Bridge Graphical Models: Coupling, Projection, and Current-Preserving Dynamics for Generative Modeling

Tiantian Zhang
Columbia University
New York, NY, USA
t.zhang8@columbia.edu

Abstract

Continuous-time generative models commonly train on endpoint-conditioned paths but deploy non-anticipative samplers with restricted information. We study this *bridge-to-sampler compression* through Bridge Graphical Models (BGMs), which separate endpoint coupling and bridge law from the local training target, sampler-visible information, time weighting, and dynamics representation. For a full bridge record $L \sim \mathscr{Q}$, target $U_t = u_t(L)$, visible state $S_t = [\mathcal{O}_U(L)]_t$, and training-time law μ, we define the target- and observation-relative Markovization gap

$$\mathfrak{G}_\mu(\mathscr{Q}, U, \mathcal{O}_U) = \int \mathbb{E}_{\mathscr{Q}}[\mathrm{Tr}\,\mathrm{Var}(U_t \mid S_t, t)]\,\mu(\,\mathrm{d}t).$$

The underlying L^2 projection identity is classical; our contribution is to identify the generative-model compression problem to which it applies and to expose the design variables that control the irreducible decoder floor. In an exactly solvable three-dimensional Gaussian–OT study, coupling, bridge target, and visible information independently alter the gap, and a trained MLP reaches the analytic floor across 11 configurations (Spearman and Pearson correlations 1.0000). A cross-fitted feature-space proxy is rank-calibrated on a higher-dimensional Gaussian family, perfectly ranks a six-design Fashion-MNIST coupling continuum, and correlates strongly across a broader fixed-cost CIFAR-10 bridge family (Spearman 0.893). A CIFAR-100 pilot shows the boundary of the claim: lower proxy predicts lower training loss but not necessarily better FID/KID/IS. The resulting diagnostic is therefore a computationally inexpensive predictor of absolute fixed-budget Markov-regression difficulty, not a universal sample-quality score.

1 Introduction

The immediate technical lineage of this work is flow matching and rectified flow. These methods first choose how base and data endpoints are coupled, then choose conditional paths between paired endpoints, and finally regress a Markov vector field from the resulting local targets [20, 21, 33]. Stochastic interpolants, diffusion flow matching, Schrödinger bridge matching, and generator matching extend this pattern to stochastic paths and broader Markov generators [2, 3, 8, 11, 27]. The conditional path used for training, however, may know endpoint or latent-path information that the deployed sampler cannot observe.

This paper isolates that mismatch. Let L denote the full bridge record: endpoints, bridge noise, latent coordinates, augmented field variables, and any relevant history. A training objective extracts a local target $U_t = u_t(L)$. A sampler receives only an information state $S_t = [\mathcal{O}_U(L)]_t$ and must predict U_t non-anticipatively. If distinct bridge records collapse to the same S_t while requiring incompatible

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

targets, no increase in decoder capacity can recover the discarded information. This is a property of the bridge design *relative to a target, observation interface, and clock*; it is not solely a property of the path law.

Bridge Graphical Models (BGMs) make this interface explicit. A BGM begins with an endpoint coupling and conditional bridge kernel, which together define a reference path law. It then specifies (i) the local target to be learned, (ii) the information exposed to the sampler, (iii) the time-sampling law used by the objective, and (iv) a dynamics representation for realizing the resulting projected current. The broader framework admits a path-space variational interpretation and field-based bridge kernels, but the present paper studies one universal edge in this graph: the compression of endpoint-conditioned bridge targets into causal sampler representations.

Our central quantity is

$$\mathfrak{G}_\mu(\mathscr{Q}, U, \mathcal{O}_U) = \int \mathbb{E}_{\mathscr{Q}}[\operatorname{Tr} \operatorname{Var}(U_t \mid S_t, t)]\, \mu(\mathrm{d}t), \qquad S_t = [\mathcal{O}_U(L)]_t. \tag{1}$$

For squared loss, this is the Bayes risk of predicting U_t from (S_t, t). We explicitly acknowledge that the conditional-expectation projection identity is classical. The substantive contribution is to identify the correct bridge-to-sampler prediction problem, make its design arguments explicit, derive consequences for coupling, bridge and information-state design, and validate the quantity against exact and learned decoder floors.

Contributions.

1. We introduce a target-, observation-, and clock-relative BGM formalism. It resolves three ambiguities in a bridge-only definition: stochastic models may use different local targets; stateful samplers may expose memory or latent variables; and time reparameterization changes the objective.
2. We define the Markovization gap in eq. (1), establish the projection decomposition and information monotonicity, and specialize a standard stability argument to separate irreducible bridge-design error from decoder approximation error.
3. We provide exact three-dimensional Gaussian–OT controls across five couplings, two local targets, and three observation maps. A neural decoder reaches the analytic floor across 11 trained configurations.
4. We calibrate a cross-fitted k-NN proxy against exact Gaussian gaps, then test a controlled Fashion-MNIST coupling continuum and a broader CIFAR-10 bridge family. We report endpoint cost, target variance, estimator sensitivity, complexity, and negative controls rather than treating the proxy as a certified population estimate.

Scope. The paper does not claim that the gap explains all generative performance. Optimization, finite-network approximation, discretization, score estimation, and sampler stability remain separate. Nor do we claim that every BGM instance has a finite exact path-space ELBO. Those broader representations are organized in the appendices; the core theory and experiments concern bridge-to-Markov compression.

2 Bridge graphical models as a compression graph

Let π_0 be a base distribution, $\pi_1 = \pi_{\mathrm{data}}$ the data distribution, and $\mathcal{M}$ a state space. A path lives in $\mathcal{X} = C([0,1], \mathcal{M})$ when continuous paths are appropriate.

Definition 1 (Endpoint coupling and bridge kernel)**.** *An endpoint coupling is* $\Gamma(\mathrm{d}z, \mathrm{d}x) \in \Pi(\pi_0, \pi_1)$. *For each* (z, x)*, a bridge kernel* $\mathsf{R}_\lambda^{z,x}(\mathrm{d}L)$ *is a probability law over a full bridge record* L *whose endpoint coordinates are* (z, x)*. The aggregated reference law is*

$$\mathscr{Q}_{\Gamma,\lambda}(\mathrm{d}L) = \int \mathsf{R}_\lambda^{z,x}(\mathrm{d}L)\, \Gamma(\mathrm{d}z, \mathrm{d}x). \tag{2}$$

The record L is deliberately more general than $X_{0:1}$. It may include stochastic bridge noise, a branch identity in an augmented field, or history used by a stateful sampler.

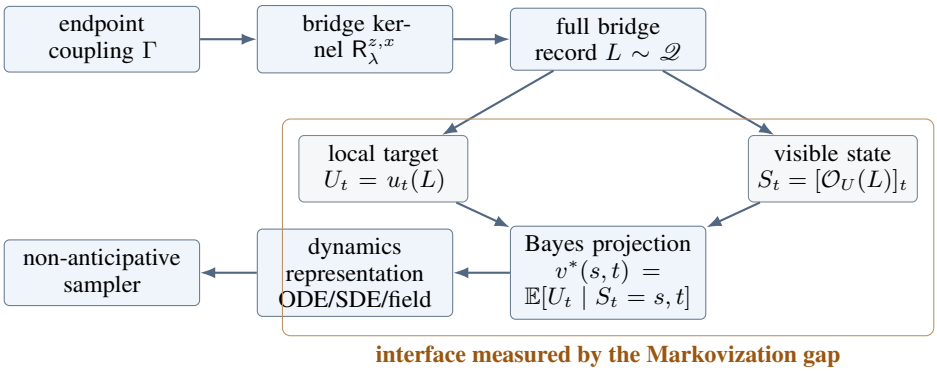

Figure 1: BGM as a compression graph. Coupling and bridge design create a full latent record. The current paper measures the highlighted target-to-visible-state projection. Path-space variational objectives and dynamics representations concern other parts of the graph and are treated separately.

Definition 2 (BGM compression specification)**.** *A compression specification is $(\mathscr{Q}, U, \mathcal{O}_U, \mu)$, where $U_t = u_t(L)$ is a square-integrable local target, $S_t = [\mathcal{O}_U(L)]_t$ is the information available to the decoder, and μ is the training-time law. A decoder is a measurable map $v(S_t, t)$ (or a generator parameterized by the same information in a non-Euclidean setting).*

Figure 1 clarifies the relation to VAEs and flow matching. Coupling and bridge kernels are native objects in flow matching and stochastic-interpolant constructions. They become a path-space VAE inference model only when promoted to a learned conditional law and paired with a probabilistic generative path law and terminal likelihood; see section E. Thus BGM is not simply "a larger VAE." It generalizes the coupling–bridge–projection structure of flow matching and admits a variational interpretation under additional modeling commitments.

Examples of U_t. For an absolutely continuous deterministic interpolation, $U_t = \partial_t X_t$. For stochastic models, the target must be specified: a forward or controlled drift, current velocity, probability-flow target, score-derived target, or another objective-specific local quantity. Different targets define different gaps. In the Brownian experiments below, we use the terminal-drift target $U_t = (X - X_t)/(1-t)$; it is not a pathwise derivative.

Examples of S_t. A memoryless ODE decoder has $S_t = X_t$. A feature-space diagnostic may use $S_t = (\phi(X_t), t)$. A stateful sampler may include memory, latent variables, branch labels, or augmented field coordinates. The definition is therefore relative to the actual sampler interface under analysis.

3 Target- and observation-relative Markovization

Define the population matching objective

$$\mathcal{L}_\mu(v; \mathscr{Q}, U, \mathcal{O}_U) = \int \mathbb{E}_{\mathscr{Q}}\left[\|U_t - v(S_t, t)\|_2^2\right] \mu(\mathrm{d}t). \tag{3}$$

Definition 3 (Markovization gap)**.** *Assume $\int \mathbb{E}\|U_t\|_2^2 \mu(\mathrm{d}t) < \infty$. The Markovization gap is*

$$\mathfrak{G}_\mu(\mathscr{Q}, U, \mathcal{O}_U) := \inf_v \mathcal{L}_\mu(v; \mathscr{Q}, U, \mathcal{O}_U), \tag{4}$$

where the infimum ranges over measurable decoders of the permitted information (S_t, t).

Proposition 1 (Projection decomposition)**.** *The population minimizer is $v^*(s,t) = \mathbb{E}[U_t \mid S_t = s, t]$, and*

$$\mathfrak{G}_\mu(\mathscr{Q}, U, \mathcal{O}_U) = \int \mathbb{E}[\mathrm{Tr}\,\mathrm{Var}(U_t \mid S_t, t)]\, \mu(\mathrm{d}t), \tag{5}$$

$$\mathcal{L}_\mu(v) = \mathfrak{G}_\mu(\mathscr{Q}, U, \mathcal{O}_U) + \int \mathbb{E}\|v(S_t, t) - v^*(S_t, t)\|_2^2 \mu(\mathrm{d}t). \tag{6}$$

The proof is the standard Hilbert-space orthogonal projection argument and is included in section B. We do not claim novelty for this identity in isolation. The BGM contribution is the specification of $(\mathscr{Q}, U, \mathcal{O}_U, \mu)$ as a generative-model design object and the use of the residual as a pre-training bridge diagnostic.

Proposition 2 (Information monotonicity)**.** *Suppose S'_t is at least as informative as S_t for μ-almost every t, i.e., $\sigma(S_t, t) \subseteq \sigma(S'_t, t)$. Then*

$$\mathfrak{G}_\mu(\mathscr{Q}, U, \mathcal{O}'_U) \leq \mathfrak{G}_\mu(\mathscr{Q}, U, \mathcal{O}_U).$$

This resolves the scope of the bottleneck. Additional memory or latent state can reduce or eliminate the gap; the lower bound is fundamental only relative to the permitted information state. The exact observation experiment in section 4 is a finite-dimensional illustration.

Clock dependence. The measure μ is part of the objective. If a geometric path is traversed under a new clock, both the target magnitude and the weighting of path locations change. Consequently, deterministic schedules that trace similar curves need not have identical gaps or optimization behavior. This dependence is visible in the CIFAR-10 schedule stress test.

Relation to terminal error. When $S_t = X_t$, U_t is a velocity target, and v^* generates the reference marginal path, a standard coupling–Gronwall argument gives the following specialization.

Theorem 1 (Terminal stability above the projection floor)**.** *Let ρ_t be transported by v^* and ρ^θ_t by a uniformly L-Lipschitz vector field v_θ from the same initial law. If both ODEs are well posed and*

$$\epsilon^2 = \int_0^1 \mathbb{E}_{X_t \sim \rho_t} \|v_\theta(X_t, t) - v^*(X_t, t)\|_2^2 \, \mathrm{d}t < \infty,$$

then $W_2(\rho^\theta_1, \rho_1) \leq e^L \epsilon$, and the excess matching loss equals ϵ^2.

The theorem deliberately controls only the error *above* the structural floor. It does not imply that a lower gap guarantees better finite-network optimization, numerical stability, or perceptual quality.

4 Exact Gaussian controls

The exact study is designed to answer three questions independently: Does coupling matter under fixed marginals? Does the selected stochastic target matter at fixed endpoint cost? Does the visible state matter while cost and target variance remain fixed?

4.1 Setup

Let $Z \sim \mathcal{N}(0, I_3)$ and $X \sim \mathcal{N}(0, \Sigma)$, where

$$\Sigma = R \, \mathrm{diag}(4, 2, 1.5) R^\top, \qquad R = R_z(0.70) R_y(-0.45) R_x(0.85).$$

All coupling rows preserve these marginals. They are parameterized by $X = AZ + B\xi$ with independent $\xi \sim \mathcal{N}(0, I_3)$:

- Monge OT: $A = \Sigma^{1/2}$, $B = 0$;
- noisy OT: $A = (\Sigma - \sigma^2 I)^{1/2}$, $B = \sigma I$, $\sigma \in \{0.3, 0.6, 1.0\}$;
- independent: $A = 0$, $B = \Sigma^{1/2}$.

We use $t \sim \mathrm{Unif}[0.05, 0.95]$. The straight target is

$$X_t = (1 - t)Z + tX, \qquad U_t = X - Z.$$

The Brownian target uses $\epsilon = 0.5$,

$$X_t = (1 - t)Z + tX + \sqrt{\epsilon t(1 - t)}\, \eta, \qquad U_t = \frac{X - X_t}{1 - t},$$

with independent $\eta \sim \mathcal{N}(0, I_3)$. Visible states are the full X_t, the first two coordinates, or the first coordinate. Gaussian conditional covariance yields the exact instantaneous gap.

Table 1: Exact full-state coupling continuum. Endpoint marginals are fixed; only cross-covariance and local target change. Average gaps are comparable with held-out sum-MSE.

Target	Coupling	Cost	$\mathrm{Var}(U)$	Gap	η
Straight	Monge	1.222	1.222	0.000	0.000
Straight	Noisy OT 0.3	1.406	1.406	0.182	0.129
Straight	Noisy OT 0.6	1.988	1.988	0.749	0.377
Straight	Noisy OT 1.0	3.622	3.622	2.270	0.627
Straight	Independent	10.500	10.500	7.484	0.713
Brownian	Monge	1.222	4.630	3.882	0.839
Brownian	Noisy OT 0.3	1.406	4.814	4.072	0.846
Brownian	Noisy OT 0.6	1.988	5.395	4.666	0.865
Brownian	Noisy OT 1.0	3.622	7.029	6.272	0.892
Brownian	Independent	10.500	13.907	11.969	0.861

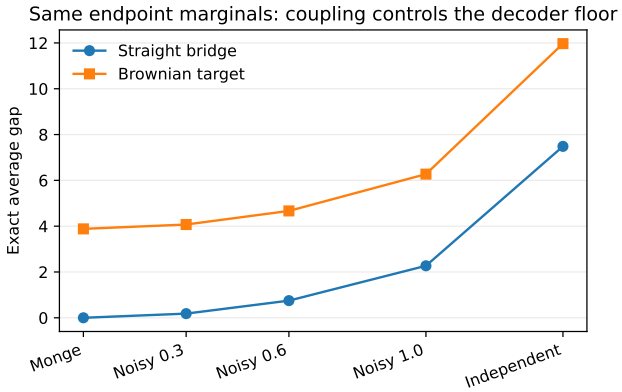

Figure 2: Degrading endpoint pairing raises the exact decoder floor under both selected targets.

4.2 Coupling and bridge target

For the straight target, the gap increases $0 \to 0.182 \to 0.749 \to 2.270 \to 7.484$ as the endpoint pairing is progressively degraded. At fixed Monge coupling and endpoint cost, replacing the straight target with the Brownian terminal drift raises the floor from 0 to 3.882. The latter is a target-relative statement, not evidence that one pathwise "velocity" is canonical.

4.3 Observation map and learned decoder

The full-state target is deterministic given (X_t, t), while projection collapses bridge records that require different targets. The experiment therefore holds simple energy statistics fixed and changes only the sampler interface. Across 11 trained configurations spanning coupling, target and observation, exact gap versus mean MLP test MSE has Spearman and Pearson correlation 1.0000 to four decimals; the MLP matches the Gaussian oracle nearly exactly.

5 A calibrated finite-sample proxy

The exact gap is rarely available in image space. We therefore estimate the held-out conditional-mean error

$$\widehat{\mathfrak{G}}_{\mathrm{CF}} = \frac{1}{N_{\mathrm{te}}} \sum_{i=1}^{N_{\mathrm{te}}} \left\| U_i - \widehat{\mathbb{E}}[U_i \mid \phi(S_i), t_i] \right\|_2^2, \tag{7}$$

where ϕ is a common PCA feature map fitted on pooled candidate states and the conditional mean is a k-nearest-neighbor regressor trained on a disjoint split. This is a *ranking proxy*: it includes finite-sample conditional-mean estimation error and is neither unbiased nor a certified lower bound.

Table 2: Exact observation ablation for the same straight Monge bridge. Endpoint cost and target variance are unchanged.

Visible state	Cost	$\mathrm{Var}(U)$	Exact gap	Oracle MSE	MLP MSE
Full 3D	1.222	1.222	0.000	0.000	0.000004
2D projection	1.222	1.222	0.227	0.226	0.226
1D projection	1.222	1.222	0.575	0.577	0.577

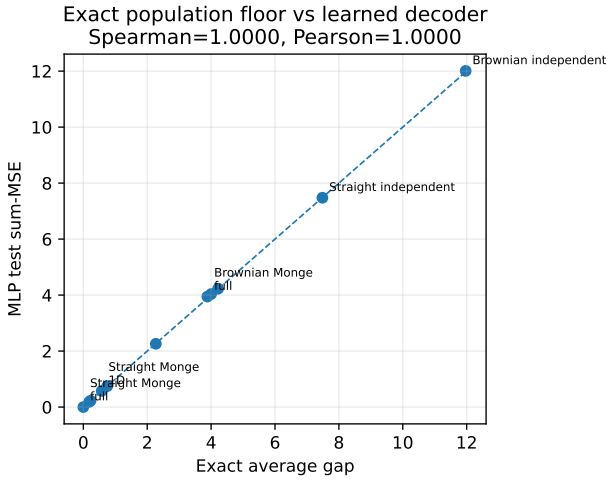

Figure 3: A trained Markov decoder reaches the target- and observation-relative analytic floor. The small Brownian discrepancy comes from evaluating sampled times against a continuous-time integral, not optimization failure.

Calibration. We apply exactly the same pipeline to a 32-dimensional Gaussian family with known population gaps. The sweep varies $N \in \{2000, 5000, 10000\}$, $k \in \{16, 32, 64\}$, PCA dimension $\{8, 16, 32\}$, seven bridge/coupling configurations, and three seeds. At $N = 10000$, the median exact-versus-proxy Spearman correlation is 1.000 for both cross-fitted MSE and local variance. This supports rank calibration in the tested family, not numerical unbiasedness.

Estimator choice. On the broader CIFAR-10 sweep, cross-fitted MSE has mean pairwise rank stability 0.993 and minimum 0.983 across $d_{\mathrm{PCA}} \in \{16, 32, 64\}$ and $k \in \{16, 32, 64\}$. Local neighborhood variance is less stable, especially across PCA dimensions (mean 0.763, minimum 0.617). We therefore use eq. (7) as the primary practical score and report local variance only as a sensitivity analysis.

Complexity. With N_{tr} reference samples, N_{te} evaluation samples, feature dimension r, and M sampled times, brute-force search costs $O(MN_{\mathrm{tr}}N_{\mathrm{te}}r)$ time. Query batching of size b reduces working distance memory from $O(N_{\mathrm{tr}}N_{\mathrm{te}})$ to $O(bN_{\mathrm{tr}})$. Independent coupling is linear in N; Hungarian matching in batches of size B costs $O(B^3)$ per batch, or $O(NB^2)$ over the dataset. Additional details and pseudocode appear in section H.

6 Real-image interventions

The image experiments deliberately use validation sum-MSE as the primary downstream quantity because it is linked directly to eq. (6). Sample-quality metrics are reported only as secondary scope checks.

6.1 Fashion-MNIST coupling continuum

One minibatch-OT assignment is progressively corrupted by re-permuting 25%, 50%, or 75% of pairs before reaching an independent permutation. A sixth row keeps the OT endpoints but changes

Table 3: Fashion-MNIST coupling continuum. Validation sum-MSE is mean ± SD over three seeds.

Design	Val. MSE	$\widehat{\mathfrak{G}}_{\mathrm{CF}}$	$\mathrm{Var}(U)$	Endpoint cost
Straight OT	140.62 ± 0.55	877.81	985.49	1247.94
Scramble 25%	152.61 ± 0.48	891.66	1002.68	1265.58
Scramble 50%	163.64 ± 0.52	906.83	1021.46	1282.99
Scramble 75%	173.95 ± 0.92	920.43	1039.38	1300.70
Independent	183.70 ± 0.60	933.71	1058.82	1318.34
Brownian OT	250.23 ± 0.51	1791.57	1866.94	1247.94

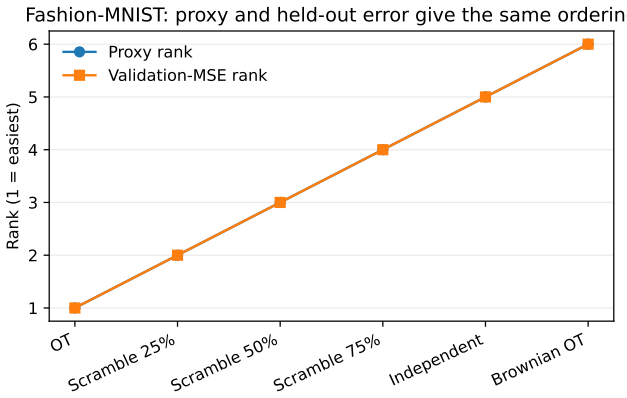

Figure 4: The cross-fitted proxy and held-out error induce the same six-design ranking.

to the Brownian terminal-drift target. Architecture, optimizer, batch size, 3500 updates, observation (X_t, t), and validation protocol are fixed; results use three optimization seeds.

Across all six designs, the proxy and validation MSE have Spearman correlation 1.000 and Pearson correlation 0.940. Across the five straight couplings alone, Spearman remains 1.000 and Pearson is 0.9995. Brownian OT and straight OT have the same endpoint cost but a 78% validation-MSE difference, showing that endpoint cost alone does not rank bridge targets. Target variance is nevertheless also strongly predictive; the image experiment supports absolute regression difficulty, not scale-independent ambiguity.

6.2 CIFAR-10 bridge-family stress test

The CIFAR-10 sweep includes linear, power-2, smoothstep, and VP-cosine deterministic paths; Brownian targets at three noise levels; a linear independent-coupling reference; and a VE corruption control. The seven principal OT-coupled rows share endpoint cost 3738.17 and use the same decoder and 6000-update budget with three seeds.

Across these seven designs, $\rho_{\mathrm{S}} = 0.893$ and $r_{\mathrm{P}} = 0.897$. Within the Brownian-noise family, the ordering is monotone. Yet the proxy does not perfectly rank deterministic schedules: VP cosine has lower proxy than smoothstep but higher validation error. This is a useful counterexample to a universal predictor and motivates the explicit dependence on μ, optimization, and numerical conditioning. Total target variance is comparably predictive ($\rho = 0.893$, $r = 0.902$).

6.3 CIFAR-100 negative control for final sample metrics

A two-configuration pilot compares independent and minibatch-OT straight bridges under matched compute. Minibatch OT lowers the proxy from 115.61 to 98.86 and training loss from 0.1625 to 0.1418, while FID worsens from 47.16 to 48.60, KID from 0.0350 to 0.0378, and IS falls from 5.97 to 5.79. The retained pilot summary has no seed-level uncertainty and is not presented as a powered sample-quality benchmark. Its role is conceptual: easier target regression need not dominate all downstream sources of error.

Table 4: Seven directly comparable OT-coupled CIFAR-10 bridge designs.

Bridge design	Val. MSE	$\widehat{\mathfrak{G}}_{\mathrm{CF}}$	$\mathrm{Var}(U)$
Linear	464.56 ± 1.60	3445.49	3725.90
Brownian $\epsilon = 0.15$	483.82 ± 5.95	3605.08	3882.34
Brownian $\epsilon = 0.35$	566.83 ± 3.38	4315.79	4578.83
Power-2 schedule	582.57 ± 6.54	4350.20	4743.42
Smoothstep	632.11 ± 8.31	4627.97	4943.76
VP cosine	675.80 ± 8.07	4347.11	4640.20
Brownian $\epsilon = 0.60$	728.28 ± 11.76	5999.54	6228.52

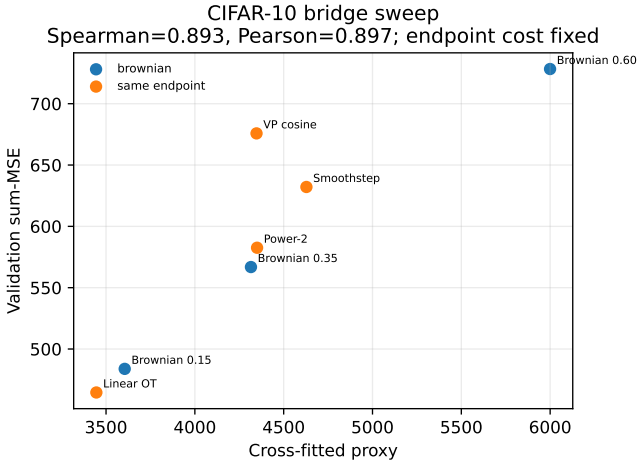

Figure 5: The proxy generalizes strongly, but not perfectly, across broader bridge choices at fixed endpoint cost.

6.4 Computational value

The full Fashion proxy sweep takes 14.1 seconds on one T4, compared with 3429.5 seconds for 18 downstream runs. The CIFAR-10 proxy sweep takes 49.2 seconds, compared with 11925.3 seconds for 27 runs. The diagnostic is therefore useful as a screening tool, although coupling construction and feature extraction add their own costs.

7 Positioning, limitations, and implications

Closest frameworks. Stochastic interpolants and diffusion flow matching already expose coupling, interpolation, and Markov projection [2, 3, 8]. Generator matching provides a broader infinitesimal-generator view [1, 11]. Schrödinger methods study entropy-minimizing reciprocal and Markov path laws [14, 19, 27, 34]. BGM is complementary: it starts from the endpoint-coupled bridge target and measures the residual error of the conditional-to-sampler information projection. The scalar residual, its target/observation dependence, and its use as a bridge-design diagnostic are the focus here.

What is supported. The exact gap is an irreducible squared-loss floor for the specified target and information state. Coupling, target choice, and visible information can change that floor independently. Cross-fitted feature-space error is rank-calibrated in the tested Gaussian family and is useful for ranking substantial coupling or bridge interventions under a fixed training budget.

What is not supported. The practical proxy is not an unbiased population estimate or a certified lower bound. The image studies do not establish predictive information beyond total target variance. Normalized η is representation- and estimator-sensitive. Lower gap does not universally imply better FID, KID, or IS. Dynamics representation is structurally separated by BGM but is not empirically validated as extensively as coupling, target, and observation.

Broader BGM program. Current-preserving ODE/SDE realizations, path-space variational inference, and field-line kernels remain useful components of the larger framework. They are moved to sections D to F to keep the core narrative focused. The present result should be read as the first rigorously measured interface inside a broader latent-path graphical model: non-causal bridge information must be compressed into a causal sampler representation.

8 Conclusion

Bridge design is not finished when a path connects the correct endpoint marginals. A deployed sampler must recover the objective-specific local target from the information it is allowed to observe. The Markovization gap makes that compression explicit. The refined quantity $\mathfrak{G}_\mu(\mathscr{Q}, U, \mathcal{O}_U)$ resolves ambiguity about stochastic targets, memory, latent state, and clock choice. Exact Gaussian controls show that coupling, target, and visible information independently alter the irreducible floor, while calibrated image experiments demonstrate practical ranking value and expose the diagnostic's limits. The resulting message is deliberately narrow: the gap predicts the difficulty of the bridge-to-decoder regression problem, not every aspect of final generation.

References

[1] Eddie Aamari and Arthur Stéphanovitch. Statistical analysis of markovian generative modeling. *arXiv preprint*, 2026.

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

[3] Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint*, 2023.

[4] Joe Benton, George Deligiannidis, and Arnaud Doucet. Error bounds for flow matching methods. *Transactions on Machine Learning Research*, 2024. arXiv:2305.16860.

[5] Gerard Brunick and Steven Shreve. Mimicking an itô process by a solution of a stochastic differential equation. *The Annals of Applied Probability*, 23(4):1584–1628, 2013.

[6] Hans Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 101–203. 1988.

[7] Mark I. Freidlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, 3 edition, 2012.

[8] Marta Gentiloni Silveri, Giovanni Conforti, and Alain Durmus. Theoretical guarantees in KL for diffusion flow matching. In *Advances in Neural Information Processing Systems*, 2024.

[9] István Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an itô differential. *Probability Theory and Related Fields*, 71(4):501–516, 1986.

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

[11] Peter Holderrieth, Marton Havasi, Jason Yim, Neta Shaul, Itai Gat, Tommi Jaakkola, Brian Karrer, Ricky T. Q. Chen, and Yaron Lipman. Generator matching: Generative modeling with arbitrary markov processes. In *International Conference on Learning Representations*, 2025.

[12] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[13] Mohammad Jalali, Cheuk Ting Li, and Farzan Farnia. An information-theoretic evaluation of generative models in learning multi-modal distributions. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

[14] Sergei Kholkin, Grigoriy Ksenofontov, David Li, Nikita Maksimovich Kornilov, Nikita Gushchin, Alexandra Suvorikova, Alexey Kroshnin, Evgeny Burnaev, and Alexander Korotin. Diffusion and adversarial schrödinger bridges via iterative proportional markovian fitting. In *International Conference on Learning Representations*, 2026.

[15] Minyoung Kim. A unified framework for diffusion bridge problems: Flow matching and schrödinger matching into one. *arXiv preprint*, 2025.

[16] Diederik P. Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

[17] Alexander Kolesov, Stepan I. Manukhov, Vladimir V. Palyulin, and Alexander Korotin. Field matching: An electrostatic paradigm to generate and transfer data. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025.

[18] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[19] Christian Léonard. A survey of the schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems-A*, 34(4):1533–1574, 2014.

[20] 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.

[21] 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*, 2023.

[22] Stepan I. Manukhov, Alexander Kolesov, Vladimir V. Palyulin, and Alexander Korotin. Interaction field matching: Overcoming limitations of electrostatic models. In *International Conference on Learning Representations*, 2026.

[23] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 6 edition, 2003.

[24] Grigorios A. Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*. Springer, 2014.

[25] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[26] Hannes Risken. *The Fokker–Planck Equation: Methods of Solution and Applications*. Springer, 2 edition, 1996.

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

[28] Daniil Shlenskii, Alexander Varlamov, Nazar Buzun, and Alexander Korotin. Unlocking the duality between flow and field matching. *arXiv preprint*, 2026.

[29] Saurabh Singh and Dmitry Lagun. Latent stochastic interpolants. *arXiv preprint*, 2025.

[30] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 2256–2265, 2015.

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

[32] Sophia Tang. Foundations of schrödinger bridges for generative modeling. *arXiv preprint arXiv:2603.18992*, 2026.

[33] 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*, 2024.

[34] Alexander 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 *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, 2024.

[35] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. Scikit-image: Image processing in python. *PeerJ*, 2:e453, 2014.

[36] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

[37] Yilun Xu, Ziming Liu, Max Tegmark, and Tommi Jaakkola. Poisson flow generative models. In *Advances in Neural Information Processing Systems*, 2022.

[38] Yilun Xu, Ziming Liu, Yonglong Tian, Shangyuan Tong, Max Tegmark, and Tommi Jaakkola. PFGM++: Unlocking the potential of physics-inspired generative models. In *Proceedings of the 40th International Conference on Machine Learning*, 2023.

A Formal path-space setup

Let $(\mathcal{M}, \mathcal{B})$ be Polish and $\mathcal{X} = C([0,1], \mathcal{M})$ with the Borel sigma-algebra induced by the uniform topology. The endpoint map is $e(\omega) = (X_0(\omega), X_1(\omega))$. If $\mathscr{Q}$ is any path measure with endpoint law $\Gamma = e_{\#}\mathscr{Q}$, disintegration gives a probability kernel $(z, x) \mapsto \mathsf{R}^{z,x}$ such that

$$\mathscr{Q}(\,\mathrm{d}\omega) = \int \mathsf{R}^{z,x}(\,\mathrm{d}\omega)\,\Gamma(\,\mathrm{d}z,\,\mathrm{d}x).$$

Conversely, any coupling and bridge kernel define an aggregated path law. This justifies treating coupling and bridge as distinct graphical-model components.

The full record L may enlarge ω with auxiliary variables. The target U_t need only be square-integrable under $\mathscr{Q} \otimes \mu$. The projection results therefore do not require smooth densities or a unique sampler. Smoothness enters only when the target is interpreted as a velocity that transports a marginal path or when terminal stability is invoked.

B Proofs for the core results

B.1 Proof of proposition 1

For fixed t, let $\mathcal{H}_t$ be the closed subspace of $L^2(\mathscr{Q})$ containing functions measurable with respect to $\sigma(S_t, t)$. The conditional expectation $\mathbb{E}[U_t \mid S_t, t]$ is the orthogonal projection of U_t onto $\mathcal{H}_t$. Therefore, for every admissible v,

$$\begin{aligned}\mathbb{E}\|U_t - v(S_t,t)\|_2^2 &= \mathbb{E}\|U_t - \mathbb{E}[U_t \mid S_t, t]\|_2^2 \\ &\quad + \mathbb{E}\|\mathbb{E}[U_t \mid S_t, t] - v(S_t,t)\|_2^2,\end{aligned}$$

with zero cross term. The first term equals $\mathbb{E}\,\mathrm{Tr}\,\mathrm{Var}(U_t \mid S_t, t)$. Integrating against μ proves eqs. (5) and (6). Uniqueness is up to (S_t, t)-null sets for μ-almost every t.

B.2 Proof of proposition 2

By the law of total variance,

$$\mathbb{E}\,\mathrm{Tr}\,\mathrm{Var}(U_t \mid S_t, t) = \mathbb{E}\,\mathrm{Tr}\,\mathrm{Var}(U_t \mid S'_t, t) + \mathbb{E}\|\mathbb{E}[U_t \mid S'_t, t] - \mathbb{E}[U_t \mid S_t, t]\|_2^2$$

when $\sigma(S_t, t) \subseteq \sigma(S'_t, t)$. Integrating yields the claim.

B.3 Proof sketch for theorem 1

Couple the flows by using the same initial point. If X_t follows v^* and Y_t follows v_θ, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\|X_t - Y_t\| \le L\|X_t - Y_t\| + \|v_\theta(X_t,t) - v^*(X_t,t)\|.$$

Gronwall, Cauchy–Schwarz, and expectation give $\mathbb{E}\|X_1 - Y_1\|^{21/2} \le e^L \epsilon$, hence the Wasserstein bound. The loss identity is eq. (6) with $S_t = X_t$.

C Exact Gaussian derivation

Write $X = AZ + B\xi$, with independent standard Gaussian Z, ξ. Let $C = \mathrm{Cov}(Z, X) = A^\top$. For a straight bridge,

$$\begin{aligned}\Sigma_t &= (1-t)^2 I + t^2 \Sigma + t(1-t)(C + C^\top), \\ \Sigma_U &= I + \Sigma - C - C^\top, \\ \Sigma_{U,X_t} &= (1-t)(C^\top - I) + t(\Sigma - C).\end{aligned}$$

For a visible linear projection $S_t = PX_t$, Gaussian conditioning gives

$$\mathrm{Var}(U_t \mid S_t) = \Sigma_U - \Sigma_{U,X_t} P^\top (P\Sigma_t P^\top)^\dagger P \Sigma_{X_t,U}. \tag{8}$$

The instantaneous gap is the trace of eq. (8). For the Brownian target, add $\epsilon t(1-t)I$ to Σ_t, $\epsilon t/(1-t)I$ to Σ_U, and $-\epsilon tI$ to Σ_{U,X_t}. Numerical integration uses 801 points over $[0.05, 0.95]$.

The MLP has three hidden layers of width 128 with SiLU activations, receives (S_t, t), and predicts the full $U_t \in \mathbb{R}^3$. The rebuttal-scale run uses three optimization seeds and a held-out test set; sum-MSE is reported so that it is directly comparable to $\operatorname{Tr}\operatorname{Var}(U_t \mid S_t, t)$.

D Current-preserving dynamics representation

This component is a representation choice after a marginal path and current are specified. Let ρ_t satisfy

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0, \qquad J_t = \rho_t v_t.$$

For an SDE $\mathrm{d}X_t = b_t(X_t)\,\mathrm{d}t + \sigma_t(X_t)\,\mathrm{d}B_t$ with covariance $a = \sigma\sigma^\top$, the Fokker–Planck current is

$$J_{t,i} = \rho_t b_{t,i} - \frac{1}{2}\sum_j \partial_j(a_{ij}\rho_t).$$

If $\rho_t > 0$ and the equations are well posed, then

$$b_i^{(a)}(x,t) = v_i(x,t) + \frac{1}{2\rho_t(x)}\sum_j \partial_j(a_{ij}(x,t)\rho_t(x))$$

realizes the same marginal path. For $a = g(t)^2 I$, this becomes $b^{(a)} = v + (g^2/2)\nabla \log \rho_t$. This standard identity separates the projected current from its ODE/SDE representation; it is not a claim that score estimation is automatically well conditioned [24, 26, 31].

E Path-space variational interpretation

A coupling becomes a VAE-style encoder only after it is represented as a conditional inference law $q_\phi(\mathrm{d}z \mid x)\pi_{\mathrm{data}}(\mathrm{d}x)$ and paired with a generative path law. Let

$$\mathscr{Q}_{\phi,\lambda}(\mathrm{d}Y_{0:1} \mid x) = \int \mathsf{R}_\lambda^{z,x}(\mathrm{d}Y_{0:1}) q_\phi(\mathrm{d}z \mid x).$$

Formally, if $\mathscr{Q}_{\phi,\lambda}(\cdot \mid x) \ll \mathsf{P}_\theta$ and the terminal likelihood is integrable,

$$\log p_{\theta,\psi}(x) \geq \mathbb{E}_{\mathscr{Q}_{\phi,\lambda}}\left[\log p_\psi(x \mid Y_1) - \log \frac{\mathrm{d}\mathscr{Q}_{\phi,\lambda}(\cdot \mid x)}{\mathrm{d}\mathsf{P}_\theta}(Y_{0:1})\right].$$

For exact point-conditioned bridges on a continuous state space, the closed-interval KL is generally singular. A finite objective requires truncation to $[0, 1-\delta]$ with a terminal decoder, or non-degenerate terminal noise. Shared-covariance SDEs admit a Girsanov drift-matching reduction under support and integrability assumptions [23]; deterministic flow matching is a singular/optimal-control relaxation. This interpretation organizes the broader BGM program but is not needed for the Markovization-gap theory.

F Field-line bridges and augmented observations

Poisson and electrostatic models can be represented by a potential Φ, field $F = -\nabla\Phi$, and flow $\dot{\varphi}_s = F(\varphi_s)$. A hitting map $T_\Phi(z)$ induces a coupling $(\mathrm{Id}, T_\Phi)_\# \pi_0$ and deterministic field-line bridge kernel. Singularities, nonunique hitting paths, and time parameterization are part of the model specification; small-noise field-biased bridges give one regularization route [17, 22, 28, 37, 38].

A controlled toy example isolates the role of augmented state. Let $H \in \{-1, +1\}$ be equiprobable, $S \sim \mathrm{Unif}(\mathbb{S}^1)$, $X_t = S + Ht \pmod 1$, and $U_t = H$. The observed X_t is independent of H, so $\operatorname{Var}(U_t \mid X_t) = 1$. If the state is augmented to (X_t, H, t), the conditional variance is zero. This is precisely proposition 2: augmented field coordinates may preserve branch information that is lost in data space.

G Gaussian score, noise, and velocity reparameterization

For $X_t = \alpha_t X + \sigma_t \varepsilon$ with independent $\varepsilon \sim \mathcal{N}(0, I)$, Tweedie's identity gives

$$\mathbb{E}[\varepsilon \mid X_t = x] = -\sigma_t \nabla \log \rho_t(x).$$

Consequently,

$$\mathbb{E}[\dot{X}_t \mid X_t = x] = \frac{\dot{\alpha}_t}{\alpha_t} x + \left(\frac{\dot{\alpha}_t \sigma_t^2}{\alpha_t} - \dot{\sigma}_t \sigma_t \right) \nabla \log \rho_t(x).$$

Thus score, noise, denoiser, and velocity targets are schedule-dependent linear reparameterizations in this Gaussian setting. Outside it, conditioning and numerical properties may differ even when the marginal current agrees [10, 16, 31].

H Estimator algorithm and complexity details

Given bridge samples and times, form training tuples $(\phi(S_i), t_i, U_i)$ and a disjoint evaluation split. Fit one common PCA basis on pooled candidate states. For each query, retrieve k nearest reference points in the joint standardized feature-time metric, average their targets, and accumulate held-out squared residuals. Using a common feature basis is essential: fitting a separate PCA basis per design would change the metric and confound ranking.

Table 5: Asymptotic costs of the canonical diagnostic pipeline.

Operation	Time	Working memory
PCA projection (after fitting)	$O(Ndr)$	$O(Nr)$
Brute-force cross-fitted k-NN	$O(MN_{\mathrm{tr}}N_{\mathrm{te}}r)$	$O(bN_{\mathrm{tr}})$ with query chunk b
Independent endpoint pairing	$O(Nd)$	$O(Nd)$
Hungarian minibatch OT	$O(NB^2)$ overall	$O(B^2)$
Sinkhorn minibatch OT (I iterations)	$O(NIB)$ overall	$O(B^2)$
Learned encoder inference	$O(NC_{\mathrm{enc}})$	implementation-dependent

The Sinkhorn row follows from N/B batches, each costing $O(IB^2)$. Approximate-neighbor search can reduce practical cost but does not remove high-dimensional statistical difficulty.

I Additional experimental protocol and legacy pilots

Fashion-MNIST. The canonical run uses the full $60{,}000/10{,}000$ split, minibatch-OT batch size 128, a pooled common PCA basis, $N = 12{,}000$ proxy samples, $d_{\mathrm{PCA}} \in \{16, 32, 64\}$, $k \in \{16, 32, 64\}$, 3500 neural updates, batch size 256, and seeds $11, 29, 47$.

CIFAR-10. The canonical bridge sweep uses minibatch OT within batches of 128, $N = 12{,}000$ proxy samples, $d_{\mathrm{PCA}} \in \{16, 32, 64\}$, $k \in \{16, 32, 64\}$, 6000 updates, batch size 64, and seeds $11, 29, 47$. Correlations in the main text use the seven OT-coupled deterministic/Brownian rows; independent coupling and VE corruption are external controls.

Earlier pilots. The original manuscript also reported an eight-Gaussian synthetic comparison, PCA-latent Digits/LFW decoder experiments, and two-coupling CIFAR-10/Fashion-MNIST sample-quality pilots. They are consistent with the narrower claim that minibatch OT often reduces absolute regression burden, but the new exact controls, graded coupling continuum, proxy calibration, and bridge-family sweep supersede them as primary evidence. We omit the older tables from the core paper to avoid conflating different proxy normalizations and experimental scopes.

J Claim-to-evidence summary

Claim	Evidence / qualification	Status
Exact gap is the irreducible decoder floor	Gaussian conditional covariance, oracle MSE, 11-config MLP study	Supported
Coupling changes the attainable regression problem	Noisy-OT exact continuum and Fashion scramble continuum	Supported
Bridge target matters beyond endpoint cost	Straight/Brownian exact control and fixed-cost image rows	Supported
Visible information changes the gap for the same bridge	Full/2D/1D exact ablation and monotonicity	Supported
Cross-fitted proxy preserves exact rankings	Gaussian calibration over N, PCA dimension, and k	Qualified
Proxy ranks absolute fixed-budget image regression difficulty	Fashion $\rho = 1$; CIFAR-10 $\rho = 0.893$	Supported
Proxy adds image predictive power beyond target variance	Target variance remains comparably predictive	Not established
Normalized η universally ranks designs	Estimator- and representation-sensitive	Not established
Lower gap universally improves FID/KID/IS	CIFAR-100 provides a counterexample	Not established

NeurIPS Paper Checklist

1. **Claims.** [Yes] The abstract and introduction state the target- and observation-relative population claim and distinguish it from the finite-sample proxy and final sample quality. Supporting evidence appears in Sections 3–6 and Appendix I.
2. **Limitations.** [Yes] Section 7 discusses the classical nature of the projection identity, target and observation dependence, estimator bias, target-scale confounding, limited datasets, and failure to predict all perceptual metrics.
3. **Theory assumptions and proofs.** [Yes] Assumptions are stated with each result. Proofs of the projection, information-monotonicity, and terminal-stability results appear in Appendix B.
4. **Experimental reproducibility.** [Yes] Architectures, sample sizes, time windows, seeds, optimization budgets, bridge targets, coupling construction, and evaluation conventions are given in Sections 4–6 and Appendix H.
5. **Open access to data and code.** [Yes] All datasets are public benchmarks, and the camera-ready artifact includes the experiment notebooks, exported tables, and plotting data needed to reproduce the reported analyses.
6. **Experimental settings.** [Yes] Hyperparameters and data splits are reported in Appendix H; common architectures and fixed compute are used within each comparison.
7. **Statistical significance.** [Yes] Neural experiments report means and standard deviations over three seeds where applicable. Rank and linear correlations are reported over the pre-specified design families; small-sample interpretations are qualified.
8. **Compute resources.** [Yes] The exact and proxy experiments and downstream image sweeps were run on a single NVIDIA Tesla T4. Wall-clock times are reported in Section 5 and Appendix G.
9. **Code of ethics.** [Yes] The work follows the NeurIPS Code of Ethics and studies general-purpose generative-model diagnostics using public datasets.
10. **Broader impacts.** [Yes] The diagnostic may reduce wasted training compute by screening bridge designs, but it does not certify sample quality, fairness, privacy, or safety. These limitations are stated explicitly.
11. **Safeguards.** [N/A] The paper does not release a high-risk deployed system or a model intended for sensitive decision making.
12. **Licenses for existing assets.** [Yes] The paper cites the public datasets and software libraries used. Authors should verify and preserve the applicable upstream licenses when distributing the artifact.
13. **New assets.** [Yes] The released notebooks and tables are research artifacts for reproducing the reported results; their documentation states the intended use and limitations.
14. **Human subjects or crowdsourcing.** [N/A] No human-subject or crowdsourcing study was conducted.
15. **Institutional review board.** [N/A] No human-subject research requiring IRB review was conducted.
16. **Use of large language models.** [Yes] Language-model assistance was used for editing and code drafting. The author verified the mathematics, experiments, citations, and final claims and remains responsible for all content.