# SIMPLICIAL SHEAF DIFFUSION: LEARNED ANISOTROPIC TRANSPORT ON 2-COMPLEXES

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Simplicial neural networks propagate through Hodge Laplacians whose coupling between coincident simplices is a fixed incidence sign, so the transport they learn is isotropic. Sheaf neural networks learn a map per edge but operate only on graphs. We introduce **Simplicial Sheaf Diffusion (SSD)**, which diffuses node, edge and face cochains jointly through the Hodge Laplacians of a cellular sheaf whose restriction maps are learned from the features and recomputed at every layer. Placing one map on each vertex, one on each triangle and, optionally, a unit-determinant gauge on each edge makes successive coboundaries compose to zero for every value the learner produces, so the operator is the Hodge Laplacian of a sheaf at every layer and its equilibria are the sheaf cohomology. We prove permutation and orientation equivariance for this learned operator and an exact per-coordinate formula for its harmonic dimension, which the normalization we use preserves. In one shared harness with ten neural baselines across five benchmarks, SSD has the highest accuracy on SHREC mesh classification and the lowest error among the neural models on a real ocean-current complex, where removing its triangles raises its error to that of the zero predictor; on the remaining three benchmarks it has the highest mean or is level with the best model. A seed-paired ablation on the two trajectory sets attributes $13.5$ and $8.4$ points of accuracy to the learned maps. Under a random node relabeling the trained model's outputs change only at floating-point noise on every benchmark; on PROTEINS, where we run the same test on the baselines, four simplicial models' outputs do not.

## 1 INTRODUCTION

A triangulated mesh records which edges bound a face as well as which vertices are joined, and a flow on its edges reverses sign when the canonical vertex order flips. *Simplicial neural networks* (Roddenberry et al., 2021; Yang et al., 2022; Yang & Isufi, 2023; Einizade et al., 2025) reach that structure by lifting computation to nodes, edges and triangles through Hodge Laplacians, but the coupling between coincident simplices is a fixed incidence sign with no learnable geometry, so the transport is isotropic. *Sheaf neural networks* (Bodnar et al., 2022; Borgi et al., 2025; Ribeiro et al., 2025) learn exactly that geometry, a restriction map per edge, but operate on the 1-skeleton and have no access to triangles. Learning maps at two incidence levels requires more than adding a level: the maps must satisfy a compatibility condition on every triangle, or successive coboundaries stop composing to zero, $H^1$ is undefined, and the diffusion operator is no longer the Hodge Laplacian of any sheaf.

We introduce **Simplicial Sheaf Diffusion (SSD)**, which attaches restriction maps to *cells* rather than to incidences, one map $R_v$ per vertex, one $F_t$ per triangle and, optionally, a diagonal unit-determinant gauge $S_e$ per edge, and diffuses node, edge and face cochains jointly through the resulting sheaf Hodge Laplacians (Figure 1). Per-cell placement is sufficient for the cochain condition to hold identically, with no constraint, penalty or projection. The node maps are *polar*, an orthogonal factor times a learned per-coordinate scale, because orthogonal maps leave the lower edge Laplacian with no learned content. The face maps are diagonal, which lets the cochain complex split by stalk coordinate. Every generator reads only quantities that relabeling and reorientation leave unchanged and every odd rank is updated through an odd activation without bias, which together give equivariance for an operator that is itself learned.

**Contributions.**



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.

- **A learned operator that is a sheaf Hodge Laplacian at every layer.** Per-cell maps make $\delta_1^{\mathcal{F}} \circ \delta_0^{\mathcal{F}} = 0$ an identity of the parameterization, gauge included (Proposition 2), and the normalized rank-1 operator keeps the Hodge split and, in exact arithmetic, the harmonic dimension (Lemma 4).
- **Equivariance and harmonic control.** Theorem 3 gives permutation and orientation equivariance at all three ranks under stated conditions on the generators and the initialization. Theorem 5 gives $\dim H^1$ as a per-coordinate support-rank count, which fixes the family's endpoints; a trained model sits at the simplicial endpoint, so the theorem bounds what the face map can do to the harmonic space. Proposition 6 states where the learned anisotropy enters and why the node maps are polar.
- **Best or level with the best neural model on five benchmarks.** Against ten neural baselines run in the same harness under one selection rule, SSD leads on SHREC, has the lowest error among them on the North Pacific complex of Yang et al. (2024), where it needs its triangles, has the highest mean on PROTEINS and synthetic by gaps inside the seed spread, and is level with the best model on ocean-drifter. A seed-paired ablation attributes $13.5$ and $8.4$ points of trajectory accuracy to the learned maps (Section 6.5).
- **Invariance, measured.** For every reported row, SSD gives the same outputs, up to float32 summation-order residuals, under a random node relabeling with every derived structure rebuilt, and on PROTEINS its cochains transform correctly at all three ranks; there, where all nine baselines were tested, four of the simplicial ones change their outputs.

## 2 Related Work

**Sheaf neural networks.** Cellular sheaves (Hansen & Ghrist, 2019; Hansen & Gebhart, 2020; Bodnar et al., 2021) equip each graph edge with a *restriction map* between endpoint stalks. Hansen & Ghrist (2019) develop the spectral theory, Hansen & Gebhart (2020) learn the maps, NSD (Bodnar et al., 2022) predicts them from node features, and later work adds polynomial filters (Borgi et al., 2025), complex-valued maps (Fiorini et al., 2025), two maps per edge (Ribeiro et al., 2025) and maps on hyperedge incidences (Duta et al., 2023). All place their maps on one incidence level, so no composition of coboundaries arises and the cochain condition does not arise.

**Simplicial neural networks.** SCoNe (Roddenberry et al., 2021) learns edge-flow representations with the full Hodge Laplacian and sets out the equivariance properties we adopt. SAN (Giusti et al., 2022) adds attention on lower and upper adjacency. Of the two models of Yang et al. (2022), SCN convolves each rank independently while SCCN convolves jointly over one adjacency relating all ranks. SCCNN (Yang & Isufi, 2023) couples ranks through the boundary maps, and Yang et al. (2025) analyze the family through the Hodge spectrum. GSAN (Battiloro et al., 2024) combines graph-level and simplicial attention, and COSIMO (Einizade et al., 2025) replaces discrete filtering with matrix-exponential diffusion.

**Learned transport without a cochain complex, and other symmetries.** Copresheaf topological neural networks (Hajij et al., 2025) assign a learned map to each directed adjacency and set the sheaf consistency condition aside, so their operator is a weighted quiver Laplacian rather than the Laplacian of a cochain complex; SSD retains a well-defined $H^1$, a Hodge Laplacian split into upper and lower parts, and the equivariance of the signed incidence structure. ETNN (Battiloro et al., 2025) makes message passing on combinatorial complexes $E(n)$-equivariant in node positions (Satorras et al., 2021) but uses no boundary matrices, so orientation is not the symmetry it addresses.

## 3 Preliminaries

**Simplicial complexes.** An abstract simplicial complex $\mathcal{X}$ over a vertex set $V$ is a downward-closed collection of subsets; we work in dimension $\leq 2$, with edges $E$ and triangles $T$, each edge stored in its canonical orientation $e = (u, v)$ with $u < v$ and each triangle as $[i, j, k]$ with $i < j < k$. The incidence signs $s_{v,e}$ ($+1$ at the head $v$) and $s_{e,t}$ fill the signed incidence matrices $B_1 \in \mathbb{R}^{|V| \times |E|}$ and $B_2 \in \mathbb{R}^{|E| \times |T|}$. The Hodge Laplacian on edges is $L_1 = B_1^\top B_1 + B_2 B_2^\top$, with

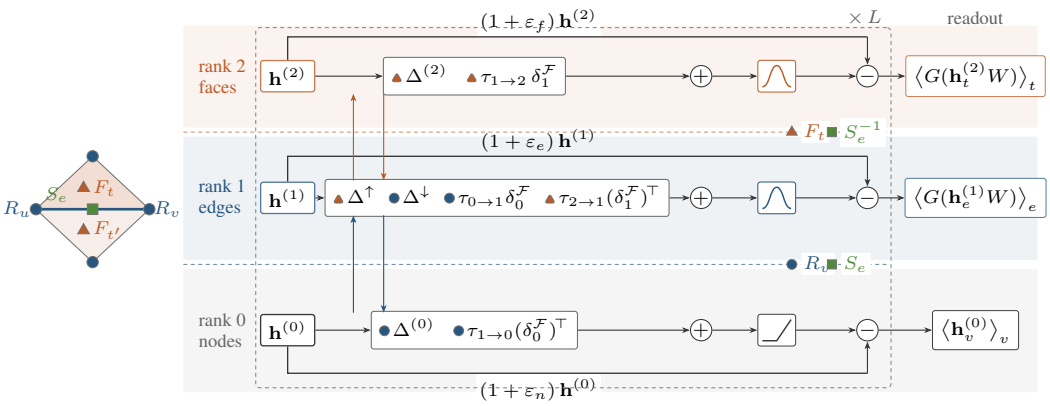



Figure 1: One SSD layer, applied $L$ times. Each rank's box lists its update terms (Eqs. 6 to 8): a $K$-hop Chebyshev filter $\Delta$ in the rank's normalized Laplacian and one coupling per neighboring rank (arrows between lanes); the sum passes through an odd $\operatorname{asinh}$ (a GELU at rank 0) and is subtracted from the residual $(1+\varepsilon)\mathbf{h}^{(k)}$. Glyphs mark where each learned map enters: $R_v$ (circles) through $\delta_0^{\mathcal{F}}$ and its adjoint, $F_t$ (triangles) through $\delta_1^{\mathcal{F}}$ and its adjoint, the optional gauge $S_e$ (squares) on the two rails; the maps sit on cells (left) and are regenerated from $\mathbf{h}^{(0)}$ at every layer. Readout: the mean $\langle\cdot\rangle_\sigma$ over each rank's simplices of $\mathbf{h}^{(0)}$ and, for the odd ranks, of the Gram matrices $G(X) = (X^\top X, XX^\top)$ of $X = \mathbf{h}_\sigma^{(k)} W$, which a flipped simplex leaves unchanged (Section 4.3).

$\ker(L_1) \cong H_1(\mathcal{X};\mathbb{R})$ (Lim, 2020); write $\beta_1 := \dim\ker(L_1)$ and $\beta_1^{\text{graph}} := \dim\ker(B_1)$, which differ by $\operatorname{rank}(B_2)$.

**Cellular sheaves.** A *cellular sheaf* $\mathcal{F}$ on $\mathcal{X}$ (Hansen & Ghrist, 2019; Hansen & Gebhart, 2020; Hu, 2025) assigns a *stalk* $\mathcal{F}(\sigma) \cong \mathbb{R}^d$, of a fixed dimension $d$, to each simplex and a *restriction map* $\mathcal{F}_{\sigma\trianglelefteq\tau}$ to each incident pair, a change of frame from the smaller cell into the larger one. The *coboundary* $\delta_k^{\mathcal{F}}$ transports each $k$-simplex's features into every incident $(k{+}1)$-simplex and takes the signed sum there, so $\delta_0^{\mathcal{F}}$ measures the disagreement between an edge's endpoints in the edge's frame and $\delta_1^{\mathcal{F}}$ the circulation around a triangle:

$$(\delta_0^{\mathcal{F}}\mathbf{x})_e = \sum_{v\trianglelefteq e} s_{v,e}\,\mathcal{F}_{v\trianglelefteq e}\,\mathbf{x}_v, \qquad (\delta_1^{\mathcal{F}}\mathbf{x})_t = \sum_{e\trianglelefteq t} s_{e,t}\,\mathcal{F}_{e\trianglelefteq t}\,\mathbf{x}_e.$$

The sheaf Hodge Laplacians are $L_0^{\mathcal{F}} = (\delta_0^{\mathcal{F}})^\top\delta_0^{\mathcal{F}}$, $L_1^{\mathcal{F}} = L_1^{\downarrow,\mathcal{F}} + L_1^{\uparrow,\mathcal{F}}$ with $L_1^{\downarrow,\mathcal{F}} = \delta_0^{\mathcal{F}}(\delta_0^{\mathcal{F}})^\top$ and $L_1^{\uparrow,\mathcal{F}} = (\delta_1^{\mathcal{F}})^\top\delta_1^{\mathcal{F}}$, and $L_2^{\mathcal{F}} = \delta_1^{\mathcal{F}}(\delta_1^{\mathcal{F}})^\top$. The *harmonic space* is $\ker(L_1^{\mathcal{F}})$. All of this needs the cochain condition

$$\delta_1^{\mathcal{F}} \circ \delta_0^{\mathcal{F}} = 0, \tag{1}$$

which makes $H^1(\mathcal{X};\mathcal{F}) = \ker\delta_1^{\mathcal{F}}/\operatorname{im}\delta_0^{\mathcal{F}}$ well defined; on a functorial sheaf it follows from the agreement of the restriction maps (Hansen & Ghrist, 2019), and in the learned setting nothing forces a network to output maps that agree.

**Theorem 1** (Sheaf Hodge theorem; Hansen & Ghrist, 2019, Theorem 3.1)**.** *Under Eq.* (1), $\ker(L_1^{\mathcal{F}}) \cong H^1(\mathcal{X};\mathcal{F})$ *and* $C^1 = \operatorname{im}\delta_0^{\mathcal{F}} \oplus \ker L_1^{\mathcal{F}} \oplus \operatorname{im}(\delta_1^{\mathcal{F}})^\top$ *orthogonally.*

## 4 Simplicial Sheaf Diffusion

SSD processes a complex $\mathcal{X} = (V, E, T)$ in three stages: projection to node, edge and face cochains, $L$ layers of sheaf Hodge diffusion with $K$-hop polynomial filters, and a task readout. Cochains carry a stalk of dimension $d$ and $C$ channels, $\mathbf{h}^{(k)} \in \mathbb{R}^{|\mathcal{X}_k|\times d\times C}$, and two parities organize what follows: $\mathbf{h}^{(1)}$ is *odd* under edge reorientation, $\mathbf{h}^{(2)}$ under face reorientation, and $\mathbf{h}^{(0)}$ is even.

### 4.1 Restriction maps on cells

We attach a map to each *cell*: one $R_v$ per vertex, reused by every edge containing $v$, and one $F_t$ per triangle, reused by every edge of $t$. The alternative, one map per incident pair, is the general definition

of a cellular sheaf, and it turns Eq. (1) into a constraint on the learner rather than a property of the parameterization (Section 5). An optional third family, one diagonal gauge $S_e$ per edge, changes the frame of the edge stalk consistently on both sides and cancels in the composite. All three are recomputed from the current rank-0 state at every layer, so the diffusion operator is itself learned:

$$R_v = Q_v\ \mathrm{diag}(e^{\mathbf{s}_v}), \qquad F_t = \mathrm{diag}(\mathbf{f}_t), \qquad S_e = \mathrm{diag}(e^{\boldsymbol{\sigma}_e}),\ \ \textstyle\sum_j \sigma_{e,j} = 0, \tag{2}$$

with $Q_v \in O(d)$ a product of $d$ Householder reflections whose vectors are predicted from $\mathbf{h}_v^{(0)}$, $\mathbf{s}_v = c\tanh(g_R(\mathbf{h}_v^{(0)}))$, $\mathbf{f}_t = \tanh(g_F(\mathbf{m}_t))$ with $\mathbf{m}_t$ the *mean* of the three vertex states of $t$, and $\boldsymbol{\sigma}_e = c\tanh(g_S(\mathbf{u}_e))$ centered to sum to zero, where $\mathbf{u}_e$ is an endpoint-symmetric descriptor of the edge built from orientation-invariant quantities and $g_R, g_F, g_S$ are linear or two-layer maps (Appendix B). The node map is *polar*: $\mathbf{s}_v = 0$ recovers an orthogonal map, and the scale head starts at zero so every layer begins there. The scale is what puts learned content into the lower edge Laplacian, since orthogonal maps collapse $L_1^{\downarrow,\mathcal{F}}$ to the isotropic $B_1^\top B_1 \otimes I_d$ (Proposition 6). The face map is diagonal, which makes the complex split by stalk coordinate (Theorem 5); an orthogonal one would collapse $L_1^{\uparrow,\mathcal{F}}$ to $B_2 B_2^\top$. The maps define the two coboundaries,

$$(\delta_0^{\mathcal{F}}\mathbf{x})_e = S_e\big(R_v\mathbf{x}_v - R_u\mathbf{x}_u\big)\ \ (e=(u,v)), \qquad (\delta_1^{\mathcal{F}}\mathbf{y})_t = F_t \sum_{e \trianglelefteq t} s_{e,t}\, S_e^{-1}\mathbf{y}_e, \tag{3}$$

and $S_e = I$ recovers the per-cell family exactly; whether the gauge is on is part of the search.

### 4.2 Normalized operators and spectral filters

We normalize each Laplacian before it enters the layer to bound its spectrum by a constant determined by the complex, as required by a Chebyshev filter, while preserving the Hodge split. Applying a diagonal to both sides of each half, whether shared or not, inserts it between $\delta_1^{\mathcal{F}}$ and $\delta_0^{\mathcal{F}}$ in the product and breaks the split; only a pair conjugated by $A$ and $A^{-1}$ cancels. We instead take adjoints with respect to a family-aware degree inner product, with $D_k$ a positive scalar per $k$-cell (the mean diagonal of its degree block, plus one), and conjugate the rank-1 pair by *one* symmetric $d \times d$ block per edge and one common scalar:

$$\widehat{L}_1^{\downarrow} = \gamma\, A\, \delta_0^{\mathcal{F}} D_0^{-1} (\delta_0^{\mathcal{F}})^\top A, \qquad \widehat{L}_1^{\uparrow} = \gamma\, A^{-1} (\delta_1^{\mathcal{F}})^\top D_2\, \delta_1^{\mathcal{F}}\, A^{-1}, \qquad A_e = \big(P_e^{-1} \,\#\, Q_e\big)^{1/2}, \tag{4}$$

where $P_e$ and $Q_e$ are the diagonal blocks of the two unnormalized operators, $\#$ is the matrix geometric mean, and $\gamma = 1/\max_e \mathrm{tr}(A_e P_e A_e)$. This $A$ gives both halves the same block, $AP_eA = A^{-1}Q_eA^{-1}$ (equal to $P_e \,\#\, Q_e$ when the two blocks commute), so neither dominates, and $\widehat{L}_1^{\uparrow}\widehat{L}_1^{\downarrow}$ contains $\delta_1^{\mathcal{F}}\delta_0^{\mathcal{F}}$ in its middle and vanishes (Lemma 4). On an edge with no incident triangle $Q_e = 0$, so the normalized lower Laplacian there vanishes up to the clamp that keeps $A$ positive definite, and Lemma 4 holds in exact arithmetic only; on a complex with no triangle at all, $\gamma$ restores its scale (Appendix B). The rank-0 and rank-2 operators have no partner to stay orthogonal to and are whitened by their own diagonal block. An incidence count then bounds every spectrum by a constant of the complex (2 at rank 0, the maximum degree and 3 for the lower and upper halves at rank 1, the maximum face count of an edge at rank 2), with no learned quantity involved (Appendix B). Each filter is a shifted Chebyshev polynomial with a vanishing constant,

$$\Delta^L\mathbf{h} = \sum_{j=1}^{K} \tau_j\, q_j(\widehat{L})\, \mathbf{h}, \qquad q_j(\widehat{L}) = T_j\big(\tfrac{2}{b}\widehat{L} - I\big) - T_j(-1)\, I, \tag{5}$$

with learnable $\tau_j$ initialized by least squares to $1 - e^{-\lambda}$ on $[0, b]$; $q_j(0) = 0$, so every filter annihilates the harmonic space.

### 4.3 Diffusion layers, initialization and readout

We first mix each rank by left and right projections, $\tilde{\mathbf{h}}^{(k)} = W_L^{(k)}\mathbf{h}^{(k)}W_R^{(k)}$, as in Hansen & Gebhart (2020), with a bias at rank 0 only. Ranks then exchange information through the coboundaries and

their adjoints, each carrying one incidence sign:

$$\mathbf{h}^{(1)} \leftarrow (1+\varepsilon_e)\,\mathbf{h}^{(1)} - \operatorname{asinh}\Big(\Delta^{\uparrow}\tilde{\mathbf{h}}^{(1)} + \Delta^{\downarrow}\tilde{\mathbf{h}}^{(1)} + \tau_{0\to1}\,\mathbf{n}\odot\delta_0^{\mathcal{F}}\tilde{\mathbf{h}}^{(0)} + \tau_{2\to1}\,(\delta_1^{\mathcal{F}})^{\top}\tilde{\mathbf{h}}^{(2)}\Big), \quad (6)$$

$$\mathbf{h}^{(0)} \leftarrow (1+\varepsilon_n)\,\mathbf{h}^{(0)} - \mathrm{GELU}\Big(\Delta^{(0)}\tilde{\mathbf{h}}^{(0)} + \tau_{1\to0}\,(\delta_0^{\mathcal{F}})^{\top}\tilde{\mathbf{h}}^{(1)}\Big), \quad (7)$$

$$\mathbf{h}^{(2)} \leftarrow (1+\varepsilon_f)\,\mathbf{h}^{(2)} - \operatorname{asinh}\Big(\Delta^{(2)}\tilde{\mathbf{h}}^{(2)} + \tau_{1\to2}\,\delta_1^{\mathcal{F}}\tilde{\mathbf{h}}^{(1)}\Big), \quad (8)$$

with residual weights $\varepsilon_\cdot$ and couplings $\tau_\cdot$ initialized at zero, so every layer starts at the identity. The node-to-edge term carries $n_e = (\deg u + \deg v)^{-1/2}$ because its scale would otherwise dominate the edge pre-activation. The parities fix the activations: every term entering ranks 1 and 2 is odd and takes an odd activation with no bias, rank 0 is even, and $\operatorname{asinh}$ is odd without saturating.

**Initialization.** Given node features $\mathbf{x}_v$, a covariant edge attribute $\mathbf{a}_e^{\mathrm{cov}}$ where the data provides one (a signed flow), and invariant edge and face descriptors $\mathbf{u}_e$ and $\mathbf{u}_t$,

$$\mathbf{h}_v^{(0)} = \mathrm{GELU}(W_{\mathrm{in}}\mathbf{x}_v), \quad \mathbf{h}_e^{(1)} = \boldsymbol{\gamma}_e \odot \operatorname{asinh}\big(W_E[\mathbf{x}_v - \mathbf{x}_u;\, \mathbf{a}_e^{\mathrm{cov}}]\big), \quad \mathbf{h}_t^{(2)} = \boldsymbol{\rho}_t \odot \operatorname{asinh}\big((\delta_1^{\mathcal{F}}\mathbf{h}^{(1)})_t\big), \quad (9)$$

where $W_E$ has no bias, $\boldsymbol{\gamma}_e \in [0.1, 1]$ is a gate from $\mathbf{u}_e$ that attenuates but never silences an edge, and $\boldsymbol{\rho}_t = \mathrm{sigmoid}(W_F\mathbf{u}_t)$. The carrier, the $\operatorname{asinh}$ term, is odd because both blocks of its argument flip under $u \leftrightarrow v$, and the gate is even; the face cochain is odd under face reorientation because $\delta_1^{\mathcal{F}}$ carries one sign. An edge with indistinguishable endpoints and no covariant attribute starts at zero, the only oriented value equivariance allows, and becomes active once diffusion separates its endpoints. Since $\mathbf{h}^{(2)}$ is the sheaf curl of a sheaf gradient, at initialization it is a remainder about one percent as large as $\mathbf{h}^{(1)}$, so SSD rescales it to the root-mean-square of $\mathbf{h}^{(1)}$ per sample by a detached ratio times a learned even gain.

**Readout.** For graph classification we pool each rank on its own simplices: the mean of $\mathbf{h}^{(0)}$ and, for the two odd ranks, the mean of the Gram matrices of a projected cochain, unchanged when a simplex flips. For edge regression a bias-free linear head reads $\mathbf{h}_e^{(1)}$ and $((\delta_1^{\mathcal{F}})^{\top}\mathbf{h}^{(2)})_e$, both odd; trajectory prediction uses the incidence readout of the SCoNe pipeline (Roddenberry et al., 2021).

## 5 Theoretical Analysis

### 5.1 The cochain condition holds by construction

**Proposition 2** (Cochain condition under per-cell placement)**.** *Let $R_v \in \mathbb{R}^{d\times d}$ be one map per node, $F_t \in \mathbb{R}^{d\times d}$ one per triangle, and $S_e$ one invertible diagonal matrix per edge. With the coboundaries of Eq.* (3)*, $\delta_1^{\mathcal{F}} \circ \delta_0^{\mathcal{F}} = 0$ for every choice of $\{R_v\}$, $\{F_t\}$, $\{S_e\}$.*

The proof is in Appendix A. The gauge cancels on every edge, since $\delta_1^{\mathcal{F}}$ applies $S_e^{-1}$ to what $\delta_0^{\mathcal{F}}$ multiplied by $S_e$. Because the node map depends only on the vertex, the lower coboundary becomes the ordinary coboundary of the transported cochain $\mathbf{y}_v = R_v\mathbf{x}_v$. Because the face map depends only on the triangle, it factors out of the upper coboundary. What remains is $F_t$ applied to the boundary of a boundary. Both steps break under per-incidence maps. Per-cell placement is sufficient, not necessary.

### 5.2 Permutation and orientation equivariance

We adopt the framework of Roddenberry et al. (2021), writing $\mathrm{SSD}_{W,\partial}$ for the network with parameters $W$ on a complex with boundary maps $\partial = \{B_1, B_2\}$; their two properties (Definitions 1 and 2, Appendix A) require that relabeling the simplices permutes the output the same way, and that reversing the orientation of an edge or a face negates its cochain while node features stay fixed. For an operator fixed by the complex, orientation equivariance holds if and only if the activation is odd (Roddenberry et al., 2021; Yang et al., 2025); SSD's operator depends on the features, so oddness is not enough.

**Theorem 3** (Permutation and orientation equivariance)**.** *Suppose the generators of $R_v$, $F_t$ and $S_e$ read only quantities invariant under relabeling and reorientation ($\mathbf{h}_v^{(0)}$; the mean $\mathbf{m}_t$; an endpoint-symmetric $\mathbf{u}_e$), the edge and face cochains are initialized by Eq.* (9) *with $\mathbf{a}^{\mathrm{cov}}$ odd and $\mathbf{u}_e, \mathbf{u}_t$ invariant, and the rank-1 and rank-2 paths carry no bias. Then at every layer:* ***(a)*** *for permutation matrices $\{P_k\}$, $\mathrm{SSD}_{W,\partial}(c_j) = P_\ell\, \mathrm{SSD}_{W,P\partial}(P_j c_j)$;* ***(b)*** *under reorientations $\epsilon_1 : E \to \{\pm 1\}$, $\epsilon_2 : T \to \{\pm 1\}$, node states are unchanged, $\hat{\mathbf{h}}_e^{(1)} = \epsilon_1(e)\mathbf{h}_e^{(1)}$ and $\hat{\mathbf{h}}_t^{(2)} = \epsilon_2(t)\mathbf{h}_t^{(2)}$; the graph readout is invariant and the edge head odd.*

We prove it by a parity count over Eqs. (6) to (8): a term landing in an edge or face slot carries an odd number of sign factors and one in a node slot an even number; the $\Delta$ terms and the couplings keep this count as long as the maps themselves are invariant, which the hypothesis on the generators secures. Under node relabeling at inference, which reverses some edge orientations and re-sorts some triangles, the initialization tracks the new orientations and part (b) propagates them (Corollary 7; Section 6.4).

### 5.3 Harmonic dimension and anisotropy

**Lemma 4** (The normalized pair keeps the split)**.** *For any symmetric positive definite $A$ and positive diagonal $D_0, D_2$, the operators of Eq.* (4) *satisfy $\widehat{L}_1^{\uparrow}\widehat{L}_1^{\downarrow} = \widehat{L}_1^{\downarrow}\widehat{L}_1^{\uparrow} = 0$ and $\dim\ker(\widehat{L}_1^{\downarrow} + \widehat{L}_1^{\uparrow}) = \dim\ker(L_1^{\mathcal{F}}) = \dim H^1(\mathcal{X};\mathcal{F})$.*

**Theorem 5** (Sheaf-controlled harmonic dimension)**.** *With node maps $R_v \in \mathrm{GL}(d)$, face maps $F_t = \mathrm{diag}(\alpha_1^t, \dots, \alpha_d^t)$ and invertible diagonal gauges $S_e$, the cochain complex splits by stalk coordinate after the change of basis $\mathbf{y}_v = R_v\mathbf{x}_v$, and*

$$\dim H^1(\mathcal{X};\mathcal{F}) \;=\; \beta_1^{\mathrm{graph}}\, d \;-\; \sum_{j=1}^{d} \mathrm{rank}\big(B_2^{S_j}\big), \qquad \beta_1\, d \;\le\; \dim H^1(\mathcal{X};\mathcal{F}) \;\le\; \beta_1^{\mathrm{graph}}\, d, \tag{10}$$

*where $S_j = \{t \in T : \alpha_j^t \neq 0\}$ and $B_2^{S_j}$ is $B_2$ restricted to those columns. For $S_e = I$ the harmonic space is the direct sum over stalk coordinates $j$ of $\ker(B_1) \cap \ker\big((B_2^{S_j})^\top\big)$, and the gauge, an isomorphism of cochain complexes, leaves the dimension unchanged (Appendix A).*

Polar node maps are invertible by construction, so the hypothesis always holds. Each stalk coordinate decides independently which triangles constrain it: invertible face maps give the simplicial endpoint $\beta_1 d$, vanishing ones the graph-only endpoint $\beta_1^{\mathrm{graph}} d$, and a vanishing coordinate moves the count one triangle at a time. With $F_t = \mathrm{diag}(\tanh(\cdot))$ an exact zero is not reached in training, so every trained model reported here sits at the simplicial endpoint; a coordinate near zero leaves a small eigenvalue rather than a harmonic direction, and the theorem bounds what the face map can do to the harmonic space. Since $q_j(0) = 0$, the filters leave that space untouched.

**What the maps change.** For invertible $R_v$ and $F_t$ the maps $\phi_v = R_v$, $\phi_e = S_e^{-1}$, $\phi_t = F_t^{-1}$ commute with every restriction map, so $\mathcal{F}$ is isomorphic to the constant sheaf $\mathbb{R}^d$ and $H^1(\mathcal{X};\mathcal{F}) \cong H^1(\mathcal{X};\mathbb{R})^d$, the simplicial endpoint above. The model learns the inner product used to take the adjoints of the coboundaries: with $B_k$ standing for $B_k \otimes I_d$, $L_1^{\mathcal{F}} = SB_1^\top W_0 B_1 S + S^{-1}B_2 W_2 B_2^\top S^{-1}$ with $W_0 = \mathrm{diag}_v(R_v R_v^\top)$, $W_2 = \mathrm{diag}_t(F_t^2)$ and $S = \mathrm{diag}_e(S_e)$, a Hodge Laplacian with learned matrix-valued cell weights, whose spectrum and diffusion differ from the constant sheaf's while its harmonic dimension does not (its harmonic representatives move with the gauge and under the normalization of Eq. (4)); Theorem 5 describes the boundary of that regime. Appendix E reads the learned maps and this spectrum off two trained models.

**Proposition 6** (Where the learned anisotropy enters)**.** *For adjacent $u \neq v$ joined by $e$, for edges $e, e'$ sharing the vertex $v$, and for any two edges,*

$$\begin{aligned} (L_0^{\mathcal{F}})_{u,v} &= s_{u,e}s_{v,e}\, R_u^\top S_e^2 R_v, & (L_1^{\downarrow,\mathcal{F}})_{e,e'} &= s_{v,e}s_{v,e'}\, S_e R_v R_v^\top S_{e'}, \\ (L_1^{\uparrow,\mathcal{F}})_{e,e'} &= \textstyle\sum_{t \ni e,e'} s_{e,t}s_{e',t}\, S_e^{-1} F_t^2 S_{e'}^{-1}. \end{aligned} \tag{11}$$

*At $R_v = \epsilon I$ with one sign $\epsilon$ for every vertex, $S_e = I$ and $F_t = cI$ these are $L_0 \otimes I_d$, $B_1^\top B_1 \otimes I_d$ and $c^2 B_2 B_2^\top \otimes I_d$, so the isotropic operators are a point of the family. With orthogonal node maps and no gauge, $L_1^{\downarrow,\mathcal{F}} = B_1^\top B_1 \otimes I_d$ holds for every value of the maps.*

Table 1: Graph classification under one protocol, ten seeds, selection on validation. $|\Delta|$ is the mean absolute change in test AUROC under a random node relabeling of the test graphs, and *logit residual* is the largest absolute change in any test logit under it (0 is exact invariance, $10^{-6}$ float32 summation noise; Section 6.4). SHREC chance is $0.033$. Per-model notes in Appendix C.

| | PROTEINS | | | SHREC |
|---|---|---|---|---|
| **Model** | AUROC | $\lvert\Delta\rvert$ | logit residual | accuracy |
| MLP | $0.601 \pm 0.036$ | 0.0000 | $3.3\times10^{-7}$ | $0.294 \pm 0.042$ |
| GCN | $0.610 \pm 0.046$ | 0.0000 | $4.2\times10^{-7}$ | $0.847 \pm 0.035$ |
| NSD | $0.669 \pm 0.035$ | 0.0002 | $9.5\times10^{-6}$ | $0.839 \pm 0.034$ |
| SCN | $0.646 \pm 0.034$ | 0.0001 | $2.2\times10^{-6}$ | $0.446 \pm 0.048$ |
| SCCN | $0.669 \pm 0.033$ | 0.0656 | 1.8 | $0.602 \pm 0.047$ |
| SCCNN | $0.577 \pm 0.043$ | 0.0372 | $1.1\times10^{3}$ | $0.607 \pm 0.037$ |
| SAN | $0.504 \pm 0.035$ | 0.0342 | $1.7\times10^{-5}$ | $0.807 \pm 0.031$ |
| GSAN | $\underline{0.769 \pm 0.033}$ | 0.0146 | 10.6 | $0.742 \pm 0.043$ |
| COSIMO | $0.736 \pm 0.031$ | 0.0218 | 6.9 | $\underline{0.853 \pm 0.027}$ |
| **SSD (ours)** | $\mathbf{0.773 \pm 0.023}$ | 0.0001 | $2.3\times10^{-6}$ | $\mathbf{0.883 \pm 0.028}$ |

The first block has the form of NSD's coupling $\mathcal{F}_{u\trianglelefteq e}^{\top}\mathcal{F}_{v\trianglelefteq e}$ (Hansen & Ghrist, 2019; Bodnar et al., 2022) with both maps drawn from per-vertex ones, so its transport around any cycle is the identity: SSD learns a per-vertex metric at rank 0, not the holonomy NSD uses. The last block is the degree-one analog, where an isotropic model couples two edges of a triangle through the scalar $s_{e,t}s_{e',t}$ and SSD through $s_{e,t}s_{e',t}S_e^{-1}F_t^2S_{e'}^{-1}$. The middle block is why the node maps are polar: with orthogonal maps it carries no learned parameter and the node maps reach the edge cochain only through $\delta_0^{\mathcal{F}}\tilde{\mathbf{h}}^{(0)}$.

## 6 Experiments

We obtain every result below in one shared harness per benchmark: SSD and each baseline are searched and evaluated with the same code on the same splits under the same selection rule (validation on the graph and North Pacific benchmarks; Section 6.3 for the trajectory sets). Every $\pm$ is the standard deviation over seeds (divisor $n$). A seed-paired contrast is the mean over seeds of the per-seed difference, its $t$ the paired $t$ statistic (Student, 1908) on $n-1$ degrees of freedom, and "$k$ of $n$" counts the seeds on which the difference has the stated sign. Appendix C gives budgets, search spaces and per-model provenance. Baselines are SCoNe (Roddenberry et al., 2021), SCN and SCCN (Yang et al., 2022), SCCNN (Yang & Isufi, 2023), SAN (Giusti et al., 2022), GSAN (Battiloro et al., 2024), COSIMO (Einizade et al., 2025), NSD (Bodnar et al., 2022), GCN (Kipf & Welling, 2017) and an MLP, and every model receives the invariant edge and face attributes each dataset provides.

### 6.1 Graph classification

PROTEINS (Borgwardt et al., 2005) has no intrinsic triangles, so its 3-cliques serve as triangles; each of ten seeds draws a stratified 64/16/20 split, and the score is test AUROC. SHREC (the SHREC'11 meshes, Lian et al., 2011, as TopoNetX's `shrec_16`, Hajij et al., 2024) has 600 meshes over 30 classes with faces intrinsic to the surface, the official 480/120 split and a 20% validation cut of the training meshes per seed. Every model gets 50 search trials of three repetitions; the top three configurations are retrained over ten seeds and the best validation mean is reported on test.

On SHREC, SSD leads at $0.883$ against $0.853$ for COSIMO (paired by seed, $t = 2.9$; six seeds of ten, three ties, one against), with GCN at $0.847$ and NSD at $0.839$ next, both behind SSD on eight seeds of ten ($t = 2.6$ and $3.2$). On PROTEINS, SSD has the highest mean AUROC, $0.773$ against $0.769$ for GSAN, a gap inside both spreads (paired by seed, $t = 0.4$, six seeds of ten) that we read as a tie, with COSIMO third. PROTEINS is weak evidence about higher-order structure: on its molecular attributes a structure-agnostic baseline matches the graph networks (Errica et al., 2020; Coupette et al., 2025), and our protocol feeds the node degree only (Appendix C), so its numbers are not comparable with published ones; we use it for the relabeling measurement rather than as evidence about rank 2 (Appendix D retrains SSD without triangles).

Table 2: North Pacific edge-flow imputation, test RMSE over five seeds (zero predictor 1.00). Neural models only; best in bold, second underlined.

| Model | RMSE |
|---|---|
| SCN | $\underline{0.596 \pm 0.015}$ |
| SCCN | $0.599 \pm 0.015$ |
| SCCNN | diverged |
| SAN | $1.024 \pm 0.024$ |
| COSIMO | $0.619 \pm 0.015$ |
| **SSD (ours)** | $\mathbf{0.444 \pm 0.024}$ |

Table 3: Trajectory prediction, SWA accuracy in percent over five seeds. One test trajectory is one point on synthetic and 2.5 on ocean-drifter. Best mean in bold, second underlined; a tie is broken by the seed spread.

| Model | synthetic | ocean-drifter |
|---|---|---|
| SCoNe | $66.20 \pm 1.47$ | $39.50 \pm 2.45$ |
| SCN | $64.00 \pm 3.41$ | $41.00 \pm 4.90$ |
| SCCN | $\underline{67.00 \pm 3.90}$ | $47.00 \pm 7.97$ |
| COSIMO | $66.20 \pm 2.71$ | $\mathbf{48.00 \pm 5.79}$ |
| **SSD (ours)** | $\mathbf{67.40 \pm 4.13}$ | $\underline{48.00 \pm 8.43}$ |

### 6.2 Edge-flow imputation on an independent ocean-current complex

Every complex above is clique-lifted or a small mesh. Yang et al. (2024) release a simplicial complex over North Pacific drifter observations, built following Chen et al. (2021) on NOAA records, with the standardized edge flow they evaluate on (1,500 nodes, 19,023 edges, 90,551 triangles). We use their split and metric: per seed, 20% of the edges are observed and the other 80% are predicted, scored by RMSE. Each neural model selects its hyperparameters (20 search trials) and stopping step on a fifth of the observed edges, runs five seeds, and at test is given every observed edge (Appendix C).

Among the neural models SSD has the lowest error, $0.444 \pm 0.024$ against $0.596 \pm 0.015$ for SCN (Table 2).[1] Without its triangles SSD scores 0.994 (Table A1), the zero predictor's error, so its accuracy here depends on rank 2. SAN does not improve on the zero predictor, and SCCNN, run as released, diverges in every trial.

### 6.3 Trajectory prediction

The two trajectory sets of Roddenberry et al. (2021) predict the next edge of a trajectory from its edge-flow prefix: **synthetic** (500 shortest paths on a Delaunay complex) and **ocean-drifter** (200 buoy tracks on a hexagonal complex; Appendix C). Every model follows the published pipeline with 30 search trials over its own hyperparameter space, whose objective is the accuracy of the stochastic weight average (SWA) over three seeds, then five seeds at 1,000 epochs. The pipeline's 80/20 split is both the selection and the reporting split for every model, so the column is comparable across models but not held out for any of them.

SSD has the highest synthetic mean, 67.40 against 67.00 for SCCN (seed-paired $t = 0.2$), and ties COSIMO on ocean-drifter at 48.00 (Table 3), the only model to lead or tie for the lead on both. Every gap is inside one seed spread (100 and 40 test trajectories); we claim no margin on either.

### 6.4 Relabeling test

Theorem 3 promises that relabeling the nodes, which reverses about half the edge orientations and re-sorts the triangles, leaves the trained model's output unchanged. For each reported row we retrain the selected configuration through the same harness and seeds, and score it on the test data and on a copy with a random node relabeling and every derived structure (orientations, triangles, eigenpairs) rebuilt; an identity control rebuilds the test set unpermuted.

On three rows the metric is unchanged on every seed; it moves by at most $5 \times 10^{-7}$ on North Pacific and, through a near-tie in the logits, $4 \times 10^{-4}$ on PROTEINS. The output residuals are float32 summation-order effects, at most $5.7 \times 10^{-6}$ on the graph sets, $5.2 \times 10^{-4}$ at output scale 8 to 20 on North Pacific, $3.0 \times 10^{-4}$ at scale 62 to 98 on synthetic and $3.9 \times 10^{-3}$ at scale 5.8 on ocean-drifter (Table A2); the identity control is exactly zero, and on PROTEINS the node, edge and face states also transform correctly at every rank (Appendix D). Of the PROTEINS baselines in Table 1, five are output-invariant and four are not: GSAN, COSIMO, SCCN and SCCNN move their test logits by 1.8

[1] The edge Gaussian processes of Yang et al. (2024), run on our splits from their released code, score $0.343 \pm 0.002$ (HC diffusion). They regress one edge signal on a fixed complex with a spectral prior from 500 eigenpairs of its Hodge Laplacian and seven hyperparameters, and are not defined for our other tasks (Appendix C).

Table 4: Ablation of the map families on the trajectory sets: SWA accuracy in percent, mean and standard deviation over the five seeds of Table 3 (40 and 100 test trajectories), node family by face family. The full model is the (polar, per-coordinate) cell; the map-free cell, both maps fixed at the identity under the same normalization and recipe, is (identity, identity).

**ocean-drifter**

| node \ face | per-coord. | scalar | identity |
|---|---|---|---|
| polar | $48.0 \pm 8.4$ | $44.5 \pm 6.2$ | $31.5 \pm 9.9$ |
| orthogonal | $38.0 \pm 10.4$ | $41.5 \pm 7.5$ | $31.0 \pm 3.4$ |
| identity | $42.0 \pm 7.5$ | $32.5 \pm 13.7$ | $34.5 \pm 7.6$ |

**synthetic**

| node \ face | per-coord. | scalar | identity |
|---|---|---|---|
| polar | $67.4 \pm 4.1$ | $66.2 \pm 2.3$ | $63.6 \pm 4.1$ |
| orthogonal | $63.8 \pm 1.9$ | $68.4 \pm 2.3$ | $64.8 \pm 2.7$ |
| identity | $68.2 \pm 1.2$ | $66.0 \pm 2.9$ | $59.0 \pm 2.4$ |

to $10^3$ and their AUROC by $0.015$ to $0.066$ on average, COSIMO by $3.0$ already under the identity control, where only the recomputed eigenpairs can differ. As run, these implementations lack the equivariance available to them (Yang et al., 2025).

### 6.5 What the learned maps carry

We ask whether the learned maps of Proposition 6 improve accuracy on the two trajectory sets, where all samples share one complex. A $3 \times 3$ grid crosses the node family {polar, orthogonal, identity} with the face family {per-coordinate, scalar, identity} (Table 4; *scalar*: one gate per face, averaged over coordinates before the $\tanh$, so without the selection of Theorem 5; *identity*: fixed, not learned). Every cell trains SSD's selected configuration over the five seeds of Table 3 with the same initialization and split per seed; the other eight cells are not re-tuned, which favors the full cell. The primary contrast is the full model against the map-free cell; the others change one family from the full cell or from the map-free cell (Appendix D averages over the other family). All are seed-paired differences with $t$ on four degrees of freedom ($t = 2.8$ is the two-sided 5% point, uncorrected for the number of contrasts).

On ocean-drifter, SSD's accuracy depends on the learned maps: the full model beats the map-free cell by $13.5$ points on five seeds of five ($t = 10.6$; the cell spreads are mostly the shared seed, which pairing removes), and the map-free $34.5$ is below every row of Table 3. The families interact. The face gate alone recovers $7.5$ of the $13.5$ ($t = 3.2$, four of five), the node map alone none ($31.5$, within noise of the map-free $34.5$, $t = -0.7$), and the two together $13.5$. From the full cell, fixing the face gate at the identity therefore costs $16.5$ ($t = 5.5$, five of five) and fixing the node map costs $6.0$ ($t = 3.5$, four of five), each conditional on the other family being learned. The node map's gain needs its scale: orthogonal, the same family without it, costs $10.0$ from the full cell ($t = 5.7$, five of five) and scores below the identity node map ($38.0$ against $42.0$; Appendix D). The scalar gate keeps $13$ of the face gate's $16.5$ points; the per-coordinate selection adds $3.5$ ($t = 1.5$, three of five), unresolved at five seeds, although Appendix E shows the trained gate using it.

On synthetic, the learned maps also raise accuracy by $8.4$ points over map-free on five seeds of five ($t = 4.3$), and the face gate alone carries that gain: (identity, per-coordinate) scores $68.2$, $9.2$ over map-free ($t = 6.8$, five of five), adding the node map then costs $0.8$, and the node map alone recovers $4.6$ ($63.6$, $t = 2.6$, four of five). Appendix D gives the axis-averaged contrasts and interprets the gate patterns on both sets.

## 7 Discussion and Conclusion

SSD diffuses at all three ranks of a 2-complex through a learned operator that is a sheaf Hodge Laplacian for every parameter value and is equivariant. Its outputs are invariant under relabeling, whereas those of four simplicial baselines on PROTEINS are not.

**Limitations and future work.** SSD is $2.8\times$ slower per step than COSIMO and uses $7\times$ its peak memory on the same meshes (Appendix D), because the maps are regenerated at every layer and every edge carries a $d \times d$ block whose roots are iterated; layer checkpointing recovers about half of the memory, most of which is the autograd tape, for 35% more time. The construction stops at triangles; a sheaf on $k$-simplices would need the condition of Proposition 2 at every consecutive pair of ranks, which per-cell placement suggests but we have not built. Meshes carrying a physical field, as in mesh-based simulation, are a natural next domain.

### AI USE STATEMENT

In this work, we used generative AI tools to implement methods and experiment scripts, to prepare the datasets, to run and monitor experiments, to analyze and help interpret results, to give feedback on the experimental protocol, and to assist with writing, including the text of the proofs. The authors designed the method and the experiments and formulated the mathematical claims; we have not used generative AI tools to generate synthetic datasets, and translation and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to create figures, search the literature, format references and edit the paper for readability. We have reviewed all AI-assisted work: the authors checked the proofs and verified the reported results, and AI-written code was tested against reference implementations. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

### ETHICS STATEMENT

This work uses only publicly released benchmark datasets and involves no human subjects, no personally identifying information and no data collection of our own. We are not aware of a way in which the methods here create a direct risk of harm.

### REPRODUCIBILITY STATEMENT

Section 4 and Appendix B give the full architecture, including the parameterization of the three map families, the normalization, the filters, the initialization and the readouts. Appendix C states the splits, the search spaces and budgets, the selection procedure and the provenance of every baseline implementation, and Appendix D the relabeling protocol. The code is included in the supplementary material.

## REFERENCES

Claudio Battiloro, Lucia Testa, Lorenzo Giusti, Stefania Sardellitti, Paolo Di Lorenzo, and Sergio Barbarossa. Generalized simplicial attention neural networks. *IEEE Transactions on Signal and Information Processing over Networks*, 2024. doi: 10.1109/TSIPN.2024.3487618.

Claudio Battiloro, Ege Karaismailoğlu, Mauricio Tec, George Dasoulas, Michelle Audirac, and Francesca Dominici. E(n) equivariant topological neural networks. In *International Conference on Learning Representations (ICLR)*, 2025.

Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Liò, Guido Montufar, and Michael Bronstein. Weisfeiler and Lehman Go Cellular: CW networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Cristian Bodnar, Francesco Di Giovanni, Benjamin Paul Chamberlain, Pietro Liò, and Michael M Bronstein. Neural Sheaf Diffusion: A topological perspective on heterophily and oversmoothing in GNNs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Alessio Borgi, Fabrizio Silvestri, and Pietro Liò. Polynomial neural sheaf diffusion: A spectral filtering approach on cellular sheaves. *arXiv preprint arXiv:2512.00242*, 2025.

Karsten M Borgwardt, Cheng Soon Ong, Stefan Schönauer, SVN Vishwanathan, Alex J Smola, and Hans-Peter Kriegel. Protein function prediction via graph kernels. In *Bioinformatics*, volume 21, pp. i47–i56, 2005.

Yu-Chia Chen, Weicheng Wu, Marina Meilă, and Ioannis G. Kevrekidis. Helmholtzian eigenmap: Topological feature discovery and edge flow learning from point cloud data. *arXiv preprint arXiv:2103.07626*, 2021.

Corinna Coupette, Jeremy Wayland, Emily Simons, and Bastian Rieck. No metric to rule them all: Toward principled evaluations of graph-learning datasets. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 2025.

Iulia Duta, Giulia Cassarà, Fabrizio Silvestri, and Pietro Liò. Sheaf hypergraph networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Aref Einizade, Dorina Thanou, Fragkiskos D Malliaros, and Jhony H Giraldo. Continuous simplicial neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli. A fair comparison of graph neural networks for graph classification. In *International Conference on Learning Representations (ICLR)*, 2020.

Stefano Fiorini, Hakan Aktas, Iulia Duta, Stefano Coniglio, Pietro Morerio, Alessio Del Bue, and Pietro Liò. Sheaves reloaded: A directional awakening. *arXiv preprint arXiv:2506.02842*, 2025.

L Giusti, C Battiloro, P Di Lorenzo, S Sardellitti, and S Barbarossa. Simplicial attention neural networks. *arXiv preprint arXiv:2203.07485*, 2022.

Mustafa Hajij, Mathilde Papillon, Florian Frantzen, Jens Agerberg, Ibrahem AlJabea, Rubén Ballester, Claudio Battiloro, Guillermo Bernárdez, Tolga Birdal, et al. TopoX: A suite of Python packages for machine learning on topological domains. *arXiv preprint arXiv:2402.02441*, 2024.

Mustafa Hajij, Lennart Bastian, Sarah Osentoski, et al. Copresheaf topological neural networks: A generalized deep learning framework. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. MeshCNN: A network with an edge. *ACM Transactions on Graphics*, 38(4):1–12, 2019.

Jakob Hansen and Thomas Gebhart. Sheaf neural networks. In *NeurIPS 2020 Workshop on Topological Data Analysis and Beyond*, 2020. arXiv:2012.06333.

Jakob Hansen and Robert Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3(4):315–358, 2019. arXiv:1808.01513.

Chuan-Shen Hu. Cellular sheaves on higher-dimensional structures. *arXiv preprint arXiv:2505.23993*, 2025.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In *Advances in Neural Information Processing Systems*, 2020.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL `https://openreview.net/forum?id=SJU4ayYgl`.

Zhouhui Lian, Afzal Godil, Benjamin Bustos, Mohamed Daoudi, Jeroen Hermans, Shun Kawamura, Yukinori Kurita, Guillaume Lavoué, Hien Van Nguyen, Ryutarou Ohbuchi, et al. SHREC'11 Track: Shape retrieval on non-rigid 3D watertight meshes. In *Eurographics Workshop on 3D Object Retrieval (3DOR)*, 2011.

Lek-Heng Lim. Hodge Laplacians on graphs. *SIAM Review*, 62(3):685–715, 2020.

André Ribeiro, Ana Luiza Tenório, Juan Belieni, Amauri H Souza, and Diego Mesquita. Cooperative sheaf neural networks. In *International Conference on Learning Representations (ICLR)*, 2025.

T. Mitchell Roddenberry, Nicholas Glaze, and Santiago Segarra. Principled simplicial neural networks for trajectory prediction. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

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

Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.

Maosheng Yang and Elvin Isufi. Convolutional learning on simplicial complexes. *arXiv preprint arXiv:2301.11163*, 2023.

Maosheng Yang, Viacheslav Borovitskiy, and Elvin Isufi. Hodge-compositional edge Gaussian processes. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2024.

Maosheng Yang, Geert Leus, and Elvin Isufi. Hodge-aware convolutional learning on simplicial complexes. *Transactions on Machine Learning Research (TMLR)*, 2025.

Ruochen Yang, Frederic Sala, and Paul Bogdan. Efficient representation learning for higher-order data with simplicial complexes. In *Learning on Graphs Conference (LoG)*, 2022.

# A PROOFS

Throughout, $D_1$ and $D_2$ in this appendix denote diagonal $\pm 1$ matrices when the context is reorientation (Definition 2) and positive degree scalars when the context is normalization (Eq. (4)); each use says which.

## A.1 EQUIVARIANCE DEFINITIONS

We restate the two properties of Roddenberry et al. (2021) in our notation.

**Definition 1** (Permutation equivariance (Roddenberry et al., 2021, Property 1))**.** *Let $\mathcal{X}$ be a simplicial complex with boundary maps $\{\partial_k\}$, let $\{P_k\}$ be permutation matrices matching the dimensions of the cochain spaces $\{C_k\}$, and define $[P\partial]_k := P_{k-1}\partial_k P_k^\top$. A simplicial network $f$ is* permutation equivariant *if $f_{W,\partial}(c_j) = P_\ell\, f_{W,P\partial}(P_j\, c_j)$.*

**Definition 2** (Orientation equivariance (Roddenberry et al., 2021, Property 2))**.** *Let $\{D_k\}_{k=0}^2$ be diagonal $\pm 1$ matrices with $D_0 = I$, and define $[D\partial]_k := D_{k-1}\partial_k D_k$. A simplicial network $f$ is* orientation equivariant *if $f_{W,\partial}(c_j) = D_\ell\, f_{W,D\partial}(D_j\, c_j)$.*

The face signs $D_2$ are part of the definition, so the face cochain must also be odd under face reorientation; this is why the theorem covers the rank-2 update.

## A.2 PROOF OF PROPOSITION 2 (COCHAIN CONDITION)

That coboundaries compose to zero when the restriction maps agree along every triangle is the standard property of cellular sheaves (Hansen & Ghrist, 2019). What the proposition adds is that the per-cell parameterization makes the maps agree for every value a generator outputs, so the property needs no constraint on the learner.

*Proof.* Fix a node cochain $\mathbf{x}$ and a triangle $t = [i, j, k]$ with $i < j < k$, whose boundary edges in canonical orientation are $e_{ij}, e_{ik}, e_{jk}$ with signs $s_{e_{ij},t} = +1$, $s_{e_{ik},t} = -1$, $s_{e_{jk},t} = +1$. By Eq. (3), for every edge $e = (u, v)$, $S_e^{-1}(\delta_0^{\mathcal{F}}\mathbf{x})_e = R_v\mathbf{x}_v - R_u\mathbf{x}_u =: \mathbf{y}_v - \mathbf{y}_u$, so the gauge cancels on every edge and

$$(\delta_1^{\mathcal{F}}\delta_0^{\mathcal{F}}\mathbf{x})_t = F_t\big[(\mathbf{y}_j - \mathbf{y}_i) - (\mathbf{y}_k - \mathbf{y}_i) + (\mathbf{y}_k - \mathbf{y}_j)\big] = F_t\,\mathbf{0} = \mathbf{0}.$$

The proof uses two facts: $R_v$ depends only on $v$, so the same transported value $\mathbf{y}_v$ appears on both edges of $t$ containing $v$, once with each sign; and $F_t$ depends only on $t$, so it factors out of the sum. Since $t$ and $\mathbf{x}$ were arbitrary, $\delta_1^{\mathcal{F}} \circ \delta_0^{\mathcal{F}} = 0$. □

**Remark 1.** *With per-incidence node maps $R_{v \trianglelefteq e}$ the bracket becomes $R_{j \trianglelefteq e_{ij}}\mathbf{x}_j - R_{j \trianglelefteq e_{jk}}\mathbf{x}_j + \cdots$, which vanishes only if the two maps at $j$ agree, and with per-incidence face maps nothing factors out. The exact requirement is that for every triangle and every vertex of it, the composite map through either edge be the same; per-cell placement satisfies it for every value of the maps, and it is the only requirement the gauge has to respect, which is why $S_e$ and $S_e^{-1}$ must be the same tensor.*

## A.3 PROOF OF THEOREM 3 (PERMUTATION AND ORIENTATION EQUIVARIANCE)

We write $D_1 = \mathrm{diag}(\epsilon_1)$ and $D_2 = \mathrm{diag}(\epsilon_2)$ for the reorientation of edges and faces, acting on a cochain by $\hat{\mathbf{h}}^{(1)} = D_1\mathbf{h}^{(1)}$, $\hat{\mathbf{h}}^{(2)} = D_2\mathbf{h}^{(2)}$ (each sign multiplies the whole $d \times C$ block of its cell), and $\hat{\mathbf{h}}^{(0)} = \mathbf{h}^{(0)}$. Reorienting an edge $e = (u, v)$ swaps which endpoint is head and tail, so $s_{v,e} \mapsto -s_{v,e}$ for both endpoints, and it flips $s_{e,t}$ for every face containing $e$; reorienting a face flips $s_{e,t}$ for its three edges. Hence, as operators on cochains,

$$\hat{\delta}_0 = D_1\delta_0^{\mathcal{F}}, \qquad \hat{\delta}_0^\top = (\delta_0^{\mathcal{F}})^\top D_1, \qquad \hat{\delta}_1 = D_2\delta_1^{\mathcal{F}} D_1, \qquad \hat{\delta}_1^\top = D_1(\delta_1^{\mathcal{F}})^\top D_2, \tag{12}$$

*provided the maps $R_v, F_t, S_e$ are unchanged by the reorientation.* That is the first ingredient and it is the hypothesis on the generators: $R_v$ is a function of $\mathbf{h}_v^{(0)}$, which reorientation does not touch; $F_t$ is a function of the mean of the three vertex states of $t$, which is symmetric in them and independent of every sign; and $S_e$ is a function of an endpoint-symmetric descriptor of invariant quantities. Any odd-degree dependence in a generator, for instance on $\mathbf{h}_e^{(1)}$ or on $\mathbf{x}_v - \mathbf{x}_u$, would make the operator itself orientation dependent, and the count below would fail at exactly the coupling terms.

**Step 1: the initialization is odd.** Under $u \leftrightarrow v$ the carrier argument $[\mathbf{x}_v - \mathbf{x}_u;\, \mathbf{a}_e^{\mathrm{cov}}]$ maps to its negative by the hypothesis on $\mathbf{a}^{\mathrm{cov}}$; $W_E$ has no bias, so the carrier is odd, $\mathrm{asinh}$ is odd, and the gate $\boldsymbol{\gamma}_e$ is a function of the invariant $\mathbf{u}_e$, so $\hat{\mathbf{h}}_e^{(1)} = \epsilon_1(e)\mathbf{h}_e^{(1)}$ after Eq. (9). Then $\hat{\delta}_1\hat{\mathbf{h}}^{(1)} = D_2\delta_1^{\mathcal{F}}D_1D_1\mathbf{h}^{(1)} = D_2\delta_1^{\mathcal{F}}\mathbf{h}^{(1)}$, and since $\mathrm{asinh}$ is odd and the face gate $\boldsymbol{\rho}_t$ is invariant, $\hat{\mathbf{h}}^{(2)} = D_2\mathbf{h}^{(2)}$. The rank-2 rescaling multiplies each face block by a positive scalar (a ratio of root mean squares, both even) times an even channel gain, so it preserves the sign.

**Step 2: the normalized operators conjugate by the signs.** The degree scalars $D_k$ of Eq. (4) are sums of traces of $R_vR_v^\top$, $F_t^2$ and $S_e^{\pm 2}$ and carry no sign. The diagonal blocks $P_e = S_e(\sum_{v\in e} R_vR_v^\top/D_{0,v})S_e$ and $Q_e = S_e^{-1}(\sum_{t\ni e} D_{2,t}F_t^2)S_e^{-1}$ are products of two incidence signs at the same cell and are unchanged, so $A_e$ and $\gamma$ are unchanged. Therefore, by Eq. (12), $\widehat{L}_1^{\downarrow}$ and $\widehat{L}_1^{\uparrow}$ become $D_1\widehat{L}_1^{\downarrow}D_1$ and $D_1\widehat{L}_1^{\uparrow}D_1$ (the face signs cancel in the second), $\widehat{L}_0$ is unchanged, and $\widehat{L}_2$ becomes $D_2\widehat{L}_2D_2$. A polynomial in a conjugated operator is the conjugated polynomial, $q_j(DLD) = D\,q_j(L)\,D$, and the Chebyshev shift $b$ is a count over the complex.

**Step 3: parity count of the three updates.** Mixing $\tilde{\mathbf{h}}^{(k)} = W_L\mathbf{h}^{(k)}W_R$ is linear and, at ranks 1 and 2, bias free, so it commutes with a sign on each cell. In Eq. (6), $\Delta^{\uparrow}\tilde{\mathbf{h}}^{(1)}$ and $\Delta^{\downarrow}\tilde{\mathbf{h}}^{(1)}$ map to $D_1(\cdot)$ by Step 2 (two signs on the operator, one on the argument); the coupling $\delta_0^{\mathcal{F}}\tilde{\mathbf{h}}^{(0)}$ maps to $D_1\delta_0^{\mathcal{F}}\tilde{\mathbf{h}}^{(0)}$, and $\mathbf{n}$ is a degree count; the coupling $(\delta_1^{\mathcal{F}})^\top\tilde{\mathbf{h}}^{(2)}$ maps to $D_1(\delta_1^{\mathcal{F}})^\top D_2D_2\tilde{\mathbf{h}}^{(2)} = D_1(\delta_1^{\mathcal{F}})^\top\tilde{\mathbf{h}}^{(2)}$. Every term is odd, $\mathrm{asinh}$ is odd and elementwise, $\varepsilon_e$ is a scalar, so the updated edge cochain is $D_1$ times the unreoriented one. In Eq. (7), $\Delta^{(0)}\tilde{\mathbf{h}}^{(0)}$ is unchanged and $(\delta_0^{\mathcal{F}})^\top\tilde{\mathbf{h}}^{(1)}$ maps to $(\delta_0^{\mathcal{F}})^\top D_1D_1\tilde{\mathbf{h}}^{(1)}$, so every term is even and the GELU may be any function. In Eq. (8), $\Delta^{(2)}\tilde{\mathbf{h}}^{(2)}$ maps to $D_2(\cdot)$ and $\delta_1^{\mathcal{F}}\tilde{\mathbf{h}}^{(1)}$ to $D_2\delta_1^{\mathcal{F}}D_1D_1\tilde{\mathbf{h}}^{(1)} = D_2\delta_1^{\mathcal{F}}\tilde{\mathbf{h}}^{(1)}$, so every term is odd and $\mathrm{asinh}$ closes it. Induction over layers gives part (b) at every layer.

**Step 4: readouts.** The Gram matrix of a projected cochain, $\mathbf{u}\mathbf{u}^\top$ with $\mathbf{u} = \mathbf{h}_\sigma^{(k)}W$ and $W$ bias free, is unchanged when the block of $\sigma$ is multiplied by $\pm 1$, and the mean of $\mathbf{h}^{(0)}$ is even, so the graph readout is invariant. The edge head is bias free and linear in $\mathbf{h}_e^{(1)}$ and $((\delta_1^{\mathcal{F}})^\top\mathbf{h}^{(2)})_e$, both of which map to $\epsilon_1(e)$ times themselves, so the head is odd, which is the correct covariance for a signed flow.

**Part (a): permutation equivariance.** Every operation above is defined through incidence: scatter and gather along $(u, v)$ and along the three edges of each triangle, with $R_v$ computed from the state at $v$, $F_t$ from the states at the vertices of $t$, and $S_e$ from the descriptor of $e$. None references a global index, the pooled readout is index free, and the Chebyshev shifts are maxima of counts. Relabeling the cells of each rank by $\{P_k\}$ therefore permutes every cochain by the same matrices. □

**Corollary 7** (Node relabeling at inference)**.** *A node relabeling $\pi$ acts on a stored complex as a permutation $P_0$ of the nodes, a permutation $P_1$ of the edges composed with $D_1 = \mathrm{diag}(\epsilon_1)$ where $\epsilon_1(e) = -1$ exactly when the canonical order of the endpoints of $e$ reverses, and a permutation $P_2$ of the triangles composed with $D_2$ where $\epsilon_2(t)$ is the parity of the permutation that re-sorts the vertex triple of $t$. Rebuilding the complex and the inputs from the relabeled data reproduces exactly these signs, since $\mathbf{x}_v - \mathbf{x}_u$ and a stored signed flow are recomputed in the new canonical order. By Theorem 3 the cochains of the relabeled model are $P_kD_k$ times the original ones at every rank, and the graph readout and the reported metrics are unchanged.*

### A.4 Proof of Lemma 4 (the normalized pair keeps the split)

*Proof.* Write $M = \widehat{L}_1^{\downarrow} = \gamma A\delta_0^{\mathcal{F}}D_0^{-1}(\delta_0^{\mathcal{F}})^\top A$ and $N = \widehat{L}_1^{\uparrow} = \gamma A^{-1}(\delta_1^{\mathcal{F}})^\top D_2\delta_1^{\mathcal{F}}A^{-1}$ with $A$ symmetric positive definite (block diagonal over edges) and $D_0, D_2$ positive diagonal. Both are symmetric positive semidefinite. Then

$$NM = \gamma^2 A^{-1}(\delta_1^{\mathcal{F}})^\top D_2\,(\delta_1^{\mathcal{F}}A^{-1}A\delta_0^{\mathcal{F}})\,D_0^{-1}(\delta_0^{\mathcal{F}})^\top A = \gamma^2 A^{-1}(\delta_1^{\mathcal{F}})^\top D_2\,(\delta_1^{\mathcal{F}}\delta_0^{\mathcal{F}})\,D_0^{-1}(\delta_0^{\mathcal{F}})^\top A = 0$$

by Proposition 2, and $MN = (NM)^\top = 0$. For symmetric $M, N \succeq 0$ with $MN = 0$, $\langle M\mathbf{x}, N\mathbf{y}\rangle = \mathbf{x}^\top MN\mathbf{y} = 0$, so $\mathrm{im}\,M \perp \mathrm{im}\,N$ and $\mathrm{rank}(M + N) = \mathrm{rank}\,M + \mathrm{rank}\,N$.

Now $\operatorname{rank} M = \operatorname{rank}(D_0^{-1/2}(\delta_0^{\mathcal{F}})^\top A) = \operatorname{rank}\delta_0^{\mathcal{F}}$ and $\operatorname{rank} N = \operatorname{rank}(D_2^{1/2}\delta_1^{\mathcal{F}}A^{-1}) = \operatorname{rank}\delta_1^{\mathcal{F}}$ because $A$, $D_0$, $D_2$ are invertible. Hence $\dim\ker(M+N) = d|E| - \operatorname{rank}\delta_0^{\mathcal{F}} - \operatorname{rank}\delta_1^{\mathcal{F}} = \dim\ker\delta_1^{\mathcal{F}} - \dim\operatorname{im}\delta_0^{\mathcal{F}} = \dim H^1(\mathcal{X};\mathcal{F})$, the last equality using $\operatorname{im}\delta_0^{\mathcal{F}} \subseteq \ker\delta_1^{\mathcal{F}}$. The unnormalized $L_1^{\mathcal{F}}$ is the case $A = I$, $D_0 = D_2 = I$, $\gamma = 1$. □

The same argument does not go through for a normalization that scales $L_1^{\downarrow,\mathcal{F}}$ and $L_1^{\uparrow,\mathcal{F}}$ by diagonals on both sides, shared or not, because a matrix then sits between $\delta_1^{\mathcal{F}}$ and $\delta_0^{\mathcal{F}}$ in $NM$; that is the mechanism Section 4.2 describes. Numerically, applying the normalized pair to a gradient field $\mathbf{y} = \widehat{L}_1^{\downarrow}\mathbf{x}$, the ratio $\|\widehat{L}_1^{\uparrow}\mathbf{y}\|/\|\mathbf{y}\|$ is 0.049 to 0.059 with mismatched diagonals and $1.3\times10^{-4}$ to $3.6\times10^{-3}$ with one shared diagonal, whereas the conjugated pair gives zero to float precision.

### A.5 Proof of Theorem 5 (sheaf-controlled harmonic dimension)

*Proof.* Let $S = \operatorname{blockdiag}(S_e)$, $R = \operatorname{blockdiag}(R_v)$ and $F = \operatorname{blockdiag}(F_t)$, the first two invertible, and write a $k$-cochain with $d$ stalk coordinates as $d$ scalar cochains $\mathbf{z}_{\cdot j}$. By Eq. (3), $\delta_0^{\mathcal{F}} = S\,(B_1^\top \otimes I_d)\,R$ and $\delta_1^{\mathcal{F}} = F\,(B_2^\top \otimes I_d)\,S^{-1}$. Since $\dim H^1 = \dim\ker\delta_1^{\mathcal{F}} - \dim\operatorname{im}\delta_0^{\mathcal{F}}$ is a property of the cochain complex alone, and $S$, $R$ are invertible,

$$\dim\operatorname{im}\delta_0^{\mathcal{F}} = \operatorname{rank}(B_1^\top \otimes I_d) = d\,\operatorname{rank} B_1, \qquad \dim\ker\delta_1^{\mathcal{F}} = \dim\ker\big(F(B_2^\top \otimes I_d)\big).$$

Because $F$ is diagonal, $F(B_2^\top \otimes I_d)\mathbf{w} = 0$ holds iff for every triangle $t$ and every coordinate $j$ with $\alpha_j^t \neq 0$ the scalar circulation $(B_2^\top \mathbf{w}_{\cdot j})_t$ vanishes, that is, iff $(B_2^{S_j})^\top \mathbf{w}_{\cdot j} = 0$ for every $j$. Hence $\dim\ker\delta_1^{\mathcal{F}} = \sum_j \big(|E| - \operatorname{rank} B_2^{S_j}\big)$, and

$$\dim H^1(\mathcal{X};\mathcal{F}) = \sum_{j=1}^{d}\big(|E| - \operatorname{rank} B_1 - \operatorname{rank} B_2^{S_j}\big) = \sum_{j=1}^{d}\big(\beta_1^{\mathrm{graph}} - \operatorname{rank} B_2^{S_j}\big),$$

which is Eq. (10); the bounds follow from $0 \le \operatorname{rank} B_2^{S_j} \le \operatorname{rank} B_2 = \beta_1^{\mathrm{graph}} - \beta_1$. For the harmonic space with $S = I$, Theorem 1 gives $\ker L_1^{\mathcal{F}} = \ker\delta_1^{\mathcal{F}} \cap \ker(\delta_0^{\mathcal{F}})^\top$, and $(\delta_0^{\mathcal{F}})^\top \mathbf{z} = R^\top(B_1 \otimes I_d)\mathbf{z} = 0$ iff $B_1\mathbf{z}_{\cdot j} = 0$ for every $j$, so the harmonic space is $\bigoplus_j \big(\ker B_1 \cap \ker(B_2^{S_j})^\top\big)$, one copy of the graph cycle space cut down by the triangles that coordinate $j$ keeps. The gauge is an isomorphism of cochain complexes (a change of basis on the edge stalks), so it moves the harmonic representatives but not the dimension. □

Two remarks. Invertibility of $R_v$ is what removes the node maps from the count entirely: the polar family has $\det R_v = \det Q_v\, e^{\sum_j s_{v,j}} \neq 0$ for every parameter value. And by Lemma 4 the count is also the harmonic dimension of the operator the layer applies.

### A.6 Proof of Proposition 6 (blocks)

The block formulas are those of a sheaf Laplacian (Hansen & Ghrist, 2019) evaluated at per-cell maps; the proposition records where each learned family enters them.

*Proof.* From Eq. (3) the adjoints are $((\delta_0^{\mathcal{F}})^\top \mathbf{y})_v = \sum_{e\ni v} s_{v,e} R_v^\top S_e \mathbf{y}_e$ and $((\delta_1^{\mathcal{F}})^\top \mathbf{z})_e = \sum_{t\ni e} s_{e,t} S_e^{-1} F_t \mathbf{z}_t$. Composing, the off-diagonal block of $L_0^{\mathcal{F}} = (\delta_0^{\mathcal{F}})^\top \delta_0^{\mathcal{F}}$ at $(u,v)$ collects the single edge $e = (u,v)$: $s_{u,e}s_{v,e}R_u^\top S_e^2 R_v$. The block of $L_1^{\downarrow,\mathcal{F}} = \delta_0^{\mathcal{F}}(\delta_0^{\mathcal{F}})^\top$ at $(e,e')$ collects the vertices shared by $e$ and $e'$: for $e \neq e'$ sharing $v$ it is $s_{v,e}s_{v,e'}S_e R_v R_v^\top S_{e'}$, and the diagonal block is $\sum_{v\in e} S_e R_v R_v^\top S_e$. The block of $L_1^{\uparrow,\mathcal{F}} = (\delta_1^{\mathcal{F}})^\top \delta_1^{\mathcal{F}}$ at $(e,e')$ collects the triangles containing both: $\sum_t s_{e,t}s_{e',t} S_e^{-1} F_t^2 S_{e'}^{-1}$. At $R_v = \epsilon I$ with one sign $\epsilon$ shared by every vertex (so $\epsilon^2 = 1$ cancels in every product $R_u^\top R_v$ and $R_v R_v^\top$; with mixed signs $L_0^{\mathcal{F}}$ is a signed Laplacian instead), $S_e = I$ and $F_t = cI$, the blocks reduce to $s_{u,e}s_{v,e}I$, $s_{v,e}s_{v,e'}I$ and $c^2\sum_t s_{e,t}s_{e',t}I$, which are the blocks of $L_0 \otimes I_d$, $B_1^\top B_1 \otimes I_d$ and $c^2 B_2 B_2^\top \otimes I_d$. If every $R_v$ is orthogonal and $S_e = I$ then $R_v R_v^\top = I$ in every block of $L_1^{\downarrow,\mathcal{F}}$, so $L_1^{\downarrow,\mathcal{F}} = B_1^\top B_1 \otimes I_d$ whatever the maps. □

Both endpoints are reachable. The Householder product has determinant $(-1)^d$, which excludes $I$ for odd $d$ but not $-I$, and $-I$ gives the same Laplacians; $\mathbf{s}_v = 0$ is the initialization of the scale head; and $F_t = cI$ with $|c| < 1$ is in the range of the $\tanh$. What the $\tanh$ excludes is $F_t = \pm I$ exactly, so the simplicial endpoint of Theorem 5 is reached by any non-vanishing constant $c$, which the normalization then rescales.

## B Architectural Details

**Map generators.** $Q_v$ is the product of $d$ Householder reflections $I - 2\mathbf{v}_k\mathbf{v}_k^\top/\|\mathbf{v}_k\|^2$ whose vectors are the output of a linear map (or a two-layer map with a GELU, the *map generator width* of the search) applied to the flattened $\mathbf{h}_v^{(0)} \in \mathbb{R}^{dC}$. No squashing function is applied to the pre-parameters, because a bounded pre-parameter keeps the reachable set away from $\pm I$. The log-scale head $\mathbf{s}_v = \tanh(\cdot)$ is a second linear map from the same input with weights and bias initialized at zero, so training starts in the orthogonal family, and the scale cap $c = 1$ bounds the singular values of $R_v$ to $[e^{-1}, e]$. The face generator is a two-layer map from the mean of the three flattened vertex states with a $\tanh$ output. The gauge generator is a two-layer map with a $\tanh$ output and a zero-initialized last layer, applied to the invariant edge descriptor $\mathbf{u}_e$, and centered so that $\sum_j \sigma_{e,j} = 0$. Its condition number is at most $e^2$.

**Invariant descriptors and gates.** On every benchmark $\mathbf{u}_e$ contains the magnitude of the covariant edge attribute where one exists, an observation indicator, a constant, and the dataset's own invariant edge attributes (the endpoint-degree sum on PROTEINS; the ten geometric quantities of SHREC), each standardized per sample; $\mathbf{u}_t$ contains the means over the three edges of $t$ of the first two and the dataset's face attributes (seven on SHREC). The edge gate is $\boldsymbol{\gamma}_e = 0.1 + 0.9\,\mathrm{sigmoid}(g_\gamma(\mathbf{u}_e))$, one value per edge on the shared-complex benchmarks and one per channel on the graph benchmarks, where a scalar gate would let SHREC's ten geometric quantities reach the model through a single number per edge while the baselines take all ten as input channels. The gate multiplies rather than concatenates because $\mathbf{h}^{(1)}$ is odd and the descriptors are even.

**Degrees and normalization.** The degree scalars are one positive number per cell, the mean diagonal of the cell's degree block plus one: $D_{0,v} = \frac{1}{d}\sum_{e\ni v} \bar{g}_e^2 \operatorname{tr}(R_v R_v^\top) + 1$, $D_{1,e} = \frac{1}{d}\big[\bar{g}_e^2(\operatorname{tr} R_u R_u^\top + \operatorname{tr} R_v R_v^\top) + \bar{g}_e^{-2}\sum_{t\ni e}\operatorname{tr} F_t^2\big] + 1$ and $D_{2,t} = \frac{1}{d}\sum_{e\in t}\bar{g}_e^{-2}\operatorname{tr} F_t^2 + 1$, with $\bar{g}_e^2$ the mean squared diagonal of $S_e$ (one for the ungauged model). The additive one is a floor that keeps the normalization defined when a face map vanishes. The scalar graph degree equals the sheaf degree only for orthogonal maps, so the family-aware form is required for the polar family. The rank-1 blocks are $P_e = S_e\big(R_u R_u^\top/D_{0,u} + R_v R_v^\top/D_{0,v}\big)S_e$ and $Q_e = S_e^{-1}\big(\sum_{t\ni e} D_{2,t}F_t^2\big)S_e^{-1}$, the diagonal blocks of $\delta_0^{\mathcal{F}} D_0^{-1}(\delta_0^{\mathcal{F}})^\top$ and $(\delta_1^{\mathcal{F}})^\top D_2 \delta_1^{\mathcal{F}}$. The block $A_e = G_e^{1/2}$ with $G_e = P_e^{-1} \# Q_e = P_e^{-1/2}(P_e^{1/2}Q_eP_e^{1/2})^{1/2}P_e^{-1/2}$ is the solution of $A_eP_eA_e = A_e^{-1}Q_eA_e^{-1}$, so both halves have the same block (the geometric mean $P_e \# Q_e$ when the two commute); $\gamma = 1/\max_e \operatorname{tr}(A_eP_eA_e)$ puts that block below one in operator norm, since the trace bounds the largest eigenvalue of a positive semidefinite block. $Q_e$ can be singular, since a coordinate of $F_t$ can be zero, which is the mechanism of Theorem 5; $G_e$ is therefore ridged relative to its own trace ($10^{-6}$) before the root, which bounds $A_e^{-1}$ while leaving the directions that carry signal alone. On an edge with no incident triangle, $Q_e = 0$ (its diagonal is clamped at $10^{-12}$ before the mean), so the common block is of order $10^{-6}$: the clamp keeps $A_e$ positive definite; $A_e$ is of order $10^{-3}$ relative to a balanced edge. The normalized lower Laplacian is numerically switched off on that edge (at unit maps its diagonal block is $5 \times 10^{-8}$ on the pendant edge of a triangle, against $0.5$ on the triangle's edges), while the couplings $\delta_0^{\mathcal{F}}\tilde{\mathbf{h}}^{(0)}$ and $(\delta_1^{\mathcal{F}})^\top\tilde{\mathbf{h}}^{(2)}$ do not pass through $A_e$ and are unaffected. Lemma 4 therefore holds in exact arithmetic, and in float32 the operator the layer applies has a numerical kernel larger than $\dim H^1$ by up to $d$ per face-free edge (fewer when a cycle runs through such an edge, Appendix E). $\gamma$ is taken per sample; on a complex with no triangle at all it is taken over these small blocks and restores the scale, so a triangle-free graph keeps its lower Laplacian, whereas a single triangle switches it off on every other edge (measured on one triangle joined to a path of six edges: $5 \times 10^{-8}$ on each path edge). Every reported result was measured with this operator; on PROTEINS, where 1,109 of 1,113 graphs carry a triangle and most edges lie in none, rank-1 diffusion is confined to the neighborhood of the triangles. All square roots and inverse roots are Newton–Schulz iterations (8 steps in float32, 14 in float64),

never eigendecompositions, because the blocks are routinely degenerate at initialization ($R_v R_v^\top = I$) and the eigendecomposition backward divides by eigenvalue gaps. The rank-0 and rank-2 operators are whitened by the inverse root of their own diagonal blocks, $M_{0,v} = R_v^\top \operatorname{diag}(\sum_{e \ni v} D_{1,e}\, \mathbf{g}_e^2) R_v$ and $M_{2,t} = \sum_{e \in t} F_t^2 / (D_{1,e} \mathbf{g}_e^2)$.

**Spectral bounds and filters.** With every operator normalized so that its own diagonal block has operator norm at most one, the incidence-count lemma $L \preceq k \cdot \operatorname{blockdiag}(L)$, with $k$ the number of cells one co-cell couples, bounds the spectrum: an edge couples its two endpoints ($\widehat{L}_0$, $b = 2$), a node couples its incident edges ($\widehat{L}_1^{\downarrow}$, $b =$ maximum degree), a triangle couples its three edges ($\widehat{L}_1^{\uparrow}$, $b = 3$), and an edge couples the faces holding it ($\widehat{L}_2$, $b =$ maximum number of faces on an edge). No learned quantity enters, so the bounds hold for whatever the maps become; on the trajectory complexes the rank-1 lower half measures $4.4$ against a bound of $11$, and a loose bound reduces filter resolution without affecting correctness. In a batch of different complexes each block takes its own count, so a prediction never depends on which samples share its batch. Each filter's $\tau \in \mathbb{R}^K$ is initialized by least squares on a 256-point grid of $[0, b]$ against $1 - e^{-\lambda}$ in the basis $\{q_j\}$. The monomial basis has condition number about $2{,}590$ at $K = 3$ against $15$ for the Chebyshev basis, and $\tau = (1, 0, \ldots)$ amplifies the top modes of the upper operator by $1.33$ per layer, which the least-squares start does not.

**Rank-2 scale matching.** $\mathbf{h}^{(2)}$ is the sheaf curl of a sheaf gradient, and $\delta_1^{\mathcal{F}} \delta_0^{\mathcal{F}} = 0$ holds to numerical precision, so what reaches rank 2 at initialization is a remainder. Measured on SHREC meshes at the reported configuration, the mean per-face $|\mathbf{h}_t^{(2)}|$ is $1.1\%$ of the mean per-edge $|\mathbf{h}_e^{(1)}|$, and in a quadratic readout that becomes one part in ten thousand. SSD multiplies $\mathbf{h}^{(2)}$ by the detached ratio of the per-sample root mean squares of $\mathbf{h}^{(1)}$ and $\mathbf{h}^{(2)}$ times a learned even channel gain, which moves rank 2's share of the pooled graph vector's norm at initialization from $1.3 \times 10^{-4}\%$ to $0.9\%$ (rank 1 holds $0.8\%$ either way). On SHREC the matched and unmatched variants score $0.877 \pm 0.026$ and $0.897 \pm 0.023$ over ten seeds at the search's best trial, which is not the configuration Table 1 reports; the gap is one seed spread (paired $t = 2.1$ on test, $-0.6$ on validation), so the term restores the scale but does not improve accuracy.

**Readouts.** Graph classification pools each rank on its own simplices: the mean of the flattened $\mathbf{h}^{(0)}$; for each of the two odd ranks, the mean over its simplices of the upper triangles of the two Gram matrices of a bias-free projection $\mathbf{h}_\sigma^{(k)} W \in \mathbb{R}^{d \times r}$ ($r = 8$), one over stalk directions and one over channels, passed through $\operatorname{asinh}$; and three size features, $\log(1 + |E|)$, $\log(1 + |T|)$ and an indicator of whether the sample has faces. The concatenation feeds a two-layer head. A readout that pushes the odd ranks to the vertices through a signed incidence operator and then averages over vertices telescopes edge by edge, and at a trained configuration rank 2 contributes $0.00\%$ of the pooled vector; the Gram readout avoids this because a quadratic form pools on the simplex itself. The edge regression head is $\mathbf{w}_1^\top \mathbf{h}_e^{(1)} + \mathbf{w}_2^\top ((\delta_1^{\mathcal{F}})^\top \mathbf{h}^{(2)})_e$ with no bias. The trajectory pipeline's incidence readout $\mathbf{z}_v = \sum_{e \ni v} s_{v,e} \mathbf{h}_e^{(1)}$ is shared by every model in that table.

**Dropout, activation, determinism.** Dropout is elementwise on the cochains entering the diffusion (the map generators read the clean state), which preserves parity because a non-negative mask commutes with a sign. Every sum over real values is a sorted segment sum rather than an atomic scatter, and the filter initialization solves its normal equations in float64 rather than calling a threaded least-squares routine. Four independent processes then agree bitwise on logits, loss and every gradient, and two 100-epoch runs of one configuration agree to all printed digits. The cost is about $8\%$ of a step.

## C Experimental Details

**Graph classification.** PROTEINS (Borgwardt et al., 2005) has 1,113 graphs. Triangles are the 3-cliques of the graph, 30,501 in all, and 1,109 graphs carry at least one. All models receive the node degree as the node feature and the endpoint-degree sum as an invariant edge attribute. We do not use an OGB molecular set (Hu et al., 2020): under the same lift, ogbg-molhiv has 1,106 triangles across its 41,127 molecules and $98\%$ of the molecules have none, so its rank-2 level would be nearly empty.

SHREC is the mesh collection TopoNetX (Hajij et al., 2024) ships as `shrec_16`: the SHREC'11 non-rigid watertight meshes (Lian et al., 2011) in the MeshCNN preprocessing (Hanocka et al., 2019), 600 meshes over 30 classes, 480 training and 120 test in the collection's own partition, each simplified to 750 edges and 500 faces. All models receive the same 6-dimensional node features (coordinates, normals), 10-dimensional invariant edge features (dihedral angle, span, angles, ratios) and 7-dimensional invariant face features (area, normal, angles). On PROTEINS each seed draws a stratified 64/16/20 split. On SHREC the test set is fixed and each seed cuts 20% of the training meshes for validation. Training is AdamW with a cosine schedule, batch size 64, at most 100 epochs, early stopping on the validation metric with patience 30 before stochastic weight averaging over the last quarter, for every model. Search: 50 trials of three repetitions per model with the validation metric as the objective, then the top three configurations retrained over seeds 0 to 9 and the one with the highest validation mean reported on test. The SSD search space is $d \in \{2, 3, 4, 5\}$, layers $\{2, 3, 4\}$, channels $\{16, 24, 32, 48\}$, map generator width $\{16, 32\}$, gauge on or off, dropout $[0, 0.5]$, and learning rate, weight decay, sheaf-map weight decay and label smoothing on log or linear ranges, and $K \in \{1, 2, 3\}$; the reported models have 51,768 (PROTEINS) and 123,900 (SHREC) parameters. The baseline spaces cover each model's own width, depth and optimizer ranges.

**Notes on two rows of Table 1.** GSAN's SHREC row uses $\tanh$ in place of its published sigmoid cell activation; with the sigmoid it scores $0.138 \pm 0.044$ under the same protocol. SAN's PROTEINS $|\Delta|$ of $0.0342$ is not an equivariance failure: its logits are invariant to $1.7 \times 10^{-5}$, and the metric moves because a predictor at AUROC $0.50$ reorders near-ties.

**Per-model provenance.** SCN, SCCN, SCCNN and SAN are the TopoModelX (Hajij et al., 2024) implementations, GSAN is our own port of the published layer, and NSD, COSIMO, GCN and MLP run from their official codebases. COSIMO's SHREC entry is our re-run on this 30-class task: the COSIMO paper reports SHREC-16 as a regression on the label index with a normalized squared error, so its published number is not comparable and is not what we report. When a baseline's published update leaves the elementwise nonlinearity open, we use the choice under which training succeeds. GSAN is absent from North Pacific. Its attention is local to the support of each Laplacian, but the implementation we run, a port of the published layer, stores each attention matrix densely, so the rank-2 one alone has $90{,}551^2$ entries (30.5 GiB in float32) and a layer forms several, whatever the width; a sparse implementation would remove this limit and we did not write one. SAN does run there: TopoModelX's SAN forms its attention Laplacians and its harmonic projection as dense $19{,}023^2$ matrices at every step, so we apply the same modules with the same parameters through sparse products instead. In float64 on a 300-node subcomplex the two passes agree to within $10^{-14}$ in outputs and gradients; in float32 on the full complex the outputs agree to $3 \times 10^{-6}$ relative. NSD, GCN and the MLP are graph models and run on graph classification only.

**North Pacific.** The complex and the edge flow are the artifact released with Yang et al. (2024), built following Chen et al. (2021) on publicly archived NOAA drifter records: their boundary matrices, from which our preprocessing derives only the edge list and the face node-triples, and the edge flow their experiments evaluate on, standardized over the edges (mean 0, standard deviation 1). Nothing is rebuilt from the raw records and no data is added. The split is theirs: per seed a random 20% of the edges is observed and the other 80% is the test set. A fifth of the observed edges is held out to select hyperparameters and the stopping step. A neural model's input is the flow on known edges: the signed flow (covariant), its magnitude and an observation indicator (invariant), with the standardized node degree as the rank-0 input for every model. The known edges are the other four fifths of the observed ones at validation and every observed edge at test, the same 20% the Gaussian processes condition on; Table A1 also reports the test error with only the four fifths as input. Training minimizes the mean squared error on a random 25% sub-mask of the four fifths hidden at each step (masked self-supervision), with Adam, a cosine schedule, evaluation every 100 steps and early stopping on validation RMSE with patience 10. Search: 20 trials at 1,500 steps on seed 0 selecting on validation RMSE, then five seeds at 3,000 steps on identical splits. Seed 0's split is therefore seen by the search for every model. Validation and test do not overlap, so this is not test leakage, and the comparison is symmetric. The SSD search space is the learning rate, $d \in \{2, 4, 6\}$, channels $\{8, 16, 32\}$, layers 2 to 4 and $K \in \{1, 2, 3\}$, with a linear map generator and the gauge on or off; the selected model has 10,918 parameters, with $K$ and the channel count at the top of their ranges. The row "SSD, 1-skeleton only" of Table A1 is the same model and search space with every triangle removed and its

own 20 trials; its selected model has 3,161 parameters. SCCNN diverges in all 20 trials and all five seeds: its reference implementation filters with powers of the unnormalized Laplacians, whose largest eigenvalues on this complex are $38.3$ (rank 0 and the lower rank-1 part) and $30.4$ (the upper rank-1 part and rank 2), and we run it as released. The Gaussian processes are the kernels and training code Yang et al. (2024) release: the 500 released eigenpairs split into gradient and curl modes, a constant mean and a scale kernel, conditioned on every observed edge. We train them with the recipe of their released ocean-current notebook, Adam at learning rate 0.1 for 200 iterations, from five initializations, keeping the one with the highest marginal likelihood. On their own seed-5 split this recipe gives test MSE $0.119$ (RMSE $0.345$), their reported HC diffusion result. Beside the fit-only column, Table A1 gives each Gaussian process conditioned on the same four fifths, so in each column every model predicts from the same known edges; the neural models also use the held-out fifth for search and early stopping.

**Trajectory prediction.** Both sets come from Roddenberry et al. (2021): synthetic trajectories are shortest paths on a Delaunay triangulation of 150 nodes, 370 edges and 234 faces (500 trajectories, 400 train and 100 test), and ocean-drifter trajectories are GPS buoy tracks on a hexagonal complex of 133 nodes, 320 edges and 186 faces (200 trajectories, 160 and 40). All models use the 2D coordinates as node features and the signed 1-hop edge flow $\in \{-1, 0, +1\}$ as the covariant edge attribute; SSD's invariant descriptor is its magnitude. Training is 1,000 epochs with a cosine schedule, batch size 64, and SWA from epoch 750, reported as the SWA model's 1-hop accuracy. Search: 30 trials per model whose objective is the mean SWA accuracy of seeds 0 to 2, then seeds 0 to 4 at the selected configuration. The SSD space is the learning rate, weight decay, sheaf-map weight decay, $d \in \{2, \dots, 6\}$, channels $\{4, 6, 8, 12, 16\}$, layers 2 to 5, $K \in \{1, 2, 3\}$, dropout $[0, 0.5]$, generator width $\{16, 32\}$ and the gauge on or off; the selected models have 5,213 (synthetic) and 5,322 (ocean-drifter) parameters. The baseline rows are the models' own searches at thirty trials under the same pipeline.

**Compute.** All experiments ran on NVIDIA A100 80 GB GPUs.

## D Additional Results

**PROTEINS without triangles.** Removing all 30,501 triangles, and the face attributes with them, and retraining SSD's reported PROTEINS configuration on the same ten splits without a new search gives test AUROC $0.766 \pm 0.021$ against $0.773 \pm 0.023$ for the full model. The full model is higher on nine seeds of ten by $0.007$ on average (paired $t = 2.7$), a third of the seed spread. The comparison favors the full model, whose configuration was selected with the triangles present, so the gap is not the rank-2 contribution under a separate graph-only search, which we did not run.

**North Pacific, every model.** Table A1 lists every model of Section 6.2, every Gaussian process kernel of Yang et al. (2024), and both amounts of known flow at test. The Gaussian processes are the most accurate models on this task, under both amounts. They take no input features: the covariance of two edges is a spectral function of the complex,

$$k(e, e') = \sum_{i \in G} u_i(e)\, \phi_G(\lambda_i)\, u_i(e') + \sum_{i \in C} u_i(e)\, \phi_C(\lambda_i)\, u_i(e'),$$

where $G$ and $C$ are the gradient and the curl modes among the 500 released eigenpairs $(\lambda_i, u_i)$ of the Hodge Laplacian, and each profile is a diffusion, a Matérn or an inverse-Laplacian one ($\phi(\lambda) \propto 1/\lambda$, the Hodge Laplacian kernels). The HC variants parameterize $\phi_G$ and $\phi_C$ separately, the non-HC variants share one profile, and the prediction is the posterior mean given the known flows; the HC diffusion kernel fits seven scalars by marginal likelihood. We keep them out of Table 2 because they are a different kind of method, a regressor for one edge signal on one fixed complex with no learned representation, not defined for the classification and trajectory tasks that every neural model here also runs.

**Relabeling protocol.** For the graph benchmarks the test graphs are rebuilt under a random node relabeling (permutation seed 0): edges are re-sorted into canonical order, which reverses the orientation of about half of them, triangles are re-sorted, and every derived structure a baseline uses, including

Table A1: North Pacific edge-flow imputation, test RMSE over five seeds on the 80% of edges not observed (the zero predictor scores 1.00). *20% known*: a neural model is given every observed edge at test and a Gaussian process is conditioned on every observed edge, the protocol of Table 2. *16% known*: a neural model is given only the four fifths of the observed edges it was trained on, and a Gaussian process is conditioned on the same four fifths.

| **Model** | **20% known** | **16% known** |
|---|---|---|
| GP, HC diffusion | $0.3427 \pm 0.0025$ | $0.3472 \pm 0.0022$ |
| GP, HC Matérn | $0.3433 \pm 0.0025$ | $0.3493 \pm 0.0032$ |
| GP, HC Hodge Laplacian | $0.3445 \pm 0.0024$ | $0.3498 \pm 0.0018$ |
| GP, non-HC diffusion | $0.3452 \pm 0.0019$ | $0.3510 \pm 0.0009$ |
| GP, non-HC Matérn | $0.3441 \pm 0.0024$ | $0.3481 \pm 0.0020$ |
| GP, non-HC Hodge Laplacian | $0.3445 \pm 0.0024$ | $0.3497 \pm 0.0017$ |
| SCN | $0.5964 \pm 0.0147$ | $0.5642 \pm 0.0225$ |
| SCCN | $0.5987 \pm 0.0151$ | $0.6156 \pm 0.0098$ |
| SCCNN | diverged | diverged |
| SAN | $1.0243 \pm 0.0241$ | $1.0206 \pm 0.0229$ |
| COSIMO | $0.6194 \pm 0.0147$ | $0.6109 \pm 0.0091$ |
| SSD, 1-skeleton only | $0.9943 \pm 0.0035$ | $0.9950 \pm 0.0031$ |
| SSD | $0.4441 \pm 0.0241$ | $0.4068 \pm 0.0092$ |

Laplacian eigenpairs, is recomputed. The trained model is scored on the stored and on the relabeled test set. $\Delta$ is the difference in the metric and the residual is the largest absolute change in a test logit. An identity control rebuilds the same test set with the identity relabeling, which recomputes eigenpairs and nothing else. For the shared-complex benchmarks, where every sample lives on one complex, the relabeled model is a second instance of the same architecture on the relabeled complex with the trained parameters copied across (and, for the baselines that carry batch normalization, their running statistics; SSD has no normalization layer), fed the relabeled observations and pulled back with the induced signs. For SSD every row reproduces its recorded per-seed score to $10^{-6}$ before the test is applied. For the PROTEINS baselines the retrained score differs from the recorded row within the seed spread on six of nine models (their training is not bitwise reproducible), so their $\Delta$ is a statement about a model retrained on the same configuration. Five PROTEINS baselines are output-invariant: MLP, GCN and NSD read nodes only, SCN pools its node output, and SAN's logits move by $1.7 \times 10^{-5}$, enough to reorder its chance-level near-ties.

Table A2: SSD under a random node relabeling, every reported row: the metric on the original test data, the largest absolute change $\max|\Delta|$ of the metric over seeds, and the largest absolute change in any output over the test set (with the output scale where it is not order one).

| **Benchmark** | **metric** | **original** | $\max\|\Delta\|$ | **output residual** |
|---|---|---|---|---|
| PROTEINS (10 seeds) | AUROC | $0.7729 \pm 0.0226$ | $4\times10^{-4}$ | $2.32 \times 10^{-6}$ |
| SHREC (10 seeds) | accuracy | $0.8833 \pm 0.0284$ | 0, every seed | $5.7 \times 10^{-6}$ |
| North Pacific (5 seeds) | RMSE | $0.4441 \pm 0.0241$ | $4.8 \times 10^{-7}$ | $5.2 \times 10^{-4}$ (scale 8 to 20) |
| synthetic (5 seeds) | SWA acc. | $67.40 \pm 4.13$ | 0, every seed | $3.0 \times 10^{-4}$ (scale 62 to 98) |
| ocean-drifter (5 seeds) | SWA acc. | $48.00 \pm 8.43$ | 0, every seed | $3.9 \times 10^{-3}$ (scale 5.8) |

**Feature level.** Table A3 pulls the internal states of each model back through the relabeling on PROTEINS. SSD's three ranks are at float32 precision once the odd cochains are pulled back with the sign; the unsigned edge and face residuals, 4.5 and 1.8, are due only to the sign flips induced by the relabeling (the absolute values agree to $10^{-7}$). SCN keeps its node states while its edge and face states change by their full scale, which does not reach its output because its classifier pools node states only. The four output-variant models change every rank by an amount on the order of that rank's scale.

**Per-seed relabeling residuals.** On SHREC the ten logit residuals are between $3.6 \times 10^{-6}$ and $5.7 \times 10^{-6}$, with the identity control at exactly zero on all ten. On North Pacific the five output residuals are $4.4 \times 10^{-5}$ to $5.2 \times 10^{-4}$, or $4.2 \times 10^{-6}$ to $3.0 \times 10^{-5}$ relative to the largest predicted flow (7.7 to 20.3). On synthetic the log-probability residuals are $5.5 \times 10^{-5}$ to $3.0 \times 10^{-4}$ against output scales of 62 to 98, and on ocean-drifter 1.2 to $4.1 \times 10^{-5}$ on four seeds and $3.9 \times 10^{-3}$ on

Table A3: Feature-level residuals on PROTEINS, seed 0, largest over 223 test graphs of the per-graph maximum. *unsigned* pulls the relabeled features back onto the original simplex order without the orientation sign, *signed* with it; *scale* is the largest absolute feature value. An equivariant model holds the signed residual at machine precision.

| | **node** | | **edge** | | | **face** | |
|---|---|---|---|---|---|---|---|
| **Model** | residual | scale | unsigned | signed | scale | signed | scale |
| SSD | $2.26\times10^{-6}$ | 3.2 | 4.5 | $4.77\times10^{-7}$ | 2.2 | $4.17\times10^{-7}$ | 0.88 |
| NSD | $6.1\times10^{-5}$ | 41 | — | — | — | — | — |
| GCN | $4.8\times10^{-7}$ | 8.3 | — | — | — | — | — |
| MLP | 0 | 29 | — | — | — | — | — |
| SCN | $1.5\times10^{-5}$ | 121 | 28 | 32 | 28 | 53 | 51 |
| GSAN | 0.99 | 1.0 | 0.62 | 2.0 | 1.0 | 1.4 | 0.75 |
| COSIMO | $3.7\times10^{3}$ | $5.6\times10^{3}$ | $1.3\times10^{3}$ | $1.3\times10^{3}$ | $1.3\times10^{3}$ | $3.9\times10^{2}$ | $1.8\times10^{2}$ |
| SCCN | 1.0 | 1.0 | 1.0 | 2.0 | 1.0 | 2.0 | 1.0 |
| SCCNN | $3.0\times10^{13}$ | $2.5\times10^{13}$ | $7.1\times10^{11}$ | $7.1\times10^{11}$ | $6.2\times10^{11}$ | $1.0\times10^{9}$ | $4.6\times10^{9}$ |
| SAN | $1.9\times10^{5}$ | $1.6\times10^{5}$ | $7.8\times10^{3}$ | $7.9\times10^{3}$ | $6.6\times10^{3}$ | — | — |

the seed with the smallest output scale (5.8). With deterministic accumulation the relabeling changes only the float32 summation order, and that reaches the output at a size that varies with the seed; the metric moves only on PROTEINS (by at most $4 \times 10^{-4}$) and, at $5 \times 10^{-7}$, on North Pacific.

**Ablation on the trajectory sets, details.** The contrasts of Section 6.5 change one family from the full cell or from the map-free cell of Table 4. The full-cell contrasts not quoted there are: ocean polar against orthogonal 10.0 ($t = 5.7$, five of five); synthetic face per-coordinate against identity 3.8 ($t = 2.1$, three of five), polar against orthogonal 3.6 ($t = 1.6$, four of five), per-coordinate against scalar 1.2 ($t = 1.0$, two of five). Averaged over the other family instead (fifteen seed-pairs each), the ocean-drifter contrasts are: face per-coordinate against identity $+10.3$ ($t = 4.9$), against scalar $+3.2$ ($t = 1.3$); node polar against identity $+5.0$ ($t = 1.7$), against orthogonal $+4.5$ ($t = 2.1$). They are smaller than the full-cell contrasts because the families interact: the node map contributes only in the presence of a learned face gate ((polar, identity) 31.5 against the map-free 34.5, (identity, per-coordinate) 42.0, full 48.0). Since the layer-1 node maps are nearly isotropic there ($\kappa \leq 1.3$, Appendix E), what the orthogonal family loses is close to a per-node magnitude; no cell with a per-node scale and no rotation was run, so the grid shows that the polar map needs its scale, not that the scale alone would do. On synthetic the axis-averaged contrasts are face per-coordinate against identity $+4.0$ ($t = 2.9$), against scalar $-0.4$; node polar against identity $+1.3$ ($t = 1.3$), against orthogonal $+0.1$; and the full model is not the best cell, though the three lie within noise of each other ((orthogonal, scalar) 68.4, (identity, per-coordinate) 68.2, full 67.4). The full-model cells reproduce the rows of Table 3 seed by seed (37.5, 62.5, 50.0, 42.5, 47.5 and 68, 65, 63, 75, 66). The per-cell standard deviations of Table 4 are population standard deviations over the five seeds, as in Table 3; the paired standard errors behind the $t$ statistics use the sample standard deviation of the five differences.

Every cell has the parameter shapes of the full model: a fixed map keeps its generator's parameters and does not use them, so seed $k$ draws the same initialization in every cell. The gains do not follow the generators' sizes. On ocean-drifter the face-map generator has 282 of the model's 5,322 parameters and the node-map generator 1,254; on synthetic, 522 and 2,570 of 5,213. The face gate, the smaller of the two, carries most of the gain on both sets, and on ocean-drifter the node map alone recovers none of it.

Appendix E offers a possible explanation for the difference between the two sets; it does not test that explanation. On synthetic the layer-2 face gate opens on the faces the sample itself crosses (1.2 to 2.3 times its value elsewhere; the test-mean field follows the training traffic, $r = 0.60$), a sample-dependent selection that a node map could also express through node states that already encode the prefix. This would be consistent with the face gate alone recovering the full gain and the node map alone recovering about half of it. Whether the node maps of the ablated cells take over that gating was not measured. On ocean-drifter the gate does not follow the path (ratios 0.6 to 1.6 on the stored trajectories) and responds to the geometry of the complex, one stalk coordinate driven toward the boundary of Theorem 5 and the other opening on the 66 boundary triangles (0.23 against 0.18 inside), and there the face family is decisive.

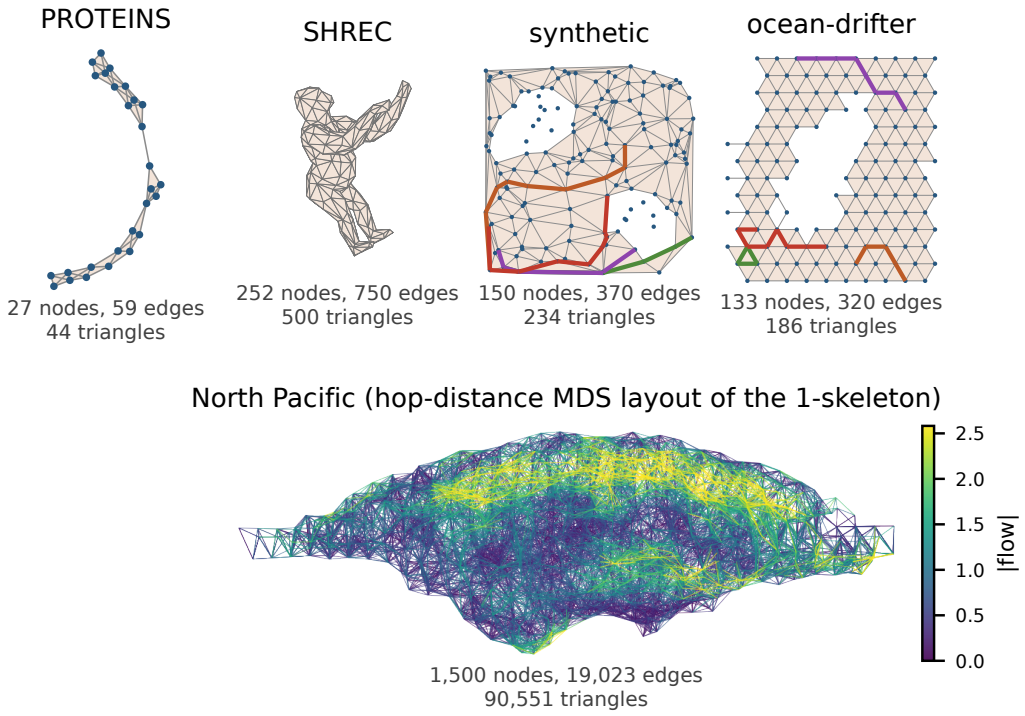



Figure A1: One complex per benchmark. PROTEINS: one graph with its 3-cliques as triangles. SHREC: one of the 600 meshes. Synthetic and ocean-drifter: the complex every trajectory lives on, at its node coordinates, with four test trajectories drawn as colored edge paths (15 of the 150 synthetic nodes lie inside the two holes and touch no edge). North Pacific: the 1-skeleton in a classical multidimensional-scaling layout of its hop distance (the artifact carries no coordinates), edges colored by the magnitude of the edge flow of Yang et al. (2024), color clipped at the 98th percentile.

**Runtime and memory.** Table A4 reports the per-step cost of SSD and four of the SHREC baselines on the same 64 training meshes (48,000 edges, 32,000 faces), using the batch of 64 from the searches and one process per model at its reported configuration on an otherwise idle A100. It reports the mean of five steps after two warm-up steps and the allocator's peak, excluding the CUDA context. SSD's forward alone takes 1.05 GiB. The rest is the autograd tape of the diffusion stack, about half of which layer checkpointing recovers for 35% more time. Per-cell maps recomputed at every layer, the block normalization and its Newton–Schulz roots account for the difference from COSIMO.

Table A4: Per-step cost on a batch of 64 SHREC meshes, one process per model.

| Model | ms / step | peak GiB | params |
|---|---|---|---|
| SCCN | 54.8 | 0.13 | 18,138 |
| SCCNN | 80.8 | 0.41 | 42,714 |
| COSIMO | 96.7 | 1.33 | 126,510 |
| SCN | 109.2 | 17.25 | 1,335 |
| SSD | 266.0 | 9.27 | 123,900 |
| SSD, layer checkpointing | 358.4 | 5.23 | 123,900 |

## E The Complexes and What the Maps Learn

Figure A1 shows one complex from each benchmark. On the two trajectory sets every sample lives on the same complex, so the learned maps are a field over it that can be drawn. We read them off the two trajectory rows, retrained for seed 0; each run reproduces its recorded seed-0 score (SWA accuracy 0.680 on synthetic, 0.375 on ocean-drifter). The generator runs from the rank-0 state before each layer, so "layer 1" is the map acting in the first diffusion layer, which depends on the node

coordinates alone and is shared by every sample; deeper layers see the sample's flow through the couplings and are reported per test trajectory or as their mean. Throughout, $R_v = Q_v \operatorname{diag}(e^{s_v})$ with $s_v \in [-1, 1]^d$ and $F_t = \operatorname{diag}(\mathbf{f}_t)$ with $\mathbf{f}_t \in (-1, 1)^d$ (Appendix B); the operator depends on $F_t$ through $F_t^2$, so the figures show $|\mathbf{f}_t|$.

**Ocean-drifter ($d = 2$, three layers).** The gate approaches the boundary of Theorem 5 without crossing it (Figure A2, bottom). The first stalk coordinate is coupled to the faces at $|f_{t,1}| < 0.02$ on every one of the 186 triangles (median $0.008$), the second at $|f_{t,2}|$ between $0.08$ and $0.19$ (median $0.12$), a median ratio of 15 on the same triangle; no gate is zero, so the harmonic dimension stays at $\beta_1 d$ (below). The deeper layers keep the split (layer-3 medians $0.013$ and $0.17$) and open the second coordinate further where the complex has a boundary, along its outer rim and around the hole (Figure A2, bottom right): the test mean is $0.23$ on the 66 triangles that touch a boundary edge, up to $0.61$, against $0.18$ on the 120 interior ones. The node maps are almost isotropic at layer 1, $\kappa(R_v) = \sigma_{\max}/\sigma_{\min}$ between $1.00$ and $1.30$ in a smooth gradient across the complex, and anisotropic on a scattered subset of nodes in the last layer ($\kappa > 2$ on 20 of the 133 nodes, up to $4.8$).

**Synthetic ($d = 6$, two layers).** No coordinate is switched off: the layer-1 gate means per coordinate are $0.28$, $0.31$, $0.13$, $0.10$, $0.22$ and $0.23$, a factor of three between the largest and smallest means. The anisotropy sits in the node maps instead, $\kappa(R_v)$ between $3.2$ and $6.7$ (mean $4.8$) at layer 1 with singular values per coordinate of $0.63$, $0.80$, $0.59$, $0.69$, $2.32$ and $0.49$, in a smooth gradient across the square. The layer-2 gate, which sees the trajectory prefix, opens along the sample's own path (Figure A2, top right): on each of the six stored test trajectories the gate on the faces that trajectory crosses is $1.2$ to $2.3$ times its value elsewhere ($0.30$ to $0.55$ against $0.24$), whereas the layer-1 gate, which sees only the coordinates, does not vary with the path. Averaged over the test set this becomes a field that follows the traffic ($r = 0.60$ with the number of training trajectories through a face). On ocean-drifter, the same test gives no consistent opening along the path (ratios $0.6$ to $1.6$ on the six stored trajectories): there the deeper gate responds to the geometry of the complex, not to the sample.

**Spectrum of the normalized rank-1 pair.** Figure A3 sorts the eigenvalues of $\widehat{L}_1 = \widehat{L}_1^{\downarrow} + \widehat{L}_1^{\uparrow}$ of Eq. (4) on each trajectory complex under the identity maps and under the learned layer-1 maps of the same models. Both complexes have $\beta_1 = 2$, and on both the number of zero eigenvalues is $\beta_1 d$ for the identity and the learned maps alike (12 on synthetic, 4 on ocean-drifter), as Lemma 4 and Theorem 5 say for invertible maps; on ocean-drifter, eight further directions lie at the normalization's floor on its five edges in no triangle, as described in Appendix B; this holds for both maps.

Above the kernel the two operators differ. In the unnormalized $L_1^{\mathcal{F}}$ of Section 5.3 the learned ocean-drifter gate does what Theorem 5 describes: the mean diagonal of $B_2 W_2 B_2^{\top}$ in the first coordinate is $2 \times 10^{-4}$ against $2.9$ for $B_1^{\top} W_0 B_1$ (second coordinate: $0.03$ against $3.6$), so that coordinate is coupled along the graph and almost not through the faces, and 183 eigenvalues lie below $10^{-3}$ against 4 for the identity maps, close to the $\beta_1^{\text{graph}} = 188$ harmonic directions the coordinate would have at the graph endpoint. The operator the model runs is the normalized one, and Eq. (4) re-expresses this selection. The normalization conjugates the two halves by a mutually inverse pair $A_e$, $A_e^{-1}$, which is what the Hodge split allows: Lemma 4 holds for any symmetric positive definite $A_e$, whereas a diagonal applied to both sides of each half, shared or not, breaks the split (Section 4.2). Within that family it takes the balanced $A_e = (P_e^{-1} \# Q_e)^{1/2}$, which puts both halves on one common block, $A_e P_e A_e = A_e^{-1} Q_e A_e^{-1}$ (equal to $P_e \# Q_e$ when the two blocks commute), so that neither half dominates. Every member of the family, the identity maps included, keeps the split through such a conjugate pair, and the balanced choice scales down both halves of a coordinate's rank-1 operator when that coordinate has a small face weight: on the edges in a triangle the diagonal of $\widehat{L}_1^{\downarrow}$ in the first coordinate is $0.032$ and that of $\widehat{L}_1^{\uparrow}$ is $0.041$, against $0.629$ for both in the second coordinate and $0.425$ for both under the identity maps, and the median eigenvalue of the first coordinate's block is $0.058$ against $1.22$ for the second and $0.876$ under the identity maps. That is where the small eigenvalues come from: 52 below $10^{-2}$ against 12 for the identity maps, 240 below $10^{-1}$ against 34, and a median of $0.18$ against $0.88$. Through the face gate alone, the model has reduced rank-1 diffusion by a factor of about fifteen in one stalk coordinate and increased it by about half in the other (identity node maps with the learned gate give 52 and 235; the learned node maps with the identity gate give 12 and 34). On synthetic, where every coordinate stays open, the change is a reweighting: the smallest nonzero eigenvalue moves from $0.0038$ to $0.0012$, the median from $0.27$ to $0.16$, and the

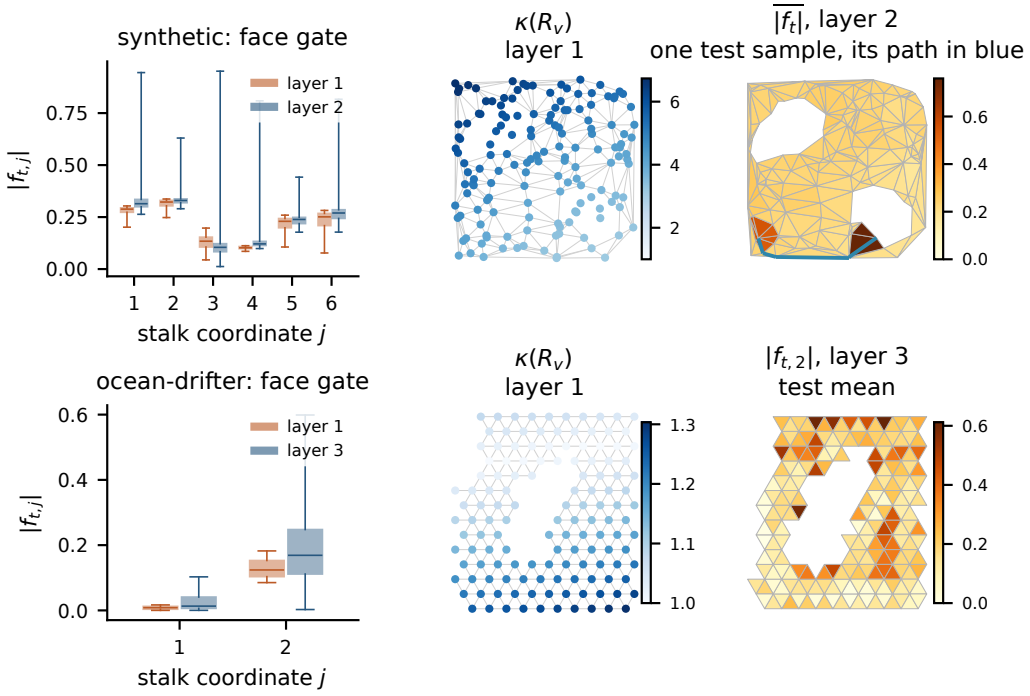



Figure A2: Learned maps on the two trajectory sets (seed 0 of the reported rows). Left: face gate $|f_{t,j}|$ per stalk coordinate over all triangles at layer 1 (one field, shared by every sample) and at the last layer (pooled over the test trajectories); boxes are quartiles, whiskers the 1st to 99th percentile. Middle: condition number of the layer-1 node map on the complex. Right: what the deeper layers do on each set. Synthetic: the layer-2 gate of one test sample, averaged over coordinates, with that sample's trajectory drawn in blue; the gate opens along the path. Ocean-drifter: the layer-3 gate of the open coordinate averaged over the test trajectories; it opens along the outer rim and around the hole, independently of the sample.

top from $0.63$ to $1.95$. The learned maps change the spectrum and leave the harmonic dimension at the value predicted by the theory.

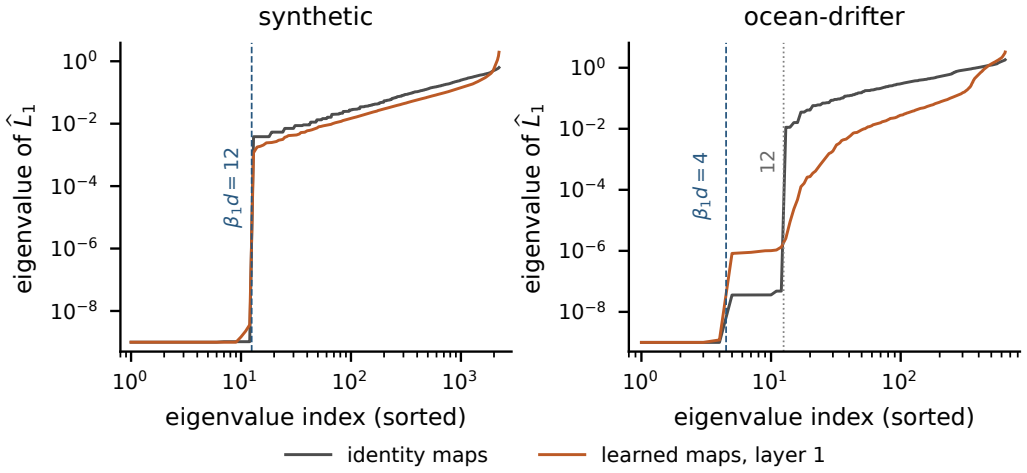



Figure A3: Sorted spectrum of the normalized rank-1 operator $\widehat{L}_1$ on the two trajectory complexes under the identity maps and under the learned layer-1 maps; both axes logarithmic, zeros drawn at $10^{-9}$. The dashed line marks $\beta_1 d$; on ocean-drifter the dotted line marks the twelve numerically null directions, the harmonic ones plus the floor the block normalization leaves on the five edges in no triangle (Appendix B).