GRAPH RESIDUAL CONJUGATE DIFFUSION: SNR-EQUALIZED HEAT FLOW FOR GRAPH SIGNALS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models generate data by reversing a forward corruption process that typically approaches a simple Gaussian prior. Recent work has extended this framework to signals supported on fixed graphs, e.g., road-network traffic and sensor-network measurements. Many graph signals have nonuniform spectral energy across graph frequencies, whereas isotropic corruption adds the same conditional noise variance to every graph-frequency mode. Driving all modes to near-zero terminal signal-to-noise ratio (SNR) requires strong corruption, which increases the noise range that must be covered under a fixed sampling budget. We introduce *Graph Residual Conjugate Diffusion* (GRCD), which replaces the shared clock of graph heat diffusion with a mode-dependent clock that gives every graph-Fourier mode the same conditional signal-to-noise ratio. GRCD fits a zero-mean graph-spectral Gaussian reference on the training split and, rather than driving the corruption to near-zero terminal SNR, stops at a finite terminal SNR at which the propagated reference still carries the fitted spectral variances, narrowing the SNR range that sampling must cover. The Gaussian component has an exact modewise propagator in the probability-flow ODE, so sampling advances it analytically and integrates only the learned residual score numerically. We evaluate GRCD on five graph-signal settings (the METR-LA traffic and Molene weather datasets, and three synthetic settings from stochastic block models) against seven comparators under a matched protocol: Graph-Aware Diffusion (GAD), EDM (Karras et al., 2022) adapted to the graph backbone, two graph adaptations of Whitened Score Diffusion (WSD), and three preconditioning controls. At four function evaluations (NFEs), GRCD lowers averaged maximum mean discrepancy (aMMD) by 22–$36\times$ over the best comparator on all five settings, reaching 0.054 on METR-LA, where it clears an aMMD 0.1 target with 87% less sampling wall-clock time than the cheapest comparator that reaches it. Within GRCD, fitting the terminal reference reduces aMMD by factors of 2.7–7.3 at finite terminal SNR, while the corresponding factors shrink to 1.00–1.01 near zero, matching the analytic limit in which the propagated reference loses dependence on the fitted covariance.

1 INTRODUCTION

Diffusion models combine a prescribed forward corruption process with learned reverse dynamics, while the forward process is chosen so that its terminal marginal approaches a simple tractable reference distribution, most commonly a Gaussian distribution (Ho et al., 2020; Song et al., 2021).

Graph signals can have strongly nonuniform spectral energy, especially when they are smooth with respect to the underlying graph topology (Ortega et al., 2018; Rozada et al., 2026). In our benchmarks this appears as a large variation in graph-Fourier coefficient variance across frequencies. However, isotropic corruption injects the same noise variance into every graph-Fourier mode (Ho et al., 2020; Song et al., 2021) independently of the spectral graph characteristics. Graph-Aware Diffusion (GAD) (Rozada et al., 2026) incorporates the graph structure through graph heat diffusion, while Whitened Score Diffusion (WSD) (Alido et al., 2025) uses anisotropic Gaussian corruption and structured spectral priors. In this work, we study a complementary question: whether a graph-signal diffusion process should be driven to an endpoint near zero SNR when a structured finite-SNR reference can instead be fitted and propagated analytically.

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.

We derive a mode-dependent clock, i.e., a frequency-dependent diffusion rate, for which the time-dependent graph-filter transformation $\widetilde{\mathbf{x}}_t = \boldsymbol{A}_t^{-1}\mathbf{x}_t$ reduces the forward process to isotropic Gaussian corruption. In these conjugate coordinates every graph-Fourier mode has unit signal coefficient and conditional noise variance $q(t)^2$, hence the same conditional SNR $q(t)^{-2}$. Thus, a single scalar $q(t)$ controls the spectral corruption level. We call this construction a *conjugate forward process*.

PriorGrad showed that data-dependent Gaussian priors can improve conditional diffusion (gil Lee et al., 2022). Instead of a prior that varies with the conditioning input, we fit one zero-mean Gaussian reference shared by every sample: we estimate a Ledoit–Wolf shrinkage covariance (Ledoit & Wolf, 2004), retain its diagonal variances in the graph-Fourier basis, and propagate the resulting reference to the chosen finite terminal SNR. This defines the initial distribution for reverse sampling.

We use the probability-flow ODE (Song et al., 2021) and decompose its score into an analytic Gaussian reference component and a learned residual. Related decompositions are known: Remy et al. (2023) combine an analytic Gaussian score with a learned residual trained by residual denoising score matching, while Wang & Vastola (2024) show that learned diffusion scores are well approximated by a Gaussian linear score at moderate to high noise levels and exploit the corresponding analytic dynamics for sampling acceleration. In our method, the Gaussian reference is fitted in the graph-Fourier basis and propagated under conjugate graph diffusion, so its score is available in closed form at every time. This reference also yields our sampler's exact modewise propagator.

Contributions: We propose *Graph Residual Conjugate Diffusion* (GRCD), decomposing the score into a Gaussian reference score and a learned residual correction. In graph-Fourier coordinates the Gaussian contribution has a closed-form modewise propagator, so only the residual is integrated numerically. The forward process, the fitted reference, the residual parameterization, and the exponential-residual solver are separable components of the diffusion process, allowing their effects to be evaluated through controlled ablations. Our empirical evaluation focuses on low-budget graph-signal generation and evaluates these components individually. Our main contributions are:

Conjugate graph diffusion. We derive a mode-dependent clock whose conjugate coordinates have equal conditional noise variance across graph frequencies, providing a single scalar corruption coordinate while retaining graph-dependent dynamics in the original signal space.
Fitted finite-SNR reference and residual propagation. We fit a Gaussian reference in the graph-Fourier basis to training data only and propagate it analytically to finite terminal SNR for reverse initialization. The same reference yields a closed-form propagator for the Gaussian component of the probability-flow ODE, leaving only the learned residual to be integrated numerically.
Asymptotic behavior at zero terminal SNR. We show analytically that the propagated reference becomes independent of the fitted spectral covariance as the terminal noise grows, converging to $\mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{L}_\delta^{-1})$, where $\boldsymbol{L}_\delta$ is the shifted, spectrally normalized graph Laplacian from Section 2. The experiments match this prediction: across five settings, fitting the terminal reference reduces aMMD by factors of 2.7–7.3 at finite terminal SNR, while the factors shrink to 1.00–1.01 near zero.
Low-budget generation. On METR-LA, GRCD reaches an aMMD of 0.054 with only 4 function evaluations (NFE), which is lower than every external baseline and preconditioning control at evaluated NFE budgets from 4 to 64. For an aMMD target of 0.10, GRCD reduces sampling wall-clock time by 87% relative to the cheapest of these that reaches the target.

2 Background and Setting

Graph signals. We consider a fixed, undirected weighted graph $\mathcal{G}$ on $N \in \mathbb{N}$ nodes. Let $\boldsymbol{W}$ denote the symmetric weighted adjacency matrix and $\boldsymbol{D}$ the diagonal degree matrix with $\boldsymbol{D}_{i,i} = \sum_{j=1}^{N} \boldsymbol{W}_{i,j}$. Following the standard graph-signal-processing construction (von Luxburg, 2007; Shuman et al., 2013; Ortega et al., 2018; Rozada et al., 2026), we use the symmetric normalized graph Laplacian $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$. For an undirected graph with nonnegative weights, $\boldsymbol{L}$ is symmetric positive semidefinite (von Luxburg, 2007), and therefore admits the eigendecomposition $\boldsymbol{L} = \boldsymbol{U}\,\mathrm{diag}(\boldsymbol{\lambda})\boldsymbol{U}^\top, \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}, \lambda_i \geq 0$. For the diffusion dynamics, we normalize the Laplacian by its largest eigenvalue $\lambda_{\max}$ to place the graph spectrum on a common scale across datasets, and add a small positive spectral shift $\delta > 0$ so that all eigenvalues are strictly positive: $\boldsymbol{L}_\delta = \frac{\boldsymbol{L}}{\lambda_{\max}} + \delta\boldsymbol{I} = \boldsymbol{U}\,\mathrm{diag}(\boldsymbol{\mu})\boldsymbol{U}^\top, \mu_i = \frac{\lambda_i}{\lambda_{\max}} + \delta > 0$. Because $\boldsymbol{L}_\delta$ is an affine function of $\boldsymbol{L}$, it has the same eigenvectors and preserves the ordering of the Laplacian eigenvalues.

Score-based diffusion. Let p_t denote the forward diffusion probability density at time t. A score-based diffusion model learns the time-dependent score vector $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$. For the corresponding SDE, there exists a deterministic probability-flow ODE that, when initialized from the same distribution, has the same marginal density p_t at every time t (Song et al., 2021). We use this probability-flow formulation for sampling and measure sampling cost in NFE.

Terminal SNR. The SNR provides a natural parameterization of diffusion noise levels (Kingma et al., 2021). Lin et al. (2024) showed for image diffusion models that a nonzero terminal SNR can create a training–inference mismatch when inference is initialized from pure noise. In GRCD, terminal SNR plays a different role: it controls how much fitted graph-spectral covariance remains in the propagated terminal reference. We define the corresponding quantity in Section 3.2.

3 METHODOLOGY

3.1 CONJUGATE GRAPH DIFFUSION

Motivated by graph heat-diffusion models such as Graph-Aware Diffusion (GAD) (Rozada et al., 2026), we seek a graph diffusion with a mode-dependent clock that is exactly conjugate to an isotropic variance-exploding (VE) process. Throughout this section $\mathbf{x}_t \in \mathbb{R}^N$ denotes the graph signal at diffusion time t and $x_{t,i} := [\boldsymbol{U}^\top \mathbf{x}_t]_i$ denotes its i-th graph-Fourier coefficient. μ_i is the eigenvalue of the shifted normalized Laplacian $\boldsymbol{L}_\delta$ associated with this mode. We model each graph-Fourier coefficient with the linear stochastic differential equation (SDE)

$$dx_{t,i} \;=\; -\mu_i c_i(t) x_{t,i}\, dt + \sqrt{2\sigma^2 c_i(t)}\, dW_{t,i}, \tag{1}$$

where $W_{t,i} := [\boldsymbol{U}^\top \boldsymbol{W}_t]_i$ is a standard one-dimensional Brownian motion, independent across modes, $c_i(t) \geq 0$ is the mode-dependent clock rate defined below, and $\sigma > 0$ is a fixed diffusion scale. The drift term $-\mu_i c_i(t) x_{t,i}\, dt$ damps mode i toward zero, while the stochastic term injects Gaussian noise with variance rate $2\sigma^2 c_i(t)$. The clock $c_i(t)$ controls each spectral mode's evolution.

Let $q(t) = \kappa t$ for $t \in [0, t_{\max}]$ and $\kappa > 0$, with training and sampling restricted to $t \in [t_{\min}, t_{\max}]$. In the conjugate coordinates, we use the isotropic VE process $d\widetilde{\mathbf{x}}_t \;=\; \sqrt{2q(t)q'(t)}\, d\boldsymbol{W}_t$, whose conditional marginals at each fixed t admit the reparameterization

$$\widetilde{\mathbf{x}}_t \;=\; \widetilde{\mathbf{x}}_0 + q(t)\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{2}$$

Proposition 1 (Graph-filter conjugate representation)**.** *For the isotropic VE process introduced above with $\widetilde{\mathbf{x}}_0 = \mathbf{x}_0$, define*

$$\boldsymbol{A}_t \;:=\; \left(\boldsymbol{I} + \frac{q(t)^2}{\sigma^2}\boldsymbol{L}_\delta\right)^{-1/2}, \tag{3}$$

and let $\mathbf{x}_t \;=\; \boldsymbol{A}_t \widetilde{\mathbf{x}}_t$. The matrix $\boldsymbol{A}_t$ has the same eigenvectors as $\boldsymbol{L}_\delta$, with eigenvalue $a_i(t) = \left(1 + \frac{\mu_i q(t)^2}{\sigma^2}\right)^{-1/2}$ in graph-Fourier mode i. The resulting process satisfies

$$x_{t,i} \;=\; a_i(t)\,(x_{0,i} + q(t)\epsilon_i)\,, \qquad \epsilon_i := [\boldsymbol{U}^\top \boldsymbol{\epsilon}]_i, \tag{4}$$

and hence $\mathrm{Var}(x_{t,i} \mid x_{0,i}) \;=\; a_i(t)^2 q(t)^2$. In conjugate coordinates, $\mathrm{Var}(\widetilde{x}_{t,i} \mid \widetilde{x}_{0,i}) \;=\; q(t)^2$ for every mode i, so all graph-Fourier modes carry the same conditional noise variance.

Corollary 1 (Induced mode-dependent clock)**.** *The graph-Fourier modes of the process described in Proposition 1 satisfy Equation 1 with*

$$c_i(t) \;=\; \frac{q(t)q'(t)}{\sigma^2 + \mu_i q(t)^2}. \tag{5}$$

The conditional noise variance induced in graph coordinates is $a_i(t)^2 q(t)^2 \;=\; \frac{\sigma^2 q(t)^2}{\sigma^2 + \mu_i q(t)^2}$, which approaches σ^2/μ_i as $q(t) \to \infty$, whereas the corresponding variance $q(t)^2$ in conjugate coordinates grows without bound. At the same time, the signal coefficient $a_i(t)$ tends to zero, and the conditional SNR, in which $a_i(t)$ cancels, equals $1/q(t)^2$ and likewise tends to zero. Thus, although the noise

variance remains bounded in graph coordinates, the forward process still removes dependence on the initial signal in the large-$q(t)$ limit. Section 3.2 instead uses a finite terminal noise level so that data-dependent spectral structure can remain in the terminal reference. The proof of Proposition 1 and the derivation of Corollary 1 are provided in Appendix A.1.1.

3.2 Gaussian reference and finite-SNR terminal initialization

We fit a zero-mean Gaussian reference using only the training split. Let $\mathbf{\Sigma}_{\mathrm{LW}}$ denote the Ledoit–Wolf shrinkage covariance (Ledoit & Wolf, 2004). We transform this covariance to the graph-Fourier basis and retain only its diagonal entries, discarding cross-mode covariances:

$$v_i \;=\; \left[\boldsymbol{U}^\top \left(\mathbf{\Sigma}_{\mathrm{LW}} + \varepsilon_{\mathrm{ref}} \boldsymbol{I}\right) \boldsymbol{U}\right]_{ii}, \qquad \mathbf{\Sigma}_{\mathrm{ref}} \;=\; \boldsymbol{U}\,\mathrm{diag}(\boldsymbol{v})\boldsymbol{U}^\top. \tag{6}$$

Here, $\varepsilon_{\mathrm{ref}} > 0$ is a small numerical shift. Since $\mathbf{\Sigma}_{\mathrm{LW}}$ is positive semidefinite and $\boldsymbol{U}$ is orthonormal, $v_i = \left[\boldsymbol{U}^\top \mathbf{\Sigma}_{\mathrm{LW}} \boldsymbol{U}\right]_{ii} + \varepsilon_{\mathrm{ref}} \geq \varepsilon_{\mathrm{ref}}$, so every spectral variance is strictly positive and $\mathbf{\Sigma}_{\mathrm{ref}}$ is non-singular. Thus, $\mathbf{\Sigma}_{\mathrm{ref}}$ defines the Gaussian reference distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathrm{ref}})$. Further details of the spectral diagonal approximation are given in Appendix A.2.

Because $\boldsymbol{A}_t$ and $\mathbf{\Sigma}_{\mathrm{ref}}$ are diagonal in the same graph-Fourier basis, the reference propagates analytically under the conjugate forward process. Its variance in mode i is

$$\gamma_i(t) \;=\; a_i(t)^2 \left(v_i + q(t)^2\right) \;=\; \frac{\sigma^2\left(v_i + q(t)^2\right)}{\sigma^2 + \mu_i q(t)^2}. \tag{7}$$

Therefore, we get the exact marginal of the chosen Gaussian reference distribution

$$p_{\mathrm{ref},t} \;=\; \mathcal{N}\big(\mathbf{0},\, \boldsymbol{U}\,\mathrm{diag}(\gamma_1(t), \ldots, \gamma_N(t))\,\boldsymbol{U}^\top\big)\,.$$

The forward kernel itself does not depend on the fitted spectral variances v_i. By Proposition 1 mode i has signal coefficient $a_i(t)$ and conditional variance $a_i(t)^2 q(t)^2$, so

$$\mathrm{SNR}_i^{\mathrm{cond}}(t) \;=\; \frac{a_i(t)^2}{a_i(t)^2 q(t)^2} \;=\; \frac{1}{q(t)^2} \tag{8}$$

is the same for every graph-Fourier mode; this is the SNR definition of Kingma et al. (2021) applied modewise. The coefficient $a_i(t)$ cancels, so the mode-dependent clock equalizes conditional corruption across graph frequencies.

A second, reference-level signal-to-noise ratio describes the fitted reference rather than the forward kernel: it measures how much fitted data-dependent signal remains relative to the injected noise, and therefore depends on the data. Under the reference, the propagated signal and injected-noise variances in graph-Fourier mode i are $a_i(t)^2 v_i$ and $a_i(t)^2 q(t)^2$, respectively, so the reference SNR in mode i at time t is given by

$$\mathrm{SNR}_i^{\mathrm{ref}}(t) \;=\; \frac{a_i(t)^2 v_i}{a_i(t)^2 q(t)^2} \;=\; \frac{v_i}{q(t)^2}. \tag{9}$$

For reporting a single terminal value we use the energy-weighted average $\overline{\mathrm{SNR}}^{\mathrm{ref}}(t) = \sum_{i=1}^{N} \frac{v_i}{\sum_{j=1}^{N} v_j} \mathrm{SNR}_i^{\mathrm{ref}}(t)$, so that modes carrying little signal energy cannot dominate the summary. Therefore, as the terminal noise level $q(t_{\max}) = \kappa t_{\max}$ grows, $\mathrm{SNR}_i^{\mathrm{ref}}(t_{\max}) \to 0$, and we consequently have

$$\gamma_i(t_{\max}) \;=\; \sigma^2 \frac{v_i/q(t_{\max})^2 + 1}{\sigma^2/q(t_{\max})^2 + \mu_i} \quad \xrightarrow{q(t_{\max})\to\infty} \quad \frac{\sigma^2}{\mu_i}. \tag{10}$$

Thus, since $\mu_i > 0$ holds for every mode, dependence on the fitted spectral variance v_i vanishes in the limit, leaving the limiting variance σ^2/μ_i.

3.3 Residual score parameterization

The propagated Gaussian reference distribution has the following graph-Fourier score

$$s_{\mathrm{ref},i}(\boldsymbol{x}, t) \;=\; -\frac{x_i}{\gamma_i(t)}, \qquad x_i \;=\; [\boldsymbol{U}^\top \boldsymbol{x}]_i.$$

Let $s_i^*(\boldsymbol{x}, t) := \left[\boldsymbol{U}^\top \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})\right]_i$ denote the unknown marginal data score in mode i. We define the residual r_i^* as $r_i^*(\boldsymbol{x}, t) := s_i^*(\boldsymbol{x}, t) - s_{\mathrm{ref},i}(\boldsymbol{x}, t)$. We parameterize this residual with a neural network while keeping the Gaussian reference score in closed form.

Let $\eta_i(t) = q(t)a_i(t)$ denote the conditional standard deviation in mode i. Using the definition of $a_i(t)$, we get $\eta_i(t) = q(t)\left(1 + \frac{\mu_i q(t)^2}{\sigma^2}\right)^{-1/2} = \frac{\sigma q(t)}{\sqrt{\sigma^2 + \mu_i q(t)^2}}$. For $q(t) \geq 0$, $\eta_i(t)$ increases with $q(t)$ and approaches the finite positive limit $\eta_i(t) \longrightarrow \frac{\sigma}{\sqrt{\mu_i}}$ as $q(t) \to \infty$, where $\mu_i > 0$ is ensured by the shift $\delta > 0$. Thus, the residual scaling $1/\eta_i(t)$ remains finite in the large-noise limit.

Let $f_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ denote the network output in the node basis and $f_{\boldsymbol{\theta},i}(\boldsymbol{x}, t) = [\boldsymbol{U}^\top f_{\boldsymbol{\theta}}(\boldsymbol{x}, t)]_i$ its i-th graph-Fourier coefficient. We parameterize the residual as $r_{\boldsymbol{\theta},i}(\boldsymbol{x}, t) = -\frac{f_{\boldsymbol{\theta},i}(\boldsymbol{x},t)}{\eta_i(t)}$. The reconstructed score is then given as

$$s_{\boldsymbol{\theta},i}(\boldsymbol{x}, t) = -\frac{x_i}{\gamma_i(t)} - \frac{f_{\boldsymbol{\theta},i}(\boldsymbol{x}, t)}{\eta_i(t)}. \tag{11}$$

For the forward corruption of graph signals induced by the conjugate construction, the corresponding noise-prediction target is $f_i^* = \epsilon_i - \eta_i(t)\frac{x_{t,i}}{\gamma_i(t)}$. We train with the mean-squared denoising objective (Vincent, 2011; Ho et al., 2020)

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\epsilon}}\left[\frac{1}{N}\sum_{i=1}^{N}\left(f_{\boldsymbol{\theta},i}(\mathbf{x}_t, t) - f_i^*\right)^2\right], \tag{12}$$

where $t \sim \mathcal{U}[t_{\min}, t_{\max}]$. Appendix A.3 shows that the population minimizer of Equation 12, together with Equation 11, recovers the marginal score $s_i^*(\boldsymbol{x}, t)$ exactly. Thus, the residual parameterization changes the denoising target's representation rather than approximating the marginal score.

3.4 Exact Gaussian propagation and the exponential-residual sampler

For the modewise SDE in Equation 1, define $b_i(t) := -\mu_i c_i(t)$ and $g_i(t)^2 := 2\sigma^2 c_i(t)$. Substituting Equation 11 into the probability-flow ODE (Song et al., 2021) gives

$$\frac{dx_i}{dt} = \underbrace{\left[b_i(t) + \frac{g_i(t)^2}{2\gamma_i(t)}\right]x_i}_{\text{Gaussian reference}} \underbrace{- \frac{1}{2}g_i(t)^2 r_{\boldsymbol{\theta},i}(\boldsymbol{x}, t)}_{\text{learned residual}}. \tag{13}$$

The Gaussian component has the exact modewise propagator mapping a state at time τ to time t via

$$\phi_i(t, \tau) = \sqrt{\frac{\gamma_i(t)}{\gamma_i(\tau)}}. \tag{14}$$

Its derivation and composition property are given in Appendix A.4.

For a backward step $t_1 < t_0$ variation of constants gives

$$x_i(t_1) = \phi_i(t_1, t_0)x_i(t_0) + \int_{t_0}^{t_1} \phi_i(t_1, \tau)\left[-\frac{1}{2}g_i(\tau)^2\right] r_{\boldsymbol{\theta},i}(\boldsymbol{x}(\tau), \tau)\, d\tau. \tag{15}$$

Thus, the Gaussian component is propagated exactly, while only the learned residual is approximated numerically. The sampler uses two network evaluations per step and implementation details are given in Appendix A.5. This is related to the general strategy of diffusion-specific exponential integrators, which propagate analytically tractable components exactly while approximating the learned contribution numerically (Lu et al., 2022; Zhang & Chen, 2023). Here, the exact propagator is mode dependent and is induced by the fitted Gaussian reference.

Time grid. For the even NFE budgets used in our experiments each sampling step uses two network evaluations, so $K = \mathrm{NFE}/2$. Following the power-law noise discretization of Karras et al. (2022), applied here to the conjugate noise scale q with $q_{\min} = q(t_{\min})$ and $q_{\max} = q(t_{\max})$, we use

$$q_j = \left[q_{\max}^{1/\rho} + \frac{j}{K}\left(q_{\min}^{1/\rho} - q_{\max}^{1/\rho}\right)\right]^{\rho}, \qquad j = 0, \ldots, K. \tag{16}$$

We then set $t_j = q_j/\kappa$. Because $q(t) = \kappa t$ is linear, $\rho = 1$ recovers the uniform grid used by the base exponential-residual sampler. We select a single ρ for all budgets by mean validation aMMD, with the uniform grid $\rho = 1$ as the baseline. This yields $\rho = 3$, which is fixed before any test evaluation. Unless a row is explicitly labelled $\rho = 1$, all reported exponential-residual results use $\rho = 3$ at every budget. Table 5 reports the corresponding test aMMD.

Heun sampler. "Heun" denotes the deterministic second-order predictor–corrector solver (Ascher & Petzold, 1998) for the probability-flow ODE. On the uniform-in-t grid used for all reported Heun results, each step evaluates the velocity at the current state, takes an Euler predictor step, and then evaluates the velocity again at the predicted state at the next time point. The two velocities are averaged for the corrected update. Thus, each step requires two network evaluations, so $\mathrm{NFE} = 2k$ corresponds to k sampling steps. The exponential-residual solver likewise uses two evaluations per step, so the two solvers are compared at equal NFE. EDM and the three preconditioning controls instead follow the EDM implementation, which omits the final correction (Appendix A.6.5).

4 Experimental setup

We evaluate five graph settings spanning three dataset families: METR-LA traffic speeds (Li et al., 2018), Molene temperature signals (Girault, 2015), and three synthetic stochastic block model (SBM) regimes with spectral concentration $c \in \{1, 4, 16\}$. The directed METR-LA adjacency matrix is symmetrized before constructing the graph Laplacian. We evaluate two terminal-noise regimes: $\kappa = 2$ and $\kappa \approx 32.6$ (targeting near-zero terminal SNR). All graph methods use the same backbone, optimization, batch size, random streams, and evaluation. In the central fitted-versus-scalar comparison the score model is fixed and only the Gaussian terminal initialization changes. Results are reported at the validation-converged stage, averaged over three random seeds.

Data generation quality is measured by aMMD (Gretton et al., 2012), the mean of the MMDs of quadratic variation, spectral centroid and degree correlation, based on the statistics and multi-bandwidth RBF estimator of Rozada et al. (2026) on normalized signals. For a common evaluation operator, graph-based statistics are computed using the combinatorial Laplacian $\boldsymbol{D} - \boldsymbol{W}$, fixed across all methods.

We compare our proposed GRCD method against seven comparators: Elucidating Diffusion Models (EDM) (Karras et al., 2022), two Whitened Score Diffusion (WSD)-style graph adaptations (Alido et al., 2025), and GAD (Rozada et al., 2026) under the matched graph protocol. GAD uses our matched protocol, so its reported aMMD values are not a reproduction of those in Rozada et al. (2026). The other three are preconditioning controls: isotropic, static graph, and dynamic conjugate preconditioning. Details are in Appendix A.6.

5 Results

5.1 Overall comparison

Table 1 shows the comparison on METR-LA. At NFE 4 GRCD reaches an aMMD of 0.0538. Among external baselines EDM reaches an aMMD of 2.7757, while the strongest method at this budget, the matched graph-prior WSD-style variant, reaches 1.2431, which is still $23\times$ higher than GRCD. The gap narrows with sampling budget: at NFE 64 the strongest external baseline or preconditioning control, i.e., conjugate preconditioning, reaches an aMMD of 0.0906, while GRCD without the residual score reaches an aMMD of 0.0499. The practical advantage of our proposed method is therefore largest in the low-NFE regime. Sampling cost is reported separately in Section 5.5, where quality is plotted against measured wall-clock time.

GAD's results reflect a sampler run far outside its intended regime: at NFE 4, 96.5% of its generated readings fall outside the 0–120 mph range a loop detector can report. As a native-budget diagnostic we also ran it at the authors' own 1000 and 3000 steps, where it produces no out-of-range readings across three seeds; its aMMD there is 1.53 ± 0.63 and 1.81 ± 0.63, respectively, so the sampler is stable at those budgets but the variance across random seeds remains large. Appendix A.7 repeats this comparison on the other four settings. GRCD is best at NFE 4 on all of them by a factor of 22–$36\times$, while the isotropic and static preconditioning controls and EDM overtake it at NFE 64 on

Table 1: **Matched comparison on METR-LA.** aMMD at the validation-converged stage and a fixed nominal NFE budget (Appendix A.6.5). Lower numbers are better and bold denotes the best mean in each column. Results are mean ± sample standard deviation over three random seeds. The last two rows use the exponential-residual solver.

METHOD	NFE 4	NFE 8	NFE 16	NFE 64
EDM	2.7757 ± 0.0040	1.3996 ± 0.0605	0.2640 ± 0.0139	0.1155 ± 0.0053
GAD	2.9091 ± 0.0016	2.9093 ± 0.0056	2.8878 ± 0.0146	2.8785 ± 0.0163
WSD-style, scalar	1.8752 ± 0.2182	0.3941 ± 0.1355	0.2695 ± 0.0587	0.2572 ± 0.0183
WSD-style, graph prior	1.2431 ± 0.2440	0.2410 ± 0.1347	0.1980 ± 0.0928	0.0966 ± 0.0325
Isotropic VE	2.7581 ± 0.0084	0.7810 ± 0.0240	0.1816 ± 0.0139	0.1152 ± 0.0087
Static precond.	2.7728 ± 0.0088	0.7993 ± 0.0155	0.1755 ± 0.0070	0.1134 ± 0.0043
Conjugate precond.	2.7581 ± 0.0198	0.6867 ± 0.0344	0.1385 ± 0.0025	0.0906 ± 0.0016
GRCD w/o residual score	0.1463 ± 0.0083	0.0603 ± 0.0029	0.0513 ± 0.0021	$\mathbf{0.0499 \pm 0.0018}$
GRCD, ρ=1	0.5298 ± 0.1303	0.0519 ± 0.0057	$\mathbf{0.0499 \pm 0.0032}$	0.0527 ± 0.0032
GRCD, ρ=3	$\mathbf{0.0538 \pm 0.0037}$	$\mathbf{0.0503 \pm 0.0028}$	0.0509 ± 0.0015	0.0529 ± 0.0028

the three SBM settings, as does the conjugate preconditioner on SBM c=16. Because the evaluation statistics are second order, we also compare GRCD with four non-learned Gaussian controls drawn directly from covariances fitted on the training split. The best control consistently has a higher aMMD value than GRCD at NFE 4 on all five settings (cf. Appendix A.8.1).

5.2 Terminal-reference effects across SNR regimes

Section 3.2 predicts that a fitted terminal reference matters only while the propagated terminal reference retains data-dependent spectral variance. As terminal SNR approaches zero, this dependence vanishes by Equation 10. NFE 16 is used for all mechanism comparisons in this paper: it is the smallest budget at which the Heun sampler shared by both arms has converged, so the contrast measures the terminal reference rather than discretization error.

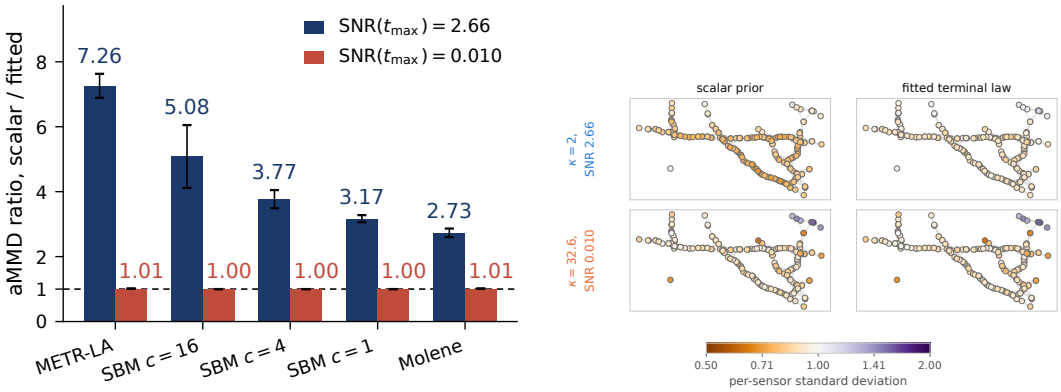


Figure 1: **Terminal-reference effects are gated by terminal SNR. Left:** Seed-paired scalar-to-fitted aMMD ratio at NFE 16 across five graph settings. The dashed rule at 1 marks no effect and values above it favor the fitted reference. **Right:** Per-sensor standard deviation on METR-LA for one of the three random seeds with each dot placed at its sensor's longitude and latitude. Color is the generated standard deviation relative to the correct value of 1.0. Lower color contrast is better.

Figure 1 confirms this prediction empirically. At $\kappa = 2$ (i.e., terminal SNR 2.66 on METR-LA) the fitted reference improves aMMD by 2.7×–7.3× across the five graph settings. At $\kappa = 32.6$ (i.e., terminal SNR 0.010 on METR-LA) the ratios collapse to 1.00–1.01×. Generated dispersion follows this pattern: on METR-LA the mean per-sensor standard deviation changes from 0.795 to 0.913 at finite SNR, but only from 0.916 to 0.918 at near-zero SNR. Thus, the benefit of fitting the terminal reference vanishes as the propagated reference reaches near-zero SNR.

Table 2: **Ablation ladder on METR-LA.** aMMD at $\kappa = 2$, validation-converged stage, mean $\pm$ s.d. over three random seeds. Each row adds one component to the row above. Lower numbers are better and bold denotes the best mean in each column. All rows use the uniform grid in t. The first and third rows are ablation-only configurations.

CONFIGURATION	NFE 8	NFE 12	NFE 16	NFE 64
Mode-dependent clock	0.4205 ± 0.0085	0.3817 ± 0.0067	0.3722 ± 0.0069	0.3632 ± 0.0076
+ fitted terminal ref.	0.0603 ± 0.0029	0.0530 ± 0.0023	0.0513 ± 0.0021	$\mathbf{0.0499 \pm 0.0018}$
+ residual score	0.0622 ± 0.0028	0.0549 ± 0.0034	0.0537 ± 0.0034	0.0531 ± 0.0033
+ exp. solver	$\mathbf{0.0519 \pm 0.0057}$	$\mathbf{0.0493 \pm 0.0036}$	$\mathbf{0.0499 \pm 0.0032}$	0.0527 ± 0.0032

5.3 Component contributions in GRCD

Table 2 summarizes an ablation study that demonstrates the impact of GRCD's main components sequentially. The dominant quality improvement comes from fitting the terminal reference: at NFE 16, aMMD drops from 0.3722 to 0.0513, which is a $7.25\times$ reduction.

The fitted terminal reference provides the main quality gain. The residual-score parameterization is structural rather than a direct improvement: at NFE 16 it changes aMMD from 0.0513 to 0.0537, but it separates the analytic Gaussian score from the learned residual and thereby enables the propagator in Equation 14. The exponential-residual solver then shifts comparable quality toward smaller NFE.

5.4 Solver and grid comparison

The previous ablation study separates the fitted terminal reference from the numerical solver. We therefore compare reverse integrators seed by seed with all solvers following identical evaluation pipelines. NFE is counted from actual network evaluations, and every configuration below attains its requested budget exactly. Heun and our exponential-residual solver use two evaluations per step, whereas DPM-Solver++ (Lu et al., 2025) multistep uses one.

For DPM-Solver++ we report all four tested configurations: $\{$2M, 3M$\} \times \{$uniform half-log-SNR grid, $\rho = 3$ grid$\}$. Our solver uses only the validation-selected $\rho = 3$ grid. DPM-Solver++ remains applicable because, by Equation 8, the conditional SNR is identical across all graph-Fourier modes. Its half-log-SNR coordinate $\ell(t) = \frac{1}{2} \log \mathrm{SNR}_i^{\mathrm{cond}}(t) = -\log q(t)$ is therefore independent of i: the shrinkage factor $a_i(t)$ cancels between signal and conditional noise scales, so step size and extrapolation weights remain scalar. This quantity is distinct from the mode-dependent reference terminal SNR in Equation 9. The graph-frequency dependence enters the update only through $a_i(t)$.

The exponential-residual solver is competitive with DPM-Solver++ but does not outperform it consistently. The strongest DPM-Solver++ configuration is 2M with the $\rho = 3$ grid at every tested budget. That configuration beats ours at NFE 4, while ours has lower means at NFE 8, 16, and 32. The differences at larger budgets are small in absolute terms. At NFE 16 and 32 the mean differences from the best DPM-Solver++ configuration are 0.0010 and 0.0003. The sampling grid has a larger effect at low NFE. Replacing the uniform half-log-SNR grid with the validation-selected $\rho = 3$ grid improves DPM-Solver++ at NFE 4 from 0.0659 to 0.0523 for 2M and from 0.0750 to 0.0708 for 3M. For 2M, this improvement (0.0136) is over four times the difference between the best DPM-Solver++ configuration and our solver at the same budget. At NFE 32 the solver differences (≤ 0.0005) are an order of magnitude smaller than the shared-seed variation (≈ 0.003).

5.5 Wall-clock efficiency

NFE counts function evaluations, but methods with the same NFE can have different sampling times because non-network computations differ. Figure 2 therefore compares sample quality against measured wall-clock time on the same machine. GRCD with the fitted terminal reference, residual score parameterization, exponential-residual solver, and validation-selected $\rho = 3$ grid generates 1000 samples in 0.5229 seconds at NFE 4 and attains aMMD 0.0538 there. No comparison method is close at that cost: the best at NFE 4 reaches 1.2431, which is $23\times$ worse. The comparison methods become competitive only at NFE 32–64, where sampling costs 4.01 (EDM at NFE 32) to 8.27 (WSD

Table 3: **Solver comparison.** aMMD on METR-LA, mean ± s.d. over three random seeds. Lower numbers are better and **bold** marks the best mean at each NFE. DPM denotes DPM-Solver++.

SOLVER	NFE 4	NFE 8	NFE 16	NFE 32
Heun	0.1608 ± 0.0205	0.0606 ± 0.0014	0.0514 ± 0.0027	0.0523 ± 0.0029
DPM 2M, uniform ℓ	0.0659 ± 0.0028	0.0523 ± 0.0022	0.0509 ± 0.0022	0.0522 ± 0.0029
DPM 2M, ρ=3	$\mathbf{0.0523 \pm 0.0010}$	0.0516 ± 0.0020	0.0507 ± 0.0026	0.0522 ± 0.0029
DPM 3M, uniform ℓ	0.0750 ± 0.0023	0.0552 ± 0.0021	0.0516 ± 0.0021	0.0524 ± 0.0029
DPM 3M, ρ=3	0.0708 ± 0.0023	0.0539 ± 0.0022	0.0510 ± 0.0023	0.0523 ± 0.0029
Exp. residual (ρ=3, ours)	0.0552 ± 0.0024	$\mathbf{0.0488 \pm 0.0028}$	$\mathbf{0.0497 \pm 0.0020}$	$\mathbf{0.0519 \pm 0.0029}$

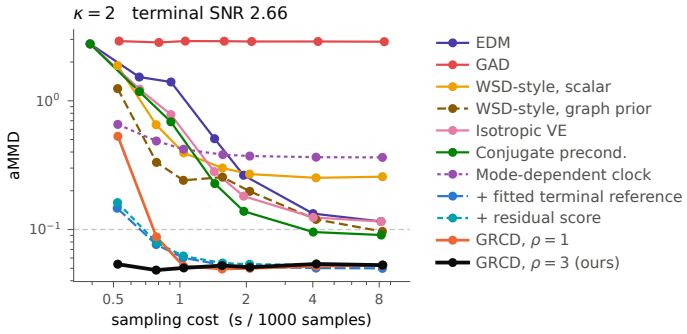


Figure 2: **Wall-clock sampling efficiency on METR-LA.** aMMD versus measured seconds per 1000 generated samples. Lower and left are better.

at NFE 64) seconds per 1000 samples, 7.7×–15.8× more than GRCD's NFE-4 point. The main advantage is therefore that GRCD reaches useful sample quality much earlier on the wall-clock curve. The exponential-residual sampler uses two function evaluations per reverse step, the same as Heun, so both pay the same price per step. Its wall-clock advantage comes from taking fewer steps: the analytic propagator absorbs the Gaussian part of each step, leaving the learned residual as the only piece integrated numerically. GAD's curve is nearly flat because aMMD has saturated there. Its samples continue to change: between NFE 4 and NFE 64 the median per-sensor dispersion ratio falls from 166 to 15, an elevenfold drop, while aMMD moves from 2.9091 to 2.8785, so the metric registers only a small fraction of the change. We read the flat segment as a metric limit at large aMMD. Appendix A.8.2 reports the cheapest measured cost at which each configuration reaches a fixed aMMD target. At the 0.1 target GRCD needs 0.52 seconds against 4.02 for the cheapest external baseline or preconditioning control that reaches it, a reduction of $1 - 0.52/4.02 \approx 87\%$.

6 Conclusion

We study whether diffusion on structured signals needs to reach near-zero terminal SNR when a finite-SNR terminal reference can be estimated from the data. GRCD uses a mode-dependent graph diffusion, fits a graph-spectral Gaussian reference from the training set and lets a neural network learn only the residual score. At finite terminal SNR, the fitted reference improves generation because it preserves data-dependent spectral structure, a benefit that disappears as the terminal SNR approaches zero. The same Gaussian reference also gives an exact modewise propagator, which enables accurate sampling with few function evaluations.

Limitations. Our construction assumes a fixed symmetric graph operator and loses directed structure, while the Gaussian reference models only second-order dependencies and permits invalid values for bounded signals. Performance gains peak in low-budget regimes, and alternative controls can outperform GRCD at higher NFE. For non-Gaussian data, a residual terminal-distribution mismatch may persist. Furthermore, the focus on unconditional generation leaves extensions to conditional tasks and higher-dimensional domains open.

AI USE STATEMENT

In this work, we used generative AI tools to improve the grammar and readability of the manuscript, identify potentially relevant literature, and assist with coding. We did not use generative AI tools for any other tasks requiring disclosure, and the remaining required-disclosure tasks were not applicable to this work. All AI-assisted content was manually reviewed and verified. We take responsibility for the final content of this work.

ETHICS STATEMENT

This work uses two public sensor datasets and one synthetic family. METR-LA (Li et al., 2018) contains loop-detector speed readings and Molene (Girault, 2015) contains weather-station temperatures. Both are aggregate physical measurements released for research and contain no personally identifying information. The stochastic block model settings are synthetic. This work was supported by a publicly funded research grant providing doctoral student funding. The funder had no role in the study design, data collection, analysis, or writing of the manuscript.

REPRODUCIBILITY STATEMENT

All details required for reproducing the experiments are provided in the Appendix. All preprocessing and reference estimation use training data only, grid selection uses validation data, and experiments use fixed seeds $0, 1, 2$ with matched random draws for paired comparisons. Code will be released upon acceptance.

REFERENCES

Jeffrey Alido, Tongyu Li, Yu Sun, and Lei Tian. Whitened score diffusion: A structured prior for imaging inverse problems. *Advances in Neural Information Processing Systems*, 38:65168–65195, 2025.

Uri M Ascher and Linda R Petzold. *Computer Methods for Ordinary Differential Equations and Differential-algebraic Equations*. SIAM, 1998.

Sang gil Lee, Heeseung Kim, Chaehun Shin, Xu Tan, Chang Liu, Qi Meng, Tao Qin, Wei Chen, Sungroh Yoon, and Tie-Yan Liu. Priorgrad: Improving conditional denoising diffusion models with data-dependent adaptive prior. In *International Conference on Learning Representations*, 2022.

Benjamin Girault. Stationary graph signals using an isometric graph translation. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pp. 1516–1520. IEEE, 2015.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35:26565–26577, 2022.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in Neural Information Processing Systems*, 34:21696–21707, 2021.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*, 2018.

Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common diffusion noise schedules and sample steps are flawed. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5392–5399. IEEE, 2024.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, 22: 730–751, 2025.

Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst. Graph signal processing: Overview, challenges and applications. *Proceedings of the IEEE*, 106 (5):808–828, 2018.

Benjamin Remy, Francois Lanusse, Niall Jeffrey, Jia Liu, J-L Starck, Ken Osato, and Tim Schrabback. Probabilistic mass-mapping with neural score estimation. *Astronomy & Astrophysics*, 672: A51, 2023.

Sergio Rozada, Vimal K B, Andrea Cavallo, Antonio G Marques, Hadi Jamali-Rad, and Elvin Isufi. Graph-aware diffusion for signal generation. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 461–465, 2026.

David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

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.

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

Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

Binxu Wang and John Vastola. The unreasonable effectiveness of gaussian score approximation for diffusion models and its applications. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. In *The Eleventh International Conference on Learning Representations*, 2023.

A Appendix

A.1 Derivations for conjugate graph diffusion

A.1.1 Derivation of mode equalization and the mode-dependent clock

Throughout this section, write

$$\boldsymbol{L}_\delta = \boldsymbol{U}\,\mathrm{diag}(\mu_1,\dots,\mu_N)\boldsymbol{U}^\top.$$

We use the same convention as in the main text: $\mathbf{x}_t$ denotes the graph-space process and $\widetilde{\mathbf{x}}_t$ its conjugate-coordinate representation. A subscript i denotes the i-th graph-Fourier coefficient of the corresponding vector, e.g., $x_{t,i} := [\boldsymbol{U}^\top \mathbf{x}_t]_i$ and $\widetilde{x}_{t,i} := [\boldsymbol{U}^\top \widetilde{\mathbf{x}}_t]_i$. Likewise, $\epsilon_i := [\boldsymbol{U}^\top \boldsymbol{\epsilon}]_i$.

We derive the shrinkage operator and the mode-dependent clock from the mode-equalization construction in Proposition 1. Consider one graph-Fourier mode with $\mu_i > 0$. We seek a linear graph-diffusion process of the form

$$dx_{t,i} = -\mu_i c_i(t) x_{t,i}\,dt + \sqrt{2\sigma^2 c_i(t)}\,dW_{t,i}, \tag{17}$$

with $c_i(t) \geq 0$, such that the change of coordinates

$$\widetilde{x}_{t,i} = \frac{x_{t,i}}{a_i(t)} \tag{18}$$

follows isotropic VE diffusion with noise amplitude $q(t)$:

$$d\widetilde{x}_{t,i} = \sqrt{2q(t)q'(t)}\, dW_{t,i}. \tag{19}$$

Here $a_i(0) = 1$ and $q(0) = 0$.

Writing

$$x_{t,i} = a_i(t)\widetilde{x}_{t,i}$$

and applying the product rule gives

$$dx_{t,i} = \frac{a_i'(t)}{a_i(t)} x_{t,i}\, dt + a_i(t)\sqrt{2q(t)q'(t)}\, dW_{t,i}. \tag{20}$$

Matching the drift and diffusion coefficients in equation 17 and equation 20 yields

$$\frac{a_i'(t)}{a_i(t)} = -\mu_i c_i(t), \qquad \sigma^2 c_i(t) = q(t)q'(t)a_i(t)^2. \tag{21}$$

Substituting the second identity into the first gives

$$\frac{a_i'(t)}{a_i(t)} = -\frac{\mu_i q(t)q'(t)}{\sigma^2} a_i(t)^2. \tag{22}$$

Therefore

$$\frac{d}{dt} a_i(t)^{-2} = \frac{2\mu_i q(t)q'(t)}{\sigma^2} = \frac{\mu_i}{\sigma^2}\frac{d}{dt} q(t)^2. \tag{23}$$

Integrating from 0 to t and using $a_i(0) = 1$ and $q(0) = 0$ gives

$$a_i(t)^{-2} = 1 + \frac{\mu_i q(t)^2}{\sigma^2}, \tag{24}$$

hence

$$a_i(t) = \left(1 + \frac{\mu_i q(t)^2}{\sigma^2}\right)^{-1/2}. \tag{25}$$

Since

$$\boldsymbol{L}_\delta = \boldsymbol{U}\, \mathrm{diag}(\mu_1, \ldots, \mu_N)\boldsymbol{U}^\top,$$

spectral functional calculus gives

$$\boldsymbol{A}_t = \boldsymbol{U}\, \mathrm{diag}(a_1(t), \ldots, a_N(t))\, \boldsymbol{U}^\top = \left(\boldsymbol{I} + \frac{q(t)^2}{\sigma^2}\boldsymbol{L}_\delta\right)^{-1/2}, \tag{26}$$

which is exactly the shrinkage operator in equation 3.

Finally, substituting equation 25 into the diffusion coefficient identity in equation 21 gives

$$c_i(t) = \frac{q(t)q'(t)}{\sigma^2 + \mu_i q(t)^2}, \tag{27}$$

which is the mode-dependent clock in equation 5.

Modewise conditional marginals. Equation equation 4 acts in graph-Fourier mode i as

$$x_{t,i} = a_i(t)\,(x_{0,i} + q(t)\epsilon_i)\,, \qquad \epsilon_i \sim \mathcal{N}(0,1). \tag{28}$$

Since

$$\widetilde{x}_{t,i} = a_i(t)^{-1} x_{t,i},$$

the conjugate coordinates satisfy

$$\widetilde{x}_{t,i} = x_{0,i} + q(t)\epsilon_i.$$

Because $\boldsymbol{A}_0 = \boldsymbol{I}$, we also have $\widetilde{\mathbf{x}}_0 = \mathbf{x}_0$. Therefore

$$\mathrm{Var}(\widetilde{x}_{t,i} \mid \widetilde{x}_{0,i}) = q(t)^2 \qquad \text{for every mode } i. \tag{29}$$

Uniqueness within the linear graph-diffusion family. For a prescribed differentiable nondecreasing $q(t)$ with $q(0) = 0$, coefficient matching in equation 21, together with $a_i(0) = 1$, uniquely determines both $a_i(t)$ and $c_i(t)$. Equivalently, the integrated mode-dependent clock is

$$C_i(t) := \int_0^t c_i(u)\, du = \frac{1}{2\mu_i} \log\left(1 + \frac{\mu_i q(t)^2}{\sigma^2}\right). \tag{30}$$

Thus the mode dependence is induced by exact conjugacy to the prescribed isotropic VE corruption within this linear graph-diffusion family.

A.1.2 Why a shared scalar clock cannot exactly equalize modes

To isolate the role of the mode-dependent clock, consider instead a single shared scalar clock $c(t)$ for every graph-Fourier mode:

$$dx^{\mathrm{sh}}_{t,i} = -\mu_i c(t) x^{\mathrm{sh}}_{t,i}\, dt + \sqrt{2\sigma^2 c(t)}\, dW_{t,i}. \tag{31}$$

Let

$$C(t) = \int_0^t c(u)\, du. \tag{32}$$

The deterministic attenuation of mode i is

$$a^{\mathrm{sh}}_i(t) = \exp(-\mu_i C(t)),$$

and its conditional variance in the native graph-Fourier coordinates is

$$\mathrm{Var}\big(x^{\mathrm{sh}}_{t,i} \mid x^{\mathrm{sh}}_{0,i}\big) = \frac{\sigma^2}{\mu_i}\left[1 - \exp(-2\mu_i C(t))\right]. \tag{33}$$

Undoing the deterministic attenuation defines the transformed shared-clock coordinate

$$\widetilde{x}^{\mathrm{sh}}_{t,i} = \frac{x^{\mathrm{sh}}_{t,i}}{a^{\mathrm{sh}}_i(t)}. \tag{34}$$

Its conditional variance in these coordinates is

$$\mathrm{Var}\big(\widetilde{x}^{\mathrm{sh}}_{t,i} \mid \widetilde{x}^{\mathrm{sh}}_{0,i}\big) = \frac{\sigma^2}{\mu_i}\left[\exp(2\mu_i C(t)) - 1\right]. \tag{35}$$

Unlike equation 29, this expression depends explicitly on μ_i.

Matching the common target variance $q(t)^2$ in mode i would require

$$C^\star_i(t) = \frac{1}{2\mu_i} \log\left(1 + \frac{\mu_i q(t)^2}{\sigma^2}\right). \tag{36}$$

For fixed t with $q(t) > 0$, this quantity is strictly decreasing in μ_i. To see this, let

$$b = \frac{q(t)^2}{\sigma^2} > 0.$$

Then

$$\frac{d}{d\mu_i} C^\star_i(t) = \frac{\frac{b\mu_i}{1+b\mu_i} - \log(1 + b\mu_i)}{2\mu_i^2}. \tag{37}$$

For $x > 0$,

$$\log(1 + x) - \frac{x}{1 + x} = \int_0^x \frac{u}{(1 + u)^2}\, du > 0, \tag{38}$$

so the derivative is strictly negative. Hence, for two graph modes with distinct eigenvalues, a single shared scalar clock cannot realize the common conjugate variance $q(t)^2$ in both modes, which is why the construction requires a mode-dependent clock.

A.2 GAUSSIAN-REFERENCE DETAILS

A.2.1 SPECTRAL DIAGONAL APPROXIMATION

The reference used in the main text is obtained from the Ledoit–Wolf covariance $\mathbf{\Sigma}_{\mathrm{LW}}$ fitted on the training split. Before retaining only its diagonal, define the full covariance in the graph-Fourier basis as

$$\widetilde{\mathbf{\Sigma}} = \boldsymbol{U}^{\top} \left(\mathbf{\Sigma}_{\mathrm{LW}} + \varepsilon_{\mathrm{ref}} \boldsymbol{I}\right) \boldsymbol{U}. \tag{39}$$

The reference retains only its diagonal entries,

$$v_i = \widetilde{\mathbf{\Sigma}}_{ii}, \qquad \mathbf{\Sigma}_{\mathrm{ref}} = \boldsymbol{U} \operatorname{diag}(\boldsymbol{v}) \boldsymbol{U}^{\top}. \tag{40}$$

Thus, cross-mode covariances are discarded. The discarded off-diagonal energy, $\|\widetilde{\mathbf{\Sigma}} - \operatorname{diag}(\widetilde{\mathbf{\Sigma}})\|_F / \|\widetilde{\mathbf{\Sigma}}\|_F$, is 0.6677 on METR-LA, so the reference is a deliberately coarse covariance model whose remaining discrepancy is left to the learned residual.

A.2.2 EXACT PROPAGATED MARGINAL OF THE GAUSSIAN REFERENCE

Suppose, for this derivation only, that

$$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathrm{ref}}) \,. \tag{41}$$

In graph-Fourier mode i, equation 4 gives

$$x_{t,i} = a_i(t) \left(x_{0,i} + q(t)\epsilon_i\right). \tag{42}$$

Since

$$x_{0,i} \sim \mathcal{N}(0, v_i), \qquad \epsilon_i \sim \mathcal{N}(0,1),$$

and the two are independent,

$$\begin{aligned} \gamma_i(t) &= \operatorname{Var}(x_{t,i}) && (43)\\ &= a_i(t)^2 \left(v_i + q(t)^2\right) && (44)\\ &= \frac{\sigma^2 \left(v_i + q(t)^2\right)}{\sigma^2 + \mu_i q(t)^2}. && (45) \end{aligned}$$

Therefore

$$p_{\mathrm{ref},t} = \mathcal{N}\big(\mathbf{0},\, \boldsymbol{U} \operatorname{diag}(\gamma_1(t), \ldots, \gamma_N(t)) \, \boldsymbol{U}^{\top}\big) \,, \tag{46}$$

which proves equation 7 for the propagated Gaussian reference.

A.2.3 REFERENCE MARGINAL VERSUS THE TRUE DATA MARGINAL

The Gaussian calculation above does not imply that the true marginal p_t is Gaussian. For an arbitrary initial data distribution p_0, the conjugate marginal is

$$\widetilde{p}_t = p_0 * \mathcal{N}\big(\mathbf{0}, q(t)^2 \boldsymbol{I}\big) \,, \tag{47}$$

and the graph-space marginal is its pushforward under $\boldsymbol{A}_t$:

$$p_t = (\boldsymbol{A}_t)_{\#} \left[p_0 * \mathcal{N}\big(\mathbf{0}, q(t)^2 \boldsymbol{I}\big)\right]. \tag{48}$$

This distribution is generally non-Gaussian. Hence $p_{\mathrm{ref},t}$ is the exact propagated marginal of the chosen Gaussian reference, but is not identified with the true data marginal unless the initial data distribution actually equals the Gaussian reference.

A.3 RESIDUAL-SCORE TARGET DERIVATION

Let

$$\eta_i(t) = q(t) a_i(t) \tag{49}$$

denote the conditional standard deviation in graph-Fourier mode i. From equation 4,

$$x_{t,i} - a_i(t) x_{0,i} = \eta_i(t) \epsilon_i. \tag{50}$$

Hence the conditional distribution is Gaussian with score

$$\frac{\partial}{\partial x_{t,i}} \log p(x_{t,i} \mid x_{0,i}) = -\frac{x_{t,i} - a_i(t)x_{0,i}}{\eta_i(t)^2} \tag{51}$$

$$= -\frac{\epsilon_i}{\eta_i(t)}. \tag{52}$$

The propagated Gaussian reference has graph-Fourier score

$$s_{\mathrm{ref},i}(\boldsymbol{x}, t) = -\frac{x_i}{\gamma_i(t)}, \qquad x_i = [\boldsymbol{U}^\top \boldsymbol{x}]_i. \tag{53}$$

Let $s_i^*(\boldsymbol{x}, t)$ denote the true marginal score in mode i, and define

$$r_i^*(\boldsymbol{x}, t) = s_i^*(\boldsymbol{x}, t) - s_{\mathrm{ref},i}(\boldsymbol{x}, t). \tag{54}$$

Equivalently,

$$r_i^*(\boldsymbol{x}, t) = \left[\boldsymbol{U}^\top \nabla_{\boldsymbol{x}} \log \frac{p_t(\boldsymbol{x})}{p_{\mathrm{ref},t}(\boldsymbol{x})}\right]_i, \tag{55}$$

so the residual is the i-th graph-Fourier coefficient of the density-ratio score between the true marginal and the propagated Gaussian reference.

Let $f_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ denote the network output in the node basis and

$$f_{\boldsymbol{\theta},i}(\boldsymbol{x}, t) = [\boldsymbol{U}^\top f_{\boldsymbol{\theta}}(\boldsymbol{x}, t)]_i$$

its i-th graph-Fourier coefficient. We parameterize the residual as

$$r_{\boldsymbol{\theta},i}(\boldsymbol{x}, t) = -\frac{f_{\boldsymbol{\theta},i}(\boldsymbol{x}, t)}{\eta_i(t)}. \tag{56}$$

The reconstructed score is therefore

$$s_{\boldsymbol{\theta},i}(\boldsymbol{x}, t) = s_{\mathrm{ref},i}(\boldsymbol{x}, t) - \frac{f_{\boldsymbol{\theta},i}(\boldsymbol{x}, t)}{\eta_i(t)}. \tag{57}$$

Using the conditional score in equation 52 as the denoising target gives

$$s_{\mathrm{ref},i}(\mathbf{x}_t, t) - \frac{f_i^*}{\eta_i(t)} = -\frac{\epsilon_i}{\eta_i(t)}. \tag{58}$$

Multiplying by $\eta_i(t)$ yields

$$f_i^* = \epsilon_i + \eta_i(t) s_{\mathrm{ref},i}(\mathbf{x}_t, t) \tag{59}$$

$$= \epsilon_i - \eta_i(t)\frac{x_{t,i}}{\gamma_i(t)}. \tag{60}$$

This is the target used in equation 12.

Consistency with the marginal score. Under the squared loss in equation 12, the population minimizer is the conditional expectation of the target:

$$f_i^{\mathrm{opt}}(\boldsymbol{x}, t) = \mathbb{E}[f_i^* \mid \mathbf{x}_t = \boldsymbol{x}] \tag{61}$$

$$= \mathbb{E}[\epsilon_i \mid \mathbf{x}_t = \boldsymbol{x}] - \eta_i(t)\frac{x_i}{\gamma_i(t)}. \tag{62}$$

The marginal score satisfies

$$s_i^*(\boldsymbol{x}, t) = \mathbb{E}\big[\big[\boldsymbol{U}^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_0)\big]_i \;\big|\; \mathbf{x}_t = \boldsymbol{x}\big] = -\frac{\mathbb{E}[\epsilon_i \mid \mathbf{x}_t = \boldsymbol{x}]}{\eta_i(t)}. \tag{63}$$

Substituting equation 62 into equation 57 gives

$$s_{\boldsymbol{\theta},i}(\boldsymbol{x}, t) = s_{\mathrm{ref},i}(\boldsymbol{x}, t) - \frac{f_i^{\mathrm{opt}}(\boldsymbol{x}, t)}{\eta_i(t)} = -\frac{x_i}{\gamma_i(t)} - \frac{\mathbb{E}[\epsilon_i \mid \mathbf{x}_t = \boldsymbol{x}]}{\eta_i(t)} + \frac{x_i}{\gamma_i(t)} \tag{64}$$

$$= -\frac{\mathbb{E}[\epsilon_i \mid \mathbf{x}_t = \boldsymbol{x}]}{\eta_i(t)} \tag{65}$$

$$= s_i^*(\boldsymbol{x}, t). \tag{66}$$

Thus, at the population optimum, the residual parameterization exactly recovers the marginal score.

If the Gaussian reference score is disabled, $s_{\mathrm{ref},i} \equiv 0$, then

$$f_i^* = \epsilon_i, \tag{67}$$

recovering the corresponding standard ϵ-prediction objective.

Gaussian-reference exactness. If the data distribution equals the fitted Gaussian reference at $t = 0$, then

$$p_t = p_{\mathrm{ref},t} \tag{68}$$

for every t, and therefore

$$r_i^*(\boldsymbol{x}, t) = 0 \tag{69}$$

for every mode. In this special case, the Gaussian reference supplies the complete marginal score analytically and no learned residual is required.

A.4 Derivation of the exact Gaussian propagator

For the modewise forward SDE, define

$$b_i(t) = -\mu_i c_i(t), \qquad g_i(t)^2 = 2\sigma^2 c_i(t). \tag{70}$$

The probability-flow ODE is

$$\frac{dx_i}{dt} = b_i(t)x_i - \frac{1}{2}g_i(t)^2 s_{\boldsymbol{\theta},i}(\boldsymbol{x}, t). \tag{71}$$

Substituting

$$s_{\boldsymbol{\theta},i}(\boldsymbol{x}, t) = -\frac{x_i}{\gamma_i(t)} + r_{\boldsymbol{\theta},i}(\boldsymbol{x}, t)$$

gives

$$\frac{dx_i}{dt} = \left[b_i(t) + \frac{g_i(t)^2}{2\gamma_i(t)}\right] x_i - \frac{1}{2}g_i(t)^2 r_{\boldsymbol{\theta},i}(\boldsymbol{x}, t). \tag{72}$$

Because $\gamma_i(t)$ is the variance of the Gaussian reference propagated under the linear SDE, it satisfies the standard linear variance equation

$$\frac{d\gamma_i(t)}{dt} = 2b_i(t)\gamma_i(t) + g_i(t)^2. \tag{73}$$

Dividing by $2\gamma_i(t)$ gives

$$b_i(t) + \frac{g_i(t)^2}{2\gamma_i(t)} = \frac{1}{2}\frac{d}{dt}\log\gamma_i(t). \tag{74}$$

The exact propagator of the Gaussian component from time s to time t is therefore

$$\phi_i(t, s) = \exp\left(\int_s^t \left[b_i(u) + \frac{g_i(u)^2}{2\gamma_i(u)}\right] du\right) \tag{75}$$

$$= \exp\left(\frac{1}{2}\int_s^t \frac{d}{du}\log\gamma_i(u)\, du\right) \tag{76}$$

$$= \sqrt{\frac{\gamma_i(t)}{\gamma_i(s)}}. \tag{77}$$

This is exactly equation 14.

Substituting the closed-form reference variance from equation 7 gives

$$\phi_i(t, s) = \sqrt{\frac{v_i + q(t)^2}{v_i + q(s)^2}}\sqrt{\frac{\sigma^2 + \mu_i q(s)^2}{\sigma^2 + \mu_i q(t)^2}}. \tag{78}$$

The propagator has the exact composition property

$$\phi_i(t,u)\phi_i(u,s) = \sqrt{\frac{\gamma_i(t)}{\gamma_i(u)}}\sqrt{\frac{\gamma_i(u)}{\gamma_i(s)}} \tag{79}$$

$$= \sqrt{\frac{\gamma_i(t)}{\gamma_i(s)}} \tag{80}$$

$$= \phi_i(t,s). \tag{81}$$

Thus the linear Gaussian component is propagated exactly between arbitrary time points.

If $r_{\boldsymbol{\theta},i} \equiv 0$, then equation 72 reduces to the known linear Gaussian flow and

$$x_i(t) = \phi_i(t,s)x_i(s) \tag{82}$$

exactly. The exponential-residual sampler is therefore exact for the probability-flow dynamics in the zero-residual Gaussian-reference case.

A.5 Exponential-residual solver details

For a backward interval $t_1 < t_0$, write the residual integral in Equation 15 using

$$B_i(\tau;t_1) = -\frac{1}{2}\phi_i(t_1,\tau)g_i(\tau)^2, \qquad \alpha(\tau) = \frac{\tau - t_0}{t_1 - t_0}. \tag{83}$$

We approximate the learned residual linearly between its endpoint values,

$$r_{\boldsymbol{\theta},i}(\boldsymbol{x}(\tau),\tau) \approx (1-\alpha(\tau))r_{0,i} + \alpha(\tau)r_{1,i}, \tag{84}$$

with corresponding weights

$$w_{0,i} = \int_{t_0}^{t_1} B_i(\tau;t_1)(1-\alpha(\tau))\,d\tau, \qquad w_{1,i} = \int_{t_0}^{t_1} B_i(\tau;t_1)\alpha(\tau)\,d\tau. \tag{85}$$

At the beginning of the step,

$$r_{0,i} = r_{\boldsymbol{\theta},i}(\boldsymbol{x}(t_0),t_0).$$

Using $r_{0,i}$ at both endpoints gives the predictor

$$x_{i,\mathrm{pred}} = \phi_i(t_1,t_0)x_i(t_0) + (w_{0,i}+w_{1,i})r_{0,i}. \tag{86}$$

We then evaluate

$$r_{1,i} = r_{\boldsymbol{\theta},i}(\boldsymbol{x}_{\mathrm{pred}},t_1),$$

where $\boldsymbol{x}_{\mathrm{pred}}$ denotes the signal with graph-Fourier coefficients $x_{i,\mathrm{pred}}$, and apply the corrected update

$$x_i(t_1) = \phi_i(t_1,t_0)x_i(t_0) + w_{0,i}r_{0,i} + w_{1,i}r_{1,i}. \tag{87}$$

A.6 Experimental details

A.6.1 Datasets and splits

Table 4 summarizes the five graph-signal settings. The two real datasets use chronological splits, while the synthetic SBM datasets use independently generated train, validation and test sets. All normalization, covariance fitting and reference estimation use the training split only.

The chronological rule is $70/10/20$ on both real datasets. On Molene it gives $520/74/150$ signals after integer rounding. On METR-LA it gives $23{,}990/3{,}427/6{,}855$, which the caps listed in Table 6 then reduce to the sizes shown, by evenly spaced subsampling so that each split still spans its full time period. This is why the METR-LA row is not exactly $70/10/20$.

METR-LA. We use the adjacency matrix distributed with DCRNN (Li et al., 2018). The source graph is strongly directional: $1{,}111$ of its $1{,}313$ edges occur in one direction only and carry 75.0% of the total edge weight. We therefore symmetrize the graph by the half-sum for the common protocol. Repeating the exponential-residual arm and the conjugate preconditioning control under $\max(\boldsymbol{W},\boldsymbol{W}^\top)$ leaves the ordering unchanged: at NFE 16 the former moves from 0.0499 to 0.0536 and the latter from 0.1385 to 0.1295, so the gap between them is preserved under either symmetrizer.

Table 4: **Datasets.** N denotes the number of graph nodes. Split sizes are the numbers of graph signals used in each experiment.

SETTING	SOURCE	N	TRAIN	VAL	TEST
METR-LA	traffic, 5-min (Li et al., 2018)	207	20,000	3,000	5,000
Molene	temperature (Girault, 2015)	37	520	74	150
SBM (c=1)	synthetic	32	4,000	1,000	5,000
SBM (c=4)	synthetic	32	4,000	1,000	5,000
SBM (c=16)	synthetic	32	4,000	1,000	5,000

Molene. The dataset contains 744 hourly temperature measurements from 37 stations in Brittany. We build a 5-nearest-neighbor Gaussian graph from the station coordinates, with bandwidth equal to the median pairwise distance. The original Molene study does not define a train/validation/test split for generative modeling, so we apply the same chronological rule as for METR-LA. The resulting 150 test signals make this the smallest evaluation set in the study, so its aMMD estimates carry visibly larger seed-to-seed spread than the $5{,}000$-sample settings (Table 7).

SBM. We generate a connected two-block stochastic block model with $N = 32$, $p_{\text{in}} = 0.4$ and $p_{\text{out}} = 0.04$. Signals are sampled in the eigenbasis of its combinatorial Laplacian $\boldsymbol{L}_{\text{comb}}$, with spectral variance

$$v_0(\nu_i) = 0.2 + \frac{0.8}{1 + c\,\nu_i}, \qquad c \in \{1, 4, 16\},$$

together with a bimodal ± 3 shift in the second graph-Fourier mode, the Fiedler vector, which separates the two blocks. Here $\nu_i = \lambda_i(\boldsymbol{L}_{\text{comb}})/\lambda_{\max}(\boldsymbol{L}_{\text{comb}}) + \delta \in [\delta, 1 + \delta]$ are the scaled and shifted eigenvalues of $\boldsymbol{L}_{\text{comb}}$. The parameter c controls how sharply signal energy concentrates at low graph frequencies, while the bimodal component makes the data non-Gaussian, so that its distribution cannot be represented exactly by a covariance-matched Gaussian. The graph and the dataset are generated once and held fixed. The three experimental seeds vary only model initialization, batch order and sampling noise.

A.6.2 Training protocol

All methods in the common protocol use the `DenoiserGNN` architecture released with GAD (Rozada et al., 2026), with three $K = 5$ graph-filter blocks and a 64-dimensional time embedding. In our common protocol, we set the hidden width to 128. Training uses Adam with the hyperparameters in Table 6. Validation denoising error is evaluated every 250 updates using EMA weights, and training stops after eight consecutive evaluations without a relative improvement greater than 0.2%, subject to a maximum of $40{,}000$ updates. All methods therefore share one stopping criterion. No arm reaches the $40{,}000$-update cap.

For paired comparisons, configurations with the same experimental seed receive the same batches, forward-noise draws, and time draws at each update.

A.6.3 Noise regimes

GRCD and the WSD-style adaptations use

$$q(t) = \kappa t, \qquad t \in [0.02, 1].$$

The finite-SNR setting uses $\kappa = 2$, giving $q \in [0.04, 2.0]$ and an energy-weighted terminal SNR of 2.663 on METR-LA. The near-zero-SNR setting uses $\kappa = 32.6356$ on METR-LA, calibrated to terminal SNR 0.010, and $\kappa = 32.64$ on the other settings, where the resulting terminal SNR ranges from 0.003 to 0.027. On all five settings this is two to three orders of magnitude below the corresponding finite-SNR value.

The isotropic VE controls use

$$y_t = x_0 + q\boldsymbol{\epsilon}, \qquad q \in [0.02, 32.64],$$

with $P_{\text{mean}} = -1.2$, $P_{\text{std}} = 1.2$ during training and a $\rho_{\text{VE}} = 7$ power-law grid during sampling.

A.6.4 BASELINE IMPLEMENTATION

EDM (Karras et al., 2022) uses its published preconditioning, loss weighting, noise schedule, and second-order Heun sampler, with the shared graph backbone substituted for the image network. We use $P_{\text{mean}} = -1.2$, $P_{\text{std}} = 1.2$, $\sigma_{\min} = 0.002$, $\sigma_{\max} = 80$, and $\rho = 7$. GAD (Rozada et al., 2026) uses its published graph-aware process and native Euler–Maruyama sampler. Following the authors' released implementation, we use two Newton iterations to invert the heat-time map and a sampling grid descending to $s = 0$.

Trained instead at the authors' own configuration (3,985 parameters, 5,000 epochs, no early stopping) and sampled at their evaluation script's 5,000 steps, GAD (Rozada et al., 2026) reaches aMMD 2.10 ± 0.15 over three seeds, so it is not better under its own settings than under ours. That configuration costs 785,000 optimizer updates, against at most 25,000 for any arm in the matched protocol, and sampling at its native budget takes 261 seconds per 1000 samples against GRCD's 0.52.

For the WSD-style adaptations (Alido et al., 2025), we transfer the whitened-score parameterization to the conjugate forward process and evaluate both a scalar variance and the same Ledoit–Wolf graph-spectral covariance used by GRCD.

The isotropic, static-graph, and dynamic-conjugate preconditioning controls use the same isotropic VE corruption and deterministic Heun sampler, differing only in the preconditioner.

A.6.5 SAMPLER DETAILS

The exponential-residual solver propagates the Gaussian component analytically and evaluates the residual integral with composite Simpson quadrature on 65 nodes after linear interpolation of the residual in time. The quadrature adds no network evaluations, so each reverse step uses two evaluations. Following the reference EDM Heun implementation, a k-step trajectory uses $2k - 1$ network evaluations because the second-order correction is omitted on the final step to zero noise. On our nominal even-NFE grid, EDM and the three Heun preconditioning controls therefore use one evaluation fewer than the nominal budget. All other configurations use the nominal budget exactly.

Reverse steps use

$$q_j = \left(q_{\max}^{1/\rho} + \frac{j}{K} \left(q_{\min}^{1/\rho} - q_{\max}^{1/\rho} \right) \right)^{\rho}, \qquad j = 0, \ldots, K. \tag{88}$$

The base sampler uses $\rho = 1$. We select $\rho = 3$ once on the METR-LA validation split and use it unchanged for all sampling budgets and datasets. Table 5 reports the corresponding test aMMD.

Table 5: **Time-grid sweep for the exponential-residual solver on METR-LA.** Test aMMD at $\kappa = 2$, converged, mean $\pm$ standard deviation over three random seeds.

GRID	NFE 4	NFE 6	NFE 8	NFE 12	NFE 16
$\rho = 1$	0.5298 ± 0.1303	0.0876 ± 0.0257	0.0519 ± 0.0057	0.0493 ± 0.0036	0.0499 ± 0.0032
$\rho = 2$	0.0702 ± 0.0127	0.0482 ± 0.0019	0.0498 ± 0.0030	0.0522 ± 0.0015	0.0508 ± 0.0016
$\rho = 3$	0.0538 ± 0.0037	0.0484 ± 0.0011	0.0503 ± 0.0028	0.0525 ± 0.0015	0.0509 ± 0.0015
$\rho = 5$	0.0558 ± 0.0028	0.0482 ± 0.0006	0.0504 ± 0.0027	0.0525 ± 0.0016	0.0509 ± 0.0015
$\rho = 7$	0.0589 ± 0.0031	0.0482 ± 0.0006	0.0502 ± 0.0026	0.0524 ± 0.0016	0.0508 ± 0.0015

A.6.6 EVALUATION PROTOCOL

Generation quality is measured by averaged Maximum Mean Discrepancy (aMMD), the mean MMD over quadratic variation, spectral centroid, and degree correlation. Each MMD sums five RBF kernels whose bandwidths are log-spaced multiples 10^{-1}, $10^{-1/2}$, 1, $10^{1/2}$, 10 of the median heuristic, following the released implementation of Rozada et al. (2026). Unlike the released GAD evaluation, which uses the symmetric normalized Laplacian for the spectral centroid, our matched protocol uses the combinatorial Laplacian for all graph-based statistics. Each reported cell compares 5,000 generated signals with 5,000 held-out test signals, except for Molene, where both sets contain 150. All methods use the same normalized signals, graph operators, and metric implementation. All experiments run on a single Apple M5 Pro (48 GB unified memory).

A.6.7 HYPERPARAMETERS

Table 6: **Experimental hyperparameters.** Values are shared across methods and datasets unless stated otherwise.

GROUP	PARAMETER	VALUE
data	train / val / test caps	$20{,}000/3{,}000/5{,}000$
	normalization	per-node z-score, training statistics
graph	symmetrizer	$\frac{1}{2}(\boldsymbol{W}+\boldsymbol{W}^\top)$
	self-loops	removed after symmetrization
	Laplacian (model basis)	symmetric normalized
	Laplacian (evaluation)	combinatorial
	diffusion operator $\boldsymbol{L}_\delta$	$\boldsymbol{L}/\lambda_{\max}+\delta\boldsymbol{I}$
	Tikhonov shift δ	0.05
	backbone graph shift	$\boldsymbol{L}/\lambda_{\max}$
backbone	architecture	`DenoiserGNN` (Rozada et al., 2026)
	width	128
	blocks $\times$ taps	3×5
	time embedding	64
	trainable parameters	$61{,}185$, identical for every arm
optimization	optimizer	Adam
	learning rate	3×10^{-4}
	weight decay	10^{-4}
	batch size	256
	gradient clipping	1.0
	EMA decay	0.999
	validation interval	250 updates
stopping	patience	8 evaluations
	relative improvement	0.002
	maximum updates	$40{,}000$
conjugate	κ, finite SNR	2.0
forward	κ, near-zero SNR	32.6356 (METR-LA), 32.64 (others)
	t range	$[0.02, 1.0]$
	σ	1.0
VE forward	$q_{\min}, q_{\max}$	$0.02,\ 32.64$
	training draw	log-normal, $P_{\text{mean}}=-1.2$, $P_{\text{std}}=1.2$
	sampling grid exponent ρ_{VE}	7.0
reference	covariance estimator	Ledoit–Wolf (Ledoit & Wolf, 2004)
	retained covariance	graph-Fourier diagonal v_i
	floor ε_{ref}	10^{-6}
EDM	$P_{\text{mean}}, P_{\text{std}}, \rho$	$-1.2,\ 1.2,\ 7.0$
	$\sigma_{\min}, \sigma_{\max}$	$0.002,\ 80.0$
GAD process	γ, S, σ, T	$0.8,\ 7.0,\ 1.0,\ 1.0$
	$\alpha, c_{\min}$	$4.0,\ 0.1$
	total-variation loss weight	5×10^{-3}
	time-sampling exponent	0.65
GAD sampler	Newton iterations	2
	terminal s	0.0
GRCD sampler	residual quadrature	composite Simpson, 65 nodes
	selected grid exponent ρ	3
evaluation	RBF bandwidth multipliers	$\{0.1, 0.32, 1, 3.2, 10\}\times$ median
	generated / test samples	$5{,}000/5{,}000$ ($150/150$ on Molene)
	experimental seeds	$0, 1, 2$

A.7 OTHER BENCHMARK RESULTS

GRCD is the best method at NFE 4 on all four settings. The margin narrows with budget: by NFE 64 the isotropic and static preconditioning controls and EDM match or beat it on the three SBM settings, as does the conjugate preconditioner on SBM c=16, where they converge to a lower floor. On Molene, EDM's NFE-16 entry is a crossing point and it is not a converged value. Its

per-node dispersion is still contracting from $45\times$ toward 1, and the same arm degrades to 0.176 and 0.213 at NFE 32 and 64. The three preconditioning controls (Isotropic VE, Static precond. and Conjugate precond.) share this shape. In Table 1 and the tables below, "GRCD w/o residual score" uses the fitted terminal reference with ϵ-prediction and the Heun sampler. The last two rows add the residual score parameterization and the exponential-residual solver.

Table 7: **Matched comparison on Molene.** aMMD at the validation-converged stage and a fixed nominal NFE budget (Appendix A.6.5). Lower is better and bold denotes the best mean in each column. Results are mean $\pm$ sample standard deviation over three random seeds.

METHOD	NFE 4	NFE 8	NFE 16	NFE 64
EDM	2.6268 ± 0.0378	1.3517 ± 0.0869	$\mathbf{0.0951 \pm 0.0110}$	0.2126 ± 0.0330
GAD	2.7836 ± 0.0116	2.7613 ± 0.0558	2.7147 ± 0.0155	2.7303 ± 0.0261
WSD-style, scalar	2.6712 ± 0.0060	1.5278 ± 0.0730	1.1318 ± 0.0384	1.0181 ± 0.0134
WSD-style, graph prior	2.5191 ± 0.0101	0.8458 ± 0.0235	1.4101 ± 0.0507	1.1717 ± 0.0493
Isotropic VE	2.6264 ± 0.0319	0.5675 ± 0.0779	0.1243 ± 0.0304	0.2159 ± 0.0354
Static precond.	2.6326 ± 0.0287	0.7461 ± 0.0474	0.1325 ± 0.0163	0.2208 ± 0.0209
Conjugate precond.	2.6308 ± 0.0133	0.6139 ± 0.0307	0.1377 ± 0.0130	0.2284 ± 0.0295
GRCD w/o residual score	0.1545 ± 0.0295	0.1207 ± 0.0198	0.1556 ± 0.0055	0.1716 ± 0.0078
GRCD, ρ=1	0.5514 ± 0.0003	$\mathbf{0.1018 \pm 0.0275}$	0.1816 ± 0.0181	$\mathbf{0.1599 \pm 0.0245}$
GRCD, ρ=3	$\mathbf{0.1114 \pm 0.0497}$	0.1731 ± 0.0308	0.1725 ± 0.0414	0.1704 ± 0.0286

Table 8: **Matched comparison on SBM (c=1).** aMMD at the validation-converged stage and a fixed nominal NFE budget (Appendix A.6.5). Lower is better and bold denotes the best mean in each column. Results are mean $\pm$ sample standard deviation over three random seeds.

METHOD	NFE 4	NFE 8	NFE 16	NFE 64
EDM	1.7509 ± 0.0082	1.5238 ± 0.0101	0.4285 ± 0.0026	0.0119 ± 0.0017
GAD	2.2376 ± 0.0018	2.0090 ± 0.0085	1.9618 ± 0.0034	1.9395 ± 0.0078
WSD-style, scalar	1.3046 ± 0.1772	1.4082 ± 0.0704	1.3388 ± 0.0665	0.8715 ± 0.0535
WSD-style, graph prior	1.2084 ± 0.2054	1.4998 ± 0.1438	1.4533 ± 0.0949	0.8845 ± 0.0644
Isotropic VE	1.7524 ± 0.0067	1.2200 ± 0.0027	0.1704 ± 0.0014	$\mathbf{0.0113 \pm 0.0019}$
Static precond.	1.7522 ± 0.0083	1.2339 ± 0.0152	0.1465 ± 0.0255	0.0210 ± 0.0073
Conjugate precond.	1.7521 ± 0.0062	1.1663 ± 0.0203	0.1835 ± 0.0158	0.0621 ± 0.0455
GRCD w/o residual score	0.0527 ± 0.0092	$\mathbf{0.0426 \pm 0.0050}$	$\mathbf{0.0427 \pm 0.0050}$	0.0429 ± 0.0051
GRCD, ρ=1	$\mathbf{0.0485 \pm 0.0124}$	0.0502 ± 0.0095	0.0508 ± 0.0086	0.0510 ± 0.0083
GRCD, ρ=3	0.0497 ± 0.0062	0.0494 ± 0.0029	0.0526 ± 0.0069	0.0502 ± 0.0030

Table 9: **Matched comparison on SBM (c=4).** aMMD at the validation-converged stage and a fixed nominal NFE budget (Appendix A.6.5). Lower is better and bold denotes the best mean in each column. Results are mean $\pm$ sample standard deviation over three random seeds.

METHOD	NFE 4	NFE 8	NFE 16	NFE 64
EDM	1.9471 ± 0.0082	1.6387 ± 0.0103	0.4467 ± 0.0045	0.0161 ± 0.0028
GAD	2.3262 ± 0.0017	2.1619 ± 0.0091	2.1281 ± 0.0028	2.1100 ± 0.0066
WSD-style, scalar	1.5867 ± 0.2069	1.3181 ± 0.1055	1.4083 ± 0.1201	0.9278 ± 0.0792
WSD-style, graph prior	1.4519 ± 0.2599	1.3647 ± 0.2030	1.5479 ± 0.1677	0.9285 ± 0.1027
Isotropic VE	1.9492 ± 0.0062	1.2787 ± 0.0074	0.1797 ± 0.0081	0.0164 ± 0.0041
Static precond.	1.9518 ± 0.0061	1.3265 ± 0.0086	0.1766 ± 0.0115	$\mathbf{0.0132 \pm 0.0035}$
Conjugate precond.	1.9491 ± 0.0073	1.1984 ± 0.0346	0.1927 ± 0.0115	0.0696 ± 0.0487
GRCD w/o residual score	0.0707 ± 0.0203	$\mathbf{0.0525 \pm 0.0105}$	$\mathbf{0.0521 \pm 0.0100}$	0.0522 ± 0.0101
GRCD, ρ=1	$\mathbf{0.0574 \pm 0.0141}$	0.0641 ± 0.0128	0.0665 ± 0.0118	0.0673 ± 0.0114
GRCD, ρ=3	0.0616 ± 0.0080	0.0646 ± 0.0037	0.0685 ± 0.0092	0.0655 ± 0.0046

Table 10: **Matched comparison on SBM (c=16).** aMMD at the validation-converged stage and a fixed nominal NFE budget (Appendix A.6.5). Lower is better and bold denotes the best mean in each column. Results are mean ± sample standard deviation over three random seeds.

METHOD	NFE 4	NFE 8	NFE 16	NFE 64
EDM	2.1231 ± 0.0086	1.7378 ± 0.0141	0.4586 ± 0.0183	0.0204 ± 0.0049
GAD	2.4178 ± 0.0017	2.2966 ± 0.0091	2.2726 ± 0.0023	2.2574 ± 0.0064
WSD-style, scalar	1.5899 ± 0.4909	1.6443 ± 0.5671	1.4122 ± 0.2295	0.8368 ± 0.2536
WSD-style, graph prior	1.5082 ± 0.4521	1.7000 ± 0.7449	1.6301 ± 0.2184	0.8894 ± 0.1747
Isotropic VE	2.1257 ± 0.0056	1.3307 ± 0.0081	0.1824 ± 0.0181	0.0202 ± 0.0054
Static precond.	2.1269 ± 0.0058	1.3865 ± 0.0251	0.1758 ± 0.0201	0.0136 ± 0.0072
Conjugate precond.	2.1273 ± 0.0064	1.2838 ± 0.0019	0.1665 ± 0.0239	$\mathbf{0.0093 \pm 0.0072}$
GRCD w/o residual score	0.0918 ± 0.0298	$\mathbf{0.0503 \pm 0.0210}$	$\mathbf{0.0500 \pm 0.0196}$	0.0500 ± 0.0195
GRCD, ρ=1	$\mathbf{0.0270 \pm 0.0121}$	0.0530 ± 0.0089	0.0630 ± 0.0081	0.0663 ± 0.0078
GRCD, ρ=3	0.0421 ± 0.0144	0.0595 ± 0.0113	0.0657 ± 0.0071	0.0638 ± 0.0096

A.8 Other Results

A.8.1 Non-learned Gaussian controls

The three evaluation statistics are strongly influenced by second-order structure, so a Gaussian with a well-estimated covariance could potentially achieve a low aMMD without learning the non-Gaussian data distribution. We test this directly. The interpretation was fixed before the numbers existed: a control landing near the trained models would mean these three statistics cannot separate covariance matching from learned generation.

We evaluate four zero-mean Gaussian controls, sampled directly without a network, reverse process, or sampling budget: a per-node diagonal Gaussian $\mathcal{N}(\mathbf{0}, \mathrm{diag}(\mathbf{\Sigma}_{\mathrm{LW}}))$, the full Ledoit–Wolf Gaussian $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathrm{LW}})$, the GMRF $\mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{L}_\delta^{-1})$ corresponding to the near-zero-SNR limit in Equation 10, and $\mathcal{N}(\mathbf{0}, \boldsymbol{U}\,\mathrm{diag}(\boldsymbol{v})\boldsymbol{U}^\top)$, the fitted graph-spectral Gaussian reference before forward propagation. Their samples are evaluated with the same aMMD estimator and held-out test sets as the trained models. Because these controls have no training randomness, their reported spread is computed over 5 independent draws.

Table 11 shows that covariance matching alone does not account for GRCD's low-NFE performance: on every setting, the best Gaussian control has higher aMMD than GRCD at NFE 4. On METR-LA the strongest control reaches 0.1921 against GRCD's 0.0538, and the graph-spectral reference alone reaches 0.4134. Within the three SBM settings, the gap increases with spectral concentration, from $1.08\times$ at $c = 1$ to $2.93\times$ at $c = 16$.

On the SBM settings, GRCD reaches a plateau near its low aMMD values by moderate NFE, while alternative numerical solvers plateau at similar levels, suggesting that this behavior is not primarily caused by the numerical integrator. At $c = 1$, the fitted graph-spectral Gaussian reference itself reaches 0.0535, close to GRCD's 0.0497, indicating that the learned residual provides only a small additional improvement under this metric in that regime. In contrast, EDM and the isotropic and static preconditioning controls continue improving at larger budgets and eventually outperform GRCD at NFE 64.

A.8.2 Cost to reach a quality target

Table 11: Gaussian controls versus GRCD at NFE 4. aMMD. Lower is better. Gaussian-control values are mean ± sample standard deviation over 5 independent draws. "Best control" denotes the lowest aMMD among the four non-learned Gaussian controls. In the brackets, "full" is the full Ledoit–Wolf Gaussian and "spectral" is the fitted graph-spectral Gaussian reference.

SETTING	BEST CONTROL	GRCD	RATIO
METR-LA	0.1921 ± 0.0014 (full)	$\mathbf{0.0538}$	$\mathbf{3.57}\times$
SBM (c=16)	0.1232 ± 0.0041 (spectral)	$\mathbf{0.0421}$	$\mathbf{2.93}\times$
Molene	0.1822 ± 0.0242 (spectral)	$\mathbf{0.1114}$	$\mathbf{1.64}\times$
SBM (c=4)	0.0799 ± 0.0033 (spectral)	$\mathbf{0.0616}$	$\mathbf{1.30}\times$
SBM (c=1)	0.0535 ± 0.0027 (spectral)	$\mathbf{0.0497}$	$\mathbf{1.08}\times$

Table 12: **Cost to reach a quality target on METR-LA.** Cheapest measured sampling cost, in seconds per 1000 samples, at which each configuration first reaches an aMMD target at $\kappa = 2$, converged. "never" means the target is not reached at any budget up to NFE 64. "+ fitted terminal reference" is the row called "GRCD w/o residual score" in Table 1.

CONFIGURATION	aMMD ≤ 0.2	aMMD ≤ 0.1	aMMD ≤ 0.06
EDM	4.01 (NFE 32)	never	never
GAD	never	never	never
WSD-style, scalar	never	never	never
WSD-style, graph prior	2.07 (NFE 16)	8.27 (NFE 64)	never
Isotropic VE	1.94 (NFE 16)	never	never
Static precond.	1.94 (NFE 16)	never	never
Conjugate precond.	1.95 (NFE 16)	4.02 (NFE 32)	never
Mode-dependent clock	never	never	never
+ fitted terminal reference	**0.52** (NFE 4)	0.78 (NFE 6)	1.56 (NFE 12)
+ residual score	**0.52** (NFE 4)	0.78 (NFE 6)	1.56 (NFE 12)
GRCD, ρ=1	0.78 (NFE 6)	0.78 (NFE 6)	1.04 (NFE 8)
GRCD, ρ=3 (ours)	**0.52** (NFE 4)	**0.52** (NFE 4)	**0.52** (NFE 4)