AURA: GRAPH LEARNING VIA SUPERPOSED NEURAL FIELDS BEYOND MESSAGE PASSING

Anonymous authors
Paper under double-blind review

ABSTRACT

Message-passing graph neural networks update each node from a fixed-radius neighbourhood, so a depth-L model cannot use information more than L hops away. Many relational systems, molecules among them, instead interact through fields that superpose contributions from distributed sources. We introduce AURA, a graph layer built on *source-medium-response* computation: nodes emit source coefficients, a learned sparse medium operator propagates them through a truncated resolvent solve, and each node updates by probing the value, gradient, and curvature of the resulting field. Local linear message passing is a degenerate case of AURA, an exact solve aggregates walks of every length, and for every depth L we construct tasks on which all depth-L local message-passing networks are at chance while a constant-depth AURA model is exact. We test the mechanism with controls. On three interference tasks, local baselines stay at chance, full-attention graph transformers solve every task (a positive control), and AURA reaches 79–90% mean accuracy; replacing only AURA's learned medium with a uniform one, at identical depth, receptive field, and parameter count, returns all three tasks to chance. On a shared QM9 protocol and a QM9-derived field-reconstruction diagnostic, which bound the scope of our claims, AURA has the lowest mean error among the evaluated models, at 1.6$\times$ the per-run wall-clock cost of a GPS-style transformer on QM9.

1 INTRODUCTION

Message-passing neural networks (MPNNs) compute node representations by repeated aggregation over graph neighbourhoods (Gilmer et al., 2017; Kipf & Welling, 2017; Hamilton et al., 2017; Veličković et al., 2018; Xu et al., 2019; Battaglia et al., 2018). A depth-L MPNN is therefore *local*: the representation of a node is a function of its rooted L-hop neighbourhood (Lemma 3.4). Long-range dependencies must be relayed through intermediate nodes, which motivates work on bottlenecks and over-squashing (Alon & Yahav, 2021; Topping et al., 2022; Di Giovanni et al., 2023; Black et al., 2023), graph rewiring (Karhadkar et al., 2023; Gutteridge et al., 2023), and graph transformers that score all node pairs (Ying et al., 2021; Rampášek et al., 2022).

Many relational systems do not exchange messages at all. Atoms in a molecule jointly create electrostatic and polarisation fields, and a quantity measured at one atom superposes contributions from every other atom, attenuated and shaped by the medium between them. When a label depends on such a superposition at a probe node, for example on the sign of the net contribution of remote charges, a fixed-depth local model cannot observe the inputs that decide it, while a model that compares all node pairs can, at quadratic cost. This raises a concrete question: *can a sparse graph layer read remote superpositions in constant depth without pairwise scores, and if so, which ingredient does the work?*

We answer with AURA, a layer that replaces neighbourhood aggregation by *source-medium-response* computation (Figure 1). Each node emits source coefficients; a learned, edge-supported medium operator T_η propagates them, the field is the resolvent $\Phi = (I - \beta T_\eta)^{-1} S$ computed by a truncated Neumann solve, and each node reads the value, gradient, and curvature of Φ at its own position through a local differential stencil. No pairwise scores are formed and no matrix is inverted. The closest prior mechanisms differ in what is learned and what is propagated. Personalised-PageRank and diffusion propagation (Gasteiger et al., 2019a;b) and implicit GNNs (Gu et al., 2020) apply a

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.

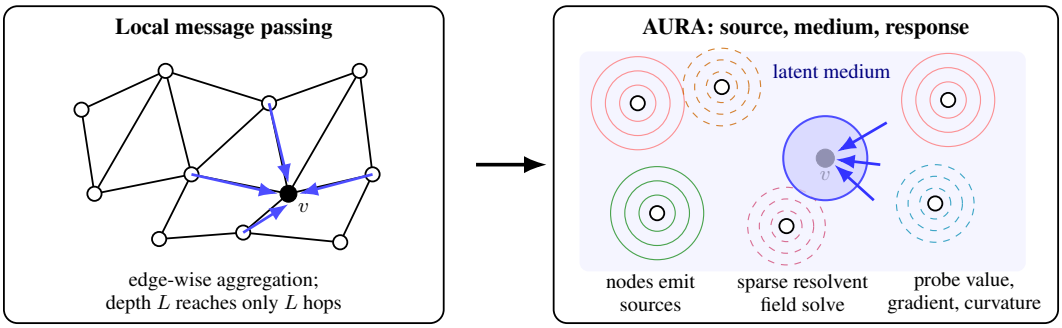

Figure 1: **From message passing to field superposition.** Left: a depth-L MPNN updates v from messages along edges, so information from farther than L hops cannot reach v. Right: AURA treats each node as a field source (solid rings: scalar and dipole channels; dashed rings: higher-order channels), solves one field over a learned sparse medium, and updates v by probing the field's value, gradient, and curvature at v. A dependency between two nodes exists only through the solved field; no pairwise scores are formed.

propagation operator of fixed form to node features; AURA learns a state- and geometry-dependent medium in every layer, re-emits sources from the probed state, and reads out differential responses. Graph transformers (Ying et al., 2021; Rampášek et al., 2022) obtain global reach through $O(n^2)$ pairwise scores; AURA forms none.

Our experiments are designed as tests of this mechanism rather than as a leaderboard. On interference tasks whose probe-local views are identical across classes, local MPNNs sit at chance, as the theory requires; full-attention transformers solve every task, which verifies that the tasks are learnable; and AURA reaches 79–90% mean accuracy. The decisive control keeps AURA's depth, 21-hop receptive field, differential probes, and exact parameter count and replaces only the learned medium by a uniform one: all three tasks return to chance (Table 1). On molecular data, AURA has the lowest mean error among the evaluated models on four QM9 targets and on a derived field-reconstruction diagnostic, and its field predictions degrade in the pattern expected of a model that resolves where each source sits (Section 4.3). Our contributions are:

- **A source-medium-response layer.** We define AURA and the instantiation we implement, show that local linear message passing is a degenerate special case (Theorem 3.1), and show that an exact solve aggregates walks of every length with a controlled truncation error (Theorem 3.2, Lemma 3.3); Sections 2–3.
- **A scoped separation.** For every depth L and any width, we construct interference tasks on which every depth-L local MPNN has error at least $1/2$ while a constant-depth AURA model is exact (Theorem 3.5). The obstruction is receptive-field depth, not Weisfeiler–Leman power, and the result is a constructive separation rather than a general expressivity characterisation; Section 3.
- **Controlled evidence with attribution.** Negative, positive, and matched controls on interference tasks (Table 1); shared-protocol QM9 comparisons reported with parameter counts and wall-clock cost (Table 2); and field-reconstruction controls that destroy source identity or geometry (Table 3); Section 4.

Our evaluation is controlled rather than broad: three constructed tasks, QM9, and a QM9-derived surrogate field. We do not evaluate on community long-range benchmarks (Dwivedi et al., 2022; Zhou et al., 2025; Mathys et al., 2026) and make no general-purpose claim (Section 6).

2 Source-medium-response layers

Setting. We consider maps $f : (G, X, R) \mapsto Y$ on a graph $G = (V, E)$ with $n = |V|$ nodes, node attributes $X = \{x_i\}$, and node coordinates $R = \{r_i\}$: 3D atom positions for molecules, or any fixed

embedding otherwise (on our synthetic tasks, i.i.d. Gaussian draws that carry no information). A standard MPNN layer computes

$$h_i^{(\ell+1)} = U^{(\ell)}\Big(h_i^{(\ell)}, \mathrm{AGG}_{j\in\mathcal{N}(i)}\, M^{(\ell)}\big(h_i^{(\ell)}, h_j^{(\ell)}, e_{ij}\big)\Big), \tag{1}$$

so depth L reaches L hops. An AURA layer instead has three stages, trained jointly end to end: *emission* (each node emits sources from its state), *medium propagation* (a learned sparse operator superposes all sources into one field), and *response* (each node probes the field locally). Appendix A.1 collects all symbols with their physical readings.

Sources. At layer ℓ, node i with state $h_i^{(\ell)} \in \mathbb{R}^d$ emits

$$\theta_i^{(\ell)} = \Psi_{\mathrm{emit}}^{(\ell)}\big(h_i^{(\ell)}\big) = \big(q_i,\, d_i,\, M_i\big), \qquad q_i \in \mathbb{R}^B,\ d_i \in \mathbb{R}^{3\times B},\ M_i \in \mathbb{R}^{5\times B}, \tag{2}$$

a scalar (charge-like), a directional (dipole-like), and a second-order (quadrupole-like) source in each of B field channels; M_i stores a traceless symmetric tensor as five real spherical-harmonic coefficients. Formally, the source at location u_i is $S_i(x) = q_i\,\delta(x-u_i) + d_i^\top \nabla\delta(x-u_i) + \big\langle M_i, \nabla^2\delta(x-u_i)\big\rangle$; in all experiments $u_i = r_i$, and the discrete layer below works with the coefficient vectors directly, so no distributional calculus enters training.[1]

Medium and field solve. The field solves $\mathcal{L}_\eta\Phi = \sum_i S_i$ for a learned medium operator $\mathcal{L}_\eta$. On a graph we use the discrete resolvent

$$\Phi = \big(I - \beta\, T_\eta(G)\big)^{-1} S, \tag{3}$$

where $T_\eta(G)$ is sparse, supported exactly on E, and acts independently on each channel: for every edge $(j \to i) \in E$ an MLP emits the channel weights $w_{ij} = \mathrm{MLP}_\eta\big(\rho(\|r_i - r_j\|), h_i, h_j\big) \in \mathbb{R}^B$, where ρ is a bank of 16 Gaussian radial basis functions, and $(T_\eta q)_i = \sum_{j\in\mathcal{N}(i)} w_{ij} \odot q_j$. The damping β is a single learned scalar (initialised at 0.6). We never form or invert a matrix: the scalar channel is solved by the truncated Neumann series

$$\widehat{\Phi}_K = \sum_{t=0}^{K} \beta^t T_\eta^t\, q, \qquad K = 6, \tag{4}$$

that is, K sparse products. The directional and second-order channels enter as one medium-weighted hop, $\Phi_i = \widehat{\Phi}_{K,i} + \frac{1}{2}\sum_j w_{ij} \odot \big(u_{ij}^\top d_j\big) + \frac{1}{4}\sum_j w_{ij} \odot \langle Y_2(u_{ij}), M_j\rangle$, with $u_{ij} = (r_j - r_i)/\,\|r_j - r_i\|$ and Y_2 the degree-2 real spherical harmonics. These choices implement the low-rank basis and multipole truncation described in Appendix A.4; the Nyström-type factorisation described there is not used.

Probing as a differential feature extractor. Node i then reads the field at its own position through a finite-difference stencil on its incident edges,

$$z_i = \Big[\,\Phi_i,\ \ \tfrac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} u_{ij} \otimes \frac{\Phi_j - \Phi_i}{\|r_j - r_i\|},\ \ \tfrac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \frac{\Phi_j - \Phi_i}{\|r_j - r_i\|}\,\Big] \in \mathbb{R}^{5B}, \tag{5}$$

namely the field value, a gradient estimate ($3B$ entries), and a Laplacian-type curvature estimate, and updates $h_i^{(\ell+1)} = h_i^{(\ell)} + \Psi_{\mathrm{resp}}^{(\ell)}\big([z_i, h_i^{(\ell)}]\big)$. The probe is a fixed differential feature extractor: it enriches each node with measurements of the solved field that a standard MPNN does not compute, and it samples no extra points. For the field-reconstruction task, where values are needed at off-graph points p, the same field is read through a learned atom-to-probe kernel (Appendix A.5).

Cost and reach. Per layer, the edge MLP costs $O(|E|\, d(d+B))$, the solve $O(K|E|B)$, the emitter and response MLPs $O(n\, d(d+B))$, and the probe $O(|E|B)$; no term is quadratic in n, whereas full attention costs $O(n^2 d)$ per layer. Each layer reaches K hops through the solve and one more through the probe stencil, so our three-layer, $K = 6$ configuration reaches up to 21 hops, against 3 for three-layer MPNNs and all nodes for full attention. Matched layer count is therefore not matched receptive field; Section 4.1 isolates this with a control.

[1] The general formulation also admits oscillatory, decay, and learned-location parameters per source. They are not used in any reported experiment.

3 WHAT THE FIELD SOLVE CAN AND CANNOT DO

We state three results about the exact operator and two lemmas that connect it to the implementation. The analysis assumes (A1) bounded source coefficients, (A2) an invertible medium with bounded inverse, and (A3) Lipschitz probe and response maps; Appendix B lists the auxiliary assumptions used only by the supporting stability and approximation results, and gives all proofs.

Theorem 3.1 (Local linear message passing is a special case)**.** *Let $\Phi = A_G S$ with A_G supported on one-hop neighbourhoods, $S_i = W_s h_i$, and a linear response $h_i^+ = W_0 h_i + W_1 \Phi_i$. Then $h_i^+ = \widetilde{W}_0 h_i + \sum_{j \in \mathcal{N}(i)} \widetilde{W}_{ij} h_j$ with $\widetilde{W}_{ij} = W_1 (A_G)_{ij} W_s$ and $\widetilde{W}_0 = W_0 + W_1 (A_G)_{ii} W_s$, a linear message-passing update.*

Theorem 3.2 (One solve aggregates every walk length)**.** *If $\Phi = (I - \beta T_G)^{-1} S$ in finite dimension with spectral radius $\rho(\beta T_G) < 1$, then $\Phi = \sum_{t=0}^{\infty} \beta^t T_G^t S$.*

Theorem 3.2 makes the exact layer an implicit infinite-walk aggregator: T_G^t enumerates weighted walks of length t and β^t damps them. It gives the probe $v^\star$ an explicit influence kernel $K_G(v^\star, i) = [(I - \beta T_G)^{-1}]_{v^\star i}$ set by the resolvent rather than by an L-fold composition of local aggregations. This changes how remote sources reach a probe, but we give no curvature, effective-resistance, or Jacobian bound and make no formal over-squashing or over-smoothing claim.

Lemma 3.3 (Truncation error)**.** *If $\|\beta T_G\| \le \gamma < 1$ in a compatible norm, then $\|\Phi - \widehat{\Phi}_K\| \le \gamma^{K+1}(1-\gamma)^{-1}\|S\|$.*

What the implementation assumes. The implemented layer is the degree-K polynomial of Eq. (4), which is well defined for any spectrum, so neither training nor inference relies on (A2); the assumption is used only by the exact statements. The damping β is a learned parameter, so only its initialisation is a setting. The constant γ is not a hyperparameter and never appears in the code: it is any number satisfying $\|\beta T_\eta\| \le \gamma < 1$. This condition constrains the *learned* operator, whose support comes from E and whose values come from the MLP, not the adjacency matrix, and we do not enforce it by projection, normalisation, or penalty. A large spectral radius affects only how closely $\widehat{\Phi}_K$ approximates the resolvent: at $K = 6$ the bound of Lemma 3.3 is $0.07\,\|S\|$ for $\gamma = 0.6$, $1.05\,\|S\|$ for $\gamma = 0.8$, and $4.78\,\|S\|$ for $\gamma = 0.9$, so K should grow as the medium becomes less contractive. We did not measure the norm of the trained operator.

We now construct tasks that separate AURA from local MPNNs. Let the label depend on the sign of a superposed field at a probe node $v^\star$,

$$y = \mathbf{1}\Big\{ \textstyle\sum_{i \in \mathcal{S}_1} a_i\, K(v^\star, i) - \sum_{j \in \mathcal{S}_2} b_j\, K(v^\star, j) > 0 \Big\}, \tag{6}$$

where K is a medium-induced influence kernel and the sources may lie beyond the receptive field of the local model.

Lemma 3.4 (Locality)**.** *For any depth-L MPNN of the form of Eq.* (1)*, of any width, the representation of node v depends only on the rooted L-hop attributed neighbourhood of v.*

Theorem 3.5 (Separation on interference tasks)**.** *For every depth L there is a balanced binary distribution over attributed graphs, labelled by Eq.* (6)*, whose probe-rooted L-hop attributed neighbourhoods are isomorphic across the two classes. Every deterministic permutation-invariant depth-L local MPNN, of any width, followed by a deterministic probe-level classifier has zero-one error at least $1/2$ on it. A constant-depth AURA model with an exact field solve and a threshold probe readout represents the task exactly when the signed source attributes are node attributes, the emitter maps them to scalar source strengths, and the learned field induces positive, symmetric source-to-probe influence with a nonzero margin.*

Proof sketch. Attach two disjoint paths of length $D > L$ from the boundary of a shared radius-L neighbourhood of $v^\star$ to source leaves with strengths $(1, 1)$ in one class and $(-1, -1)$ in the other. The L-hop views of $v^\star$ coincide, so by Lemma 3.4 every depth-L local MPNN predicts the same label for both classes. For AURA, path symmetry gives $\Phi(v^\star) = \kappa(q_1 + q_2)$ with $\kappa > 0$, whose sign is the label. Full proofs are in Appendix B.

Scope of Theorem 3.5. The theorem is a constructive, task-specific separation, not a characterisation of AURA's expressive power. The two graph families are *not* indistinguishable by 1-WL: they differ

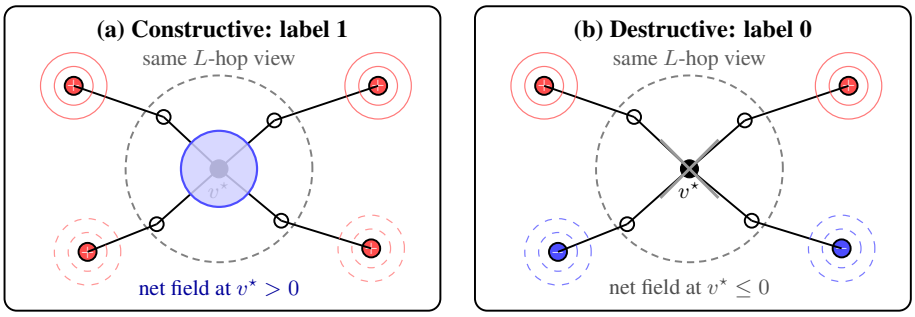

Figure 2: **The separation construction.** Both graphs present the same L-hop view around the probe $v^\star$, so every depth-L local MPNN assigns them the same representation (Lemma 3.4). Only the signs of remote sources differ; they flip the sign of the superposed field at $v^\star$ and hence the label, which a single AURA field solve exposes.

in remote node attributes, which colour refinement separates given enough rounds. The obstruction is receptive-field depth, so the result belongs with fixed-depth locality results (Alon & Yahav, 2021; Topping et al., 2022; Di Giovanni et al., 2023; Black et al., 2023), not with the Weisfeiler–Leman hierarchy (Xu et al., 2019; Morris et al., 2019). Models with global aggregation, such as full-attention transformers, have effective diameter one and are outside the quantified class; we use them as a positive control. The construction is parameterised by L: a deeper local model shifts it (a longer bridge) rather than defeating it. In short, AURA reaches in constant depth, with $O(K|E|B)$ sparse operations and no pairwise scores, a separation that fixed-depth local message passing cannot reach at any width and that full attention reaches with $O(n^2)$ pairwise scores.

4 Experiments

We ask three questions. **(Q1)** Does AURA reach remote superposed signals that fixed-depth local models cannot, and is the learned medium, rather than depth, reach, capacity, or extra features, what does the work? **(Q2)** Does the inductive bias transfer to molecular property prediction, and at what cost? **(Q3)** Does AURA predict spatial field structure, and does its prediction depend on where each source sits?

Setup. All models have three layers and are trained with AdamW (learning rate 10^{-3}, weight decay 10^{-5}, gradient clipping at norm 5) with validation-based checkpoint selection, on a single 2 GB GPU. Local baselines are GCN, GraphSAGE, GIN, and GAT. The Graphormer-style and GPS-style baselines (Ying et al., 2021; Rampášek et al., 2022) are compact reimplementations at the same depth and width: a local convolution (GIN or GAT) plus full intra-graph multi-head attention and a feed-forward block, without Graphormer's centrality and spatial encodings. The SchNet-style baseline (Schütt et al., 2017) is a compact continuous-filter convolution. AURA uses $B = 32$ channels and $K = 6$ solve steps throughout. Batch sizes are identical across models within each experiment, except on the field task (Appendix C, which lists every setting).

4.1 Q1: Is the learned medium what reaches remote sources?

Tasks. In *Remote Charge*, the probe is joined by a four-edge path to a hub from which a random tree of 35 source nodes hangs; each source carries a charge ± 1 and the label is the sign of the total charge, whose magnitude is at least 4. Every source is at least five hops from the probe, so the probe's three-hop attributed neighbourhood is identical across classes. The generator draws the topology from the same distribution for both classes and draws coordinates i.i.d. Gaussian, so only remote node attributes carry the label and no geometric prior is available. Two variants, *Resonant Cycle* (coherent versus balanced source signs) and *Shielded Source* (a remote source with or without a cancelling ring), follow the same design and are used here only in the matched control (Appendix D.1). Each task has 1,120/240/240 graphs; models read out at the probe.

Table 1: **Q1: the thesis test.** Top: Remote Charge, probe-readout accuracy over five seeds (mean $\pm$ std; median). Reach is the largest hop distance that can influence the probe; the discriminative sources sit five or more hops away. Constant predictors score 0.446 or 0.554. Cost is parameters and wall-clock per training epoch. Bottom: the matched control, seed 0; the uniform-medium model differs from AURA only in its medium weights (constant predictors: 0.446/0.554, 0.500, and 0.483/0.517).

Model	Reach (hops)	Params	s/epoch	Accuracy ↑	Median
GCN	3	17,090	0.97	0.489 ± 0.053	0.446
GraphSAGE	3	29,378	0.81	0.446 ± 0.000	0.446
GIN	3	29,570	0.90	0.446 ± 0.000	0.446
GAT	3	30,146	1.48	0.446 ± 0.000	0.446
Graphormer-style	all	129,986	3.37	1.000 ± 0.000	1.000
GPS-style	all	130,562	4.44	1.000 ± 0.000	1.000
AURA	21	193,508	4.39	0.787 ± 0.113	0.838

Matched control (seed 0)	Params	Remote Charge	Resonant Cycle	Shielded Source
AURA, learned medium	193,508	**0.846**	**1.000**	**1.000**
AURA, uniform medium	193,508	0.533	0.508	0.483

Negative and positive controls. The four local MPNNs sit at the constant-prediction level (Table 1), as Lemma 3.4 requires at any width. Both global-attention baselines reach 1.000 on every seed. Full intra-graph attention has effective diameter one, so these models lie outside the class of Theorem 3.5, and their success verifies that the task is learnable and correctly labelled: the local models fail because of their receptive field, not because of training or the generator. AURA reaches 0.787 ± 0.113 with median 0.838; four seeds reach 0.83 to 0.86 and one stays near the constant-prediction level, so its variance reflects optimisation failures of individual seeds rather than a representational limit. AURA does not match full attention on these 40-node graphs, and we claim no advantage over attention here. What it adds is constant-depth access to remote sources without pairwise scores, at a per-epoch cost comparable to the GPS-style model's (4.39 versus 4.44 s).

The matched control. AURA reaches 21 hops while the local baselines reach 3, so its success could come from depth, reach, parameter count, or the probe's extra features rather than from the learned medium. The uniform-medium control removes only the learned medium: every w_{ij} becomes the constant $1/B$, and the three layers, the six solve steps, the 21-hop reach, the differential probe, the dipole and quadrupole sources, the width, the batch size, and all 193,508 parameters stay the same. All three tasks return to the constant-prediction level (Table 1, bottom). None of depth, reach, parameter count, or probe features is sufficient; the learned medium is necessary. Two further checks agree. Structural encodings such as Laplacian eigenvectors cannot rescue a local model here: the topology is class-independent, so any function of the graph alone is independent of the label (Proposition B.12). And capacity does not drive the result: with $B = 64$ channels (290,756 parameters) AURA falls to 0.508 under the same budget (Appendix D.3). We did not train local baselines deep enough to match AURA's reach, so these results do not bound what a 21-layer MPNN could learn; Theorem 3.5 is stated per depth, and a deeper model shifts the construction rather than defeating it. The remaining components (differential probe, dipole and quadrupole sources) matter on some tasks but not on others, and only the learned medium is required by all three (Appendix D.2).

4.2 Q2: QM9 PROPERTY PREDICTION AND COST

Protocol. We fix a 30,000-molecule subset of QM9 (Ramakrishnan et al., 2014) and five molecule-level 80/10/10 splits shared by all models. Each model trains on the first 4,000 molecules of the training partition for at most eight epochs, selects its checkpoint on the 3,000-molecule validation partition, and is tested on the 3,000-molecule test partition. Targets are standardised and trained with an L1 loss. We report MAE in original units on four pre-specified targets that span electrostatic (μ), response (α), spectral ($\Delta\epsilon$), and spatial-extent ($\langle R^2 \rangle$) properties. AURA and the SchNet-style model consume 3D coordinates; GIN, GAT, and the two transformer baselines see only the molecular graph.

Table 2: **Q2: QM9 property prediction.** MAE (lower is better), mean ± std over five fixed splits of a 30k-molecule subset, 4k training molecules per split. "3D" marks models that use atom coordinates. Cost is parameters and mean wall-clock per training run. A controlled comparison, not a full-dataset leaderboard.

Model	3D	Params	s/run	μ (D)	α (a_0^3)	$\Delta\epsilon$ (eV)	$\langle R^2 \rangle$ (a_0^2)
GIN		29,953	26.0	0.859±0.019	3.422±0.064	0.645±0.034	145.4±3.1
GAT		30,529	43.3	0.815±0.025	3.345±0.129	0.672±0.032	159.4±3.9
Graphormer-style		130,369	63.1	0.750±0.020	3.184±0.097	0.503±0.032	154.3±2.7
GPS-style		130,945	87.6	0.747±0.011	3.117±0.088	0.506±0.018	135.2±3.8
SchNet-style	✓	45,697	30.0	0.782±0.023	2.907±0.059	0.443±0.026	119.0±2.0
AURA	✓	193,891	137.7	**0.638**±0.041	**2.885**±0.054	**0.397**±0.017	**115.6**±4.3

Table 3: **Q3: field reconstruction.** Top: MAE by probe distance band and Pearson r, mean ± std over five splits; all models see atom positions. Bottom: controls under one matched protocol (seed 0, 2k training molecules); ratios are relative to AURA. "Top-10%" is MAE on the probes with the largest $|V|$.

Model	Params	Overall ↓	Pearson r ↑	Near ↓	Mid ↓	Far ↓
Kernel-Sum	8,257	0.711±0.048	0.981	0.933±0.092	0.636±0.052	0.638±0.076
Probe-Token Transf.	46,145	1.124±0.037	0.967	2.083±0.044	1.040±0.057	0.569±0.025
AURA	128,771	**0.597**±0.047	**0.988**	**0.705**±0.060	**0.577**±0.041	**0.546**±0.050

Control (seed 0)	Overall	Near	Top-10%
Analytic oracle (the target formula)	1.0×10^{-6}	1.4×10^{-6}	1.7×10^{-6}
AURA	0.543	0.615	0.534
AURA, atom identities permuted within each molecule	0.600 (1.10×)	0.690 (1.12×)	0.628 (1.18×)
Summary-only model (no atom positions)	1.626 (3.0×)	2.800 (4.6×)	4.335 (8.1×)

Results and attribution. AURA has the lowest mean MAE on all four targets (Table 2). Attributing this to the architecture requires the geometry-matched comparison. Against the SchNet-style model, AURA's MAE is 18.4% lower on μ and 10.4% lower on $\Delta\epsilon$, lower on all five splits for both targets, and within one standard deviation on α (0.8%) and $\langle R^2 \rangle$ (2.9%). Against the topology-only transformer baselines the gaps are larger (7.5% to 25.1%), but they conflate the architecture with access to geometry. AURA is also the most expensive model: 137.7 s per run against 87.6 s for GPS-style and 30.0 s for SchNet-style, with $1.5\times$ and $4.2\times$ their parameters. We have no capacity-matched geometric baseline, so these margins show that the inductive bias transfers; they are not a controlled attribution. Each layer's cost is linear in $|E|$ and n, but at molecular sizes the $K \cdot B = 192$ sparse channel-steps dominate, and we have not measured where AURA becomes cheaper than dense attention on larger graphs. Absolute errors are far above full-dataset results in the literature because of the 4k-molecule budget, eight epochs, and compact models.

4.3 Q3: Field reconstruction and what the prediction depends on

Benchmark. To test whether AURA predicts spatial structure rather than graph-level scalars only, we derive a diagnostic from a 10,000-molecule QM9 subset. Around each molecule's centroid we place 64 probes in three shells (16 near, 1.5–2.5 Å; 24 mid, 2.5–4 Å; 24 far, 4–6 Å) and regress the Coulomb surrogate $V(p) = \sum_a Z_a \big(\|p - r_a\|^2 + \varepsilon^2 \big)^{-1/2}$, $\varepsilon = 0.25$ Å. The target is analytic, not a DFT electrostatic potential, and is never given to the learned models. Splits are made at the molecule level, so no molecule contributes probes to two partitions (five splits; 1,000 test molecules and 64,000 test probes per split). Both baselines see atom positions: a learned atom-to-probe kernel sum, and a probe-token Transformer whose probe tokens cross-attend to atom tokens.

Results. AURA has the lowest MAE overall and in every distance band, and the highest correlation (Table 3). The advantage is present in all three bands rather than concentrated in one; it does not grow with distance, and in the far band the Probe-Token Transformer is within one standard deviation. Two

controls test what the prediction depends on. Permuting atom identities within each molecule, while keeping coordinates and edges and computing the target from the true identities, costs AURA 10% overall, 12% in the near band, and 18% on the highest-intensity probes. The far field is dominated by the total nuclear charge, which the permutation preserves, so the loss concentrates where the field depends on which atom sits where. A model that sees only seven molecule-level statistics and the probe position, and no atom positions, is $3.0\times$ worse overall and $8.1\times$ worse on the highest-intensity probes; the analytic oracle confirms that the target is exactly the stated formula. The prediction therefore tracks resolved source structure rather than a molecule-level shortcut. As on QM9, AURA has more parameters than both baselines, so this benchmark shows what the prediction relies on, not that the gain is capacity-free. A vector-field variant and coordinate- and edge-perturbation diagnostics are in Appendix D.

5 Related work

Local message passing and its limits. MPNNs (Gilmer et al., 2017; Kipf & Welling, 2017; Hamilton et al., 2017; Veličković et al., 2018; Xu et al., 2019; Brody et al., 2022) relay long-range information through intermediate states, which causes bottlenecks and over-squashing (Alon & Yahav, 2021; Topping et al., 2022; Di Giovanni et al., 2023; Black et al., 2023) and, at depth, over-smoothing (Oono & Suzuki, 2020). Rewiring and delay (Karhadkar et al., 2023; Gutteridge et al., 2023) change which edges carry messages but keep message passing as the primitive. AURA replaces the primitive: remote influence passes through one solved field.

Propagation operators and implicit models. The mechanisms closest to AURA's solve propagate features through a resolvent-like operator. APPNP (Gasteiger et al., 2019a) truncates a personalised-PageRank series over the normalised adjacency, graph diffusion convolution (Gasteiger et al., 2019b) precomputes a diffusion kernel, implicit GNNs (Gu et al., 2020) solve for an equilibrium of node states, and DIFFormer (Wu et al., 2023) derives attention from an energy-constrained diffusion. In each, the propagation operator has a fixed form and acts on node features. AURA learns a state- and geometry-dependent, multi-channel medium in every layer, re-emits sources from the probed state, and reads out differential responses; the uniform-medium control of Section 4.1, a fixed-kernel propagation with the rest of AURA intact, fails on all three tasks.

Global attention and tokenised models. Graph transformers (Dwivedi & Bresson, 2021; Kreuzer et al., 2021; Ying et al., 2021; Mialon et al., 2021; Rampášek et al., 2022; Ma et al., 2023; Müller et al., 2024) and their sparse or scalable variants (Shirzad et al., 2023; Wu et al., 2022) define explicit pairwise scores between nodes. Graph pattern machines (Wang et al., 2025a;b) replace neighbourhood aggregation by encoding substructure patterns as tokens for a Transformer, and follow-up preprints extend this tokenisation to temporal graphs and hypergraphs (Ma et al., 2026; Zhao et al., 2026). AURA forms no pairwise scores and no tokens: a dependency between two nodes exists only as a consequence of the solved field.

Physics-inspired and long-range models. Neural diffusion, coupled-oscillator, anti-symmetric, and wave-propagation models (Chamberlain et al., 2021; Rusch et al., 2022; Gravina et al., 2023; Trenta et al., 2025) evolve node states under a continuous-time dynamic. AURA instead solves a static field from emitted sources and probes it. Long-range capability is measured by LRGB (Dwivedi et al., 2022; Tönshoff et al., 2024), GLoRa (Zhou et al., 2025), and LRIM (Mathys et al., 2026); we have not evaluated on them.

Geometric, higher-order, and operator models. Geometric molecular networks inject symmetry and physical priors into message passing (Schütt et al., 2017; Gasteiger et al., 2020; Satorras et al., 2021; Gasteiger et al., 2021; Brandstetter et al., 2022a; Batatia et al., 2022; Batzner et al., 2022; Musaelian et al., 2023; Wang et al., 2022a; Liao & Smidt, 2023; Liao et al., 2024), with expressivity analysed by Joshi et al. (2023); AURA is complementary and could host such features in its emitter. Higher-order and topological models enrich the domain on which messages pass (Morris et al., 2019; Bodnar et al., 2021; 2022; Hansen & Gebhart, 2020; Wang et al., 2022b; Lim et al., 2023). Neural operators and learned simulators (Li et al., 2020; 2021; Lu et al., 2021; Kovachki et al., 2023; Raissi et al., 2019; Brandstetter et al., 2022b; Sanchez-Gonzalez et al., 2020; Pfaff et al., 2021) apply a

learned operator to a given input field; AURA constructs its source field from the node states in every layer.

6 Conclusion and limitations

AURA replaces neighbourhood aggregation with source emission, a learned sparse resolvent solve, and local differential probing. Linear message passing is a degenerate case, the exact solve aggregates every walk length, and a constant-depth AURA model separates from every fixed-depth local MPNN on constructed interference tasks. Controlled experiments locate the working ingredient in the learned medium, and on QM9 and a derived field diagnostic AURA has the lowest mean error among the evaluated models.

Limitations. The evaluated scope is three constructed interference tasks, a 30k-molecule QM9 subset trained on 4k molecules per split, and an analytic surrogate field; we make no general-purpose claim. We have not evaluated on LRGB, GLoRa, LRIM, or large transductive graphs such as ogbn-arxiv, which would need a sampling variant of the field solve that we have not built. We did not train deeper local baselines, and we give no formal over-squashing or over-smoothing analysis. AURA is slower than the baselines at the graph sizes studied, and on QM9 it has more parameters than the geometry-matched baseline. The transformer and SchNet baselines are compact reimplementations, component ablations are single-seed, and the learned source distributions have not been analysed for interpretability (Appendix A.6 lists what can be inspected). The oscillatory and decay source parameters of the general formulation are not evaluated. The approximation and stability results are conditional on the assumptions of Appendix B.

AI USE STATEMENT

We used AI coding agents to help implement and edit parts of the model, baseline, and evaluation code released in the supplementary material. We used AI assistants as discussion partners during the research and as reviewers of drafts. We used AI tools to aid and polish the writing, including drafting revised passages when restructuring the paper from author-written material; to identify related literature; to generate plots; and to prepare slides for internal meetings. Every number in the paper was checked against the released per-run result files, AI-assisted code was reviewed and tested by the authors, every reference was checked against its publisher, proceedings, or arXiv record, and all AI-assisted text was read and edited by the authors. We take full responsibility for the content of this paper, including text, claims, code, and figures produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Section 2 and Algorithm 1 (Appendix A) specify the implemented layer, including the medium parameterisation, the truncated solve, and the probe. Appendix C lists every hyperparameter for AURA and each baseline, the data splits, the training budgets, and the hardware. The supplementary material contains the code for all models and experiments, the synthetic-task generators and pre-generated tensors, the QM9 subset indices and split files for property prediction, the scripts that build the field benchmark, and the per-run result files behind every table. All assumptions are stated in Section 3 and Appendix B, which also contains complete proofs.

REFERENCES

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 2017.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, volume 30, pp. 1024–1034. Curran Associates, Inc., 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261, 2018.

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations*, 2021.

Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations*, 2022.

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Liò, and Michael M. Bronstein. On over-squashing in message passing neural networks: The impact of width, depth, and topology. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 7865–7885. PMLR, 2023.

Mitchell Black, Zhengchao Wan, Amir Nayyeri, and Yusu Wang. Understanding oversquashing in GNNs through the lens of effective resistance. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 2528–2547. PMLR, 2023.

Kedar Karhadkar, Pradeep Kr. Banerjee, and Guido Montúfar. FoSR: First-order spectral rewiring for addressing oversquashing in GNNs. In *International Conference on Learning Representations*, 2023.

Benjamin Gutteridge, Xiaowen Dong, Michael M. Bronstein, and Francesco Di Giovanni. DRew: Dynamically rewired message passing with delay. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12252–12267. PMLR, 2023.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? In *Advances in Neural Information Processing Systems*, volume 34, pp. 28877–28888. Curran Associates, Inc., 2021.

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 14501–14515. Curran Associates, Inc., 2022.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized PageRank. In *International Conference on Learning Representations*, 2019a.

Johannes Gasteiger, Stefan Weißenberger, and Stephan Günnemann. Diffusion improves graph learning. In *Advances in Neural Information Processing Systems*, volume 32, pp. 13354–13366. Curran Associates, Inc., 2019b.

Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit graph neural networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 11984–11995. Curran Associates, Inc., 2020.

Vijay Prakash Dwivedi, Ladislav Rampášek, Mikhail Galkin, Ali Parviz, Guy Wolf, Anh Tuan Luu, and Dominique Beaini. Long range graph benchmark. In *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*, volume 35, pp. 22326–22340. Curran Associates, Inc., 2022.

Dongzhuoran Zhou, Evgeny Kharlamov, and Egor V. Kostylev. GLoRa: A benchmark to evaluate the ability to learn long-range dependencies in graphs. In *International Conference on Learning Representations*, volume 2025, pp. 92607–92627, 2025.

Joël Mathys, Henrik Christiansen, Federico Errica, Takashi Maruyama, and Francesco Alesiani. LRIM: a physics-based benchmark for provably evaluating long-range capabilities in graph learning. In *International Conference on Learning Representations*, volume 2026, pp. 43369–43406, 2026.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and Leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 4602–4609, 2019. doi: 10.1609/aaai.v33i01.33014602.

Kristof T. Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, volume 30, pp. 991–1001. Curran Associates, Inc., 2017.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, 2014. doi: 10.1038/sdata.2014.22.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *International Conference on Learning Representations*, 2022.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *International Conference on Learning Representations*, 2020.

Qitian Wu, Chenxiao Yang, Wentao Zhao, Yixuan He, David Wipf, and Junchi Yan. DIFFormer: Scalable (graph) transformers induced by energy constrained diffusion. In *International Conference on Learning Representations*, 2023.

Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. In *AAAI Workshop on Deep Learning on Graphs: Methods and Applications*, 2021.

Devin Kreuzer, Dominique Beaini, William L. Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. In *Advances in Neural Information Processing Systems*, volume 34, pp. 21618–21629. Curran Associates, Inc., 2021.

Grégoire Mialon, Dexiong Chen, Margot Selosse, and Julien Mairal. GraphiT: Encoding graph structure in transformers. arXiv preprint arXiv:2106.05667, 2021.

Liheng Ma, Chen Lin, Derek Lim, Adriana Romero-Soriano, Puneet K. Dokania, Mark Coates, Philip H. S. Torr, and Ser-Nam Lim. Graph inductive biases in transformers without message passing. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23321–23337. PMLR, 2023.

Luis Müller, Mikhail Galkin, Christopher Morris, and Ladislav Rampášek. Attending to graph transformers. *Transactions on Machine Learning Research*, 2024.

Hamed Shirzad, Ameya Velingker, Balaji Venkatachalam, Danica J. Sutherland, and Ali Kemal Sinop. Exphormer: Sparse transformers for graphs. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 31613–31632. PMLR, 2023.

Qitian Wu, Wentao Zhao, Zenan Li, David Wipf, and Junchi Yan. NodeFormer: A scalable graph structure learning transformer for node classification. In *Advances in Neural Information Processing Systems*, volume 35, pp. 27387–27401. Curran Associates, Inc., 2022.

Zehong Wang, Zheyuan Zhang, Tianyi Ma, Nitesh V. Chawla, Chuxu Zhang, and Yanfang Ye. Beyond message passing: Neural graph pattern machine. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 65496–65517. PMLR, 2025a.

Zehong Wang, Zheyuan Zhang, Tianyi Ma, Chuxu Zhang, and Yanfang Ye. Generative graph pattern machine. In *Advances in Neural Information Processing Systems*, volume 38, pp. 30068–30091. Curran Associates, Inc., 2025b.

Yijun Ma, Zehong Wang, Weixiang Sun, and Yanfang Ye. Temporal graph pattern machine. arXiv preprint arXiv:2601.22454, 2026.

Kyrie Zhao, Zehong Wang, Tianyi Ma, Fang Wu, Xiangru Tang, Pietro Liò, Sheng Wang, and Yanfang Ye. Hypergraph pattern machine: Compositional tokenization for higher-order interactions. arXiv preprint arXiv:2605.16527, 2026.

Ben Chamberlain, James Rowbottom, Maria I. Gorinova, Michael Bronstein, Stefan Webb, and Emanuele Rossi. GRAND: Graph neural diffusion. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1407–1418. PMLR, 2021.

T. Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 18888–18909. PMLR, 2022.

Alessio Gravina, Davide Bacciu, and Claudio Gallicchio. Anti-symmetric DGN: a stable architecture for deep graph networks. In *International Conference on Learning Representations*, 2023.

Alessandro Trenta, Alessio Gravina, and Davide Bacciu. SONAR: Long-range graph propagation through information waves. In *Advances in Neural Information Processing Systems*, volume 38, pp. 145054–145087. Curran Associates, Inc., 2025.

Jan Tönshoff, Martin Ritzert, Eran Rosenbluth, and Martin Grohe. Where did the gap go? Reassessing the long-range graph benchmark. *Transactions on Machine Learning Research*, 2024.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 2021.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. GemNet: Universal directional graph neural networks for molecules. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6790–6802. Curran Associates, Inc., 2021.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J. Bekkers, and Max Welling. Geometric and physical quantities improve E(3) equivariant message passing. In *International Conference on Learning Representations*, 2022a.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436. Curran Associates, Inc., 2022.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022. doi: 10.1038/s41467-022-29939-5.

Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J. Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023. doi: 10.1038/s41467-023-36329-y.

Limei Wang, Yi Liu, Yuchao Lin, Haoran Liu, and Shuiwang Ji. ComENet: Towards complete and efficient message passing for 3D molecular graphs. In *Advances in Neural Information Processing Systems*, volume 35, pp. 650–664. Curran Associates, Inc., 2022a.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3D atomistic graphs. In *International Conference on Learning Representations*, 2023.

Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations. In *International Conference on Learning Representations*, volume 2024, pp. 39282–39309, 2024.

Chaitanya K. Joshi, Cristian Bodnar, Simon V. Mathis, Taco Cohen, and Pietro Liò. On the expressive power of geometric graph neural networks. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 15330–15355. PMLR, 2023.

Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido Montúfar, Pietro Liò, and Michael M. Bronstein. Weisfeiler and Lehman go topological: Message passing simplicial networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1026–1037. PMLR, 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*, volume 35, pp. 18527–18541. Curran Associates, Inc., 2022.

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

Haorui Wang, Haoteng Yin, Muhan Zhang, and Pan Li. Equivariant and stable positional encoding for more powerful graph neural networks. In *International Conference on Learning Representations*, 2022b.

Derek Lim, Joshua Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *International Conference on Learning Representations*, 2023.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6755–6766. Curran Associates, Inc., 2020.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations*, 2022b.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. Learning to simulate complex physics with graph networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8459–8468. PMLR, 2020.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations*, 2021.

A Additional method details

A.1 Notation

Table 4: Symbols, their physical readings, and their role in the reported experiments.

Symbol	Meaning (physical reading)	In the experiments
$h_i^{(\ell)} \in \mathbb{R}^d$	state of node i at layer ℓ	$d = 64$ (synthetic, QM9), 48 (field)
r_i	node coordinate (atom position)	3D positions; i.i.d. Gaussian on synthetic tasks
$q_i \in \mathbb{R}^B$	scalar source (charge)	used
$d_i \in \mathbb{R}^{3\times B}$	directional source (dipole)	used
$M_i \in \mathbb{R}^{5\times B}$	second-order source (quadrupole)	used
$\omega_i, \phi_i, \sigma_i$	oscillation frequency, phase, decay scale	general form only; not used
u_i	source location	fixed to r_i
$T_\eta(G)$	learned medium (anisotropic conductance, screening)	edge-supported, B channels
$w_{ij} \in \mathbb{R}^B$	medium weights on edge $(j \to i)$	MLP of $\rho(\|r_i - r_j\|), h_i, h_j$
β	damping per propagation step	learned scalar, initialised at 0.6
Φ	field ($n \times B$ coefficients; potential)	solved per layer
z_i	probe features (potential, force, divergence)	value, gradient, curvature
B	number of field channels	32
K	number of Neumann steps	6
γ	constant with $\|\beta T_\eta\| \le \gamma < 1$	analysis only; not a hyperparameter

A.2 Algorithm

Algorithm 1 One AURA layer as implemented

Input: graph $G = (V, E)$, coordinates $\{r_i\}$, states $\{h_i\}$
for each edge $(j \to i) \in E$ **do**
  $w_{ij} \leftarrow \mathrm{MLP}_\eta\big(\rho(\|r_i - r_j\|), h_i, h_j\big) \in \mathbb{R}^B$; $u_{ij} \leftarrow (r_j - r_i)/\|r_j - r_i\|$
end for
$(q_i, d_i, M_i) \leftarrow \Psi_{\mathrm{emit}}(h_i)$ for all i
$\Phi \leftarrow q$; $c \leftarrow q$
for $t = 1, \dots, K$ **do** ▷ truncated Neumann solve, K sparse products
  $c_i \leftarrow \beta \sum_{j \in \mathcal{N}(i)} w_{ij} \odot c_j$; $\Phi \leftarrow \Phi + c$
end for
$\Phi_i \leftarrow \Phi_i + \frac{1}{2}\sum_j w_{ij} \odot (u_{ij}^\top d_j) + \frac{1}{4}\sum_j w_{ij} \odot \langle Y_2(u_{ij}), M_j \rangle$
$z_i \leftarrow$ value, gradient, and curvature probe of Eq. (5)
$h_i \leftarrow h_i + \Psi_{\mathrm{resp}}([z_i, h_i])$
return $\{h_i\}$

A.3 Continuous medium operator

A prototypical continuous medium is the elliptic equation $-\nabla \cdot \big(\kappa_\eta(x)\nabla\Phi(x)\big) + \mu_\eta(x)\Phi(x) = \sum_i S_i(x)$, where κ_η controls anisotropic propagation and μ_η controls attenuation or screening. All experiments use the discrete form of Eq. (3).

A.4 Efficient parameterisations

A dense global solve is expensive. The general framework admits (i) a low-rank basis expansion of the field over B channels, (ii) a Nyström-type low-rank factorisation of the solver, and (iii) a multipole truncation at order P. The implemented model uses (i) with $B = 32$ and (iii) at quadrupole order, combined with the truncated Neumann solve. The Nyström factorisation is not implemented or evaluated. Appendix D.3 reports the basis-size sweep, and the multipole term is ablated in Appendix D.2.

A.5 Off-graph probing for field reconstruction

For a point p that is not a node, AURA evaluates the final-layer field through a learned atom-to-probe kernel: $V(p) = \mathrm{head}\big(\sum_a w(p,a) \odot \Phi_a\big)$ with $w(p,a) = \mathrm{MLP}\big([\rho(\|p - r_a\|),\ (p - r_a)/\|p - r_a\|]\big) \in \mathbb{R}^B$.

A.6 Quantities available for inspection

We have not analysed the learned source distributions and make no interpretability claim. The released model exposes, per layer and per node, the scalar, dipole, and quadrupole channels q_i, d_i, M_i, and, per edge, the medium weights w_{ij}. Natural analyses are whether the emitted scalar strength on QM9 correlates with partial charge or electronegativity, whether the dipole channel aligns with local bond dipoles, and whether the medium weights concentrate on the shielding ring in Shielded Source.

B Supporting theory and full proofs

B.1 Assumptions

Assumption B.1 (Bounded source coefficients)**.** *There is $C_s > 0$ with $\left\|\theta_i^{(\ell)}\right\| \le C_s$ for all layers ℓ and nodes i.*

Assumption B.2 (Well-posed medium)**.** *For each layer, $\mathcal{L}_{\eta^{(\ell)}}$ is invertible with bounded inverse.*

Assumption B.3 (Lipschitz probes)**.** *The probe map is Lipschitz in all its arguments.*

Assumption B.4 (Stable response)**.** *Each response map $\Psi_{\mathrm{resp}}^{(\ell)}$ is Lipschitz.*

Assumption B.5 (Lipschitz emission)**.** *Each emission map $\Psi_{\mathrm{emit}}^{(\ell)}$ is L_{emit}-Lipschitz.*

Assumption B.6 (Lipschitz medium generator)**.** *There is $L_T > 0$ with $\|T_\eta(G) - T_{\eta'}(G')\| \le L_T\big(d_G(G,G') + \|\eta - \eta'\|\big)$, where d_G is a fixed metric on attributed graphs with shared node indexing, and the medium parameters are a Lipschitz function of the node states.*

Assumption B.7 (Uniform resolvent bound)**.** *For all admissible inputs, $\left\|(I - \beta T_\eta(G))^{-1}\right\| \le M$ for some finite M.*

Assumption B.8 (Target factorisation)**.** *The target operator is $\mathcal{F}(G,X,R) = \Gamma(\Phi^\star)$ with $\Phi^\star = \mathcal{L}_{\eta^\star}^{-1} S^\star$, Γ Lipschitz, $S^\star$ structured, and $\mathcal{L}_{\eta^\star}^{-1}$ bounded.*

Assumption B.9 (Approximation errors)**.** *There are source truncations $S^{(P)}$ and basis-approximated inverses $\widetilde{\mathcal{L}}_B^{-1}$ with $\left\|S^\star - S^{(P)}\right\|_{\mathcal{S}} \le \varepsilon_P$ and $\left\|\mathcal{L}_{\eta^\star}^{-1} g - \widetilde{\mathcal{L}}_B^{-1} g\right\|_{\mathcal{H}} \le \varepsilon_B \|g\|_{\mathcal{S}}$ for all $g \in \mathcal{S}$.*

The main-text assumptions (A1), (A2), and (A3) are Assumptions B.1, B.2, and B.3–B.4; Assumptions B.5–B.9 are used only by Theorems B.10 and B.11.

B.2 Supporting results

Theorem B.10 (Conditional approximation under a source-medium-response factorisation)**.** *Under Assumptions B.1–B.9, for a target that admits the factorisation of Assumption B.8, there is an AURA model with multipole order P and basis size B such that $\|\mathcal{F} - \mathcal{F}_{\mathrm{AURA}}\| \le C_1\varepsilon_P + C_2\varepsilon_B$. If $\varepsilon_P = O(P^{-a})$ and $\varepsilon_B = O(B^{-b})$, the error is $O(P^{-a} + B^{-b})$. This is not a universal approximation theorem: it is a rate statement for targets that already admit a compatible factorisation.*

Theorem B.11 (Stability under aligned perturbations)**.** *For a fixed number of AURA layers, under Assumptions B.1–B.7 and with Lipschitz input encoders and readouts, there is $C > 0$, depending on the depth, the resolvent bound, and the Lipschitz constants, such that for aligned inputs with shared node indexing $\|f_{\mathrm{AURA}}(G,X,R) - f_{\mathrm{AURA}}(G',X',R')\| \le C\big(d_G(G,G') + \|X - X'\| + \|R - R'\|\big)$.*

Proposition B.12 (Structural encodings carry no label information on the interference tasks)**.** *Suppose the graph G is drawn independently of the label y, and the attributes of the probe's rooted L-hop*

neighbourhood are identical across classes. Then for any function π of the graph alone (for example Laplacian eigenvectors, random-walk encodings, or shortest-path distances) and any depth-L local MPNN that receives $\pi(G)$ as extra node features, the probe representation is independent of y, so no classifier on it exceeds the majority-class rate.

Proof. By Lemma 3.4, the probe representation is a function of the probe's rooted L-hop neighbourhood together with the node features inside it, here the original attributes, which are constant across classes, and the encodings $\pi(G)$ restricted to that neighbourhood. This is a measurable function of G alone, and G is independent of y, so the representation is independent of y. A classifier whose input is independent of the label cannot beat the best constant predictor. □

In the three generators the topology is sampled before, and independently of, the class-dependent step, which assigns only node attributes (the remote charges, the sign field, or the ring charges; in Shielded Source the topology is identical across classes), so the proposition applies with $L = 3$.

B.3 Proofs

Proof of Theorem 3.1. Under the hypotheses, $\Phi_i = \sum_j (A_G)_{ij} W_s h_j$, and $(A_G)_{ij} = 0$ unless $j \in \mathcal{N}(i) \cup \{i\}$. The linear response gives $h_i^+ = W_0 h_i + \sum_{j \in \mathcal{N}(i)\cup\{i\}} W_1 (A_G)_{ij} W_s h_j$. Setting $\widetilde{W}_{ij} = W_1 (A_G)_{ij} W_s$ for $j \in \mathcal{N}(i)$ and $\widetilde{W}_0 = W_0 + W_1 (A_G)_{ii} W_s$ yields the claimed update. □

Proof of Theorem 3.2. Since $\rho(\beta T_G) < 1$ in finite dimension, the Neumann series $\sum_{t\geq 0} (\beta T_G)^t$ converges in operator norm to $(I - \beta T_G)^{-1}$. Multiplying by S gives the claim. □

Proof of Lemma 3.3. The residual is $\sum_{t=K+1}^{\infty} (\beta T_G)^t S$; the triangle inequality and the geometric series give the bound. □

Proof of Lemma 3.4. By induction on depth. At depth 0 the representation is the input feature. At depth $\ell + 1$ it is a function of the depth-ℓ representations of v and of $\mathcal{N}(v)$ and of the incident edge attributes, each of which depends only on a rooted ℓ-hop neighbourhood. No step depends on the width. □

Proof of Theorem 3.5. Fix L. Let H_L be a rooted radius-L graph centred at $v^\star$ shared by all instances. Attach two disjoint paths of equal length $D > L$ from boundary vertices of H_L to source leaves s_1, s_2, with intermediate node features identical across instances. Let $G^{(+)}$ have source strengths $(q_1, q_2) = (1, 1)$ and $G^{(-)}$ have $(-1, -1)$. Their rooted L-hop attributed neighbourhoods of $v^\star$ coincide, so by Lemma 3.4 every depth-L local MPNN, of any width, gives both the same representation and hence the same prediction; with probability $1/2$ on each graph the error is at least $1/2$. For AURA, choose a field operator whose influence kernel at the probe is positive and symmetric along the two paths. Then $\Phi(v^\star) = \kappa(q_1 + q_2)$ for some $\kappa > 0$, so $\Phi_{G^{(+)}}(v^\star) = 2\kappa > 0$ and $\Phi_{G^{(-)}}(v^\star) = -2\kappa < 0$, and thresholding at 0 recovers the labels with margin 2κ. □

Proof of Theorem B.10. Write $\mathcal{F} = \Gamma(\Phi^\star)$ and $\Phi^{(P,B)} = \widetilde{\mathcal{L}}_B^{-1} S^{(P)}$. By the triangle inequality, $\|\Phi^\star - \Phi^{(P,B)}\|_{\mathcal{H}} \leq \|\mathcal{L}_{\eta^\star}^{-1}(S^\star - S^{(P)})\|_{\mathcal{H}} + \|(\mathcal{L}_{\eta^\star}^{-1} - \widetilde{\mathcal{L}}_B^{-1}) S^{(P)}\|_{\mathcal{H}}$. Boundedness of $\mathcal{L}_{\eta^\star}^{-1}$ bounds the first term by $C_{\mathcal{L}} \varepsilon_P$; Assumption B.9 with $\|S^{(P)}\|_{\mathcal{S}} \leq C_S$ bounds the second by $C_S \varepsilon_B$. Lipschitz continuity of Γ gives $\|\mathcal{F} - \mathcal{F}_{\mathrm{AURA}}\| \leq L_\Gamma C_{\mathcal{L}} \varepsilon_P + L_\Gamma C_S \varepsilon_B$, and the power-law assumptions give the rate. □

Proof of Theorem B.11. Let $\delta = d_G(G, G') + \|X - X'\| + \|R - R'\|$ and $\Delta_\ell = \max_i \|h_i^{(\ell)} - h_i'^{(\ell)}\|$. We show $\Delta_{\ell+1} \leq A\Delta_\ell + B'\delta$. *Sources:* by Assumption B.5, $\|S^{(\ell)} - S'^{(\ell)}\| \leq C_{S,1}\Delta_\ell + C_{S,2}\delta$. *Resolvent:* with $R_\ell = (I - \beta T_{\eta^{(\ell)}}(G))^{-1}$ and R'_ℓ its perturbed counterpart, the resolvent identity gives $R_\ell - R'_\ell = R_\ell\, \beta \big(T_{\eta^{(\ell)}}(G) - T_{\eta'^{(\ell)}}(G')\big) R'_\ell$, so by Assumptions B.6 and B.7, $\|R_\ell - R'_\ell\| \leq \beta M^2 (C_{T,1}\delta + C_{T,2}\Delta_\ell)$. *Fields:* combining $\Phi^{(\ell)} = R_\ell S^{(\ell)}$ with the bounds above

and bounded sources gives $\left\|\Phi^{(\ell)} - \Phi'^{(\ell)}\right\| \le A_1\Delta_\ell + B_1\delta$. *Probes and response:* Assumptions B.3 and B.4 give $\Delta_{\ell+1} \le A\Delta_\ell + B'\delta$. Unrolling from $\Delta_0 \le C_0\delta$ with a Lipschitz readout gives the claim. □

B.4 Remark on information flow

In a depth-L MPNN, remote influence on a probe passes through L successive compressions into intermediate node states. In AURA, the probe field $\Phi(v^\star) = \sum_i K_G(v^\star, i)\, q_i$ accumulates remote contributions through a solved field with kernel K_G. This changes the mechanism by which remote nodes affect the probe, but it is a qualitative observation, not a curvature, effective-resistance, or Jacobian bound, and not a claim of dominance over nonlocal architectures.

C Experimental protocol

Table 5: Training settings for every reported run, as recorded in the released code and per-run result files.

	Synthetic	QM9 property	QM9 field
Data	1,600 graphs of 40 nodes per task; 1,120/240/240	30k-molecule subset; five 80/10/10 splits; first 4k training molecules used	10k-molecule subset; five 80/10/10 splits; 2k training molecules; 64 probes per molecule
Seeds / splits	5 seeds	5 splits	5 splits
Layers / hidden	3 / 64	3 / 64	3 / 48
Batch size	32 (all models)	32 (all models)	8 (AURA, Probe-Token), 16 (Kernel-Sum)
Epochs	at most 40	at most 8	8
Early stopping	validation accuracy; patience 12 after 15 epochs	validation MAE; patience 4 after 4 epochs	best validation checkpoint
Loss	cross-entropy at the probe	L1, standardised target	L1, standardised V
Optimiser	AdamW, learning rate 10^{-3}, weight decay 10^{-5}, gradient clipping at norm 5		
AURA	$B = 32$, $K = 6$, β initialised at 0.6, 16 Gaussian RBFs on $[0, 6]$ Å, residual response		

Baselines. GCN, GraphSAGE, GIN, and GAT (two heads) use a linear input layer, three convolutions with ReLU, and a two-layer head, reading out at the probe (synthetic) or by mean pooling (QM9). The Graphormer-style model adds, in each layer, a GIN convolution, a full intra-graph multi-head attention block (two heads, pre-norm, residual), and a feed-forward block; the GPS-style model is identical with a GAT convolution. Neither uses Graphormer's centrality or spatial encodings, and neither uses coordinates. The SchNet-style model applies three continuous-filter convolutions with 16 Gaussian radial basis functions (cutoff 5 Å) to coordinates. For the field task, Kernel-Sum sums a learned function of atom type and atom-to-probe distance over atoms, and the Probe-Token Transformer lets probe tokens (position and distance to the centroid) cross-attend to atom tokens (type and position) through two attention layers.

Ablation budgets. The Remote Charge ablations and the basis-size sweep cap training at 25 epochs; the ablations on the other two tasks use the 40-epoch cap of the main runs. Within each task all variants share the same budget.

Controls on the field task. The controls in Table 3 (bottom) use one protocol: seed 0, 2k training molecules, 8 epochs with early stopping on 300 validation molecules. The permutation control shuffles atom types across atoms within each molecule while keeping coordinates and edges, and computes the target from the true atoms. The summary-only model regresses $V(p)$ from seven molecule-level statistics (centroid, mean and standard deviation of atomic number, and two coordinate spreads) and the probe's offset and distance to the centroid.

Hardware. All runs use a single NVIDIA Quadro P520 GPU (2 GB) with 30 GB of system memory. Total wall-clock for the QM9 table is about 2.2 hours.

D Additional results

D.1 All three interference tasks

Table 6 reports every model on all three tasks. The per-seed AURA accuracies are 0.846, 0.563, 0.833, 0.838, 0.858 (Remote Charge); 1.000, 0.500, 0.988, 1.000, 0.996 (Resonant Cycle); and 1.000, 1.000, 1.000, 0.792, 0.517 (Shielded Source). Four of five seeds exceed 0.75 on each task; the failures are individual seeds that stay near the constant-prediction level. Remote Charge is the hardest task for AURA: it is never solved exactly, while most seeds solve the other two.

Table 6: All models on all three interference tasks: probe-readout accuracy, mean $\pm$ std over five seeds, and mean wall-clock per epoch (s).

	Remote Charge		Resonant Cycle		Shielded Source	
Model	Acc.	s/ep.	Acc.	s/ep.	Acc.	s/ep.
GCN	0.489 ± 0.053	0.97	0.500 ± 0.000	1.01	0.510 ± 0.013	1.07
GraphSAGE	0.446 ± 0.000	0.81	0.500 ± 0.000	1.08	0.503 ± 0.016	1.11
GIN	0.446 ± 0.000	0.90	0.500 ± 0.000	1.01	0.503 ± 0.016	1.04
GAT	0.446 ± 0.000	1.48	0.500 ± 0.000	1.48	0.503 ± 0.016	1.09
Graphormer-style	1.000 ± 0.000	3.37	1.000 ± 0.000	2.18	1.000 ± 0.000	1.74
GPS-style	1.000 ± 0.000	4.44	1.000 ± 0.000	2.88	1.000 ± 0.000	2.68
AURA	0.787 ± 0.113	4.39	0.897 ± 0.198	4.24	0.862 ± 0.190	5.29

D.2 Component ablations

Table 7 removes one component at a time (seed 0). Only the learned medium is required by all three tasks; the differential probe matters on Remote Charge and Resonant Cycle but not on Shielded Source, scalar-only sources suffice for Resonant Cycle only, and removing the multipole term leaves Remote Charge intact. The three tasks therefore probe different components.

Table 7: Component ablations, seed-0 accuracy; parameter counts in parentheses.

Variant	Params	Remote Charge	Resonant Cycle	Shielded Source
Full AURA	193,508	0.846	1.000	1.000
Uniform (unlearned) medium	193,508	0.533	0.508	0.483
No differential probe (value only)	168,932	0.646	0.500	1.000
Scalar sources and value probe only	125,252	0.533	1.000	0.483
No multipole term	149,828	0.858	0.500	0.496

D.3 Basis-size sweep

Table 8 varies B on Remote Charge (seed 0, 25-epoch cap). The sweep is non-monotone: too few channels fail, 16 and 32 succeed, and 64 channels fail under the same budget despite 50% more parameters than $B = 32$, so the result is not a capacity effect. The $B = 32$ configuration is identical to the full model of Table 7 but was run separately and reached 0.908 rather than 0.846; GPU scatter aggregation is not bitwise deterministic, and differences of this size are consistent with the seed-level spread in Table 1. We therefore read single-seed diagnostics only for large effects.

D.4 Field reconstruction: qualitative example and vector field

Figure 3 shows a representative test molecule. In a vector-valued variant, we replace the scalar head with a three-component head and regress $E(p) = \sum_a Z_a (p - r_a) \big(\|p - r_a\|^2 + \varepsilon^2 \big)^{-3/2}$ at

Table 8: Basis-size sweep on Remote Charge, seed 0.

B	Accuracy	Params	Wall-clock (s)
8	0.446	120,572	62.2
16	0.913	144,884	87.1
32	0.908	193,508	89.9
64	0.508	290,756	54.8

the same probes (1,000 training molecules, 8 epochs, two seeds). AURA's component MAE is 0.344 ± 0.017 against 0.400 ± 0.024 for Kernel-Sum, and its magnitude MAE is 0.444 ± 0.012 against 0.580 ± 0.031; cosine similarity is near saturation for both (0.996 and 0.994).

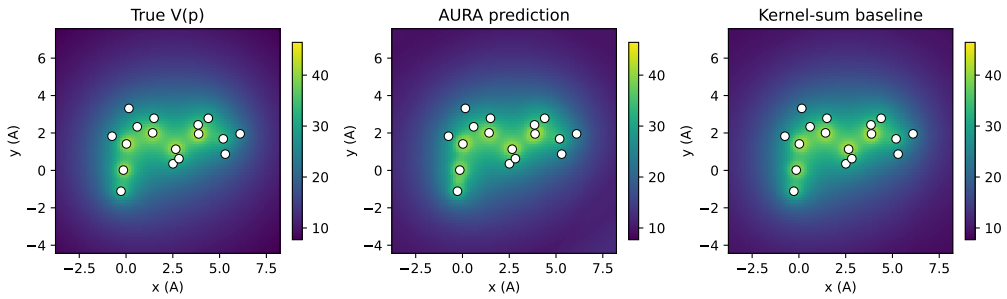

Figure 3: **Qualitative field reconstruction.** For one test molecule, from left to right: the Coulomb surrogate $V(p)$ on a plane through the centroid, AURA's prediction, and the Kernel-Sum prediction. White markers indicate atoms within 1.2 Å of the plane.

D.5 Perturbation diagnostics

These single-run diagnostics check whether the ordering persists under small test-time perturbations; they do not claim certified robustness. The QM9 runs here use an auxiliary setting (10k-molecule subset, 4k training molecules, hidden width 48, 15 epochs, μ only), and the field runs use 1,500 training molecules. Tables 9 and 10 add isotropic Gaussian noise of standard deviation σ (Å) to coordinates, or rewire a fraction p of each graph's edges.

Table 9: MAE under coordinate noise σ (Å). The GIN baseline does not use coordinates.

Task	Model	$\sigma = 0$	0.01	0.03	0.05	0.10
QM9 μ	GIN	0.815	0.815	0.815	0.815	0.815
	SchNet-style	0.757	0.759	0.810	0.828	0.957
	AURA	**0.603**	**0.603**	**0.606**	**0.606**	**0.656**
Field	Kernel-Sum	0.801	0.801	0.803	0.815	0.868
	AURA	**0.631**	**0.632**	**0.639**	**0.657**	**0.727**

Table 10: MAE when a fraction p of each graph's edges is rewired at test time.

Task	Model	$p = 0$	0.01	0.03	0.05
QM9 μ	GIN	0.815	0.913	0.913	0.935
	SchNet-style	0.757	0.791	0.791	0.808
	AURA	**0.603**	**0.682**	**0.682**	**0.711**
Field	Kernel-Sum	0.801	0.801	0.801	0.801
	AURA	**0.631**	**0.635**	**0.635**	**0.634**