, β are not fitted in this paper; Eq. (24) records one possible structure consistent with the gate and capacity interventions, and the normalized FM baseline of Sec. 4.2 is an exploratory proxy for this interpretation.

L Proof of the Minimizer of Model D

Complete proof of Proposition 6. Let $f(\varepsilon) = E^2(\varepsilon; d)$ from Eq. (9), so $f'(\varepsilon) = d\hat{c}_0 - A\beta_0/(1+\beta_0\varepsilon)^2$ and $f'(0) = d\hat{c}_0 - A\beta_0 = \hat{c}_0(d-d^*)$. Because $(1+\beta_0\varepsilon)^{-2}$ is strictly decreasing, f' is strictly increasing on $[0,\infty)$, and $f''(\varepsilon) = 2A\beta_0^2/(1+\beta_0\varepsilon)^3 > 0$, so f is strictly convex.

Part (a): high dimension. For $d > d^*$, $f'(0) > 0$ and f' increasing give $f' > 0$ on $[0,1]$; f is strictly increasing and $\varepsilon = 0$ is the global minimizer on $[0,1]$. At $d = d^*$, $f'(0) = 0$ and $f'' > 0$, so $\varepsilon = 0$ remains the strict global minimizer.

Part (b): low dimension. For $d < d^*$, $f'(0) < 0$; by convexity the unique stationary point solves $(1+\beta_0\varepsilon_{\text{stat}})^2 = A\beta_0/(d\hat{c}_0) = d^*/d$, giving Eq. (10), and it is the unconstrained global minimizer.

It lies below 1 iff $\sqrt{d^*/d} < 1 + \beta_0$, i.e. $d > d^*/(1+\beta_0)^2$; otherwise f is decreasing on all of $[0,1]$ and the constrained minimizer is the boundary $\varepsilon^* = 1$.

Monotonicity in the SB regime: For $d < d^*$, Eq. (10) is strictly decreasing in d: lower dimension calls for more bridge noise under the model. □

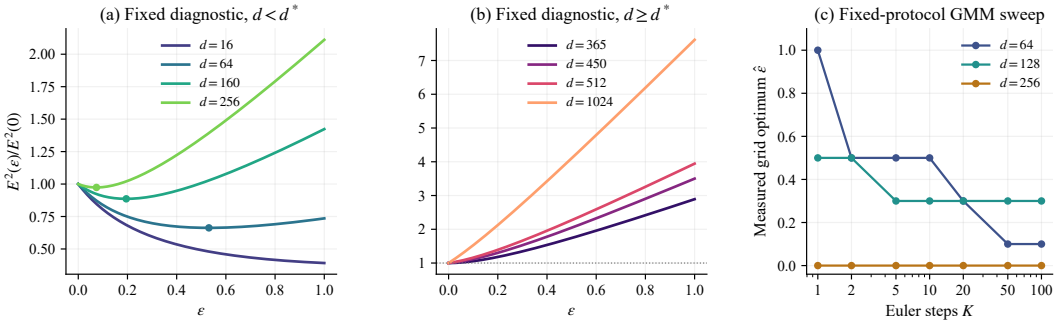


Figure 8: Shape of Model D and measured budget trend. (a) Below d^*, positive-noise minima move toward zero as d grows; dots mark interior optima. (b) At and above d^*, E^2 is strictly increasing, so $\varepsilon = 0$ is the global minimizer. (c) Fixed-protocol GMM grid optima at $d = 64, 128, 256$ across budgets; the selected values decline with K at $d = 64$ and 128, while $d = 256$ selects $\varepsilon = 0$ at every tested K (App. M).

Remark 1 ($\hat{c}_0$-free form of $\varepsilon_{\text{stat}}$)**.** The closed form in Eq. (10) can be rewritten purely in terms of (d^*, β_0):

$$\varepsilon_{\text{stat}}(d) \;=\; \frac{\sqrt{d^*} - \sqrt{d}}{\sqrt{d}\,\beta_0}, \qquad d < d^*, \tag{25}$$

interior for $d > d^*/(1+\beta_0)^2$ and clipped to $[0,1]$ otherwise. This reparameterization removes the separate coefficient once d^* is specified; it does not prove invariance to the training cutoff or experimental protocol. Its clipped form is exactly the diagnostic used in Tab. 5.

M Sampling-Budget Conventions

The paper reports three sampling-budget conventions, denoted separately below.

The first convention, K_{ODE}, counts ODE solver steps. This is the K used in the fixed-protocol GMM sweep and multi-budget table, in PCA-CIFAR, and in tabular experiments: we sweep $K \in \{1, 2, 5, \ldots, 100\}$ at inference only on the same trained network. The error model Eq. (9) contains no budget term.

In the fixed-protocol multi-budget sweep, selected optima move toward smaller ε as K_{ODE} grows at $d = 64$ and 128, while $d = 256$ selects FM at every tested budget. Smoothing of the learned field is one possible explanation for a benefit under coarse integration. This is not a general theorem that optimal noise decreases with K: the measured whitened PCA-CIFAR curves behave differently (App. E), and Proposition B alone gives no budget-independent sample-error penalty.

The second, K_{distill}, is the number of function evaluations of a distilled student. This is the K used in SnapFlow (Luan et al., 2026b) and other 1-NFE methods, and is fundamentally different from K_{ODE}: the student is trained on a shortcut objective rather than by integrating the teacher K_{distill} times. Published one-step distillation results are reported separately in App. AB and are not predictions of the error model.

The third, K_{solver}, is the solver step count of published case studies. In RSBM (Luan et al., 2026a) and similar published results, K denotes the number of Heun steps under a different sampler rather than Euler steps. These case studies appear in the corresponding appendices as context; their ε^* values are not predictions of the fixed numerical diagnostic.

Throughout the main text, "K" refers to K_{ODE} unless otherwise stated. Both the local sign change and global minimizer of Model D are K-independent by construction; the measured optima need not be. Under distillation, ε^* is set by a different objective.

N Ambient and Operational Effective Dimension

Proposition B uses the ambient dimension of the isotropic bridge process. It becomes a d_{eff} statement when both the model and noise prior are restricted to a d_{eff}-dimensional subspace, as in PCA-CIFAR. FashionMNIST, ShapeNet, and robotics instead use PCA or action-chunk coordinates while retaining different ambient processes. Their entries in Table 2 are operational summaries; Table 11 reports both PCA-90% and PCA-99% values where available.

O Additional Derivation for Proposition B

We expand here the algebra omitted from the main-text proof of Proposition 4.

O.1 Conditional Velocity in Closed Form

Starting from Eq. (2), substitute $\mathbf{x}_\tau = (1-\tau)\mathbf{x}_0 + \tau\mathbf{x}_1 + \sigma_\tau \mathbf{z}$ with $\sigma_\tau = \sqrt{\varepsilon\tau(1-\tau)}$:

$$\begin{aligned}\mathbf{v}_\tau &= (\mathbf{x}_1 - \mathbf{x}_0) + \frac{1-2\tau}{2\tau(1-\tau)}\,\sigma_\tau\,\mathbf{z} \\ &= (\mathbf{x}_1 - \mathbf{x}_0) + \frac{(1-2\tau)\sqrt{\varepsilon}}{2\sqrt{\tau(1-\tau)}}\,\mathbf{z}. \end{aligned} \tag{26}$$

The deterministic part $(\mathbf{x}_1 - \mathbf{x}_0)$ is the FM target; the second term vanishes as $\sqrt{\varepsilon}$.

O.2 Scope of the Raw Target Calculation

Proposition B does not decompose $\mathrm{Var}(\mathbf{v}_\tau \mid \mathbf{x}_\tau)$: conditioning couples $\mathbf{x}_0$, $\mathbf{x}_1$, and $\mathbf{z}$ in a distribution-dependent way. The phenomenological $\hat{c}_0$ used later is therefore a separate empirical coefficient, not an estimate of the analytical integral below.

O.3 Integrated Correction Trace

The integrated drift cost over the truncated interval $[\delta, 1-\delta]$ is $C_{\text{bridge}}(\delta) = \int_\delta^{1-\delta} \mathrm{Tr}[\varepsilon \boldsymbol{\Sigma}_{\text{bridge}}(\tau)]\,d\tau$, where $\boldsymbol{\Sigma}_{\text{bridge}}(\tau) = \frac{(1-2\tau)^2}{4\tau(1-\tau)}\mathbf{I}_d$. Trace is $d \cdot \frac{(1-2\tau)^2}{4\tau(1-\tau)}$, so

$$C_{\text{bridge}}(\varepsilon;\delta) = \varepsilon\, d \int_\delta^{1-\delta} \frac{(1-2\tau)^2}{4\tau(1-\tau)}\,d\tau = \varepsilon\, d\, c_0(\delta). \tag{27}$$

Evaluating $c_0(\delta)$. Using the identity $(1-2\tau)^2 = 1-4\tau(1-\tau)$ and partial fractions $1/[4\tau(1-\tau)] = 1/(4\tau) + 1/(4(1-\tau))$:

$$c_0 = \frac{1}{4}\int_{-1}^{1} \frac{u^2}{(1-u^2)/4}\cdot\frac{du}{2} = \frac{1}{2}\int_{-1}^{1}\frac{u^2}{1-u^2}\,du. \tag{28}$$

Using the partial fraction $\frac{u^2}{1-u^2} = -1 + \frac{1}{1-u^2} = -1 + \frac{1}{2}\left[\frac{1}{1-u} + \frac{1}{1+u}\right]$:

$$\int \frac{u^2}{1-u^2}\,du = -u + \tfrac{1}{2}\log\left|\tfrac{1+u}{1-u}\right| + \text{const.} \tag{29}$$

The formal limits $u \to \pm 1$ produce a logarithmic divergence, since the integrand $\frac{(1-2\tau)^2}{4\tau(1-\tau)} \sim \frac{1}{4\tau}$ as $\tau \to 0$ (and symmetrically at $\tau = 1$).

Truncated-time objective. Sampling $\tau \sim \mathcal{U}[\delta, 1-\delta]$ with $\delta > 0$ gives the finite value

$$c_0(\delta) \;=\; \int_\delta^{1-\delta} \frac{(1-2\tau)^2}{4\tau(1-\tau)}\,d\tau \;=\; \tfrac{1}{2}\log\frac{1-\delta}{\delta} - 1 + 2\delta. \tag{30}$$

This gives $c_0 \approx 0.57$ at $\delta = 0.05$ and $c_0 \approx 2.45$ at $\delta = 10^{-3}$. The phenomenological $\hat{c}_0$ in Model D has a different operational definition. The algebra does not guarantee cutoff invariance of empirical parameters, so the numerical operating boundary must be reported together with the protocol.

Remark 2 (Truncated loss)**.** The corresponding training objective is $\mathcal{L}_\varepsilon^{(\delta)}(\theta) = \mathbb{E}_{\tau\sim\mathcal{U}[\delta,1-\delta]}\mathbb{E}_{\mathbf{x}_0,\mathbf{x}_1,\mathbf{z}}\|\mathbf{v}_\theta(\mathbf{x}_\tau,\tau) - \mathbf{v}_\tau\|^2$, under which c_0 is finite and the $O(d)$-vs-$O(1)$ asymmetry of Proposition B / Theorem C is preserved.

P Additional Derivation for Theorem C

P.1 Stein-Style Identity for Kernel Regression Gradient

Recall $\mathbf{u}_\tau(x;\varepsilon) = \mathbb{E}[\mathbf{v}_\tau \mid \mathbf{x}_\tau = x]$ as in Eq. (6), with Gaussian kernel q_ε. Using the Bayes/Tweedie identity $\nabla_x \log p_\varepsilon(\mathbf{x}_\tau = x) = (\boldsymbol{\mu}_\tau^* - x)/\sigma_\tau^2$ (where $\boldsymbol{\mu}_\tau^* = \mathbb{E}[\boldsymbol{\mu}_\tau \,|\, \mathbf{x}_\tau = x]$):

$$\begin{aligned}
\nabla_x \mathbf{u}_\tau(x;\varepsilon) &= \nabla_x \mathbb{E}[\mathbf{v}_\tau \,|\, x] \\
&= \frac{1}{\sigma_\tau^2}\big(\mathbb{E}[\mathbf{v}_\tau\,(\boldsymbol{\mu}_\tau - \mathbf{x}_\tau)^\top \,|\, x] - \mathbf{u}_\tau(x)\,\mathbb{E}[(\boldsymbol{\mu}_\tau - \mathbf{x}_\tau)^\top \,|\, x]\big) \\
&\quad + \frac{1-2\tau}{2\tau(1-\tau)}\mathbf{I}_d \\
&= \frac{1}{\sigma_\tau^2}\operatorname{Cov}(\mathbf{v}_\tau, \boldsymbol{\mu}_\tau \,|\, \mathbf{x}_\tau = x;\varepsilon) + \frac{1-2\tau}{2\tau(1-\tau)}\mathbf{I}_d. \qquad (31)
\end{aligned}$$

P.2 Operator-Norm Bound

By the operator-norm Cauchy–Schwarz inequality for outer-product expectations derived in Step 3 of App. J,

$$\|\operatorname{Cov}(\mathbf{v}_\tau, \boldsymbol{\mu}_\tau \,|\, \mathbf{x}_\tau = x;\varepsilon)\|_{\mathrm{op}} \le \sqrt{\|\operatorname{Var}(\mathbf{v}_\tau \,|\, x)\|_{\mathrm{op}} \cdot \|\operatorname{Var}(\boldsymbol{\mu}_\tau \,|\, \mathbf{x}_\tau = x)\|_{\mathrm{op}}}. \qquad (32)$$

Both factors are operator norms rather than traces. Under the two assumptions of Theorem 5,

$$\|\operatorname{Cov}\|_{\mathrm{op}} \le V_0\sqrt{C_\mu}\,\sigma_\tau \quad \Longrightarrow \quad \|\nabla_x \mathbf{u}_\tau\|_{\mathrm{op}} \le \frac{V_0\sqrt{C_\mu}}{\sqrt{\varepsilon\tau(1-\tau)}} + \frac{|1-2\tau|}{2\tau(1-\tau)}. \qquad (33)$$

No explicit d appears, but the posterior constants may depend on d. Lemma 7 supplies the contraction factor for log-concave latent means; it does not apply to arbitrary multimodal sub-Gaussian data. For the controlled GMM family the constants can be measured exactly; see the next subsection.

P.3 Measured Posterior Constants on the Controlled GMM Family

For the isotropic 8-mode GMM family (centers on the radius-3 sphere, component std 0.3) with $x_1 \sim \mathcal{N}(0, I)$ and bridge latent $z \sim \mathcal{N}(0, I)$, the posterior of (x_0, x_1, z) given $\mathbf{x}_\tau = x$ is an exact 8-component Gaussian mixture, and both $\operatorname{Var}(\boldsymbol{\mu}_\tau \mid \mathbf{x}_\tau = x)$ and $\operatorname{Var}(\mathbf{v}_\tau \mid \mathbf{x}_\tau = x)$ decompose into a scalar within-mode term plus a rank-8 mode-spread term whose operator norm is computed exactly from an 8×8 Gram matrix at any ambient dimension. We approximate the supremum over Ω by the maximum over 4096 samples from the time-τ marginal, for each (d, ε, τ) with $d \in \{16, 64, 256\}$, $\varepsilon \in \{0.1, 0.5, 1.0\}$, $\tau \in \{0.25, 0.5, 0.75\}$. Table 13 reports $\hat{C}_\mu = \max_x \|\operatorname{Var}(\boldsymbol{\mu}_\tau \mid x)\|_{\mathrm{op}}/h_\tau^2$ and $\hat{V}_0^2 = \max_x \|\operatorname{Var}(\mathbf{v}_\tau \mid x)\|_{\mathrm{op}}$.

The worst cases occur at $\tau = 0.25$, where mode responsibilities mix most strongly; at fixed (ε, τ) the measured constants do not grow from $d = 16$ to $d = 256$. This finite-sample check is consistent with the assumed scaling on the tested grid, but proves neither a uniform bound over x nor dimension-uniformity outside that grid.

P.4 Relation to the Empirical Proxy

The measured gradient proxy is approximately flat in the controlled GMM sweep at fixed ε; on its own it does not establish the pointwise posterior assumptions. The exact posterior computation above

Table 13: Measured Theorem C posterior constants on the coarse GMM family. Exact mixture-posterior operator norms, maximized over 4096 marginal samples. Worst case is over the tested (ε, τ) grid. The constants are approximately flat (mildly decreasing) in d, consistent with the dimension-uniformity assumed by Theorem C on this family.

	Worst case over (ε, τ) grid		$\varepsilon = 1, \tau = 0.5$	
d	$\hat{C}_\mu$	$\hat{V}_0^2$	$\hat{C}_\mu$	$\hat{V}_0^2$
16	6.60	22.76	1.83	12.47
64	6.50	22.39	1.74	11.68
256	6.06	21.87	1.63	10.63

(App. P.3) provides a finite-grid diagnostic for the GMM family, not a proof of uniform posterior bounds.

Q Scope and Boundary Cases of Model D

Model D is the two-term error model of Eq. (9). We discuss its failure modes.

Q.1 Multimodal Data with Sharp Modes

Well-separated modes can loosen the regularity bound because posterior moments vary across modal boundaries. The 8-mode and 16-mode GMM experiments retain the same directional decline in optimal noise, but the transition location should be estimated under the target mixture geometry rather than inferred from mode count alone.

Q.2 Heavy-Tailed Conditional Distributions

Heavy tails can invalidate the finite-moment assumptions of Theorem A and the posterior constants of Theorem C. Applying the phenomenological error model in this setting requires separate moment control; preprocessing alone changes neither theorem assumption.

Q.3 Anisotropic Data Geometry

For PCA-CIFAR the model and noise are explicitly restricted to a k-dimensional subspace, so Proposition B uses k. For raw ShapeNet and FashionMNIST the ambient bridge remains higher-dimensional; a PCA proxy is descriptive and cannot replace d in Proposition B without changing the modeled process.

Q.4 Budget Dependence of ε^*

The model contains no budget term, so its optimum is K-independent. In the fixed-protocol GMM sweep, selected noise decreases with K at $d = 64, 128$ and stays zero at $d = 256$. Smoothing that offsets coarse-integration error is a possible explanation, not an identified cause. Whitened PCA-CIFAR instead benefits from larger noise at high budgets (App. M). Model D should therefore be read as a fixed-budget heuristic, not a quantitative law for budget-dependent selection.

Q.5 Behavior at d^*

At $d = d^*$, $f'(0) = 0$ and $f''(\varepsilon) = 2A\beta_0^2/(1 + \beta_0\varepsilon)^3 > 0$, so $\varepsilon = 0$ is the strict global minimizer exactly at the boundary as well.

Q.6 Assumption Sensitivity

Table 14 separates theorem conditions from empirical transfer questions. Failure to verify a sufficient condition limits what the theorem establishes; it does not by itself prove that Model D's predictions fail.

Table 14: Scope of the analytical results and empirical selection rule. These are verification requirements, not measured robustness ratings. Multimodality, anisotropy, and budget changes do not automatically violate the stated theorem conditions.

Condition or setting	Scope and verification requirement
Posterior contraction	Theorem C requires the stated bound; a failed sufficient test is inconclusive.
Posterior velocity moment	A dimension-uniform bound is needed for the trace-free comparison.
Multimodal data	Unimodality is not required; verify the posterior conditions directly.
Anisotropic data	Proposition B uses modeled bridge dimension; a PCA proxy alone cannot replace it.
Sampling budget	Re-evaluate noise at the intended K; the GMM and PCA trends differ.
Heavy-tailed data	Check finite second moments and the posterior bounds separately.

R Training and Sampling Algorithms

Algorithm 1 ε-Bridge Training

Require: p_{data}, $\varepsilon \geq 0$, cutoff δ, lr η

for each step **do**
  $\mathbf{x}_0 \sim p_{\mathrm{data}}$, $\mathbf{x}_1, \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\tau \sim \mathcal{U}[\delta, 1-\delta]$
  $\mathbf{x}_\tau = (1-\tau)\mathbf{x}_0 + \tau \mathbf{x}_1 + \sqrt{\varepsilon \tau (1-\tau)}\, \mathbf{z}$
  $\mathbf{v}_\tau = (\mathbf{x}_1 - \mathbf{x}_0) + \frac{1-2\tau}{2\tau(1-\tau)} \sqrt{\varepsilon \tau (1-\tau)}\, \mathbf{z}$
  $\theta \leftarrow \theta - \eta \nabla_\theta \|\mathbf{v}_\theta(\mathbf{x}_\tau, \tau) - \mathbf{v}_\tau\|^2$
end for

Algorithm 2 ε-Bridge Sampling (Euler, K steps)

Require: trained $\mathbf{v}_\theta$, cutoff δ, step count K

$\mathbf{x}^{(0)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\Delta = (1-2\delta)/K$
for $k = 0, \ldots, K-1$ **do**
  $\tau_k = 1 - \delta - k\Delta$
  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \Delta\, \mathbf{v}_\theta(\mathbf{x}^{(k)}, \tau_k)$
end for
return $\mathbf{x}^{(K)}$

S Experimental Details: 2D

The 2D experiments use a 4-layer MLP with hidden width 256, SiLU activations, and AdaLN time conditioning with 32-dimensional sinusoidal features. Each model is trained for 20k steps with Adam ($\beta_1 = 0.9, \beta_2 = 0.999$), learning rate 3×10^{-4}, batch size 1024, and 1k warmup steps, over 3 seeds. Sample quality is measured by W_2 (POT library) against 4096 target samples.

T Experimental Details: GMM Dimension Scaling

The GMM targets are 8-component mixtures whose mode centers are generated from a fixed distribution seed, with component standard deviation 0.3; each condition has one target mixture shared by all model seeds. Every model is a six-layer width-512 MLP trained for 30k AdamW updates with batch size 512, learning rate 10^{-3}, and weight decay 10^{-4}, over 3 model seeds. Sample quality is measured by sliced-W_2 with 1,024 common random projections and 8,192 samples: at $K = 20$, ε^* is selected on validation data and the selected checkpoint is evaluated once on an independent test sample, and a separate inference sweep under the same recipe evaluates $K \in \{1, 2, 5, 10, 20, 50, 100\}$ at $d \in \{64, 128, 256\}$. The complete results of the fixed-protocol sweep and of the multi-budget sweep are reported in App. AK.

U Experimental Details: Tabular Data

California Housing, Breast Cancer, and Digits are standardized using all available records and used in full for training. The retained implementation has no train/test split. The ISOLET run in App. D uses the same procedure. Thus these experiments measure matching to an empirical training distribution, not held-out generalization. We generate $n_g = \min(5000, n)$ samples and compare them with the first n_g standardized records: 5000, 569, 1797, and 5000 for California, Breast Cancer, Digits, and ISOLET, respectively.

The model is a four-layer SiLU MLP, with hidden width 128 for $d < 30$ and 256 otherwise. Each configuration uses 30k Adam updates, learning rate 10^{-3} with cosine decay, batch size 256, and three model seeds $(42, 123, 456)$. Training samples τ uniformly from $[0.05, 0.95]$; generation starts from a standard Gaussian at $\tau = 0.95$ and uses K reverse Euler steps of size $0.9/K$, ending at 0.05. The interpolant code adds 10^{-8} inside the bridge-variance square root for numerical stability, including at the nominal FM endpoint.

The metric is sliced Wasserstein-2 (SW_2), not full W_2: 200 unit-normalized Gaussian projection directions are fixed with NumPy seed 0, the sorted projected differences are squared and averaged over samples and directions, and the final average is square-rooted. We select the lowest seed-mean score at $K = 20$ on this same empirical reference. Retained JSON files contain aggregate means and standard deviations computed with divisor 3; individual seed scores are not retained. Table 15 summarizes the three recovered datasets, and Table 29 gives their complete grids. No matched result file was recovered for the 178-sample, 13-feature sklearn Wine sweep, so its earlier performance numbers are excluded from both the tables and the gate analysis; no success or failure is inferred from that missing result.

Table 15: Tabular SW_2 at $K = 20$, averaged over three seeds and evaluated against training-data references. Full-grid minima are descriptive; a nonzero argmin alone does not establish a useful gain.

Dataset	d	SW_2 ($\varepsilon = 0$)	SW_2 ($\varepsilon = \varepsilon^*$)	ε^*
California Housing	8	0.446	0.445	0.10
Breast Cancer	30	0.158	0.144	1.00
Digits	64	0.157	0.153	0.30

V Experimental Details: CIFAR-10

The primary CIFAR-10 comparison uses the 11.5M-parameter U-Net of App. Z. Table 19 evaluates EMA checkpoints saved after 200k updates of a 500k-step cosine schedule, with 10,000 generated samples per FID. It is separate from the v3 300k SNR-weighted ablation in Table 20; their checkpoints and losses are not interchangeable.

W Experimental Details: RSBM Navigation

RSBM (Luan et al., 2026a) uses an ε-rectified bridge for visual navigation. Its 16-dimensional action representation and selection $\varepsilon = 0.5$ are summarized here together with the corresponding navigation results.

For reference, Table 16 summarizes the published RSBM configuration.

For the published comparison against navigation baselines, Table 17 reports RSBM at $k = 3$ and $k = 10$ alongside CFM and NaviBridger.

X Protocol Separation and Evidence Scope

The primary GMM conclusions use the fixed-distribution protocol of App. AK: distribution construction is independent of model seeds, architecture and training length are fixed, and validation selection is separated from testing. Capacity interventions change the stated capacity or training setting at fixed

Table 16: RSBM hyperparameters, summarized from Luan et al. (2026a).

Component	**Value**
Action space	8-step waypoints, $d=16$ ($d_{\text{eff}}=16$)
Observation	96×96 RGB + language goal embedding
Prior network g_ψ	Variational prior, latent $z\sim\mathcal{N}(0,I)$ at inference
Bridge network f_θ	Conditional U-Net-1D, FiLM time conditioning
Prediction target	v-prediction (velocity)
ε (optimal)	0.5 (selected on Custom Indoor, fixed for all datasets)
Solver	Heun 2nd-order, NFE$=2k-1$
Steps k (default)	10 (NFE$=19$); few-step at $k=3$ (NFE$=5$)
Training steps	100k, Adam lr 10^{-4}, batch 256
Hardware	RTX 4090, 3 seeds
Datasets (train/test)	
Custom Indoor (Gazebo)	500 / 100 episodes
CitySim (outdoor)	400 / 80 episodes
HuRoN / Recon / SACSoN / SCAND / GoStanford	$\sim$60k trajectories total (public)

Table 17: RSBM and baselines on Custom Indoor (from Luan et al., 2026a, Table I). RSBM at $k=3$ uses 5 NFE versus NaviBridger's 19 NFE at $k=10$.

Method	k	NFE	MSE↓	CosSim↑	Suc.%↑	FDE↓
CFM	10	10	2.80	0.910	82	1.45
NaviBridger	10	19	1.82	0.942	88	0.82
NaviBridger	3	5	12.00	0.710	35	4.20
RSBM ($\varepsilon=0.5$)	10	19	1.72	0.949	93	0.75
RSBM ($\varepsilon=0.5$)	3	5	**1.90**	**0.945**	**92**	**0.80**

dimension. The endpoint control in App. AL reuses these checkpoints and changes only initialization and evaluation target. Neither diagnostic is pooled into the primary validation/test table.

CIFAR uses two separate 11.5M-parameter protocols: unweighted-MSE v2, evaluated at update 200k of a 500k cosine schedule, and SNR-weighted v3, evaluated after a complete 300k schedule. Each has its own FM comparator and sample counts (App. Z). Their absolute FID values are not treated as a single noise sweep. PCA results use the stated whitened representation; they are not pooled with unwhitened runs.

Historical GMM runs used dimension-dependent widths and training lengths, Adam, and 100 projections. Their training seeds changed mixture centers while evaluation used a fixed mixture. This target mismatch excludes those results from performance comparisons; they cannot serve as a valid second protocol-sensitivity test. The simultaneous changes also prevent attributing a shifted crossover to one factor. Comparisons lacking a traceable matched baseline are excluded; we claim no performance ranking against optimized DDBM or DSBM solvers.

The original sweep overview in Fig. 6 retains its graphical labels. In that figure, tabular/PCA "W_2" denotes SW_2, ShapeNet Chamfer values use units of 10^{-3}, and LIBERO-Goal's $+5\%$ denotes $+5$ percentage points. The PCA-512 "FM" label does not describe the current whitened $K = 20$ result ($\varepsilon^* = 0.1$, Table 2). SnapFlow has no zero-noise comparator in that panel. QM9 varies post-distillation reverse-step noise with DDIM at zero. The CIFAR panel is not the matched v2 comparison in Fig. 9 and is not used to infer a monotone noise response. These panels are retained as the original task-specific overview, not as additional evidence for the controlled GMM crossover.

Y LIBERO VLA ε-SWEEP

We run a four-suite, four-ε LIBERO (Liu et al., 2023a) VLA sweep without any consistency distillation. Each model is the ε-bridge SmolVLA (Shukor et al., 2025) backbone trained for 30k steps on LIBERO mixed action-chunks at $\varepsilon\in\{0.0, 0.1, 0.3, 0.5\}$ (final training loss converged to ≈0.025 across all four), and evaluated with an action-chunk length of 10, $K_{\text{ODE}}=10$ Euler steps per chunk, with $n_{\text{ep}}=10$ episodes per task on each of the 10 tasks per suite (100 episodes per cell).

Table 18: LIBERO success rate (%) by suite and ε, $n_{\text{ep}} = 100$ episodes per cell. Goal increases over the grid, Object saturates, and Spatial/Long-10 vary within three percentage points.

Suite	$\varepsilon=0.0$	$\varepsilon=0.1$	$\varepsilon=0.3$	$\varepsilon=0.5$	Verdict
LIBERO-Object	100.0	100.0	100.0	100.0	saturated
LIBERO-Spatial	97.0	96.0	96.0	96.0	flat (within noise)
LIBERO-Long-10	96.0	93.0	96.0	96.0	flat (within noise)
LIBERO-Goal	95.0	97.0	98.0	**100.0**	monotone $\varepsilon^*=0.5$ (+5pp)

Using the executed action chunk gives a descriptive dimension of 70. LIBERO-Goal improves from 95 to 100, whereas Object saturates and Spatial/Long-10 fluctuate within three percentage points. The four suites together show the range of behaviors across the LIBERO tasks.

Distilled one-step policies optimize a different objective from this ODE-step sweep and are not comparable through the ODE-step budget (App. M).

Z CIFAR-10 Details and FID-vs-NFE Curve

The primary v2 comparison uses an 11,507,523-parameter U-Net (base channels 64), batch size 256, AdamW with initial learning rate 3×10^{-4}, unweighted MSE, and uniform training time on $[0.05, 0.95]$. The evaluated EMA checkpoints were saved after 200k updates of a 500k-step cosine schedule (EMA decay 0.9999); there is one checkpoint per $\varepsilon \in \{0, 0.1\}$. FID uses `pytorch_fid` with 2048-dimensional Inception features, 10,000 generated images and the first 10,000 CIFAR-10 training images as reference. These are 10k/10k within-study comparisons, not 50k-sample benchmark FIDs.

Both CIFAR protocols use noise-to-data time $s = 1 - \tau$ and Euler updates of size $1/K$ over $[0, 1]$, with the network time clamped to $[0.05, 0.95]$. This differs from the GMM truncated-interval sampler with step size $0.9/K$. Outputs are rescaled to $[0, 1]$, clipped and saved as PNGs before FID. Table 19 favors $\varepsilon = 0.1$ at NFE$= 1$ and FM from NFE$= 2$ onward.

The separate v3 ablation uses the same architecture, batch size, initial learning rate, EMA, image sampler and FID setup, but a complete 300k cosine schedule and SNR-weighted MSE for positive ε (ordinary MSE at $\varepsilon = 0$). Uniform time sampling is retained. Its FM models have two seeds; the $\varepsilon = 0.05, 0.30$ models have three. Table 20 compares only these v3 models.

Table 19: CIFAR-10 v2 FID: 11.5M U-Net, step-200k EMA checkpoints, one checkpoint per noise level, 10k generated and 10k reference images. The $\varepsilon = 0.1$ checkpoint is slightly better at NFE$= 1$; FM is better from NFE$= 2$ onward.

ε	NFE$=1$	NFE$=2$	NFE$=3$	NFE$=5$	NFE$=10$	NFE$=20$	NFE$=50$	NFE$=100$	NFE$=200$
0.0 (FM)	379.3	**138.7**	**75.6**	**42.1**	**26.7**	**20.5**	**15.6**	**14.8**	**14.6**
0.1	**366.2**	248.2	180.7	115.6	70.7	48.0	42.6	41.5	40.9

Table 20: CIFAR-10 v3 ablation, separately trained for 300k updates. FID uses 10k generated and 10k reference images. Entries are means $\pm$ sample standard deviations: FM has $n = 2$ seeds; positive-noise models have $n = 3$. Each comparison uses the v3 loss and checkpoint protocol, not the v2 values in Table 19.

ε	NFE$=5$	NFE$=10$	NFE$=50$
0.00 (FM)	$\mathbf{45.9 \pm 0.4}$	$\mathbf{28.6 \pm 0.1}$	$\mathbf{16.1 \pm 0.1}$
0.05	79.4 ± 0.9	41.0 ± 1.1	31.2 ± 0.9
0.30	167.7 ± 0.8	119.6 ± 0.9	82.0 ± 0.1

At NFE$= 200$, FM gives 15.8 ± 0.1 ($n = 2$); positive-noise results are available for seed 0 only: 31.6 for $\varepsilon = 0.05$ and 79.1 for $\varepsilon = 0.30$. Figure 9 reports the separate matched-capacity comparison at $\varepsilon \in \{0, 0.1\}$.

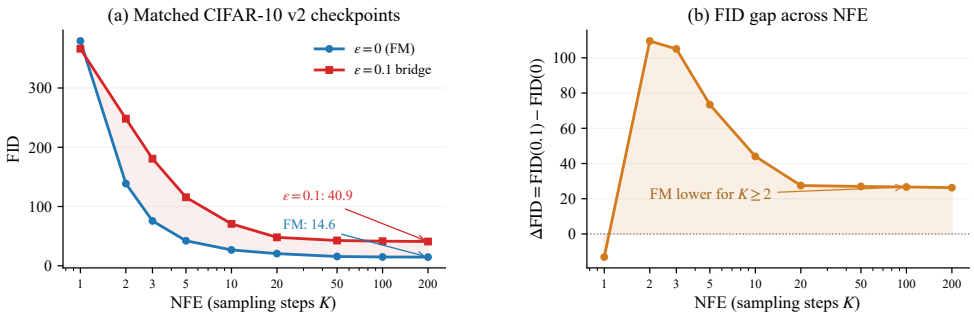


Figure 9: CIFAR-10 v2 (11.5M, step-200k EMA, 10k/10k FID). (a) FID vs NFE for $\varepsilon = 0$ and $\varepsilon = 0.1$; $\varepsilon = 0.1$ is slightly lower at NFE$=1$, while FM is lower from NFE$=2$ onward. (b) The FID gap $\Delta = \mathrm{FID}(0.1) - \mathrm{FID}(0)$ becomes positive after NFE$=1$ and reaches about 26 FID points at NFE$=200$.

AA STL10 ε-Sweep at $d = 3072$

To test the FM regime on a second, distributionally different natural-image dataset we ran a six-point ε sweep on STL10 (Coates et al., 2011). The unlabeled split is 96×96; we resize to 32×32 so that the same U-Net used for CIFAR-10 applies unchanged, giving $d = 3072$. STL10 therefore varies the data distribution at fixed ambient dimension rather than raising the dimension. Models: a 64-channel U-Net, AdamW with learning rate 2×10^{-4}, and batch size 128. Training length differs across cells (300k steps for $\varepsilon \in \{0, 0.1\}$ seed 0, 200k for $\varepsilon \in \{0.03, 0.3, 0.5, 1.0\}$ and for the second seed), which we note because the $\varepsilon = 0$ mean below averages a 300k and a 200k run. The boundary rows $\varepsilon = 0$ and $\varepsilon = 1.0$ are evaluated over 2 independent seeds; remaining rows are single-seed (compute-limited).

Table 21: STL10 FID by ε and NFE. The measured grid is monotone increasing in ε at every tested NFE.

ε	NFE$=5$	NFE$=10$	NFE$=50$	NFE$=100$
0.00 (FM)	**51.4**	**31.6**	**19.5**	**19.6**
0.03	63.0	37.0	32.8	33.5
0.10	104.5	68.0	49.7	49.7
0.30	147.6	105.3	75.4	73.8
0.50	176.2	126.5	90.0	86.8
1.00	220.0	163.0	115.0	110.3

At every tested NFE, the measured FID grid increases with ε. The endpoint gap is substantial across all four budgets, and the result supports FM on this protocol while leaving untested ε, budgets, and training schedules for future evaluation.

Both pixel-space datasets use 32×32 images and therefore test different data distributions at the same ambient dimension $d = 3072$. The dimension ladders are supplied by the GMM and PCA-CIFAR experiments; conclusions beyond the tested image resolution are outside the present empirical scope.

AB Published One-Step VLA Results

SnapFlow (Luan et al., 2026b) distills multi-step flow-matching VLA policies into a single forward pass with a consistency objective built on the $\pi_{0.5}$ backbone (Black et al., 2025a). We cite it as published context for few-step action generation; all numbers in this appendix are taken from Luan et al. (2026b), and no additional experiment on their models is reported in this paper. Distilled one-step budgets follow a different objective from the ODE-step budget K_{ODE} of the error model (App. M).

Table 22: Published one-step VLA comparison (from Luan et al., 2026b, Table I). SnapFlow at 1 NFE achieves 98.75% average success, compared with 97.75% for its 10-NFE teacher.

Method	NFE	L-spatial	L-object	L-goal	L-10	**Avg**
$\pi_{0.5}$ teacher	10	98	100	96	97	97.75
Naïve 1-step	1	96	99	98	94	96.75
SnapFlow	1	**99**	**100**	**99**	**97**	**98.75**

AC Experimental Details: QM9 ε-BMD

This appendix reports a related molecular-generation case study. We distill a pre-trained EDM (Hoogeboom et al., 2022) teacher from 1000 steps to students evaluated at 50 and 10 steps via iterative step-halving, then vary reverse-step noise strength ε; we call this configuration ε-BMD. Here $\varepsilon = 0$ is a DDIM sampling endpoint, not the FM training endpoint of Eq. (2). This case study tests sampling stochasticity after distillation rather than the interpolant parameter of Model D. The ambient dimensionality is $D = 3N + K_{\text{type}}N = 232$ for $N = 29$ and $K_{\text{type}} = 5$; the reported PCA-90% proxy is $d_{\text{eff}} \approx 40$. These dimensions are descriptive and are not used to validate d^*.

The full configuration used for QM9 molecular generation is listed in Table 23.

Table 23: ε-BMD hyperparameters.

Component	**Value**
Teacher model	EDM (Hoogeboom et al., 2022), 1000-step DDPM
Distillation	Iterative step-halving from the 1000-step teacher; students at 50 and 10 steps
Architecture	EGNN (Satorras et al., 2021) (equivariant GNN), same as teacher
Dataset	QM9 (133k molecules; max 9 heavy atoms, 29 including hydrogen)
D (ambient)	232 ($3N$ coords + $5N$ one-hot atom types, $N=29$); $d_{\text{eff}} \approx 40$ (PCA-90%)
Bridge noise	$\sqrt{\varepsilon} \cdot \sigma_{t\|s} \cdot \eta$ injected at each reverse step
ε^*	0.1 (all 3 reported metrics peak simultaneously)
Eval samples	5000 molecules, 3 seeds
Metrics	Atom Stability%, Molecular Stability%, Validity%
Training	Teacher init, lr 10^{-4}, AdamW, batch 64 per round
Hardware	RTX 4090, $\sim$8h per distillation round

The full ε-ablation at 50 distilled steps is reported in Table 24; all three molecular quality metrics peak at $\varepsilon^* = 0.1$.

Table 24: Full ε-ablation at 50 distilled steps. Mean$\pm$std over 3 seeds, 5000 sampled molecules per cell. All three metrics peak at $\varepsilon^* = 0.1$, so the optimum is not metric-specific.

ε	Atom Stab.%↑	Mol Stab.%↑	Validity%↑
0.0 (DDIM)	97.6±0.1	80.6±0.6	89.7±0.4
0.1	**99.6**±0.0	**94.1**±0.2	**98.3**±0.3
0.2	99.4±0.0	92.3±0.4	97.2±0.2
0.3	99.2±0.0	89.7±0.5	95.6±0.0
0.5	98.5±0.0	83.1±0.1	92.1±0.4
0.8	97.3±0.0	70.8±0.2	84.8±0.2
1.0 (DDPM)	96.5±0.0	63.0±0.2	80.3±0.3

AD Experimental Details: FashionMNIST, ShapeNet, PCA-CIFAR

AD.1 FashionMNIST

FashionMNIST provides 60k training and 10k test images of size 28×28 (grayscale), flattened to $d = 784$; the descriptive PCA-90% proxy is ≈ 80 and the PCA-99% value is 428 (Tab. 11). We train a SimpleUNet with 64 base channels for 30k steps with Adam (learning rate 3×10^{-4}, batch size 128)

over 10 seeds, and evaluate sliced-W_2 (512 projections, 8192 samples) with $K=50$ Euler steps. The sweep selects $\varepsilon^*=1.0$, with $SW_2=0.01347\pm0.00256$ against 0.01605 ± 0.00253 at FM (-16.0%); the full ε-sweep at $K=50$ is shown in Table 25.

Table 25: FashionMNIST ε-sweep at $K=50$ (sliced-W_2 $\downarrow$; mean and std over 10 seeds).

ε	0.0 (FM)	0.01	0.05	0.1	0.3	0.5	1.0
SW_2 mean	0.01605	0.01660	0.01645	0.01592	0.01540	0.01462	**0.01347**
SW_2 std	0.00253	0.00275	0.00266	0.00299	0.00277	0.00288	0.00256

AD.2 ShapeNet 3D Point Cloud

We use three ShapeNet-Core categories, Airplane ($\sim$4k shapes), Chair ($\sim$7k), and Car ($\sim$4k), each represented by 2048 points, so the raw dimension is $d=3\times2048=6144$ and the effective dimension is $d_{\text{eff}}\approx250$ (PCA capturing 90% of the variance of training-set point differences). The model is a PointNet-based denoising network with 8 layers, hidden width 512, and self-attention at resolutions 256 and 128, trained with $T=10{,}000$ continuous-time diffusion steps, AdamW (learning rate 10^{-4}, batch size 64), 100k iterations, and 2 seeds. Quality is measured by Chamfer distance (CD-L_2, $\times10^3$) on 2048-point clouds against the test set. The sweep selects $\varepsilon^*=1.0$ for Chair and Car and $\varepsilon^*=0.5$ for Airplane (where $\varepsilon=0.5$ and 1.0 differ by 0.6%), so all three categories sit in the SB regime; the CD improvement over FM at the per-category optimum is 43%, 50%, and 53%.

The full ε-sweep for the three categories at $T=10$k steps is given in Table 26. These are within-configuration comparisons; the PCA dimension is descriptive and does not replace the ambient bridge dimension in Proposition B or calibrate Model D.

Table 26: ShapeNet ε-sweep (Chamfer Dist.$\downarrow$, $\times10^{-3}$) at $T=10$k.

Category	$\varepsilon=0$	$\varepsilon=0.01$	$\varepsilon=0.05$	$\varepsilon=0.1$	$\varepsilon=0.3$	$\varepsilon=0.5$	$\varepsilon=1.0$
Airplane	24.86	24.35	21.90	20.46	17.01	**14.18**	14.27
Chair	20.77	20.04	18.30	16.98	13.50	10.95	**10.35**
Car	16.07	16.19	15.81	15.14	11.99	10.02	**7.54**

AD.3 PCA-CIFAR Experiments

The first $20{,}000$ CIFAR-10 training images ($32\times32\times3$) are standardized per pixel and projected to $d\in\{32,64,128,256,512\}$ principal components; the PCA scores are then standardized per component, so the model and the noise prior both live in an isotropically scaled d-dimensional subspace. Without this per-component standardization the leading components dominate the sliced-W_2 metric and the sweep selects $\varepsilon^*=0$ at both tested dimensions ($d=32,64$; sliced-W_2 2.38 and 1.68 at $K=20$), so the ladder below is a statement about the whitened subspace.

This is a task-specific dimensionality study. Its selected ε values provide cross-domain evidence and are not additional estimates of the fixed-protocol GMM operating boundary.

The network is a SiLU MLP with hidden width 256 and 4 layers for $d\le64$, width 512 and 5 layers for $64<d\le256$, and width 512 and 6 layers for $d>256$. Training uses Adam with learning rate 10^{-3} and cosine decay, batch size 512, $\min(\max(200d,\ 2\times10^4),\ 8\times10^4)$ steps, $\tau\in[0.05,0.95]$, and 3 seeds. Quality is measured by sliced-W_2 (200 random projections) between $5{,}000$ generated samples and $5{,}000$ training projections at $K\in\{1,5,10,20,50,100\}$ Euler steps, and ε^* is the grid argmin of the seed mean at $K=20$.

The selected ε^* by dimension at $K=20$ is summarized in Table 27; it shifts toward the FM endpoint from $d=128$ ($\varepsilon^*=0.5$) to $d\ge256$ ($\varepsilon^*=0.1$).

Table 27: PCA-CIFAR ε^* summary by dimension (at $K=20$). The selected value moves toward the FM endpoint as the modeled dimension increases; percentages are computed from unrounded values.

d	32	64	128	256	512
Variance explained	79.40%	86.18%	91.72%	95.82%	98.42%
ε^* ($K=20$)	1.0	0.3	0.5	0.1	0.1
SW_2 improvement vs FM	-23.0%	-26.7%	-4.9%	-4.8%	-7.1%

AE Broader Impact and Limitations

AE.1 Broader Impact

The ε-bridge framework provides an empirical criterion for choosing between Flow Matching and Brownian-bridge-style stochastic interpolants in continuous generative models. Positive applications include acceleration of VLA robot control, faster molecular drug design, and more efficient tabular data modeling. We do not foresee direct negative societal impacts specific to this theoretical contribution. The framework is agnostic to the underlying data modality and does not introduce new capabilities beyond those of existing FM/SB methods.

AE.2 Limitations

Our analysis has several limitations. First, Proposition B uses the ambient modeled dimension; the PCA and action-chunk proxies used for heterogeneous raw-domain cells are descriptive and can change with the variance threshold. Second, Theorem C requires dimension-uniform posterior moment and contraction bounds. Sampled exact posterior computations support approximately flat norms on a finite GMM grid (App. P.3), but verification for general data distributions remains open, and the reported PCA scales do not provide it. Third, Model D is a fixed-budget heuristic: it neither predicts the GMM decrease of ε^* with K nor the different PCA trend. Smoothing is a possible explanation, not an isolated cause (App. M). Fourth, the raw target admits an exact conditional-mean alternative with the same marginal velocity (App. AM); its training and sample-quality effects are untested. Finally, we train one model per ε value; conditioning a single model on ε could enable adaptive inference-time selection.

The precision of the operating boundary is limited. The primary $K = 20$ grid brackets the canonical transition by $(160, 192]$. The independent 16-mode family selects 0.01 at $d = 192$, operationally near FM under the stated 0.05 threshold; family C has a different bracket, $(128, 176]$. The illustrative Model D rule has an exact-zero threshold near 364 and a 0.05-regime threshold near 285; neither is fitted to these GMM results. Its transfer discrepancies do not establish failure within the earlier protocol. Capacity interventions support protocol dependence, but do not isolate which historical protocol changes moved the crossover.

The gate itself has a limited scope. Its 0.09 threshold on FM/s_{ref} is exploratory and uses the declared reference scales of App. E; it is not derived or externally calibrated. The 34 cells share families and dimension ladders, and PCA/tabular scores use training-data references. They are not 34 independent external validation tasks. At $K \geq 50$, whitened PCA-CIFAR benefits from larger noise, and a severely reduced PCA-CIFAR network can retain a low SW_2 baseline while gaining nothing from noise (I6). Thus baseline fit, capacity, and noise benefit are not interchangeable. The above-threshold helping cell has 569 records; this observation does not establish a sample-size cutoff for using the diagnostic. Wine performance is excluded because its matching result file was not recovered, rather than assigned an empirical gate label. We have not calibrated the diagnostic against data-to-data metric floors or additional distributional metrics.

Finally, the experiments establish comparisons within the independent-endpoint interpolant family, not against optimized DDBM or DSBM algorithms. The endpoint diagnostic is also limited to three dimensions and existing noise grids: exact initialization requires data samples, and its exact terminal reference depends on ε. It measures truncated-path transport, not deployable generation against a common target. The original calibration procedure for the fixed diagnostic constants has not been recovered, so they are presented only as illustrative parameters.

AF Extended Theoretical Analysis

AF.1 Optimization Domain for ε

The stationary-point and endpoint analysis of the error model is carried out in App. L. We restrict ε to $[0, 1]$ because this interval connects the chosen zero-noise and unit Brownian-bridge endpoints. For $d\hat{c}_0 > 0$, the linear penalty makes the model diverge as $\varepsilon \to \infty$; the chosen interval is a modeling domain, not a finite-limit argument.

AF.2 Relation to Schrödinger Bridge Optimal Transport

The $\varepsilon = 1$ member uses independent endpoints and a single Gaussian latent $\mathbf{z}$ across time. Its conditional marginal at each fixed τ is $\mathcal{N}(\boldsymbol{\mu}_\tau, \tau(1-\tau)\mathbf{I}_d)$, matching a unit-diffusivity Brownian bridge. Its conditional path law is different: Def. 1 specifies a fixed-latent interpolant. An entropy-regularized optimal transport (EOT) Schrödinger bridge with the same Brownian reference has a Brownian-bridge conditional path law and an entropic-OT endpoint coupling. Table 28 separates these properties.

Table 28: ε-bridge ($\varepsilon = 1$) vs. entropy-regularized OT Schrödinger bridge.

	ε-bridge ($\varepsilon = 1$)	EOT-SB
Endpoint coupling	Independent $(\mathbf{x}_0, \mathbf{x}_1)$	Entropic-OT plan
Single-time conditional marginal	Brownian-bridge kernel	Same kernel
Conditional path law	Fixed-latent interpolant	Brownian bridge
Training	Conditional FM regression	IPF/IMF or score-based
Endpoint marginals	Matched by construction	Matched at the fixed point

The raw target-correction energy $\varepsilon d\, c_0(\delta)$ concerns the regression target of the fixed-latent construction; it is not a path-space cost or sampling-error bound for EOT-SB. The $\varepsilon = 1$ member remains trainable by conditional FM regression, with independent endpoints and a Brownian-bridge kernel at each fixed time.

AG Full ε^* Tables for All Experiments

AG.1 Additional GMM Results

The complete fixed-protocol dimension and budget sweeps are in App. AK. Capacity interventions are reported separately in App. E, and endpoint sensitivity in App. AL; these do not replace the primary validation/test results.

AG.2 Tabular Datasets: Full Results at All ε

Table 29 gives the retained tabular SW_2 grids at $K = 20$. Each entry is a mean and the saved standard deviation over three seeds (divisor 3), against the training-data reference of App. U. California and Digits have only small best-grid gains; these nominal positive-noise argmins do not establish reliable interior optima.

Table 29: Retained tabular SW_2 grids ($K = 20$, three seeds, mean$\pm$standard deviation with divisor 3). The reference records are part of the training data; the selected minima are not held-out test estimates.

Dataset	d	$\varepsilon = 0$	$\varepsilon = 0.01$	$\varepsilon = 0.03$	$\varepsilon = 0.1$	$\varepsilon = 0.3$	$\varepsilon = 0.5$	$\varepsilon = 1.0$
California	8	$0.446 \pm .014$	$0.446 \pm .013$	$0.446 \pm .012$	$\mathbf{0.445 \pm .015}$	$0.449 \pm .011$	$0.452 \pm .009$	$0.449 \pm .007$
Breast Ca.	30	$0.158 \pm .015$	$0.155 \pm .015$	$0.154 \pm .014$	$0.154 \pm .017$	$0.149 \pm .014$	$0.144 \pm .012$	$\mathbf{0.144 \pm .014}$
Digits	64	$0.157 \pm .012$	$0.158 \pm .012$	$0.156 \pm .011$	$0.156 \pm .009$	$\mathbf{0.153 \pm .005}$	$0.155 \pm .007$	$0.164 \pm .017$

AH NOISE-GRID SENSITIVITY

We recompute the grid argmin for subsets of the retained California Housing $K = 20$ scores in Table 29. This is a post-hoc subgrid analysis, with no retraining or new evaluation. The seven-point and four-point subsets select $\varepsilon = 0.1$; the three-point subset $\{0, 0.5, 1\}$ selects FM. The latter follows from scores 0.4460, 0.4524, and 0.4486, respectively. The full-grid improvement over FM is only 0.25%, and its mean gap is much smaller than the retained seed standard deviations. This example shows how a nominal argmin changes with the grid, not robustness of a resolved positive-noise optimum.

Table 30: Post-hoc subgrid selection from the same California SW_2 sweep at $K = 20$. Values are recomputed from retained aggregate scores, not independent grid experiments.

Score subset	Selected ε	Mean SW_2
$\{0, 0.01, 0.03, 0.1, 0.3, 0.5, 1\}$	0.1	0.4449
$\{0, 0.1, 0.3, 1\}$	0.1	0.4449
$\{0, 0.5, 1\}$	0	0.4460

Capacity and training interventions with retained configurations and results are reported in App. E. Auxiliary architecture, training-duration, solver, and batch-size/learning-rate tables from earlier drafts are excluded because matching run outputs could not be recovered; they supply no robustness claim here.

AI PER-SEED RESULTS

To show the seed-level variation behind the reported means, Table 31 lists the three seed-level validation values for GMM $d=64$ at $K=20$ under the fixed-protocol sweep of App. AK. With $n = 3$ and grid-based model selection, we report raw variation rather than a significance claim.

Table 31: Per-seed validation sliced-W_2 for GMM $d=64$, $K=20$ (fixed-protocol sweep, App. AK). The selected $\varepsilon^* = 0.3$ is evaluated once on the independent test sample per seed.

Seed	Val. SW_2 (FM)	Val. SW_2 ($\varepsilon^*=0.3$)	Test SW_2 (ε^*)
42	0.0312	0.0190	0.0193
123	0.0316	0.0198	0.0196
456	0.0317	0.0192	0.0194
Mean	0.0315 ± 0.0003	0.0193 ± 0.0004	0.0194 ± 0.0002

AJ VISUALIZATIONS AND QUALITATIVE RESULTS

AJ.1 2D INTERPOLANT GEOMETRY

This appendix collects qualitative observations. Figure 1(a–c) shows fixed-latent arcs on the 8-Gaussian distribution. The $\varepsilon = 0$ paths are straight, $\varepsilon = 0.5$ introduces mild curvature, and $\varepsilon = 1$ produces stronger bridge perturbations.

AJ.2 VELOCITY FIELD SMOOTHNESS

At $\tau = 0.5$, the maximum gradient-norm proxy is 4.82, 3.91, 2.47, and 1.83 for $\varepsilon \in \{0, 0.1, 0.5, 1\}$, respectively. This is an empirical diagnostic, not a test of Theorem C's uniform posterior assumptions; App. P.3 provides a separate finite-grid posterior check.

AK FIXED-PROTOCOL GMM SWEEP

This appendix reports the fixed-protocol GMM sweep behind the main-text crossover. Each condition uses one target GMM shared by all three model seeds. Every model has six layers of width 512 and

is trained for 30k AdamW updates. At $K = 20$, ε is selected with 8,192 validation samples and 1,024 common random projections. The selected checkpoint is then evaluated on an independent test sample. The initial grid is $\{0, 0.01, 0.03, 0.1, 0.3\}$, extended to $\{0.5, 1\}$ when 0.3 is selected.

Table 32: Fixed-protocol GMM sweep: validation-selected results. The first two score columns support validation gains. The final column evaluates the selected checkpoint on independent test samples; it does not report a paired test gain over FM.

Family	d	Rank	Val. FM	Val. selected	ε^*	Selected test SW_2
8-mode GMM	64	64	0.03151	**0.01933** (-38.6%)	0.3	0.0194 ± 0.0002
	128	128	0.02988	**0.01925** (-35.6%)	0.3	0.0191 ± 0.0009
	160	160	0.03132	**0.02521** (-19.5%)	0.1	0.0270 ± 0.0011
	192	192	**0.08342**	0.08342 ($\pm 0\%$)	0	0.0836 ± 0.0054
	224	224	**0.17393**	0.17393 ($\pm 0\%$)	0	0.1710 ± 0.0026
	256	256	**0.24611**	0.24611 ($\pm 0\%$)	0	0.2436 ± 0.0018
	300	300	**0.31424**	0.31424 ($\pm 0\%$)	0	0.3146 ± 0.0016
	350	350	**0.37446**	0.37446 ($\pm 0\%$)	0	0.3720 ± 0.0017
	400	400	**0.41725**	0.41725 ($\pm 0\%$)	0	0.4152 ± 0.0004
	450	450	**0.45035**	0.45035 ($\pm 0\%$)	0	0.4488 ± 0.0013
Held-out 16-GMM	160	160	0.03682	**0.02890** (-21.5%)	0.3	0.0294 ± 0.0004
	192	192	0.04674	**0.04646** (-0.606%)	0.01	0.0476 ± 0.0044
	350	350	**0.28288**	0.28288 ($\pm 0\%$)	0	0.2839 ± 0.0019
	400	400	**0.32328**	0.32328 ($\pm 0\%$)	0	0.3231 ± 0.0006
Fixed $d = 400$	400	100	0.56347	0.56340 (-0.0125%)	0.01	0.5615 ± 0.0020
	400	200	0.51687	0.51680 (-0.0151%)	0.01	0.5165 ± 0.0005
	400	400	**0.41671**	0.41671 ($\pm 0\%$)	0	0.4160 ± 0.0017

Table 32 and Figure 4 give the results. The canonical sweep selects $\varepsilon = 0.3$ at $d = 64, 128$, $\varepsilon = 0.1$ at $d = 160$, and FM from $d = 192$ onward. The independent 16-mode family selects 0.3 at $d = 160$ and 0.01 at $d = 192$, placing it on the FM side of the stated 0.05 regime threshold in $(160, 192]$, without claiming equality of the grid optima. At fixed ambient $d = 400$, reducing active rank to 100 or 200 changes the selected grid point only from 0 to 0.01, with validation improvements below 0.02%.

The companion budget sweep uses the same fixed-target training protocol and three seeds at $d \in \{64, 128, 256\}$. Every trained noise level is evaluated at seven budgets against an independently drawn sample; Table 33 reports argmins of mean SW_2. Noise decreases with K at $d = 64, 128$ and remains zero at $d = 256$. The $K = 20$ selections agree with Table 32. These descriptive minima do not extend the budget trend to unmeasured dimensions.

Table 33: Fixed-protocol $\varepsilon^*(d, K)$ matrix (argmin of mean sliced-W_2 over three seeds).

d	$K=1$	$K=2$	$K=5$	$K=10$	$K=20$	$K=50$	$K=100$
64	1.0	0.5	0.5	0.5	0.3	0.1	0.1
128	0.5	0.5	0.3	0.3	0.3	0.3	0.3
256	0.0	0.0	0.0	0.0	0.0	0.0	0.0

AL Endpoint-Marginal Sensitivity with Fixed Checkpoints

The primary GMM sampler integrates from 0.95 to 0.05 but initializes from $N = \mathcal{N}(0, I)$ and evaluates against data. These approximate the exact marginals $p^{\varepsilon}_{0.95}$ and $p^{\varepsilon}_{0.05}$. The terminal bridge-noise trace $0.0475\varepsilon d$ can affect this comparison, but is not itself a Wasserstein error.

We cross the two initializations with the two reference targets. Exact marginals use

$$x_\tau = (1-\tau)x_0 + \tau x_1 + \sqrt{\varepsilon\tau(1-\tau)}z, \quad x_0 \sim p_{\mathrm{data}}, \quad x_1, z \overset{\mathrm{ind}}{\sim} \mathcal{N}(0, I),$$

with the fixed training mixture. Each initialization is integrated separately with $K = 20$ reverse Euler steps of size $0.9/20$. Initial and reference draws are independent; 8,192 samples and 1,024

unit projections are used. Base draws and projections are paired across noise levels and arms. All evaluation seeds differ from the original training, validation and test seeds.

We reuse the final 30k-step checkpoints, three seeds per setting, without retraining. After an initial $\{0, 0.1, 0.3\}$ diagnostic, we expanded to all existing checkpoints on the stated grids: $\{0, 0.01, 0.03, 0.1, 0.3, 0.5, 1\}$ at $d = 64$ and $\{0, 0.01, 0.03, 0.1, 0.3\}$ at $d = 192, 256$. These 51 checkpoints yield 204 scores. This post-hoc sensitivity check does not replace primary validation/test selection.

Table 34: Endpoint sensitivity under fixed GMM training. Mean ± sample standard deviation of SW_2 over three runs. The runs differ in model seed and initialization draws; reference draws and projections are common. Exact-reference columns measure transport to an ε-dependent terminal marginal, not quality against a common data target.

d	ε	$N \to p_{\text{data}}$	$N \to p^{\varepsilon}_{0.05}$	$p^{\varepsilon}_{0.95} \to p_{\text{data}}$	$p^{\varepsilon}_{0.95} \to p^{\varepsilon}_{0.05}$
64	0	0.03174 ± 0.00026	0.01453 ± 0.00039	0.04214 ± 0.00030	0.02295 ± 0.00050
64	0.01	0.03101 ± 0.00044	0.01430 ± 0.00046	0.04133 ± 0.00017	0.02267 ± 0.00036
64	0.03	0.02954 ± 0.00032	0.01402 ± 0.00040	0.03972 ± 0.00021	0.02211 ± 0.00037
64	0.1	0.02524 ± 0.00047	0.01352 ± 0.00088	0.03447 ± 0.00037	0.02041 ± 0.00060
64	0.3	0.01995 ± 0.00060	0.01384 ± 0.00105	0.02322 ± 0.00084	0.01718 ± 0.00120
64	0.5	0.02557 ± 0.00034	0.01550 ± 0.00075	0.02017 ± 0.00061	0.01480 ± 0.00109
64	1	0.05381 ± 0.00042	0.02090 ± 0.00013	0.04534 ± 0.00039	0.01546 ± 0.00054
192	0	0.08295 ± 0.00449	0.09562 ± 0.00465	0.06312 ± 0.00396	0.07498 ± 0.00425
192	0.01	0.08636 ± 0.00287	0.09849 ± 0.00289	0.06618 ± 0.00269	0.07761 ± 0.00274
192	0.03	0.09272 ± 0.00256	0.10366 ± 0.00258	0.07218 ± 0.00243	0.08257 ± 0.00248
192	0.1	0.10718 ± 0.00225	0.11377 ± 0.00224	0.08602 ± 0.00219	0.09235 ± 0.00219
192	0.3	0.13682 ± 0.00313	0.13071 ± 0.00312	0.11566 ± 0.00304	0.10957 ± 0.00303
256	0	0.24430 ± 0.00172	0.25752 ± 0.00173	0.21483 ± 0.00164	0.22802 ± 0.00165
256	0.01	0.24682 ± 0.00100	0.25931 ± 0.00100	0.21735 ± 0.00101	0.22981 ± 0.00101
256	0.03	0.24953 ± 0.00112	0.26059 ± 0.00112	0.22019 ± 0.00105	0.23123 ± 0.00105
256	0.1	0.25722 ± 0.00200	0.26333 ± 0.00200	0.22844 ± 0.00189	0.23454 ± 0.00189
256	0.3	0.28010 ± 0.00148	0.27249 ± 0.00148	0.25304 ± 0.00120	0.24544 ± 0.00121

At $d = 64$, the four columns select $\varepsilon = 0.3, 0.1, 0.5, 0.5$ by their lowest grid mean. Standard-protocol FM/best scores are $0.03174/0.01995$; doubly matched scores are $0.02295/0.01480$. At $d = 192, 256$, every arm has its lowest mean at $\varepsilon = 0$. At $d = 192$, however, $\varepsilon = 0.01$ improves on FM in one of three seeds; no significance or unique-optimum claim follows.

Endpoint handling changes errors, gains, and the positive-noise optimum, but does not fully explain the ordering at these three dimensions. This control omits $d = 160$ and untested noise levels, so it does not re-estimate the crossover. Exact initialization requires data; exact terminal references vary with ε. The matched-endpoint scores therefore measure truncated-path transport, not deployable generation against a common target.

AM Conditioning the Gaussian Regression Target

Proposition B describes the raw target in Eq. (2), not an unavoidable sampling-error cost. Under Def. 1, fix $\tau \in (0, 1)$ and write

$$a_\tau = \sqrt{\tau^2 + \sigma_\tau^2}, \qquad \boldsymbol{\eta} = \frac{\tau \mathbf{x}_1 + \sigma_\tau \mathbf{z}}{a_\tau}, \qquad \boldsymbol{\zeta} = \frac{-\sigma_\tau \mathbf{x}_1 + \tau \mathbf{z}}{a_\tau}. \tag{34}$$

At each fixed τ, this orthogonal rotation makes $\boldsymbol{\eta}, \boldsymbol{\zeta}$ independent standard Gaussians, independent of $\mathbf{x}_0$. Separately, draw a time-independent standard Gaussian $\widetilde{\boldsymbol{\eta}}$ independent of $\mathbf{x}_0$. The alternative interpolant $\mathbf{y}_s = (1 - s)\mathbf{x}_0 + a_s \widetilde{\boldsymbol{\eta}}$ has the same single-time marginals as $\mathbf{x}_s$. Its state–velocity pair at $s = \tau$ has the same law as $(\mathbf{x}_\tau, \mathbf{w}_\tau)$, where $\mathbf{w}_\tau = -\mathbf{x}_0 + \dot{a}_\tau \boldsymbol{\eta}$. Substitution into the original target yields

$$\mathbf{v}_\tau = \mathbf{w}_\tau + r_\tau \boldsymbol{\zeta}, \qquad r_\tau = \frac{\tau \dot{\sigma}_\tau - \sigma_\tau}{a_\tau}, \qquad r_\tau^2 = \frac{\varepsilon}{4(1 - \tau)[\tau + \varepsilon(1 - \tau)]}. \tag{35}$$

Since $\boldsymbol{\eta} = (\mathbf{x}_\tau - (1 - \tau)\mathbf{x}_0)/a_\tau$, independence gives $\mathbf{w}_\tau = \mathbb{E}[\mathbf{v}_\tau \mid \mathbf{x}_0, \mathbf{x}_\tau, \tau]$ and hence $\mathbb{E}[\mathbf{w}_\tau \mid \mathbf{x}_\tau, \tau] = \mathbf{u}_\tau(\mathbf{x}_\tau)$. Thus this conditional-mean target preserves the marginal velocity, although the two fixed-latent constructions need not share a joint path law.

For the same $\tau \sim \mathcal{U}[\delta, 1-\delta]$, $0 < \delta < 1/2$, assume $\mathbb{E}\|\mathbf{x}_0\|^2 < \infty$ and finite predictor second moment. Orthogonality of the residual gives

$$\mathcal{L}_\varepsilon(\theta) = \mathbb{E}\|\mathbf{v}_\theta(\mathbf{x}_\tau, \tau) - \mathbf{w}_\tau\|^2 + d\,\mathbb{E}_\tau[r_\tau^2]. \tag{36}$$

The time expectation is normalized by $1-2\delta$, unlike the unnormalized integral defining $c_0(\delta)$. At $\varepsilon = 0$, $a_\tau = \tau$ and $r_\tau = 0$. The removable residual differs from the correction $\dot{\sigma}_\tau \mathbf{z}$ in Proposition B: at $\tau = 1/2$ that correction vanishes, whereas $r_\tau^2 = \varepsilon/(1+\varepsilon)$.

This identity separates properties of the marginal path from those of the regression target. It does not establish improved trained sample quality or removal of the empirical crossover; we do not evaluate this alternative target. Nor does either target-energy calculation imply Model D's SW_2 penalty: isotropic residual energy along any unit projection is r_τ^2, without the ambient trace factor d. Relating regression noise to the downstream metric requires additional analysis or controlled experiments.