
Quadruple Attention in Many-body Systems for Accurate Molecular Property Predictions

Jiahua Rao^{1*} Dahao Xu^{1*} Wentao Wei¹ Yicong Chen¹ Mingjun Yang² Yuedong Yang^{1,3}

Abstract

While Graph Neural Networks and Transformers have shown promise in predicting molecular properties, they struggle with directly modeling complex many-body interactions. Current methods often approximate interactions like three- and four-body terms in message passing, while attention-based models, despite enabling direct atom communication, are typically limited to triplets, making higher-order interactions computationally demanding. To address the limitations, we introduce MABNet, a geometric attention framework designed to model four-body interactions by facilitating direct communication among atomic quartets. This approach bypasses the computational bottlenecks associated with traditional triplet-based attention mechanisms, allowing for the efficient handling of higher-order interactions. MABNet achieves state-of-the-art performance on benchmarks like MD22 and SPICE. These improvements underscore its capability to accurately capture intricate many-body interactions in large molecules. By unifying rigorous many-body physics with computational efficiency, MABNet advances molecular simulations for applications in drug design and materials discovery, while its extensible framework paves the way for modeling higher-order quantum effects.

1. Introduction

Molecular representation is pivotal in cheminformatics, critically influencing the accuracy of property prediction in drug discovery and materials science (Fourches et al., 2010).

*Equal contribution ¹School of Computer Science and Engineering, Sun Yat-sen University, China ²Shenzhen Jingtai Technology Co., Ltd. (XtalPi), Shenzhen, China ³Key Laboratory of Machine Intelligence and Advanced Computing, Sun Yat-sen University, China. Correspondence to: Yuedong Yang <yangyd25@mail.sysu.edu.cn>.

Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

Traditional approaches, including structural fingerprints and descriptor-based approaches (Rogers & Hahn, 2010; Jaeger et al., 2018), encode molecules using simplified heuristics that emphasize one- and two-body interactions (Song et al., 2020; Rao et al., 2022a), for instance, atomic connectivity or pairwise bond relationships. While these representations efficiently capture coarse structural patterns, they inherently neglect complex quantum mechanical phenomena. Such simplifications limit their utility in modeling properties rooted in the synergistic interplay of multiple atoms or bonds (Hansen et al., 2015).

To address these limitations, modern machine learning frameworks increasingly rely on quantum mechanical principles to predict molecular properties, necessitating explicit modeling of many-body interactions (Unke et al., 2021). These include not only classical covalent and non-covalent atom-atom interactions (two-body) but also higher-order terms such as bond-angle distortions (three-body), torsional potentials (four-body), and electronic delocalization effects (e.g., in conjugated π -systems). For example, at the quantum level, properties such as dipole moments, polarizability, and excitation energies emerge from entangled multi-atom correlations that resist decomposition into pairwise terms. Ignoring these interactions can lead to inaccurate predictions of molecular properties (Yang & Zhou, 2008).

To mimic the many-body terms, Graph Neural Networks (GNNs) (Schütt et al., 2018; Gasteiger et al., 2020b) and Transformers (Liao & Smidt, 2023) have emerged as foundational architectures by leveraging node-edge frameworks to represent atoms and bonds as discrete entities. However, their reliance on pairwise correlations inherently limits their capacity to encode multi-atom interactions, such as three-body angle dependencies or four-body torsional terms. While these architectures approximate higher-order effects through the iterative aggregation of local interactions, this inductive bias introduces systematic errors in capturing global quantum behaviors, particularly in systems with strong electronic cooperativity or long-range correlations.

Various architectures have recently been proposed to mitigate these limitations. For example, SE3Set (Wu et al., 2024) operates on hypergraphs constructed from overlapping fragments to provide a natural representation of many-

body phenomena. ViSNet (Wang et al., 2024) and QuinNet (Wang et al., 2023b) implicitly capture many-body geometric features in message passing, aligning with the force fields used in classical molecular dynamics (MD). Despite their strengths, these methods primarily rely on message-passing mechanisms that approximate multi-body interactions rather than explicitly modeling them, often introducing trade-offs in accuracy or computational efficiency.

To address these challenges, we propose **MABNet** (**M**any-**B**ody interaction **N**etwork), a novel four-body direct communication attention mechanism specifically designed for quantum many-body systems. Unlike traditional methods that rely on approximations or implicit representations of high-order interactions, our approach explicitly models complex interactions involving up to four bodies. This explicit modeling not only captures the intricate geometric and structural dependencies within molecular systems but also ensures computational efficiency, making it suitable for larger molecular simulations. These geometric features also enhance the E(3) equivariant message-passing process by providing a more expressive and accurate representation of the molecular system’s symmetries. By explicitly integrating higher-order interactions into the architecture, our approach overcomes the limitations of existing methods, paving the way for more precise molecular property predictions.

Our approach achieves state-of-the-art (SOTA) performance on challenging benchmarks, including MD22 and SPICE, showcasing its capability to accurately capture intricate many-body interactions, even in larger and more complex molecular systems. By excelling across diverse molecular structures, our method highlights its robustness and adaptability, addressing the limitations of existing models that struggle with high-order interactions. This exceptional performance not only validates its effectiveness but also underscores its transformative potential in computational chemistry. By providing a more accurate and physically grounded framework for modeling complex systems, our method paves the way for advances in molecular dynamics simulations, quantum chemistry, and related fields.

The contributions can be summarized as follows:

- We introduce a novel attention mechanism that explicitly models complex higher-order interactions, enabling the direct communication of four-body interactions in molecular quantum many-body systems.
- Our model generates rich geometric features through the four-body direct communication attention mechanism in the E(3) equivariant message-passing process, improving the expressivity and accuracy of our model.
- The proposed method achieves SOTA performance on challenging benchmarks, including MD22 and SPICE,

demonstrating its ability to accurately model intricate many-body interactions across diverse and complex molecular systems.

2. Related Work

2.1. Molecular Property Predictions

Graph neural networks (GNNs) (Gilmer et al., 2017; Song et al., 2020; Rao et al., 2022b; 2024a) have revolutionized molecular property predictions by enabling accurate modeling of atomic interactions. They (Schütt et al., 2018; Gasteiger et al., 2020b) naturally encode atomic systems as graphs, preserving permutation invariance through message passing to predict molecular energies directly from quantum states. However, their limited ability to incorporate geometric symmetries (e.g., SE(3)/E(3) equivariance) restricts generalization across conformations and coordinate systems.

Recent advances address this by integrating geometric inductive biases (Satorras et al., 2021; Thölke & Fabritiis, 2022; Rao et al., 2024b). Equivariant architectures (Schütt et al., 2021) use irreducible representations to enforce symmetry-aware tensor transformations, while invariant methods (Montavon et al., 2012) augment features with descriptors like distances or angles. Graph transformers (Liao & Smidt, 2023) have also been adapted to geometric domains: Equiformer combines equivariant irreps with attention mechanisms, achieving accurate force and energy predictions.

2.2. Many-body Interactions Modeling

Traditional approaches approximate many-body interactions via pairwise decompositions (atom-atom, atom-bond, etc.) (Unke et al., 2021), but struggle to capture cooperative phenomena like dipole moments, polarizability, and excitation energies, which require explicit multi-atom correlation modeling. Neural networks must preserve SE(3)/E(3)-equivariance while directly representing hierarchical relationships, avoiding reliance on pairwise approximations.

Recent methods show partial progress: SE3Set (Wu et al., 2024) employs hypergraphs to model delocalized effects but introduces computational overhead from fragment decomposition, while VisNet/QuinNet (Wang et al., 2024; 2023b) approximate multi-body terms via geometric message passing, inheriting classical force fields’ limitations. Emerging approaches like TGT (Hussain et al., 2024) enable direct triplet communication but lack SE(3)/E(3)-equivariance and scalability limited to three-node interactions. Other strategies, such as GEM-2’s (Liu et al., 2022) axial attention over edge pairs and AlphaFold’s (Jumper et al., 2021) scalar-based triangular updates, either constrain interaction granularity to triplets or rely on scalar approximations. Collectively,

Table 1. Comparison of different many-body interactions.

| Methods | Axial Attn. | Tria. Update | Triplet Attn. | MABNet |
|---------|--------------------------------|-------------------------------------|--------------------------------|--------------------------------|
| Ops. | | | | |
| Attn. | 3-body ✓ | 3-body ✗ | 3-body ✓ | 4-body ✓ |
| Equi. | ✗ | ✗ | ✗ | ✓ |
| Values | Vectors | Scalars | Vectors | Vectors |
| Comp. | $\mathcal{O}(\mathcal{N} ^3)$ | $\mathcal{O}(\mathcal{N} ^{2.37})$ | $\mathcal{O}(\mathcal{N} ^3)$ | $\mathcal{O}(\mathcal{N} ^3)$ |

these methods fall short of enabling explicit and equivariant modeling of many-body effects (e.g., torsions, covalent hybridization). In general, these limitations underscore the need for architectures that unify E(3)-equivariance preservation with scalable many-body interaction modeling.

3. Preliminaries

3.1. Many-body Interactions

The properties of a molecular system, such as energy, can be expressed as the sum of contributions from various factors, including bonds, bond angles, torsion angles, and nonbonding interactions (such as electrostatic and van der Waals forces) (Unke et al., 2021). These include three-body interactions (e.g., bond angles, where the energy depends on three linked atoms), four-body interactions (e.g., torsion angles, involving four sequentially bonded atoms), and higher-order collective effects (e.g., electronic polarization).

Therefore, the total energy E of an N -atom system is:

$$E = \sum_{i < j} E_{ij}^{(2)} + \sum_{i < j < k} E_{ijk}^{(3)} + \sum_{i < j < k < l} E_{ijkl}^{(4)} + \dots, \quad (1)$$

where $E^{(n)}$ denotes n -body terms. Classical force fields approximate $E^{(3)}$, $E^{(4)}$ with handcrafted functions, while MLFFs learn them directly.

3.2. Complexity Analysis

Node-to-Node Attention (Pairwise) Standard attention mechanisms (Vaswani, 2017; Chen et al., 2021) in Transformers operate on node pairs (i, j) , yielding a computational complexity of $\mathcal{O}(|\mathcal{N}|^2)$, where N is the number of atoms. While efficient, this approach restricts expressivity to 1-GWL (Graph Weisfeiler-Lehman) equivalence (Joshi et al., 2023), failing to distinguish molecular substructures requiring higher-order interactions. This limitation arises because pairwise attention cannot explicitly model multi-body geometric relationships like angles or torsions.

Axial Attention Axial attention (Liu et al., 2022) generalizes self-attention to pairs of edges (i, j) and (j, k) , com-

puting interactions as:

$$s_{ij} = \sum_k (a_{ijk} v_{jk}), a_{ijk} = \text{Softmax}(\mathbf{q}_{ij}^T \mathbf{k}_{jk}), \quad (2)$$

where $\mathbf{q}_{ij}^T, \mathbf{k}_{jk}$ are queris and keys for edge pairs. Though this extends complexity to $\mathcal{O}(|\mathcal{N}|^3)$, it neglects critical 3rd-order geometric features (e.g., the angle θ_{ijk}) and the direct influence of the (i, k) pair. GEM-2 (Liu et al., 2022) partially addresses this by 3rd-order positional encodings, but it remains limited to implicit angular dependencies, reducing accuracy in many-body systems.

Triangular Update (AlphaFold) AlphaFold’s triangular update aggregates scalar features (Jumper et al., 2021; Lu et al., 2022; Abramson et al., 2024) from edge triplets (ij, jk, ik) via tensor products, achieving complexity $\mathcal{O}(|\mathcal{N}|^{2.37})$ through optimized matrix multiplication (Ambainis et al., 2015). However, this method lacks attention-based gating: all triplets contribute equally, regardless of chemical relevance. Additionally, scalar features discard directional information (e.g., vectorial forces), and unbounded summation over variable-sized graphs introduces instability.

Triplet Graph Transformers (TGT) Triplet-based methods (Hussain et al., 2024) explicitly model 3rd-order interactions (i, j, k) , enabling direct information flow between edges (i, j) and (j, k) . The attention mechanism becomes:

$$s_{ij} = \sum_k a_{ijk} s_{jk}, a_{ijk} = \text{Softmax}(\mathbf{q}_{ij}^T \mathbf{k}_{jk} + \phi(\theta_{ijk})), \quad (3)$$

where $\mathbf{q}_{ij}^T, \mathbf{k}_{jk}$ are queries and keys for the triplet of nodes, and $\phi(\theta_{ijk})$ encodes the angular features. While this improves expressivity, the $\mathcal{O}(|\mathcal{N}|^3)$ complexity persists, and equivariance is sacrificed by scalarizing geometric features.

Proposed Many-body Attention Network (MABNet) Our method generalizes attention to four-body interactions (i, j, k, l) , critical for modeling torsional angles ϕ_{ijkl} and non-local polarization effects. To mitigate the combinatorial explosion $\mathcal{O}(|\mathcal{N}|^4)$, we employ: (1) spatial sparsity: restrict interactions to quartets within a cutoff radius r_c ; (2) quantum-inspired pruning: prioritizes edges critical for multi-body interactions (e.g., dihedrals).

The resulting complexity scales as $\mathcal{O}(|\mathcal{N}|^2 \cdot |E|)$, where $|E| \leq 6|N|$ is the average neighborhood size. Crucially, our attention retains SE(3)-equivariance by operating on vector-valued queries, keys, and values:

$$s_{ij} = \sum_{k,l} a_{ijkl} \cdot v_{jkl}, \quad (4)$$

$$a_{ijkl} = \text{Softmax} \left(\frac{\langle \mathbf{q}_{ij}, \mathbf{k}_{jkl} \rangle + \psi(\phi_{ijkl})}{\sqrt{d}} \right), \quad (5)$$

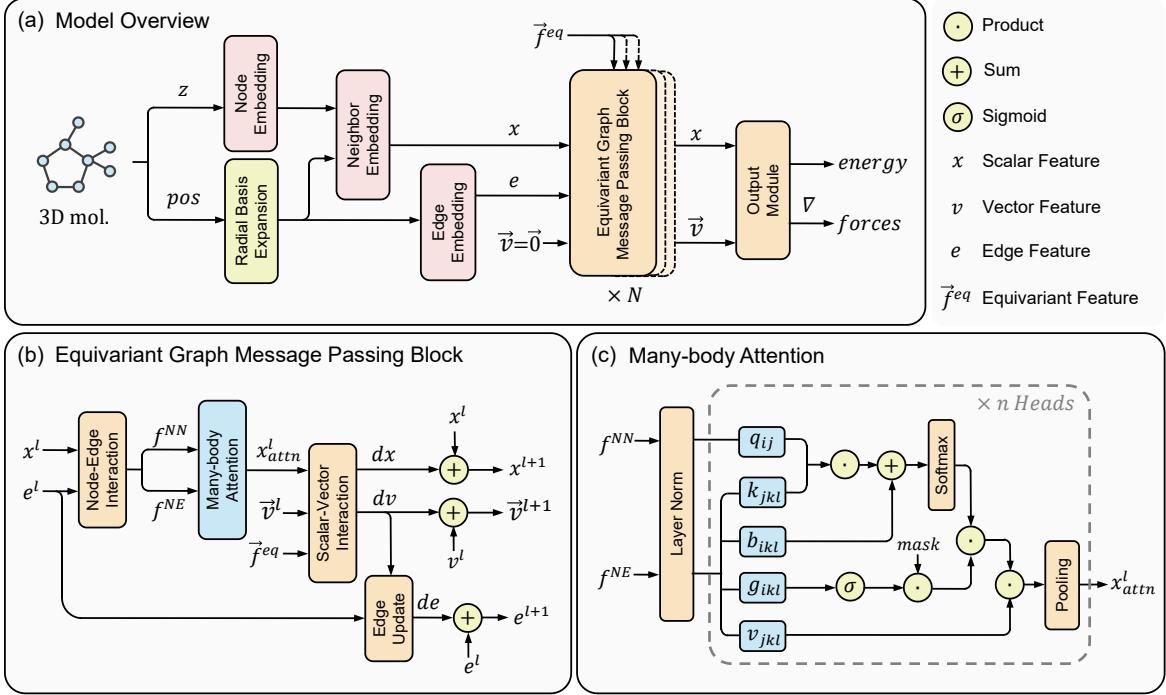


Figure 1. The overall framework of our method.

where $\psi(\phi_{ijkl})$ encodes torsional features via spherical harmonics. This framework combines the expressivity of explicit four-body terms with the efficiency of equivariant sparsity, addressing the limitations of prior works (Table 1).

4. Methodology

4.1. Model Architecture

Figure 1 illustrates our framework. The input is a 3D molecular structure defined by atomic types and coordinates. The model first embeds atomic and edge features into scalar features , vector features , and edge features . Inspired by VisNet (Wang et al., 2024), these features are iteratively updated through equivariant graph message passing (EquiMP) blocks, which enforce $E(3)$ -equivariance during all transformations. The key innovation to our model is the many-body attention module within each EquiMP block, which explicitly models four-body interactions to capture complex quantum effects. Finally, the updated scalar and vector features are processed to predict total energy, and atomic forces are derived as the energy gradient with respect to atomic coordinates, ensuring physical consistency.

4.2. Feature Embedding

As shown in Figure 1(a), the feature embedding layer transforms 3D molecular structures into three components: scalar features $x \in \mathbb{R}^{N \times D}$, vector features $\vec{v} \in \mathbb{R}^{N \times L \times D}$, and edge features $e \in \mathbb{R}^{E \times D}$, where N is the number of atoms,

E represents the number of edges, $L = (l_{\max} + 1)^2 - 1$ corresponds to the spherical harmonics expansion level, and D is the feature dimensionality.

Scalar and Vector Node Feature Embedding Scalar features are derived through two sequential stages: (1) atomic embedding directly encodes atomic types into high-dimensional vectors, followed by (2) neighborhood aggregation that integrates adjacent atom data via radial basis function (RBF)-expanded distances, filtered through a cutoff function and combined with neighbor features via element-wise multiplication. The scalar features are generated by concatenating the initial atomic embeddings with the aggregated neighborhood representations:

$$x_{nbh} = \sum_{j \in \mathcal{N}(i)} (\phi(d_{ij}) \cdot (\mathbf{W}_n d_{ij}^{rbf} + \mathbf{b}_n) \odot h_j), \quad (6)$$

where \odot denotes element-wise multiplication. \mathbf{W}_n and \mathbf{b}_n are learnable parameters, and $\mathcal{N}(i)$ denotes the set of neighboring nodes of node i . The scalar feature x_i of node i is computed by concatenating its initial feature h_i with the aggregated neighbor features x_{nbh} , followed by a linear transformation:

$$x_i = \mathbf{W}_s(h_i \parallel x_{nbh}) + \mathbf{b}_s, \quad (7)$$

where the \parallel denotes the concatenation operation. \mathbf{W}_s and \mathbf{b}_s are learnable parameters. The vector features \vec{v} are initialized as zero tensors to preserve equivariance.

Edge Feature Embedding Edge features are computed by combining the scalar features of connected nodes with edge attributes transformed via RBF expansion into a higher-dimensional space. This process captures complex interaction patterns, enhancing the model’s ability to understand intricate network relationships.

Full implementation details for scalar, vector, and edge embeddings are provided in Appendix A.1.

4.3. Many-body Attention Module

As shown in Fig. 1(b), the scalar features \mathbf{x} are updated through a many-body attention module in each EquiMP layer. Let $x^l \in \mathbb{R}^{N \times D}$ and $e^l \in \mathbb{R}^{E \times D}$ denote the node and edge features at layer l , respectively. These features interact through two distinct relationship tensors:

$$f_{ij}^{NN} = x_i^l \cdot x_j^l, \quad f_{jkl}^{NE} = x_j^l \cdot e_{kl}^l, \quad (8)$$

where N , E , and D correspond to the number of atoms, edges, and feature dimensions. The node-node interaction tensor f^{NN} captures pairwise atomic interactions (two-body terms), while the node-edge interaction tensor f^{NE} encodes three-body relationships through edge-mediated connections between node triples (j, k, l) .

Inspired by the cross-attention mechanisms (Vaswani, 2017), we establish asymmetric information fusion by designating different relationship tensors as query (Q), key (K), and value (V) components. As shown in Fig. 1(c), the interaction process begins with layer normalization of both tensors, followed by learned linear projections:

$$\mathbf{q}_{ij} = W^Q f_{ij}^{NN}, \mathbf{k}_{jkl} = W^K f_{jkl}^{NE}, \mathbf{v}_{jkl} = W^V f_{jkl}^{NE}, \quad (9)$$

$$b_{ikl} = W^B f_{ikl}^{NE}, g_{ikl} = W^G f_{ikl}^{NE}, \quad (10)$$

where W^Q, W^K, W^V, W^B, W^G are learnable projection matrices. The bias term b_{ikl} and gating vector g_{ikl} enable adaptive modulation of attention patterns.

For each attention head, the interaction weights are computed through a gated softmax operation:

$$a_{ijkl} = \text{Softmax} \left(\frac{\mathbf{q}_{ij} \mathbf{k}_{jkl}}{\sqrt{d}} + b_{ikl} \right) \times \mathcal{G}(g_{ikl}), \quad (11)$$

$$x_{ij}^l = \sum_{k,l} a_{ijkl} \mathbf{v}_{jkl}, \quad (12)$$

where $\mathcal{G}(g_{ikl}) = \sigma(g_{ikl})$ denotes the gate function. This gating mechanism allows the model to selectively control the flow of information, enhancing its ability to learn complex, adaptive representations by emphasizing specific interactions.

The final node update aggregates information from all neighboring components:

$$x_{attn}^l = \sum_{j \in \mathcal{N}(i)} \sum_{k,l} a_{ijkl} \mathbf{v}_{jkl}, \quad (13)$$

where $\mathcal{N}(i)$ denotes the neighborhood of node i . This multi-head attention formulation enables simultaneous modeling of both direct many-body (node-node, node-edge-node) interactions, effectively capturing complex many-body relationships while maintaining permutation equivariance through invariant tensor operations.

4.4. Equivariant graph neural message passing

As shown in Fig. 1(b), each EquiMP layer performs feature updates through an additive residual scheme that preserves geometric equivariance:

$$x^{l+1} = x^l + \Delta x^l, \quad \vec{v}^{l+1} = \vec{v}^l + \Delta \vec{v}^l, \quad e^{l+1} = e^l + \Delta e^l. \quad (14)$$

The update terms $\Delta x^l, \Delta \vec{v}^l, \Delta e^l$ are computed through equivariant transformation blocks and respect the geometric constraints.

This architecture ensures all feature updates transform equivariantly under SE(3) transformations while enabling rich information flow between scalar and geometric representations. The residual connections help preserve physical constraints while allowing deep feature integration across multiple interaction scales.

Scalar-Vector Interaction The many-body attention module is applied to the scalar features x to directly enable four-body communication for feature updates. The updated scalar feature x_{attn} of node i is divided into three components: x_i^1, x_i^2 , and x_i^3 . Meanwhile, the vector feature \vec{v}_i is split into two parts: \vec{v}_i^1 and \vec{v}_i^2 . The updated scalar feature Δx_i is derived by integrating \vec{v}_i^1 :

$$\Delta x_i = (\mathbf{W}_{v1} \cdot \vec{v}_i^1 + \mathbf{b}_{v1}) \odot x_i^1 + x_i^2, \quad (15)$$

where \mathbf{W}_{v1} and \mathbf{b}_{v1} are learnable parameters, and \odot denotes the Hadamard product.

The update of vector features employs the Hadamard product of the updated scalar feature x_i^3 and the linearly transformed v_i^2 , in addition to incorporating the equivariant feature f_i^{eq} :

$$\begin{aligned} \Delta \vec{v}_i &= (\mathbf{W}_{v2} \cdot \vec{v}_i^2 + \mathbf{b}_{v2}) \odot x_i^3 + f_i^{eq}, \\ f_i^{eq} &= s^1 \odot \vec{v}_j + s^2 \odot \frac{\vec{r}_{ij}}{\|\vec{r}_{ij}\|}, \end{aligned} \quad (16)$$

where \mathbf{W}_{v2} and \mathbf{b}_{v2} are learnable parameters. \vec{r}_{ij} represents the position vector from nodes i to j . The equivariant feature f_i^{eq} is derived from the Hadamard product of s^1 with \vec{v}_j

and the Hadamard product of s^2 with the equivariant information, where s^1 and s^2 are projections of the pre-update x used as residuals.

Edge Update The update of edge features employs vector rejection, which calculates the component of node features perpendicular to the edge direction, thereby capturing the influence of relative node positions on edge features. The dot product w_{dot} of vector rejections in two directions is then computed:

$$\begin{aligned} w_{dot} &= R(\vec{v}_i, \vec{r}_{ij}) \cdot R(\vec{v}_j, -\vec{r}_{ij}), \\ R(\vec{v}_i, \vec{r}_{ij}) &= L(\vec{v}_i) - \left(\frac{L(\vec{v}_i) \cdot \vec{r}_{ij}}{\|\vec{r}_{ij}\|^2} \right) \cdot \vec{r}_{ij}, \end{aligned} \quad (17)$$

where $L(\vec{v}_i)$ denotes a linear transformation of \vec{v}_i . The computation for $R(\vec{v}_j, -\vec{r}_{ij})$ follows a similar process.

Finally, the edge feature e_{ij} undergoes a nonlinear transformation, and the Hadamard product with the dot product result w_{dot} yields the updated edge feature Δe_{ij} :

$$\Delta e_{ij} = \text{SiLU}(\mathbf{W}_e \cdot e_{ij} + \mathbf{b}_e) \odot w_{dot}, \quad (18)$$

where \mathbf{W}_e and \mathbf{b}_e are learnable parameters.

4.5. Output Module

Through iterative processing via EquiMP layers that integrate Many-body Attention mechanisms and Equivariant Graph Message Passing operations, the system generates updated scalar features x and geometrically aware vector features v . These dual-modal features emerge as disentangled latent descriptors encoding both invariant chemical properties and directional spatial interactions within the molecular system. These features are processed by the Output Module to ultimately predict the energy E and atomic forces F of the molecular system. The Output Module comprises multiple equivariant output blocks, where scalar and vector features interact within each block as follows:

$$\mathbf{G} = \text{MLP}(x || (\mathbf{W}_o v + \mathbf{b}_o)). \quad (19)$$

Here, the MLP consists of two linear layers with a SiLU activation function, $||$ denotes the concatenation operation, and \mathbf{W}_o and \mathbf{b}_o are learnable parameters. Following pooling across multiple output blocks, the final output of the Output Module yields the energy E . The force \vec{F}_i acting on atom i is computed as the negative gradient of E with respect to its initial input coordinates pos_i , expressed as:

$$\vec{F}_i = -\frac{\partial E}{\partial \text{pos}_i}. \quad (20)$$

5. Experiments

5.1. Experimental Settings

In this section, we briefly introduce the benchmarks, baselines, and implementation details related to our experiment. Further details are presented in the Appendix A.3-A.4.

Datasets We consider two challenging benchmarks of MD22 and SPICE, following (Wang et al., 2024) and (Eastman et al., 2023). **MD22** was introduced by (Chmiela et al., 2023), containing a 42-atom peptide to a 370-atom nanotube, with high-resolution sampling at 400–500 K using the PBE+MBD (Perdew et al., 1996; Tkatchenko et al., 2012) framework for energy and force computations. The training and testing splits used in **MD22** are consistent with those in methods such as LSR-MP (Li et al., 2023). **SPICE**, collected by (Eastman et al., 2023), offers approximately one million conformations for pharmaceutical molecules, dipeptides, and solvated amino acids. **SPICE** spans a wider range of chemical elements compared to **MD22** and includes both covalent and non-covalent interactions, with energy and force calculations performed at the ω B97M-D3(BJ)/def2-TZVPPD level of theory. Unlike **MD22**, which focuses on training and testing within a single molecule, **SPICE** enables cross-molecular training and testing, making it a more comprehensive benchmark for assessing the generalizability of molecular dynamics simulation models. The **SPICE** dataset follows a consistent 8:1:1 split for training, validation, and testing. For more details about the datasets, please refer to Appendix A.3.

Baselines We compared our method with the state-of-the-art baselines, including both GNN/Transformers-based methods such as SchNet (Schütt et al., 2018), sGMLM (Chmiela et al., 2023), PaiNN (Schütt et al., 2021), ET (Thölke & De Fabritiis, 2022), So3krates (Frank et al., 2022), MACE (Batatia et al.), and Equiformer (Liao & Smidt, 2023), as well as methods designed for modeling many-body interactions, such as ViSNet (Wang et al., 2024) and LSR-MP (Li et al., 2023). Additional baseline methods and their detailed descriptions are provided in Appendix A.4.

Implementation details All models are implemented in PyTorch (Paszke et al., 2019) and trained using the Adam optimizer (Kingma & Ba) with Mean Squared Error (MSE) loss, unless otherwise specified. Training begins with a linear learning rate warm-up phase, followed by a systematic reduction of the learning rate using a decay factor whenever the validation loss stagnates. All experiments are conducted on either NVIDIA A800 Tensor Core GPUs. Our code is publicly available at <https://github.com/biomed-AI/MABNet>. Additional details on hyperparameter settings are provided in Appendix A.5.

Table 2. Mean absolute errors of energy (kcal/mol) and force (kcal/mol/Å) for 5 large-scale molecules on MD22.

| Model | Ac-Ala3-NHMe | | AT-AT | | AT-AT-CG-CG | | DHA | | Stachyose | |
|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | Energy | Forces |
| sGDML | 0.3902 | 0.7968 | 0.7235 | 0.6911 | 1.3885 | 0.7028 | 1.3117 | 0.7474 | 4.0497 | 0.6744 |
| PaiNN | 0.1168 | 0.2302 | 0.1673 | 0.2384 | 0.2638 | 0.3696 | 0.1151 | 0.1355 | 0.1517 | 0.2329 |
| ET | 0.1121 | 0.1879 | 0.1120 | 0.2036 | 0.2072 | 0.3259 | 0.1205 | 0.1209 | 0.1393 | 0.1921 |
| So3krates | 0.337 | 0.224 | 0.178 | 0.216 | 0.345 | 0.332 | 0.379 | 0.242 | 0.442 | 0.435 |
| Allegro | 0.1019 | 0.1068 | 0.1428 | 0.0952 | 0.3933 | 0.1280 | 0.1153 | 0.0732 | 0.2485 | 0.0971 |
| MACE | <u>0.0620</u> | 0.0876 | 0.1093 | 0.0992 | 0.1578 | 0.1153 | 0.1317 | 0.0646 | 0.1244 | 0.0876 |
| Equiformer | 0.0828 | <u>0.0804</u> | 0.1309 | 0.0960 | 0.1510 | 0.1252 | 0.1788 | <u>0.0506</u> | 0.1404 | <u>0.0635</u> |
| ViSNet | 0.0796 | 0.0972 | 0.1688 | 0.1070 | 0.1995 | 0.1563 | 0.1526 | 0.0668 | 0.1283 | 0.0869 |
| Equiformer-LSRM | 0.0780 | 0.0887 | 0.1007 | 0.0881 | <u>0.1335</u> | <u>0.1065</u> | 0.0878 | 0.0534 | 0.1252 | 0.0632 |
| ViSNet-LSRM | 0.0654 | 0.0902 | <u>0.0772</u> | <u>0.0781</u> | 0.1135 | 0.1063 | <u>0.0873</u> | 0.0598 | <u>0.1055</u> | 0.0767 |
| MABNet | 0.0534 | 0.0773 | 0.0637 | 0.0731 | 0.1862 | 0.1251 | 0.0618 | 0.0502 | 0.0895 | 0.0752 |

5.2. Results on MD22 Dataset

To evaluate the capacity of higher-order many-body interaction modeling in complex molecular systems, we first benchmarked our approach against the comprehensive MD22 dataset. As demonstrated in Table 2, MABNet achieved state-of-the-art performance, outperforming leading methods such as ViSNet and LSR-MP with an average 20% reduction in mean absolute errors (MAEs) across both energy and force predictions. Notably, MABNet attained the lowest MAEs in 7 of 10 metrics, including critical benchmarks for the glycoside DHA (56 atoms) and the 189-atom Stachyose system. This advancement arose from our many-body attention module, which directly encoded high-order interactions, bypassing the approximation errors inherent in existing frameworks that relied on sequential message-passing or aggregated lower-order expansions.

While MABNet achieved superior performance across most systems, its energy MAE for the 370-atom carbon nanotube (0.1862 kcal/mol) trailed ViSNet-LSRM (0.1135 kcal/mol). We attributed this gap to the dominance of pairwise interactions in symmetric systems with periodic geometries. In such cases, methods like LSR-MP, which prioritize scalable approximations of lower-order effects, gained an advantage by focusing on dominant pairwise terms.

MABNet achieved superior energy predictions across all systems, but force MAEs occasionally lagged behind the baselines (e.g., Equiformer-LSR on Stachyose forces: 0.0632 vs. MABNet’s 0.0752). This divergence reflected the sensitivity of force calculations, which depended on energy surface gradients, in high-order expansions. While our module captured global energy landscapes robustly, fine-grained force accuracy might have benefited from local geometric descriptors with subtle atomic displacements.

5.3. Results on SPICE Dataset

The SPICE dataset’s diversity provided a rigorous test for evaluating a model’s ability to generalize across molecular scales and interaction types. As shown in Table 3, our framework achieved state-of-the-art performance in all categories, demonstrating its capacity to unify many-body interaction modeling within a single architecture. Specifically, for dipeptides, where torsional flexibility and hydrogen bonding introduce complex many-body effects, our model achieved a force MAE of 7.7 meV/Å, outperforming TensorNet by 67%. This improvement highlighted the limitations of methods that approximate angular and torsional terms through sequential message-passing. In solvated amino acids, a regime demanding explicit modeling of solvent-solute polarization, our framework reduced force errors by 52% (31.9 vs. 66.4 meV/Å) compared to models like DimeNet++ and ET. These baselines, which lacked explicit higher-order interaction terms, struggled to resolve the dynamic interplay between solvent molecules and solute charge distributions. Critically, the cross-molecule robustness of our approach stemmed from the many-body attention module, which adaptively resolved both local (e.g., covalent bonds) and non-local (e.g., solvation shells) interactions through direct four-body communication. Unlike message-passing frameworks, our method avoided accuracy degradation in heterogeneous systems by explicitly encoding multi-body physics.

5.4. Ablation Study

To better understand our model, we conducted ablation studies focusing on model architecture and the combinations of many-body nodes. Table 4 summarized performance across three chemically distinct systems: dipeptides, sol-

Table 3. Mean absolute errors of energy (meV) and force (meV/Å) for 8 large-scale molecules on SPICE.

| Molecular | | SchNet | DimeNet++ | ET | TensorNet | MABNet |
|----------------------|--------|--------|-----------|-------------|--------------|-------------|
| PubChem Set 1 | Energy | 220.5 | 59.0 | <u>57.6</u> | 68.1 | 32.4 |
| | Forces | 150.7 | 59.1 | 88.3 | <u>49.4</u> | 26.5 |
| PubChem Set 2 | Energy | 222.5 | 60.6 | <u>55.8</u> | <u>32.5</u> | 27.2 |
| | Forces | 144.1 | 58.5 | 80.8 | <u>39.2</u> | 21.7 |
| PubChem Set 3 | Energy | 217.0 | 47.7 | 45.2 | <u>34.1</u> | 28.9 |
| | Forces | 138.9 | 45.9 | 65.9 | <u>36.2</u> | 20.5 |
| PubChem Set 4 | Energy | 299.9 | 44.6 | 40.6 | <u>32.8</u> | 26.7 |
| | Forces | 153.3 | 42.3 | 58.8 | <u>35.1</u> | 20.3 |
| PubChem Set 5 | Energy | 266.7 | 53.3 | 45.6 | <u>31.4</u> | 29.2 |
| | Forces | 144.8 | 52.7 | 69.7 | <u>35.1</u> | 20.6 |
| PubChem Set 6 | Energy | 242.9 | 32.8 | 30.7 | <u>23.8</u> | 17.3 |
| | Forces | 135.6 | 31.2 | 43.7 | <u>27.7</u> | 14.1 |
| Dipeptides | Energy | 386.2 | 169.9 | <u>26.0</u> | 28.7 | 9.6 |
| | Forces | 118.2 | 45.8 | 34.8 | <u>23.1</u> | 7.7 |
| Solvated Amino Acids | Energy | 1438.3 | 460.1 | 224.4 | <u>149.1</u> | 98.2 |
| | Forces | 318.1 | 173.7 | 118.2 | <u>66.4</u> | 31.9 |
| Average | Energy | 411.8 | 116.0 | 65.7 | <u>50.1</u> | 33.7 |
| | Forces | 163.0 | 63.7 | 70.0 | <u>39.0</u> | 20.4 |

Table 4. Mean absolute errors of energy and force for different attention mechanisms on Dipeptides and Solvated Amino Acids (energy in meV and force in meV/Å), and for Ac-Ala3-NHMe (energy in kcal/mol and force in kcal/mol/Å).

| Methods | Dipeptides | | Sol. Amino Acids | | Ac-Ala3-NHMe | |
|---------------------------|-------------|------------|------------------|-------------|--------------|--------------|
| | Energy | Forces | Energy | Forces | Energy | Forces |
| 3-body Attn. (w/o Equi.) | - | - | 151.5 | 59.4 | - | - |
| 3-body Attn. | 12.3 | 10.1 | 130.2 | 48.4 | 0.061 | 0.082 |
| 4-body Attn. (1.5N Edges) | 11.2 | 8.4 | 119.3 | 36.3 | 0.058 | 0.080 |
| 4-body Attn. (3N Edges) | <u>10.0</u> | <u>7.9</u> | <u>98.6</u> | <u>32.5</u> | <u>0.055</u> | <u>0.078</u> |
| 4-body Attn. (6N Edges) | 9.6 | 7.7 | 98.2 | 31.9 | 0.053 | 0.077 |

vated amino acids, and the tripeptide Ac-Ala3-NHMe.

Removing equivariance from the 3-body attention module (“3-body Attn. (w/o Equi.)”, similar to TGT) degraded force prediction accuracy on solvated amino acids by 23% (59.4 vs. 48.4 meV/Å), demonstrating that roto-translation equivariance was critical for modeling orientation-dependent interactions like hydrogen bonding and solvation. Extending to 4-body attention while maintaining equivariance further reduced force errors by 33% (48.4 vs. 32.5 meV/Å), confirming that higher-order terms were essential for resolving multi-center polarization effects.

Increasing the edge density in 4-body attention—from 1.5N to 6N edges—progressively improved performance, with force MAEs decreasing by 8% (7.9 vs. 7.7 meV/Å) for dipeptides and 2% (32.5 vs. 31.9 meV/Å) for sol-

vated systems. This showed that while even sparse 4-body graphs (1.5N edges) captured most many-body dependencies, denser connectivity refined accuracy by explicitly modeling weaker but chemically relevant interactions (e.g., van der Waals contacts in Ac-Ala3-NHMe). Crucially, our efficient attention mechanism minimized computational overhead, with 6N edges incurring only a 1.1× runtime increase over 1.5N, as shown in Table 5.

5.5. Computational Cost and Analysis

To further validate the efficiency of our approach, a tripeptide Ac-Ala3-NHMe was used to quantify computational efficiency. Table 5 compared the efficiencies ('3-body Attn. (w/o Equi.)', similar to TGT), 3-body attention, and various edge counts of 4-body attention in terms of memory consumption, training speed, and inference speed. The results

of the ablation study, presented in Table 4, revealed that the mean absolute error (MAE) for the energy and forces prediction using 4-body attention with $6N$ edges is markedly lower than that of the 3-body Attn. (w/o Equi.). Notably, this improvement in prediction accuracy came at a modest cost, with memory consumption increasing by only 6.8%, and no substantial degradation in training speed. Additionally, we performed a comparative analysis of training and inference costs with ViSNet. Our results indicated that the improvement in prediction accuracy came at a modest cost, with no significant degradation in training speed. A detailed analysis of computational efficiency is provided in Appendix B.2.

Table 5. Comparison of memory consumption, training speed for different attention mechanisms on the Ac-Ala3-NHMe dataset.

| Methods | Type of Many-body Attn. | Memory (MiB) | Training (it/s) |
|---------|-------------------------|--------------|-----------------|
| ViSNet | - | 15962 | 4.44 |
| | 3-body | 16542 | 3.93 |
| MABNet | 4-body (1.5N Edges) | 17264 | 2.78 |
| | 4-body (3N Edges) | 17400 | 2.66 |
| | 4-body (6N Edges) | 17758 | 2.35 |

6. Conclusion

By explicitly modeling four-body interactions through a geometric attention mechanism, MABNet advances molecular property prediction with quantum-mechanical accuracy, as demonstrated by state-of-the-art results on MD22 and SPICE. Future work will extend this framework to five-body and higher-order terms, enabling precise modeling of systems governed by entangled electronic states (e.g., transition metal catalysts) or cooperative solvation effects. To balance accuracy and scalability, we will explore adaptive quantum-inspired pruning strategies that prioritize dominant interactions in large systems. These advancements could redefine computational chemistry workflows, accelerating drug discovery through simulations that close the gap between empirical and ab initio accuracy.

Acknowledgements

This study has been supported by Guangdong S&T Program (2024B0101040005), Guangdong S&T Program (2023B1111030002), Shenzhen Science and Technology Plan Project (CJGJZD20220517142201004), and Young Science and Technology Talent Support Program of Guangdong Precision Medicine Application Association (YSTTGDPMAA202502).

Impact Statement

Our work aims to improve chemical property prediction, which can accelerate the discovery of new beneficial materials and drugs. This has the potential to greatly benefit society by enabling the development of cheaper, safer, and more effective medicines. We would make the code open-source so that domain experts can further improve and validate the models before real-world deployment.

References

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.
- Ambainis, A., Filmus, Y., and Le Gall, F. Fast matrix multiplication: limitations of the coppersmith-winograd method. In *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*, pp. 585–593, 2015.
- Batatia, I., Kovács, D. P., Simm, G. N., Ortner, C., and Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields (2022). URL <https://arxiv.org/abs/2206.07697>.
- Chen, J., Zheng, S., Song, Y., Rao, J., and Yang, Y. Learning attributed graph representation with communicative message passing transformer. In Zhou, Z.-H. (ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 2242–2248. International Joint Conferences on Artificial Intelligence Organization, 8 2021.
- Chmiela, S., Tkatchenko, A., Sauceda, H. E., Poltavsky, I., Schütt, K. T., and Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Science advances*, 3(5):e1603015, 2017.
- Chmiela, S., Vassilev-Galindo, V., Unke, O. T., Kabylda, A., Sauceda, H. E., Tkatchenko, A., and Müller, K.-R. Accurate global machine learning force fields for molecules with hundreds of atoms. *Science Advances*, 9 (2):eadf0873, 2023.
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simonett, A. C., Harrigan, M. P., Stern, C. D., et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7): e1005659, 2017.
- Eastman, P., Behara, P. K., Dotson, D. L., Galvelis, R., Herr, J. E., Horton, J. T., Mao, Y., Chodera, J. D., Pritchard, B. P., Wang, Y., et al. Spice, a dataset of drug-like

- molecules and peptides for training machine learning potentials. *Scientific Data*, 10(1):11, 2023.
- Fourches, D., Muratov, E., and Tropsha, A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and qsar modeling research. *Journal of chemical information and modeling*, 50(7):1189, 2010.
- Frank, T., Unke, O., and Müller, K.-R. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. *Advances in Neural Information Processing Systems*, 35:29400–29413, 2022.
- Gasteiger, J., Giri, S., Margraf, J. T., and Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020a.
- Gasteiger, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020b. URL <https://openreview.net/forum?id=B1eWbxStPH>.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. PMLR, 2017.
- Hansen, K., Biegler, F., Ramakrishnan, R., Pronobis, W., Von Lilienfeld, O. A., Muller, K.-R., and Tkatchenko, A. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *The journal of physical chemistry letters*, 6(12):2326–2331, 2015.
- Hussain, M. S., Zaki, M. J., and Subramanian, D. Triplet interaction improves graph transformers: accurate molecular graph learning with triplet graph transformers. In *Proceedings of the 41st International Conference on Machine Learning*, ICML’24. JMLR.org, 2024.
- Jaeger, S., Fulle, S., and Turk, S. Mol2vec: unsupervised machine learning approach with chemical intuition. *Journal of chemical information and modeling*, 58(1):27–35, 2018.
- Joshi, C. K., Bodnar, C., Mathis, S. V., Cohen, T., and Lio, P. On the expressive power of geometric graph neural networks. In *International conference on machine learning*, pp. 15330–15355. PMLR, 2023.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.
- Kingma, D. P. and Ba, J. In Bengio, Y. and LeCun, Y. (eds.), *ICLR (Poster)*.
- Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.
- Li, Y., Wang, Y., Huang, L., Yang, H., Wei, X., Zhang, J., Wang, T., Wang, Z., Shao, B., and Liu, T.-Y. Long-short-range message-passing: A physics-informed framework to capture non-local interaction for scalable molecular dynamics simulation. *arXiv preprint arXiv:2304.13542*, 2023.
- Liao, Y.-L. and Smidt, T. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=KwmPfARgOTD>.
- Liu, L., He, D., Fang, X., Zhang, S., Wang, F., He, J., and Wu, H. Gem-2: Next generation molecular property prediction network by modeling full-range many-body interactions. *arXiv preprint arXiv:2208.05863*, 2022.
- Liu, T., Lin, Y., Wen, X., Jorissen, R. N., and Gilson, M. K. Bindingdb: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic acids research*, 35(suppl_1):D198–D201, 2007.
- Lu, W., Wu, Q., Zhang, J., Rao, J., Li, C., and Zheng, S. Tankbind: Trigonometry-aware neural networks for drug–protein binding structure prediction. *Advances in neural information processing systems*, 35:7236–7249, 2022.
- Maier, J., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K., and Simmerling ff14SB, C. Improving the accuracy of protein side chain and backbone parameters from ff99sb., 2015, 11. DOI: <https://doi.org/10.1021/acs.jctc.5b00255>, pp. 3696–3713.
- Montavon, G., Hansen, K., Fazli, S., Rupp, M., Biegler, F., Ziehe, A., Tkatchenko, A., Lilienfeld, A., and Müller, K.-R. Learning invariant representations of molecules for atomization energy prediction. *Advances in neural information processing systems*, 25, 2012.
- Musaelian, A., Batzner, S., Johansson, A., Sun, L., Owen, C. J., Kornbluth, M., and Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raihan, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative

- style, high-performance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.
- Perdew, J. P., Burke, K., and Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.
- Rao, J., Zheng, S., Lu, Y., and Yang, Y. Quantitative evaluation of explainable graph neural networks for molecular property prediction. *Patterns*, 3(12), 2022a.
- Rao, J., Zheng, S., Mai, S., and Yang, Y. Communicative subgraph representation learning for multi-relational inductive drug-gene interaction prediction. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pp. 3919–3925, 2022b.
- Rao, J., Xie, J., Lin, H., Zheng, S., Wang, Z., and Yang, Y. Incorporating retrieval-based causal learning with information bottlenecks for interpretable graph neural networks. *arXiv preprint arXiv:2402.04710*, 2024a.
- Rao, J., Xie, J., Yuan, Q., Liu, D., Wang, Z., Lu, Y., Zheng, S., and Yang, Y. A variational expectation-maximization framework for balanced multi-scale learning of protein and drug interactions. *Nature Communications*, 15(1):4476, 2024b.
- Rogers, D. and Hahn, M. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.
- Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.
- Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.
- Schütt, K. T., Saucedo, H. E., Kindermans, P.-J., Tkatchenko, A., and Müller, K.-R. Schnet—a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24), 2018.
- Simeon, G. and De Fabritiis, G. Tensornet: Cartesian tensor representations for efficient learning of molecular potentials. *Advances in Neural Information Processing Systems*, 36, 2024.
- Song, Y., Zheng, S., Niu, Z., Fu, Z.-H., Lu, Y., and Yang, Y. Communicative representation learning on attributed molecular graphs. In *29th International Joint Conference on Artificial Intelligence and the 17th Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI2020)*. International Joint Conferences on Artificial Intelligence Organization, 2020.
- Thölke, P. and De Fabritiis, G. Torchmd-net: equivariant transformers for neural network based molecular potentials. *arXiv preprint arXiv:2202.02541*, 2022.
- Thölke, P. and Fabritiis, G. D. Equivariant transformers for neural network based molecular potentials. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=zNHqZ9wrRB>.
- Tkatchenko, A., DiStasio Jr, R. A., Car, R., and Scheffler, M. Accurate and efficient method for many-body van der waals interactions. *Physical review letters*, 108(23):236402, 2012.
- Unke, O. T. and Meuwly, M. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of chemical theory and computation*, 15(6):3678–3693, 2019.
- Unke, O. T., Chmiela, S., Saucedo, H. E., Gastegger, M., Poltavsky, I., Schütt, K. T., Tkatchenko, A., and Müller, K.-R. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021.
- Vaswani, A. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- Wang, Y., Li, S., Wang, T., Shao, B., Zheng, N., and Liu, T.-Y. Geometric transformer with interatomic positional encoding. *Advances in Neural Information Processing Systems*, 36:55981–55994, 2023a.
- Wang, Y., Wang, T., Li, S., He, X., Li, M., Wang, Z., Zheng, N., Shao, B., and Liu, T.-Y. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nature Communications*, 15(1):313, 2024.
- Wang, Z., Liu, G., Zhou, Y., Wang, T., and Shao, B. Quinnet: efficiently incorporating quintuple interactions into geometric deep learning force fields. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pp. 77043–77055, 2023b.
- Wu, H., Wu, L., Liu, G., Liu, Z., Shao, B., and Wang, Z. Se3set: Harnessing equivariant hypergraph neural networks for molecular representation learning. *arXiv preprint arXiv:2405.16511*, 2024.
- Yang, Y. and Zhou, Y. Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins: Structure, Function, and Bioinformatics*, 72(2):793–803, 2008.

A. Additional Details

A.1. Further Details on Feature Embedding

Radial Basis Expansion Interatomic distances are projected into a higher-dimensional space using exponential normal radial basis functions (RBFs), coupled with a cosine cutoff function to ensure smooth behavior near the cutoff distance.

The distance d_{ij} between atoms i and j is first expanded into a higher-dimensional space using a set of exponential normal radial basis functions (RBF) (Unke & Meuwly, 2019; Thölke & De Fabritiis, 2022). Additionally, a cosine cutoff function is applied to d_{ij} to ensure smoothness. The vector d^{rbf} consists of the values of n radial basis functions defined as:

$$d_{ij}^{rbf} = \phi(d_{ij}) \cdot \exp\left(-\beta_n (\exp(\alpha \cdot (-d_{ij})) - \mu_n)^2\right), \quad (21)$$

$$\phi(d_{ij}) = \begin{cases} \frac{1}{2} \cdot \cos\left(\pi \cdot \frac{d_{ij}}{d_{cut}}\right) + 1, & \text{if } d_{ij} < d_{cut}, \\ 0, & \text{if } d_{ij} \geq d_{cut}. \end{cases} \quad (22)$$

where β_n and μ_n represent the width and mean parameters of the radial basis function n , respectively. The parameter β_n controls the sensitivity of the radial basis function, while μ_n defines its center. α is a scaling factor that adjusts the rate of exponential decay. The width parameter β_n is set to $(2n^{-1}(1 - \exp(-d_{cut})))^{-2}$, following the approach described in Unke & Meuwly (2019).

Scalar and Vector Node Feature Embedding The scalar feature x_i of node i is derived from the initial embedding which encodes the atom type z_i , and the aggregated features from its neighboring nodes. First, node i maps its atom type z_i to a high-dimensional space using the embedding function f_i^{emb} , resulting in the initial node feature $h_i = f_i^{emb}(z_i)$.

Next, node i gathers information from all its neighboring nodes j . The distance d_{ij} between nodes i and j is first truncated using the cutoff function $\phi(d_{ij})$, as described in Equation (22), to limit the influence of long-range interactions. A linear transformation is then applied to the edge features d_{ij}^{rbf} , as defined in Equation (21), followed by multiplying the result with the initial feature h_j of the neighboring node j . The aggregated neighbor features x_{nbh} for node i are obtained by summing the contributions from all its neighbors:

$$x_{nbh} = \sum_{j \in \mathcal{N}(i)} (\phi(d_{ij}) \cdot (\mathbf{W}_n d_{ij}^{rbf} + \mathbf{b}_n) \odot h_j), \quad (23)$$

where \odot denotes element-wise multiplication. \mathbf{W}_n and \mathbf{b}_n are learnable parameters, and $\mathcal{N}(i)$ denotes the set of neighboring nodes of node i . Finally, the scalar feature x_i of node i is computed by concatenating its initial feature h_i with the aggregated neighbor features x_{nbh} , followed by a linear transformation:

$$x_i = \mathbf{W}_s(h_i \parallel x_{nbh}) + \mathbf{b}_s, \quad (24)$$

where the \parallel denotes the concatenation operation. \mathbf{W}_s and \mathbf{b}_s are learnable parameters.

The vector feature \vec{v}_i of node i is initialized as $\vec{0}$.

Edge Feature Embedding The edge features e_{ij} are computed by combining the scalar features of nodes i and j , obtained from Equation (24), with the projected edge attributes d_{ij}^{rbf} , derived from Equation (21). First, d_{ij}^{rbf} is transformed into a higher-dimensional space using a linear projection. Then, the edge features are calculated as:

$$e_{ij} = (x_i + x_j) \odot (\mathbf{W}_r d_{ij}^{rbf} + \mathbf{b}_r), \quad (25)$$

where \mathbf{W}_r and \mathbf{b}_r are learnable parameters.

A.2. Quadruple Attention for Dihedral Angles

Our Many-body Interaction method employs Quadruple Attention to directly capture four-body interactions, specifically the relationships among nodes i, j, k, l . These four atoms form a dihedral angle, as illustrated in Figure 2. In chemistry, a dihedral angle is defined as the angle between two planes, each formed by three atoms, involving a total of four atoms. The

common edge of these planes corresponds to a chemical bond (atoms k and l), while the planes themselves are determined by the additional atoms i and j interacting with this bond.

Dihedral torsional potential serves as a crucial descriptor of energy variations in a non-collinear four-atom molecular system. Compared to bond stretching and bond angle bending potentials, the dihedral torsional potential is relatively weaker but plays a decisive role in molecular conformational changes. This is because variations in dihedral angles can lead to large-scale rearrangements of molecular chains, ultimately influencing the overall shape and function of a molecule. For instance, in biomacromolecules such as proteins and DNA, dihedral angle torsion dictates their three-dimensional conformations, which are essential for their biological functions. Our approach enables comprehensive dihedral angle information capture through direct four-body interactions, making it highly relevant for molecular dynamics simulations.

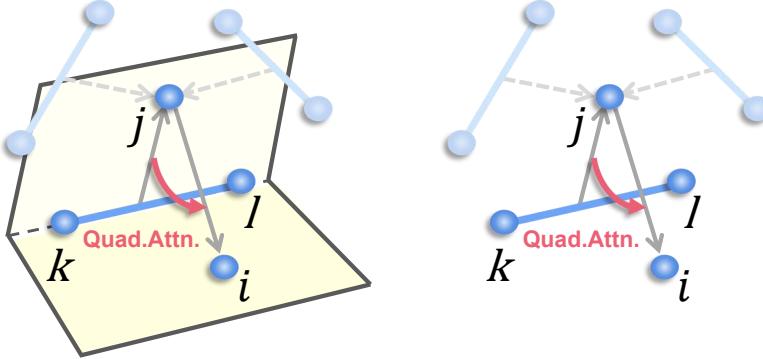


Figure 2. Quadruple Attention-Based Visualization of the Dihedral Angle Formation by Atoms i, j, k, l .

A.3. Datasets

A.3.1. MD22

MD22 is a newly developed molecular dynamics benchmark dataset introduced by (Chmiela et al., 2023), designed to provide novel challenges for the research and development of Machine Learning Force Fields (MLFFs). The dataset includes a variety of molecular systems ranging from tens to hundreds of atoms, encompassing biomolecules such as simple peptide chains and complex carbohydrates. These systems provide diverse training and testing scenarios for MLFF models. The primary goal of MD22 is to enhance the accuracy and scalability of MLFF models, particularly in dealing with large molecular systems that exhibit long-range interactions and complex molecular dynamics behaviors. The molecular dynamics trajectories in the dataset are sampled within a temperature range of 400 K to 500 K, with a time resolution of 1 femtosecond (fs). The potential energy and atomic forces for each system are computed using the PBE+MBD (Perdew et al., 1996; Tkatchenko et al., 2012) theory level. Compared to the earlier MD17 dataset, the molecular systems in MD22 are larger and more flexible, making them better suited for simulating long-range interactions and complex molecular dynamics behaviors. Furthermore, MD22 offers greater molecular flexibility, allowing for the simulation of complex dynamic processes such as molecular vibrations, rotations, and conformational changes. Table 6 provides an overview of the atomic counts, molecular formulas, number of conformations, and the ranges of energy and force values for the systems in MD22.

Table 6. Properties of the MD22 datasets. Energies are in kcal/mol, forces in kcal/mol/Å.

| Dataset | Atoms | Formula | Conformations | Energies | | Forces | |
|--------------|-------|--|---------------|----------|----------|--------|----------|
| | | | | Range | Variance | Range | Variance |
| Ac-Ala3-NHMe | 42 | C ₁₂ H ₂₂ N ₄ O ₄ | 85,109 | 102.19 | 67.30 | 437.99 | 678.02 |
| DHA | 56 | C ₂₂ H ₃₂ O ₂ | 69,753 | 75.71 | 93.31 | 420.07 | 673.99 |
| Stachyose | 83 | C ₂₄ H ₄₂ O ₂₁ | 27,272 | 106.15 | 189.33 | 426.08 | 655.55 |
| AT-AT | 60 | C ₂₀ H ₂₂ N ₁₄ O ₄ | 20,001 | 139.29 | 120.09 | 444.01 | 779.81 |
| AT-AT-CG-CG | 118 | C ₃₈ H ₄₂ N ₃₀ O ₈ | 10,153 | 243.50 | 246.64 | 407.55 | 768.13 |

A.3.2. MD17

MD17, proposed by (Chmiela et al., 2017), is a molecular dynamics dataset specifically designed to support the development of MLFFs. It includes ab initio molecular dynamics (AIMD) trajectories for a variety of small organic molecules across multiple conformations. These molecules include, but are not limited to, benzene, toluene, naphthalene, ethanol, uracil, and aspirin, offering a range of sizes and complex potential energy surfaces. Each molecular dataset contains between 150,000 and nearly 1 million conformational geometries, sampled at a time resolution of 0.5 femtoseconds (fs). The energy and force labels for all systems were computed using the PBE+vdW-TS electronic structure method. The MD17 dataset was designed to facilitate the development and validation of novel machine learning algorithms aimed at constructing accurate and energy-conserving molecular force fields. It has been widely used to demonstrate the efficiency and accuracy of machine learning models in predicting forces and energies, particularly when achieving high predictive accuracy with a limited number of training samples. Table 7 summarizes the atomic counts, molecular formulas, number of conformations, as well as the energy and force ranges for the molecules included in MD17.

Table 7. Properties of the MD17 datasets. Energies are in kcal/mol, forces in kcal/mol/Å.

| Dataset | Atoms | Formula | Conformations | Energies | | Forces | |
|----------------|-------|---|---------------|----------|----------|--------|----------|
| | | | | Range | Variance | Range | Variance |
| Ethanol | 9 | C ₂ H ₆ O | 555,092 | 36.61 | 17.57 | 432.00 | 689.43 |
| Malonaldehyde | 9 | C ₃ H ₄ O ₂ | 993,237 | 43.83 | 17.14 | 570.65 | 820.10 |
| Uracil | 12 | C ₄ H ₄ N ₂ O ₂ | 133,770 | 39.92 | 24.24 | 476.63 | 901.43 |
| Toluene | 15 | C ₇ H ₈ | 442,790 | 52.08 | 26.01 | 425.60 | 746.91 |
| Salicylic acid | 16 | C ₇ H ₆ O ₃ | 320,231 | 47.47 | 29.34 | 453.77 | 817.37 |
| Aspirin | 21 | C ₉ H ₈ O ₄ | 211,762 | 55.29 | 35.36 | 423.87 | 779.99 |

A.3.3. SPICE

SPICE is a quantum chemistry dataset specifically designed for training machine learning potentials, introduced by (Eastman et al., 2023). It focuses on modeling interactions between drug molecules and proteins. The dataset includes diverse conformations of drug molecules, dimers, dipeptides, and solvated amino acids, with tens of thousands to hundreds of thousands of conformations for each category, as shown in Table 8. It spans more than a dozen chemical elements and incorporates both charged and neutral molecules, providing a comprehensive sampling of covalent and non-covalent interactions. All energies and forces are calculated using the ω B97M-D3(BJ)/def2-TZVPPD level of theory, alongside additional useful information such as multipole moments and bond orders. Additionally, SPICE includes other valuable quantum chemical properties, such as partial charges and multipole moments, which can be used to train different types of models or enhance potential models through multitask learning.

SPICE comprises multiple subsets, each designed to provide specific types of information. To validate the efficacy of the proposed many-body interaction approach in capturing the intricate details of large molecular systems, subsets with a higher number of atoms were selected from the SPICE dataset for training and testing. Specifically, these subsets include Dipeptides, Solvated Amino Acids, and PubChem Molecules. These subsets encompass a variety of chemical structures ranging from dipeptides to complex drug-like molecules, thereby providing a comprehensive evaluation of the model’s performance in handling large molecular systems.

Dipeptides The Dipeptides subset provides extensive coverage of covalent interactions in proteins. It includes 676 dipeptides composed of all possible combinations of 20 natural amino acids and their common protonation states, as well as one pair of cysteine residues linked by a disulfide bond, resulting in 677 dipeptides. Each dipeptide is capped with ACE and NME groups and has 50 conformations: half representing low-energy states and the other half high-energy states. These conformations were generated using RDKit to create initial structures, followed by molecular dynamics simulations with OpenMM 7.6 (Eastman et al., 2017) and the Amber14 (Maier et al.) force field at varying temperatures. This subset is sufficient to sample all covalent interaction types found in naturally occurring proteins, excluding special cases such as post-translational modifications.

Solvated Amino Acids The Solvated Amino Acids subset focuses on sampling non-covalent interactions between proteins and water, as well as water-water interactions, which are crucial for protein simulations. This subset includes the 26 amino

acid variants mentioned earlier, also capped with ACE and NME groups, surrounded by 20 TIP3P-FB water molecules. For each amino acid, 50 conformations were generated. Each amino acid was placed in a cubic water box with a side length of 2.2 nanometers, and a 1-nanosecond molecular dynamics simulation was performed at 300 K, saving conformations every 20 picoseconds. The 20 water molecules closest to the amino acid were retained for each conformation, while the rest were discarded. This allows researchers to study how the local water environment around amino acids influences protein behavior effectively.

PubChem Molecules The PubChem Molecules subset comprises a large and diverse collection of drug-like small molecules. Approximately 1.5 million entries were downloaded from the PubChem database, specifically from BindingDB (Liu et al., 2007) or ChemIDplus (cited in the U.S. National Library of Medicine databases). A filtering process was applied to select as diverse a sample as possible based on Tanimoto similarity using ECFP4 fingerprints. For each molecule, 50 conformations were generated using the same procedure as the dipeptides to obtain both low- and high-energy states. This subset is designed to provide broad chemical diversity, enabling models to better understand and predict molecular behavior in complex and varied chemical environments.

Table 8. Properties of the SPICE datasets.

| Subset | Atoms (avg.) | Molecules | Conformations | Elements |
|----------------------|--------------|-----------|---------------|-------------|
| Solvated Amino Acids | 88.12 | 26 | 1,300 | |
| Dipeptides | 44.23 | 677 | 33,850 | H,C,N,O,S |
| PubChem Set 1 | 33.78 | 2372 | 114,359 | |
| PubChem Set 2 | 37.05 | 2431 | 115,997 | |
| PubChem Set 3 | 37.46 | 2446 | 118,931 | H,C,N,O,F |
| PubChem Set 4 | 37.70 | 2455 | 118,486 | P,S,Cl,Br,I |
| PubChem Set 5 | 37.02 | 2463 | 117,841 | |
| PubChem Set 6 | 38.49 | 2476 | 123,786 | |

A.4. Baselines

A.4.1. BASELINES OF MD22

ViSNet (Wang et al., 2024) is an equivariant geometry-enhanced graph neural network designed to efficiently extract and utilize geometric features for molecular modeling. By leveraging equivariance principles, ViSNet effectively integrates geometric information into its architecture, enabling precise modeling of molecular structures with reduced computational costs. This approach enhances the representation of molecular conformations and provides interpretability by mapping geometric representations to molecular structures, making it a powerful tool for addressing challenges in molecular dynamics and related fields.

Equiformer (Liao & Smidt, 2023) is a graph neural network that integrates the strengths of Transformer architectures with SE(3)/E(3)-equivariant features based on irreducible representations (irreps). By replacing standard Transformer operations with their equivariant counterparts and incorporating tensor products, Equiformer effectively encodes equivariant information within irreps feature channels without adding complexity to graph structures. Furthermore, it introduces a novel equivariant graph attention mechanism, which replaces traditional dot product attention with multi-layer perceptron attention and incorporates non-linear message passing, enhancing its ability to capture geometric and relational information in 3D atomistic graphs. With these innovations, Equiformer provides a simple yet powerful architecture for modeling 3D molecular systems.

MACE (Batatia et al.) is a state-of-the-art machine learning force field architecture designed for a wide range of in-domain, extrapolation, and low-data regime tasks. By employing a strictly local atom-centered model, MACE achieves high data efficiency and is capable of accurately modeling diverse systems, including amorphous carbon, universal materials, small organic molecules, large molecular systems, and weakly interacting molecular assemblies. Its architecture excels in tasks such as constrained geometry optimization and molecular dynamics simulations, showcasing its versatility and ability to handle complex physical and chemical phenomena.

LSR-MP (Li et al., 2023) is a novel framework that generalizes existing equivariant graph neural networks (EGNNs) by efficiently incorporating long-range interactions while maintaining an effective description of many-body interactions.

Inspired by fragmentation-based methods, LSR-MP aims to address the limitations of traditional machine learning and fragmentation approaches in modeling chemical and biological systems. By extending the message-passing paradigm to capture both short- and long-range interactions, this framework enhances the modeling of complex molecular interactions and can be seamlessly integrated into existing EGNN architectures. Its general applicability and robustness are demonstrated through consistent performance improvements across various EGNNs, including its application to ViSNet(ViSNet-LSRM) and Equiformer(Equiformer-LSRM).

Allegro (Musaelian et al., 2023) is a strictly local equivariant deep neural network interatomic potential architecture designed to achieve both high accuracy and scalability. Unlike traditional atom-centered message passing neural networks (MPNNs), which are limited by the range of their information propagation, Allegro avoids message passing altogether. Instead, it uses iterated tensor products of learned equivariant representations to model many-body interactions. This approach allows Allegro to retain the accuracy of many-body potentials while maintaining the scalability of local methods, making it a powerful tool for modeling the potential energy surfaces of molecules and materials.

So3krates (Frank et al., 2022) is a self-attention-based message passing neural network designed to capture non-local quantum mechanical effects in molecules and materials. It introduces spherical harmonic coordinates (SPHCs) to encode higher-order geometric information, enabling a non-local attention mechanism in SPHC space. By decoupling geometric information from atomic features, So3krates constructs spherical filters that extend continuous filters to SPHC space, forming the foundation for its spherical self-attention mechanism. This allows So3krates to model non-local effects over arbitrary length scales with high data efficiency and generalization. It achieves state-of-the-art performance while being significantly faster and more parameter-efficient than other models.

TorchMD-NET(ET) (Thölke & De Fabritiis, 2022) is a versatile framework for molecular dynamics simulations that integrates classical and machine learning potentials. By expressing all force computations—such as bond, angle, dihedral, Lennard-Jones, and Coulomb interactions—as PyTorch operations, TorchMD enables seamless compatibility with machine learning techniques. TorchMD’s capabilities are validated through diverse applications, including standard Amber all-atom simulations, learning ab initio potentials, and coarse-grained protein folding models.

PaiNN (Schütt et al., 2021) is a message passing neural network that extends traditional formulations by incorporating rotationally equivariant representations, addressing limitations of invariant approaches in data efficiency. By leveraging equivariant atomwise representations, PaiNN improves the prediction of chemical properties and tensorial quantities while achieving state-of-the-art performance on molecular benchmarks. Its architecture reduces model size and inference time, making it both accurate and efficient. PaiNN demonstrates its capability by simulating molecular spectra with speedups of 4–5 orders of magnitude compared to electronic structure methods, showcasing its potential for accelerating molecular dynamics and quantum chemistry studies.

sGML (Chmiela et al., 2023) is a global machine learning force field designed to capture collective many-atom interactions in molecular systems without introducing locality assumptions or uncontrolled approximations. Unlike traditional approaches, sGML preserves full correlation among all atomic degrees of freedom, allowing it to accurately model complex molecules and materials with far-reaching interaction lengths. Using an exact, iterative, and parameter-free training method, sGML scales to systems with up to several hundred atoms while retaining the accuracy of global force fields. Its performance is demonstrated on the MD22 benchmark dataset (42–370 atoms) and in robust nanosecond-scale path-integral molecular dynamics simulations for supramolecular complexes.

A.4.2. BASELINES OF SPICE

TensorNet (Simeon & De Fabritiis, 2024) is an O(3)-equivariant message-passing neural network designed for efficient and accurate representation of molecular systems. It leverages Cartesian tensor atomic embeddings, simplifying feature mixing through matrix product operations. By decomposing tensors into rotation group irreducible representations, TensorNet processes scalars, vectors, and tensors separately when needed, making it highly efficient. Compared to spherical tensor models, TensorNet achieves state-of-the-art performance with significantly fewer parameters, even with a single interaction layer for small molecular potential energies. Its framework also enables accurate predictions of vector and tensor molecular properties alongside potential energies and forces, all while reducing computational cost. TensorNet thus provides a flexible and powerful foundation for designing advanced equivariant models.

DimeNet++ (Gasteiger et al., 2020a) is a machine learning model for molecular property prediction that builds upon a directional message-passing mechanism to capture intricate angular dependencies in molecular systems. It improves

Table 9. Hyperparameters used for MD22, SPICE and MD17.

| Parameter | MD22 | SPICE | MD17 |
|-------------------------|------|-------|------|
| initial LR | 1e-4 | 1e-4 | 3e-4 |
| min LR | 1e-7 | 1e-7 | 1e-7 |
| LR warm up steps | 1000 | 1000 | 1000 |
| LR decay factor | 0.8 | 0.8 | 0.8 |
| LR patience (epochs) | 30 | 30 | 30 |
| optimizer | Adam | Adam | Adam |
| energy loss weight | 0.05 | 1.0 | 0.05 |
| forces loss weight | 0.95 | 10.0 | 0.95 |
| embedding dimension | 256 | 256 | 256 |
| attention heads | 8 | 8 | 8 |
| batch size | 2,4 | 2,4 | 4 |
| number of layers | 9 | 9 | 9 |
| number of RBFs | 32 | 32 | 32 |
| cutoff (\AA) | 5.0 | 5.0 | 5.0 |

 Table 10. Mean absolute errors of energy (meV) and force (meV/ \AA) for 7 small molecules on MD17.

| Molecular | SchNet | DimeNet++ | PaiNN | SpookyNet | GemNet | ET | NequIP | SO3KRATES | ViSNet | Equiformer | MABNet |
|----------------|--------|-----------|-------|-----------|--------|-------|--------------|--------------|--------|--------------|--------------|
| Aspirin | Energy | 0.37 | 0.204 | 0.167 | 0.151 | - | 0.123 | 0.131 | 0.139 | <u>0.116</u> | 0.122 |
| | Forces | 1.35 | 0.499 | 0.338 | 0.258 | 0.217 | 0.253 | 0.184 | 0.236 | <u>0.155</u> | 0.152 |
| Ethanol | Energy | 0.08 | 0.064 | 0.064 | 0.052 | - | 0.052 | <u>0.051</u> | 0.061 | <u>0.051</u> | 0.039 |
| | Forces | 0.39 | 0.230 | 0.224 | 0.094 | 0.085 | 0.109 | 0.071 | 0.096 | 0.060 | <u>0.067</u> |
| Malonaldehyde | Energy | 0.13 | 0.104 | 0.091 | 0.079 | - | 0.077 | 0.076 | 0.077 | 0.075 | <u>0.074</u> |
| | Forces | 0.66 | 0.383 | 0.319 | 0.167 | 0.155 | 0.169 | 0.129 | 0.147 | 0.100 | <u>0.125</u> |
| Naphthalene | Energy | 0.16 | 0.122 | 0.116 | 0.116 | - | 0.085 | 0.113 | 0.115 | <u>0.085</u> | <u>0.085</u> |
| | Forces | 0.58 | 0.215 | 0.077 | 0.089 | 0.051 | 0.061 | 0.039 | 0.074 | 0.039 | <u>0.046</u> |
| Salicylic Acid | Energy | 0.20 | 0.134 | 0.116 | 0.114 | - | 0.093 | 0.106 | 0.106 | <u>0.092</u> | 0.099 |
| | Forces | 0.85 | 0.374 | 0.195 | 0.180 | 0.125 | 0.129 | 0.090 | 0.145 | 0.084 | <u>0.087</u> |
| Toluene | Energy | 0.12 | 0.102 | 0.095 | 0.094 | - | 0.074 | 0.092 | 0.095 | <u>0.074</u> | 0.085 |
| | Forces | 0.57 | 0.216 | 0.094 | 0.087 | 0.060 | 0.067 | 0.046 | 0.073 | 0.039 | <u>0.048</u> |
| Uracil | Energy | 0.14 | 0.115 | 0.106 | 0.105 | - | <u>0.095</u> | 0.104 | 0.103 | <u>0.095</u> | 0.099 |
| | Forces | 0.56 | 0.301 | 0.139 | 0.119 | 0.097 | <u>0.095</u> | 0.076 | 0.111 | 0.062 | <u>0.067</u> |

computational efficiency and simplifies the architecture compared to its predecessor, DimeNet. Additionally, DimeNet++ incorporates ensembling and mean-variance estimation techniques to enable uncertainty quantification, making it suitable for exploring complex molecular configurations, including non-equilibrium structures.

SchNet (Schütt et al., 2018) is a deep learning architecture designed for modeling quantum interactions in molecules, leveraging continuous-filter convolutional layers to capture local correlations without relying on grid-structured data. By directly utilizing the precise locations of atoms, SchNet preserves essential physical information and ensures rotationally invariant energy predictions. The model jointly predicts total energies and interatomic forces, yielding a smooth, differentiable potential energy surface according to quantum-chemical principles.

A.5. Hyper-parameters

To account for the differences in dataset sizes and data distribution, distinct hyperparameter configurations were selected for the MD22, SPICE, and MD17 datasets, as summarized in Table 9. The Adam optimizer was chosen as the unified optimization strategy, and the energy loss weight and forces loss weight were adjusted based on the specific characteristics of each dataset. In particular, for the SPICE dataset, which involves more intricate intermolecular interactions, the force loss weight was increased to enhance the model’s sensitivity to subtle variations in the force field. This adjustment aims to improve the model’s performance in handling complex molecular systems.

B. Additional Experiments

B.1. Results of MD17

The performance of MABNet on the **MD17** dataset was presented in Table 10. It achieved higher accuracy than models like So3krates and is comparable to ViSNet and Equiformer. MABNet showed no significant improvements compared to other state-of-the-art (SOTA) models, as higher-order many-body interactions were not pronounced in small molecular systems.

B.2. Computational Cost and Analysis

To further validate the efficiency of our approach, a tripeptide Ac-Ala3-NHMe was used to quantify computational efficiency. Table 11 summarized a comparative study across different configurations: 3-body attention without equivariance (denoted as '3-body Attn. (w/o Equi.)', akin to TGT), standard 3-body attention, and several variants of 4-body attention with varying edge counts. All experiments were conducted on an NVIDIA GeForce RTX 4090 GPU with a batch size of 4. The results, presented in Table 4, showed that 4-body attention with $6N$ edges significantly reduces the mean absolute error (MAE) for both energy and force predictions, compared to the 3-body Attn. (w/o Equi.). This improvement in accuracy was achieved with only a 6.8% increase in memory consumption and no notable degradation in training speed. Furthermore, we conducted a comparative analysis of training and inference costs with two representative baselines, Geoformer (Wang et al., 2023a) and ViSNet (Wang et al., 2024). Our method exhibited a favorable trade-off between accuracy and computational overhead, demonstrating that the proposed 4-body attention mechanism offered improved predictive performance without compromising training or inference efficiency.

Table 11. Comparison of memory consumption, training and inference speeds for different attention mechanisms on the Ac-Ala3-NHMe dataset.

| Methods | Type of Many-body Attn. | Memory | Training | Inference |
|-----------|----------------------------|--------|----------|-----------|
| | | (MiB) | (it/s) | (it/s) |
| Geoformer | - | 14362 | 5.09 | 13.58 |
| ViSNet | - | 15962 | 4.44 | 11.66 |
| MABNet | 3-body Attn. | 16542 | 3.93 | 10.65 |
| | 4-body Attn. (1.5N Edges) | 17264 | 2.78 | 6.77 |
| | 4-body Attn. (3N Edges) | 17400 | 2.66 | 6.62 |
| | 4-body Attn. (6N Edges) | 17758 | 2.35 | 5.91 |

B.3. Additional Ablation Atudy Comparing ViSNet and MABNet on Ac-Ala3-NHMe.

To further investigate the impact of the many-body attention mechanism in MABNet, we conducted ablation studies comparing its performance with ViSNet on Ac-Ala3-NHMe. ViSNet used a runtime geometry calculation (RGC) mechanism to encode angular and torsional information via pairwise message passing, indirectly modeling higher-order interactions. MABNet, on the other hand, introduced a direct many-body attention mechanism, enabling joint updates among multiple atoms (e.g., four atoms involved in a dihedral angle). This direct interaction allowed our model to more effectively capture complex molecular geometries. Table 12 presented the comparison between the two models. MABNet achieved a significant reduction in energy prediction error, improving over ViSNet by 32.9%.

Table 12. Comparison of energy and force prediction errors between ViSNet and MABNet with different many-body attention mechanisms on Ac-Ala3-NHMe (energy in kcal/mol and force in kcal/mol/Å).

| Methods | Ac-Ala3-NHMe | |
|-----------------|--------------|--------------|
| | Energy | Forces |
| ViSNet | 0.080 | 0.097 |
| MABNet (2-body) | 0.072 | 0.092 |
| MABNet (3-body) | 0.061 | 0.082 |
| MABNet (4-body) | 0.053 | 0.077 |