Environment-Aware Multiscale Geometric Interaction for Equivariant Molecular Spectral Prediction

Haoran Li[1,2*], **Weiran Cui**[1,2] and **Minghui Li**[2]

[1]Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, China
[2]University of Chinese Academy of Sciences, Beijing, China
lihaoran@fjirsm.ac.cn

Abstract

Predicting molecular spectra requires modeling 3D conformation and solvent modulation. However, $E(3)$-equivariant networks based on local message passing exhibit limited sensitivity to long-range geometric dependencies, affecting the discrimination of globally distinct conformers. We introduce the *Multiscale Geometric Interaction Layer* (MGIL), which integrates global context by augmenting local features with centroid-referenced anchors, geometric moments, and virtual nodes. This design explicitly encodes global anisotropy while maintaining equivariance. Furthermore, we propose a *Solvent Field Modulator* (SFM) to encode solvent topology for conditional feature adaptation. Experiments demonstrate that MGIL enhances the capture of global structural variations, yielding consistent performance gains across spectral prediction benchmarks while maintaining linear computational efficiency.

1 Introduction

Predicting absorption and emission maxima in solution is central to screening organic emitters and chromophores. Large experimental collections now enable data-driven modeling [Joung *et al.*, 2020; Ju *et al.*, 2021; Beard *et al.*, 2019; Taniguchi and Lindsey, 2018]. Yet solution-phase spectra depend on two intertwined factors: the global 3D arrangement of the solute and the environment that perturbs it. This creates a dual requirement for learning systems: respect $SE(3)$ symmetry while remaining sensitive to long-range geometry and solvent modulation.

Despite rapid progress, most predictors still operate on a single static geometry and treat solvent as metadata. In practice, experimental spectra reflect conformational ensembles and site-specific solvation effects. A reliable model must therefore encode global 3D organization and offer a mechanism for environment-dependent feature modulation, rather than relying on implicit correlations in the training set. In this benchmark, each solute–solvent measurement is paired with a single optimized S0 geometry as a proxy. We adopt this static proxy setting to explore how much predictive gain can be recovered via explicit global geometric modeling and topological solvent conditioning, without relying on computationally expensive ensemble sampling.

Geometry gap: local cutoffs vs. global conjugation. Equivariant GNNs such as EGNN and PaiNN are effective for local 3D interactions [Satorras *et al.*, 2021; Schütt *et al.*, 2021]. In practice, however, most architectures rely on finite-radius message passing. As a result, long-range anisotropy and conjugation patterns are absorbed only after many hops and often fade through smoothing. Distinct global conformations with nearly identical local neighborhoods can therefore receive nearly identical representations. Attention-based backbones improve connectivity [Liao and Smidt, 2023; Ying *et al.*, 2021], but they still lack explicit global geometric reference cues.

Environment gap: passive tags vs. active fields. Solvent effects are anisotropic and site-specific. Many models reduce the solvent to a categorical label or a small set of descriptors, implicitly treating it as a passive tag. This loses the topological information that drives hydrogen bonding, polarity, and π-stacking, and it cannot express how different solute regions are stabilized differently.

Figure 1 illustrates the two gaps motivating our design. As shown in Figure 1(a), under a fixed cutoff r_c, globally distinct conformers may exhibit nearly identical local neighborhoods, leading to representation collapse for locality-limited message passing. Figure 1(b) highlights a receptive-field bottleneck: long-range geometry must propagate over multiple hops ($\leq kr_c$) and may be diluted by over-smoothing. Finally, Figure 1(c) contrasts the environment gap: global solvent tags provide uniform conditioning, while solvent-graph conditioning can produce solute-node-specific contexts via cross-attention.

We address both gaps with a mechanism-driven framework. The Multiscale Geometric Interaction Layer (MGIL) injects global reference and shape statistics into local edges, while the Solvent Field Modulator (SFM) treats the solvent graph as a conditional field that actively recalibrates solute features. We model this anisotropy through solute-node-specific conditioning from the solvent graph, without assuming an explicit 3D solute–solvent relative pose.

*Corresponding author.

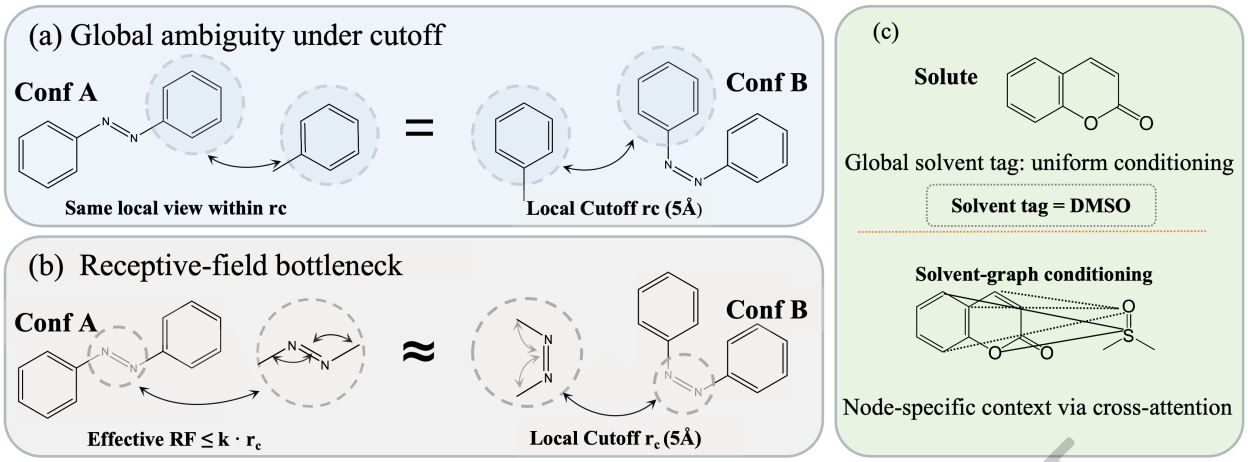

Figure 1: (a) Fixed cutoff r_c yields global ambiguity. (b) Receptive-field bottleneck: long-range geometry needs multiple hops ($\leq kr_c$) and can be diluted. (c) Environment gap: global solvent tags are uniform, while solvent-graph conditioning provides solute-node-specific context.

1.1 Contributions

- **Multiscale Geometric Interaction Layer (MGIL).** We enrich equivariant message passing with centroid-referenced anchors and moment-based shape statistics, plus an efficient virtual-node pathway for long-range broadcast.
- **Solvent Field Modulator (SFM).** We encode solvent molecules as 2D graphs and inject their influence through cross-attention and FiLM-style modulation, enabling solute-node-specific conditioning.
- **Systematic evaluation and ablations.** Across multiple equivariant backbones, MGIL provides consistent gains, while SFM adds complementary improvements when solvent information is available. Controlled ablations isolate the contribution of anchors, moments, and virtual-node communication.

Supplementary material. Additional proofs, pseudocode, feature definitions, extended ablations, and source code are available at https://github.com/ironlossli/flu.

2 Related Work

2.1 Equivariant Molecular GNNs

Early equivariant models such as Tensor Field Networks and Cormorant rely on higher-order spherical representations [Thomas *et al.*, 2018; Anderson *et al.*, 2019]. SchNet introduced continuous-filter convolutions on distance features [Schütt *et al.*, 2017], while DimeNet and SphereNet injected explicit angular and spherical geometry for higher fidelity [Klicpera *et al.*, 2020; Liu *et al.*, 2022]. EGNN and PaiNN provide efficient equivariant formulations without heavy spherical harmonics, using coordinate updates and vector channels [Satorras *et al.*, 2021; Schütt *et al.*, 2021]. Attention-based equivariant backbones such as Equiformer further broaden context via global tokenization and attention [Liao and Smidt, 2023; Ying *et al.*, 2021]. EquiformerV2 scales equivariant transformers to higher-degree representations [Liao *et al.*, 2023], while LeftNet emphasizes local frame transition encoding for expressive yet efficient equivariant GNNs [Du *et al.*, 2023]. Our MGIL complements these models by injecting explicit global geometric references with linear complexity.

2.2 Machine Learning for Molecular Spectra

Deep learning has been increasingly applied to predict excited-state properties. Early benchmarks like Deep4Chem [Joung *et al.*, 2020] utilized 2D fingerprints (ECFP) or descriptors to predict absorption peaks. While effective for simple scaffolds, 2D methods fundamentally fail to capture conformer-dependent properties. Recent efforts have moved towards 3D-GNNs. For instance, [Hung *et al.*, 2023] applied a featurized SchNet to predict photophysical data, and [Joung *et al.*, 2021] extended Deep4Chem with deep learning models for solution-phase properties. Complementary efforts also predict absorption maxima using alternative representations, such as convolutional models over molecular depictions [Jung *et al.*, 2024]. However, these works often focus on implicit solvent models or limited solute-solvent pairs, neglecting the explicit topological structure of the solvent environment. Very recently, a solvent-aware benchmark (nablaColors) was proposed to pair optical properties with solvent-conditioned GNNs [Potapov *et al.*, 2025], underscoring the need for explicit solute–solvent interaction modeling. Recently, [Zhu *et al.*, 2025] proposed a modular framework (FLAME) for fluorophore design, highlighting the need for specialized architectures that go beyond generic property predictors.

2.3 Solvent Modeling and Conditional Modulation

In learned spectral prediction, solvent effects are often encoded as categorical labels or global descriptors, which fail to express anisotropic, site-specific interactions. Conditional modulation layers such as FiLM provide a general mechanism to inject auxiliary signals into neural fea-

tures [Perez *et al.*, 2018], and GNN-FiLM extends this idea to graph settings [Brockschmidt, 2020]. Our SFM builds on this paradigm by using a solvent graph encoder and cross-attention to deliver node-specific conditioning, bridging the gap between passive tags and structured solvent-graph conditioning.

3 Method

Figure 2 illustrates the overall architecture, which integrates the Multiscale Geometric Interaction Layer (MGIL) for solute encoding and the Solvent Field Modulator (SFM) for environment conditioning.

3.1 Preliminaries and Problem Formulation

Notations. We denote the solute as a geometric graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{H}, \mathbf{R})$, where $\mathbf{H} \in \mathbb{R}^{N\times d}$ are node features and $\mathbf{R} \in \mathbb{R}^{N\times 3}$ are coordinates. For node i, $\mathbf{r}_i$ and $\mathbf{h}_i$ denote its position and feature vector. The solvent environment is encoded as a graph $\mathcal{S}$ and injected as conditioning features (Section 3.4). We use d_{ij} for interatomic distance, $\mathbf{u}_{ij}$ for its unit direction, and $\mathbf{c}$ for the global centroid. Our goal is to learn a mapping $\Phi : (\mathcal{G}, \mathcal{S}) \to \mathbb{R}$ to predict spectral targets (absorption or emission). For Abs and EM, the target y is the peak wavelength ($\lambda_{\max}$) in nanometers. The prediction must satisfy the $SE(3)$-invariance constraint $\Phi(\mathbf{QR}+\mathbf{t}, \mathbf{H}, \mathcal{S}) = \Phi(\mathbf{R}, \mathbf{H}, \mathcal{S})$ for any rotation $\mathbf{Q} \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$, although intermediate representations are allowed to be equivariant. We impose invariance to rigid motions; we do not model chirality (parity-odd features).

3.2 Bottleneck: Locality-Dominated Equivariant Modeling

While $SE(3)$-equivariant networks can theoretically represent any 3D geometry, practical implementations usually enforce strict locality via radial cutoffs (typically 5–10Å) to minimize computational costs. This creates a *receptive-field bottleneck*. Spectral properties often depend on global organization—conjugation length, long-range steric hindrance, or the alignment of distant dipolar motifs. In local models, this information must propagate through multiple hops, often becoming diluted before reaching the readout. Consequently, distinct global conformations can become indistinguishable to the model if their local neighborhoods share identical topologies under the cutoff. Even topological attention mechanisms do not guarantee that the model captures the explicit geometric frame.

These issues motivate our two-pronged approach: (i) an explicit multiscale geometric interaction to inject global reference info into local edges, and (ii) environment-aware feature modulation that treats solvent as a conditioning field.

3.3 Multiscale Geometric Interaction Layer (MGIL)

MGIL is designed to augment local equivariant interactions with explicit multiscale geometric descriptors. Crucially, it is engineered as a lightweight plugin, maintaining linear complexity $O(|\mathcal{E}|)$ to ensure scalability. Figure 2(a) details the architecture, where MGIL enriches each local edge with a unified multiscale geometry descriptor, combining distance, moment invariants, and centroid-referenced anchors. Let $d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ and $\mathbf{u}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/(d_{ij} + \epsilon)$.

Local distance channel. We use K radial basis features $\mathrm{rbf}(d_{ij}) \in \mathbb{R}^K$.

Global anchors: restoring a molecule-level reference. The necessity of global anchors arises from the limitations of strict cutoffs. With a strict cutoff, local edge geometry does not inform the architecture *where* an interaction occurs within the larger molecule. This can lead to representation collapse: two configurations may share nearly identical local neighborhoods yet differ significantly in long-range arrangement. Centroid-referenced anchors mitigate this by providing a molecule-level reference direction, allowing each local interaction to carry information about its *global placement* and helping downstream layers distinguish globally distinct conformations.

Let $\mathbf{c}$ be the per-molecule centroid and $\mathbf{g}_i = (\mathbf{r}_i - \mathbf{c})/(\|\mathbf{r}_i - \mathbf{c}\| + \epsilon)$.

Choice of global reference. We employ the geometric centroid (coordinate mean) as a single global reference due to its simplicity, differentiability, and consistency across molecules. Alternative oriented references such as PCA principal axes can be unstable: (i) the axis direction has an inherent $\pm v$ ambiguity, and (ii) when eigenvalues are close (near-spherical shapes), small geometric perturbations may cause large axis flips, injecting noise into learning. A center-of-mass reference is also viable, but our formulation does not require mass-weighting and remains purely geometry-based. In our setting with optimized S_0 conformers, the geometric centroid is stable and sufficient to convey global shape and elongation cues.

We define anchor features $f_{\mathrm{radial}} = \langle \mathbf{u}_{ij}, \mathbf{g}_i \rangle$ and $f_{\mathrm{radius}} = \|\mathbf{r}_i - \mathbf{c}\|$, so $\mathrm{anchors}_{ij} \in \mathbb{R}^2$. The radial term encodes orientation, while the radius captures scale relative to the global shape without redundant trigonometric coupling.

Moment invariants: local shape statistics. We compute node-level moments from neighbor directions under a smooth cutoff weight. For node i, we form a first moment and a second moment:

$$\mathbf{s}_i = \sum_{k\in\mathcal{N}(i)} w(d_{ik})\,\mathbf{u}_{ik}, \quad \mathbf{C}_i = \sum_{k\in\mathcal{N}(i)} w(d_{ik})\,\mathbf{u}_{ik}\mathbf{u}_{ik}^\top. \tag{1}$$

We use a smooth cutoff weight $w(d) = \exp(-d^2/\sigma^2) \cdot \mathrm{cut}(d; r_c)$, where $\mathrm{cut}(\cdot)$ is a standard cosine cutoff within radius r_c. We then form edge-level invariants by projecting these moments along $\mathbf{u}_{ij}$. We define the moment invariants used in this work as:

$$\mathrm{moments}_{ij} = \Big[\log \sum_{k\in\mathcal{N}(i)} w(d_{ik}),\ \|\mathbf{s}_i\|,\ \mathbf{u}_{ij}^\top \mathbf{s}_i, \\ \mathbf{u}_{ij}^\top \mathbf{C}_i \mathbf{u}_{ij},\ \mathrm{tr}(\mathbf{C}_i^2)\Big] \in \mathbb{R}^5. \tag{2}$$

Compared to enumerating angles/dihedrals (which scales combinatorially with neighbor tuples), low-order moments provide a compact, permutation-agnostic summary of neighborhood anisotropy. We do not claim these five scalars are

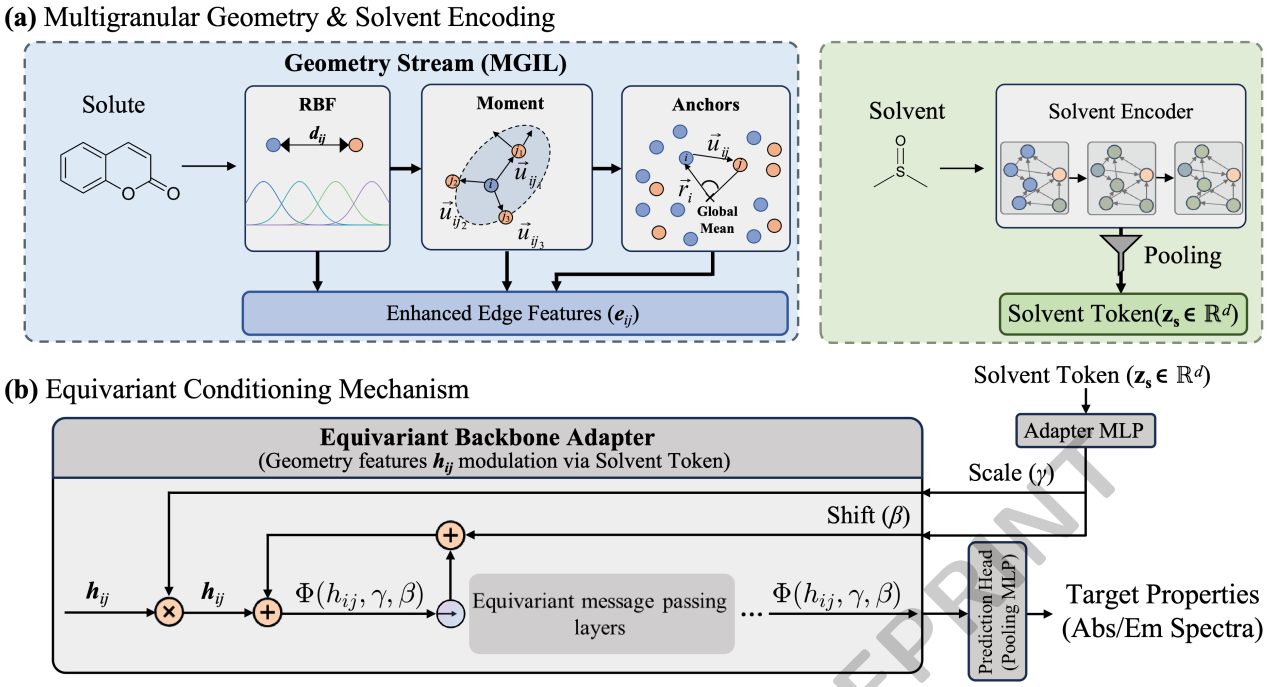

Figure 2: Framework overview: MGIL adds distance channels, moment statistics, and centroid anchors (optional virtual node), while SFM encodes the solvent graph and modulates solute features via cross-attention and FiLM.

mathematically necessary; they are a stable, lightweight set that captures density, directional bias, and anisotropy (see the supplementary material for full definitions).

Unified multiscale interaction feature. We concatenate these components:

$$\mathbf{e}_{ij} = [\mathrm{rbf}(d_{ij}),\ \mathrm{moments}_{ij},\ \mathrm{anchors}_{ij}], \tag{3}$$

optionally applying a bounded nonlinearity (such as $\tanh$ clipping) for stability.

Virtual node: efficient long-range pathway. We add a graph-level token that is updated by pooling and broadcasting once per layer. This creates a long-range information path with near-linear cost.

Backbone coupling. We feed $\mathbf{e}_{ij}$ into the backbone's edge/message functions. Where available, we also include a lightweight nonlocal + LN mixing step. Global anchors provide the 3D *spatial reference*, while the virtual node handles *semantic broadcast*.

Complexity. For a graph with N nodes and $|\mathcal{E}|$ edges, MGIL adds $O(|\mathcal{E}|d)$ time (for RBFs, anchors, and moments) and $O(Nd)$ for virtual-node operations. Memory usage remains $O(|\mathcal{E}|d)$. Since the optional nonlocal mixing uses a kernelized attention approximation with $O(Nd)$ complexity, MGIL preserves linear scaling with respect to the number of edges.

MGIL features are $SE(3)$-invariant because they are built from distances and inner products of vectors that rotate together. For example, under $\mathbf{r}'_i = \mathbf{Q}\mathbf{r}_i + \mathbf{t}$, we have $\mathbf{u}'_{ij} = \mathbf{Q}\mathbf{u}_{ij}$ and $\mathbf{g}'_i = \mathbf{Q}\mathbf{g}_i$, so $\mathbf{u}'^{\top}_{ij}\mathbf{g}'_i = \mathbf{u}^{\top}_{ij}\mathbf{g}_i$. The full proof and MGIL pseudocode are provided in the supplementary material.

3.4 Solvent Field Modulator (SFM)

Solvent encoder. We encode solvent information from a 2D solvent graph using a message passing network with mean pooling. This produces a solvent embedding $\mathbf{z}_s$ and, optionally, solvent node embeddings. The solvent graph is constructed from the solvent SMILES using RDKit with standard atom/bond categorical features; full feature definitions are given in the supplementary material.

Field modulation mechanisms. We view the solvent as a modulation field acting on the solute. Fig. 2(b) shows how the solvent token modulates intermediate solute representations via cross-attention and FiLM-style gating, yielding solute-node-specific conditioning. Here the "field" is a topology-derived *latent* conditioning signal computed from the solvent 2D graph, rather than a spatially resolved electrostatic or steric field. "Site-specific" refers to *solute-node-specific* attention over solvent atoms (Eqs. (3)–(4)), not an explicit 3D solute–solvent configuration.

Anisotropy without explicit 3D solvent poses. Although we do not input an explicit 3D solute–solvent relative pose, conditioning can still be anisotropic because the *receptivity* is node-specific: each atom in the 3D solute graph queries the solvent embedding independently through cross-attention, producing distinct solvent contexts for different solute regions. Consequently, different solute regions can attend to

different solvent motifs, yielding site-dependent solvent relevance in a topology-conditioned (semantic) sense without claiming a spatially resolved solvation shell. Given solute node features $\mathbf{h}_i$ and solvent node embeddings $\mathbf{s}_j$, we compute a site-specific solvent context via cross-attention:

$$\mathbf{q}_i = \mathbf{W}_q \mathbf{h}_i, \quad \mathbf{k}_j = \mathbf{W}_k \mathbf{s}_j, \quad \mathbf{v}_j = \mathbf{W}_v \mathbf{s}_j, \tag{4}$$

$$\alpha_{ij} = \mathrm{softmax}_j \left(\frac{\mathbf{q}_i^\top \mathbf{k}_j}{\sqrt{d}} \right), \quad \mathbf{z}_{s,i} = \sum_j \alpha_{ij} \, \mathbf{v}_j. \tag{5}$$

Here $\mathrm{softmax}_j(\cdot)$ normalizes over solvent atoms j, and d denotes the attention hidden dimension. We then apply a FiLM-style scalar gating to modulate solute features:

$$(\gamma_i, \beta_i) = \mathrm{MLP}(\mathbf{z}_{s,i}), \quad \tilde{\mathbf{h}}_i = \gamma_i \odot \mathbf{h}_i + \beta_i. \tag{6}$$

FiLM is applied to scalar channels; for vector features we use scalar gates (scaling) without additive bias to preserve equivariance. When only global conditioning is desired, we replace $\mathbf{z}_{s,i}$ with the graph-level $\mathbf{z}_s$ and broadcast (γ, β) to all nodes. As a weak baseline, we include *tag conditioning* that concatenates $\mathbf{z}_s$ to the pooled solute embedding; this bypasses cross-attention and keeps message passing solvent-agnostic, so the solvent only influences the final head.

3.5 Instantiation in Equivariant Backbones

We implement backbone-specific adapters that expose a shared interface. MGIL contributes the geometry features $\mathbf{e}_{ij}$ and defines where they enter each backbone. SFM then modulates intermediate features via scalar FiLM and optional cross-attention, all while preserving the native message-passing structure of the host backbone.

3.6 Training Objective

Given a dataset $\mathcal{D} = \{(\mathcal{G}_i, \mathcal{S}_i, y_i)\}$, we train the model end-to-end using mean-squared error:

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{|\mathcal{D}|} \sum_i \|\Phi(\mathcal{G}_i, \mathcal{S}_i) - y_i\|_2^2. \tag{7}$$

3.7 Complexity Analysis

Standard geometric Transformers, such as Graphormer [Ying *et al.*, 2021], often incur $O(N^2)$ complexity due to global attention. In contrast, MGIL operates on a sparse connectivity defined by a radial cutoff, maintaining $O(|\mathcal{E}|)$ complexity for local interactions. The virtual node mechanism introduces a global pathway with $O(N)$ cost for aggregation and broadcasting. The SFM module computes cross-attention between solute nodes (N) and solvent embeddings (1 or N_{solv}), resulting in $O(N \cdot N_{\mathrm{solv}})$ complexity. Since $N_{\mathrm{solv}} \ll N$ in typical spectral datasets, our framework maintains linear scalability with respect to solute size, making it applicable to larger molecules.

4 Experiments

4.1 Main Results Across Backbones

Table 1 summarizes overall performance across backbones, where starred entries denote the full configuration (MGIL + SFM) on top of each backbone. All results use the same fixed 70/10/20 split and evaluation protocol described in Section 4.2. Dataset construction details are given in Section 4.2.

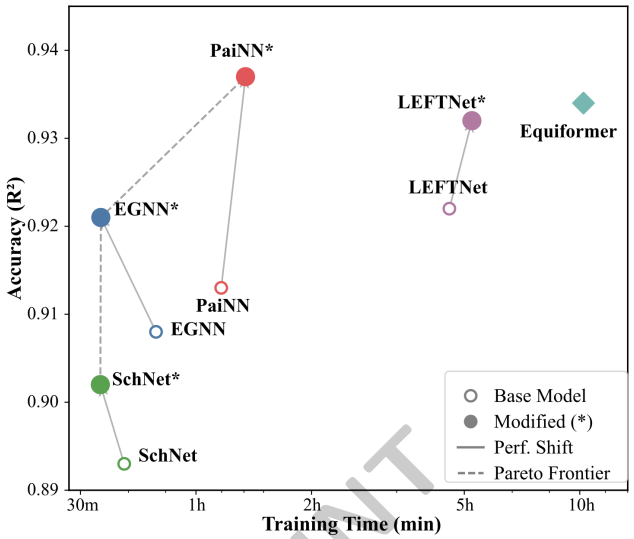

Figure 3: Accuracy–efficiency trade-off (R^2 vs training time). Hollow markers are base backbones; filled markers are MGIL + SFM; dashed curve is the Pareto frontier.

Key Takeaway. Our full model (MGIL + SFM) consistently improves over the original backbones across architectures and targets. Table 1 summarizes the main results across backbones and targets. While the effect size varies depending on the baseline capability, the improvement is universal. On Abs, the best entry (PaiNN*, i.e., backbone + MGIL + SFM) achieves 12.72 MAE, a 22.5% relative reduction compared to vanilla EGNN (16.42). On EM, LEFTNet* achieves the lowest RMSE (25.63), while LEFTNet obtains the lowest MAE (15.46). We further disentangle MGIL vs. SFM contributions via controlled ablations in Table 2.

Validation via Transformer Comparison. We observe that MGIL yields smaller marginal gains on Equiformer backbones (Table 1). Since Equiformer already employs global attention ($O(N^2)$), this result serves as a validation that MGIL successfully captures similar long-range global dependencies, but within a linear-complexity framework. Importantly, MGIL substantially boosts linear-complexity backbones ($O(E)$) such as PaiNN/EGNN, enabling lightweight message passing to match or surpass computationally intensive global-attention architectures under comparable settings. This highlights MGIL's value as an efficient geometric enhancement for scalable message passing.

Figure 3 highlights the accuracy–efficiency Pareto trade-off across backbones. As shown by the upward shifts, MGIL+SFM consistently improves the efficiency–accuracy profile of linear-complexity message passing backbones (such as EGNN/PaiNN), moving them closer to the Pareto frontier. In contrast, gains on global-attention Transformer backbones (such as Equiformer) are marginal, consistent with the validation noted above.

Model	Abs		EM	
	MAE	RMSE	MAE	RMSE
EGNN	16.42 ± 0.39	31.20 ± 0.43	25.83 ± 0.28	39.22 ± 0.33
EGNN*	13.78 ± 0.42	28.54 ± 0.55	23.57 ± 0.47	36.55 ± 0.53
SchNet	18.08 ± 0.33	33.62 ± 0.36	21.48 ± 0.27	34.48 ± 0.33
SchNet*	16.20 ± 0.21	32.18 ± 0.24	19.07 ± 0.39	31.42 ± 0.50
PaiNN	16.36 ± 0.27	30.04 ± 0.34	19.43 ± 0.44	30.46 ± 0.49
PaiNN*	$\mathbf{12.72 \pm 0.44}$	$\mathbf{25.17 \pm 0.58}$	16.65 ± 0.30	26.66 ± 0.35
LEFTNet	12.99 ± 0.49	26.85 ± 0.59	$\mathbf{15.46 \pm 0.23}$	25.95 ± 0.26
LEFTNet*	12.82 ± 0.45	25.57 ± 0.58	15.72 ± 0.44	$\mathbf{25.63 \pm 0.58}$
Equiformer	13.59 ± 0.36	26.44 ± 0.50	15.81 ± 0.31	25.78 ± 0.40
Equiformer*	13.69 ± 0.45	26.62 ± 0.58	15.92 ± 0.46	25.95 ± 0.58
Equiformer V2	15.44 ± 0.41	32.24 ± 0.46	16.44 ± 0.27	26.65 ± 0.32
Equiformer V2*	15.60 ± 0.22	32.45 ± 0.26	16.56 ± 0.23	26.87 ± 0.27

Table 1: Performance across backbones (Abs/EM MAE, RMSE in nm). "*" denotes MGIL + SFM; non-starred entries use the backbone defaults.

4.2 Datasets and Splits

Our dataset follows the FluoDB-style curation paradigm, aggregating solute–solvent measurements from multiple open-source databases (including Deep4Chem and ChemFluor). After filtering and range clipping, the final dataset contains Absorption: 21,903 samples and Emission: 19,125 samples, with a fixed 70/10/20 (train/val/test) split. We drop invalid SMILES, canonicalize with RDKit, remove small solutes (fewer than 10 atoms), and filter rare elements/solvents (frequency < 10). Targets are the peak wavelengths ($\lambda_{\mathrm{abs,max}}$ / $\lambda_{\mathrm{em,max}}$) in nm, clipped to 200–1100 nm. For 3D coordinates, we generate stereo-aware geometries from RDKit (manually built when embedding fails) and optimize all solute geometries at the S0 level using xTB (GFN2-xTB); the optimized conformer serves as a single-geometry proxy paired with each solute–solvent measurement.

4.3 Backbones and Variants

We evaluate five backbones: EGNN, PaiNN, SchNet, LeftNet, and Equiformer (v1/v2). We define MGIL Full/Off per backbone to maintain consistent semantics (see the supplementary material for the capability matrix). In this context, MGIL refers to the unified geometry stack: edge features, backbone coupling, virtual node, and nonlocal + LN where supported. Solvent conditioning is treated as a separate add-on. In Table 1, "*" denotes the full configuration with both MGIL and SFM enabled on top of the backbone.

Training protocol. We use the training hyperparameters specified in each model config. Unless otherwise stated, we train for 200 epochs with AdamW, batch sizes in the 32–64 range, and learning rates from the config (typically 3×10^{-4} to 5×10^{-4}); early stopping is disabled.

Backbone	MGIL	SFM	Abs(RMSE) ↓
EGNN	–	–	31.20 ± 0.43
	✓	–	29.34 ± 0.43
	✓	✓	$\mathbf{28.54 \pm 0.55}$
PaiNN	–	–	30.04 ± 0.34
	✓	–	26.56 ± 0.53
	✓	✓	$\mathbf{25.17 \pm 0.58}$

Table 2: Incremental impact of MGIL and SFM. Values report mean ± std.

Configuration	Abs (RMSE)	EM (RMSE)
Full Model	$\mathbf{28.54 \pm 0.55}$	$\mathbf{36.55 \pm 0.53}$
w/o Anchors	29.04 ± 0.55	38.76 ± 0.53
w/o Moments	28.96 ± 0.36	37.57 ± 0.33
w/o VNode	30.43 ± 0.24	38.96 ± 0.50
Off (Baseline)	31.20 ± 0.43	39.23 ± 0.34

Table 3: EGNN ablation of Anchors, Moments, and VNode. Values report mean ± std.

4.4 MGIL Component Gains

To investigate which geometry signal drives accuracy, we focus on additive gains. Our component ablations keep the MGIL stack fixed and toggle specific signals. Table 2 reports additive gains by enabling MGIL and SFM on representative backbones, isolating geometry and environment contributions. Table 3 further ablates individual MGIL components, including anchors, moments, and the virtual node, under a fixed EGNN backbone.

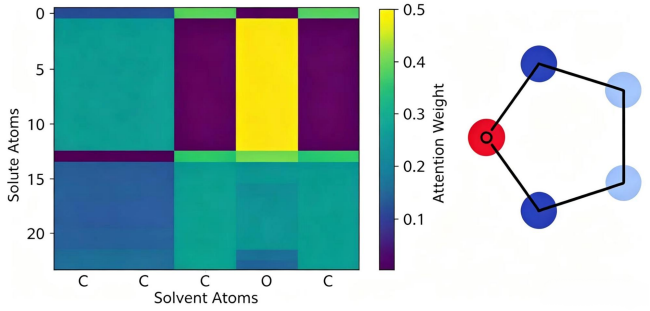

Figure 4: Solvent-graph attention for 1,4-dioxane. Left: heatmap α_{ij}; right: top-attended atoms on the 2D structure.

Mechanism	Abs (RMSE)	EM (RMSE)
Concatenation	29.11 ± 0.49	36.97 ± 0.58
Cross-Attn.	$\mathbf{28.54 \pm 0.55}$	$\mathbf{36.55 \pm 0.53}$

Table 4: Solvent conditioning: tag/concatenation vs cross-attention SFM. Values report mean $\pm$ std.

The results align with our hypothesis: anchors recover the global reference, moments correct local shape statistics, and the virtual node opens a long-range communication pathway. Removing the virtual node increases EGNN Abs RMSE from 28.54 to 30.43 and EM RMSE from 36.55 to 38.96, confirming its role in facilitating long-range communication. Combining all components yields the largest and most stable gains.

4.5 Qualitative Analysis: Interpreting Solvation

To validate whether SFM captures physically meaningful interactions beyond simple tagging, we visualize the cross-attention weights α_{ij} extracted from the trained model.

Figure 4 shows solvent-graph attention for 1,4-dioxane. Attention concentrates on a small subset of solvent atoms, highlighting chemically informative motifs (notably ether oxygen atoms). We interpret this as topology-conditioned relevance rather than a spatially resolved solvation shell. The full two-solvent comparison is provided in the supplementary material.

4.6 Solvent Conditioning

To isolate the effect of environment modeling, we evaluated conditioning adapters on top of a fixed MGIL Full configuration. Table 4 compares a global tag/concatenation baseline with cross-attention-based SFM to isolate the benefit of solute-node-specific conditioning.

5 Conclusion

We have proposed a geometry-first perspective for equivariant spectral prediction. By explicitly modeling geometry signals—anchors, moment invariants, and an efficient virtual node—we achieved controllable improvements across multiple backbones. Solvent conditioning adapters offered complementary gains without altering the core geometry semantics. Future work will focus on broader splits and scaling up interpretability analyses.

Ethical Statement

This work contributes to the advancement of molecular property prediction, with potential applications in material science and drug discovery. The datasets used are publicly available and do not contain personally identifiable information. We acknowledge the dual-use risk inherent in generative chemistry models and advocate for responsible usage in downstream applications. There are no conflicts of interest to declare.

References

[Anderson *et al.*, 2019] Brandon M. Anderson, Truong-Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 14510–14519, 2019.

[Beard *et al.*, 2019] Edward J. Beard, Ganesh Sivaraman, Álvaro Vázquez-Mayagoitia, Venkatram Vishwanath, and Jacqueline M. Cole. Comparative dataset of experimental and computational attributes of UV/Vis absorption spectra. *Scientific Data*, 6(1), 2019.

[Brockschmidt, 2020] Marc Brockschmidt. Gnn-film: Graph neural networks with feature-wise linear modulation. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 1144–1152. PMLR, 2020.

[Du *et al.*, 2023] Weitao Du, Yuanqi Du, Limei Wang, Dieqiao Feng, Guifeng Wang, Shuiwang Ji, Carla Gomes, and Zhi-Ming Ma. A new perspective on building efficient and expressive 3d equivariant graph neural networks. *arXiv preprint arXiv:2304.04757*, 2023.

[Hung *et al.*, 2023] Sheng-Hsuan Hung, Zong-Rong Ye, Chi-Feng Cheng, Berlin Chen, and Ming-Kang Tsai. Enhanced predictions for the experimental photophysical data using the featurized schnet-bondstep approach. *Journal of Chemical Theory and Computation*, 19(14):4559–4567, 2023.

[Joung *et al.*, 2020] Joonyoung F. Joung, Minhi Han, Minseok Jeong, and Sungnam Park. Experimental database of optical properties of organic compounds. *Scientific Data*, 7(1), 2020.

[Joung *et al.*, 2021] Joonyoung F Joung, Minhi Han, Jinhyo Hwang, Minseok Jeong, Dong Hoon Choi, and Sungnam Park. Deep learning optical spectroscopy based on experimental database: Potential applications to molecular design. *JACS Au*, 1(4):427–438, 2021.

[Ju *et al.*, 2021] Cheng-Wei Ju, Hanzhi Bai, Bo Li, and Rizhang Liu. Machine learning enables highly accurate predictions of photophysical properties of organic fluorescent materials: Emission wavelengths and quantum yields. *Journal of Chemical Information and Modeling*, 61(3):1053–1065, 2021.

[Jung *et al.*, 2024] Son Gyo Jung, Guwon Jung, and Jacqueline M. Cole. Automatic prediction of peak optical absorption wavelengths in molecules using convolutional neural networks. *Journal of Chemical Information and Modeling*, 2024.

[Klicpera *et al.*, 2020] Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[Liao and Smidt, 2023] Yi-Lun Liao and Tess E. Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.

[Liao *et al.*, 2023] Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.

[Liu *et al.*, 2022] Yi Liu, Limei Wang, Meng Liu, Yuchao Lin, Xuan Zhang, Bora Oztekin, and Shuiwang Ji. Spherical message passing for 3d molecular graphs. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

[Perez *et al.*, 2018] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron C. Courville. Film: Visual reasoning with a general conditioning layer. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 3942–3951. AAAI Press, 2018.

[Potapov *et al.*, 2025] Denis Potapov, Sergei Rogovoi, Kuzma Khrabrov, Konstantin Ushenin, Alexey Korovin, Anton Ber, Artur Kadurin, and Artem Tsypin. nablacolors: A 3d benchmark for optical property prediction with solvent-aware graph neural networks. *Research Square*, 2025.

[Satorras *et al.*, 2021] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 9323–9332. PMLR, 2021.

[Schütt *et al.*, 2017] Kristof 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 Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 991–1001, 2017.

[Schütt *et al.*, 2021] Kristof Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 9377–9388. PMLR, 2021.

[Taniguchi and Lindsey, 2018] Masahiko Taniguchi and Jonathan S. Lindsey. Database of absorption and fluorescence spectra of >300 common compounds for use in PhotochemCAD. *Photochemistry and Photobiology*, 94(2):290–327, 2018.

[Thomas *et al.*, 2018] Nathaniel Thomas, Tess E. Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *CoRR*, abs/1802.08219, 2018.

[Ying *et al.*, 2021] 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 Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 28877–28888, 2021.

[Zhu *et al.*, 2025] Yuchen Zhu, Jiebin Fang, Shadi Ali Hassen Ahmed, Tao Zhang, Su Zeng, Jia-Yu Liao, Zhongjun Ma, and Linghui Qian. A modular artificial intelligence framework to facilitate fluorophore design. *Nature Communications*, 16(1), 2025.