

Learnable Hodge–Laplacian Driven Mixture of Experts for Steady-PDE Surrogates on Coarse Unstructured Meshes

Stavros Nousias¹ Akis Nousias² Konstantinos Gkrispanis¹

¹Technical University of Munich ²K3Y Labs



Talk Outline

1. **Problem:** steady-flow prediction on coarse unstructured meshes
2. **Prediction:** Hodge controller, transformer experts and field decoder
3. **Supervision:** spectral smoothing, telescoping atoms and learned assignment
4. **Evidence:** AirFRANS comparisons, ablations and routing diagnostics
5. **Scope:** experimental qualifications and structural limitations

Technical backup follows the main talk.

Steady-PDE Surrogates on Unstructured Meshes

Given a mesh and flow conditions

- Vertex coordinates and geometric features
- Mesh connectivity and local neighborhoods
- Inlet velocity and boundary indicators

Predict four fields at every vertex

$$\hat{u}(v) = [u_x(v), u_y(v), p(v), \nu_t(v)].$$

AirFRANS task

Steady Reynolds-averaged Navier–Stokes airfoil solutions.

Reported discretization

Approximately 4.9k vertices and 9.6k triangular faces per working mesh.

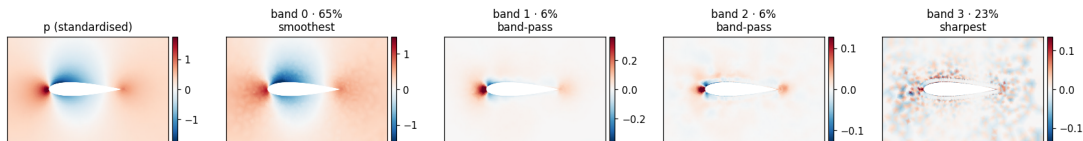
Learn directly from simulated solution fields. No PDE-residual objective is used in this model.

Paper: abstract and experimental setup. AirFRANS: Bonnet et al., NeurIPS 2022.

A Flow Field Contains Several Spatial Scales

Smooth pressure structure and fine residuals coexist on the same mesh.

Frequency bands (fixed mesh-graph smoothing, ladder (2, 8, 32)) — sum = field



- A single mixing rule must fit broad structure and localized detail.
- Separate expert targets can encourage different scale-dependent predictions.

Fixed reference illustration: adjacency smoothing ladder (2, 8, 32), four overlapping components. The reported Hodge analyzer instead uses eight rungs and nine atoms.

The Proposed Model

1. **Eight transformer experts** receive shared features and different Fourier positional inputs.
2. A **Hodge controller** computes per-vertex weights from features and mesh structure.
3. A **training-only analyzer** constructs scale-dependent auxiliary targets from reference fields.

31%

lower mean relative L_2

47%

lower interior MSE

3.02M

registered parameters

Relative to the all-input Transolver baseline on the same working meshes. Training allocation and checkpoint selection differ.

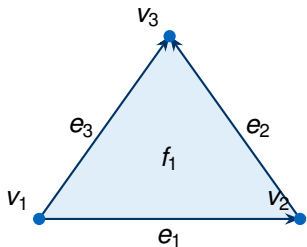
How Mesh Structure Reaches the Prediction

Model family	Role of mesh structure	Example
Point-set predictors	No explicit mesh operators	Plain transformer
Band-supervised mixtures	Mesh defines auxiliary targets	Learned-band MoE
Hodge predictors	Hodge operator updates features	PSD-star HodgeFormer
Hodge-routed mixtures	Hodge operator computes gate logits	Reported model

Geometric and multiscale gating have precedents in GNOT and GAOT. The specific construction here combines **Hodge-based routing** with **mesh-spectral auxiliary supervision** of full transformer experts.

Paper: related work and model family. This is a dense mixture, with every expert evaluated.

Incidence Matrices Encode Mesh Topology



$$d_0 = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \quad d_1 = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}.$$

A simplicial complex $\mathcal{M} = (V, E, F)$

- $d_0 \in \{-1, 0, 1\}^{N_e \times N}$ maps vertex values to oriented edge differences.
- $d_1 \in \{-1, 0, 1\}^{N_f \times N_e}$ maps edge values to signed face sums.
- Connectivity is fixed for a mesh, independent of network weights.

$$d_1 d_0 = 0$$

Hodge Operators Separate Topology and Feature Mixing

$$L_v = \underbrace{\star_0^{-1}(x_v)}_{\text{vertex star}} \underbrace{d_0^\top}_{\text{incidence}} \underbrace{\star_1(x_e)}_{\text{edge star}} \underbrace{d_0}_{\text{incidence}} .$$

Controller

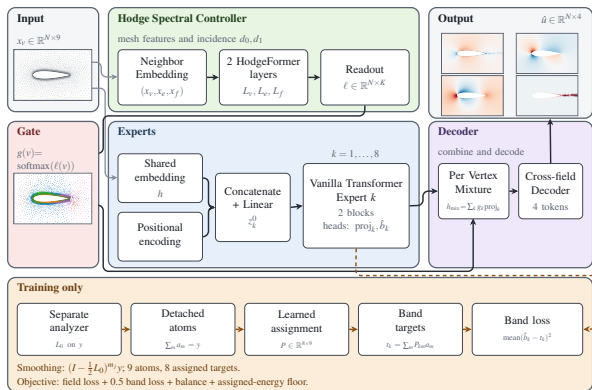
- Feature-dependent, non-diagonal stars
- Produces routing weights during prediction
- Spectral guarantees require structural assumptions

Analyzer

- Separate, positive diagonal stars
- Smooths reference fields during training
- Random-walk normalized vertex Laplacian

The controller and analyzer have different parameters, roles and guarantees.

Hodge-Routed Mixture of Transformer Experts



Prediction path

- Shared vertex embedding
- Eight experts
- Per-vertex gate
- Cross-field decoder

Dashed path

Auxiliary targets and band loss are used only in training.

The diagram includes the VEF extension. The reported best model updates the vertex controller stream only.

Shared Features and Independent Transformer Experts

$$h = W_2 \text{GELU}(W_1 x_v) \in \mathbb{R}^{N \times 192}.$$

Shared embedding

- Nine standardized vertex inputs
- The same embedding h is supplied to all experts
- Expert-specific input projections have independent parameters

Each of eight experts

- Width 88, two pre-norm transformer blocks
- Eight heads, head dimension 11
- Full softmax attention over all N vertices
- Heads for a latent representation and an auxiliary four-field prediction

Experts communicate globally within a mesh. They do not receive incidence matrices or a mesh attention mask directly.

Fourier Inputs Bias Experts Toward Different Scales

For standardized coordinates $\text{pos} \in \mathbb{R}^{N \times 2}$,

$$B_k = \sigma_k G_k \in \mathbb{R}^{2 \times 8}, \quad (G_k)_{ij} \sim \mathcal{N}(0, 1),$$

$$\gamma_k(\text{pos}) = [\sin(2\pi \text{pos} B_k); \cos(2\pi \text{pos} B_k)], \quad z_k^0 = W_{\text{in}}^k[h; \gamma_k(\text{pos})].$$

Expert k	1	2	3	4	5	6	7	8
σ_k	1.0	1.7	2.8	4.8	8.0	13.5	22.6	38.0

- B_k is sampled at initialization and remains fixed.
- Each expert gets $192 + 16 = 208$ channels, projected to width 88.
- Bandwidths bias the inputs but do not confine an expert to one spectral band.

The Gate Mixes Representations Before Field Decoding

$$g_k(v) = \frac{e^{\ell_k(v)}}{\sum_{j=1}^8 e^{\ell_j(v)}}, \quad h_{\text{mix}}(v) = \sum_{k=1}^8 g_k(v) \text{proj}_k(z_k)(v).$$

Mixture

- A convex combination in latent space
- Per-vertex weights shared across output fields
- Dropout and layer normalization before decoding

Cross-field decoder

- Four learned field embeddings at each vertex
- Two transformer blocks over the four field tokens
- Shared readout $192 \rightarrow 64 \rightarrow 1$

Auxiliary band predictions train the experts. They are not added together to produce the final fields.

Controller Embeddings and Vertex Readout

Separate controller embeddings

- Input widths $(c_v, c_e, c_f) = (9, 21, 13)$
- Normalized adjacency aggregation on vertex and edge inputs
- Stream-specific three-layer MLPs
- Controller width 64

Two HodgeFormer layers

- Four heads, width 16 per head
- Pre-normalization and residual updates
- Feed-forward width 256
- Only the vertex stream is read out

$$\ell = \text{Linear}_{64 \rightarrow 8}(\text{GELU}(\text{Linear}_{64 \rightarrow 64}(\text{LN}(h_v)))) .$$

Reported model: vertex updates only. VEF additionally evolves the edge and face streams.

Gaussian Hodge Stars Mix Supported Neighbors

Projected normalized features define a positive kernel on the retained support:

$$\varphi_i = W_\varphi \text{LN}(h_s)_i, \quad \Omega_{ij} = \exp\left(-\frac{\|\varphi_i - \varphi_j\|^2}{2\sigma^2}\right), \quad j \in \mathcal{N}(i).$$

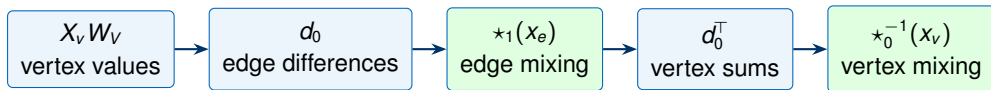
The normalized neighborhood operator is mixed with the identity:

$$D_i = \sum_j \Omega_{ij}, \quad \hat{S}_{ij} = \frac{\Omega_{ij}}{\sqrt{D_i D_j + 10^{-8}}}, \quad \star x = (1 - \beta)x + \beta \hat{S}x,$$

$$\sigma = \text{softplus}(\gamma) + 10^{-8}, \quad \beta = \frac{1}{2} \text{sigmoid}(\alpha) < \frac{1}{2}.$$

- Similar projected features receive larger interaction weights.
- The identity term preserves an element's own values.
- Effective support is capped. It is not guaranteed to be reciprocal.

The Vertex Operator Is Applied Factor by Factor



$$A_v = L_v V_v(X_v), \quad \tilde{h}_v = h_v + \text{Drop}(\text{MultiHead}(A_v)),$$

$$h_v^{\text{out}} = \tilde{h}_v + \text{Drop}\left(\text{FFN}(\text{LN}(\tilde{h}_v))\right).$$

- Incidence products use sparse matrix multiplication.
- Stars use bounded neighbor supports. No dense $N \times N$ Laplacian is required.
- The overall neural block includes projections, nonlinearities and residuals.

Controller Spectral Guarantees Are Conditional

With reciprocal effective support

- The ideal normalized kernel is symmetric with eigenvalues in $[-1, 1]$.
- The identity-mixed star has eigenvalues in $[1 - 2\beta, 1]$.
- For SPD vertex and edge stars, L_v is similar to a symmetric PSD matrix.

In the configured implementation

- Support caps and index-prefix truncation can remove reciprocal pairs.
- Stars labeled as inverses are independently parameterized.
- Accuracy does not establish symmetry or a spectral guarantee.

The analyzer's reconstruction identity holds independently of these controller assumptions.

A Separate Analyzer Smooths the Reference Fields

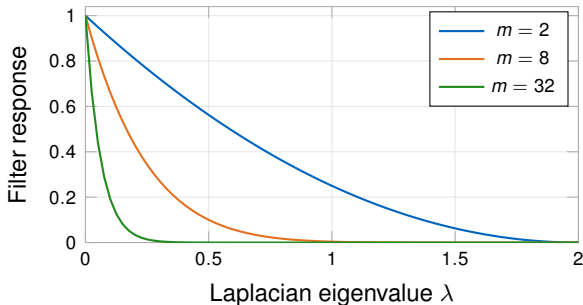
$$w_e = \text{softplus}(\text{MLP}_e(x_e)) + 10^{-4}, \quad W = \text{diag}(w), \quad \mu = |d_0|^\top w, \quad M = \text{diag}(\mu),$$

$$\boxed{L_0 = M^{-1} d_0^\top W d_0}, \quad S = I - \frac{1}{2} L_0.$$

- Conductances are positive and masses are the incident weighted degrees.
- L_0 is self-adjoint in the degree-weighted inner product.
- It is similar to a symmetric PSD normalized Laplacian, with spectrum in $[0, 2]$.
- Therefore S has spectrum in $[0, 1]$ and repeated application smooths fields.

This analyzer is separate from the controller. Its configured conductances receive no objective gradients.

Repeated Smoothing Suppresses Higher Graph Frequencies



Analytic response of the specified analyzer, not an empirical spectral measurement.

For $L_0\phi = \lambda\phi$,

$$S^m\phi = (1 - \lambda/2)^m\phi.$$

- Small λ : smoother structure persists.
- Large λ : stronger attenuation.
- Differences between smoothing levels produce overlapping scale filters.

Nine Telescoping Atoms Reconstruct the Reference

Let $q_j = S^{m_j} y$, for $m = (2, 4, 6, 8, 12, 16, 24, 32)$.

$$a_0 = q_7, \quad a_i = q_{7-i} - q_{8-i} \quad (1 \leq i \leq 7), \quad a_8 = y - q_0.$$

$$\underbrace{q_7}_{\text{coarsest}} + \underbrace{(q_6 - q_7) + \cdots + (q_0 - q_1)}_{\text{intermediate differences}} + \underbrace{(y - q_0)}_{\text{finest residual}} = y.$$

- Reconstruction is algebraic, subject to floating-point rounding.
- No eigendecomposition is used to construct these atoms.
- The filters overlap. They are not orthogonal spectral projections.

A Learned Assignment Builds Eight Expert Targets

$$P_{km} = \frac{\exp(\theta_{km}/T)}{\sum_{j=1}^8 \exp(\theta_{jm}/T)}, \quad T = 1, \quad P \in \mathbb{R}^{8 \times 9}, \quad t_k = \sum_{m=0}^8 P_{km} a_m.$$

Column normalization

$$\sum_{k=1}^8 P_{km} = 1 \quad \text{for every atom } m.$$

Target completeness

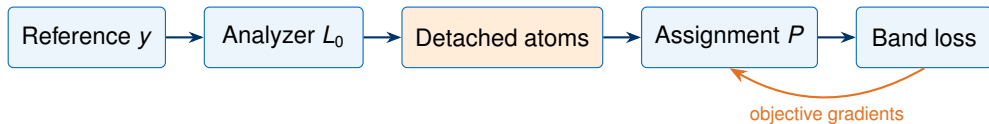
$$\sum_{k=1}^8 t_k = \sum_{m=0}^8 a_m = y.$$

Expert supervision

- Auxiliary head $\hat{b}_k$ predicts the four-channel target t_k .
- Assignment logits are trainable.
- No strict low-to-high expert ordering is imposed.

Target completeness does not impose additivity on the final predictions.

Detachment Determines Which Components Learn



- Band loss updates auxiliary heads, expert features and assignment logits.
- The floor penalty also trains the assignment through assigned-target norms.
- Neither path updates analyzer conductances because atoms are detached.
- The controller learns through the prediction loss and routing regularizer.

The configured analyzer network stays at its initialized parameters. The assignment and routing remain trainable.

The Objective Combines Field and Auxiliary Losses

$$\mathcal{L} = \mathcal{L}_{\text{vol}} + \mathcal{L}_{\text{surf}} + 0.5 \text{mean}_{v,c,k} (\hat{b}_{vck} - t_{vck})^2 \\ + 0.01K \sum_k (\bar{g}_k - 1/K)^2 + \sum_k [\text{ReLU}(0.02 - p_k)]^2.$$

$$e_{vc} = \frac{\hat{u}_{vc} - y_{vc}}{\max(\text{std}_v(y_{vc}), 0.1)}, \quad \mathcal{L}_{\text{vol}} = \text{mean}_{v \notin V_{\text{wall},c}} e_{vc}^2, \quad \mathcal{L}_{\text{surf}} = \text{mean}_{v \in V_{\text{wall},c}} e_{vc}^2.$$

- Volume and surface losses use disjoint vertex sets.
- $\bar{g}_k$ is the mean routing weight over batch and vertices.
- $p_k = \|t_k\|^2 / \max(\sum_j \|t_j\|^2, 10^{-9})$ is an assigned-target norm share.
- The one-sided floor discourages small shares. Entropy is logged, not optimized here.

AirFRANS Is Evaluated on Remeshed Working Grids

Data and discretization

- Official split: 800 training and 200 test cases
- Anisotropic triangular grids generated with gmsh
- Approximately 4.9k vertices, refined near the airfoil
- Reference fields interpolated from CFD to each working mesh

Vertex inputs

$$x_v^{\text{raw}} = [x, y, A_v^{\text{dual}}, \\ u_{\text{in}}, v_{\text{in}}, \text{SDF}, \\ n_x, n_y, \text{is_wall}].$$

- One of up to eight remeshes per training access
- One fixed evaluation mesh per case

Every comparison uses the same working-mesh targets. Scores are not directly comparable with original-grid AirFRANS results.

Target Scaling and Standardization

$$U_\infty = \sqrt{u_{\text{in}}^2 + v_{\text{in}}^2}, \quad \tilde{\mathbf{y}} = \left[\frac{u_x}{U_\infty}, \frac{u_y}{U_\infty}, \frac{p}{U_\infty^2}, \frac{\nu_t}{U_\infty} \right],$$
$$y_{vc} = \frac{\tilde{y}_{vc} - \mu_c^{\text{norm}}}{s_c^{\text{norm}} + 10^{-8}}.$$

- All nine vertex channels, including the wall indicator, use training-set standardization.
- Edge and face features are not standardized at dataset level.
- Expert Fourier features use standardized coordinate columns.
- Relative L_2 and R^2 are evaluated after returning predictions to physical units.

The turbulent-viscosity target stays linear. Dividing ν_t by speed leaves length units unless a reference length is also specified.

Training and Checkpoint Selection Differ Across Models

Setting	Hodge-routed MoE	External baselines
Training allocation	All 800 training cases	720 training, 80 validation
Checkpoint selection	Lowest test-set mean relative L_2 among 20 checkpoints	Validation-based selection
Input arm	Nine vertex channels, including wall tag	All-input arm has eight channels, without wall tag
Repetitions	One run per configuration	One run per configuration

Reported model: AdamW, 400 epochs, one whole mesh per batch, OneCycleLR with peak 10^{-3} , dropout 0.1, one RTX 4090 (24 GB).

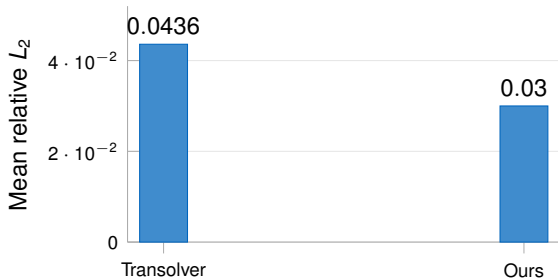
The shared evaluation meshes support score comparisons. These protocol differences limit attribution of gains to architecture alone.

Metrics Measure Different Aspects of the Flow

Metric	Meaning
Per-field relative L_2	Relative error in physical units, averaged over test cases
Mean relative L_2	Mean of the four per-field relative errors
R^2	Coefficient of determination in physical units
Volume	Globally normalized MSE on non-wall vertices
Surf	Globally normalized MSE on wall vertices
Relative lift error	Mesh-native lift-coefficient magnitude error
Lift rank correlation ρ_L	Spearman correlation across the 200 test cases

Better interior field error does not guarantee better wall pressure, integrated lift or force magnitudes.

Field Errors Decrease Relative to Transolver



	Transolver	Ours
Volume	0.00265	0.00140
R^2	0.9953	0.9983
Params	6.00M	3.02M

- **31%** lower mean relative L_2
- **47%** lower interior MSE
- About half the registered parameters

Paper: main comparison. Same 200 working-mesh cases, different training allocations and checkpoint-selection procedures.

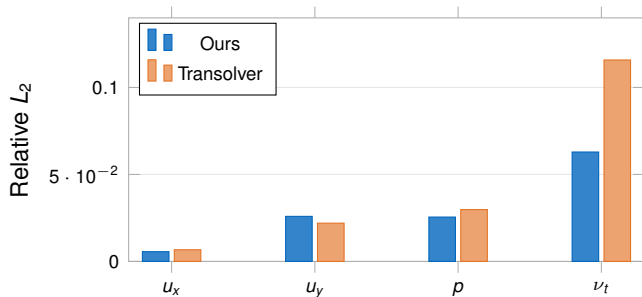
Surface and Lift Errors Favor Transolver

Model	Params	rel- $L_2 \downarrow$	Volume $\downarrow$	Surf $\downarrow$	$C_L \downarrow$	$\rho_L \uparrow$
Transolver	6.00M	0.0436	0.00265	0.00143	0.0103	0.99989
PointNet	0.08M	0.0689	0.00451	0.00440	0.0149	0.99965
Pointwise MLP	0.02M	0.1743	0.03576	0.03969	0.0860	0.99338
HodgeFormer	0.47M	0.0800	0.01028	0.01048	0.0681	0.99883
Ours	3.02M	0.0300	0.00140	0.00164	0.0225	0.99988

- Our Surf error is approximately $1.14\times$ Transolver's.
- Our relative lift error is approximately $2.2\times$ Transolver's.
- Both preserve lift case ordering well, with $\rho_L > 0.9998$.

HodgeFormer row is the best PSD-star HodgeFormer surrogate. C_L denotes the relative lift-coefficient error.

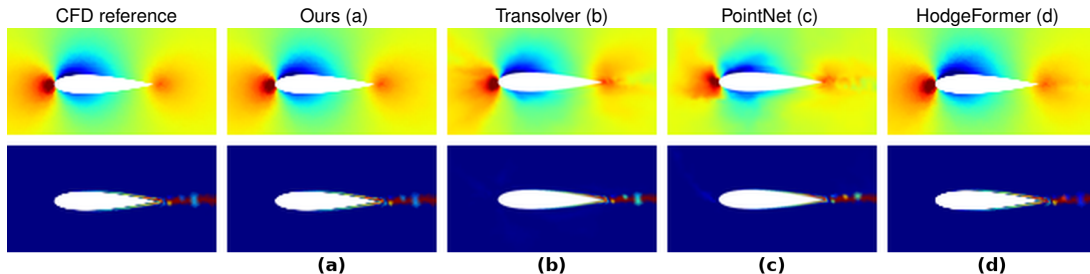
The Largest Per-Field Gain Is in Turbulent Viscosity



Per-field means over the 200-case test set. Paper: complete results, per-field table.

- Lower error on three of the four fields.
- Transolver is better on u_y .
- ν_t : 0.1158 to 0.0629, a **46% decrease**.
- ν_t remains the least accurate output of both models.

Qualitative Comparison of Pressure and Turbulent Wakes

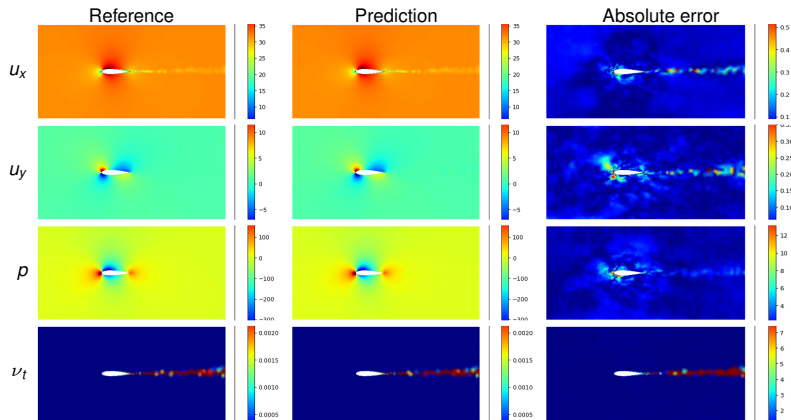


Pressure on the top row and turbulent viscosity on the bottom row, using the reference colour scales.

The models differ more in the turbulent wake than in the broad pressure field.

Illustrative case from the paper. Quantitative comparisons use all 200 test cases.

One Test Case at the Reported Checkpoint



Case relative L_2

u_x 0.0047

u_y 0.0260

p 0.0281

ν_t 0.0676

Epoch 380.

Errors are visible near the airfoil and along the wake.

Physical units on the working mesh. Reference and prediction share each field's scale. This single case does not establish representativeness.

Band Supervision and Hodge Routing Reduce Mean Error

Variant	Params	Mean $\text{rel-}L_2 \downarrow$	Volume $\downarrow$
Plain transformer	0.31M	0.0544	0.00374
Wide plain transformer	2.72M	0.0446	0.00265
Fixed-band MoE	2.75M	0.0351	0.00178
Learned-band MoE	2.75M	0.0345	0.00165
Hodge-routed MoE	3.02M	0.0300	0.00140
Hodge-routed, wide experts	5.68M	0.0341	0.00163

- Hodge routing lowers mean relative error by **13%** versus learned-band MoE.
- Increasing expert capacity does not improve the reported selected run.

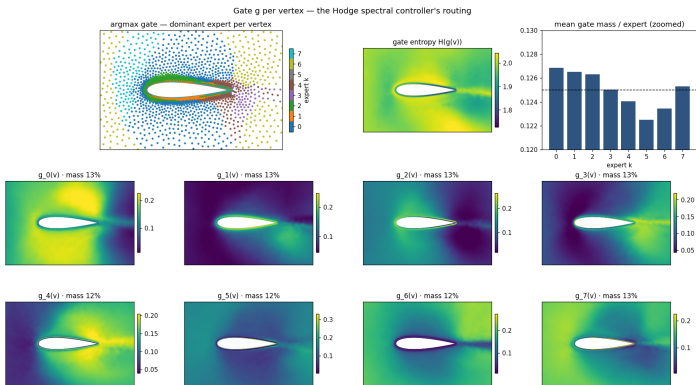
One run per configuration. Fixed-band MoE has four experts and a different analyzer, so that comparison changes several design choices.

The Selected VEF Run Has Higher Error

Setting	Vertex-only controller	VEF controller
Updated streams	Vertices	Vertices, edges, faces
Support caps	Vertex/edge: 64/64	Vertex/edge/face: 64/32/1
Epochs and checkpoints	400 epochs, 20 checkpoints	600 epochs, 30 checkpoints
Selected epoch	380	560
Parameters	3.02M	3.11M
Mean relative L_2	0.0300	0.0359
Volume	0.00140	0.00182

Support, training duration and selection also differ. This comparison does not isolate cochain degree alone.

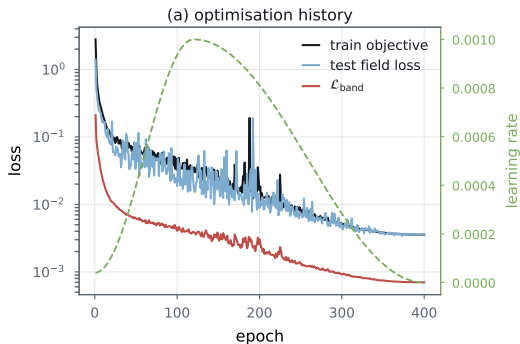
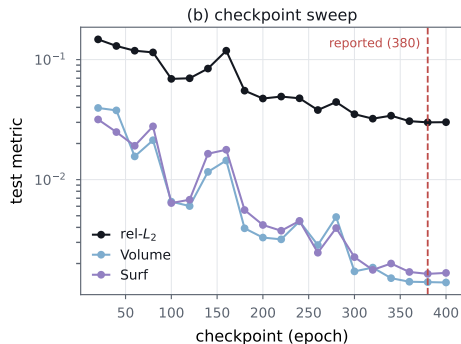
Routing Weights Vary Spatially but Remain Dense



One case at epoch 380. The mean-mass plot zooms its vertical axis to 0.120–0.130.

- Dominant experts vary across the mesh.
- Every vertex uses a soft mixture.
- Mean expert masses are close to $1/8$.
- Argmax regions do not imply hard routing.

Training Curves and the Checkpoint Sweep



- Training objective includes auxiliary terms. Test field loss excludes them.
- Best mean rel- L_2 : epoch 380. Best Volume: epoch 400.
- Checkpoint selection uses the test set, despite the original logging key `val_loss`.

Limitations and Next Experiments

What the current evidence establishes

- Lower mean field and interior errors on the reported working meshes
- A useful local comparison against the learned-band parent model
- Exact algebraic completeness of assigned auxiliary targets

What remains to be tested

- Matched data allocation, input features and validation-only selection
- Multiple random seeds and uncertainty estimates
- Original-grid accuracy and generalization across mesh resolutions
- Enforced reciprocal supports and efficient expert attention

The current model improves coarse-mesh field prediction, while surface accuracy and quadratic expert attention remain important constraints.

Conclusions

1. **Mesh-based routing** combines full transformer experts at every vertex.
2. **Multiscale auxiliary targets** come from a separate smoothing analyzer and learned assignment.
3. The reported AirFRANS run achieves **0.0300 mean relative L_2** with **3.02M parameters**.
4. Improvements concern field and interior errors. Surface and lift errors favor Transolver.



Code and
reproducibility

github.com/snousias/hodge-moe-pde

Selected References



F. Bonnet et al. *AirFRANS: High Fidelity Computational Fluid Dynamics Dataset for Approximating Reynolds-Averaged Navier–Stokes Solutions*. NeurIPS, 2022.



H. Wu et al. *Transolver: A Fast Transformer Solver for PDEs on General Geometries*. ICML, 2024.



A. Nousias and S. Nousias. *HodgeFormer: Transformers for Learnable Operators on Triangular Meshes through Data-Driven Hodge Matrices*. WACV, 2026.



Z. Hao et al. *GNOT: A General Neural Operator Transformer for Operator Learning*. ICML, 2023.



M. Tancik et al. *Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains*. NeurIPS, 2020.

Backup: Architecture Constants

Component	Configuration
Shared embedding	$9 \rightarrow 192 \rightarrow 192$, GELU
Experts	8 experts, width 88, 2 blocks, 8 heads, FFN 352
Fourier inputs	8 sampled directions, sine and cosine, $208 \rightarrow 88$ projection
Controller	Width 64, 2 Hodge layers, 4 heads, FFN 256
Controller readout	LayerNorm, $64 \rightarrow 64$, GELU, $64 \rightarrow 8$
Analyzer	Edge MLP $21 \rightarrow 64 \rightarrow 1$, eight smoothing rungs, nine atoms
Field decoder	Four field tokens, width 192, 2 blocks, 4 heads, FFN 768
Shared field readout	LayerNorm, $192 \rightarrow 64$, GELU, $64 \rightarrow 1$
Registered parameters	3,023,962 (vertex-only), 3,106,490 (VEF)

Dropout 0.1. Counts include unused parameters and analyzer parameters without objective gradients.

Backup: Neighborhood Construction and Effective Supports

1. Build same-stream graph-hop neighborhoods, with up to eight rings and 256 stored entries.
2. Allocate ring samples using decay 0.55 and fill unused capacity from nearby cells.
3. Add reverse relations, then cap the stored set again and sort by element index.
4. Retain index-prefix supports inside the controller:

Controller	Vertex cap	Edge cap	Face cap
Vertex-only	64	64	—
VEF	64	32	1

Capping after symmetrization and prefix truncation can remove reverse relations.

Backup: Operators on Vertices, Edges and Faces

$$L_v = \star_0^{-1}(x_v)d_0^\top \star_1(x_e)d_0,$$

$$L_e = d_0 \star_0^{-1}(x_v)d_0^\top \star_1(x_e) + \star_1^{-1}(x_e)d_1^\top \star_2(x_f)d_1,$$

$$L_f = d_1 \star_1^{-1}(x_e)d_1^\top \star_2(x_f).$$

- VEF updates all three streams simultaneously from the same pre-update features.
- In two layers, face-dependent edge updates can influence the final vertex readout.
- Stars labeled as inverses have independent parameters.
- SPD individual stars do not by themselves establish a shared metric for the summed edge operator.

Backup: Self-Only Face Star and Vertex Spectrum

VEF face scaling

$$z_i = \frac{\|\varphi_i\|^2}{d_h \sigma^2}, \quad (\star_2 x)_i = (1 - \beta + \beta e^{-z_i}) x_i.$$

- One face sees only itself.
- The diagonal scaling remains positive.
- It is not constrained to a geometric face area or inverse area.

Conditional vertex argument

Let $A = \star_0^{-1}$ and $Q = d_0^\top \star_1 d_0$. If A and $\star_1$ are SPD,

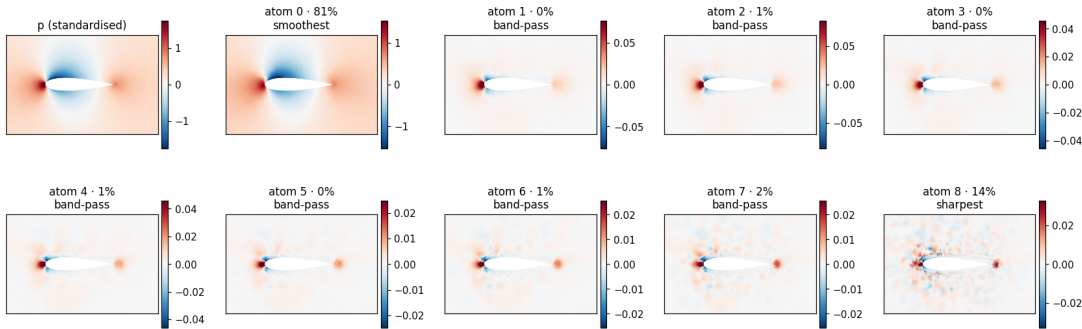
$$L_v = AQ, \quad A^{-1/2} L_v A^{1/2} = A^{1/2} Q A^{1/2}.$$

The right-hand side is symmetric PSD.

L_v can have a real nonnegative spectrum while being nonsymmetric in the Euclidean inner product.

Backup: Analyzer Atoms at Epoch 380

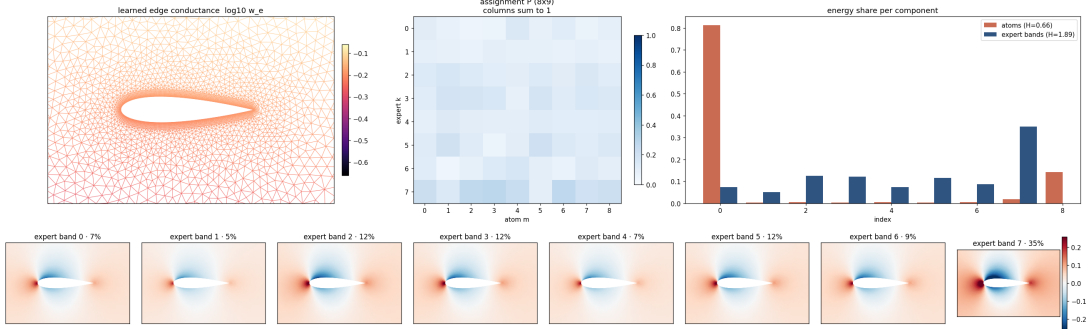
Atoms of the LEARNED Hodge analyzer ($L0=0^{-1}$ $d0^T * 1$ $d0$, $\tau=1/2$, ladder (2, 4, 6, 8, 12, 16, 24, 32)) — atom energy entropy 0.66/2.20



Coarsest and finest atoms carry about 81% and 14% of summed squared norms in this example. These are overlapping components. Conductances remain at initialization despite the legacy “learned” label.

Backup: Conductance, Assignment and Assigned Targets

How the analyzer works — p: learned conductance -> L0 -> atoms -> P -> expert bands ($\lambda_{\max} \sim 1.51$, expert bands share range ± 0.259)



Each assignment column sums to one. Smoothing uses degree-normalized conductance. No prescribed expert ordering survives global initialization.

Single-case diagnostic. Conductances stay at initialization despite the legacy labels.

Backup: Mixture and Capacity Ablations

Model	Params	rel- L_2	Volume	Surf	C_L
Hodge-routed MoE	3.02M	0.0300	0.00140	0.00164	0.0225
Wide experts	5.68M	0.0341	0.00163	0.00219	0.0286
VEF router	3.11M	0.0359	0.00182	0.00180	0.0277
Learned-band MoE	2.75M	0.0345	0.00165	0.00192	0.0265
Fixed-band MoE	2.75M	0.0351	0.00178	0.00184	0.0280
Fixed-band + wall branch	3.00M	0.0378	0.00196	0.00227	0.0247
Entropy, detached	2.75M	0.0402	0.00206	0.00214	0.0270
Entropy-balanced	2.75M	0.0412	0.00198	0.00233	0.0322

All error columns: lower is better. One selected run per configuration, on common working-mesh targets.

Backup: Predictor and Baseline Input Ablations

Model / input arm	Params	rel- L_2	Volume	Surf	C_L
Transolver: position + inlet	6.00M	0.0439	0.00270	0.00140	0.0078
Transolver: + area	6.00M	0.0592	0.00418	0.00282	0.0115
Transolver: + SDF + normals	6.00M	0.0440	0.00286	0.00137	0.0098
Transolver: all inputs	6.00M	0.0436	0.00265	0.00143	0.0103
Wide plain transformer	2.72M	0.0446	0.00265	0.00192	0.0245
Wide plain + lift loss	2.72M	0.0458	0.00248	0.00296	0.0378
Plain transformer	0.31M	0.0544	0.00374	0.00300	0.0173
PSD-star HodgeFormer	0.45M	0.1154	0.01446	0.03174	0.1747
HodgeFormer + decoder	0.47M	0.0822	0.00803	0.01205	0.0567
HodgeFormer + surface loss	0.47M	0.0800	0.01028	0.01048	0.0681

Transolver arms contain 4, 5, 7 and 8 channels respectively. None includes the wall-indicator channel used by our predictor.

Backup: Per-Field Volume and Surface Errors

Metric	Model	u_x	u_y	p	ν_t
Volume	Hodge-routed MoE	0.00011	0.00052	0.00031	0.00466
	Learned-band MoE	0.00019	0.00063	0.00072	0.00506
	Fixed-band MoE	0.00019	0.00086	0.00069	0.00539
	Wide plain transformer	0.00017	0.00109	0.00074	0.00859
Surf	Hodge-routed MoE	0.00073	0.00174	0.00248	0.00160
	Learned-band MoE	0.00087	0.00195	0.00307	0.00181
	Fixed-band MoE	0.00090	0.00198	0.00296	0.00152
	Wide plain transformer	0.00062	0.00191	0.00313	0.00201

Turbulent viscosity accounts for approximately **83%** of the Hodge-routed model's summed Volume error across the four fields.

Globally normalized MSE. Lower is better. Selected checkpoints use mean relative L_2 .

Backup: Checkpoint Minima Depend on the Metric

Model	Checkpoints	Best $\text{rel-}L_2$ (epoch)	Best Volume (epoch)
Hodge-routed MoE	20	0.0300 (380)	0.00139 (400)
Wide experts	20	0.0341 (400)	0.00163 (400)
VEF router	30	0.0359 (560)	0.00182 (560)
Learned-band MoE	20	0.0345 (380)	0.00165 (380)
Fixed-band MoE	20	0.0351 (400)	0.00178 (380)
Plain transformer	20	0.0544 (320)	0.00339 (400)
Best HodgeFormer	20	0.0800 (320)	0.00651 (400)

The main tables use the checkpoint selected by **unrounded mean relative L_2** , not independent per-column minima.

Backup: Reconstruction Identity and Its Limits

$$\begin{aligned}\sum_k t_k &= \sum_k \sum_m P_{km} a_m = \sum_m \left(\sum_k P_{km} \right) a_m \\ &= \sum_m a_m = y.\end{aligned}$$

- Requires column sums of P to equal one and a telescoping atom construction.
- Does not require a trained analyzer, orthogonal atoms or ordered experts.
- Does not imply $\sum_k \|t_k\|^2 = \|y\|^2$ because cross terms need not vanish.
- Does not imply that summing auxiliary predictions yields the final prediction.

Completeness is an algebraic target property. Predictive accuracy remains an empirical question.