How Do Stochastic Graph Flows Estimate Effective Geometry?

Anonymous authors
Paper under double-blind review

Abstract

Observed graphs are often noisy proxies for the geometry that a learning system should trust, yet continuous graph models often operate over a prescribed geometry without jointly adapting it and quantifying its uncertainty. We introduce *Stochastic Graph Flows* (SGFs), an energy-consistent framework that unifies adaptive graph geometry, representation dynamics, and model-induced geometry uncertainty. Positive edge conductances are eliminated exactly from an entropy-regularized joint energy, yielding a bounded robust energy whose gradient generates the same state-dependent Laplacian that drives the flow. Thus, geometry adaptation and representation dynamics are governed by a single scalar energy within a unified construction. At positive temperature, this energy further induces a unique Gibbs law over representation states, whose pushforward through the conductance map yields an effective mean geometry and edge-level uncertainty on the observed support. We establish free-energy dissipation and Gibbs invariance, derive conditional stability guarantees for the effective operator and its spectral subspace, and explicitly separate modeling, temperature, discretization, mixing, and sampling errors. Experiments on synthetic and real graphs support the adaptive-geometry mechanism and highlight stochasticity as a principled source of geometry uncertainty, alongside modest improvements in point estimation.

1 Introduction

Graphs supplied to learning systems are rarely exact. Their edges may be spurious, misweighted, adversarially added, or simply imperfect discretizations of an underlying geometry (Jin et al., 2020; Kalofolias, 2016). In practice, many graph learning pipelines treat the observed graph operator as fixed once computation begins. Message passing, graph diffusion, and continuous-depth dynamics propagate information according to the graph geometry (Poli et al., 2019; Chamberlain et al., 2021; Thorpe et al., 2022). This creates a basic mismatch: filtering a prescribed Laplacian is not the same as estimating the effective geometry that the Laplacian is meant to represent.

The mismatch is more consequential in stochastic graph models. Stochastic graph dynamics quantify uncertainty in representations or predictions conditioned on a given graph (Bergna et al., 2023; 2025), while adaptive graph methods revise edge weights or latent structure through prescribed criteria (Jin et al., 2020; Manenti et al., 2025). What remains less developed is a unified formulation that couples geometry adaptation, state evolution, and geometry uncertainty through a single energy.

Against this backdrop, SGF formulates effective graph geometry as a state-dependent object coupled directly to the representation dynamics. Given a representation state $\mathbf{H}_t$, exact elimination of positive edge conductances yields $c^\star_{ij}(\mathbf{H}_t) = A_{ij}\exp(-\|\mathbf{h}_i(t)-\mathbf{h}_j(t)\|_2^2/(2\sigma^2))$, inducing the closed feedback $\mathbf{H}_t \to \mathbf{C}^\star(\mathbf{H}_t) \to \mathbf{L}(\mathbf{H}_t) \to \mathrm{d}\mathbf{H}_t$. The resulting bounded robust energy governs both geometry adaptation and representation drift, placing SGF within the classical latent-weight view of robust penalties (Black and Rangarajan, 1996). The state remains centered and full-rank while evolving on the scaled Stiefel manifold $\mathcal{M}_{n,p}$ (Absil et al., 2008) under intrinsic Langevin dynamics (Hsu, 2002). At positive temperature, the induced Gibbs law is pushed through the conductance map to yield an effective mean geometry and model-based edge uncertainty on the observed support. Thus, SGF follows a single chain: variational elimination defines the adaptive geometry, stochastic dynamics induce a Gibbs state law, and the resulting effective operator admits conditional stability analysis and finite-step estimation. Figure 1 summarizes this geometry–dynamics closure.

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.

To our knowledge, SGF is the first formulation in which a single variationally reduced energy determines both the adaptive conductance geometry and the stochastic state drift, while inducing an equilibrium conductance law through its Gibbs state distribution. Accordingly, our contributions are:

- **Energy-consistent adaptive geometry.** We derive positive edge conductances by exact variational elimination, obtaining a bounded robust energy whose gradient generates the state-dependent Laplacian governing the representation flow.
- **Geometry uncertainty from stochastic state dynamics.** We formulate intrinsic Langevin dynamics on the constrained representation space, establish free-energy dissipation and a unique frozen-parameter Gibbs law, and push this law through the conductance map to obtain an effective mean geometry and edge-level uncertainty.
- **Conditional stability of the effective geometry.** We relate representation-dependent edge attenuation to perturbation of the equilibrium mean operator and its spectral subspace under explicit assumptions, and decompose the finite-step estimation error into modeling, temperature, discretization, mixing, and Monte Carlo components.

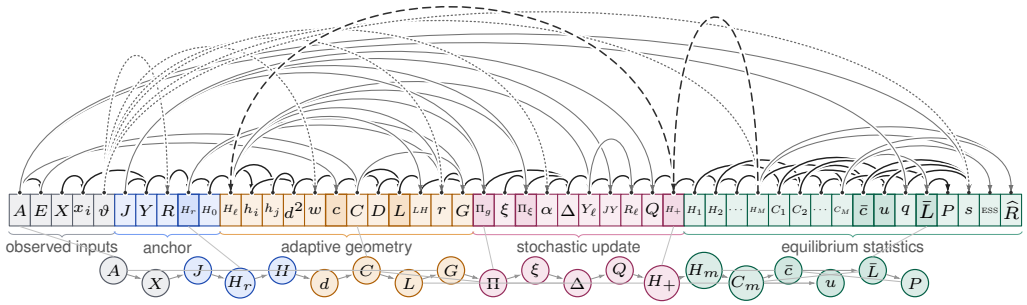


Figure 1: **Computation and induced geometry in SGF.** The upper cell graph traces SGF from observed inputs through adaptive geometry, stochastic state evolution, and equilibrium estimation. Retained states induce conductance samples whose aggregation yields the effective mean geometry, edge-level uncertainty, spectral quantities, and sampling diagnostics. The lower symbolic graph summarizes the resulting geometry–dynamics hierarchy.

2 Related Work

Continuous and stochastic graph dynamics. Graph ODE and diffusion models include GRAND and GRAND++, which study diffusion and source-driven propagation, and Beltrami neural diffusion, which jointly evolves features and positional geometry (Poli et al., 2019; Chamberlain et al., 2021; Thorpe et al., 2022; Chamberlain et al., 2021). Stochastic extensions introduce latent diffusion noise for uncertainty-aware graph prediction (Bergna et al., 2023). SGF builds on this dynamical perspective (Li et al., 2026) while coupling the evolving state to a state-dependent conductance operator, so that stochastic state variation also induces uncertainty in the effective geometry.

Latent and adaptive graph geometry. Latent graph models may alter graph support directly; Manenti et al. model a distribution over latent graphs to represent predictive and structural uncertainty (Manenti et al., 2025). Related graph-learning methods adapt edge weights or connectivity using criteria based on smoothness, task objectives, attention, spectral properties, or structural priors (Kalofolias, 2016; Veličković et al., 2018; Jin et al., 2020; Runwal et al., 2022; Liu and Wang, 2026; Zhang et al., 2023). In SGF, the object of interest is the state-dependent conductance geometry: conductances arise by exact variational elimination as a deterministic map $\mathbf{C}^\star(\mathbf{H})$, and their uncertainty is induced by pushing the Gibbs state law through this map on the declared candidate support.

Robust and constrained formulations. SGF builds on latent-weight robust penalties and classical spectral/Stiefel formulations (Black and Rangarajan, 1996; Ng et al., 2001; Absil et al., 2008; Holtz et al., 2023), using these ingredients within a reduced-energy model that couples adaptive geometry and stochastic representation dynamics. The distinction is that a single reduced energy governs the conductance map, the stochastic state drift, and the induced Gibbs conductance law.

3 PROBLEM SETTING

Let $\mathcal{G}_{\text{obs}} = (\mathcal{V}, \mathcal{E}_{\text{obs}}, \mathbf{A})$ be an undirected observed graph with $n = |\mathcal{V}|$, symmetric nonnegative weights, and $A_{ii} = 0$ for every node. Any self-loops introduced by preprocessing are excluded from the geometric energy and handled separately by the encoder. We define the unordered observed support as $\mathcal{E}_{\text{obs}} = \{\{i, j\} : i < j,\ A_{ij} > 0\}$, so every edge in the energy has strictly positive reference weight. Node attributes are $\mathbf{X} \in \mathbb{R}^{n\times d}$. Following robust and graph-structure learning settings (Jin et al., 2020; Kalofolias, 2016), we treat the observed graph as a noisy candidate support: $A_{ij} > 0$ makes an edge available for reweighting but does not require its full weight to be trusted.

An encoder Φ_θ produces $\mathbf{Y}_{\text{ref}} = \Phi_\theta(\mathbf{X}) \in \mathbb{R}^{n\times p}$. Define the centering projector $\mathbf{J} = \mathbf{I}_n - \mathbf{1}\mathbf{1}^\top/n$. We require $\mathbf{Y}_{\text{ref}}^\top \mathbf{J}\mathbf{Y}_{\text{ref}} \succ 0$; the encoded fidelity anchor is $\mathbf{H}_{\text{ref}} = \sqrt{n}\,\mathbf{J}\mathbf{Y}_{\text{ref}}\big(\mathbf{Y}_{\text{ref}}^\top \mathbf{J}\mathbf{Y}_{\text{ref}}\big)^{-1/2}$. Then $\mathbf{1}^\top\mathbf{H}_{\text{ref}} = 0$ and $\mathbf{H}_{\text{ref}}^\top\mathbf{H}_{\text{ref}} = n\mathbf{I}_p$. The state space is $\mathcal{M}_{n,p} = \{\mathbf{H} \in \mathbb{R}^{n\times p} : \mathbf{1}^\top\mathbf{H} = 0,\ \mathbf{H}^\top\mathbf{H} = n\mathbf{I}_p\}$. We assume $1 \le p < n - 1$ (Absil et al., 2008), under which $\mathcal{M}_{n,p}$ is compact and connected. Appendix A.1 gives the corresponding geometric details.

For edge variables $\{c_{ij}\}_{\{i,j\}\in\mathcal{E}_{\text{obs}}}$, define the symmetric conductance matrix by $C_{ij} = C_{ji} = c_{ij}$ on $\mathcal{E}_{\text{obs}}$, $C_{ii} = 0$, and $C_{ij} = 0$ outside $\mathcal{E}_{\text{obs}}$. Let $\mathbf{D}_{\mathbf{C}} = \operatorname{diag}(\mathbf{C}\mathbf{1})$ and $\mathbf{L}_{\mathbf{C}} = \mathbf{D}_{\mathbf{C}} - \mathbf{C}$. Because each unordered edge is counted once, $\operatorname{tr}(\mathbf{H}^\top\mathbf{L}_{\mathbf{C}}\mathbf{H}) = \sum_{\{i,j\}\in\mathcal{E}_{\text{obs}}} c_{ij}\,\|\mathbf{h}_i - \mathbf{h}_j\|_2^2$. The derivation uses this combinatorial operator; fully differentiated normalized variants are discussed in Appendix C.2.

4 SELF-CONSISTENT STOCHASTIC GRAPH FLOWS

4.1 LATENT CONDUCTANCE ENERGY

For every observed edge, introduce a positive latent conductance $c_{ij} > 0$. Given bandwidth $\sigma > 0$ and fidelity coefficient $\eta \ge 0$, define

$$\mathcal{E}(\mathbf{H}, \mathbf{C}; \mathbf{H}_{\text{ref}}) = \frac{1}{2}\sum_{\{i,j\}\in\mathcal{E}_{\text{obs}}} c_{ij}\,\|\mathbf{h}_i - \mathbf{h}_j\|_2^2 + \sigma^2 \sum_{\{i,j\}\in\mathcal{E}_{\text{obs}}} \left[c_{ij}\log\frac{c_{ij}}{A_{ij}} - c_{ij} + A_{ij}\right] + \frac{\eta}{2}\,\|\mathbf{H} - \mathbf{H}_{\text{ref}}\|_F^2 . \tag{1}$$

The first term is a conductance-weighted Dirichlet energy. The second is the generalized KL divergence between positive edge measures $\mathbf{C}$ and $\mathbf{A}$ (Csiszár, 1975); its derivative diverges to $-\infty$ as $c_{ij} \downarrow 0$, ensuring that the minimizer lies in the positive interior. The final term retains information from the anchor. Because $\mathcal{M}_{n,p}$ fixes centered covariance to $\mathbf{I}_p$, no covariance penalty is required.

Proposition 1 (Variational conductance and robust reduction)**.** *For fixed $\mathbf{H} \in \mathcal{M}_{n,p}$, the objective in Eq. (1) is strictly convex in every $c_{ij} > 0$ and has the unique minimizer $c_{ij}^\star(\mathbf{H}) = A_{ij}\exp\!\left(-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|_2^2}{2\sigma^2}\right)$. After eliminating $\mathbf{C}$, the reduced energy is*

$$\mathcal{E}_{\text{red}}(\mathbf{H}; \mathbf{H}_{\text{ref}}) = \sigma^2 \sum_{\{i,j\}\in\mathcal{E}_{\text{obs}}} A_{ij}\left[1 - \exp\!\left(-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|_2^2}{2\sigma^2}\right)\right] + \frac{\eta}{2}\,\|\mathbf{H} - \mathbf{H}_{\text{ref}}\|_F^2 . \tag{2}$$

Its Euclidean gradient is $\nabla\mathcal{E}_{\text{red}}(\mathbf{H}) = \mathbf{L}(\mathbf{H})\mathbf{H} + \eta(\mathbf{H} - \mathbf{H}_{\text{ref}})$, $\quad \mathbf{L}(\mathbf{H}) = \mathbf{L}_{\mathbf{C}^\star(\mathbf{H})}$.

The exponential conductance resembles Gaussian affinity weighting (Ng et al., 2001), but arises by exact variational elimination and evolves with the state. This variational reduction is closely related to classical latent-variable representations of robust penalties (Black and Rangarajan, 1996). The edge penalty $\rho_\sigma(s) = \sigma^2(1 - e^{-s/(2\sigma^2)})$ is locally quadratic because $\rho_\sigma(s) = s/2 - s^2/(8\sigma^2) + O(s^3/\sigma^4)$ as $s \downarrow 0$, but it is bounded above by σ^2 (Black and Rangarajan, 1996). Consequently, locally consistent edges behave like Dirichlet smoothing, while the influence $\rho'_\sigma(s) = \frac{1}{2}e^{-s/(2\sigma^2)}$ decays exponentially with squared discrepancy. The bandwidth σ controls the transition between these regimes.

4.2 GEOMETRY-CONSTRAINED STOCHASTIC DYNAMICS

For $\mathbf{H} \in \mathcal{M}_{n,p}$, the tangent space is $T_{\mathbf{H}}\mathcal{M}_{n,p} = \{\mathbf{Z} : \mathbf{1}^\top\mathbf{Z} = 0,\ \mathbf{H}^\top\mathbf{Z} + \mathbf{Z}^\top\mathbf{H} = 0\}$. With $\operatorname{sym}(\mathbf{M}) = (\mathbf{M} + \mathbf{M}^\top)/2$, the Frobenius-orthogonal tangent projection (Absil et al., 2008) is

$$\Pi_{\mathbf{H}}(\mathbf{Z}) = \mathbf{J}\mathbf{Z} - \frac{1}{n}\mathbf{H}\operatorname{sym}(\mathbf{H}^\top\mathbf{J}\mathbf{Z}). \tag{3}$$

The Riemannian gradient under the induced metric is $\operatorname{grad}\mathcal{E}_{\mathrm{red}}(\mathbf{H}) = \Pi_{\mathbf{H}}(\nabla\mathcal{E}_{\mathrm{red}}(\mathbf{H}))$ (Edelman et al., 1998; Absil et al., 2008; Boumal, 2023).

Fix the observed weights $\mathbf{A}$, the fidelity anchor $\mathbf{H}_{\mathrm{ref}}$, and hyperparameters (σ, η, τ). Let $\mathbf{H}_{\mathrm{init}} \in \mathcal{M}_{n,p}$ be an initial state and let $\mathbf{B}_t$ be standard matrix Brownian motion in $\mathbb{R}^{n\times p}$. The canonical SGF is the Stratonovich SDE

$$\mathrm{d}\mathbf{H}_t = -\operatorname{grad}\mathcal{E}_{\mathrm{red}}(\mathbf{H}_t; \mathbf{H}_{\mathrm{ref}})\,\mathrm{d}t + \sqrt{2\tau}\,\Pi_{\mathbf{H}_t} \circ \mathrm{d}\mathbf{B}_t, \qquad \mathbf{H}_{t=0} = \mathbf{H}_{\mathrm{init}}. \tag{4}$$

Here $\tau > 0$ is the temperature. The projected constant ambient vector fields form a Parseval frame of each tangent space and generate the Laplace–Beltrami operator (Hsu, 2002); Appendix A.2 gives the extrinsic identity. The equivalent Itô equation contains the mean-curvature drift. Omitting that correction while interpreting the projected equation in the Itô sense would define a different process.

The bidirectional dependence is explicit: $\mathbf{H}_t$ determines $\mathbf{C}^\star(\mathbf{H}_t)$ and $\mathbf{L}(\mathbf{H}_t)$, while the induced Laplacian enters the drift of $\mathbf{H}_t$. The guarantees in Section 5 therefore concern the full state-dependent reduced energy rather than a frozen-Laplacian surrogate.

4.3 Learned geometry and edge uncertainty

Let π_τ denote the unique invariant law of the frozen-parameter diffusion in Eq. (4). We use the term *Gibbs posterior* only in the generalized energy-based sense; no explicit likelihood model is asserted (Bissiri et al., 2016). Define the Gibbs-equilibrium conductance moments

$$\bar{c}_{ij}^{(\tau)} = \mathbb{E}_{\mathbf{H}\sim\pi_\tau}[c_{ij}^\star(\mathbf{H})], \qquad u_{ij}^{(\tau)} = \mathrm{Var}_{\mathbf{H}\sim\pi_\tau}[c_{ij}^\star(\mathbf{H})]. \tag{5}$$

The effective learned operator on the observed support is $\bar{\mathbf{L}}_\tau = \mathbf{L}_{\bar{\mathbf{C}}^{(\tau)}} = \mathbb{E}_{\pi_\tau}[\mathbf{L}(\mathbf{H})]$. The support ratio $q_{ij} = \bar{c}_{ij}^{(\tau)}/A_{ij} \in (0, 1]$ measures equilibrium support for an observed edge. One declared corruption score is $s_{ij}^{\mathrm{corr}} = 1 - q_{ij} + \kappa\,\frac{\sqrt{u_{ij}^{(\tau)}}}{A_{ij}+\varepsilon_{\mathrm{stab}}}$, $\varepsilon_{\mathrm{stab}} > 0$, where $\kappa \geq 0$ is fixed in advance or selected without test labels.

Directly averaging $\mathbf{H}$ is not invariant to right multiplication by an orthogonal matrix. Geometry evaluation therefore uses the gauge-invariant projector $\mathbf{P}_{\mathbf{H}} = \mathbf{H}\mathbf{H}^\top/n$ (Absil et al., 2008). If node features are averaged across equilibrium samples, the samples are first aligned to a reference state by an orthogonal Procrustes transformation (Schönemann, 1966).

4.4 Retraction-based discretization

For any $\mathbf{Y}$ satisfying $\mathbf{Y}^\top\mathbf{J}\mathbf{Y} \succ 0$ (equivalently, $\mathbf{J}\mathbf{Y}$ has full column rank), define the centered polar map (Absil et al., 2008; Boumal, 2023)

$$\mathcal{Q}(\mathbf{Y}) = \sqrt{n}\,\mathbf{J}\mathbf{Y}\big(\mathbf{Y}^\top\mathbf{J}\mathbf{Y}\big)^{-1/2}. \tag{6}$$

Given step size h, draw $\mathbf{\Xi}^{(\ell)}$ with independent standard normal entries and form

$$\mathbf{\Delta}^{(\ell)} = -h\,\operatorname{grad}\mathcal{E}_{\mathrm{red}}(\mathbf{H}^{(\ell)}; \mathbf{H}_{\mathrm{ref}}) + \sqrt{2\tau h}\,\Pi_{\mathbf{H}^{(\ell)}}(\mathbf{\Xi}^{(\ell)}), \tag{7}$$

followed by $\mathbf{H}^{(\ell+1)} = \mathcal{Q}(\mathbf{H}^{(\ell)} + \mathbf{\Delta}^{(\ell)})$. Appendix A.3 proves that $\mathcal{Q}$ is a second-order retraction and that the one-step weak generator converges to the generator of Eq. (4) as $h \downarrow 0$ (Boumal, 2023). Because $\mathbf{\Delta}^{(\ell)} \in T_{\mathbf{H}^{(\ell)}}\mathcal{M}_{n,p}$, $\big(\mathbf{H}^{(\ell)} + \mathbf{\Delta}^{(\ell)}\big)^\top\mathbf{J}\big(\mathbf{H}^{(\ell)} + \mathbf{\Delta}^{(\ell)}\big) = n\mathbf{I}_p + \big(\mathbf{\Delta}^{(\ell)}\big)^\top\mathbf{\Delta}^{(\ell)} \succ 0$. Hence the polar update is admissible for every realization and preserves centering and scaled orthogonality exactly. It does not preserve the Gibbs law exactly at finite h. Figure 2 summarizes the resulting one-step mechanism from $\mathbf{H}^{(\ell)}$ to $\mathbf{H}^{(\ell+1)}$, while Algorithm 1 gives the corresponding finite-depth procedure.

For K chains and T transitions, the total work is $O\big(KT(|\mathcal{E}_{\mathrm{obs}}|p + np^2 + p^3)\big)$, with $O(|\mathcal{E}_{\mathrm{obs}}| + np)$ working memory per chain. For fixed small p, each transition is approximately linear in sparse graph size. Chains are parallelizable, although total work still scales linearly with K. Multiple chains are needed only for sampling-based geometry estimation; a finite-depth representation layer does not require them.

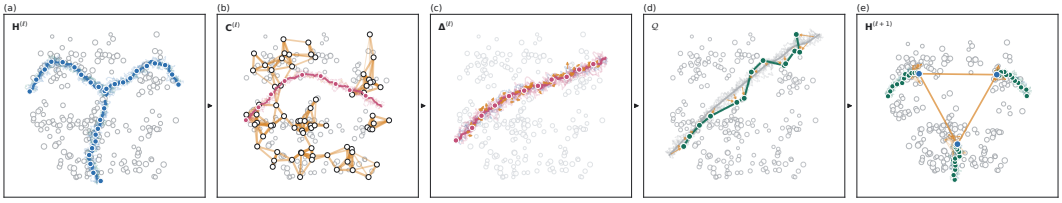


Figure 2: **Finite-depth discretization of SGF.** Starting from the current representation state $\mathbf{H}^{(\ell)}$, SGF recomputes the adaptive conductances $\mathbf{C}^{(\ell)}$, forms the stochastic tangent increment $\mathbf{\Delta}^{(\ell)}$, and retracts the perturbed state through the centered polar map $\mathcal{Q}$ to obtain $\mathbf{H}^{(\ell+1)}$. This closes the state–geometry feedback loop at each finite-depth step; the displayed trajectories are schematic.

Algorithm 1 Finite-Depth Self-Consistent Stochastic Graph Flow

Require: Observed weights $\mathbf{A}$, features $\mathbf{X}$, encoder Φ_θ, width p, steps L, step size h, bandwidth σ, fidelity η, temperature τ.

$\mathbf{H}_{\mathrm{ref}} \leftarrow \mathcal{Q}(\Phi_\theta(\mathbf{X}))$; $\mathbf{H}^{(0)} \leftarrow \mathbf{H}_{\mathrm{ref}}$.
for $\ell = 0, \dots, L-1$ **do**
  $c_{ij}^{(\ell)} \leftarrow A_{ij} \exp(-\left\|\mathbf{h}_i^{(\ell)} - \mathbf{h}_j^{(\ell)}\right\|_2^2 / (2\sigma^2))$ for $\{i,j\} \in \mathcal{E}_{\mathrm{obs}}$.
  Form $\mathbf{L}^{(\ell)} = \mathbf{L}_{\mathbf{C}^{(\ell)}}$ and $\mathbf{G}^{(\ell)} = \mathbf{L}^{(\ell)}\mathbf{H}^{(\ell)} + \eta(\mathbf{H}^{(\ell)} - \mathbf{H}_{\mathrm{ref}})$.
  Draw $\mathbf{\Xi}^{(\ell)}$ and set $\mathbf{\Delta}^{(\ell)} \leftarrow -h\Pi_{\mathbf{H}^{(\ell)}}(\mathbf{G}^{(\ell)}) + \sqrt{2\tau h}\Pi_{\mathbf{H}^{(\ell)}}(\mathbf{\Xi}^{(\ell)})$.
  $\mathbf{H}^{(\ell+1)} \leftarrow \mathcal{Q}(\mathbf{H}^{(\ell)} + \mathbf{\Delta}^{(\ell)})$.
end for
$c_{ij}^{(L)} \leftarrow A_{ij} \exp(-\left\|\mathbf{h}_i^{(L)} - \mathbf{h}_j^{(L)}\right\|_2^2 / (2\sigma^2))$ for $\{i,j\} \in \mathcal{E}_{\mathrm{obs}}$.
return $\mathbf{H}^{(L)}$ and $\mathbf{C}^{(L)}$.

Finite-depth learning versus equilibrium estimation. When SGF is used as a differentiable representation layer, no exact sampling claim is made and no Metropolis correction is required. Gibbs-equilibrium geometry estimation is a separate frozen-parameter phase with an independently specified initial state, burn-in, sampling interval, and retained sample count. Its finite-step bias and incomplete-mixing error are explicit in Corollary 6 and Appendix C.1.

Use as a representation layer. SGF can be used directly as a transductive representation method or inserted as a differentiable layer. The mechanism under test is the replacement of a fixed propagation operator by the self-consistent update induced by Eq. (4).

5 Theoretical Guarantees

5.1 Structural properties

Proposition 2 (Permutation equivariance of the SGF flow)**.** *Let* $\mathbf{S}$ *be any* $n \times n$ *permutation matrix. Simultaneously replace* $(\mathbf{A}, \mathbf{H}_{\mathrm{ref}}, \mathbf{H}_{\mathrm{init}})$ *by* $(\mathbf{SAS}^\top, \mathbf{SH}_{\mathrm{ref}}, \mathbf{SH}_{\mathrm{init}})$*. If* $\mathbf{H}_t$ *solves Eq. (4), then* $\mathbf{SH}_t$ *has the same law as the SGF driven by the permuted inputs. Moreover,* $\mathbf{C}^\star(\mathbf{SH}; \mathbf{SAS}^\top) = \mathbf{SC}^\star(\mathbf{H}; \mathbf{A})\mathbf{S}^\top$, $\mathbf{L}(\mathbf{SH}; \mathbf{SAS}^\top) = \mathbf{SL}(\mathbf{H}; \mathbf{A})\mathbf{S}^\top$*. End-to-end equivariance with respect to* $(\mathbf{A}, \mathbf{X})$ *additionally requires* Φ_θ *to be permutation equivariant.*

Theorem 3 (Well-posedness, free-energy dissipation, and Gibbs invariance)**.** *Assume* $1 \le p < n-1$, $\sigma > 0$, $\eta \ge 0$, $\tau > 0$, *and finite observed weights. For fixed* $\mathbf{A}$ *and* $\mathbf{H}_{\mathrm{ref}}$*, Eq. (4) admits a unique global strong solution on the compact manifold* $\mathcal{M}_{n,p}$*. For every* $t > 0$*, elliptic smoothing gives a smooth strictly positive density* ρ_t *with respect to Riemannian volume* μ*. Define* $\mathcal{F}_\tau(\rho) = \int_{\mathcal{M}_{n,p}} \mathcal{E}_{\mathrm{red}}(\mathbf{H}; \mathbf{H}_{\mathrm{ref}})\rho(\mathbf{H})\,\mathrm{d}\mu + \tau \int_{\mathcal{M}_{n,p}} \rho(\mathbf{H}) \log \rho(\mathbf{H})\,\mathrm{d}\mu$*. Along this density evolution, for every* $t > 0$ *(Hsu, 2002),*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}_\tau(\rho_t) = -\int_{\mathcal{M}_{n,p}} \rho_t \left\|\mathrm{grad}(\mathcal{E}_{\mathrm{red}} + \tau \log \rho_t)\right\|_F^2 \,\mathrm{d}\mu \le 0. \tag{8}$$

The Gibbs density $\pi_\tau(\mathbf{H}) = Z_\tau^{-1}\exp(-\mathcal{E}_{\mathrm{red}}(\mathbf{H};\mathbf{H}_{\mathrm{ref}})/\tau)$ is invariant. The diffusion is elliptic on the connected compact manifold, so this invariant probability measure is unique.

The theorem covers the full state-dependent Laplacian through the scalar reduced energy. The dissipated quantity is the probability-level free energy, not the expected potential energy alone; Brownian forcing may increase potential energy along individual trajectories while Eq. (8) remains non-positive at the distributional level (Jordan et al., 1998).

5.2 Suppression of inconsistent edges

Theorem 4 (Pointwise conductance separation under a representation margin)**.** *Consider two observed edges $e_{\mathrm{in}} = \{i,j\}$ and $e_{\mathrm{out}} = \{r,s\}$. At any state $\mathbf{H}$, suppose $\|\mathbf{h}_i - \mathbf{h}_j\|_2^2 \le r_{\mathrm{in}}$ and $\|\mathbf{h}_r - \mathbf{h}_s\|_2^2 \ge r_{\mathrm{out}}$, with $r_{\mathrm{out}} > r_{\mathrm{in}}$. Then*

$$\frac{c^\star_{rs}(\mathbf{H})}{c^\star_{ij}(\mathbf{H})} \le \frac{A_{rs}}{A_{ij}}\exp\left(-\frac{r_{\mathrm{out}} - r_{\mathrm{in}}}{2\sigma^2}\right). \tag{9}$$

For equal observed weights, every additional squared-distance margin Δr tightens the upper bound on the relative conductance by a factor $e^{-\Delta r/(2\sigma^2)}$.

This pointwise statement does not guarantee that the dynamics create the required representation margin. It states exactly what the conductance mechanism does once such a margin is present.

5.3 Conditional latent spectral stability

Let $\mathbf{L}_\star$ be a latent clean Laplacian. Let $\mathcal{I}_\star = [a_\star, b_\star] \subset \mathbb{R}$ be an interval containing an eigenvalue block of $\mathbf{L}_\star$ and no other eigenvalues. Define $\mathbf{P}_\star = \mathbf{1}_{\mathcal{I}_\star}(\mathbf{L}_\star), \quad k = \mathrm{rank}(\mathbf{P}_\star)$, where $\mathbf{1}_{\mathcal{I}_\star}(\mathbf{L}_\star)$ denotes the spectral projector obtained by functional calculus. Assume a standard spectral-separation condition (Davis and Kahan, 1970): the block is separated from the rest of the spectrum by $\delta_\star = \mathrm{dist}\left(\mathcal{I}_\star, \mathrm{spec}(\mathbf{L}_\star)\setminus\mathcal{I}_\star\right) > 0$. Define the enlarged interval $\mathcal{I}_\star^+ = \left\{x \in \mathbb{R} : \mathrm{dist}(x,\mathcal{I}_\star) < \frac{\delta_\star}{2}\right\}$. When $\mathbf{L}_\star$ is connected, the block may be chosen as the first k nonconstant modes by selecting an interval that excludes both the constant mode and the remaining spectrum (Chung, 1997).

For a reference state $\mathbf{H}^\dagger \in \mathcal{M}_{n,p}$, define its right-orthogonal orbit and the associated quotient distance by $\mathcal{O}^\dagger = \{\mathbf{H}^\dagger\mathbf{R} : \mathbf{R} \in \mathrm{O}(p)\}, \quad d_{\mathcal{O}}(\mathbf{H},\mathbf{H}^\dagger) = \inf_{\mathbf{R}\in\mathrm{O}(p)} \mathrm{dist}_{\mathcal{M}}(\mathbf{H},\mathbf{H}^\dagger\mathbf{R})$. The orbit formulation is necessary when $\eta = 0$, because the conductance energy is invariant under $\mathbf{H} \mapsto \mathbf{HR}$ and a unique minimizer in the ambient representation coordinates cannot then be assumed.

Assumption 1 (Equilibrium concentration modulo orthogonal gauge)**.** *There exists a geometrically relevant reference orbit $\mathcal{O}^\dagger$ such that $\mathbb{E}_{\pi_\tau}\left[d_{\mathcal{O}}(\mathbf{H},\mathbf{H}^\dagger)^2\right] \le C_{\mathrm{conc}}\tau$.*

Remark 1 (Scope of the concentration assumption)**.** *Assumption 1 is a local concentration condition that serves as the localization condition in the spectral argument. It is compatible with low-temperature concentration near a normally nondegenerate minimizing orbit, but is not derived here from the observed graph or from a data-generating model. In particular, it limits Gibbs mass far from the selected orbit. The theorem below is therefore a stability result under favorable representability and equilibrium concentration rather than a global graph-identification theorem. Appendix D.7 complements this scope with a finite-chain initialization-sensitivity diagnostic based on gauge-invariant retained-mean projectors.*

By smoothness on the compact manifold, the adaptive Laplacian map has a finite global Lipschitz constant K_L (Boumal, 2023); Appendix A.4 proves $\left\|\mathbf{L}(\mathbf{H}) - \mathbf{L}(\mathbf{H}^\dagger)\right\|_2 \le K_L d_{\mathcal{O}}(\mathbf{H},\mathbf{H}^\dagger)$.

Theorem 5 (Latent-subspace stability under explicit approximation and concentration conditions)**.** *Define $\varepsilon_{\mathrm{model}} = \left\|\mathbf{L}(\mathbf{H}^\dagger) - \mathbf{L}_\star\right\|_2, \quad \varepsilon_{\mathrm{eq}} = \varepsilon_{\mathrm{model}} + K_L\sqrt{C_{\mathrm{conc}}\tau}$. Here $\varepsilon_{\mathrm{model}}$ quantifies the approximation error of the conductance family at the selected orbit and enters the bound explicitly. Under Assumption 1, the exact equilibrium mean operator satisfies*

$$\left\|\bar{\mathbf{L}}_\tau - \mathbf{L}_\star\right\|_2 \le \varepsilon_{\mathrm{eq}}. \tag{10}$$

If $\varepsilon_{\mathrm{eq}} \le \delta_\star/4$, define $\bar{\mathbf{P}}_{\tau,k} = \mathbf{1}_{\mathcal{I}_\star^+}(\bar{\mathbf{L}}_\tau)$. Then $\mathrm{rank}(\bar{\mathbf{P}}_{\tau,k}) = k$ (Davis and Kahan, 1970), and

$$\left\|\bar{\mathbf{P}}_{\tau,k} - \mathbf{P}_\star\right\|_2 \le \frac{\varepsilon_{\mathrm{eq}}}{\delta_\star - \varepsilon_{\mathrm{eq}}} \le \frac{2\varepsilon_{\mathrm{eq}}}{\delta_\star}. \tag{11}$$

Corollary 6 (Finite-step error decomposition)**.** *Assume that the implemented Markov chain with step size h admits a stationary law $\pi_{\tau,h}$. Let $\mu_{\tau,h,t}$ be the law after t transitions from the declared initial distribution. For burn-in B, spacing S, and retained count M, define $\bar{\mathbf{L}}_{\tau,h} = \mathbb{E}_{\pi_{\tau,h}}[\mathbf{L}(\mathbf{H})]$, $\mathbf{L}^{\mathrm{mean}}_{\tau,h,B,S,M} = \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}_{\mu_{\tau,h,B+mS}}[\mathbf{L}(\mathbf{H})]$. Set $\varepsilon_{\mathrm{disc}}(h) = \left\|\bar{\mathbf{L}}_{\tau,h} - \bar{\mathbf{L}}_{\tau}\right\|_2$, $\varepsilon_{\mathrm{mix}}(B,S,M,h) = \left\|\mathbf{L}^{\mathrm{mean}}_{\tau,h,B,S,M} - \bar{\mathbf{L}}_{\tau,h}\right\|_2$. Let $\widehat{\mathbf{L}}_{\tau,h,B,S,M}$ be the average Laplacian from those M retained states, which may be dependent, and suppose $\left\|\widehat{\mathbf{L}}_{\tau,h,B,S,M} - \mathbf{L}^{\mathrm{mean}}_{\tau,h,B,S,M}\right\|_2 \le \varepsilon_{\mathrm{MC}}$ with probability at least $1-\delta$. Then, with the same probability,*

$$\left\|\widehat{\mathbf{L}}_{\tau,h,B,S,M} - \mathbf{L}_\star\right\|_2 \le \varepsilon_{\mathrm{model}} + K_L\sqrt{C_{\mathrm{conc}}\tau} + \varepsilon_{\mathrm{disc}}(h) + \varepsilon_{\mathrm{mix}}(B,S,M,h) + \varepsilon_{\mathrm{MC}}. \tag{12}$$

Define $\varepsilon_{\mathrm{tot}} = \varepsilon_{\mathrm{model}} + K_L\sqrt{C_{\mathrm{conc}}\tau} + \varepsilon_{\mathrm{disc}}(h) + \varepsilon_{\mathrm{mix}}(B,S,M,h) + \varepsilon_{\mathrm{MC}}$. If $\varepsilon_{\mathrm{tot}} \le \delta_\star/4$, define $\widehat{\mathbf{P}}_k = \mathbf{1}_{\mathcal{I}^+_\star}\left(\widehat{\mathbf{L}}_{\tau,h,B,S,M}\right)$. Then $\mathrm{rank}(\widehat{\mathbf{P}}_k) = k$, and

$$\left\|\widehat{\mathbf{P}}_k - \mathbf{P}_\star\right\|_2 \le \frac{\varepsilon_{\mathrm{tot}}}{\delta_\star - \varepsilon_{\mathrm{tot}}} \le \frac{2\varepsilon_{\mathrm{tot}}}{\delta_\star}. \tag{13}$$

The corollary is conditional on the existence of a stationary law $\pi_{\tau,h}$ for the finite-step transition kernel. Uniqueness of $\pi_{\tau,h}$, irreducibility or geometric ergodicity of the discrete chain, and invariance of the Gibbs law π_τ at finite step size are not required for the bound (Meyn and Tweedie, 2009).

Theorem 5 gives a conditional stability guarantee for the effective operator relative to $\mathbf{L}_\star$ under representability and equilibrium concentration. It quantifies the concentration cost of finite temperature, while Corollary 6 makes the continuous-to-discrete gap explicit. Whether positive temperature improves exploration or corrupted-edge detection enough to offset this concentration cost is an empirical question.

Thus, temperature is a bias–exploration control rather than a guaranteed source of predictive improvement. The theory quantifies the concentration cost $K_L\sqrt{C_{\mathrm{conc}}\tau}$ without assuming a compensating exploration benefit. In applications, τ can be selected using validation-only criteria together with mixing and temperature-sensitivity diagnostics, without using test performance (Gelman et al., 2013).

5.4 Frozen-geometry special case

Freeze a symmetric conductance matrix $\mathbf{C}$ and remove the Stiefel constraint. Assume that $\mathbf{H}_{\mathrm{init}}$ is deterministic or independent of the driving Brownian motion. With the fixed source anchor $\mathbf{H}_{\mathrm{ref}}$ and initial state $\mathbf{H}_{\mathrm{init}}$, consider $\mathrm{d}\mathbf{H}_t = -\left[(\mathbf{L}_{\mathbf{C}} + \eta\mathbf{I}_n)\mathbf{H}_t - \eta\mathbf{H}_{\mathrm{ref}}\right]\mathrm{d}t + \sqrt{2\tau}\,\mathrm{d}\mathbf{B}_t$. If $\mathbf{L}_{\mathbf{C}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$, with $\boldsymbol{\phi}_i$ denoting the ith column of $\mathbf{Q}$, $\mathbf{a}_i(t) = \boldsymbol{\phi}_i^\top\mathbf{H}_t$, and $\mathbf{a}_{\mathrm{ref},i} = \boldsymbol{\phi}_i^\top\mathbf{H}_{\mathrm{ref}}$, then for every mode with $\alpha_i = \lambda_i + \eta > 0$ (Pavliotis, 2014),

$$\mathbb{E}[\mathbf{a}_i(t)] = e^{-\alpha_i t}\mathbb{E}[\mathbf{a}_i(0)] + \frac{\eta}{\alpha_i}\left(1 - e^{-\alpha_i t}\right)\mathbf{a}_{\mathrm{ref},i}, \tag{14}$$

$$\mathrm{Cov}[\mathbf{a}_i(t)] = e^{-2\alpha_i t}\,\mathrm{Cov}[\mathbf{a}_i(0)] + \frac{\tau}{\alpha_i}\left(1 - e^{-2\alpha_i t}\right)\mathbf{I}_p. \tag{15}$$

For deterministic initialization, $\mathbb{E}[\mathbf{a}_i(0)] = \mathbf{a}_i(0)$. If $\eta > 0$, every mode is stable and the stationary mean applies $g(\lambda) = \eta/(\lambda+\eta)$, while the stationary covariance is $\tau\mathbf{I}_p/(\lambda+\eta)$. If $\eta = 0$, every zero-Laplacian mode is an unconstrained Brownian motion and has no stationary distribution. These formulas are diagnostic only: $\mathbf{L}_{\mathbf{C}}$ is fixed and no geometry is learned.

6 Experimental Study

In this section, we evaluate reference-geometry recovery, injected-edge detection, and linear-probe robustness under controlled edge corruption, following standard graph-structure learning and robustness settings (Kalofolias, 2016; Jin et al., 2020), with results reported over five graph seeds or public data splits. Corruptions add uniformly sampled nonedges at rates of 10%, 30%, and 50% relative to the clean edge count, without using class labels. Synthetic generators provide known clean supports; for Cora (Yang et al., 2016) and Roman-Empire (Platonov et al., 2023), the released graph serves as the pre-corruption reference rather than an assumed latent ground truth.

Table 1: **Reference-geometry and injected-edge recovery at 30% corruption.** Lower L error and higher AUROC. Fixed diffusion has tied edge scores, hence AUROC $= 0.500$. Bold/underlined denote best/second-best; red indicates negative transfer relative to fixed diffusion in L error.

Dataset	Method	Relative L error ↓	Added-edge AUROC ↑
SBM	Fixed diffusion	0.354 ± 0.007	0.500 ± 0.000
	Feature reweighting	0.205 ± 0.004	0.573 ± 0.032
	SGF-$\tau = 0$	0.197 ± 0.009	0.704 ± 0.026
	SGF	**0.186 ± 0.012**	**0.729 ± 0.024**
Random geometric	Fixed diffusion	0.366 ± 0.017	0.500 ± 0.000
	Feature reweighting	0.211 ± 0.015	0.681 ± 0.028
	SGF-$\tau = 0$	**0.192 ± 0.008**	0.872 ± 0.020
	SGF	0.201 ± 0.014	**0.886 ± 0.023**
Swiss roll	Fixed diffusion	0.368 ± 0.004	0.500 ± 0.000
	Feature reweighting	0.201 ± 0.010	0.748 ± 0.020
	SGF-$\tau = 0$	0.195 ± 0.013	0.886 ± 0.025
	SGF	**0.183 ± 0.016**	**0.899 ± 0.027**
Cora	Fixed diffusion	0.283 ± 0.001	0.500 ± 0.000
	Feature reweighting	0.309 ± 0.003	0.716 ± 0.006
	SGF-$\tau = 0$	0.219 ± 0.002	0.798 ± 0.005
	SGF	**0.205 ± 0.004**	**0.814 ± 0.009**
Roman-Empire	Fixed diffusion	0.450 ± 0.002	0.500 ± 0.000
	Feature reweighting	0.486 ± 0.001	0.519 ± 0.005
	SGF-$\tau = 0$	0.439 ± 0.003	0.574 ± 0.007
	SGF	**0.428 ± 0.005**	**0.585 ± 0.014**

Table 2: **Linear-probe robustness to random edge additions.** Accuracy after adding sampled nonedges at the indicated fraction of clean edges. Bold/underlined denote best/second-best means.

Dataset	Method	Clean ↑	Add 10% ↑	Add 30% ↑	Add 50% ↑
SBM	Fixed diffusion	0.949 ± 0.051	0.928 ± 0.055	0.884 ± 0.062	0.852 ± 0.048
	Feature reweighting	0.933 ± 0.054	0.911 ± 0.057	0.873 ± 0.066	0.839 ± 0.037
	SGF-$\tau = 0$	0.955 ± 0.040	0.937 ± 0.042	0.903 ± 0.050	0.875 ± 0.049
	SGF	**0.957 ± 0.038**	**0.940 ± 0.044**	**0.912 ± 0.047**	**0.882 ± 0.045**
Random geometric	Fixed diffusion	0.937 ± 0.042	0.916 ± 0.047	0.908 ± 0.043	0.887 ± 0.048
	Feature reweighting	0.925 ± 0.057	0.908 ± 0.041	0.898 ± 0.049	0.884 ± 0.054
	SGF-$\tau = 0$	**0.942 ± 0.031**	0.924 ± 0.035	0.914 ± 0.038	0.900 ± 0.041
	SGF	0.941 ± 0.034	**0.927 ± 0.036**	**0.918 ± 0.035**	**0.904 ± 0.039**
Swiss roll	Fixed diffusion	0.893 ± 0.053	0.861 ± 0.050	0.820 ± 0.067	0.784 ± 0.059
	Feature reweighting	0.880 ± 0.068	0.845 ± 0.055	0.798 ± 0.061	0.778 ± 0.072
	SGF-$\tau = 0$	0.898 ± 0.046	0.876 ± 0.044	0.840 ± 0.052	0.805 ± 0.053
	SGF	**0.902 ± 0.042**	**0.881 ± 0.046**	**0.848 ± 0.048**	**0.813 ± 0.051**
Cora	Fixed diffusion	0.763 ± 0.004	0.752 ± 0.009	0.735 ± 0.003	0.720 ± 0.005
	Feature reweighting	0.749 ± 0.006	0.740 ± 0.008	0.732 ± 0.004	0.724 ± 0.007
	SGF-$\tau = 0$	0.767 ± 0.009	0.757 ± 0.002	0.743 ± 0.008	0.731 ± 0.001
	SGF	**0.768 ± 0.008**	**0.759 ± 0.007**	**0.746 ± 0.003**	**0.734 ± 0.004**
Roman-Empire	Fixed diffusion	0.364 ± 0.007	0.361 ± 0.006	0.356 ± 0.008	0.351 ± 0.004
	Feature reweighting	0.397 ± 0.006	0.392 ± 0.004	0.389 ± 0.007	0.388 ± 0.005
	SGF-$\tau = 0$	0.400 ± 0.008	0.395 ± 0.010	0.390 ± 0.003	**0.389 ± 0.005**
	SGF	**0.401 ± 0.002**	**0.397 ± 0.001**	**0.393 ± 0.004**	0.388 ± 0.006

Table 3: **Mechanism comparison and diagnostics.** (a) Corrupted-SBM mechanism comparisons; bold/underlined denote best/second-best means. (b) Runtime and finite-chain diagnostics: ESS and split-$\widehat{R}$ use the retained mean-conductance trace (Vehtari et al., 2021); Δ_h, D_{init}, and ΔAUROC_u quantify step-size, initialization, and variance-contribution effects. Reported $\pm$ are SDs across five repeats, with each diagnostic computed within repeat. † marks values shared with main tables.

(a) Mechanism comparison

Variant	Adaptive C	Noise	Stiefel	Fidelity	Rel. L error ↓	Edge AUROC ↑	LP accuracy ↑
Fixed diffusion†	×	×	×	✓	0.354 ± 0.007	0.500 ± 0.000	0.884 ± 0.062
SGF-$\tau = 0^\dagger$	✓	×	✓	✓	0.197 ± 0.009	0.704 ± 0.026	0.903 ± 0.050
Euclidean noise	✓	Euclidean	×	✓	0.272 ± 0.014	0.681 ± 0.023	0.896 ± 0.051
SGF without fidelity	✓	intrinsic	✓	×	0.219 ± 0.008	0.712 ± 0.029	0.901 ± 0.048
SGF†	✓	intrinsic	✓	✓	**0.186 ± 0.012**	**0.729 ± 0.024**	**0.912 ± 0.047**

(b) Efficiency and finite-chain diagnostics

Dataset	Method	Time/run (s) ↓	ESS ↑	Split-$\widehat{R} \to 1$	Δ_h ↓	D_{init} ↓	ΔAUROC_u
Cora	Fixed diffusion	0.008 ± 0.000	–	–	–	–	–
	Feature reweighting	0.007 ± 0.000	–	–	–	–	–
	SGF-$\tau = 0$	0.113 ± 0.002	–	–	–	–	–
	SGF	0.31 ± 0.02	22.4 ± 2.8	1.02 ± 0.01	0.031 ± 0.007	0.046 ± 0.013	0.009 ± 0.005
Roman-Empire	Fixed diffusion	0.02 ± 0.00	–	–	–	–	–
	Feature reweighting	0.02 ± 0.00	–	–	–	–	–
	SGF-$\tau = 0$	1.27 ± 0.01	–	–	–	–	–
	SGF	3.55 ± 0.13	19.7 ± 2.6	1.04 ± 0.02	0.044 ± 0.010	0.062 ± 0.018	0.007 ± 0.003

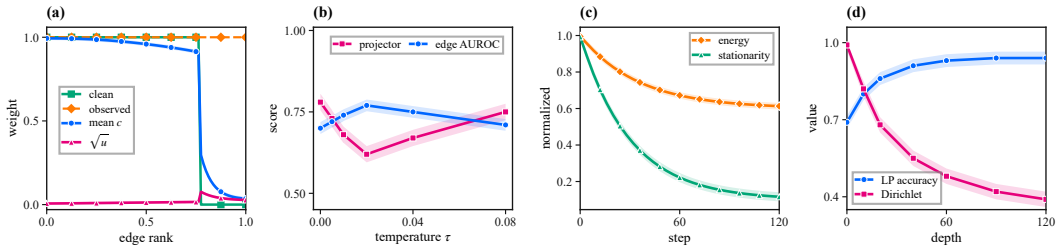


Figure 3: **SGF diagnostics.** Means and ranges over five SBM seeds: (a) edge profiles, (b) temperature sensitivity, (c) energy and stationarity, and (d) early-depth response.

Table 4: **Multi-baseline comparison at** 30% **random-edge addition.** Linear-probe accuracy on matched splits/labels; baselines are contextual. Bold/underlined denote best/second-best means; formatting follows unrounded means. † marks results transcribed from Table 2.

Method	SBM	Random geometric	Swiss roll	Cora	Roman-Empire
GRAND (Chamberlain et al., 2021)	0.892 ± 0.055	0.910 ± 0.041	0.826 ± 0.057	0.739 ± 0.003	0.371 ± 0.000
GRAND++ (Thorpe et al., 2022)	0.899 ± 0.052	0.910 ± 0.037	0.832 ± 0.053	0.742 ± 0.006	0.378 ± 0.001
LGNSDE (Bergna et al., 2025)	0.901 ± 0.046	0.913 ± 0.032	0.835 ± 0.051	0.744 ± 0.004	0.382 ± 0.006
Pro-GNN (Jin et al., 2020)	$\underline{0.906 \pm 0.044}$	$\underline{0.915 \pm 0.036}$	$\underline{0.842 \pm 0.049}$	$\underline{0.746 \pm 0.005}$	0.388 ± 0.002
SGF-$\tau = 0^{\dagger}$	0.903 ± 0.050	0.914 ± 0.038	0.840 ± 0.052	0.743 ± 0.008	$\underline{0.390 \pm 0.003}$
SGF†	$\mathbf{0.912 \pm 0.047}$	$\mathbf{0.918 \pm 0.035}$	$\mathbf{0.848 \pm 0.048}$	$\mathbf{0.746 \pm 0.003}$	$\mathbf{0.393 \pm 0.004}$

Summary. Table 1 shows that static feature reweighting can increase operator error on Cora and Roman-Empire, whereas SGF reduces it; positive-temperature changes are smaller and dataset-dependent. Table 2 shows modest robustness gains, while Table 4 provides baseline context on matched splits/labels. Table 3 and Figure 3 report cost and finite-chain diagnostics, including step-size and initialization sensitivity and the contribution of conductance variance. The results support the adaptive-geometry mechanism without implying universal gains from positive temperature.

7 Limitations

The spectral result is a conditional stability guarantee, not an identification theorem: it assumes local operator approximation and Gibbs concentration near a selected orbit. SGF cannot recover edges outside the candidate support and remains subject to discretization, burn-in, mixing, and sampling error (Meyn and Tweedie, 2009). The chains show local between-chain agreement without establishing equilibrium convergence. Learning and geometry-estimation experiments reach 22,662 nodes, while the propagation-kernel benchmark reaches one million. Positive temperature adds sampling cost and does not consistently improve point estimates, while the fixed $p = 3$ width and Stiefel constraint may limit representational capacity. Finally, conductance variance represents model-based geometric uncertainty, not causal or automatically calibrated evidence; learned conductances therefore require validation of the features, support, assumptions, and relevant failure modes.

8 Conclusion

We presented SGF as an energy-consistent closed-loop formulation in which latent conductances are eliminated exactly, the resulting bounded robust energy generates a state-dependent graph drift, and the continuous-state Gibbs law induces a distribution over conductance geometry on the observed support. The centered scaled Stiefel manifold provides a constrained non-collapse state space (Absil et al., 2008), while intrinsic Langevin dynamics define the positive-temperature law used to estimate conductance means and variances (Hsu, 2002; Pavliotis, 2014). The theory establishes exact variational reduction, probability-level free-energy dissipation, pointwise attenuation under a representation margin, and conditional operator and subspace stability, while separating model, temperature, discretization, mixing, and Monte Carlo errors under explicit representability and concentration conditions. The contribution lies in this energy-consistent stochastic formulation, its conditional operator theory, and model-based conductance uncertainty; the finite-chain experiments support the adaptive-conductance mechanism within the reported sampling and scaling regimes.

AI USE STATEMENT

Generative AI tools were used only to assist with language editing. All AI-assisted edits were reviewed by the authors, and the mathematical claims, experimental results, and final manuscript were independently checked by the authors. The authors take full responsibility for the content of this work.

ETHICS STATEMENT

This work uses synthetic graphs and publicly available benchmark datasets and does not involve human-subject experiments or the collection of sensitive personal data. We are not aware of additional ethical concerns specific to the reported methodology or experiments.

REPRODUCIBILITY STATEMENT

We provide detailed assumptions, derivations, and proofs in the appendices, and document the experimental protocol, hyperparameters, preprocessing, sampling diagnostics, and evaluation settings in Appendix D. The supplementary material includes anonymous reference code, configuration files, validation checks, and implementation notes corresponding to the reported experiments.

REFERENCES

R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

E. P. Hsu. *Stochastic Analysis on Manifolds*. Graduate Studies in Mathematics, volume 38. American Mathematical Society, 2002.

R. Zhang, Y. Zhang, C. Lu, and X. Li. Unsupervised graph embedding via adaptive graph learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):5329–5336, 2023.

A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):1103–1130, 2016.

N. Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.

I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

W. Jin, Y. Ma, X. Liu, X. Tang, S. Wang, and J. Tang. Graph structure learning for robust graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 66–74, 2020.

P. H. Schönemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

V. Kalofolias. How to learn a graph from smooth signals. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pages 920–929, 2016.

M. Poli, S. Massaroli, J. Park, A. Yamashita, H. Asama, and J. Park. Graph neural ordinary differential equations. arXiv:1911.07532, 2019.

B. P. Chamberlain, J. Rowbottom, M. I. Gorinova, M. M. Bronstein, S. D. Webb, and E. Rossi. GRAND: Graph neural diffusion. In *Proceedings of the 38th International Conference on Machine Learning*, pages 1407–1418, 2021.

M. Thorpe, T. M. Nguyen, H. Xia, T. Strohmer, A. L. Bertozzi, S. J. Osher, and B. Wang. GRAND++: Graph neural diffusion with a source term. In *International Conference on Learning Representations*, 2022.

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

R. Bergna, F. Opolka, P. Liò, and J. M. Hernández-Lobato. Graph neural stochastic differential equations. arXiv:2308.12316, 2023.

R. Bergna, S. Calvo-Ordoñez, F. L. Opolka, P. Liò, and J. M. Hernández-Lobato. Uncertainty modeling in graph neural networks via stochastic differential equations. In *International Conference on Learning Representations*, 2025.

M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision*, 19(1):57–91, 1996.

C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

C. Holtz, P. Chen, A. Cloninger, C.-K. Cheng, and G. Mishne. Semi-supervised Laplace learning on Stiefel manifolds. arXiv:2308.00142, 2023.

Z. Yang, W. W. Cohen, and R. Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 40–48, 2016.

O. Platonov, D. Kuznedelev, A. Diskin, A. Babenko, and L. Prokhorenkova. A critical look at the evaluation of GNNs under heterophily: Are we really making progress? In *International Conference on Learning Representations*, 2023.

A. Manenti, D. Zambon, and C. Alippi. Learning latent graph structures and their uncertainty. In *Proceedings of the 42nd International Conference on Machine Learning*, pages 42882–42901, 2025.

B. Runwal, Vivek, and S. Kumar. Robust graph neural networks using weighted graph Laplacian. arXiv:2208.01853, 2022.

J. L. Doob. *Stochastic Processes*. John Wiley & Sons, 1953.

J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, 2000.

I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. 2nd edition, Springer, 1991.

R. Bhatia. *Matrix Analysis*. Springer, 1997.

F. Liu and Q. Wang. AdaSpec: Adaptive spectrum for enhanced node distinguishability. In *International Conference on Learning Representations*, 2026.

A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, 16(2):667–718, 2021.

A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 14:849–856, 2001.

P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

N. J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, 2008.

F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

Y.-A. Ma, T. Chen, and E. B. Fox. A complete recipe for stochastic gradient MCMC. In *Advances in Neural Information Processing Systems*, 2015.

M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, third edition, 2013.

Z. Li, W. Xu, X. Tong, X. Lu, and X. Zhao. Continuous limits of coupled flows in representation learning. arXiv:2604.16801, 2026.

S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, second edition, 2009.

G. A. Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*. Springer, 2014.

J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

B. P. Chamberlain, J. Rowbottom, D. Eynard, F. Di Giovanni, X. Dong, and M. M. Bronstein. Beltrami flow and neural diffusion on graphs. In *Advances in Neural Information Processing Systems*, 34:1594–1609, 2021.

C. J. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, 7(4):473–483, 1992.

P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

NOTATION

Symbols in red denote the principal SGF objects. Bold uppercase letters denote matrices unless stated otherwise.

Symbol	Meaning
$\mathcal{G}_{\mathrm{obs}} = (\mathcal{V}, \mathcal{E}_{\mathrm{obs}}, \mathbf{A})$	Observed graph, nodes, candidate edges, and weights.
n, d, p; $\mathbf{X} \in \mathbb{R}^{n \times d}$	Nodes, feature dimension, representation dimension; node features.
Φ_θ, θ; $\mathbf{Y}_{\mathrm{ref}}$	Encoder and parameters; encoder output.
$\mathbf{J}$; $\mathbf{H}_{\mathrm{ref}}$	Centering projector; centered scaled-orthogonal anchor.
$\mathbf{H}, \mathbf{h}_i$; $\mathcal{M}_{n,p}$	Representation state and row; centered scaled Stiefel manifold.
$T_{\mathbf{H}}\mathcal{M}_{n,p}$; $\Pi_{\mathbf{H}}$	Tangent space; orthogonal tangent projection.
A_{ij}; $c_{ij}, \mathbf{C}$; c_{ij}^{feat}	Observed edge weight; latent conductance/matrix; static feature-reweighting conductance.
$\mathbf{D}_{\mathbf{C}}, \mathbf{L}_{\mathbf{C}}$	Conductance degree matrix and Laplacian.
$c_{ij}^\star(\mathbf{H}), \mathbf{C}^\star(\mathbf{H})$; $\mathbf{L}(\mathbf{H})$	Eliminated conductance geometry; state-dependent Laplacian.
σ, σ_f, η; $\mathcal{E}$; $\mathcal{E}_{\mathrm{red}}$	SGF and feature-reweighting bandwidths, fidelity; joint and reduced energies.
$\rho_\sigma(s)$; ∇, grad	Robust edge penalty; Euclidean and Riemannian gradients.
$\mathrm{sym}, \mathrm{tr}, \mathrm{diag}$	Symmetrization, trace, and diagonal operators.
$\mathcal{Q}$; h, ℓ, L	Centered polar map; step size, step index, and reported integration-step count or depth.
$\mathbf{\Xi}^{(\ell)}, \mathbf{\Delta}^{(\ell)}, \mathbf{G}^{(\ell)}$	Gaussian increment, tangent increment, and drift term.
K, T	Number of chains; generic transition count in complexity and scaling estimates.
τ; $\mathbf{B}_t$	Temperature; matrix Brownian motion.
ρ_t, μ; Z_τ	State density, Riemannian volume; Gibbs normalizer.
$\mathcal{F}_\tau$; π_τ	Free energy; continuous SGF Gibbs law.
$\bar{c}_{ij}^{(\tau)}$; $u_{ij}^{(\tau)}$	Gibbs conductance mean and variance.
q_{ij}; s_{ij}^{corr}	Mean conductance ratio; edge corruption score.
$\kappa, \varepsilon_{\mathrm{stab}}$	Uncertainty weight and stabilization constant.
$\bar{\mathbf{L}}_\tau$; $\mathbf{P}_{\mathbf{H}}$	Gibbs mean Laplacian; gauge-invariant state projector.
$e_{\mathrm{in}}, e_{\mathrm{out}}$; $r_{\mathrm{in}}, r_{\mathrm{out}}, \Delta r$	Compared edges; squared-distance bounds and margin.
$\mathbf{L}_\star$; $\mathcal{I}_\star = [a_\star, b_\star]$	Latent reference Laplacian; eigenvalue interval.
$\mathbf{P}_\star, k$; $\delta_\star, \mathcal{I}_\star^+$	Target projector and rank; spectral gap and enlarged interval.
$\mathbf{H}^\dagger, \mathcal{O}^\dagger$; $d_{\mathcal{O}}$	Reference state, orthogonal orbit, and quotient distance.
C_{conc}, K_L; $\varepsilon_{\mathrm{model}}, \varepsilon_{\mathrm{eq}}$	Concentration and Lipschitz constants; model and equilibrium errors.
$\bar{\mathbf{P}}_{\tau,k}$	Equilibrium mean-operator spectral projector.
$\pi_{\tau,h}, \mu_{\tau,h,t}$; B, S, M	Discrete stationary/time-t laws; burn-in, spacing, retained samples.
$\bar{\mathbf{L}}_{\tau,h}$; $\mathbf{L}_{\tau,h,B,S,M}^{\mathrm{mean}}$	Discrete stationary and finite-schedule mean Laplacians.
$\widehat{\mathbf{L}}_{\tau,h,B,S,M}$	Monte Carlo Laplacian estimator.
$\varepsilon_{\mathrm{disc}}, \varepsilon_{\mathrm{mix}}, \varepsilon_{\mathrm{MC}}$; $\varepsilon_{\mathrm{tot}}$	Discretization, mixing, Monte Carlo, and total errors.
δ; $\widehat{\mathbf{P}}_k$	Failure probability; estimated spectral projector.
$e_{\mathrm{op},F}$; $\widehat{c}_{ij}, \widehat{u}_{ij}$	Relative operator-recovery error; empirical conductance mean and variance.
$\widehat{\mathbf{c}}_h$; $\widehat{\mathbf{P}}^{(a)}$	Vector of empirical mean conductances at step size h; retained mean projector for chain a.
$\Delta_h, D_{\mathrm{init}}, \Delta\mathrm{AUROC}_u$	Step-size sensitivity, initialization sensitivity, and incremental AUROC contribution of conductance variance.
$\mathrm{ESS}, \widehat{R}$	Effective sample size and split-$\widehat{R}$.
$\mathbf{Q}, \mathbf{\Lambda}, \lambda_i, \boldsymbol{\phi}_i$	Laplacian eigenvectors/eigenvalues and ith eigenpair.
$\mathbf{a}_i(t), \mathbf{a}_{\mathrm{ref},i}$; $\alpha_i, g(\lambda)$	Spectral coordinates; modal rate and stationary filter.

$\mathbb{E}$, Var, Cov, spec, rank, and $\mathbf{1}_{\mathcal{I}}(\cdot)$ denote expectation, variance, covariance, spectrum, rank, and spectral projection; $\|\cdot\|_2$ and $\|\cdot\|_F$ denote spectral/Euclidean and Frobenius norms. Indices i, j, r, s denote nodes, t continuous time, m retained samples, and a, b chain indices where used.

APPENDIX CONTENTS

A Geometric Preliminaries . . . 15
A.1 Geometry of the Centered Scaled Stiefel Manifold . . . 15
A.2 Projected Ambient Noise and Intrinsic Brownian Motion . . . 15
A.3 Centered Polar Map and Weak Consistency . . . 16
A.4 Regularity and Quotient Lipschitz Control . . . 18

B Proofs of the Main Results . . . 19
B.1 Proof of the Variational Conductance Result . . . 19
B.2 Proof of Permutation Equivariance . . . 20
B.3 Proof of Free-Energy Dissipation and Gibbs Invariance . . . 20
B.4 Proof of Pointwise Edge Suppression . . . 21
B.5 Proof of Latent-Subspace Stability . . . 21
B.6 Proof for the Practical Finite-Step Estimator . . . 22
B.7 Frozen-Geometry Ornstein–Uhlenbeck Derivation . . . 23

C Additional Method Details . . . 23
C.1 Scope of the Stochastic Guarantees . . . 23
C.2 Normalized Laplacian Extension . . . 24
C.3 Graph-Preconditioned Noise . . . 25
C.4 Equilibrium Estimators and Sampling Error . . . 25
C.5 Gauge-Invariant Aggregation . . . 26
C.6 Computational Complexity and Numerical Safeguards . . . 28
C.7 Empirical Kernel Scaling and Scope . . . 28

D Experimental Details . . . 29
D.1 Separation of Training and Equilibrium Estimation . . . 29
D.2 Synthetic Data Generation . . . 29
D.3 Topology Corruption Models . . . 30
D.4 Real-Graph Preprocessing . . . 30
D.5 Matched Baselines . . . 30
D.6 Evaluation Metrics . . . 31
D.7 Equilibrium-Sampling Diagnostics . . . 31
D.8 Hyperparameter Search Space and Dataset . . . 32
D.9 Mechanism Comparisons . . . 33

A GEOMETRIC PRELIMINARIES

Throughout, $\langle \mathbf{A}, \mathbf{B} \rangle_F = \mathrm{tr}(\mathbf{A}^\top \mathbf{B})$ denotes the Frobenius inner product and $\mathbf{J} = \mathbf{I}_n - \mathbf{1}\mathbf{1}^\top/n$ is the centering projector. The state space is

$$\mathcal{M}_{n,p} = \left\{ \mathbf{H} \in \mathbb{R}^{n \times p} : \mathbf{1}^\top \mathbf{H} = 0,\ \mathbf{H}^\top \mathbf{H} = n\mathbf{I}_p \right\}, \qquad 1 \le p < n-1.$$

Every observed edge is unordered and appears once in sums over $\mathcal{E}_{\mathrm{obs}}$. By definition, $A_{ij} > 0$ on $\mathcal{E}_{\mathrm{obs}}$ and $A_{ij} = 0$ outside the observed support.

A.1 GEOMETRY OF THE CENTERED SCALED STIEFEL MANIFOLD

Let $\mathbf{U} \in \mathbb{R}^{n \times (n-1)}$ have orthonormal columns spanning $\mathbf{1}^\perp$, so $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_{n-1}$ and $\mathbf{U}\mathbf{U}^\top = \mathbf{J}$. Every $\mathbf{H} \in \mathcal{M}_{n,p}$ has the unique representation $\mathbf{H} = \sqrt{n}\,\mathbf{U}\mathbf{Y}$ with $\mathbf{Y}^\top \mathbf{Y} = \mathbf{I}_p$. Hence the map $\mathbf{Y} \mapsto \sqrt{n}\,\mathbf{U}\mathbf{Y}$ is a diffeomorphism from

$$\mathrm{St}(n-1, p) = \left\{ \mathbf{Y} \in \mathbb{R}^{(n-1) \times p} : \mathbf{Y}^\top \mathbf{Y} = \mathbf{I}_p \right\}$$

onto $\mathcal{M}_{n,p}$. It multiplies Frobenius inner products by n; equivalently, $\mathcal{M}_{n,p}$ is isometric to $\mathrm{St}(n-1, p)$ equipped with the scaled metric $n\langle \cdot, \cdot \rangle_F$ (Edelman et al., 1998; Absil et al., 2008). Since $n - 1 > p$, the manifold is compact and connected.

Differentiating the constraints in a direction $\mathbf{Z}$ gives the standard tangent-space characterization (Edelman et al., 1998; Absil et al., 2008),

$$T_{\mathbf{H}} \mathcal{M}_{n,p} = \left\{ \mathbf{Z} \in \mathbb{R}^{n \times p} : \mathbf{1}^\top \mathbf{Z} = 0,\ \mathbf{H}^\top \mathbf{Z} + \mathbf{Z}^\top \mathbf{H} = 0 \right\}.$$

Lemma 7 (Orthogonal tangent projection)**.** *For any $\mathbf{Z} \in \mathbb{R}^{n \times p}$ and $\mathbf{H} \in \mathcal{M}_{n,p}$, the Frobenius-orthogonal projection onto $T_{\mathbf{H}} \mathcal{M}_{n,p}$ is*

$$\Pi_{\mathbf{H}}(\mathbf{Z}) = \mathbf{J}\mathbf{Z} - \frac{1}{n}\mathbf{H}\,\mathrm{sym}(\mathbf{H}^\top \mathbf{J}\mathbf{Z}).$$

Proof. Set $\mathbf{G} = \mathbf{H}^\top \mathbf{J}\mathbf{Z}$. Since $\mathbf{1}^\top \mathbf{J} = 0$ and $\mathbf{1}^\top \mathbf{H} = 0$, $\mathbf{1}^\top \Pi_{\mathbf{H}}(\mathbf{Z}) = 0$. Using $\mathbf{J}\mathbf{H} = \mathbf{H}$ and $\mathbf{H}^\top \mathbf{H} = n\mathbf{I}_p$,

$$\mathbf{H}^\top \Pi_{\mathbf{H}}(\mathbf{Z}) + \Pi_{\mathbf{H}}(\mathbf{Z})^\top \mathbf{H} = \mathbf{G} + \mathbf{G}^\top - 2\,\mathrm{sym}(\mathbf{G}) = 0.$$

Thus the image is tangent. The residual is

$$\mathbf{Z} - \Pi_{\mathbf{H}}(\mathbf{Z}) = (\mathbf{I}_n - \mathbf{J})\mathbf{Z} + \frac{1}{n}\mathbf{H}\,\mathrm{sym}(\mathbf{G}).$$

For any tangent $\mathbf{T}$, the first term is orthogonal to $\mathbf{T}$ because $\mathbf{1}^\top \mathbf{T} = 0$. The second is also orthogonal because $\mathrm{sym}(\mathbf{G})$ is symmetric and $\mathbf{H}^\top \mathbf{T}$ is skew-symmetric:

$$\left\langle \frac{1}{n}\mathbf{H}\,\mathrm{sym}(\mathbf{G}), \mathbf{T} \right\rangle_F = \frac{1}{n}\,\mathrm{tr}\big(\mathrm{sym}(\mathbf{G})\mathbf{H}^\top \mathbf{T}\big) = 0.$$

□

For any smooth ambient extension of a scalar function E, the induced Riemannian gradient is $\mathrm{grad}\, E(\mathbf{H}) = \Pi_{\mathbf{H}}(\nabla E(\mathbf{H}))$. These identities justify the constrained state geometry and tangent projection used in Section 4.2, in particular Eq. (3).

A.2 PROJECTED AMBIENT NOISE AND INTRINSIC BROWNIAN MOTION

Let $\{\mathbf{E}_\alpha\}_{\alpha=1}^{np}$ be a Frobenius-orthonormal basis of $\mathbb{R}^{n \times p}$ and define the tangent vector fields $\mathbf{V}_\alpha(\mathbf{H}) = \Pi_{\mathbf{H}}(\mathbf{E}_\alpha)$.

Lemma 8 (Extrinsic Brownian identity)**.** *For every $\mathbf{H} \in \mathcal{M}_{n,p}$ and $\mathbf{Z} \in T_{\mathbf{H}} \mathcal{M}_{n,p}$,*

$$\sum_{\alpha=1}^{np} \langle \mathbf{V}_\alpha(\mathbf{H}), \mathbf{Z} \rangle_F^2 = \|\mathbf{Z}\|_F^2.$$

Moreover, for every smooth $f : \mathcal{M}_{n,p} \to \mathbb{R}$,

$$\sum_{\alpha=1}^{np} \mathbf{V}_\alpha(\mathbf{V}_\alpha f) = \Delta_{\mathcal{M}} f.$$

Consequently, the Stratonovich SDE

$$\mathrm{d}\mathbf{H}_t = \sum_\alpha \mathbf{V}_\alpha(\mathbf{H}_t) \circ \mathrm{d}B_t^\alpha$$

has generator $\frac{1}{2}\Delta_{\mathcal{M}}$ and is therefore an extrinsic representation of intrinsic Brownian motion (Hsu, 2002).

Proof. Because $\Pi_{\mathbf{H}}$ is self-adjoint and $\Pi_{\mathbf{H}}\mathbf{Z} = \mathbf{Z}$ for tangent $\mathbf{Z}$,

$$\sum_\alpha \langle \mathbf{V}_\alpha, \mathbf{Z}\rangle_F^2 = \sum_\alpha \langle \mathbf{E}_\alpha, \Pi_{\mathbf{H}}\mathbf{Z}\rangle_F^2 = \|\mathbf{Z}\|_F^2 .$$

Thus the projected fields form a Parseval frame of the tangent space.

For the second identity, choose a tubular neighborhood of $\mathcal{M}_{n,p}$ and a smooth ambient extension $\widetilde{f}$ that is constant along its normal fibers (Hsu, 2002). Then $\widetilde{f}|_{\mathcal{M}_{n,p}} = f$ and $\nabla\widetilde{f}$ is tangent on $\mathcal{M}_{n,p}$. Since $\mathbf{V}_\alpha = \Pi\mathbf{E}_\alpha$, direct differentiation gives

$$\begin{aligned}\sum_{\alpha=1}^{np} \mathbf{V}_\alpha(\mathbf{V}_\alpha f) = &\sum_{\alpha=1}^{np} \mathrm{Hess}^{\mathbb{R}^{n\times p}} \widetilde{f}(\mathbf{V}_\alpha, \mathbf{V}_\alpha) \\ &+ \left\langle \nabla\widetilde{f}, \sum_{\alpha=1}^{np} D_{\mathbf{V}_\alpha}\mathbf{V}_\alpha \right\rangle_F .\end{aligned}$$

The vector $\sum_\alpha D_{\mathbf{V}_\alpha}\mathbf{V}_\alpha$ is the trace of the second fundamental form, hence is normal to $\mathcal{M}_{n,p}$. Its inner product with the tangent vector $\nabla\widetilde{f}$ therefore vanishes. Because $\{\mathbf{V}_\alpha\}_{\alpha=1}^{np}$ is a Parseval frame of the tangent space, the first term is the trace of the ambient Hessian over $T_{\mathbf{H}}\mathcal{M}_{n,p}$. For a normal-fiber-constant extension this trace equals the intrinsic Hessian trace, and therefore

$$\sum_{\alpha=1}^{np} \mathbf{V}_\alpha(\mathbf{V}_\alpha f) = \mathrm{tr}\,\mathrm{Hess}_{\mathcal{M}} f = \Delta_{\mathcal{M}} f.$$

□

Lemma 8 justifies the projected-noise term in Eq. (4); with the factor $\sqrt{2\tau}$, its diffusion generator is $\tau\Delta_{\mathcal{M}}$.

A.3 Centered polar map and weak consistency

For every $\mathbf{Y}$ satisfying $\mathbf{Y}^\top\mathbf{J}\mathbf{Y} \succ 0$, define

$$\mathcal{Q}(\mathbf{Y}) = \sqrt{n}\,\mathbf{J}\mathbf{Y}\left(\mathbf{Y}^\top\mathbf{J}\mathbf{Y}\right)^{-1/2}.$$

Lemma 9 (Constraint preservation and second-order retraction)**.** *The map $\mathcal{Q}$ takes every admissible $\mathbf{Y}$ into $\mathcal{M}_{n,p}$. For $\mathbf{H} \in \mathcal{M}_{n,p}$ and $\mathbf{Z} \in T_{\mathbf{H}}\mathcal{M}_{n,p}$,*

$$\mathcal{Q}(\mathbf{H} + t\mathbf{Z}) = \mathbf{H} + t\mathbf{Z} - \frac{t^2}{2n}\mathbf{H}\mathbf{Z}^\top\mathbf{Z} + O(t^3). \tag{16}$$

The acceleration $-\mathbf{H}\mathbf{Z}^\top\mathbf{Z}/n$ is normal to the manifold. Hence the fiberwise map

$$R_{\mathbf{H}}(\mathbf{Z}) = \mathcal{Q}(\mathbf{H} + \mathbf{Z}), \qquad \mathbf{Z} \in T_{\mathbf{H}}\mathcal{M}_{n,p},$$

defines a second-order retraction under the induced Frobenius metric (Absil et al., 2008; Boumal, 2023).

Proof. Centering follows from $\mathbf{1}^\top \mathbf{J} = 0$. With $\mathbf{G} = \mathbf{Y}^\top \mathbf{J}\mathbf{Y} \succ 0$,

$$\mathcal{Q}(\mathbf{Y})^\top \mathcal{Q}(\mathbf{Y}) = n\mathbf{G}^{-1/2}\mathbf{G}\mathbf{G}^{-1/2} = n\mathbf{I}_p.$$

For tangent $\mathbf{Z}$, $\mathbf{J}\mathbf{H} = \mathbf{H}$, $\mathbf{J}\mathbf{Z} = \mathbf{Z}$, and $\mathbf{H}^\top \mathbf{Z} + \mathbf{Z}^\top \mathbf{H} = 0$, so

$$(\mathbf{H} + t\mathbf{Z})^\top \mathbf{J}(\mathbf{H} + t\mathbf{Z}) = n\mathbf{I}_p + t^2 \mathbf{Z}^\top \mathbf{Z}.$$

Using the analytic expansion of the principal inverse square root (Higham, 2008),

$$\left(n\mathbf{I}_p + t^2 \mathbf{Z}^\top \mathbf{Z}\right)^{-1/2} = n^{-1/2}\left[\mathbf{I}_p - \frac{t^2}{2n}\mathbf{Z}^\top \mathbf{Z}\right] + O(t^4).$$

Multiplication by $\sqrt{n}(\mathbf{H} + t\mathbf{Z})$ yields Eq. (16). For any tangent $\mathbf{T}$,

$$\left\langle \mathbf{H}\mathbf{Z}^\top \mathbf{Z}, \mathbf{T} \right\rangle_F = \operatorname{tr}\left(\mathbf{Z}^\top \mathbf{Z}\, \mathbf{H}^\top \mathbf{T}\right) = 0,$$

because $\mathbf{Z}^\top \mathbf{Z}$ is symmetric and $\mathbf{H}^\top \mathbf{T}$ is skew-symmetric. Thus the second derivative is normal. □

The second-order property is the geometric ingredient needed for the discrete stochastic update: at the origin of each tangent space, the second-order pullback of a smooth test function agrees with its Riemannian Hessian. Combined with the first two moments of the Gaussian tangent increment, this yields the standard local weak-generator consistency mechanism for SDE discretizations (Boumal, 2023; Kloeden and Platen, 1992).

Proposition 10 (One-step weak generator consistency)**.** *Let* $\mathbf{H}^+$ *be obtained from* $\mathbf{H} \in \mathcal{M}_{n,p}$ *by*

$$\mathbf{H}^+ = \mathcal{Q}\Big(\mathbf{H} - h \operatorname{grad} \mathcal{E}_{\mathrm{red}}(\mathbf{H}) + \sqrt{2\tau h}\, \Pi_{\mathbf{H}}(\boldsymbol{\Xi})\Big),$$

where $\boldsymbol{\Xi}$ *has independent standard normal entries. For every* $f \in C^3(\mathcal{M}_{n,p})$*,*

$$\frac{\mathbb{E}[f(\mathbf{H}^+) \mid \mathbf{H}] - f(\mathbf{H})}{h} = -\langle \operatorname{grad} \mathcal{E}_{\mathrm{red}}, \operatorname{grad} f \rangle_F + \tau \Delta_{\mathcal{M}} f + o(1)$$

as $h \downarrow 0$*.*

Proof. Set

$$\mathbf{b}(\mathbf{H}) = \operatorname{grad} \mathcal{E}_{\mathrm{red}}(\mathbf{H}), \qquad \boldsymbol{\zeta}_h = -h\mathbf{b}(\mathbf{H}) + \sqrt{2\tau h}\, \Pi_{\mathbf{H}}(\boldsymbol{\Xi}).$$

Both $\mathbf{b}(\mathbf{H})$ and $\Pi_{\mathbf{H}}(\boldsymbol{\Xi})$ belong to $T_{\mathbf{H}}\mathcal{M}_{n,p}$, and hence $\boldsymbol{\zeta}_h \in T_{\mathbf{H}}\mathcal{M}_{n,p}$ for every realization. Therefore

$$\begin{aligned}(\mathbf{H} + \boldsymbol{\zeta}_h)^\top \mathbf{J}(\mathbf{H} + \boldsymbol{\zeta}_h) &= \mathbf{H}^\top \mathbf{H} + \mathbf{H}^\top \boldsymbol{\zeta}_h + \boldsymbol{\zeta}_h^\top \mathbf{H} + \boldsymbol{\zeta}_h^\top \boldsymbol{\zeta}_h \\ &= n\mathbf{I}_p + \boldsymbol{\zeta}_h^\top \boldsymbol{\zeta}_h \succ 0.\end{aligned}$$

Thus the centered polar map is admissible for every realization of the Gaussian increment; no rejection step is required.

Define the retraction pullback

$$g_{\mathbf{H}}(\mathbf{Z}) = f(\mathcal{Q}(\mathbf{H} + \mathbf{Z})), \qquad \mathbf{Z} \in T_{\mathbf{H}}\mathcal{M}_{n,p}.$$

By compactness of $\mathcal{M}_{n,p}$ and smoothness of the centered polar map, there exist constants $r > 0$ and $C < \infty$, independent of $\mathbf{H}$, such that, whenever $\|\mathbf{Z}\|_F \le r$,

$$g_{\mathbf{H}}(\mathbf{Z}) = g_{\mathbf{H}}(\mathbf{0}) + \mathrm{D}g_{\mathbf{H}}(\mathbf{0})[\mathbf{Z}] + \frac{1}{2}\mathrm{D}^2 g_{\mathbf{H}}(\mathbf{0})[\mathbf{Z}, \mathbf{Z}] + R_{\mathbf{H}}(\mathbf{Z}),$$

with $|R_{\mathbf{H}}(\mathbf{Z})| \le C\, \|\mathbf{Z}\|_F^3$.

Let $\mathcal{A}_h = \{\|\boldsymbol{\zeta}_h\|_F \le r\}$. The drift is uniformly bounded on the compact manifold, and $\Pi_{\mathbf{H}}$ is an orthogonal projection. Standard finite-dimensional Gaussian tail bounds therefore give constants $C_1, C_2 > 0$ such that $\Pr(\mathcal{A}_h^c) \le C_1 \exp(-C_2/h)$ for all sufficiently small h, uniformly in $\mathbf{H}$ (Kloeden and Platen, 1992). Since f is bounded on $\mathcal{M}_{n,p}$, the contribution of $\mathcal{A}_h^c$ to $\mathbb{E}[f(\mathbf{H}^+) - f(\mathbf{H}) \mid \mathbf{H}]$ is $o(h)$. Moreover, by Cauchy–Schwarz and the same Gaussian tail bound, replacing the first- and second-order moments truncated to $\mathcal{A}_h$ by their full Gaussian moments changes the expectation by $o(h)$. Hence the truncation indicators may be omitted below.

On $\mathcal{A}_h$, the second-order retraction property implies that the pullback agrees with the intrinsic first- and second-order geometry at the origin (Absil et al., 2008; Boumal, 2023):

$$\mathrm{D}g_{\mathbf{H}}(\mathbf{0})[\mathbf{Z}] = \langle \operatorname{grad} f(\mathbf{H}), \mathbf{Z}\rangle_F, \qquad \mathrm{D}^2 g_{\mathbf{H}}(\mathbf{0})[\mathbf{Z},\mathbf{Z}] = \operatorname{Hess}_{\mathcal{M}} f(\mathbf{H})[\mathbf{Z},\mathbf{Z}].$$

Write

$$\Pi_{\mathbf{H}}(\boldsymbol{\Xi}) = \sum_{\alpha=1}^{np} \xi_\alpha \mathbf{V}_\alpha(\mathbf{H}), \qquad \xi_\alpha \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$$

The first-order term satisfies

$$\mathbb{E}[\mathrm{D}g_{\mathbf{H}}(\mathbf{0})[\boldsymbol{\zeta}_h] \mid \mathbf{H}] = -h\langle \operatorname{grad} \mathcal{E}_{\mathrm{red}}, \operatorname{grad} f\rangle_F.$$

Using $\mathbb{E}[\xi_\alpha \xi_\beta] = \delta_{\alpha\beta}$, the second-order term is

$$\begin{aligned}\frac{1}{2}\mathbb{E}\big[\mathrm{D}^2 g_{\mathbf{H}}(\mathbf{0})[\boldsymbol{\zeta}_h,\boldsymbol{\zeta}_h] \mid \mathbf{H}\big] &= \tau h \sum_{\alpha=1}^{np} \operatorname{Hess}_{\mathcal{M}} f[\mathbf{V}_\alpha,\mathbf{V}_\alpha] + O(h^2)\\ &= \tau h\, \Delta_{\mathcal{M}} f + O(h^2),\end{aligned}$$

where the last equality follows from the tangent Parseval identity in Lemma 8, equivalently from the trace representation of the Laplace–Beltrami operator. Finally, $\mathbb{E}\,\|\boldsymbol{\zeta}_h\|_F^3 = O(h^{3/2})$, so the Taylor remainder is $o(h)$. Combining the local expansion and the exponentially small tail contribution yields

$$\mathbb{E}[f(\mathbf{H}^+) \mid \mathbf{H}] - f(\mathbf{H}) = h\,[-\langle \operatorname{grad} \mathcal{E}_{\mathrm{red}}, \operatorname{grad} f\rangle_F + \tau \Delta_{\mathcal{M}} f] + o(h),$$

which proves the claim. □

Together, Lemma 9 and Proposition 10 justify the retraction-based update used in Section 4.4: the centered polar map preserves the representation constraints exactly, while the one-step discrete generator converges to the intrinsic SGF generator in Eq. (4) as $h \downarrow 0$. This is an infinitesimal weak-consistency statement, not an exact finite-step invariance result. In particular, a finite-h chain need not preserve the continuous Gibbs law π_τ exactly; the resulting continuous-to-discrete discrepancy is represented by $\varepsilon_{\mathrm{disc}}(h)$ in Corollary 6.

A.4 Regularity and quotient Lipschitz control

For every observed edge $\{i,j\} \in \mathcal{E}_{\mathrm{obs}}$, the conductance

$$c_{ij}^\star(\mathbf{H}) = A_{ij} \exp\left(-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|_2^2}{2\sigma^2}\right)$$

is smooth in $\mathbf{H}$ and satisfies

$$0 < c_{ij}^\star(\mathbf{H}) \le A_{ij}.$$

Let

$$d_{\max}(\mathbf{A}) = \max_i \sum_j A_{ij}.$$

Since

$$\sum_j c_{ij}^\star(\mathbf{H}) \le \sum_j A_{ij} \le d_{\max}(\mathbf{A}),$$

the standard weighted-Laplacian bound gives

$$\|\mathbf{L}(\mathbf{H})\|_2 \le 2 d_{\max}(\mathbf{A})$$

uniformly on $\mathcal{M}_{n,p}$ (Chung, 1997).

The map $\mathbf{H} \mapsto \mathbf{L}(\mathbf{H})$ is smooth. Since $\mathcal{M}_{n,p}$ is compact, its differential has uniformly bounded operator norm; in particular, there exists $K_L < \infty$ such that

$$\sup_{\mathbf{H}\in\mathcal{M}_{n,p}} \|\mathrm{D}\mathbf{L}(\mathbf{H})\|_{\mathrm{op}} \le K_L,$$

where the operator norm is taken from the tangent-space Frobenius norm to the matrix spectral norm. Integrating this differential bound along a minimizing geodesic between $\mathbf{H}$ and $\mathbf{G}$ yields (Boumal, 2023)

$$\|\mathbf{L}(\mathbf{H}) - \mathbf{L}(\mathbf{G})\|_2 \le K_L \operatorname{dist}_{\mathcal{M}}(\mathbf{H},\mathbf{G})$$

for all $\mathbf{H}, \mathbf{G} \in \mathcal{M}_{n,p}$.

Since pairwise row distances are invariant under right multiplication by an orthogonal matrix,

$$\|(\mathbf{h}_i - \mathbf{h}_j)\mathbf{R}\|_2 = \|\mathbf{h}_i - \mathbf{h}_j\|_2 ,$$

we have

$$\mathbf{C}^\star(\mathbf{GR}) = \mathbf{C}^\star(\mathbf{G}), \qquad \mathbf{L}(\mathbf{GR}) = \mathbf{L}(\mathbf{G})$$

for every $\mathbf{R} \in \mathrm{O}(p)$. Therefore, applying the preceding Lipschitz bound with $\mathbf{G} = \mathbf{H}^\dagger\mathbf{R}$ and taking the infimum over $\mathbf{R} \in \mathrm{O}(p)$ gives

$$\left\|\mathbf{L}(\mathbf{H}) - \mathbf{L}(\mathbf{H}^\dagger)\right\|_2 \le K_L d_{\mathcal{O}}(\mathbf{H}, \mathbf{H}^\dagger).$$

This quotient-level Lipschitz estimate is the regularity input used in Section 5.3. In particular, under Assumption 1, it converts Gibbs concentration of the representation state around the orbit $\mathcal{O}^\dagger$ into operator concentration of $\mathbf{L}(\mathbf{H})$ around $\mathbf{L}(\mathbf{H}^\dagger)$. This yields the $K_L\sqrt{C_{\mathrm{conc}}\tau}$ term appearing in Theorem 5 and, subsequently, in the finite-step error decomposition of Corollary 6.

B Proofs of the Main Results

B.1 Proof of Proposition 1

Fix $\mathbf{H} \in \mathcal{M}_{n,p}$ and abbreviate $d_{ij}^2 = \|\mathbf{h}_i - \mathbf{h}_j\|_2^2$. The joint objective separates over observed edges. For one edge,

$$f_{ij}(c) = \frac{1}{2} c\, d_{ij}^2 + \sigma^2 \left(c \log \frac{c}{A_{ij}} - c + A_{ij} \right), \qquad c > 0.$$

Its derivatives are

$$f'_{ij}(c) = \frac{1}{2} d_{ij}^2 + \sigma^2 \log \frac{c}{A_{ij}}, \qquad f''_{ij}(c) = \frac{\sigma^2}{c} > 0.$$

Thus f_{ij} is strictly convex. Since $f'_{ij}(c) \to -\infty$ as $c \downarrow 0$ and $f'_{ij}(c) \to +\infty$ as $c \to \infty$, the unique stationary point is the global minimizer: $c_{ij}^\star(\mathbf{H}) = A_{ij} \exp\!\left(-\frac{d_{ij}^2}{2\sigma^2}\right)$. At this point, $\log(c_{ij}^\star / A_{ij}) = -d_{ij}^2/(2\sigma^2)$, and

$$\begin{aligned} f_{ij}(c_{ij}^\star) &= \frac{1}{2} c_{ij}^\star d_{ij}^2 + \sigma^2 \left(-\frac{c_{ij}^\star d_{ij}^2}{2\sigma^2} - c_{ij}^\star + A_{ij} \right) \\ &= \sigma^2 (A_{ij} - c_{ij}^\star) \\ &= \sigma^2 A_{ij} \left[1 - \exp\!\left(-\frac{d_{ij}^2}{2\sigma^2} \right) \right]. \end{aligned}$$

Summing over $\mathcal{E}_{\mathrm{obs}}$ and adding the fidelity term gives Eq. (2). This elimination is consistent with the classical latent-weight representation of robust penalties (Black and Rangarajan, 1996): the auxiliary conductance is optimized out exactly, leaving a bounded penalty in the representation discrepancy.

For the gradient, let $\rho_\sigma(s) = \sigma^2(1 - e^{-s/(2\sigma^2)})$. Since $\rho'_\sigma(s) = \frac{1}{2} e^{-s/(2\sigma^2)}$, the derivative of the edge term with respect to $\mathbf{h}_i$ is

$$2A_{ij}\rho'_\sigma(d_{ij}^2)(\mathbf{h}_i - \mathbf{h}_j) = c_{ij}^\star(\mathbf{H})(\mathbf{h}_i - \mathbf{h}_j).$$

Summing incident edges yields the i-th row of $\mathbf{L}(\mathbf{H})\mathbf{H}$. The fidelity derivative is $\eta(\mathbf{H} - \mathbf{H}_{\mathrm{ref}})$, proving the stated gradient. Equivalently, the envelope theorem for a unique interior optimizer gives the same derivative by differentiating the joint objective with respect to $\mathbf{H}$ at $\mathbf{C} = \mathbf{C}^\star(\mathbf{H})$ (Bonnans and Shapiro, 2000). Terms involving derivatives of $\mathbf{C}^\star(\mathbf{H})$ vanish by the first-order optimality condition.

Thus, this proof establishes both parts of the variational closure used in Section 4.1: the same optimizer $\mathbf{C}^\star(\mathbf{H})$ that defines the adaptive conductance geometry also generates the state-dependent Laplacian through

$$\nabla \mathcal{E}_{\mathrm{red}}(\mathbf{H}) = \mathbf{L}(\mathbf{H})\mathbf{H} + \eta(\mathbf{H} - \mathbf{H}_{\mathrm{ref}}).$$

Consequently, the conductance update and representation drift are not independent prescriptions; they arise from the same reduced energy, which is the closure used by the SGF dynamics in Section 4.2.

B.2 Proof of Proposition 2

Let $\widetilde{\mathbf{A}} = \mathbf{S}\mathbf{A}\mathbf{S}^\top$, $\widetilde{\mathbf{H}} = \mathbf{S}\mathbf{H}$, $\widetilde{\mathbf{H}}_{\mathrm{ref}} = \mathbf{S}\mathbf{H}_{\mathrm{ref}}$, and $\widetilde{\mathbf{H}}_{\mathrm{init}} = \mathbf{S}\mathbf{H}_{\mathrm{init}}$. Row distances are preserved up to relabeling, so

$$\mathbf{C}^\star(\widetilde{\mathbf{H}}; \widetilde{\mathbf{A}}) = \mathbf{S}\mathbf{C}^\star(\mathbf{H}; \mathbf{A})\mathbf{S}^\top.$$

Since $\mathbf{S}\mathbf{1} = \mathbf{1}$,

$$\mathbf{D}_{\mathbf{S}\mathbf{C}\mathbf{S}^\top} = \mathrm{diag}(\mathbf{S}\mathbf{C}\mathbf{S}^\top\mathbf{1}) = \mathbf{S}\mathbf{D}_{\mathbf{C}}\mathbf{S}^\top,$$

which gives the corresponding Laplacian conjugation.

The centering projector commutes with every permutation: $\mathbf{S}\mathbf{J} = \mathbf{J}\mathbf{S}$. Lemma 7 therefore gives

$$\Pi_{\mathbf{S}\mathbf{H}}(\mathbf{S}\mathbf{Z}) = \mathbf{S}\Pi_{\mathbf{H}}(\mathbf{Z}).$$

The reduced energy is invariant under simultaneous permutation of graph, state, and anchor. Moreover, the map $\mathbf{H} \mapsto \mathbf{S}\mathbf{H}$ is an isometry of $\mathcal{M}_{n,p}$ under the induced Frobenius metric. Hence the Riemannian gradient is equivariant:

$$\mathrm{grad}\,\mathcal{E}_{\mathrm{red}}(\mathbf{S}\mathbf{H}; \mathbf{S}\mathbf{H}_{\mathrm{ref}}, \mathbf{S}\mathbf{A}\mathbf{S}^\top) = \mathbf{S}\ \mathrm{grad}\,\mathcal{E}_{\mathrm{red}}(\mathbf{H}; \mathbf{H}_{\mathrm{ref}}, \mathbf{A}),$$

as follows from the standard transformation rule for gradients under Riemannian isometries (Boumal, 2023). Finally, $\widetilde{\mathbf{B}}_t = \mathbf{S}\mathbf{B}_t$ is again standard matrix Brownian motion because orthogonal transformations preserve the covariance structure of Brownian motion (Karatzas and Shreve, 1991). Using the projection equivariance above, multiplication of Eq. (4) by $\mathbf{S}$ therefore yields exactly the SGF with inputs $(\mathbf{S}\mathbf{A}\mathbf{S}^\top, \mathbf{S}\mathbf{H}_{\mathrm{ref}}, \mathbf{S}\mathbf{H}_{\mathrm{init}})$. Equality in law then follows from the uniqueness of the corresponding solution.

Thus, every component of the SGF closure commutes with node relabeling: the conductance map, state-dependent Laplacian, tangent projection, drift, and stochastic forcing transform consistently under the same permutation. This establishes Proposition 2 and shows that the state–geometry feedback introduced in Sections 4.1 and 4.2 is independent of the particular node labeling.

B.3 Proof of Theorem 3

Step 1: well-posedness. The reduced energy is smooth because $\sigma > 0$, and the drift and projected diffusion fields are smooth on the compact embedded manifold. They and their derivatives are bounded. Standard Stratonovich SDE theory on smooth manifolds therefore gives a unique maximal strong solution, and compactness excludes explosion, making the solution global (Hsu, 2002).

Step 2: generator. Write the projected noise using the vector fields of Lemma 8. The Stratonovich generator acting on smooth f is

$$\mathcal{L}f = -\langle \mathrm{grad}\,\mathcal{E}_{\mathrm{red}}, \mathrm{grad}\, f\rangle_F + \tau \sum_\alpha \mathbf{V}_\alpha(\mathbf{V}_\alpha f).$$

Lemma 8 gives $\mathcal{L}f = -\langle \mathrm{grad}\,\mathcal{E}_{\mathrm{red}}, \mathrm{grad}\, f\rangle_F + \tau\Delta_{\mathcal{M}} f$. The Parseval identity also shows that the diffusion is elliptic on every tangent space.

Step 3: Fokker–Planck equation. For every $t > 0$, smooth uniform ellipticity on the connected compact manifold implies that the transition law has a smooth strictly positive density ρ_t with respect to Riemannian volume μ (Hsu, 2002). The corresponding forward Kolmogorov equation is (Pavliotis, 2014)

$$\partial_t \rho_t = \mathrm{div}_{\mathcal{M}}\left(\rho_t\, \mathrm{grad}\,\mathcal{E}_{\mathrm{red}} + \tau\, \mathrm{grad}\,\rho_t\right) = \mathrm{div}_{\mathcal{M}}\left[\rho_t\, \mathrm{grad}(\mathcal{E}_{\mathrm{red}} + \tau \log \rho_t)\right].$$

Step 4: dissipation. The first variation of

$$\mathcal{F}_\tau(\rho) = \int \mathcal{E}_{\mathrm{red}}\rho\, \mathrm{d}\mu + \tau \int \rho \log \rho\, \mathrm{d}\mu$$

is $\delta\mathcal{F}_\tau/\delta\rho = \mathcal{E}_{\mathrm{red}} + \tau(1+\log\rho)$. Probability conservation removes the additive constant τ. Substituting the Fokker–Planck equation and integrating by parts on the compact boundaryless manifold gives the standard free-energy dissipation identity (Jordan et al., 1998; Pavliotis, 2014):

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}_\tau(\rho_t) &= \int_{\mathcal{M}_{n,p}} (\mathcal{E}_{\mathrm{red}} + \tau\log\rho_t)\,\mathrm{div}_{\mathcal{M}}\left[\rho_t\,\mathrm{grad}(\mathcal{E}_{\mathrm{red}} + \tau\log\rho_t)\right]\,\mathrm{d}\mu \\ &= -\int_{\mathcal{M}_{n,p}} \rho_t\,\|\mathrm{grad}(\mathcal{E}_{\mathrm{red}} + \tau\log\rho_t)\|_F^2\,\mathrm{d}\mu.\end{aligned}$$

Step 5: invariant law. Compactness makes $Z_\tau = \int_{\mathcal{M}_{n,p}} \exp(-\mathcal{E}_{\mathrm{red}}/\tau)\,\mathrm{d}\mu$ finite and positive. For $\pi_\tau = Z_\tau^{-1} e^{-\mathcal{E}_{\mathrm{red}}/\tau}$, $\mathrm{grad}(\mathcal{E}_{\mathrm{red}} + \tau\log\pi_\tau) = 0$, so its probability flux vanishes. Moreover, for every probability density ρ,

$$\mathcal{F}_\tau(\rho) = \tau D_{\mathrm{KL}}(\rho\,\|\,\pi_\tau) - \tau\log Z_\tau,$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}_\tau(\rho_t) = -\tau^2 \int_{\mathcal{M}_{n,p}} \rho_t \left\| \mathrm{grad}\log\frac{\rho_t}{\pi_\tau} \right\|_F^2 \mathrm{d}\mu.$$

Step 6: uniqueness. The smooth strictly positive transition density implies that the diffusion semigroup is strong Feller and topologically irreducible on the connected manifold. Since π_τ is already known to be invariant, Doob's uniqueness theorem implies that it is the unique invariant probability measure (Doob, 1953).

Together, Steps 1–6 establish Theorem 3: the continuous SGF is globally well posed, its law evolves according to the gradient-drift Fokker–Planck equation, and its probability-level free energy decreases while admitting the unique Gibbs invariant law π_τ. This justifies the equilibrium law used in Section 4.3 to define $\bar{c}_{ij}^{(\tau)}$, $u_{ij}^{(\tau)}$, and $\bar{\mathbf{L}}_\tau$. It also provides the continuous-time reference law against which the finite-h discrepancy is separated in Section 4.4 and Corollary 6.

B.4 Proof of Theorem 4

The conductance formula gives

$$\frac{c_{rs}^\star(\mathbf{H})}{c_{ij}^\star(\mathbf{H})} = \frac{A_{rs}}{A_{ij}} \exp\left[-\frac{\|\mathbf{h}_r - \mathbf{h}_s\|_2^2 - \|\mathbf{h}_i - \mathbf{h}_j\|_2^2}{2\sigma^2} \right].$$

The assumed distance bounds make the numerator of the exponent at least $r_{\mathrm{out}} - r_{\mathrm{in}}$, which proves Eq. (9). Equivalently, $c_{ij}^\star \geq A_{ij} e^{-r_{\mathrm{in}}/(2\sigma^2)}$ and $c_{rs}^\star \leq A_{rs} e^{-r_{\mathrm{out}}/(2\sigma^2)}$. Thus, conditional on a representation margin, the variational conductance map attenuates the more discrepant edge exponentially, up to the ratio of the observed reference weights. This is consistent with the classical latent-weight interpretation of robust penalties, in which large residuals receive progressively smaller effective weights (Black and Rangarajan, 1996).

This proves Theorem 4 and makes explicit the edge-attenuation mechanism induced by the variational solution in Proposition 1. In particular, the result is pointwise: it characterizes how the adaptive geometry responds once a representation margin is present, but does not assert that the SGF dynamics themselves create such a margin.

B.5 Proof of Theorem 5

Because the combinatorial Laplacian is linear in conductance,

$$\bar{\mathbf{L}}_\tau = \mathbf{L}_{\bar{\mathbf{C}}^{(\tau)}} = \mathbb{E}_{\pi_\tau}[\mathbf{L}(\mathbf{H})].$$

By the quotient Lipschitz bound in Appendix A.4, Jensen's inequality, and Cauchy–Schwarz,

$$\begin{aligned}\left\|\bar{\mathbf{L}}_\tau - \mathbf{L}(\mathbf{H}^\dagger)\right\|_2 &\le \mathbb{E}_{\pi_\tau}\left\|\mathbf{L}(\mathbf{H}) - \mathbf{L}(\mathbf{H}^\dagger)\right\|_2 \\ &\le K_L \mathbb{E}_{\pi_\tau} d_{\mathcal{O}}(\mathbf{H}, \mathbf{H}^\dagger) \\ &\le K_L \sqrt{\mathbb{E}_{\pi_\tau} d_{\mathcal{O}}(\mathbf{H}, \mathbf{H}^\dagger)^2} \\ &\le K_L \sqrt{C_{\mathrm{conc}}\tau}.\end{aligned}$$

The triangle inequality with $\varepsilon_{\mathrm{model}} = \left\|\mathbf{L}(\mathbf{H}^\dagger) - \mathbf{L}_\star\right\|_2$ proves Eq. (10). Suppose $\varepsilon_{\mathrm{eq}} \le \delta_\star/4$. By Weyl's inequality (Bhatia, 1997), target eigenvalues remain within $\varepsilon_{\mathrm{eq}} \le \delta_\star/4$ of $\mathcal{I}_\star$, whereas complementary eigenvalues remain at least $\delta_\star - \varepsilon_{\mathrm{eq}} \ge 3\delta_\star/4$ away. Hence $\mathbf{1}_{\mathcal{I}_\star^+}(\bar{\mathbf{L}}_\tau)$ has rank k and selects exactly the perturbed target block. Moreover, Weyl's bound implies that the separation between this perturbed block and the unperturbed complementary spectrum is at least $\delta_\star - \varepsilon_{\mathrm{eq}}$. The Davis–Kahan theorem (Davis and Kahan, 1970) therefore yields

$$\left\|\bar{\mathbf{P}}_{\tau,k} - \mathbf{P}_\star\right\|_2 \le \frac{\varepsilon_{\mathrm{eq}}}{\delta_\star - \varepsilon_{\mathrm{eq}}} \le \frac{2\varepsilon_{\mathrm{eq}}}{\delta_\star},$$

which is Eq. (11).

Theorem 5 separates the equilibrium operator error into the approximation term $\varepsilon_{\mathrm{model}}$ and the finite-temperature concentration term $K_L\sqrt{C_{\mathrm{conc}}\tau}$, and converts this operator control into spectral-projector stability whenever the resulting perturbation is small relative to the latent spectral gap. This is the conditional equilibrium-level guarantee stated in Section 5.3, under the stated approximation and concentration conditions. The bound also supplies the continuous-equilibrium component used in the finite-step error decomposition of Corollary 6.

B.6 Proof of Corollary 6

Insert the retained-state mean, the discrete stationary mean, and the exact continuous equilibrium mean:

$$\begin{aligned}\left\|\widehat{\mathbf{L}}_{\tau,h,B,S,M} - \mathbf{L}_\star\right\|_2 \le{}& \left\|\widehat{\mathbf{L}}_{\tau,h,B,S,M} - \mathbf{L}^{\mathrm{mean}}_{\tau,h,B,S,M}\right\|_2 \\ &+ \left\|\mathbf{L}^{\mathrm{mean}}_{\tau,h,B,S,M} - \bar{\mathbf{L}}_{\tau,h}\right\|_2 \\ &+ \left\|\bar{\mathbf{L}}_{\tau,h} - \bar{\mathbf{L}}_\tau\right\|_2 \\ &+ \left\|\bar{\mathbf{L}}_\tau - \mathbf{L}_\star\right\|_2.\end{aligned}$$

On the event defining $\varepsilon_{\mathrm{MC}}$, which holds with probability at least $1-\delta$, the four terms are bounded by $\varepsilon_{\mathrm{MC}}$, $\varepsilon_{\mathrm{mix}}(B,S,M,h)$, $\varepsilon_{\mathrm{disc}}(h)$, and Eq. (10), respectively. Hence

$$\left\|\widehat{\mathbf{L}}_{\tau,h,B,S,M} - \mathbf{L}_\star\right\|_2 \le \varepsilon_{\mathrm{tot}},$$

which is Eq. (12). If $\varepsilon_{\mathrm{tot}} \le \delta_\star/4$, Weyl's inequality (Bhatia, 1997) places target eigenvalues within $\varepsilon_{\mathrm{tot}}$ of $\mathcal{I}_\star$ and complementary ones at least $\delta_\star - \varepsilon_{\mathrm{tot}} \ge 3\delta_\star/4$ away. Hence $\mathcal{I}_\star^+$ selects exactly the rank-k perturbed target block of $\widehat{\mathbf{L}}_{\tau,h,B,S,M}$. Moreover, its separation from the unperturbed complementary spectrum is at least $\delta_\star - \varepsilon_{\mathrm{tot}}$. The Davis–Kahan sin-$\Theta$ theorem (Davis and Kahan, 1970) therefore gives

$$\left\|\widehat{\mathbf{P}}_k - \mathbf{P}_\star\right\|_2 \le \frac{\varepsilon_{\mathrm{tot}}}{\delta_\star - \varepsilon_{\mathrm{tot}}} \le \frac{2\varepsilon_{\mathrm{tot}}}{\delta_\star},$$

which proves Eq. (13). No independence assumption or scalar effective-sample-size substitution is used in this argument. Finite-chain nonstationarity is represented explicitly by $\varepsilon_{\mathrm{mix}}(B,S,M,h)$, while sampling fluctuations of the possibly dependent retained states enter only through the assumed high-probability operator deviation $\varepsilon_{\mathrm{MC}}$ (Meyn and Tweedie, 2009).

Corollary 6 lifts the equilibrium-level operator guarantee of Theorem 5 to the implemented finite-step estimator. The resulting bound separates the approximation, finite-temperature, discretization, incomplete-mixing, and Monte Carlo contributions rather than absorbing them into a single residual term. This is the practical error decomposition stated in Section 5.3, and it also makes explicit that the spectral guarantee applies only when their combined magnitude remains small relative to the latent spectral gap $\delta_\star$.

B.7 Frozen-geometry Ornstein–Uhlenbeck derivation

Set $\mathbf{A}_{\mathbf{C}} = \mathbf{L}_{\mathbf{C}} + \eta \mathbf{I}_n$. Assume that $\mathbf{H}_{\text{init}}$ is deterministic or independent of the driving Brownian motion. The unconstrained linear process is

$$\mathrm{d}\mathbf{H}_t = -\mathbf{A}_{\mathbf{C}}\mathbf{H}_t \,\mathrm{d}t + \eta \mathbf{H}_{\text{ref}} \,\mathrm{d}t + \sqrt{2\tau}\,\mathrm{d}\mathbf{B}_t, \qquad \mathbf{H}_{t=0} = \mathbf{H}_{\text{init}}.$$

Let $\mathbf{L}_{\mathbf{C}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$ and define $\mathbf{a}_i(t) = \boldsymbol{\phi}_i^\top \mathbf{H}_t$, $\mathbf{a}_{\text{ref},i} = \boldsymbol{\phi}_i^\top \mathbf{H}_{\text{ref}}$. Orthogonal invariance of Brownian motion yields (Karatzas and Shreve, 1991)

$$\mathrm{d}\mathbf{a}_i(t) = -\alpha_i \mathbf{a}_i(t)\,\mathrm{d}t + \eta \mathbf{a}_{\text{ref},i}\,\mathrm{d}t + \sqrt{2\tau}\,\mathrm{d}\boldsymbol{\beta}_i(t), \qquad \alpha_i = \lambda_i + \eta.$$

Variation of constants gives

$$\mathbf{a}_i(t) = e^{-\alpha_i t}\mathbf{a}_i(0) + \eta \int_0^t e^{-\alpha_i(t-s)}\mathbf{a}_{\text{ref},i}\,\mathrm{d}s + \sqrt{2\tau}\int_0^t e^{-\alpha_i(t-s)}\,\mathrm{d}\boldsymbol{\beta}_i(s).$$

For $\alpha_i > 0$, taking expectations and applying the Itô isometry gives Eqs. (14)–(15) (Pavliotis, 2014). In particular, the contribution of the initial condition decays exponentially, and the unique stationary law of this mode is Gaussian with mean $\eta \mathbf{a}_{\text{ref},i}/\alpha_i$ and covariance $\tau \mathbf{I}_p/\alpha_i$. When $\eta = 0$ and $\lambda_i = 0$, we have $\alpha_i = 0$, so the mode reduces to Brownian motion and has no stationary probability law. By contrast, every mode with $\alpha_i > 0$ is mean-reverting.

This derivation gives the frozen-geometry benchmark used in Section 5.4. With the conductance operator fixed, the modal equations decouple, and each mode with $\alpha_i > 0$ follows an Ornstein–Uhlenbeck process (Pavliotis, 2014) with decay rate $\alpha_i = \lambda_i + \eta$, stationary filter $g(\lambda_i) = \eta/(\lambda_i + \eta)$, and stationary variance $\tau/(\lambda_i + \eta)$. These formulas are diagnostic rather than geometry-learning guarantees, since freezing $\mathbf{L}_{\mathbf{C}}$ removes state feedback into the conductance geometry.

C Additional Method Details

C.1 Scope of the stochastic guarantees

Theorem 3 concerns the continuous-time process with fixed observed weights $\mathbf{A}$ (and hence fixed candidate support $\mathcal{E}_{\text{obs}}$), fixed fidelity anchor $\mathbf{H}_{\text{ref}}$, and fixed hyperparameters (σ, η, τ). End-to-end training changes θ and therefore changes the anchor and energy across optimization iterations; the training procedure is not itself the stationary diffusion analyzed by the theorem.

For Gibbs-equilibrium geometry estimation, freeze the encoder, decoder, observed graph, anchor, and SGF hyperparameters. Choose an initial state $\mathbf{H}_{\text{init}} \in \mathcal{M}_{n,p}$, deterministic or independent of the simulation noise, run a declared burn-in, and retain samples under a declared sampling schedule (Meyn and Tweedie, 2009). The retraction chain is weakly consistent with the continuous SGF as $h \downarrow 0$, but it need not preserve π_τ exactly at finite step size (Kloeden and Platen, 1992). Corollary 6 makes the resulting finite-step effects explicit through separate discretization, mixing, and sampling terms.

The practical estimator therefore uses a frozen-parameter Markov chain only after training has been completed. The purpose of the following routine is not to claim exact sampling from π_τ at finite h, nor to imply that a prescribed burn-in and thinning schedule by itself guarantees stationarity, independence, or adequate mixing. Rather, the routine makes the implemented sampling protocol explicit: the initial state, burn-in, spacing, and retained sample count are all declared, and each retained state is mapped through $\mathbf{C}^\star(\mathbf{H})$ before aggregation. Any agreement with the continuous Gibbs law is therefore subject to the discretization and mixing effects represented explicitly in Corollary 6, together with the finite-sample fluctuation represented by ε_{MC}.

In particular, the schedule separates the numerical chain used in practice from the continuous Gibbs diffusion analyzed in Theorem 3. The algorithm itself does not establish uniqueness or geometric ergodicity of the finite-h chain, does not certify that the retained states are effectively independent, and does not provide an additional convergence theorem beyond the assumptions and error terms already stated in Section 5.3. Burn-in, spacing, effective sample size, and between-chain diagnostics should therefore be interpreted as practical diagnostics rather than as proofs of exact equilibrium sampling. Likewise, the resulting conductance mean and variance are finite-chain estimates of the model-induced geometry quantities and should not be read as causal or automatically calibrated uncertainty measures.

Algorithm 2 summarizes this frozen-parameter estimation procedure.

Algorithm 2 Frozen-Parameter Gibbs-Geometry Estimation

Require: Frozen $\mathbf{A}$, $\mathbf{H}_{\mathrm{ref}}$, (σ, η, τ, h); initial state $\mathbf{H}_{\mathrm{init}}$; burn-in $B \geq 0$; spacing $S \geq 1$; retained count $M \geq 2$.

1: $\mathbf{H}^{(0)} \leftarrow \mathbf{H}_{\mathrm{init}}$; initialize an empty sample list $\mathcal{S}$.
2: **for** $\ell = 0, \ldots, B + SM - 1$ **do**
3: Apply one transition from Eq. (7) and Eq. (6).
4: **if** $\ell + 1 > B$ and $(\ell + 1 - B) \bmod S = 0$ **then**
5: Append $\mathbf{C}^\star(\mathbf{H}^{(\ell+1)})$ to $\mathcal{S}$.
6: **end if**
7: **end for**
8: **return** conductance sample mean, sample variance, and $\widehat{\mathbf{L}}_{\tau,h,B,S,M}$.

C.2 Normalized Laplacian extension

Let

$$\mathbf{D}(\mathbf{H}) = \mathrm{diag}(\mathbf{C}^\star(\mathbf{H})\mathbf{1}).$$

To cover isolated nodes, fix a declared degree stabilizer $\varepsilon_{\mathrm{deg}} > 0$ and define

$$\mathbf{D}_\varepsilon(\mathbf{H}) = \mathbf{D}(\mathbf{H}) + \varepsilon_{\mathrm{deg}}\mathbf{I}_n, \qquad \mathbf{S}_\varepsilon(\mathbf{H}) = \mathbf{D}_\varepsilon(\mathbf{H})^{-1/2}.$$

The regularized symmetric normalized operator is

$$\mathbf{L}_{\mathrm{sym},\varepsilon}(\mathbf{H}) = \mathbf{I}_n - \mathbf{S}_\varepsilon(\mathbf{H})\mathbf{C}^\star(\mathbf{H})\mathbf{S}_\varepsilon(\mathbf{H}).$$

This is an $\varepsilon_{\mathrm{deg}}$-regularized variant of the symmetric normalized Laplacian (Chung, 1997). For $\varepsilon_{\mathrm{deg}} > 0$, the inverse square root remains well defined even in the presence of isolated nodes; the resulting operator should therefore be viewed as a stabilized normalized operator rather than the unregularized normalized Laplacian itself.

Consider the scalar energy

$$E_{\mathrm{sym},\varepsilon}(\mathbf{H}) = \frac{1}{2}\,\mathrm{tr}\big(\mathbf{H}^\top \mathbf{L}_{\mathrm{sym},\varepsilon}(\mathbf{H})\mathbf{H}\big) + \frac{\eta}{2}\,\|\mathbf{H} - \mathbf{H}_{\mathrm{ref}}\|_F^2 .$$

For a perturbation $\mathbf{Z}$, its directional derivative is

$$\begin{aligned} \mathrm{D}E_{\mathrm{sym},\varepsilon}(\mathbf{H})[\mathbf{Z}] = {} & \langle \mathbf{L}_{\mathrm{sym},\varepsilon}(\mathbf{H})\mathbf{H} + \eta(\mathbf{H} - \mathbf{H}_{\mathrm{ref}}), \mathbf{Z}\rangle_F \\ & + \frac{1}{2}\left\langle \mathbf{H}\mathbf{H}^\top, \mathrm{D}\mathbf{L}_{\mathrm{sym},\varepsilon}(\mathbf{H})[\mathbf{Z}]\right\rangle_F . \end{aligned}$$

For every observed edge,

$$\mathrm{D}c_{ij}^\star(\mathbf{H})[\mathbf{Z}] = -\frac{c_{ij}^\star(\mathbf{H})}{\sigma^2}\,\langle \mathbf{h}_i - \mathbf{h}_j, \mathbf{z}_i - \mathbf{z}_j\rangle .$$

Consequently,

$$\mathrm{D}\mathbf{D}(\mathbf{H})[\mathbf{Z}] = \mathrm{diag}(\mathrm{D}\mathbf{C}^\star(\mathbf{H})[\mathbf{Z}]\mathbf{1}), \qquad \mathrm{D}\mathbf{D}_\varepsilon(\mathbf{H})[\mathbf{Z}] = \mathrm{D}\mathbf{D}(\mathbf{H})[\mathbf{Z}].$$

Because $\mathbf{D}_\varepsilon(\mathbf{H})$ and its directional derivative are diagonal and therefore commute, the Fréchet derivative of the inverse square-root map reduces to the entrywise scalar derivative (Higham, 2008):

$$\mathrm{D}\mathbf{S}_\varepsilon(\mathbf{H})[\mathbf{Z}] = -\frac{1}{2}\mathbf{D}_\varepsilon(\mathbf{H})^{-3/2}\mathrm{D}\mathbf{D}(\mathbf{H})[\mathbf{Z}].$$

The full operator derivative is

$$\begin{aligned} \mathrm{D}\mathbf{L}_{\mathrm{sym},\varepsilon}(\mathbf{H})[\mathbf{Z}] = {} & -\mathrm{D}\mathbf{S}_\varepsilon(\mathbf{H})[\mathbf{Z}]\,\mathbf{C}^\star(\mathbf{H})\mathbf{S}_\varepsilon(\mathbf{H}) \\ & - \mathbf{S}_\varepsilon(\mathbf{H})\,\mathrm{D}\mathbf{C}^\star(\mathbf{H})[\mathbf{Z}]\,\mathbf{S}_\varepsilon(\mathbf{H}) \\ & - \mathbf{S}_\varepsilon(\mathbf{H})\mathbf{C}^\star(\mathbf{H})\,\mathrm{D}\mathbf{S}_\varepsilon(\mathbf{H})[\mathbf{Z}]. \end{aligned}$$

Thus, $\mathbf{L}_{\mathrm{sym},\varepsilon}(\mathbf{H})\mathbf{H}$ alone is not the Euclidean gradient of $E_{\mathrm{sym},\varepsilon}$: the dependence of the normalized operator on $\mathbf{H}$ contributes the additional term involving $\mathrm{D}\mathbf{L}_{\mathrm{sym},\varepsilon}(\mathbf{H})$. For optimization on $\mathcal{M}_{n,p}$, the corresponding Riemannian gradient is therefore obtained by projecting the full Euclidean gradient,

$$\mathrm{grad}\, E_{\mathrm{sym},\varepsilon}(\mathbf{H}) = \Pi_{\mathbf{H}}\big(\nabla E_{\mathrm{sym},\varepsilon}(\mathbf{H})\big).$$

By contrast, the closed expression $\mathbf{L}(\mathbf{H})\mathbf{H}$ in the main method arises from exact variational elimination in the joint combinatorial energy: the conductance optimizer satisfies its first-order condition, so the dependence of $\mathbf{C}^\star(\mathbf{H})$ cancels through the envelope argument. It would not, in general, be obtained by simply differentiating $\frac{1}{2}\operatorname{tr}(\mathbf{H}^\top\mathbf{L}(\mathbf{H})\mathbf{H})$ while ignoring the state dependence of $\mathbf{L}(\mathbf{H})$. Consequently, an energy-consistent normalized extension must differentiate its complete scalar energy, either analytically as above or through automatic differentiation.

C.3 Graph-preconditioned noise

In Euclidean coordinates, let $\mathbf{M}(\mathbf{h})$ be a smooth symmetric positive-definite mobility. For the corresponding reversible Itô diffusion with diffusion tensor $2\tau\mathbf{M}(\mathbf{h})$, invariance of the Gibbs density proportional to $e^{-E(\mathbf{h})/\tau}$ requires the state-dependence of the mobility to be accounted for in the drift (Pavliotis, 2014; Ma et al., 2015):

$$\mathrm{d}\mathbf{h}_t = [-\mathbf{M}(\mathbf{h}_t)\nabla E(\mathbf{h}_t) + \tau\nabla\cdot\mathbf{M}(\mathbf{h}_t)]\,\mathrm{d}t + \sqrt{2\tau\,\mathbf{M}(\mathbf{h}_t)}\,\mathrm{d}\mathbf{b}_t.$$

Here $(\nabla\cdot\mathbf{M})_i = \sum_j \partial_{h_j} M_{ij}$. With this convention, the corresponding Fokker–Planck equation can be written in divergence form (Pavliotis, 2014),

$$\partial_t\rho = \nabla\cdot\left[\mathbf{M}(\rho\nabla E + \tau\nabla\rho)\right].$$

Hence the Gibbs density $\rho_\infty \propto e^{-E/\tau}$ has zero probability flux. If the mobility-divergence correction is omitted, a state-dependent diffusion matrix generally introduces an additional bias and the same Gibbs density need not remain invariant.

On $\mathcal{M}_{n,p}$, the analogue is a smooth self-adjoint positive-definite bundle endomorphism

$$\mathbf{M}_{\mathbf{H}} : T_{\mathbf{H}}\mathcal{M}_{n,p} \to T_{\mathbf{H}}\mathcal{M}_{n,p}.$$

Its state dependence must be differentiated intrinsically: the required correction involves the Riemannian divergence (equivalently, covariant derivatives of the mobility field) and depends on the reference volume and the precise Itô/Stratonovich representation (Girolami and Calderhead, 2011; Hsu, 2002).

Consequently, replacing the isotropic projected noise in Eq. (4) by a graph-dependent preconditioner without simultaneously deriving the corresponding intrinsic drift correction would define a different stochastic process and would not inherit the Gibbs-invariance argument of Theorem 3. We therefore treat graph-preconditioned noise only as a possible extension; none of the free-energy, invariant-law, or equilibrium-recovery guarantees in the main theory are asserted for that variant.

C.4 Equilibrium estimators and sampling error

Given retained states $\{\mathbf{H}^{(m)}\}_{m=1}^M$, define

$$\widehat{c}_{ij} = \frac{1}{M}\sum_{m=1}^{M} c_{ij}^\star(\mathbf{H}^{(m)}), \qquad \widehat{u}_{ij} = \frac{1}{M-1}\sum_{m=1}^{M}\left(c_{ij}^\star(\mathbf{H}^{(m)}) - \widehat{c}_{ij}\right)^2.$$

Here $\widehat{u}_{ij}$ is the empirical sample variance. For dependent samples it is a descriptive finite-chain estimator. At fixed h, its stationary target is the variance under $\pi_{\tau,h}$; recovering $u_{ij}^{(\tau)}$ additionally requires control of discretization error.

Linearity of the combinatorial Laplacian gives

$$\widehat{\mathbf{L}}_{\tau,h,B,S,M} = \mathbf{L}_{\widehat{\mathbf{C}}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{L}(\mathbf{H}^{(m)}).$$

For genuinely independent stationary samples from the exact law, let $\mathbf{X}_m = \mathbf{L}(\mathbf{H}^{(m)}) - \bar{\mathbf{L}}_\tau$, $\|\mathbf{X}_m\|_2 \le R_L$, and $v_L = \left\|\mathbb{E}[\mathbf{X}_m^2]\right\|_2$. A self-adjoint matrix Bernstein inequality (Tropp, 2012) gives, with probability at least $1-\delta$,

$$\left\|\frac{1}{M}\sum_{m=1}^{M}\mathbf{L}(\mathbf{H}^{(m)}) - \bar{\mathbf{L}}_\tau\right\|_2 \le \sqrt{\frac{2v_L\log(2n/\delta)}{M}} + \frac{2R_L\log(2n/\delta)}{3M}.$$

Indeed, $\left\|\bar{\mathbf{L}}_\tau\right\|_2 \le \mathbb{E}_{\pi_\tau}\left\|\mathbf{L}(\mathbf{H})\right\|_2 \le 2d_{\max}(\mathbf{A})$, so the triangle inequality gives $\left\|\mathbf{X}_m\right\|_2 \le 4d_{\max}(\mathbf{A})$.

This independent-sample inequality must not be reused by replacing M with a scalar effective sample size. For a Markov chain, a rigorous operator bound requires a dependence-aware concentration result or an explicit assumption such as the $\varepsilon_{\mathrm{MC}}$ condition in Corollary 6. Scalar effective sample sizes remain useful summary-specific diagnostics. For a stationary scalar functional $g(\mathbf{H})$, the formal integrated-autocorrelation-time relation is

$$M_{\mathrm{eff}}(g) = \frac{M}{1 + 2\sum_{r\ge 1}\rho_g(r)},$$

with the autocorrelation sum estimated using a declared truncation or stabilization rule (Geyer, 1992; Vehtari et al., 2021). Such diagnostics may be reported for the energy, representative conductances, and scalar operator summaries, and for projector-related summaries when a suitable reference is available. They are summary-dependent diagnostics and should not be presented as a proof of operator-norm concentration.

C.5 Gauge-invariant aggregation

When $\eta = 0$, the reduced energy and conductance map are invariant under $\mathbf{H} \mapsto \mathbf{H}\mathbf{R}$ for $\mathbf{R} \in \mathrm{O}(p)$. Direct averaging of representation matrices can therefore cancel or blur equivalent orientations. The projector

$$\mathbf{P}_{\mathbf{H}} = \frac{1}{n}\mathbf{H}\mathbf{H}^\top$$

is invariant under this right-orthogonal action and represents only the underlying p-dimensional subspace (Absil et al., 2008; Edelman et al., 1998). Its equilibrium estimator is

$$\widehat{\mathbf{P}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{P}_{\mathbf{H}^{(m)}}.$$

Although each $\mathbf{P}_{\mathbf{H}^{(m)}}$ is a rank-p orthogonal projector, their average need not itself be a projector. If a single rank-p representative is required, let $\mathbf{U}_p$ contain the leading p eigenvectors of $\widehat{\mathbf{P}}$ and set

$$\widehat{\mathbf{H}}_{\mathrm{proj}} = \sqrt{n}\,\mathbf{U}_p.$$

If $\lambda_p(\widehat{\mathbf{P}}) > \lambda_{p+1}(\widehat{\mathbf{P}})$, the leading rank-$p$ subspace is unique, with representatives differing by right-orthogonal transformations; otherwise it need not be unique.

For feature-level aggregation, fix a declared alignment reference $\mathbf{H}^{\mathrm{align}}$ and solve the orthogonal Procrustes problem (Schönemann, 1966)

$$\mathbf{R}_m = \operatorname*{arg\,min}_{\mathbf{R}^\top\mathbf{R}=\mathbf{I}_p}\left\|\mathbf{H}^{(m)}\mathbf{R} - \mathbf{H}^{\mathrm{align}}\right\|_F.$$

If

$$(\mathbf{H}^{(m)})^\top\mathbf{H}^{\mathrm{align}} = \mathbf{U}_m\mathbf{\Sigma}_m\mathbf{V}_m^\top,$$

then one optimal solution is

$$\mathbf{R}_m = \mathbf{U}_m\mathbf{V}_m^\top.$$

The aligned feature average may then be formed as

$$\widehat{\mathbf{H}}_{\mathrm{align}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{H}^{(m)}\mathbf{R}_m.$$

If $(\mathbf{H}^{(m)})^\top\mathbf{H}^{\mathrm{align}}$ is rank deficient, the Procrustes optimizer may be nonunique. Since $\widehat{\mathbf{H}}_{\mathrm{align}}$ need not satisfy the scaled Stiefel constraints, it can be mapped back to $\mathcal{M}_{n,p}$ via $\mathcal{Q}$ in Eq. (6) when $\widehat{\mathbf{H}}_{\mathrm{align}}^\top\mathbf{J}\widehat{\mathbf{H}}_{\mathrm{align}} \succ 0$.

For reference, Figure 4 summarizes the full SGF state–geometry loop and its equilibrium readout.

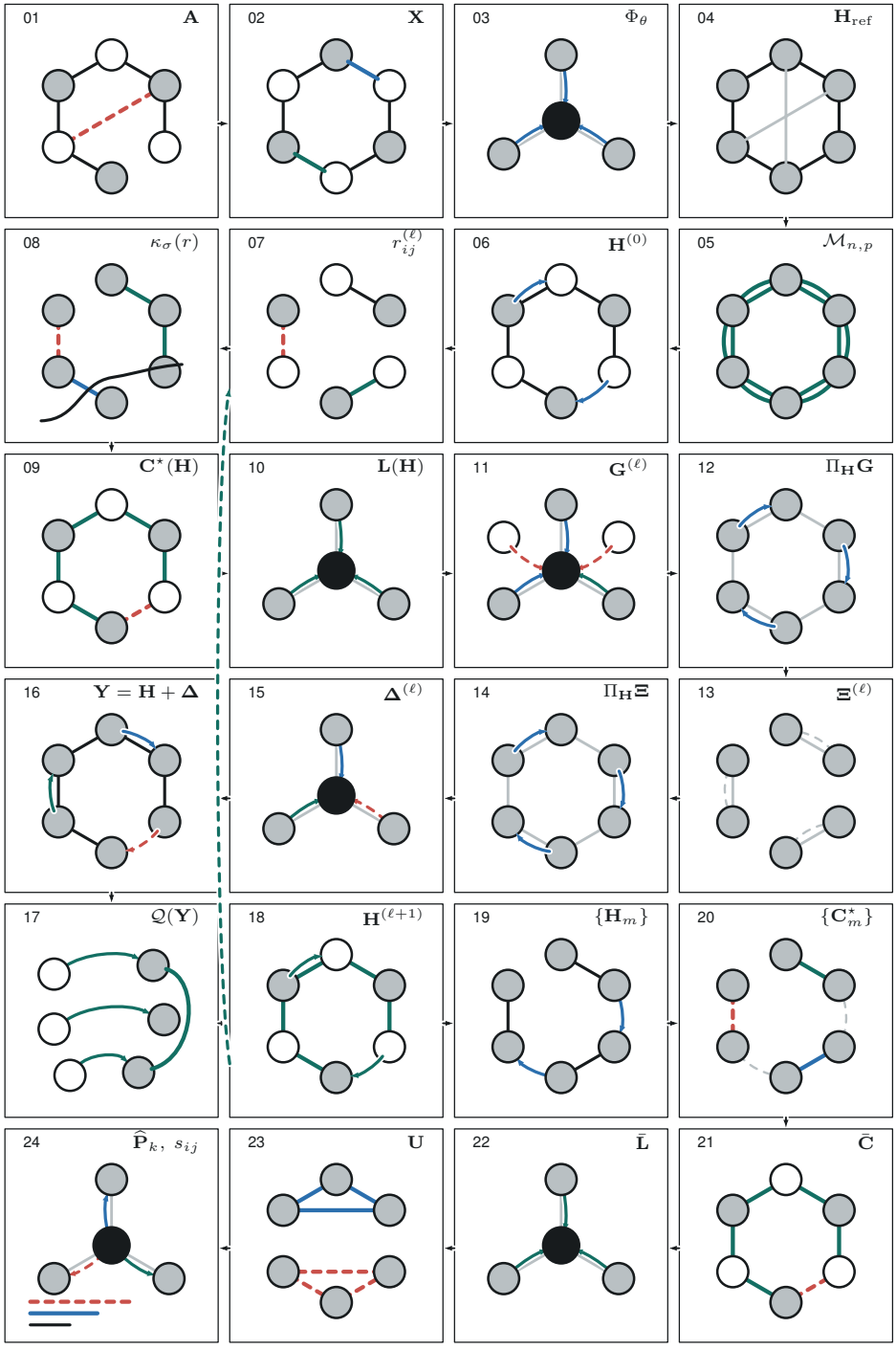


Figure 4: **Topology atlas of the SGF closed loop.** Cells 01–24 show candidate support, adaptive geometry, stochastic transitions, retraction, and pushforward statistics. Black and gray denote nodes and reference bonds; blue indicates state transport, dark green retained geometry or retraction, and red suppressed or uncertain relations.

C.6 Computational complexity and numerical safeguards

Evaluating the adaptive conductances requires a p-dimensional discrepancy on each observed edge and costs $O(|\mathcal{E}_{\mathrm{obs}}|p)$ time. Applying the resulting sparse weighted Laplacian to the representation has the same complexity. The tangent projection and the centered Gram matrix used by the polar retraction each require $O(np^2)$ work, while the inverse square root of the resulting $p \times p$ matrix costs $O(p^3)$. Hence one SGF transition has time complexity

$$O\big(|\mathcal{E}_{\mathrm{obs}}|p + np^2 + p^3\big)\,.$$

When $p \ll n$ and the observed support is sparse, the edgewise operations and the np^2 manifold operations dominate, while the cubic term is confined to a small dense matrix.

With streaming accumulation of equilibrium statistics, the per-chain working memory is

$$O(|\mathcal{E}_{\mathrm{obs}}| + np + p^2) = O(|\mathcal{E}_{\mathrm{obs}}| + np),$$

up to the lower-order $p \times p$ workspace and excluding encoder activations. This bound assumes that retained states or conductance matrices are not all stored explicitly. Storing all M retained conductance samples would require an additional $O(M|\mathcal{E}_{\mathrm{obs}}|)$ memory, while storing all retained representation states would require $O(Mnp)$.

Consequently, running K chains for T transitions requires aggregate work

$$O\big(KT\big(|\mathcal{E}_{\mathrm{obs}}|p + np^2 + p^3\big)\big)\,.$$

Independent chains can be executed in parallel when sufficient hardware is available, reducing wall-clock time but not aggregate computation. If all chains are resident simultaneously, their chain-specific working memory also scales linearly with K, apart from graph, encoder, or anchor quantities that can be shared.

The inverse square root in the centered polar map should be evaluated from a symmetric eigendecomposition, or equivalently a singular-value decomposition, after symmetrizing the Gram matrix

$$\mathbf{G} = \mathbf{Y}^\top \mathbf{J}\mathbf{Y}.$$

In exact arithmetic, $\mathbf{G}$ is positive definite whenever $\mathbf{J}\mathbf{Y}$ has full column rank. In finite precision, its smallest eigenvalue or condition number should be monitored (Higham, 2008). Small negative eigenvalues attributable only to roundoff may be clipped at machine-precision scale. By contrast, adding an undeclared ridge $\varepsilon\mathbf{I}_p$ changes the centered polar map and should not be described as the same retraction.

Near-singularity should likewise be treated explicitly rather than hidden by numerical regularization. A conservative implementation records such events and reruns the sampler with a smaller declared fixed step size. Per-step rejection, state-dependent step-size adaptation, or additional regularization changes the finite-step transition kernel and must therefore be reported as a modified numerical scheme rather than as the unaltered update analyzed in Section 4.4.

Finally, the second-order polar retraction yields the local weak-consistency property established in Appendix A.3 as $h \downarrow 0$, but it does not imply that the finite-h chain preserves the continuous Gibbs law π_τ exactly (Kloeden and Platen, 1992). Equilibrium estimates should therefore be accompanied by step-size sensitivity checks and compared at matched physical integration time whenever discretization effects are being assessed. Where feasible, additional checks using smaller step sizes or a sampler with an explicit invariant-measure correction can help distinguish discretization bias from finite-chain mixing and Monte Carlo error. These numerical diagnostics complement, rather than replace, the $\varepsilon_{\mathrm{disc}}$, $\varepsilon_{\mathrm{mix}}$, and $\varepsilon_{\mathrm{MC}}$ terms in Corollary 6.

C.7 Empirical kernel scaling and scope

Table 5 reports float32 NumPy propagation-kernel benchmarks on synthetic fixed-degree graphs using single-thread BLAS. At one million nodes and four million undirected edges, one stochastic SGF transition takes 0.963 seconds and reaches 526 MiB peak resident memory, compared with 0.621 seconds for the fixed-weight sparse propagation kernel. For fixed p and bounded degree, the measured transition cost is broadly consistent with the near-linear scaling predicted by Section C.6; at the largest tested size, the stochastic SGF kernel incurs a $1.6\times$ runtime overhead relative to the fixed-weight kernel.

This comparison isolates propagation costs under a common sparse-kernel setting. It excludes encoder and decoder computation, backpropagation, data loading, preprocessing, and hyperparameter search. The reported peak RSS likewise refers to the benchmark process, including the graph, representation states, and transition temporaries, rather than to the memory footprint of a complete training pipeline.

Sampling nevertheless multiplies the per-transition cost. For illustration, projecting the measured million-node kernel to $K = 2$ chains and $T = 120$ transitions per chain gives approximately 231 seconds of aggregate serial propagation. Independent chains can be executed in parallel when sufficient hardware is available, reducing wall-clock time but not aggregate work. This arithmetic projection is a workload estimate only: $T = 120$ is not asserted to be sufficient for burn-in, mixing, or equilibrium convergence.

The benchmark therefore demonstrates that the sparse SGF transition kernel can be executed at the million-node scale in the reported implementation and hardware setting. It does not establish end-to-end million-node training, nor does it establish converged Gibbs estimation at that scale. The measured overhead should instead be interpreted in relation to the additional quantities produced by SGF, namely adaptive effective geometry and model-induced geometry uncertainty. When only a point estimate is required, SGF-$\tau = 0$ provides a lower-cost adaptive-geometry variant without the sampling overhead associated with positive temperature.

Table 5: **Measured scaling of the propagation kernels.** Float32 synthetic 8-regular graphs use $p = 8$ and $|\mathcal{E}_{\mathrm{obs}}| = 4n$ with single-thread BLAS. Times are median seconds per transition. The $K = 2$, $T = 120$ column projects aggregate serial SGF kernel time as 240 times the measured stochastic-transition time; it is an arithmetic workload projection rather than an end-to-end training or convergence measurement. Peak RSS includes the Python process, graph, representation states, and transition temporaries.

Nodes	Edges	Fixed kernel (s/step)	SGF-$\tau = 0$ (s/step)	SGF (s/step)	SGF/fixed ratio	$K = 2, T = 120$ projected (s)	Peak RSS (MiB)
1,000	4,000	0.0002	0.0004	0.0005	2.5×	0.1	38
10,000	40,000	0.0030	0.0065	0.0083	2.8×	2.0	42
100,000	400,000	0.0568	0.0809	0.0976	1.7×	23.4	85
1,000,000	4,000,000	0.6209	0.7929	0.9631	1.6×	231.1	526

D Experimental Details

D.1 Separation of Training and Equilibrium Estimation

Two experimental regimes are used.

Finite-depth representation learning. For representation experiments, the SGF block is unrolled for $L = 420$ steps and optimized end-to-end with the encoder and inner-product decoder using binary cross-entropy with a 1:1 ratio of observed edges to uniformly sampled nonedges. Injected edges are therefore included in the positive training set; no clean-edge indicator is used by this objective. At evaluation, a single stochastic realization is used to obtain the frozen representations before fitting the linear probe. All compared methods use the same data splits and label protocol. This regime evaluates SGF as a finite-depth representation layer and does not assume that the resulting trajectory has reached equilibrium.

Gibbs-equilibrium geometry estimation. For sampling-based geometry statistics, training is completed first and the encoder, decoder, observed graph, fidelity anchor, and SGF hyperparameters are then frozen. SGF trajectories are run with a declared burn-in and sampling interval, and conductance moments are estimated from the retained states. Because the implemented chain has finite step size and finite length, these quantities are treated as finite-chain approximations to the Gibbs-induced conductance statistics rather than as evidence of exact equilibrium sampling.

D.2 Synthetic Data Generation

Three controlled synthetic benchmarks are used.

Stochastic block models. Each SBM contains 90 nodes and 3 blocks, with $p_{\text{in}} = 0.30$ and $p_{\text{out}} = 0.035$. Node attributes are noisy three-dimensional one-hot features. The clean graph is generated before topology corruption and defines the clean reference Laplacian $\mathbf{L}_\star$. Community labels are used for data generation, splitting, and evaluation, but not for constructing the random edge corruptions or SGF conductances.

Random geometric graphs. Each random geometric graph contains 90 nodes. A symmetrized 8-nearest-neighbor graph is constructed from planar latent coordinates with heat-kernel edge weights. The noisy two-dimensional coordinates are used as node attributes. The graph constructed before corruption defines the clean reference Laplacian $\mathbf{L}_\star$.

Manifold benchmarks. The reported manifold benchmark is a Swiss roll with 90 nodes. A symmetrized 8-nearest-neighbor graph is constructed in the latent (t, z) coordinates, while noisy three-dimensional ambient coordinates are used as node attributes. The observation-noise level is 0.05. The graph before corruption defines the clean reference operator.

D.3 Topology Corruption Models

Let $\mathbf{A}_\star$ denote the pre-corruption graph and $\mathbf{A}_{\text{obs}}$ the graph supplied to the methods.

Random edge addition. The reported corruption study uses random edge additions. Nonedges of $\mathbf{A}_\star$ are sampled uniformly without replacement and inserted symmetrically. Corruption levels of 10%, 30%, and 50% are defined relative to the number of clean undirected edges. The 30% setting is used for the main geometry-recovery and multi-baseline comparisons, while all three levels are used for the robustness study. The corruption procedure does not use class labels.

For each dataset, random seed, and corruption level, all compared methods receive the same corrupted graph realization. The present study evaluates suppression of injected candidate edges; it does not evaluate unrestricted recovery of missing edges outside the declared candidate support.

D.4 Real-Graph Preprocessing

Two real graphs are used. Cora contains 2,708 nodes, 5,278 unique undirected edges after preprocessing, 1,433 features, and 7 classes. The graph is symmetrized and deduplicated, self-loops are excluded from the geometric energy, and TF–IDF features are used. Results use five stratified splits.

Roman-Empire contains 22,662 nodes, 32,927 unique undirected edges, 300 features, and 18 classes. The graph is symmetrized and deduplicated, and the released z-scored GloVe features and five public splits are used.

For both real datasets, the released graph is treated as the pre-corruption reference rather than as a known latent ground-truth geometry. Random nonedges are added to this reference graph using the same corruption protocol as above. Labels are not used to construct the injected-edge corruption or the SGF conductances. Accordingly, operator errors on these datasets measure recovery toward the released pre-corruption graph after controlled corruption.

D.5 Matched Baselines

The comparisons are organized around the mechanism being tested. Geometry recovery and corruption robustness compare

1. fixed diffusion, which keeps the observed graph operator fixed;
2. feature reweighting, with $c_{ij}^{\text{feat}} = A_{ij} \exp(-\|\mathbf{h}_i^{\text{ref}} - \mathbf{h}_j^{\text{ref}}\|_2^2/(2\sigma_f^2))$, computed once at $\mathbf{H}_{\text{ref}}$ with validation-selected σ_f;
3. SGF-$\tau = 0$, which retains state-dependent conductances but removes stochastic forcing; and
4. SGF with positive-temperature intrinsic stochastic dynamics.

The mechanism comparison additionally includes adaptive deterministic dynamics, Euclidean-noise dynamics without the centered Stiefel constraint, and SGF without the fidelity term. The contextual comparison at 30% corruption includes GRAND, GRAND++, LGNSDE, and Pro-GNN together

with SGF-$\tau = 0$ and SGF (Chamberlain et al., 2021; Thorpe et al., 2022; Bergna et al., 2025; Jin et al., 2020).

All reported comparisons use the same dataset splits and label protocol. Feature reweighting uses the same encoder architecture, objective, and training settings as SGF. Within the SGF family, shared encoder, decoder, optimization, and evaluation settings are held fixed where applicable; the compared variants differ as specified in Appendix D.9.

D.6 Evaluation Metrics

Operator recovery. The main geometry-recovery metric is the relative Frobenius error

$$e_{\mathrm{op},F} = \frac{\left\| \widehat{\mathbf{L}} - \mathbf{L}_{\mathrm{ref}} \right\|_F}{\max\{\|\mathbf{L}_{\mathrm{ref}}\|_F, \varepsilon_{\mathrm{stab}}\}}.$$

For the synthetic datasets, $\mathbf{L}_{\mathrm{ref}} = \mathbf{L}_\star$ is the clean Laplacian generated before corruption. For Cora and Roman-Empire, $\mathbf{L}_{\mathrm{ref}}$ is the released pre-corruption graph Laplacian. For SGF, $\widehat{\mathbf{L}}$ is formed from the estimated mean conductances.

Corrupted-edge detection. For an observed edge, retained SGF samples give the empirical conductance mean $\widehat{c}_{ij}$ and variance $\widehat{u}_{ij}$. The corruption score follows the definition in the main text,

$$s_{ij}^{\mathrm{corr}} = 1 - \frac{\widehat{c}_{ij}}{A_{ij}} + \kappa \frac{\sqrt{\widehat{u}_{ij}}}{A_{ij} + \varepsilon_{\mathrm{stab}}}.$$

Injected edges are treated as positives, and detection quality is reported by AUROC. This score is used only for controlled injected-edge ranking. A mean-only comparison with $\kappa = 0$ isolates the incremental ranking contribution of the variance term; no calibrated-uncertainty claim is made.

Unsupervised node representations. Representation quality is evaluated using a fixed linear-probe protocol. Accuracy is reported on the clean graph and after adding sampled nonedges at 10%, 30%, and 50% of the clean edge count. The linear-probe regularization parameter is selected on the validation split from the declared grid.

D.7 Equilibrium-Sampling Diagnostics

Sampling diagnostics are reported for the positive-temperature SGF runs used to estimate conductance statistics. Each primary sampling run uses $K = 3$ chains initialized at $\mathbf{H}_{\mathrm{ref}}$ with independent noise streams and $L = 600$ transitions at $h = 0.003$. The first half of each chain is discarded as burn-in, and states are retained every 10 transitions thereafter. Thus, each chain contributes 30 retained states, for 90 retained states across the three chains.

Table 3 reports effective sample size and split-$\widehat{R}$ for the retained mean-conductance trace on Cora and Roman-Empire (Vehtari et al., 2021). These diagnostics measure local finite-chain sampling efficiency and between-chain agreement.

We additionally report three targeted diagnostics. Step-size sensitivity is measured by

$$\Delta_h = \frac{\|\widehat{\mathbf{c}}_{0.003} - \widehat{\mathbf{c}}_{0.001}\|_2}{\|\widehat{\mathbf{c}}_{0.001}\|_2},$$

where $\widehat{\mathbf{c}}_h = (\widehat{c}_{ij,h})_{\{i,j\}\in\mathcal{E}_{\mathrm{obs}}}$ stacks the empirical mean conductances over the candidate support, pooled across the retained states from the K chains at step size h. The comparison uses matched physical integration time. The $h = 0.001$ diagnostic therefore uses $L = 1800$, burn-in 900, and retention every 30 transitions.

Initialization sensitivity is measured from an auxiliary set of independently initialized chains,

$$\mathbf{H}_0^{(a)} = \mathcal{Q}(\mathbf{Z}^{(a)}), \qquad Z_{ij}^{(a)} \sim \mathcal{N}(0,1), \qquad a = 1, \dots, K,$$

using the same frozen model and sampling schedule. For chain a, define the retained mean projector

$$\widehat{\mathbf{P}}^{(a)} = \frac{1}{M} \sum_{m=1}^{M} \frac{\mathbf{H}_m^{(a)} (\mathbf{H}_m^{(a)})^\top}{n}.$$

We report initialization sensitivity as

$$D_{\text{init}} = \frac{2}{K(K-1)} \sum_{a<b} \frac{\left\| \widehat{\mathbf{P}}^{(a)} - \widehat{\mathbf{P}}^{(b)} \right\|_F}{\sqrt{2p}}.$$

Finally,

$$\Delta\text{AUROC}_u = \text{AUROC}(\kappa = 0.15) - \text{AUROC}(\kappa = 0)$$

compares the full corruption score with its mean-only counterpart using the same retained samples.

For each dataset repeat, each diagnostic is first computed as a single repeat-level quantity. In particular, Δ_h uses the conductance means pooled over the retained states from the K chains at each step size; D_{init} is computed from the K auxiliary-chain mean projectors within that repeat; and ΔAUROC_u is computed from the paired full-score and mean-only AUROCs using the same retained samples. ESS and split-$\widehat{R}$ are likewise computed within each repeat from the K primary chains. Table 3 reports the mean and standard deviation of these repeat-level quantities over the five evaluation repeats.

These diagnostics probe discretization sensitivity, initialization dependence, and the incremental ranking contribution of conductance variance within the finite-chain regime. They do not establish exact finite-step Gibbs invariance, global exploration, or calibrated uncertainty. The reported uncertainty quantities are therefore interpreted as finite-chain estimates of model-induced conductance variability.

D.8 Hyperparameter Search Space and Dataset

This section documents the experimental configuration used throughout the reported studies. Table 6 summarizes the SGF hyperparameters, training choices, numerical safeguards, and evaluation protocol, including the settings used for finite-depth learning and sampling-based estimation. Table 7 complements this information with the corresponding dataset statistics, data splits, preprocessing choices, and licenses for both synthetic and real-world benchmarks. The two tables are intended to be read together: the first specifies the model and evaluation settings, whereas the second records the data configuration under which those settings are applied. Together, they collect the implementation and dataset details referenced by the experiments in the main text and make explicit the distinctions between the finite-depth and sampling-based experimental regimes.

The reported configurations distinguish main experimental settings from diagnostic variants. In particular, $\tau \in \{0, 0.035\}$ defines the main deterministic/stochastic comparison, while the wider temperature range is used only for the sensitivity diagnostic. Likewise, the alternative step size and integration schedule are used only for the matched-time discretization check.

Table 6: **Hyperparameters.** Candidate grids and reported settings used in the experiments. Values were fixed before test-time evaluation.

Hyperparameter	Candidate values	Reported value / role
SGF core		
Representation width p	3	fixed before evaluation
Bandwidth σ	$\{0.5, 1, 2\}\,\text{median}(d_{ij}^{\text{ref}})$	$2\,\text{median}(d_{ij}^{\text{ref}})$, with $d_{ij}^{\text{ref}} = \left\|\mathbf{h}_i^{\text{ref}} - \mathbf{h}_j^{\text{ref}}\right\|_2$ computed once from $\mathbf{H}_{\text{ref}}$ and held fixed
Feature bandwidth σ_f	same grid as σ	validation-selected
Fidelity η	20 (SGF), 0 (ablation)	mechanism comparison
Temperature τ	0, 0.035	main comparison; diagnostic sweep over $[0, 0.08]$
Step size h	$\{0.001, 0.003, 0.005\}$	0.003 default; 0.001 diagnostic
Integration steps L	600 (sampling), 420 (finite-depth), 1800 (diagnostic)	matched-time step-size check
Chains K	3	primary and initialization diagnostics
Burn-in / sampling interval	300/10 default; 900/30 diagnostic	matched physical time
Architecture and training		
Encoder	64-dim ReLU	fixed across matched SGF variants
Decoder / evaluation	inner-product decoder; linear probe	fixed protocol
Optimizer	Adam, lr $\{0.005, 0.01, 0.02\}$, wd $\{0, 5\text{e}{-4}\}$	0.01 / 5e−4
Numerical safeguards		
Stabilizer	$\varepsilon_{\text{stab}} = 10^{-8}$	error and score denominators
Retraction check	$\lambda_{\min}(\mathbf{Y}^\top \mathbf{J} \mathbf{Y})$ monitored	floating-point safeguard
Dataset configuration		
SBM	90 nodes, 3 blocks, $p_{\text{in}} = 0.30$, $p_{\text{out}} = 0.035$, 3 features	controlled synthetic graph
Random geometric	90 nodes, 8-NN, heat-kernel, 2 features	symmetrized weights
Swiss roll	90 nodes, latent (t, z), 0.05 noise, 3 features	latent 8-NN graph
Evaluation protocol		
Corruption rate	$\{0, 0.10, 0.30, 0.50\}$	added nonedges / clean-edge count
Evaluation repeats	5	graph seeds or public splits
Linear probe	regularizer grid $\{10^{-4}, \ldots, 10\}$	validation split
Corruption score	$\kappa = 0.15$	fixed before evaluation

Table 7: **Dataset documentation.** Edge counts are unique undirected edges after removing self-loops; synthetic counts vary across graph seeds.

Dataset	Nodes	Edges	Features	Classes	Split	Preprocessing	License
SBM	90	498 ± 25	3	3	5 stratified splits	$p_{\text{in}} = .30$, $p_{\text{out}} = .035$; noisy one-hot	synthetic
Random geometric	90	424 ± 10	2	2	5 stratified splits	8-NN; noisy planar coordinates	synthetic
Swiss roll	90	427 ± 4	3	3	5 stratified splits	8-NN in latent (t, z); noisy ambient coordinates	synthetic
Cora	2,708	5,278	1,433	7	5 stratified splits	undirected, deduplicated; TF–IDF	CC BY-SA
Roman-Empire	22,662	32,927	300	18	5 public splits	undirected, deduplicated; z-scored GloVe	CC BY

D.9 Mechanism Comparisons

The reported comparison study focuses on variants directly associated with the SGF mechanism. On the corrupted SBM benchmark, it compares

1. fixed diffusion;
2. adaptive deterministic dynamics;
3. Euclidean-noise dynamics;
4. SGF without the fidelity term; and
5. full SGF.

These comparisons are reported jointly with relative Laplacian error, injected-edge AUROC, and linear-probe accuracy. The accompanying diagnostic study reports runtime, finite-chain ESS and split-$\widehat{R}$, step-size and initialization sensitivity, and the incremental ranking contribution of conductance variance. Normalized-Laplacian extensions and alternative candidate-support constructions are not treated as separate empirical claims in the present study.