SGBNET: SCALAR-GEOMETRIC BIDIRECTIONAL NETWORK FOR EDGE-EFFICIENT INTERATOMIC POTENTIALS

Anonymous authors
Paper under double-blind review

ABSTRACT

Machine learning interatomic potentials often suffer from restricted geometric-to-scalar information flow and edge-dominated computation. We propose SGBNet, an equivariant architecture that improves the accuracy-efficiency trade-off through scalar-geometric bidirectional (SGB) interactions and edge-efficient convolution. SGB interactions improve geometric-to-scalar information flow by introducing explicit pathways from higher-order geometric features to invariant scalar features. In the SO(2) branch, conjugate multiplication cancels shared rotation-induced phases to construct scalar features. Edge-efficient SGB convolution leverages the structure of scalar features to move channel mixing from edges to nodes, reducing edge-domain computation. Experiments demonstrate clear accuracy-efficiency gains on molecular benchmarks, slightly improved performance on materials benchmarks, and scalability to large-scale systems. On SPICE-MACE-OFF, all SGBNet variants lie on the accuracy-efficiency Pareto frontier. The compact SGBNet-N maintains comparable MLIPAudit performance to eSEN-3.2M while reducing runtime and memory consumption by 47% and 60%, respectively. Moreover, SGBNet-N is both faster and substantially more accurate than the Euclidean-based ViSNet, narrowing the divide between efficient Euclidean models and more expressive higher-order equivariant architectures.

1 INTRODUCTION

For atomistic simulations, classical force fields Cornell et al. (1995) suffer from limited accuracy, while *ab initio* molecular dynamics (AIMD) Marx & Hutter (2009) remains extremely costly for large-scale systems. Machine learning interatomic potentials (MLIPs) Schütt et al. (2017); Deringer et al. (2019); Batatia et al. (2022); Wang et al. (2024); Fu et al. (2025); Wood et al. (2025) provide an efficient alternative by approaching DFT accuracy at substantially reduced cost.

Equivariant graph neural networks (GNNs) Satorras et al. (2021); Batzner et al. (2022); Fu et al. (2025) improve MLIP accuracy by modeling geometric symmetries and angular interactions, but still face a persistent accuracy-efficiency trade-off. Euclidean models Wang et al. (2024) offer efficient computation but limited geometric expressiveness, whereas higher-order geometric models achieve improved accuracy at the cost of expensive tensor operations. SO(2)-based architectures such as eSEN Fu et al. (2025) reduce the cost of tensor products through SO(2) convolution, but their computation remains dominated by edge-domain message passing Li et al. (2025). Maintaining high accuracy with compact architectures remains essential for large-scale molecular simulations, where larger models can become prohibitively expensive for biomolecular applications.

Beyond this, conventional gated equivariant architectures exhibit restricted geometric-to-scalar information flow. As the energy prediction head consumes only scalar features, higher-order features in the final block can affect the prediction only through normalization statistics. Formally, let h denote the scalar features before the prediction head, and W denote parameters whose direct outputs contribute only to higher-order geometric features. If the influence of W on h is mediated through a K-dimensional normalization statistic s, the corresponding Jacobian is low-rank as

$$\operatorname{rank}\left(\frac{\partial h}{\partial W}\right) = \operatorname{rank}\left(\frac{\partial h}{\partial s}\frac{\partial s}{\partial W}\right) \leqslant K. \tag{1}$$

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.

We refer to this structurally constrained pathway as *low-rank readout coupling*. This does not imply that higher-order features have no effect, but rather that their influence on scalar representations is constrained to a lower-dimensional pathway.

We propose SGBNet, a MLIP architecture that improves the accuracy-efficiency trade-off in SO(2)-based Passaro & Zitnick (2023) message passing networks. SGBNet addresses these architectural limitations through two complementary designs.

First, we develop scalar-geometric bidirectional (SGB) interactions through the SGB SO(2) and SO(3) linear. These modules address the restricted geometric-to-scalar pathway caused by low-rank readout coupling by explicitly mapping higher-order geometric features to second-order invariants. In particular, the SGB SO(2) linear constructs such invariants through conjugate multiplication within the same SO(2) frequency component. Second, we introduce edge-efficient convolution to alleviate edge-dominated computation through computation reordering and feature-channel redistribution. Without edge-dependent modulation, scalar channel mixing commutes with graph indexing and aggregation, making reordered and conventional scalar transformations equivalent and allowing scalar transformations to move from edges to nodes.

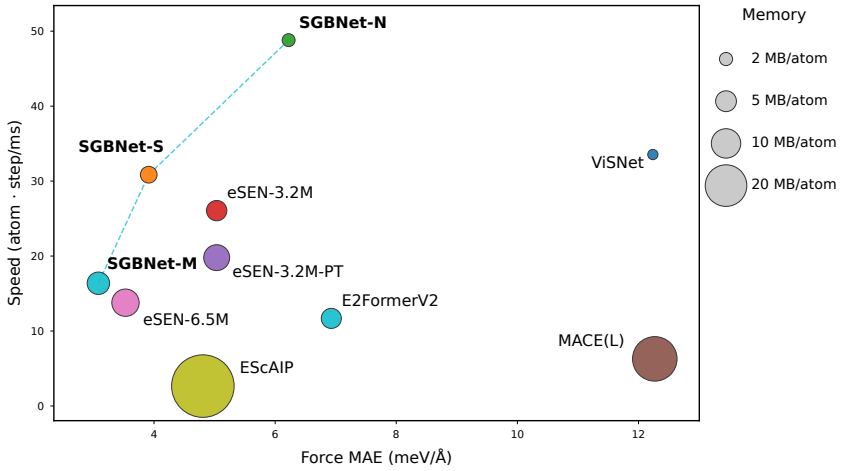

Figure 1: Accuracy-efficiency Pareto frontier. Force MAE is plotted against MD speed for small systems ($N < 4000$), where speed is the reciprocal of per-atom runtime; marker area denotes memory usage. The upper-left direction is preferred.

Experiments show that SGBNet improves the accuracy-efficiency trade-off across molecular and materials benchmarks. As shown in Figure 1, all SGBNet variants lie on the accuracy-efficiency Pareto frontier of the SPICE-MACE-OFF Kovács et al. (2025) dataset. SGBNet-N matches or surpasses eSEN-3.2M on most MLIPAudit Wehrhan et al. (2025) metrics, while reducing runtime and memory by 47% and 60%, respectively. Compared with ViSNet, SGBNet-N achieves substantially lower MAEs with higher efficiency, challenging the conventional accuracy-efficiency trade-off that motivates Euclidean alternatives to higher-order equivariant models. Beyond molecular systems, SGBNet achieves slightly improved performance on materials benchmarks while preserving its efficiency advantages, and is further evaluated under large-scale training regimes.

Our contributions are summarized as follows:

- We develop scalar-geometric bidirectional (SGB) interactions at both the SO(2) and SO(3) levels, enabling explicit geometric-to-scalar feedback through learned second-order invariants, with conjugate multiplication as a key mechanism in the SO(2) branch.
- We propose edge-efficient SGB convolution that combines scalar transformation reordering with feature-channel redistribution, moving scalar transformations from edges to nodes and reducing their complexity from $\mathcal{O}(\mathcal{E})$ to $\mathcal{O}(\mathcal{N})$.
- We provide comprehensive experiments establishing a favorable molecular accuracy-efficiency trade-off across predictive and application-level benchmarks, while demonstrating applicability to materials and large-scale systems.

2 Related work

Machine learning interatomic potentials. Machine learning interatomic potentials (MLIPs) aim to approximate the potential energy surface and typically derive forces as energy gradients, ensuring physical consistency via energy conservation. Early approaches relied on kernel methods Glielmo et al. (2017) and handcrafted descriptors Behler (2011); Bartók et al. (2013), while recent works leverage deep neural networks to improve scalability and expressivity. Representative models such as MACE Batatia et al. (2022), ViSNet Wang et al. (2024), GotenNet Aykent & Xia (2025), DPA3 Zhang et al. (2026), eSEN Fu et al. (2025), and EquiformerV3 Liao et al. (2026) have demonstrated strong performance across diverse molecular and materials datasets.

Equivariant graph neural networks. Equivariant graph neural networks (GNNs) incorporate rotational and translational symmetries, often using SO(3) irreducible representations and tensor operations. Representative approaches include higher-order tensor-product models such as MACE, Euclidean methods such as ViSNet and TACE Xu et al. (2026), and SO(2)-based models such as eSCN Passaro & Zitnick (2023) and eSEN, which replace higher-order tensor products with more efficient SO(2) convolution. Recent work such as QHNetV2 Yu et al. (2026) further explores more general SO(2)-based equivariant transformations in local frames. Our model similarly uses efficient SO(2) and SO(3) operations while retaining higher-order equivariant representations.

Efficient and scalable equivariant models. Recent efforts aim to improve efficiency by reducing the cost of equivariant operations or limiting higher-order features. One direction reduces complexity via $\mathcal{O}(\mathcal{V})$ operations instead of $\mathcal{O}(\mathcal{E})$ operations, e.g., E2Former Li et al. (2025) reduces computational complexity with Wigner 6j convolution. Another direction seeks to reduce or eliminate higher-order features altogether. So3krates Frank et al. (2024) restricts higher-order channels with enhanced scalar interactions, while EScAIP Qu & Krishnapriyan (2024) removes equivariant representations using scalar descriptors such as BOO Steinhardt et al. (1983). Our method combines these directions by moving scalar transformations to the node domain while retaining fewer but more strongly coupled higher-order geometric features.

3 Method

3.1 Overview

We propose SGBNet, an equivariant message-passing architecture with decoupled scalar and geometric pathways and explicit bidirectional interaction between them. We represent the node state as $\boldsymbol{x}_i = h_i \bigoplus_{l \geqslant 1} \boldsymbol{\chi}_i^{(l)}$, where h_i denotes invariant scalar features and $\boldsymbol{\chi}_i^{(l)} \in \mathbb{R}^{2l+1}$ denotes degree-l SO(3)-equivariant geometric features. These geometric features are also referred to as irreducible features or spherical-harmonic features in prior works. SGBNet follows the equivariant message-passing update

$$\boldsymbol{x}_i' = \boldsymbol{x}_i + \sum_{j \in \mathcal{N}(i)} \Phi\left(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{r}_{ij}, z_i, z_j\right), \tag{2}$$

where Φ is SO(3)-equivariant. As shown in Figure 2(a), each layer consists of a edge-efficient SGB convolution block (Figure 2(b)), followed by a SGB feed-forward block (Figure 2(c)). The final prediction is produced by a SGB energy head (Figure 2(d)). Nonlinearity is implemented via a SiLU-based gating mechanism Liao & Smidt (2023), and forces are obtained via automatic differentiation. We likewise decouple normalization Liao et al. (2024), using layer normalization for scalar features and RMS normalization for geometric features to preserve the separation between the two pathways.

First, we introduce scalar-geometric bidirectional (SGB) interactions. Concretely, the SGB SO(2) linear (Section 3.2) and SGB SO(3) linear (Section 3.3) establish explicit geometric-to-scalar mappings through learnable second-order invariant constructions, using conjugate multiplication and contraction, respectively, and inject the resulting scalars before the gating nonlinearity. Second, we design the edge-efficient SGB convolution block (Section 3.4), where scalar transformations are moved from the edge domain to the node domain before graph propagation. Meanwhile, we increase the scalar feature channels while reducing the higher-order geometric feature channels, thereby reallocating computation toward the more efficient scalar pathway.

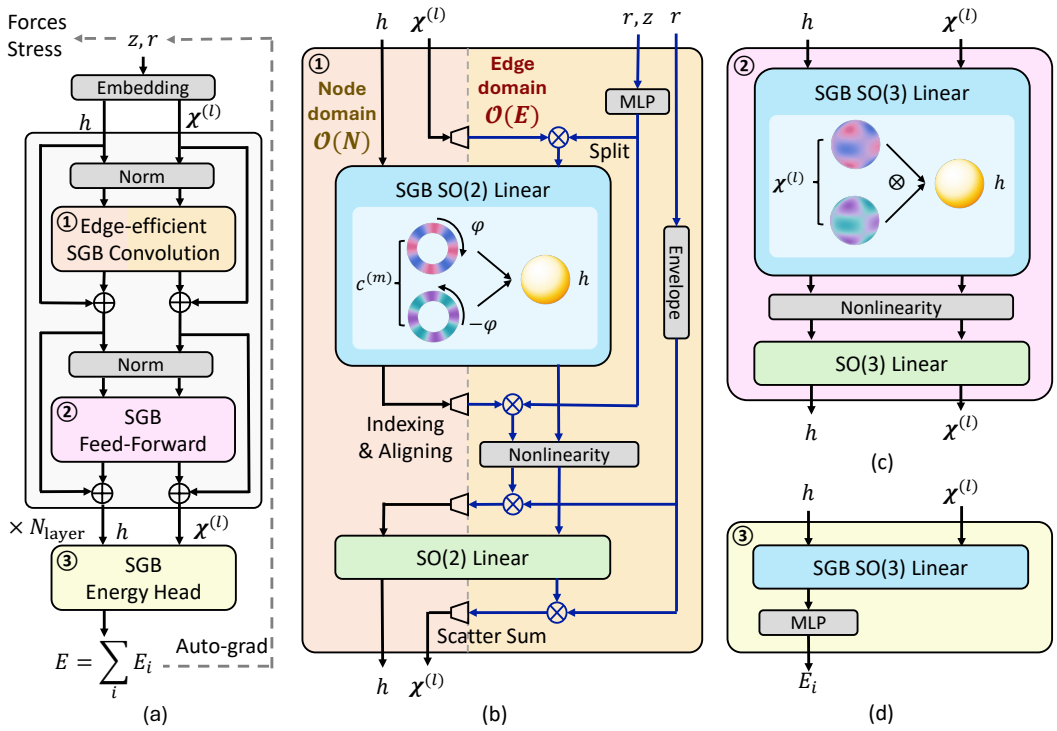

Figure 2: Overview of the proposed SGBNet architecture. (a) Overall model architecture, with separated scalar and higher-order geometric feature pathways. (b) Edge-efficient SGB convolution block. Trapezoids denote graph transition operations: indexing and aligning (edge construction) and scatter sum (edge-to-node aggregation). Blue connections indicate edge-domain operations, whereas black connections indicate node-domain operations. (c) SGB feed-forward block. (d) SGB energy head.

3.2 SGB SO(2) LINEAR

The SGB SO(2) linear extends standard SO(2) convolution by introducing an additional explicit invariant geometric-to-scalar pathway within SO(2)-based message passing.

In the standard SO(2) convolution, geometric features are first indexed and aligned to the y-axis using Wigner D-matrices Worrall et al. (2017) to obtain $\tilde{\boldsymbol{\chi}}^{(l)}$. For each frequency $m > 0$, the aligned m and $-m$ components are transformed jointly by:

$$\begin{pmatrix} \boldsymbol{\chi}'_m \\ \boldsymbol{\chi}'_{-m} \end{pmatrix} = \begin{pmatrix} W_{\mathrm{r}}^{(m)} & -W_{\mathrm{i}}^{(m)} \\ W_{\mathrm{i}}^{(m)} & W_{\mathrm{r}}^{(m)} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{\chi}}_m \\ \tilde{\boldsymbol{\chi}}_{-m} \end{pmatrix}, \tag{3}$$

where $W_{\mathrm{r}}^{(m)}$ and $W_{\mathrm{i}}^{(m)}$ are learnable linear transformations. The $m = 0$ sector is invariant in the aligned edge frame and forms rotation-invariant scalars. This form is a frequency-wise SO(2)-equivariant linear map, hence we refer to this operation as SO(2) linear throughout the paper. To expose the phase structure of the SO(2) representation, we equivalently write the paired components and linear weights in complex form as $c^{(m)} = \tilde{\chi}_m + i\tilde{\chi}_{-m}$ and $A^{(m)} = W_{\mathrm{r}}^{(m)} + iW_{\mathrm{i}}^{(m)}$. The implementation retains the equivalent real-valued representation.

We augment this transformation with an additional branch that maps geometric features to scalars. Specifically, we first apply two SO(2)-equivariant linear maps, using the complex form equivalent to the real-valued formulation above:

$$\tilde{c}^{(m)} = A^{(m)} c^{(m)}, \hat{c}^{(m)} = A'^{(m)} c^{(m)}. \tag{4}$$

Because the two transformed features within the same frequency channel acquire the same rotation-induced phase, conjugate multiplication cancels this phase:

$$h^{(m)} = \mathrm{Re}\left(\tilde{c}^{(m)}\hat{c}^{(m)*}\right), \tag{5}$$

where $(\cdot)^*$ denotes complex conjugation. Although both real and imaginary parts are invariant, we retain only the real part in the implementation.

These scalar features are concatenated and projected, then added to the original projected scalars:

$$h' = W_s h + W_t \mathrm{Concat}\left[h^{(1)}, h^{(2)}, \cdots, h^{(M)}\right]. \tag{6}$$

The resulting scalar features are injected into the scalar pathway before the gating nonlinearity.

Proof sketch of equivariance. After aligning, any global rotation R induces only a residual rotation $R_y(\varphi)$ about the aligned axis, since

$$F(R\hat{\boldsymbol{r}})RF(\hat{\boldsymbol{r}})^{-1} = R_y(\varphi). \tag{7}$$

The restriction of an $\mathrm{SO}(3)$ irreducible representation to this $\mathrm{SO}(2)$ subgroup decomposes into frequency-m components, under which

$$c^{(m)} \mapsto e^{im\varphi}c^{(m)}. \tag{8}$$

Consequently, any $\mathrm{SO}(2)$-equivariant linear map, including the SO(2) linear, acts independently within each m sector and preserves this transformation law. Hence both $\tilde{c}^{(m)}$ and $\hat{c}^{(m)}$ acquire the same phase, so that

$$\tilde{c}^{(m)}\hat{c}^{(m)*} \mapsto \tilde{c}^{(m)}\hat{c}^{(m)*}. \tag{9}$$

Its real and imaginary parts are therefore invariant under arbitrary global rotations. A complete derivation is given in Appendix A.

3.3 SGB SO(3) LINEAR AND NODEWISE BLOCKS

The SGB blocks follow the structures shown in Figure 2(c). Given node features, we first apply an SGB SO(3) linear, followed by the SiLU-based gating nonlinearity, and then a standard SO(3) linear. In the SGB energy head shown in Figure 2(d), we use an SGB SO(3) linear and only take the scalar part to feed into a MLP for energy prediction. This allows geometric information to directly influence the predicted energy without maintaining geometric channels in the head.

The SGB SO(3) linear introduces geometric-to-scalar interaction through second-order invariants. Specifically, we apply two independent SO(3)-equivariant linear transformations:

$$\tilde{\boldsymbol{\chi}}^{(l)} = W^{(l)}\boldsymbol{\chi}^{(l)}, \hat{\boldsymbol{\chi}}^{(l)} = W'^{(l)}\boldsymbol{\chi}^{(l)}, \tag{10}$$

and form a scalar via contraction:

$$h^{(l)} = \frac{1}{\sqrt{2l+1}} \sum_{m=-l}^{l} \tilde{\chi}_m^{(l)}\hat{\chi}_m^{(l)}. \tag{11}$$

This corresponds to the $l \otimes l \to 0$ tensor product, producing an SO(3)-invariant scalar. The resulting scalars are concatenated and projected:

$$h' = W_s h + W_t \mathrm{Concat}\left[h^{(1)}, h^{(2)}, \cdots, h^{(L)}\right], \tag{12}$$

and injected into the scalar pathway before the gating nonlinearity.

3.4 EDGE-EFFICIENT SGB CONVOLUTION

As shown in Figure 2(b), higher-order geometric features are first indexed and aligned as described above, and then modulated by edge-dependent inputs derived from interatomic distance r and atomic numbers before entering the SGB SO(2) linear. Scalar features follow a different ordering, with channel mixing performed in the node domain before edge indexing and edge-dependent modulation

Table 1: Per-atom computational efficiency in MD simulations. Time, memory, and FLOPs are reported in μs/atom/step, MB/atom, and MFLOP/atom, respectively. OOM indicates GPU memory exhaustion. *PyTorch implementations. FLOPs and memory are obtained using framework-specific profilers. The best and second-best results are highlighted in bold and underlined, respectively.

	MACE (L)*	ViSNet	EScAIP*	E2Former V2*	eSEN -3.2M	eSEN -3.2M-PT*	eSEN -6.5M	SGBNet -N	SGBNet -S	SGBNet -M
Small systems ($N < 4000$)										
Time	159.10	<u>29.81</u>	376.58	85.59	38.37	50.53	72.57	**20.49**	32.41	61.07
Memory	22.923	**1.212**	45.192	4.645	4.781	7.765	8.628	<u>1.932</u>	3.228	5.813
FLOPs	3075.0	**68.7**	25579.0	478.1	620.7	915.6	1222.5	<u>139.9</u>	390.3	786.5
Large systems ($N \geqslant 4000$)										
Time	OOM	<u>20.54</u>	OOM	59.92	26.81	32.93	52.12	**14.45**	22.83	43.16
Memory	OOM	**1.202**	OOM	4.512	3.597	6.976	6.502	<u>1.428</u>	2.420	4.379
FLOPs	OOM	**53.9**	OOM	478.1	462.0	681.4	909.8	<u>104.8</u>	293.2	590.3

Table 2: MAE of energy (meV/atom) and force (meV/Å) on the SPICE-MACE-OFF dataset. The best and second-best results are highlighted in bold and underlined, respectively.

		MACE(L)	ViSNet	EScAIP	E2Former V2	eSEN -3.2M	eSEN -6.5M	SGBNet -N	SGBNet -S	SGBNet -M
PubChem	E	0.88	0.50	0.53	0.69	0.22	**0.15**	0.30	0.22	<u>0.20</u>
	F	14.75	15.52	5.86	9.00	6.10	<u>4.21</u>	7.91	5.02	**3.96**
DES370K Monomers	E	0.59	0.35	0.41	0.52	<u>0.17</u>	**0.13**	0.22	0.18	0.18
	F	6.58	5.15	3.48	4.00	1.85	<u>1.24</u>	2.55	1.52	**1.20**
DES370K Dimers	E	0.54	0.34	0.38	0.30	0.20	<u>0.15</u>	0.21	0.16	**0.14**
	F	6.62	4.74	2.18	2.09	2.77	2.12	2.43	<u>1.44</u>	**1.13**
Dipeptides	E	0.42	0.28	0.31	0.37	0.10	**0.07**	0.15	0.10	<u>0.08</u>
	F	10.19	9.74	5.21	5.25	3.04	<u>2.00</u>	4.47	2.58	**1.92**
Solvated AA	E	0.98	0.60	0.61	0.56	0.30	0.25	0.33	<u>0.19</u>	**0.14**
	F	19.43	18.61	11.52	12.66	5.76	**3.68**	8.42	5.05	**3.68**
Water	E	0.83	0.64	0.72	0.70	0.24	**0.15**	0.30	<u>0.20</u>	0.22
	F	13.57	11.84	10.31	7.16	3.88	<u>2.50</u>	5.56	3.40	**2.48**
QMugs	E	0.45	0.44	0.41	0.66	0.16	**0.12**	0.22	0.14	<u>0.13</u>
	F	16.93	18.39	8.74	11.21	5.70	<u>3.78</u>	7.69	4.65	**3.46**

applied afterward. A similar reordering is applied to the subsequent SO(2) linear. Geometric features remain on edges and are modulated by the envelope function Gasteiger et al. (2021) before inverse alignment and scatter summation, whereas scalar features are first aggregated and then transformed in the node domain. In addition, we redistribute feature channels by increasing scalar channels and reducing higher-order geometric channels, further shifting computation toward the more efficient scalar pathway. We therefore refer to this block as the edge-efficient SGB convolution.

This reordering is enabled by the scalar transformation structure. For scalars, alignment is the identity because $D^{(0)}(R_{ij}) = 1$. Moreover, graph indexing and aggregation act on the node/edge dimension, whereas channel mixing acts on the feature dimension. Thus, for $H \in \mathbb{R}^{N\times C}$ and a channel transformation W,

$$WD^{(0)}(R_{ij})\mathcal{I}(H) = \mathcal{I}(WH), \mathcal{S}(D^{(0)}(R_{ij})^{-1}WX) = W\mathcal{S}(H), \tag{13}$$

where $\mathcal{I}$ and $\mathcal{S}$ denote node-to-edge indexing and edge-to-node scatter summation, respectively. These identities show that the conventional and reordered scalar transformations are exactly equivalent in the absence of channel-dependent edge modulation. Edge-dependent modulation M_{ij} breaks this exact equivalence because it does not generally commute with W. Nevertheless, we find empirically that the resulting accuracy degradation is modest, and therefore adopt the reordered formulation to move expensive scalar channel transformations from edges to nodes while retaining edge-dependent modulation and higher-order geometric operations on edges.

4 EXPERIMENTS

4.1 OVERVIEW AND SETUP

We conduct evaluations on both computational efficiency and predictive performance, covering molecular and materials benchmarks with complementary accuracy and practical assessments. Efficiency benchmarks assess the computational advantages of the proposed architecture, including runtime scaling and resource consumption (Section 4.2). For molecular systems, we evaluate predictive accuracy on the SPICE-MACE-OFF dataset (Section 4.3) and further assess practical performance through MLIPAudit (Section 4.4). For materials systems, we similarly evaluate predictive accuracy on the sAlex dataset and practical performance on Matbench Discovery (Section 4.5). Finally, we investigate the scalability of SGBNet under large-scale training regimes using the OMat24 dataset (Section 4.6).

SGBNet is implemented in JAX based on the MLIP codebase Brunken et al. (2025; 2026), whose JAX implementations of eSEN and ViSNet are also used as the primary baselines. All JAX-based models are evaluated with JAX-MD Schoenholz & Cubuk (2020) under the same simulation setup. For fair cross-framework comparison, we additionally benchmark the eSEN-3.2M PyTorch implementation to quantify the impact of implementation frameworks. PyTorch baselines use TF32 and compilation whenever compatible; E2FormerV2 retains its native Triton kernels. All efficiency benchmarks are conducted on a single NVIDIA A100 GPU.

4.2 COMPUTATIONAL EFFICIENCY EVALUATION

Table 1 reports per-atom scaling coefficients from five PDB systems ranging from 138 to 11658 atoms, grouped into small ($N < 4000$) and large ($N \geqslant 4000$) regimes. For small systems, SGBNet-N reduces runtime and memory by 47% and 60% relative to eSEN-3.2M, and reduces runtime by 31% relative to ViSNet. SGBNet-S reduces runtime, memory, and FLOPs by 16%, 33%, and 37% relative to eSEN-3.2M. These architectural gains remain nearly unchanged for large systems, where SGBNet-N reduces runtime by 46% relative to eSEN-3.2M and by 30% relative to ViSNet.

4.3 MOLECULAR ACCURACY ON SPICE-MACE-OFF

SPICE-MACE-OFF Kovács et al. (2025), derived from SPICE Eastman et al. (2023), covers diverse molecular environments ranging from small molecules and dimers to solvated systems and larger molecules. As shown in Table 2, SGBNet provides particularly strong improvements in force prediction. Compared with the similarly sized eSEN-3.2M, SGBNet-S achieves lower force MAE on *all seven* subsets, while also improving energy MAE on four subsets and matching it on two. Both SGBNet variants also substantially outperform the earlier MACE and E2FormerV2 Huang et al. (2026) baselines across all reported subsets. These accuracy gains are achieved simultaneously with the efficiency improvements reported in Table 1, where SGBNet-S is approximately 16% faster, and uses 33% less memory than eSEN-3.2M for small systems. At larger capacity, SGBNet-M outperforms eSEN-6.5M in force MAE on six subsets and ties on the remaining one, although its energy MAE is not uniformly lower. For the Pareto analysis in Figure 1, force MAEs are averaged across the seven subsets with weights proportional to their test-set sizes.

4.4 MOLECULAR PRACTICAL EVALUATION ON MLIPAUDIT

Low test-set energy and force errors do not necessarily translate into physically faithful molecular simulations Maheshwari et al. (2025); Bigi et al. (2025). MLIPAudit Wehrhan et al. (2025) therefore complements conventional MAEs with physically motivated evaluations of structural stability, conformational energetics, intermolecular interactions, reactive processes, liquid structure, and biomolecular sampling.

Following the MLIPAudit protocol, all models are trained on the neutral-molecule subset of SPICE2-curated-v2 Brunken et al. (2026), derived from OMol25 Levine et al. (2025). All baselines are retrained on the same dataset using the same loss function and training protocol, while retaining their original model hyperparameters.

As shown in Table 3, SGBNet-S achieves the best or tied-best performance on 10 MLIPAudit metrics. More importantly, the compact SGBNet-N remains comparable to eSEN-3.2M across MLIPAudit metrics, performing better on six, worse on three, and tying on four. Thus, despite its lower conventional SPICE accuracy, SGBNet-N largely preserves the practically relevant molecular behavior of eSEN-3.2M at substantially lower computational cost. For small systems, SGBNet-N reduces runtime and memory by 47% and 60% relative to eSEN-3.2M. Compared with the Euclidean-based ViSNet, it also reduces runtime by 31% while achieving substantially lower force MAE across all SPICE subsets. SGBNet-N therefore provides a particularly favorable practical operating point, showing that explicit higher-order tensor representations can retain strong simulation performance while exceeding a Euclidean model in both force accuracy and computational speed.

Table 3: MLIPAudit benchmark scores, grouped by category. All scores are normalized from underlying error metrics (e.g., MAE, RMSD) to the range (0, 1], where higher is better. A score of 1.0 indicates performance comparable to the reference DFT calculations on the corresponding task. The best and second-best results are highlighted in bold and underlined, respectively.

		ViSNet	eSEN-3.2M	SGBNet-N	SGBNet-S
Small Molecules	Bond length distribution	**1.000**	**1.000**	**1.000**	**1.000**
	Ring planarity	**1.000**	**1.000**	**1.000**	**1.000**
	Reference geometry stability	**1.000**	**1.000**	**1.000**	**1.000**
	Conformer selection	0.944	<u>0.968</u>	0.967	**0.971**
	Dihedral scan	0.711	0.709	<u>0.713</u>	**0.719**
	Noncovalent interactions	0.572	0.666	<u>0.694</u>	**0.709**
	Nudged elastic band	0.110	**0.160**	0.090	<u>0.120</u>
	Reactivity	0.259	0.301	<u>0.312</u>	**0.325**
	Tautomers	0.105	<u>0.200</u>	0.193	**0.214**
Molecular Liquids	Water radial distribution	**1.000**	0.995	**1.000**	**1.000**
	Solvent radial distribution	0.967	0.956	<u>0.984</u>	**0.988**
Biomolecules	Folding stability	**0.530**	<u>0.529</u>	<u>0.529</u>	0.528
	Sampling	0.718	<u>0.752</u>	**0.786**	0.737

4.5 Materials evaluation on Matbench Discovery and sAlex

Matbench Discovery Riebesell et al. (2025) provides a comprehensive benchmark for evaluating MLIPs on materials modeling tasks. Models are trained using the MPtrj Jain et al. (2013); Deng et al. (2023) and sAlex Barros-Luque et al. (2026) datasets combined with an 8:1 ratio. The trained models are evaluated on selected tasks from Matbench Discovery, including crystal stability prediction and geometry optimization, together with energy, force, and stress prediction performance on the sAlex validation set.

As shown in Table 4, SGBNet-S achieves slightly better overall performance than eSEN-3.2M on the Matbench Discovery and sAlex evaluation tasks, matching or exceeding eSEN-3.2M on the majority of reported metrics. Meanwhile, SGBNet-S preserves the computational advantages observed in molecular benchmarks, reducing runtime, memory usage, and FLOPs under the same implementation and evaluation protocol, consistent with the efficiency analysis in Section 4.2.

Table 4: Evaluation results on Matbench Discovery and sAlex prediction tasks. The Matbench Discovery results are reported on the unique prototype split. Energy, force, and stress errors are reported in meV/atom, meV/Å, and meV/Å^3, respectively. (↑/↓) indicates whether higher or lower values are preferred.

Models	F1↑	DAF↑	Prec↑	Acc↑	MAE↓	R2↑	RMSD↓	MAE_E	MAE_F	MAE_S
MACE-MPA-0	0.852	5.582	0.853	0.954	0.028	0.842	0.0731	-	-	-
ORB-v2-MPA	0.880	**6.041**	**0.924**	0.965	0.028	0.824	0.0973	-	-	-
eSEN-3.2M-MPA	**0.895**	5.822	0.890	<u>0.967</u>	**0.023**	**0.855**	**0.0681**	13.2	31.7	1.82
SGBNet-S-MPA	**0.895**	<u>5.838</u>	<u>0.893</u>	**0.968**	**0.023**	<u>0.849</u>	<u>0.0687</u>	13.0	30.4	1.81

4.6 LARGE-SCALE SCALING ON OMAT24

OMat24 Barros-Luque et al. (2026) is a large-scale dataset comprising over 110 million DFT calculations. We use OMat24 to evaluate the scalability of SGBNet under large-data and large-model regimes. As shown in Table 5, the results demonstrate the feasibility of applying the proposed architecture at this scale, while the performance difference compared with large-scale pretrained models indicates potential opportunities for further scaling optimization. Meanwhile, SGBNet-L reduces runtime and memory consumption by 11% and 10% compared with eSEN-30M, respectively.

Table 5: MAE on the OMat24 validation set. The best and second-best results are highlighted in bold and underlined, respectively.

	EquiformerV2-S	eSEN-30M	UMA-S	SGBNet-L
Energy MAE ($\mathrm{meV/atom}$)	11.0	**10.7**	11.3	<u>10.8</u>
Force MAE ($\mathrm{meV/Å}$)	49.2	**47.3**	57.1	<u>48.4</u>
Stress MAE ($\mathrm{meV/Å^3}$)	**2.4**	<u>2.6</u>	2.9	2.7

4.7 ABLATION STUDIES

Table 6: Ablation study on SGB interaction and edge-efficient design. Results are evaluated on the SPICE2 validation set.

	Compact	SGB interaction			Edge-efficient design	
		-SO2	-Head	-FFN	-Reorder	-Realloc
$\mathrm{MAE_F}$	3.75	4.10	5.73	6.90	3.73	3.40
Time	32.07	28.69	24.20	23.89	33.69	45.98
Memory	3.225	2.957	2.537	2.529	3.264	4.985

All ablations start from the same Compact model, with different groups representing independent analysis paths. The SGB interaction ablation progressively removes SGB components from different blocks, including the SGB SO(2) linear and the SGB SO(3) linear in the energy head and feed-forward block. Each removal leads to degraded accuracy, demonstrating the contribution of SGB-based geometric-to-scalar interactions. The larger degradation caused by removing SGB SO(3) linear is because it directly links geometric features to the final prediction. Removing it forces part of the geometric information back into the low-rank readout coupling pathway. The edge-efficient design ablation evaluates the effects of computation reordering and channel reallocation. The reordering strategy reduces computational cost with negligible impact on accuracy, while channel reallocation further improves efficiency at the cost of a controllable accuracy trade-off.

5 DISCUSSION

Our results show that improving scalar-geometric interaction and reallocating computation from edgewise geometric operations to nodewise scalar transformations can substantially improve the accuracy-efficiency trade-off of equivariant MLIPs. The gains are particularly pronounced for compact models, where SGBNet achieves strong predictive performance while substantially reducing computational overhead.

However, these advantages become less pronounced as model and data scales increase. This may result from diminishing architecture-dependent gains under scaling, while optimization challenges in larger models may also affect the achievable performance.

Long-time stability under rare or out-of-distribution configurations remains an open challenge for MLIPs. Although SGBNet demonstrates robust performance across most evaluated systems, SGBNet-S encounters similar stability issues on a small number of challenging configurations. Improving robustness in such regimes may require complementary stability-aware training strategies, such as incorporating additional stability-focused data or objectives Raja et al. (2025).

AI USE STATEMENT

In this work, we used generative AI tools to aid and polish the writing of the manuscript. Specifically, generative AI was used to revise author-written text for clarity, grammar, and presentation, and to check the revised text for consistency and correctness. We did not use generative AI tools for literature retrieval or discovery, research ideation, experimental design or execution, scientific analysis, interpretation of results, or the implementation of models and experiments. All AI-assisted content was carefully reviewed and verified by the authors for correctness. We take full responsibility for the final content of this work, including all text, claims, results, code, and artifacts produced with the aid of generative AI.

REFERENCES

Sarp Aykent and Tian Xia. GotenNet: Rethinking Efficient 3D Equivariant Graph Neural Networks. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=5wxCQDtbMo.

Luis Barros-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M. Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, Matt Uyttendaele, C. Lawrence Zitnick, and Zachary W. Ulissi. The Open Materials 2024 (OMat24) inorganic materials dataset and models. *Nature Computational Science*, June 2026. ISSN 2662-8457. doi: 10.1038/s43588-026-00996-w. URL https://doi.org/10.1038/s43588-026-00996-w.

Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Phys. Rev. B*, 87:184115, May 2013. doi: 10.1103/PhysRevB.87.184115. URL https://link.aps.org/doi/10.1103/PhysRevB.87.184115.

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/4a36c3c51af11ed9f34615b81edb5bbc-Paper-Conference.pdf.

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, May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5. URL https://doi.org/10.1038/s41467-022-29939-5.

Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of Chemical Physics*, 134(7):074106, 02 2011. ISSN 0021-9606. doi: 10.1063/1.3553717. URL https://doi.org/10.1063/1.3553717.

Filippo Bigi, Marcel F. Langer, and Michele Ceriotti. The dark side of the forces: assessing non-conservative force models for atomistic machine learning. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 4384–4414. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/bigi25a.html.

Christoph Brunken, Olivier Peltre, Heloise Chomet, Lucien Walewski, Manus McAuliffe, Valentin Heyraud, Solal Attias, Martin Maarand, Yessine Khanfir, Edan Toledo, Fabio Falcioni, Marie Bluntzer, Silvia Acosta-Gutiérrez, and Jules Tilly. Machine learning interatomic potentials: library for efficient training, model development and simulation of molecular systems, 2025. URL https://arxiv.org/abs/2505.22397.

Christoph Brunken, Titouan Cormier, Lucien Walewski, Marco Carobene, Yessine Khanfir, Zachary Weller-Davies, Miguel Bragança, Armand Picard, Adrien Pichard, Leon Wehrhan, Heloise Chomet, Eszter Varga-Umbrich, Marie Bluntzer, Massimo Bortone, Valentin Heyraud, Silvia

Acosta-Gutiérrez, Jules Tilly, and Olivier Peltre. Machine learning interatomic potentials: Advancing open-source software for efficient and scalable molecular simulation, 2026. URL https://arxiv.org/abs/2605.22698.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240):1–113, 2023. URL http://jmlr.org/papers/v24/22-1144.html.

Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995. doi: 10.1021/ja00124a002. URL https://doi.org/10.1021/ja00124a002.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, September 2023. ISSN 2522-5839. doi: 10.1038/s42256-023-00716-3. URL https://doi.org/10.1038/s42256-023-00716-3.

Volker L. Deringer, Miguel A. Caro, and Gábor Csányi. Machine learning interatomic potentials as emerging tools for materials science. *Advanced Materials*, 31(46):1902765, 2019. doi: https://doi.org/10.1002/adma.201902765. URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/adma.201902765.

Peter Eastman, Pavan Kumar Behara, David L. Dotson, Raimondas Galvelis, John E. Herr, Josh T. Horton, Yuezhi Mao, John D. Chodera, Benjamin P. Pritchard, Yuanqing Wang, Gianni De Fabritiis, and Thomas E. Markland. SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials. *Scientific Data*, 10(1):11, January 2023. ISSN 2052-4463. doi: 10.1038/s41597-022-01882-6. URL https://doi.org/10.1038/s41597-022-01882-6.

R.P. Feynman, R.B. Leighton, and M.L. Sands. *The Feynman Lectures on Physics, Vol. III: The New Millennium Edition: Quantum Mechanics*. The Feynman Lectures on Physics. Basic Books, 2011. ISBN 9780465025015.

J. Thorben Frank, Oliver T. Unke, Klaus-Robert Müller, and Stefan Chmiela. A Euclidean transformer for fast and stable machine learned force fields. *Nature Communications*, 15(1):6539, August 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-50620-6. URL https://doi.org/10.1038/s41467-024-50620-6.

Xiang Fu, Brandon M Wood, Luis Barroso-Luque, Daniel S. Levine, Meng Gao, Misko Dzamba, and C. Lawrence Zitnick. Learning smooth and expressive interatomic potentials for physical property prediction. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 17875–17893. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/fu25h.html.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. In M. Ranzato, A. Beygelzimer,

Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 6790–6802. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/35cf8659cfcb13224cbd47863a34fc58-Paper.pdf.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks, 2022. URL https://arxiv.org/abs/2207.09453.

Aldo Glielmo, Peter Sollich, and Alessandro De Vita. Accurate interatomic force fields via machine learning with covariant kernels. *Phys. Rev. B*, 95:214302, Jun 2017. doi: 10.1103/PhysRevB.95.214302. URL https://link.aps.org/doi/10.1103/PhysRevB.95.214302.

Lin Huang, Chengxiang Huang, Ziang Wang, Yiyue Du, Chu Wang, Haocheng Lu, Yunyang Li, Xiaoli LIU, Arthur JIANG, and Jia Zhang. E2former-v2: On-the-fly equivariant attention with linear activation memory. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=gIUVdEAa3l.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 07 2013. ISSN 2166-532X. doi: 10.1063/1.4812323. URL https://doi.org/10.1063/1.4812323.

Dávid Péter Kovács, J. Harry Moore, Nicholas J. Browning, Ilyes Batatia, Joshua T. Horton, Yixuan Pu, Venkat Kapil, William C. Witt, Ioan-Bogdan Magdău, Daniel J. Cole, and Gábor Csányi. MACE-OFF: Short-Range Transferable Machine Learning Force Fields for Organic Molecules. *Journal of the American Chemical Society*, 147(21):17598–17611, May 2025. ISSN 0002-7863. doi: 10.1021/jacs.4c07099. URL https://doi.org/10.1021/jacs.4c07099.

Daniel S. Levine, Muhammed Shuaibi, Evan Walter Clark Spotte-Smith, Michael G. Taylor, Muhammad R. Hasyim, Kyle Michel, Ilyes Batatia, Gábor Csányi, Misko Dzamba, Peter Eastman, Nathan C. Frey, Xiang Fu, Vahe Gharakhanyan, Aditi S. Krishnapriyan, Joshua A. Rackers, Sanjeev Raja, Ammar Rizvi, Andrew S. Rosen, Zachary Ulissi, Santiago Vargas, C. Lawrence Zitnick, Samuel M. Blau, and Brandon M. Wood. The open molecules 2025 (omol25) dataset, evaluations, and models, 2025. URL https://arxiv.org/abs/2505.08762.

Yunyang Li, Lin Huang, Zhihao Ding, Xinran Wei, Chu Wang, Han Yang, Zun Wang, Chang Liu, Yu Shi, Peiran Jin, Tao Qin, Mark Gerstein, and Jia Zhang. E2former: An efficient and equivariant transformer with linear-scaling tensor products. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 23593–23630. Curran Associates, Inc., 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/21f7b745f73ce0d1f9bcea7f40b1388e-Paper-Conference.pdf.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=KwmPfARgOTD.

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 (ICLR)*, 2024. URL https://openreview.net/forum?id=mCOBKZmrzD.

Yi-Lun Liao, Alexander J. Hoffman, Sabrina C. Shen, Alexandre Duval, Sam Walton Norwood, and Tess Smidt. Equiformerv3: Scaling efficient, expressive, and general se(3)-equivariant graph attention transformers, 2026. URL https://arxiv.org/abs/2604.09130.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Shagun Maheshwari, Zhengxian Tang, Janghoon Ock, Adeesh Kolluru, Amir Barati Farimani, and John R. Kitchin. Beyond force metrics: Pre-training mlffs for stable md simulations, 2025. URL https://arxiv.org/abs/2506.14850.

Dominik Marx and Jürg Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

Saro Passaro and C. Lawrence Zitnick. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27420–27438. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/passaro23a.html.

Eric Qu and Aditi S. Krishnapriyan. The importance of being scalable: Improving the speed and accuracy of neural network interatomic potentials across chemical domains. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 139030–139053. Curran Associates, Inc., 2024. doi: 10.52202/079017-4412. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/fad8e1915f66161581bb127ccf01092e-Paper-Conference.pdf.

Sanjeev Raja, Ishan Amin, Fabian Pedregosa, and Aditi S. Krishnapriyan. Stability-aware training of machine learning force fields with differentiable boltzmann estimators. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=ZckLMG00sO.

Janosh Riebesell, Rhys E. A. Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Gerbrand Ceder, Mark Asta, Alpha A. Lee, Anubhav Jain, and Kristin A. Persson. A framework to evaluate machine learning crystal stability predictions. *Nature Machine Intelligence*, 7(6):836–847, June 2025. ISSN 2522-5839. doi: 10.1038/s42256-025-01055-1. URL https://doi.org/10.1038/s42256-025-01055-1.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Samuel Schoenholz and Ekin Dogus Cubuk. Jax md: A framework for differentiable physics. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 11428–11441. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/83d3d4b6c9579515e1679aca8cbc8033-Paper.pdf.

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 I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/303ed4c69846ab36c2904d3ba8573050-Paper.pdf.

Ken Shoemake. Animating rotation with quaternion curves. 19(3):245–254, July 1985. ISSN 0097-8930. doi: 10.1145/325165.325242. URL https://doi.org/10.1145/325165.325242.

Muhammed Shuaibi, Abhishek Das, Anuroop Sriram, Misko, Luis Barroso-Luque, Ray Gao, Siddharth Goyal, Zachary Ulissi, Brandon Wood, Tian Xie, Junwoong Yoon, Brook Wander, Adeesh Kolluru, Richard Barnes, Ethan Sunshine, Kevin Tran, Xiang, Daniel Levine, Nima Shoghi, Ilias Chair, Janice Lan, Kaylee Tian, Joseph Musielewicz, Weihua Hu, and Kyle Michel. FAIRChem. URL https://github.com/facebookresearch/fairchem.

Paul J. Steinhardt, David R. Nelson, and Marco Ronchetti. Bond-orientational order in liquids and glasses. *Phys. Rev. B*, 28:784–805, Jul 1983. doi: 10.1103/PhysRevB.28.784. URL https://link.aps.org/doi/10.1103/PhysRevB.28.784.

Yusong Wang, Tong Wang, Shaoning Li, Xinheng He, Mingyu Li, Zun Wang, Nanning Zheng, Bin Shao, and Tie-Yan Liu. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nature Communications*, 15(1):313, January 2024. ISSN 2041-1723. doi: 10.1038/s41467-023-43720-2. URL https://doi.org/10.1038/s41467-023-43720-2.

Leon Wehrhan, Lucien Walewski, Marie Bluntzer, Heloise Chomet, Jules Tilly, Christoph Brunken, and Silvia Acosta-Gutiérrez. Mlipaudit: A benchmarking tool for machine learned interatomic potentials, 2025. URL https://arxiv.org/abs/2511.20487.

Brandon Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John Kitchin, Daniel Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Sushree Sahoo, Ammar Rizvi, Zachary Ulissi, and Larry Zitnick. Uma: A family of universal models for atoms. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 129391–129427. Curran Associates, Inc., 2025. doi: 10.52202/085713-4310. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/bbf23e81b0ad7637fe9a731d0b676ca6-Paper-Conference.pdf.

Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

Zemin Xu, Wenbo Xie, and P. Hu. Spectral/spatial tensor atomic cluster expansion with universal embeddings in cartesian space, 2026. URL https://arxiv.org/abs/2509.14961.

Haiyang Yu, Yuchao Lin, Xuan Zhang, Xiaofeng Qian, and Shuiwang Ji. Efficient prediction of SO(3)-equivariant hamiltonian matrices via SO(2) local frames. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=f9i6VLHT5v.

Duo Zhang, Anyang Peng, Chun Cai, Wentao Li, Yuanchang Zhou, Jinzhe Zeng, Mingyu Guo, Chengqian Zhang, Bowen Li, Hong Jiang, Tong Zhu, Weile Jia, Linfeng Zhang, and Han Wang. A graph neural network for the era of large atomistic models. *npj Computational Materials*, 12(1):276, May 2026. ISSN 2057-3960. doi: 10.1038/s41524-026-02146-2. URL https://doi.org/10.1038/s41524-026-02146-2.

A PROOF OF EQUIVARIANCE

We first provide an equivariance proof of the SO(2) convolution used in our formulation, and then verify the equivariance of the proposed invariant branch. We will show that SO(2) convolution is exactly an SO(2)-equivariant linear map.

Given an edge direction $\hat{\boldsymbol{r}} \in \mathbb{R}^3$, SO(2) convolution constructs a rotation matrix

$$F(\hat{\boldsymbol{r}}) \in \mathbb{R}^{3\times 3} \tag{14}$$

such that

$$F(\hat{\boldsymbol{r}})\hat{\boldsymbol{r}} = \hat{\boldsymbol{e}}_y, \tag{15}$$

where $\hat{\boldsymbol{e}}_y$ is the unit vector along the y-axis.

Let $\boldsymbol{\chi}^{(l)} \in \mathbb{R}^{2l+1}$ denote features transforming under the irreducible representation of degree l. Then, the aligned features are

$$\tilde{\boldsymbol{\chi}}^{(l)} := D^{(l)}\left(F(\hat{\boldsymbol{r}})\right)\boldsymbol{\chi}^{(l)}, \tag{16}$$

where $D^{(l)}(R)$ is the Wigner D-matrix.

Consider the action of any global rotation R, under which

$$\hat{\boldsymbol{r}} \mapsto R\hat{\boldsymbol{r}}, \boldsymbol{\chi}^{(l)} \mapsto D^{(l)}(R)\boldsymbol{\chi}^{(l)}. \tag{17}$$

Define

$$Q := F(R\hat{\boldsymbol{r}})RF(\hat{\boldsymbol{r}})^{-1}. \tag{18}$$

By construction,

$$Q\hat{\boldsymbol{e}}_y = F(R\hat{\boldsymbol{r}})R\hat{\boldsymbol{r}} = \hat{\boldsymbol{e}}_y, \tag{19}$$

which implies that Q belongs to the stabilizer subgroup of $\hat{\boldsymbol{e}}_y$, i.e., the subgroup of $\mathrm{SO}(3)$ that leaves $\hat{\boldsymbol{e}}_y$ invariant. Therefore, there exists $\varphi \in [0, 2\pi)$ such that

$$Q = R_y(\varphi). \tag{20}$$

This yields the decomposition

$$R = F(R\hat{\boldsymbol{r}})^{-1}R_y(\varphi)F(\hat{\boldsymbol{r}}). \tag{21}$$

The aligned features are then transformed by

$$\tilde{\boldsymbol{\chi}}^{(l)} \mapsto D^{(l)}\left(F(R\hat{\boldsymbol{r}})\right)D^{(l)}(R)\boldsymbol{\chi}^{(l)} = D^{(l)}\left(R_y(\varphi)\right)\tilde{\boldsymbol{\chi}}^{(l)}. \tag{22}$$

That is, the induced transformation on the aligned features is governed by a one-parameter subgroup of SO(3) corresponding to rotations about the y-axis.

Lemma 1 (SO(2) equivariance)**.** *Let $\tilde{\boldsymbol{\chi}}^{(l)}$ denote the aligned features as defined above. For each $m > 0$, define the complex-valued feature*

$$c^{(m)} = \tilde{\chi}^{(l)}_m + i\tilde{\chi}^{(l)}_{-m}. \tag{23}$$

Under any global rotation R, the complex feature $c^{(m)}$ transforms as

$$c^{(m)} \mapsto e^{im\varphi}c^{(m)}, \tag{24}$$

where φ is defined by the decomposition in Equation 21. We refer to $c^{(m)}$ as an SO(2)-equivariant feature.

Proof. From Equation 22, the transformation of aligned features is given by

$$\tilde{\boldsymbol{\chi}}^{(l)} \mapsto D^{(l)}(R_y(\varphi))\tilde{\boldsymbol{\chi}}^{(l)}, \tag{25}$$

The Wigner D-matrix of $R_y(\varphi)$ in the complex spherical harmonic basis satisfies Feynman et al. (2011)

$$D^{(l)}(R_y(\varphi))_{mm'} = \delta_{mm'}e^{im\varphi}, \tag{26}$$

which yields

$$\psi_m^{(l)} \mapsto e^{im\varphi}\psi_m^{(l)}, \tag{27}$$

where $\boldsymbol{\psi}^{(l)} \in \mathbb{C}^{2l+1}$ is in a complex basis.

The relation between the real coefficients $\tilde{\boldsymbol{\chi}}^{(l)}$ and the complex coefficients $\boldsymbol{\psi}^{(l)}$ is given by Geiger & Smidt (2022)

$$\begin{cases} \psi_m^{(l)} = \dfrac{1}{\sqrt{2}}\left(\tilde{\chi}_m^{(l)} + i\tilde{\chi}_{-m}^{(l)}\right), \\ \psi_{-m}^{(l)} = \dfrac{(-1)^m}{\sqrt{2}}\left(\tilde{\chi}_m^{(l)} - i\tilde{\chi}_{-m}^{(l)}\right). \end{cases} \tag{28}$$

With the definition of $c^{(m)}$, we have

$$c^{(m)} = \sqrt{2}\psi_m^{(l)}, \tag{29}$$

which immediately obtain

$$c^{(m)} \mapsto e^{im\varphi}c^{(m)}. \tag{30}$$

□

In eSCN, the SO(2) convolution is then constructed on the aligned features of each m as

$$\tilde{c}_{\tilde{k}}^{(m)} = \sum_k \left(W_{\tilde{k}k}^{(1,m)} + iW_{\tilde{k}k}^{(2,m)}\right) c_k^{(m)}, \tag{31}$$

which can be equivalently written in complex form as

$$\tilde{c}^{(m)} = A^{(m)}c^{(m)}, A^{(m)} \in \mathbb{C}^{\tilde{K}_m \times K_m}. \tag{32}$$

We now show that, this form exhausts all possible SO(2)-equivariant linear maps.

Theorem 1 (SO(2)-equivariant linear map)**.** *Let*

$$\mathcal{V} = \bigoplus_m \mathbb{C}^{K_m}, \tilde{\mathcal{V}} = \bigoplus_m \mathbb{C}^{\tilde{K}_m} \tag{33}$$

be SO(2)-equivariant feature spaces. Consider a linear map

$$F : \mathcal{V} \to \tilde{\mathcal{V}} \tag{34}$$

that is SO(2)-equivariant. Then F must take the form of Equation 32.

Proof. By linearity, any F can be represented as

$$F^{(\tilde{m})}(c) = \sum_m A^{(\tilde{m}\leftarrow m)}c^{(m)}, \tag{35}$$

where $A^{(\tilde{m}\leftarrow m)} \in \mathbb{C}^{\tilde{K}_{\tilde{m}} \times K_m}$. Under a rotation $R_y(\varphi)$, equivariance requires

$$\sum_m A^{(\tilde{m}\leftarrow m)}e^{im\varphi}c^{(m)} = e^{i\tilde{m}\varphi}\sum_m A^{(\tilde{m}\leftarrow m)}c^{(m)}. \tag{36}$$

Rearranging yields

$$\sum_m \left(e^{im\varphi} - e^{i\tilde{m}\varphi}\right) A^{(\tilde{m}\leftarrow m)}c^{(m)} = 0. \tag{37}$$

Since this holds for all φ and all $c^{(m)}$, we must have $A^{(\tilde{m}\leftarrow m)} = 0$ whenever $m \neq \tilde{m}$. Thus F is block-diagonal with respect to m, and take the form of Equation 32 for each m. □

It is easy to verify that $\tilde{c}^{(m)}$ is also SO(2)-equivariant, as

$$\tilde{c}^{(m)} \mapsto A^{(m)}\left(e^{im\varphi}c^{(m)}\right) = e^{im\varphi}\tilde{c}^{(m)}. \tag{38}$$

By Lemma 1 and the argument above, both $\tilde{c}^{(m)}$ and $\hat{c}^{(m)}$ transform under an arbitrary global rotation $R \in \mathrm{SO}(3)$ according to the same induced phase,

$$\tilde{c}^{(m)} \mapsto e^{im\varphi}\tilde{c}^{(m)}, \hat{c}^{(m)} \mapsto e^{im\varphi}\hat{c}^{(m)}, \tag{39}$$

where φ is determined by Equation 21. Therefore,

$$\tilde{c}^{(m)}\hat{c}^{(m)*} \mapsto \left(e^{im\varphi}\tilde{c}^{(m)}\right)\left(e^{-im\varphi}\hat{c}^{(m)*}\right) = \tilde{c}^{(m)}\hat{c}^{(m)*}. \tag{40}$$

Since this holds for every global rotation $R \in \mathrm{SO}(3)$, $\tilde{c}^{(m)}\hat{c}^{(m)*}$ is an $\mathrm{SO}(3)$-invariant complex scalar. Consequently, both its real and imaginary parts are $\mathrm{SO}(3)$-invariant real scalars, completing the proof of equivariance.

B Implementation details

B.1 Model implementations

SGBNet is implemented in JAX based on the MLIP codebase Brunken et al. (2025; 2026), and molecular dynamics benchmarks are performed with JAX-MD Schoenholz & Cubuk (2020). The JAX implementations of eSEN and ViSNet are also adopted from MLIP and aligned with their original architectures. PyTorch baselines are evaluated using their paper-aligned implementations without later model-specific acceleration not described in their papers. TF32 and compiled execution are additionally enabled whenever compatible. For E2FormerV2, we evaluate both the compiled implementation without Triton kernels (E2FormerV2-C) and the Triton accelerated implementation. The latter follows the original implementation and is evaluated without additional compilation, as their Triton kernels are incompatible with compilation. A controlled comparison with Triton disabled and compilation enabled is reported in Section C.1.

For all retrained models, we adopt the FAIRChem Shuaibi et al. loss formulation and optimizer instead of the default MLIP training configuration. Energy and stress use MAE losses, whereas force error is computed as the per-atom ℓ_2 norm over the three Cartesian components. All models are optimized with AdamW Loshchilov & Hutter (2019). To improve training stability for deeper networks, loss spikes encountered during fine-tuning are handled by restoring the previous checkpoint and skipping the corresponding batches Chowdhery et al. (2023).

Following recent FAIRChem implementations, Wigner D-matrices in eSEN and SGBNet are computed using a quaternion-based formulation rather than Euler angles to avoid gimbal lock Shoemake (1985) and improve numerical stability. We further reorder the edge channels in SGBNet to group positive and negative m components, enabling coefficient application through broadcasting and reducing tensor operations.

B.2 Efficiency profiling

We follow the scaling-benchmark setup of MLIPAudit Wehrhan et al. (2025), while using a different set of PDB systems to extend the evaluation to larger system sizes. Runtime is measured from complete MD steps after compilation and warm-up.

For PyTorch models, FLOPs are measured for the forward pass and reported as three times the forward count to approximate the additional cost of force computation through automatic differentiation. For JAX models, FLOPs are obtained from XLA static analysis of the compiled energy-and-force computation. The JAX analysis is performed on CPU because the corresponding GPU estimates are unreliable.

Peak GPU memory for PyTorch models is measured from process-level GPU usage, whereas JAX memory is obtained from XLA allocator statistics because JAX preallocates GPU memory by default. FLOPs and memory values are therefore used primarily for within-framework comparisons.

C Detailed experiments

C.1 Speed and memory

We evaluate scaling on six Protein Data Bank systems: 1UAO (chignolin, 138 atoms), 5KGZ (634), 1BIP (1818), 1VSQ (6713), 1LQ8 (11658), and 7SPC (29716). For JAX implementations, both cell-list and dense neighbor representations provided by JAX-MD are evaluated.

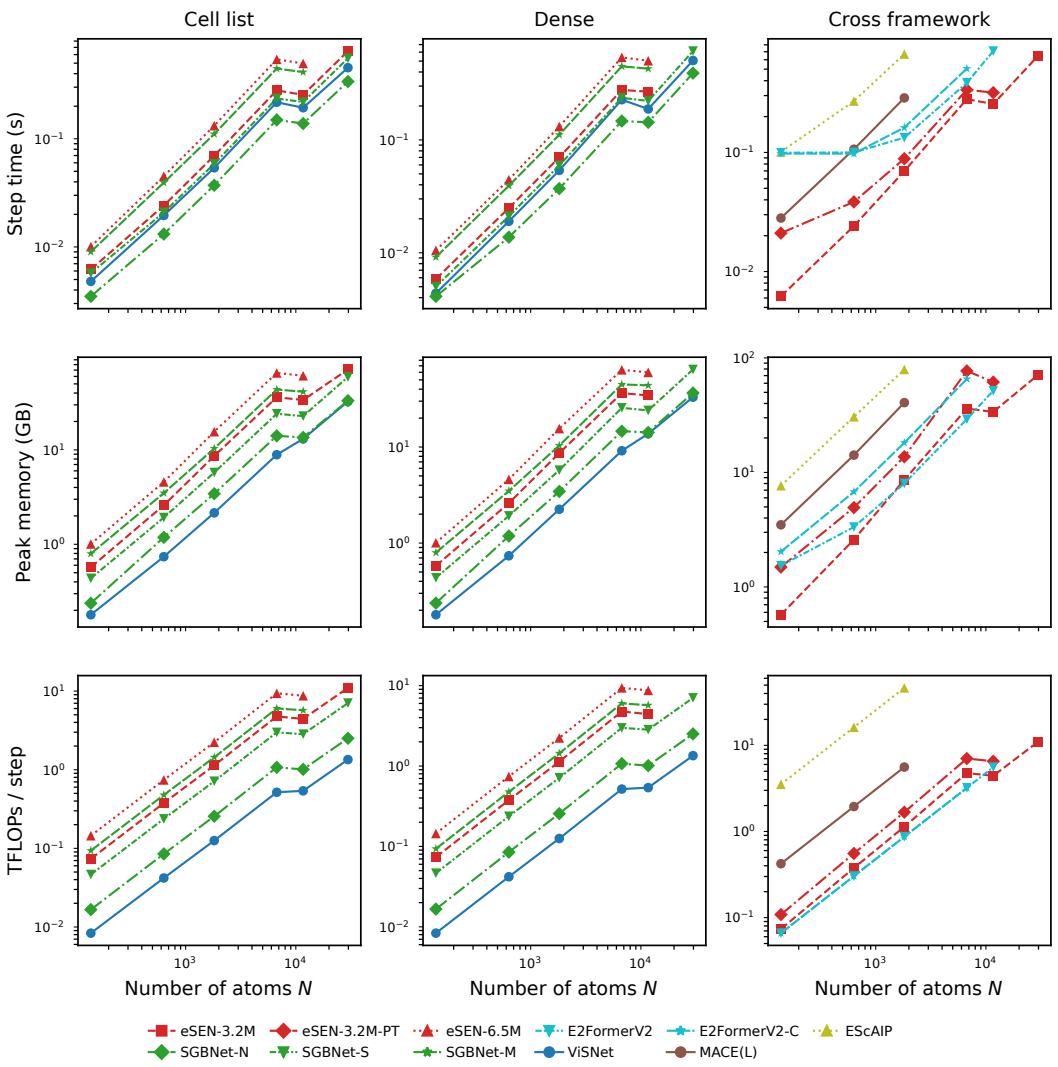

Figure 3: Scaling behavior of computational efficiency across system sizes and implementations. Columns correspond to JAX implementations using cell-list neighbors, dense neighbors, and cross-framework comparisons, respectively. Rows show step time, peak memory, and FLOPs per step.

As shown in Figure 3, the relative efficiency ranking remains consistent across the two JAX neighbor representations. SGBNet-N achieves the lowest runtime among all models, while SGBNet-S maintains lower computational cost than eSEN-3.2M with comparable or improved accuracy. The cross-framework comparison further examines implementation effects. The identical eSEN-3.2M architecture implemented in JAX and PyTorch shows similar scaling trends at large system sizes, while the PyTorch implementation exhibits higher overhead for small systems. This difference is reflected by the increased per-atom runtime and memory usage at the left side of the scaling curves. Among PyTorch baselines, compared to E2Former-C, E2FormerV2 achieves improved memory efficiency and competitive runtime, particularly for larger systems, due to kernel-level optimization. Overall, SGBNet consistently achieves favorable scaling behavior across different system sizes and implementation settings, providing improved accuracy-efficiency trade-offs with lower runtime, memory consumption, and computational cost while maintaining competitive predictive performance.

For the larger material models, eSEN-30M and SGBNet-L require 286.08 and 253.63 μs/atom/step, 30.647 and 27.644 MB/atom, and 6607.22 and 6124.73 MFLOP/atom, respectively, in the small-system regime. Both models encounter OOM in the large-system regime.

C.2 MLIPAUDIT RESULTS

Bond lengths and ring geometries are expected to exhibit thermal fluctuations during MD simulations, and deviations from a static QM reference do not directly measure physical accuracy. Therefore, lower structural deviations do not necessarily indicate improved model fidelity, as excessively constrained dynamics may instead suppress physically relevant fluctuations. We use the MAE/RMSE relative to the MLIPAudit threshold in Figure 4 only as a diagnostic for potential structural instability rather than as a direct ranking criterion. All four models exhibit comparable deviations, with none showing systematic violations of the 0.05 Å threshold, suggesting that SGBNet maintains structural stability comparable to the baseline models in bond and ring-geometry preservation.

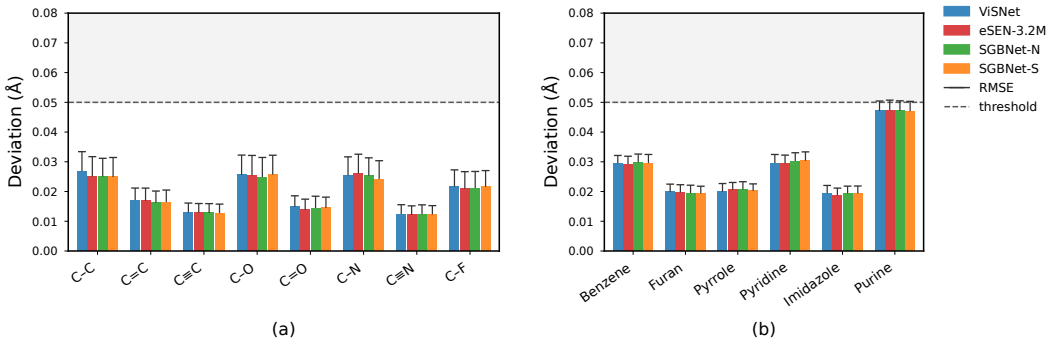

Figure 4: MD structural fidelity for small molecules. (a) Bond-length deviations from QM references along MD trajectories for different bond types. (b) Ring out-of-plane deviations (plane RMSD) for representative aromatic rings. Bars indicate MAE and caps indicate RMSE. The dashed line denotes the 0.05 Å scoring threshold.

For conformer selection, relative conformer energies are referenced to the DFT global minimum for each molecule. Spearman rank correlation evaluates the consistency of conformer ordering, whereas energy errors quantify the accuracy of relative energy differences. As shown in Figure 5, SGBNet-N and SGBNet-S achieve higher mean Spearman correlations than ViSNet and eSEN-3.2M, respectively, demonstrating improved identification of energetically favorable conformers. The corresponding RMSE values are not necessarily reduced, since rank-based metrics are insensitive to systematic scaling of energy differences and can remain high in the presence of a small number of large energetic deviations. Nevertheless, the improved ranking performance indicates that SGBNet provides a more reliable description of the conformational energy landscape.

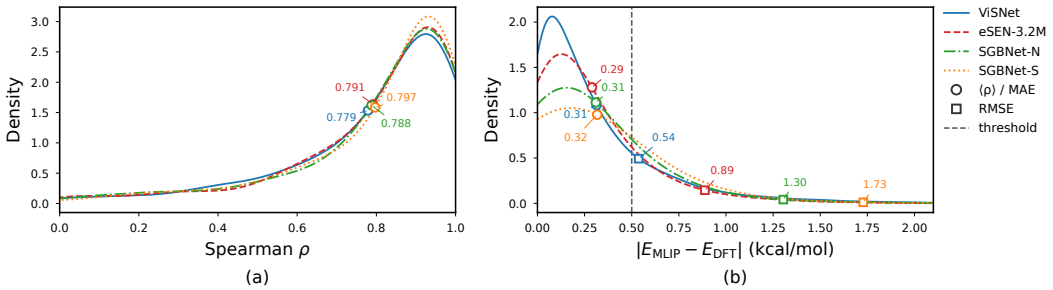

Figure 5: Conformer selection metrics. (a) Distribution of per-molecule Spearman rank correlations between MLIP and DFT relative conformer energies. (b) Distribution of absolute relative-energy errors. Open circles and squares denote mean Spearman correlation and MAE/RMSE values, respectively.

Figure 6 presents representative dihedral scans selected to cover distinct energy-landscape topologies. All models reproduce the qualitative features of the DFT torsional profiles, including different barrier shapes and minima. The remaining deviations are mainly observed in high-energy regions,

suggesting that insufficient sampling of strained conformations in the training data may limit quantitative accuracy. Consistent with the representative profiles, Figure 7 shows that all models achieve comparable performance in correlation and barrier-height prediction. SGBNet-N and SGBNet-S provide modest improvements over ViSNet and eSEN-3.2M, respectively, in terms of energy-profile MAE, while maintaining similar correlation coefficients, indicating slightly improved quantitative accuracy without compromising the overall landscape description.

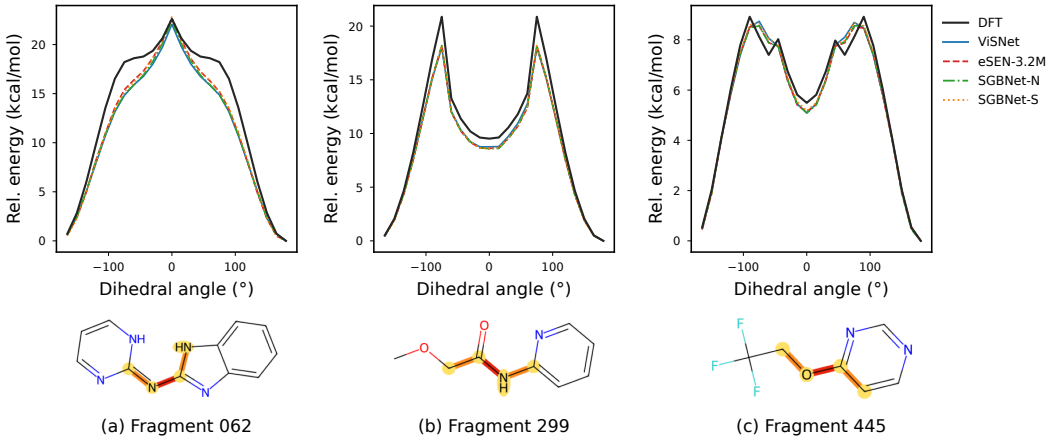

Figure 6: Relative torsional energy profiles for representative dihedral scans. (a), (b), and (c) correspond to single-, double-, and four-peak profiles, respectively. Black curves represent DFT references, and colored curves represent MLIP predictions. The highlighted region in the molecular structures indicates the scanned dihedral angle.

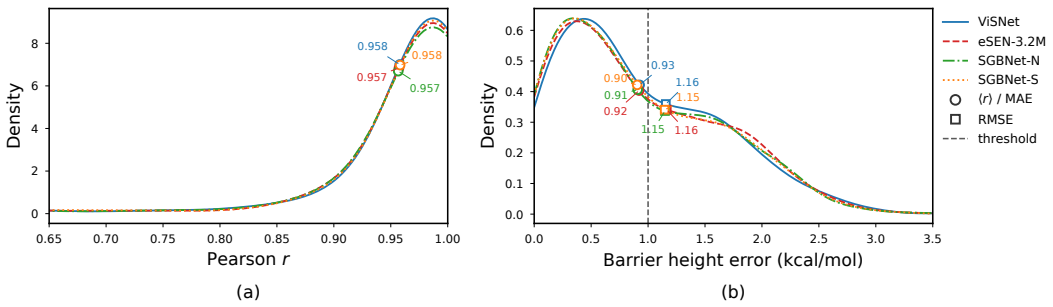

Figure 7: Dihedral-scan accuracy metrics. (a) Distribution of Pearson correlation coefficients between MLIP and DFT energy profiles. (b) Distribution of barrier-height errors. Markers indicate mean correlation coefficients and MAE/RMSE values, respectively.

Figure 8 evaluates energetic accuracy for noncovalent interactions and tautomer stability. The former measures intermolecular interaction energies obtained from bimolecular distance scans, whereas the latter assesses relative stability between molecular forms. SGBNet-N and SGBNet-S consistently achieve lower MAE and RMSE values than ViSNet and eSEN-3.2M, respectively, for both tasks. These results demonstrate that SGBNet provides improved quantitative descriptions of weak intermolecular interactions and subtle intramolecular energy differences, which are challenging for MLIPs due to their small energy scales.

Figure 9 evaluates whether the improved energetic accuracy translates into realistic liquid-phase structural properties. Radial distribution functions (RDFs) are compared for water, CCl_4, methanol, and acetonitrile. For water, the complete RDF profile is evaluated against experimental measurements, whereas for the other solvents, where experimental RDFs are unavailable, we focus on the accuracy of the first-peak position relative to literature references. SGBNet-N and SGBNet-S achieve closer agreement with reference peak positions than eSEN-3.2M and ViSNet, while also

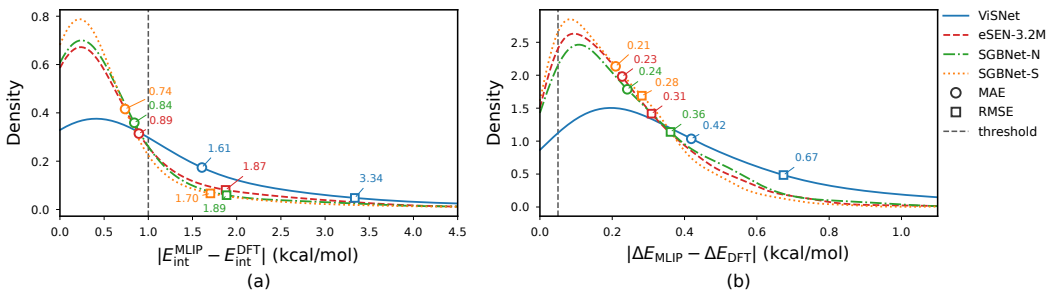

Figure 8: Energy-error distributions. (a) Absolute errors in noncovalent interaction energies. (b) Absolute errors in tautomer energy differences. All values are reported in kcal/mol; open circles and squares indicate MAE and RMSE, respectively.

reducing deviations in the water RDF profile. These results indicate that SGBNet provides improved transferability from molecular energetics to condensed-phase structural properties.

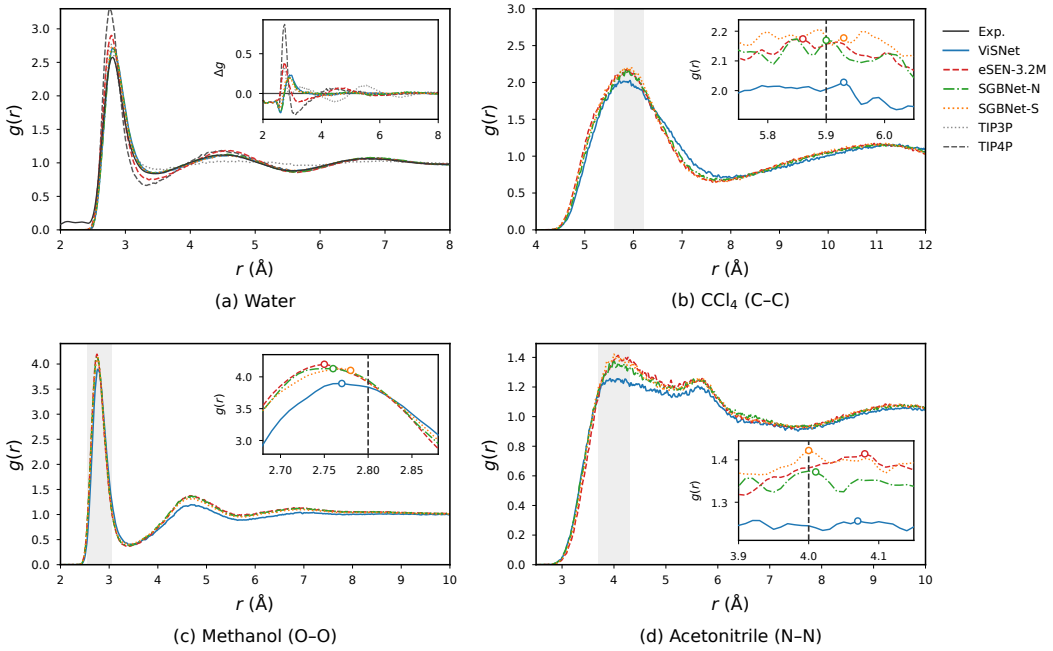

Figure 9: Radial distribution functions (RDFs). For water, the inset shows the deviation $\Delta g(r) = g_{\mathrm{MLIP}}(r) - g_{\mathrm{exp}}(r)$ relative to experiment. For other solvents, the shaded region highlights the first-peak range, with predicted peak positions shown as open circles and literature references indicated by dashed lines.

The sampling benchmark evaluates the agreement between MD-generated dihedral distributions and reference Ramachandran/rotamer libraries. Outlier ratios measure the fraction of sampled configurations outside the reference support, while RMSD and Hellinger distance quantify distribution-level agreement. As shown in Figure 10, side-chain conformations remain the dominant challenge for all models, with higher outlier rates and distribution discrepancies than backbone dihedrals. The shared difficulties for residues such as GLU and TYR suggest that these errors mainly originate from limited rotamer coverage in the training data or reference-library limitations rather than model-specific deficiencies. Among the four models, SGBNet-N achieves the highest sampling score, with improved backbone stability and closer agreement with side-chain reference distributions. The improvements of SGBNet-N are primarily reflected in reduced moderate deviations and

better ensemble-level agreement, while the most challenging residue types remain difficult for all models.

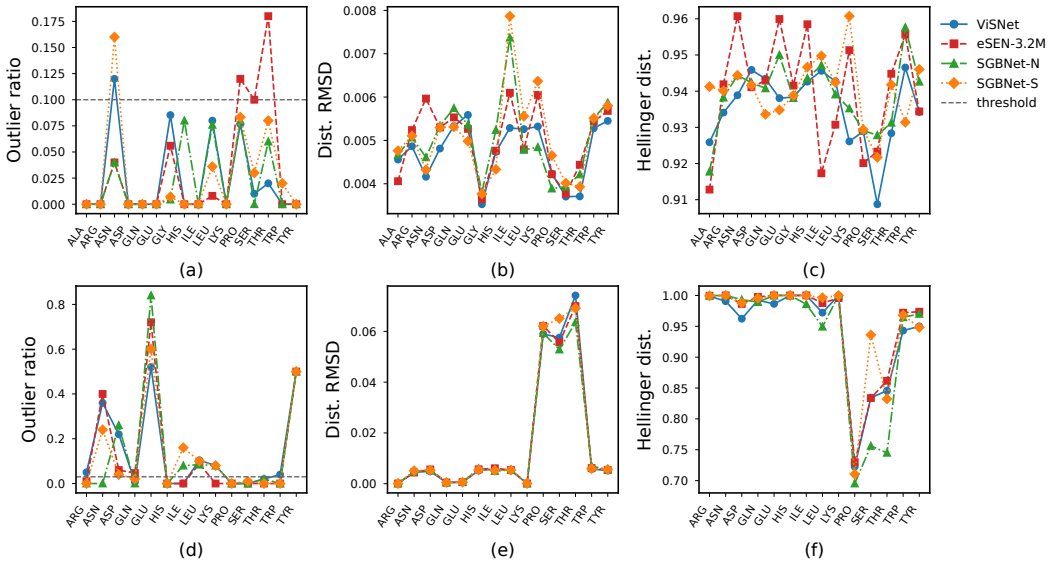

Figure 10: Per-residue protein sampling quality. Backbone metrics: (a) Ramachandran outlier ratio, (b) dihedral-distribution RMSD, and (c) Hellinger distance. Side-chain metrics: (d)-(f) the corresponding metrics for χ rotamers. Dashed lines in (a) and (d) indicate the MLIPAudit scoring thresholds (0.10 and 0.03, respectively).

D HYPER-PARAMETERS

	SGBNet-N	SGBNet-S	SGBNet-M	SGBNet-L
num_layers	2	2	4	8
scalar_channels	128	256	256	192
geometric_channels	32	64	64	96
radial_channels	128	128	128	128
l_max	2	2	2	3
m_max	2	2	2	2
num_rbf	32	32	32	32