

EFFICIENT EQUIVARIANT GRAPH NEURAL NETWORKS FOR 3D ATOMISTIC SYSTEMS

Anonymous authors

Paper under double-blind review

ABSTRACT

Understanding complex three-dimensional (3D) structures of graphs is essential for accurately modeling various properties, yet many existing approaches struggle with fully capturing the intricate spatial relationships and symmetries inherent in such systems, especially in large-scale, dynamic molecular datasets. These methods often must balance trade-offs between expressiveness and computational efficiency, limiting their scalability. To address this gap, we propose a novel Geometric Tensor Network (GotenNet) that effectively models the geometric intricacies of 3D graphs while ensuring strict equivariance under the Euclidean group $E(3)$. Our approach directly tackles the expressiveness-efficiency trade-off by leveraging effective geometric tensor representations without relying on irreducible representations or Clebsch-Gordan transforms, thereby reducing computational overhead. We introduce a unified structural embedding, incorporating geometry-aware tensor attention and hierarchical tensor refinement that iteratively updates edge representations through inner product operations on high-degree steerable features, allowing for flexible and efficient representations for various tasks. We evaluated models on QM9, rMD17, MD22, and Molecule3D datasets, where the proposed model consistently outperforms state-of-the-art methods in both scalar and high-degree property predictions, demonstrating exceptional robustness across diverse datasets, and establishes GotenNet as a versatile and scalable framework for 3D equivariant Graph Neural Networks.

1 INTRODUCTION

Accurately modeling 3D molecular systems is increasingly crucial in areas such as drug discovery (Jing et al., 2021; Nguyen et al., 2021; Huang et al., 2020; Chen et al., 2020; Yang et al., 2022; Zhang et al., 2022; 2021), materials science (Pablo-García et al., 2023; Reiser et al., 2022), and structural biology (Zhang et al., 2022; 2021). These tasks require a precise understanding of the spatial configurations and symmetries inherent in molecular structures, as these factors are fundamental to predicting molecular properties. While predicting scalar molecular properties, such as energy and stability, is challenging, predicting molecular forces is particularly difficult due to the vector nature of forces and their dependence on local geometric environments (Klicpera et al.; Liao & Smidt, 2023; Wang et al., 2024; 2023b; Du et al., 2023). Traditional graph neural networks (GNNs), while effective for general graph-structured data, face challenges in handling the geometric and topological complexities of 3D molecular systems, where achieving equivariance remains a significant challenge (Satorras et al.; Jing et al., 2021; Aykent & Xia, 2023; Thomas et al., 2018).

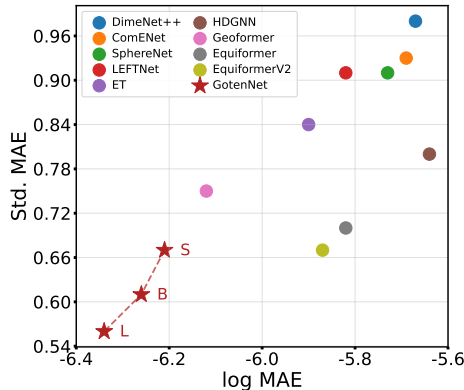


Figure 1: Comparison of GotenNet and baseline models on the QM9 dataset. The x-axis shows the logarithmic MAE across all targets, while the y-axis shows the standardized MAE. Lower values on both axes indicate better performance, with the ideal performance point being at the lower-left corner. Points marked as S, B, and L represent small, base, and large model size variations of the GotenNet, respectively.

054 Recent advances in equivariant neural networks have led to two distinct approaches (Han et al., 2024):
055 scalarization-based models and high-degree steerable models. Scalarization-based models (Satorras
056 et al.; Schütt et al., 2021; Du et al., 2023; Thölke & De Fabritiis, 2022) operate by projecting 3D
057 geometric information into scalar features before reconstruction, offering computational efficiency
058 and scalability for large-scale applications. However, this projection process may limit expressiveness
059 in capturing complex geometric patterns, particularly in scenarios requiring precise understanding of
060 spatial relationships and symmetries. On the other hand, high-degree steerable models (Batzner et al.,
061 2022; Batatia et al., 2022b; Musaelian et al., 2023; Batatia et al., 2022a; Qiao et al., 2022; Liao &
062 Smidt, 2023; Liao et al., 2024) achieve impressive performance through irreducible representations
063 (irreps) and Clebsch-Gordan (CG) transforms, enabling direct manipulation of geometric features in
064 higher-resolution representation spaces. Despite their theoretical rigor and strong performance, these
065 models incur significant computational overhead due to their reliance on complex tensor operations
066 (Cen et al., 2024; Liao & Smidt, 2023; Liao et al., 2024). This fundamental dichotomy between
067 computational efficiency and geometric expressiveness presents a critical challenge in the field: how
068 to achieve both qualities while maintaining strong performance across diverse molecular properties.

069 Fundamental breakthroughs in theoretical understanding have revealed promising directions for
070 addressing this challenge. Rather than relying on explicit CG coefficients, recent advances have
071 illuminated how inner product operations can effectively capture similar geometric relationships (Cen
072 et al., 2024) while being computationally more tractable. This insight suggests the potential for more
073 efficient architectures that maintain the expressiveness of high-degree representations without the
074 computational burden of explicit CG transforms. However, translating this theoretical understanding
075 into practical architectures remains challenging – it requires not only a novel formulation of geometric
076 operations but also careful consideration of how to maintain numerical stability and computational
077 efficiency at scale. The challenge is evident in existing models’ inability to bridge the gap between
078 scalarization-based and high-degree steerable approaches while maintaining practical applicability.
079 Most current architectures (Han et al., 2024; Wang et al., 2024; Liao & Smidt, 2023; Liao et al., 2024)
080 either compromise on expressiveness for efficiency or forfeit computational tractability for geometric
081 accuracy, leaving a clear divide between models optimized for scalar property prediction and those
082 designed for force field calculations, with few achieving strong performance in both domains.

082 To address these challenges, we propose a novel framework, the Geometric Tensor Network (Goten-
083 Net). Our approach focuses on addressing the trade-off between expressiveness and efficiency. First,
084 we introduce an efficient representation and embedding strategy designed specifically using geometric
085 tensors, eliminating the need for irreps and CG transforms, thereby reducing computational com-
086 plexity without sacrificing the expressiveness required for modeling intricate 3D structures. Second,
087 we present the geometry-aware tensor attention and hierarchical tensor refinement mechanisms.
088 These mechanisms enhance transformer-based architectures by refining edge representations through
089 high-degree steerable features, enabling the self-attention mechanism to leverage refined geometric
090 relationships in determining node interactions. This refinement process enriches the attention weights
091 with granular geometric information, allowing for more precise modeling of spatial relationships in
092 molecular structures. These innovations allow the model to represent molecular properties across
093 multiple scales, adapting to both broad patterns and fine-grained molecular details. As shown in
094 Figure 1, our model consistently outperforms baselines on the QM9 dataset, excelling in both standard
095 MAE and log MAE metrics. This highlights GotenNet’ ability to maintain accuracy for large-scale
096 properties while ensuring precision for smaller-scale ones, resulting in strong overall performance
097 across diverse molecular properties.

097 Through rigorous evaluations on benchmark datasets—QM9, Molecule3D, rMD17, and MD22—our
098 approach consistently outperforms state-of-the-art methods, even in its smallest configuration, estab-
099 lishing GotenNet as a versatile and scalable framework for future developments in 3D equivariant
100 graph neural networks. The demonstrated robustness in predicting both scalar and higher degree
101 tensor properties highlights its broad potential for applications in fields such as drug discovery,
102 materials science, and molecular dynamics simulations.

104 2 RELATED WORK

105 The field of machine learning for molecular representation learning has seen significant advancements
106 in recent years (Gasteiger et al., 2020; Liu et al.; Gasteiger et al.; Wang et al.; Liao & Smidt, 2023;
107

Liao et al., 2024), especially for the development of Graph Neural Networks (GNNs) for predicting quantum mechanical properties and simulating molecular dynamics. These approaches can be broadly categorized into two main groups: invariant GNNs and equivariant GNNs.

2.1 INVARIANT GNNs

Invariant GNNs focus on extracting rotation and translation invariant features from molecular graphs (Xie & Grossman, 2018; Unke & Meuwly; Gasteiger et al., 2020; Gasteiger et al.; Schütt et al.; Liu et al.; Wang et al.; Klicpera et al.). DimeNet (Gasteiger et al., 2020) introduced the concept of directional message passing, embedding messages between atoms instead of atoms themselves. This approach allowed the incorporation of angular information while maintaining rotational equivariance. GemNet (Klicpera et al.) extended this idea by incorporating dihedral angles and SphereNet (Liu et al.) integrated torsion information efficiently in the message passing scheme. ComENet (Wang et al.) built upon these approaches, introducing a novel message passing scheme that operates within 1-hop neighborhoods and achieves both global and local completeness in incorporating 3D information.

2.2 EQUIVARIANT GNNs

Equivariant GNNs, on the other hand, directly model rotational equivariance and translational invariance in their architectures (Thomas et al., 2018; Kondor et al., 2018; Schütt et al., 2021; Jing et al., 2021; Thölke & De Fabritiis, 2022; Wang et al., 2024; Le et al., 2022; Du et al., 2022; 2023; Aykent & Xia, 2023; Wang et al., 2023b). Current equivariant GNNs can be divided into two primary approaches based on their feature processing strategies (Han et al., 2024): scalarization-based models and high-degree steerable models. Scalarization-based models focus on deriving invariant scalar features from 3D coordinates and then reconstructs directional information for equivariant updates (Han et al., 2024; Satorras et al.). Several models have successfully implemented this strategy: PaiNN (Schütt et al., 2021) incorporated both scalar and vectorial features in its message passing framework, LEFTNet (Du et al., 2023) developed local frame-based representations with structural encodings, and TorchMD-NET (Thölke & De Fabritiis, 2022) introduced an equivariant transformer architecture that processes scalar and vector features separately.

The high-degree steerable models leverage high-degree representations and CG tensor products for molecular modeling (Batzner et al., 2022; Batatia et al., 2022b; Musaelian et al., 2023; Batatia et al., 2022a; Qiao et al., 2022; Liao & Smidt, 2023; Liao et al., 2024). SE(3)-Transformer (Fuchs et al., 2020) pioneered this direction by introducing attention mechanisms with high-degree steerable features, though computational limitations arose from tensor product operations. Subsequent works like NequIP (Batzner et al., 2022) and MACE (Batatia et al., 2022b) introduced approaches using CG coefficients for equivariance, while Allegro (Musaelian et al., 2023) introduced a local equivariant architecture using iterated tensor products. Equiformer (Liao & Smidt, 2023) introduced SE(3) equivariance into Transformers through depth-wise tensor products and MLP-based attention mechanisms. Building on this work, EquiformerV2 (Liao et al., 2024) enhanced computational efficiency for high-degree representations by incorporating eSCN (Passaro & Zitnick, 2023), although its Fibonacci grid sampling approach incurs an $\mathcal{O}(L^3)$ computational overhead in achieving quasi-equivariance. While these methods show impressive performance, their computational overhead remains significant.

Recent approaches have explored alternative methods to capture geometric information without relying on tensor products and CG coefficients, offering more computationally efficient solutions. ViSNet (Wang et al., 2024) established connections between inner products and Legendre polynomials through vector-scalar interactive message passing, though focusing on first-degree steerable features. SO3KRATES (Frank et al., 2024) demonstrated that certain applications of CG coefficients are equivalent to inner products of high-degree steerable features, achieving notable performance in property prediction tasks through their equivariant transformer architecture. HEGNN (Cen et al., 2024) further developed these concepts by introducing a scalarization approach using inner products to incorporate high-degree steerable features. This approach proved capable of capturing complete angular information between edge pairs and demonstrating enhanced model robustness in dynamics tasks. Our work advances this direction by introducing geometry-aware tensor attention, which employs a concise formulation of inner product operations combined with hierarchical refinement mechanisms. GotenNet represents a significant advancement in bridging the critical gap between scalarization-based and high-degree steerable models. Through its novel architectural design, GotenNet achieves superior

harmonics of degree l , where the degree determines the transformation behavior under rotations, and the parity of l determines the behavior under inversion. These functions chosen to capture complex spatial relationships and rotational symmetries inherent in molecular structures. Leveraging the inherent normalization property of spherical harmonics, each $\tilde{r}^{(l)}$ for $l \geq 1$ is naturally normalized, ensuring consistent scaling across different representations.

We denote the geometric node tensors into two types of features: scalar features $\mathbf{h} \in \mathbb{R}^{d_{ne}}$ which remain invariant under transformations, and high-degree steerable features $\tilde{\mathbf{X}}^{(l)} \in \mathbb{R}^{(1+2l) \times d_{ne}}$ whose transformations depend on their degree l where d_{ne} denotes node embedding dimension. These representations are initialized and updated through message passing phases using the edge tensor representation $\tilde{\mathbf{r}}_{ij}$ and edge scalar features \mathbf{t}_{ij} as input. The notation $\tilde{\mathbf{X}}$ without a specified degree l refers to the collection of features with degrees from 1 to L_{\max} . This initialization strategy enables our model to effectively capture, process, and propagate complex structural information.

Our initialization and feature design ensure equivariance throughout the network. A geometric tensor field maps 3D points to tensor quantities that transform equivariantly under geometric transformations, combining both invariant scalars and steerable features. Each layer of GotenNet processes these tensor fields through equivariant operations while preserving $E(3)$ transformations, with the final layer producing either equivariant geometric features or invariant representations as required by the task. This composition of equivariant operations ensures that the entire network maintains equivariance, with complete proofs provided in the Appendices A, B, and C.

3.2 UNIFIED STRUCTURAL EMBEDDING: INTEGRATING CONTENT AND GEOMETRY

Our approach introduces a unified structure embedding that captures intrinsic atomic properties and relational information through an integrated node-edge interaction mechanism. By employing a dual representation strategy, we incorporate local geometric structure through node-edge interaction. This allows the model to simultaneously process both semantic and geometric information, enabling efficient message passing for both nodes and edges.

Node Scalar Feature Initialization. Node scalar features are obtained through a two-step process involving message passing and representation updates. Information from neighboring nodes is aggregated as:

$$\mathbf{m}_i = \sum_{j \in \mathcal{N}(i)} \mathbf{z}_j \mathbf{A}_{\text{nbr}} \circ \left(\varphi(\tilde{\mathbf{r}}_{ij}^{(0)}) \mathbf{W}_{\text{ndp}} \circ \phi(\tilde{\mathbf{r}}_{ij}^{(0)}) \right), \quad (1)$$

where \mathbf{z} denotes the one-hot encoding of the atomic number, \circ denotes element-wise product, and $\mathbf{A}_{\text{nbr}} \in \mathbb{R}^{|\mathcal{Z}| \times d_{ne}}$ is a learnable embedding matrix for neighbor atoms with maximum atomic number $|\mathcal{Z}|$. The radial basis functions $\varphi(\tilde{\mathbf{r}}_{ij}^{(0)})$ encode the distance between nodes i and j , which are then projected through \mathbf{W}_{ndp} . A cutoff function $\phi(\tilde{\mathbf{r}}_{ij}^{(0)})$ is applied to modulate the influence of distant neighbors.

The initial node scalar feature is defined as:

$$\mathbf{h}_{i,\text{init}} = \left(\sigma \left(\text{LN} \left((\mathbf{z}_i \mathbf{A}_{\text{na}}, \mathbf{m}_i) \mathbf{W}_{\text{nrđ}} \right) \right) \right) \mathbf{W}_{\text{nru}}. \quad (2)$$

Here, $\mathbf{A}_{\text{na}} \in \mathbb{R}^{|\mathcal{Z}| \times d_{ne}}$ is a learnable embedding matrix for node atoms, σ denotes a non-linear activation function and (\cdot, \cdot) denotes concatenation operation. The concatenated node atom embedding and aggregated neighbor information undergo a series of transformations: node representation projections ($\mathbf{W}_{\text{nrđ}}$, \mathbf{W}_{nru}), and layer normalization (LN).

Edge Scalar Feature Initialization. Edge scalar features are computed by combining node features with distance-based edge attributes:

$$\mathbf{t}_{ij,\text{init}} = (\mathbf{h}_{i,\text{init}} + \mathbf{h}_{j,\text{init}}) \circ \left(\sigma \left(\text{LN} \left(\varphi(\tilde{\mathbf{r}}_{ij}^{(0)}) \mathbf{W}_{\text{erd}} \right) \right) \right) \mathbf{W}_{\text{eru}}. \quad (3)$$

Edge attributes are processed through down-projection \mathbf{W}_{erd} and up-projection \mathbf{W}_{eru} , enabling the integration of node-level features and spatial relationships. This formulation captures complex interactions between nodes while maintaining equivariance under molecular transformations.

High-degree Steerable Feature Initialization. The high-degree steerable features $\tilde{\mathbf{X}}$ initialized during initial interaction layer with the following formulation:

$$\begin{aligned} \{\mathbf{o}_{ij,\text{init}}^{(l)}\}_{l=1}^{L_{\max}} &= \text{split}\left(\text{sea}_{ij} + (\mathbf{t}_{ij,\text{init}} \mathbf{W}_{rs,\text{init}}) \circ \gamma_s(\mathbf{h}_{j,\text{init}}) \circ \phi(\tilde{\mathbf{r}}_{ij}^{(0)}), d_{ne}\right), \\ \tilde{\mathbf{X}}_{i,\text{init}}^{(l)} &= \bigoplus_{j \in \mathcal{N}(i)} \left(\mathbf{o}_{ij,\text{init}}^{(l)} \circ \tilde{\mathbf{r}}_{ij}^{(l)}\right), \end{aligned} \quad (4)$$

where sea_{ij} is self-attention with geometric encoding, $\mathbf{W}_{rs,\text{init}} \in \mathbb{R}^{d_{ed} \times d_{ne}}$ is a learnable weight matrix, $\gamma_s : \mathbb{R}^{d_{ne}} \rightarrow \mathbb{R}^{L_{\max} \times d_{ne}}$ is a differentiable function, and the split function decomposes the input tensor into d_{ne} -dimensional segments. These segments are used as different coefficients for each l -degree steerable features. \bigoplus denotes a permutation-invariant aggregation function.

3.3 GEOMETRY-AWARE TENSOR ATTENTION

We introduce a novel module called Geometry-Aware Tensor Attention (GATA), which enhances the attention mechanism in graph neural networks by incorporating spatial information. GATA captures the geometric relationships between nodes to improve attention-driven message passing.

The GATA module combines self-attention with geometric encoding to generate rich node interaction representations. We compute the query (\mathbf{q}), key (\mathbf{k}), and value (\mathbf{v}) representations:

$$\mathbf{q}_i = \mathbf{h}_i \mathbf{W}_q, \quad \mathbf{k}_j = \mathbf{h}_j \mathbf{W}_k, \quad \mathbf{v}_j = \gamma_v(\mathbf{h}_j), \quad (5)$$

where $\mathbf{W}_q, \mathbf{W}_k \in \mathbb{R}^{d_{ne} \times d_{ne}}$ are learnable weight matrices, and $\gamma_v : \mathbb{R}^{d_{ne}} \rightarrow \mathbb{R}^{S \times d_{ne}}$ is a differentiable function (e.g., MLP). The S variable introduced to generate different coefficients for each degree of steerable features and formulated as $(1 + 2 \times L_{\max})$. The attention coefficients α_{ij} between nodes i and j using the dot product of the query vector \mathbf{q}_i and a geometry-infused key vector, which is obtained via an element-wise product of \mathbf{k}_j and a transformed edge embedding:

$$\text{sea}_{ij} = \frac{\exp(\alpha_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(\alpha_{ik})} \mathbf{v}_j, \quad \text{where} \quad \alpha_{ij} = \mathbf{q}_i (\mathbf{k}_j \circ \sigma_k(\mathbf{t}_{ij} \mathbf{W}_{re}))^T. \quad (6)$$

Here, σ_k denotes a non-linear activation function, and $\mathbf{W}_{re} \in \mathbb{R}^{d_{ed} \times C_{ne}}$ is a learnable weight matrix that transforms the edge scalar features. To incorporate spatial and directional information, we augment the attention mechanism with geometric encoding. The GATA operation combines self-attention with geometric features and is then split into S components:

$$\mathbf{o}_{ij}^s, \{\mathbf{o}_{ij}^{d,(l)}\}_{l=1}^{L_{\max}}, \{\mathbf{o}_{ij}^{t,(l)}\}_{l=1}^{L_{\max}} = \text{split}(\text{sea}_{ij} + (\mathbf{t}_{ij} \mathbf{W}_{rs}) \circ \gamma_s(\mathbf{h}_j) \circ \phi(\tilde{\mathbf{r}}_{ij}^{(0)}), d_{ne}), \quad (7)$$

where $\mathbf{W}_{rs} \in \mathbb{R}^{d_{ed} \times (S \cdot d_{ne})}$ is a learnable weight matrix, $\gamma_s : \mathbb{R}^{d_{ne}} \rightarrow \mathbb{R}^{S \times d_{ne}}$ is a differentiable function, and the split function decomposes the input tensor into d_{ne} -dimensional segments. We define $\Delta \mathbf{h}_i$ and $\Delta \tilde{\mathbf{X}}$ as the residues, which are calculated by: high-degree steerable features:

$$\Delta \mathbf{h}_i = \bigoplus_{j \in \mathcal{N}(i)} (\mathbf{o}_{ij}^s), \quad \Delta \tilde{\mathbf{X}}_i^{(l)} = \bigoplus_{j \in \mathcal{N}(i)} \left(\mathbf{o}_{ij}^{d,(l)} \circ \tilde{\mathbf{r}}_{ij}^{(l)} + \mathbf{o}_{ij}^{t,(l)} \circ \tilde{\mathbf{X}}_j^{(l)}\right). \quad (8)$$

Here, each degree $l \in [1, L_{\max}]$ contributes its own component of steerable features weighted by their respective coefficients $\mathbf{o}_{ij}^{d,(l)}$ and $\mathbf{o}_{ij}^{t,(l)}$. Finally updated representations using residues calculated with:

$$\mathbf{h}_i = \mathbf{h}_i + \Delta \mathbf{h}_i, \quad \tilde{\mathbf{X}}_i^{(l)} = \tilde{\mathbf{X}}_i^{(l)} + \Delta \tilde{\mathbf{X}}_i^{(l)}, \quad (9)$$

By infusing geometric information into the attention mechanism, GATA allows the model to better capture spatial dependencies and fine-grained node interactions, leading to improved performance in molecular property predictions, as demonstrated in Section 4.

3.4 HIERARCHICAL TENSOR REFINEMENT

The Hierarchical Tensor Refinement (HTR) component processes graph-structured data through multi-scale analysis and layer-wise refinement. High-degree steerable features are projected to query

and key representations using degree-specific SO(3)-equivariant linear transformation (Deng et al., 2021; Du et al., 2023; Wang et al., 2024) as shown in Equation (10):

$$\widetilde{\mathbf{EQ}}_i^{(l)} = \tilde{\mathbf{X}}_i^{(l)} \mathbf{W}_{vq}, \quad \widetilde{\mathbf{EK}}_j^{(l)} = \tilde{\mathbf{X}}_j^{(l)} \mathbf{W}_{vk}^{(l)}, \quad \text{for } l \in \{1, \dots, L_{\max}\}, \quad (10)$$

where $\mathbf{W}_{vq}, \mathbf{W}_{vk}^{(l)} \in \mathbb{R}^{d_{ed} \times d_{xpd}}$ are tensor query and key projection matrices, respectively. Here, \mathbf{W}_{vq} is a shared projection matrix across degrees, while $\mathbf{W}_{vk}^{(l)}$ is degree-specific. To maintain equivariance, we apply uniform weights across the spatial dimensions for each representation dimension. These projections aggregate angular and magnitude information between nodes across tensor degrees, defined as:

$$\mathbf{w}_{ij} = \text{Agg}_{l=1}^{L_{\max}} \left((\widetilde{\mathbf{EQ}}_i^{(l)})^\top \widetilde{\mathbf{EK}}_j^{(l)} \right), \quad (11)$$

where $\mathbf{w}_{ij} \in \mathbb{R}^{d_{xpd}}$ represents the aggregated similarity between nodes i and j , and $\text{Agg}_{l=1}^{L_{\max}}$ denotes an aggregation operation. The aggregated information updates edge representations through a residual connection:

$$\mathbf{t}_{ij} = \mathbf{w}_{ij} \mathbf{W}_{edp} + \mathbf{t}_{ij} \mathbf{W}_{eru}, \quad (12)$$

where $\mathbf{W}_{edp} \in \mathbb{R}^{d_{xpd} \times d_{ed}}$ is an edge projection matrix, $\mathbf{W}_{eru} \in \mathbb{R}^{d_{ed} \times d_{ed}}$ is a residual update projection matrix. d_{xpd} is chosen larger than d_{ed} for richer intermediate representations. The weight matrices \mathbf{W}_{edp} and \mathbf{W}_{eru} apply the same values across the tensor dimensions, ensuring the model’s equivariance is preserved.

3.5 EQUIVARIANT FEED-FORWARD (EQFF) NETWORKS

The EQFF blocks, employed after GATA, facilitates efficient channel-wise interaction while maintaining equivariance. The module separates scalar and high-degree steerable features, combining them with non-linear mappings:

$$\begin{aligned} \text{EQFF}(\mathbf{h}, \tilde{\mathbf{X}}^{(l)}) &= \left((\mathbf{h} + \mathbf{m}_1), (\tilde{\mathbf{X}}^{(l)} + (\mathbf{m}_2 \circ \tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu})) \right), \\ \text{where } \mathbf{m}_1, \mathbf{m}_2 &= \text{split}_2 \left(\gamma_m (\|\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu}\|_2, \mathbf{h}) \right). \end{aligned} \quad (13)$$

Here γ_m denote differentiable functions such as MLPs, \mathbf{W}_{vu} denotes learnable weight matrices, (\cdot, \cdot) denotes concatenation, and $\|\cdot\|_2$ denotes L_2 norm. The EQFF module operates on the tensor representations while separating the scalar and high-degree steerable features and combining them through element-wise operations. The use of γ_m enables the model to learn complex non-linear mappings, enhancing its expressiveness (Elfving et al., 2018; Ramachandran et al., 2017).

4 EXPERIMENTS

In this section, we compare the performance of GotenNet with other state-of-the-art methods. Experiments were conducted with an NVIDIA A100 GPU with 80GB video memory, 512GB RAM, and an AMD EPYC 7713P CPU. We evaluated models on QM9, rMD17, MD22, and Molecule3D datasets. The best results are **bolded** and the second best are underlined. Additional details on hyperparameters and scalability, as well as additional experiments, can be found in the Appendix E.

4.1 QM9 DATASET

Dataset. The proposed method is evaluated against a comprehensive set of baselines using the QM9 dataset (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014). These baselines include Cormorant (Anderson et al., 2019), ClofNet (Du et al., 2022), NMP (Gilmer et al., 2017), EGNN (Satorras et al.), SEGNN (Brandstetter et al., 2022), PaiNN (Schütt et al., 2021), DimeNet++ (Gasteiger et al.), ComENet (Wang et al.), SphereNet (Liu et al.), LEFTNet (Du et al., 2023), EQGAT (Le et al., 2022), ET (Thölke & De Fabritiis, 2022), HDGNN (An et al., 2024), Geoformer (Wang et al., 2023a), Equiformer (Liao & Smidt, 2023), EquiformerV2 (Liao et al., 2024), GeoSSL-DDM (Liu et al., 2023), 3D-EMGP (Jiao et al., 2023), Coord (Zaidi et al., 2023) and Frad (Ni et al., 2024). Due to space constraints, Table 1 presents only the ten baseline methods with the lowest std. MAEs, while the complete comparison is provided in Appendix H Table 8.

Table 1: Performance comparisons on QM9 dataset. † denotes using different data partitions.

Task	α	$\Delta\varepsilon$	$\varepsilon_{\text{HOMO}}$	$\varepsilon_{\text{LUMO}}$	μ	C_V	G	H	R^2	U	U_0	ZPVE	std.	log
Units	ma_0^3	meV	meV	meV	mD	$\frac{\text{mcal}}{\text{mol}\cdot\text{K}}$	meV	meV	ma_0^2	meV	meV	meV	%	-
PaiNN†	45	45.7	27.6	20.4	12.0	24	7.35	5.98	66	5.83	5.85	1.28	1.01	-5.85
DimeNet++†	44	32.6	24.6	19.5	29.7	23	7.56	6.53	331	6.28	6.32	1.21	0.98	-5.67
SphereNet†	46	31.1	22.8	18.9	24.5	22	7.78	6.33	268	6.36	6.26	<u>1.12</u>	0.91	-5.73
LEFTNet	48	40	24	18	12	23	7	6	109	7	6	1.33	0.91	-5.82
EQGAT	53	32	20	16	11	24	23	24	382	25	25	2.00	0.86	-5.28
ET	59	36.1	20.3	17.5	11	26	7.62	6.16	33	6.38	6.15	1.84	0.84	-5.90
HDGNN†	46	32	18	16	17	23	11	10	342	8.12	8.34	1.21	0.80	-5.64
Geoformer	<u>40</u>	33.8	18.4	15.4	<u>10</u>	<u>22</u>	<u>6.13</u>	<u>4.39</u>	<u>28</u>	<u>4.41</u>	<u>4.43</u>	1.28	0.75	<u>-6.12</u>
Equiformer	46	30	15.4	14.7	12	23	7.63	6.63	251	6.74	6.59	1.26	0.70	-5.82
Equiformerv2	47	<u>29.0</u>	<u>14.4</u>	<u>13.3</u>	<u>9.9</u>	23	7.57	6.22	186	6.49	6.17	1.47	<u>0.67</u>	-5.87
GotenNet _S	35	23.2	<u>16.3</u>	<u>14.7</u>	7.5	20	5.51	3.86	26	3.76	3.82	1.15	0.62	-6.27
GotenNet _B	33	21.3	<u>15.2</u>	<u>13.5</u>	7.3	20	5.33	3.52	25	3.49	3.49	1.10	0.58	-6.33
GotenNet _L	30	19.9	13.7	12.2	7.7	19	4.98	3.36	21	3.33	3.37	1.08	0.54	-6.39

Model Performance and Size Scaling Analysis. We evaluate three model variants - small (S), base (B), and large (L) - to analyze both performance and scaling behavior, with detailed specifications in Appendix E. As shown in Table 1, even our smallest variant GotenNet_S outperforms baseline methods on ten out of twelve targets while surpassing baselines on std. MAE and log MAE. GotenNet_B demonstrates further improvements, achieving best performance on ten targets and significantly improving aggregated metrics, reducing standard MAE by over 13% and log MAE by 3% compared to the best baseline results. The largest variant GotenNet_L achieves state-of-the-art performance across all metrics, although the relative improvement decreases compared to GotenNet_B, which suggests that dataset size may become a limiting factor for larger models. To investigate scaling to larger datasets, we conduct experiments in Section 4.2 on the Molecule3D dataset, which contains more than 3 million molecules - an order of magnitude larger than QM9. These comprehensive results establish GotenNet as the new state-of-the-art while revealing important insights about model scaling behavior.

Table 2: Performance comparisons on Molecule3D dataset.

Split Task	Random						Scaffold
	μ	$\varepsilon_{\text{HOMO}}$	$\varepsilon_{\text{LUMO}}$	$\Delta\varepsilon$	std.	log	$\Delta\varepsilon$
GIN-Virtual	.0882	.0692	.0632	.1036	.0592	-2.87	.2371
SchNet	.0532	.0275	.0265	.0428	.0263	-3.66	.1511
DimeNet++	.0293	.0240	.0190	.0306	.0188	-4.01	.1214
SphereNet	.0288	.0239	.0183	.0301	.0184	-4.03	.1182
ComENet	.0345	.0288	.0252	.0326	.0220	-3.84	.1273
PaiNN	.0196	.0263	.0197	.0307	.0182	-4.08	.1208
ET	.0223	.0199	.0194	.0303	.0170	-4.13	.1282
LEFTNet	.0151	.0183	.0157	.0275	.0145	-4.32	.1317
SaVeNet-L	<u>.0136</u>	<u>.0159</u>	<u>.0143</u>	<u>.0239</u>	<u>.0128</u>	<u>-4.44</u>	<u>.1082</u>
Geoformer	-	-	-	<u>.0202</u>	-	-	.1135
GotenNet	.0103	.0108	.0112	.0165	.0103	-4.65	.1002

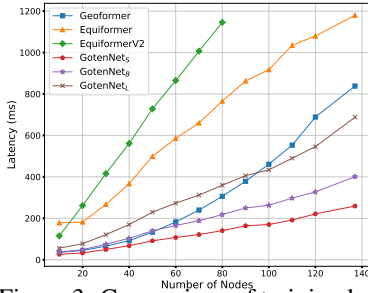


Figure 3: Comparison of training latency of the models with respect to node count on the Molecule3D dataset.

4.2 MOLECULE3D DATASET

Dataset. We further evaluate our model on the Molecule3D dataset (Xu et al., 2021). This dataset contains over $29\times$ more graphs than QM9, with approximately $1.6\times$ and $1.9\times$ increases in the average number of nodes and edges per graph, providing an ideal benchmark for both model performance and computational scaling. We compare against diverse baseline models, including GIN-Virtual (Wang et al.), SchNet (Schütt et al.), DimeNet++ (Gasteiger et al.), SphereNet (Liu et al.), ComENet (Wang et al.), PaiNN (Schütt et al., 2021), ET (Thölke & De Fabritiis, 2022), LEFTNet (Du et al., 2023), SaVeNet (Aykent & Xia, 2023), and Geoformer (Wang et al., 2023a).

Model Performance on Large-Scale Data. As shown in Table 2, GotenNet maintains its superior performance even on this larger Molecule3D dataset, achieving the lowest errors across all tasks, including μ , $\varepsilon_{\text{HOMO}}$, $\varepsilon_{\text{LUMO}}$, and $\Delta\varepsilon$. Notably, GotenNet surpasses the previous best model, SaVeNet-L, by a significant margin of 24% in μ and more than 32% in $\varepsilon_{\text{HOMO}}$. The best log error of -4.65 in the random split further demonstrates the model’s robustness on larger datasets.

Table 3: Comprehensive comparison of various molecular modeling methods on MD22 dataset. The results are reported in MAE of energy (kcal/mol) and forces (kcal/mol/Å) denoted as E and F, respectively. $|G|$ denotes the size of the graphs in terms of node count.

Molecule	$ G $		sGDML	ET	Allegro	MACE	Equiformer	ViSNet	QuinNet	E-LSRM	V-LSRM	SO3KRATES	GotenNet _S	GotenNet _B
Tetrapeptide	42	E	0.3902	0.1121	0.1019	<u>0.0620</u>	0.0828	0.0796	0.0840	0.0780	0.0654	0.337	<u>0.0589</u>	0.0501
		F	0.7968	0.1879	0.1068	0.0876	0.0804	0.0972	<u>0.0681</u>	0.0887	0.0902	0.244	<u>0.0719</u>	0.0600
DHA	56	E	1.3117	0.1205	0.1153	0.1317	0.1788	0.1526	0.1200	0.0878	<u>0.0873</u>	0.379	<u>0.0642</u>	0.0575
		F	0.7474	0.1209	0.0732	0.0646	<u>0.0506</u>	0.0668	0.0515	0.0534	0.0598	0.242	<u>0.0505</u>	0.0421
Stachyose	87	E	4.0497	0.1393	0.2485	0.1244	0.1404	0.1283	0.2300	0.1252	<u>0.1055</u>	0.442	<u>0.0751</u>	0.0673
		F	0.6744	0.1921	0.0971	0.0876	0.0635	0.0869	<u>0.0543</u>	0.0632	0.0767	0.435	<u>0.0512</u>	0.0427
AT-AT	60	E	0.7235	0.1120	0.1428	0.1093	0.1309	0.1688	0.1400	0.1007	<u>0.0772</u>	0.178	<u>0.0640</u>	0.0557
		F	0.6911	0.2036	0.0952	0.0992	0.0960	0.1070	<u>0.0687</u>	0.0881	0.0781	0.216	<u>0.0632</u>	0.0525
AT-AT-CG-CG	118	E	1.3885	0.2072	0.3933	0.1578	0.1510	0.1995	0.3800	0.1335	<u>0.1135</u>	0.345	<u>0.0964</u>	0.0959
		F	0.7028	0.3259	0.1280	0.1153	0.1252	0.1563	0.1273	0.1065	<u>0.1063</u>	0.332	<u>0.0824</u>	0.0773
Buckyball catcher	148	E	1.1962	0.5188	0.5258	-	0.3978	0.4421	0.5624	-	0.4220	0.381	<u>0.3391</u>	0.3088
		F	0.6820	0.3318	<u>0.0887</u>	-	0.1114	0.1335	0.1091	-	0.1026	0.237	<u>0.0858</u>	0.0797
Double-walled nanotube	370	E	4.0122	1.4732	2.2097	-	1.1945	<u>1.0339</u>	1.8130	-	1.8230	0.993	<u>1.0279</u>	0.6641
		F	0.5231	1.0031	0.3428	-	<u>0.2747</u>	0.3959	0.2473	-	0.3391	0.727	<u>0.2558</u>	0.1888

Computational Efficiency Analysis. Beyond performance metrics, scaling to larger datasets requires efficient handling of increased graph sizes. We analyze computational efficiency by measuring training time across varying node counts (10-140 nodes per graph). Figure 3 compares GotenNet with competitive attention-based baselines including Geformer (Wang et al., 2023a), Equiformer (Liao & Smidt, 2023), and EquiformerV2 (Liao et al., 2024). Full experimental setup details are provided in Appendix F. The x-axis shows the node count, while the y-axis shows the training time per batch in milliseconds. The results show GotenNet maintains efficient scaling at higher node counts, while baseline methods like Geformer, despite strong performance on smaller graphs, become computationally intensive due to their dense $O(n^2)$ representations. Both GotenNet_S and GotenNet_B variants maintain consistent efficiency across all node counts, demonstrating their suitability for large-scale applications where computational overhead is critical.

4.3 MD22 DATASET

Dataset. The MD22 dataset (Chmiela et al., 2023) contains molecular dynamics trajectories for seven systems, with atom counts from 42 to 370, across four biomolecule and supramolecule classes. It presents challenges in system size, flexibility, and nonlocality, making it a key benchmark for scalability and accuracy in molecular force field models. Following the data splits from (Chmiela et al., 2023), we evaluate GotenNet against several baselines, including sGDML (Chmiela et al., 2018), ET (Thölke & De Fabritiis, 2022), Allegro (Musaelian et al., 2023), MACE (Batatia et al., 2022b), Equiformer (Liao & Smidt, 2023), ViSNet (Wang et al., 2024), QuinNet (Wang et al., 2023b), SO3KRATES (Frank et al., 2024), and LSRM (Li et al., 2024) along with its variants E-LSRM (Equiformer-LSRM) and V-LSRM (ViSNet-LSRM).

Results. The MD22 dataset poses significant challenges due to its wide range of molecule sizes, requiring accurate predictions of both energy and forces. We showed results of MD22 dataset in Table 3. Our proposed model, GotenNet, consistently outperforms state-of-the-art methods across all evaluated molecules, demonstrating superior performance in both energy and force predictions.

For molecules such as Tetrapeptide and AT-AT, GotenNet achieves notable reductions in energy errors, with improvements of 19.2% and 27.8% over the previous best models, respectively. Simultaneously, force prediction errors are reduced by up to 27.2%, underscoring GotenNet’s balanced performance across both metrics. In more complex cases, such as the Buckyball catcher and Double-walled nanotube, GotenNet sets new benchmarks, reducing energy errors by over 35% and force errors by up to 31.9%. These results highlight the robustness and versatility of GotenNet in handling diverse molecular structures, establishing it as a leading model in both energy and force prediction.

4.4 RMD17 DATASET

Dataset. The rMD17 dataset (Christensen & Von Lilienfeld, 2020) is a revised version of the MD17 benchmark, featuring 10 small organic molecules with 100,000 conformations per molecule. It serves as a key benchmark for evaluating machine learning models’ ability to predict molecular energies and forces across diverse conformations. We follow the standard split (Christensen & Von Lilienfeld, 2020) of 950 training, 50 validation, and the remaining conformations for testing, the results are averaged over five predefined splits to ensure robust evaluation.

Table 4: The table presents MAE for energy (kcal/mol) and forces (kcal/mol/Å) on the rMD17 dataset.

Molecule		NequIP	ACE	UNiTE	Allegro	BOTNet	MACE	TensorNet	GotenNet
Aspirin	E	0.0530	0.1407	0.0553	0.0530	0.0530	<u>0.0507</u>	0.0553	0.0364
	F	0.1891	0.4128	0.1753	0.1683	0.1960	<u>0.1522</u>	0.2052	0.1338
Azobenzene	E	<u>0.0161</u>	0.0830	0.0254	0.0277	<u>0.0161</u>	0.0277	<u>0.0161</u>	0.0125
	F	0.0669	0.2514	0.0969	<u>0.0600</u>	0.0761	0.0692	0.0715	0.0500
Benzene	E	0.0009	0.0009	0.0016	0.0069	<u>0.0007</u>	0.0092	0.0005	0.0005
	F	0.0069	0.0115	0.0168	0.0046	0.0069	0.0069	0.0069	<u>0.0048</u>
Ethanol	E	0.0092	0.0277	0.0143	0.0092	0.0092	0.0092	0.0115	0.0071
	F	0.0646	0.1683	0.0853	<u>0.0484</u>	0.0738	<u>0.0484</u>	0.0807	0.0482
Malonaldehyde	E	0.0184	0.0392	0.0254	0.0138	0.0184	0.0184	0.0184	<u>0.0129</u>
	F	0.1176	0.2560	0.1522	0.0830	0.1338	0.0945	0.1245	0.0830
Naphthalene	E	0.0208	0.0208	0.0106	<u>0.0046</u>	<u>0.0046</u>	0.0115	<u>0.0046</u>	0.0039
	F	0.0300	0.1176	0.0600	0.0208	0.0415	0.0369	0.0369	<u>0.0240</u>
Paracetamol	E	0.0323	0.0922	0.0438	0.0346	<u>0.0300</u>	<u>0.0300</u>	<u>0.0300</u>	0.0212
	F	0.1361	0.2929	0.1637	0.1130	0.1338	<u>0.1107</u>	0.1361	0.0929
Salicylic acid	E	0.0161	0.0415	0.0168	0.0208	0.0184	0.0208	0.0184	0.0161
	F	0.0922	0.2145	0.0876	0.0669	0.0992	0.0715	0.1061	<u>0.0763</u>
Toluene	E	<u>0.0069</u>	0.0254	0.0104	0.0092	0.0092	0.0115	<u>0.0069</u>	0.0048
	F	0.0369	0.1499	0.0577	0.0415	0.0438	<u>0.0346</u>	0.0392	0.0254
Uracil	E	<u>0.0092</u>	0.0254	0.0134	0.0138	<u>0.0092</u>	0.0115	<u>0.0092</u>	0.0074
	F	0.0715	0.1522	0.0876	0.0415	0.0738	0.0484	0.0715	<u>0.0450</u>

Results. As shown in Table 4, GotenNet outperforms other models in 75% of tasks and ranks second in the remaining, excelling in both energy and force predictions. GotenNet sets new benchmarks for molecules such as Aspirin, Azobenzene, Ethanol, Paracetamol, and Toluene, demonstrating balanced improvements across energy and force predictions. These results highlight GotenNet’s robustness and its ability to accurately model molecular properties, outperforming prior methods on rMD17 dataset.

4.5 ABLATION STUDY

Table 5 presents the results of the ablation study, highlighting the impact of various components on the performance of GotenNet. The inclusion of structural embedding (SE), self-attention (SEA), geometric encoding (GE), and HTR generally leads to improved results, as shown in rows 1, 7, and 8, where the model achieves the lowest std MAE and log MAE. The removal of any one of these components results in a significant degradation in performance, particularly in the cases without geometric encoding (row 4) or reducing L_{\max} (row 2). The full model with 12 layers (row 8) achieves the best performance, with the lowest std MAE of 0.56 and log MAE of -6.34. This demonstrates the combined effectiveness of all components for model scalability.

Table 5: Ablation study on QM9 dataset.

# L	L_{\max}	SE	SEA	GE	HTR	std	log
4	2	✓	✓	✓	✓	0.67	-6.21
6	1	✓	✓	✓	✓	0.68	-6.17
6	2	✗	✓	✓	✓	0.67	-6.17
6	2	✓	✗	✓	✓	0.65	-6.23
6	2	✓	✓	✗	✓	0.83	-5.96
6	2	✓	✓	✓	✗	0.64	-6.20
6	2	✓	✓	✓	✓	0.61	-6.26
12	2	✓	✓	✓	✓	0.56	-6.34

5 CONCLUSION

We presented GotenNet, a framework for modeling 3D molecular structures that strikes a balance between expressiveness and efficiency by integrating geometric tensor representations with innovative components, including unified structure embedding, geometry-aware tensor attention, and hierarchical tensor refinement. GotenNet consistently outperforms state-of-the-art methods across four benchmark datasets. It also demonstrates scalability and computational efficiency, making it highly suitable for large-scale molecular systems. These results establish GotenNet as a versatile and powerful framework for 3D equivariant graph neural networks. Future work could further enhance its scalability to larger molecular systems and explore applications in molecular dynamics and materials science.

REFERENCES

- 540
541
542 Junyi An, Chao Qu, Zhipeng Zhou, Fenglei Cao, Xu Yinghui, Yuan Qi, and Furao Shen. Hybrid directional graph neural network for molecules. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=BBD6KXIGJL>.
- 543
544
545 Brandon Anderson, Truong Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural
546 networks. *Advances in neural information processing systems*, 32, 2019.
- 547
548 Sarp Aykent and Tian Xia. Savenet: A scalable vector network for enhanced molecular representation
549 learning. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL
550 <https://openreview.net/forum?id=00ImBCFsdf>.
- 551
552 Ilyes Batatia, Simon Batzner, Dávid Péter Kovács, Albert Musaelian, Gregor NC Simm, Ralf Drautz,
553 Christoph Ortner, Boris Kozinsky, and Gábor Csányi. The design space of e(3)-equivariant
554 atom-centered interatomic potentials. *arXiv preprint arXiv:2205.06643*, 2022a.
- 555
556 Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. Mace: Higher
557 order equivariant message passing neural networks for fast and accurate force fields. In S. Koyejo,
558 S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information
559 Processing Systems*, volume 35, pp. 11423–11436. Curran Associates, Inc., 2022b.
- 560
561 Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth,
562 Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for
563 data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022.
- 564
565 Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J Bekkers, and Max Welling. Ge-
566 ometric and physical quantities improve e(3) equivariant message passing. In *International
567 Conference on Learning Representations*, 2022. URL [https://openreview.net/forum?
568 id=_xwr8g0BeV1](https://openreview.net/forum?id=_xwr8g0BeV1).
- 569
570 Jiacheng Cen, Anyi Li, Ning Lin, Yuxiang Ren, Ziheng Wang, and Wenbing Huang. Are high-
571 degree representations really unnecessary in equivariant graph neural networks? In *The Thirty-
572 eighth Annual Conference on Neural Information Processing Systems*, 2024. URL [https://
573 openreview.net/forum?id=M0ncNVuGYN](https://openreview.net/forum?id=M0ncNVuGYN).
- 574
575 Lifan Chen, Xiaoqin Tan, Dingyan Wang, Feisheng Zhong, Xiaohong Liu, Tianbiao Yang, Xiaomin
576 Luo, Kaixian Chen, Hualiang Jiang, and Mingyue Zheng. Transformerpci: improving compound-
577 protein interaction prediction by sequence-based deep learning with self-attention mechanism and
578 label reversal experiments. *Bioinformatics*, 36(16):4406–4414, 2020.
- 579
580 Stefan Chmiela, Huziel E Saucedo, Klaus-Robert Müller, and Alexandre Tkatchenko. Towards exact
581 molecular dynamics simulations with machine-learned force fields. *Nature communications*, 9(1):
582 3887, 2018.
- 583
584 Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T. Unke, Adil Kabylda, Huziel E. Saucedo,
585 Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate global machine learning force fields for
586 molecules with hundreds of atoms. *Science Advances*, 9(2):eadf0873, 2023. doi: 10.1126/sciadv.
587 adf0873. URL <https://www.science.org/doi/abs/10.1126/sciadv.adf0873>.
- 588
589 Anders S Christensen and O Anatole Von Lilienfeld. On the role of gradients for machine learning of
590 molecular energies and forces. *Machine Learning: Science and Technology*, 1(4):045018, 2020.
- 591
592 Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J
593 Guibas. Vector neurons: A general framework for so(3)-equivariant networks. In *Proceedings of
the IEEE/CVF International Conference on Computer Vision*, pp. 12200–12209, 2021.
- 594
595 Weitao Du, He Zhang, Yuanqi Du, Qi Meng, Wei Chen, Nanning Zheng, Bin Shao, and Tie-Yan Liu.
596 Se(3) equivariant graph neural networks with complete local frames. In *International Conference
597 on Machine Learning*, pp. 5583–5608. PMLR, 2022.
- 598
599 Weitao Du, Yuanqi Du, Limei Wang, Dieqiao Feng, Guifeng Wang, Shuiwang Ji, Carla P Gomes,
600 and Zhi-Ming Ma. A new perspective on building efficient and expressive 3d equivariant graph
601 neural networks. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
602 URL <https://openreview.net/forum?id=hWPNYWkYPN>.

- 594 Stefan Elfving, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network
595 function approximation in reinforcement learning. *Neural networks*, 107:3–11, 2018.
596
- 597 J Thorben Frank, Oliver T Unke, Klaus-Robert Müller, and Stefan Chmiela. A euclidean transformer
598 for fast and stable machine learned force fields. *Nature Communications*, 15(1):6539, 2024.
599
- 600 Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-
601 translation equivariant attention networks. *Advances in neural information processing systems*, 33:
602 1970–1981, 2020.
- 603 Johannes Gasteiger, Shankari Giri, Johannes T. Margraf, and Stephan Günnemann. Fast and
604 uncertainty-aware directional message passing for non-equilibrium molecules. In *Machine Learn-
605 ing for Molecules Workshop, NeurIPS*.
606
- 607 Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular
608 graphs. In *International Conference on Learning Representations*, 2020. URL [https://
609 openreview.net/forum?id=BlEWbxStPH](https://openreview.net/forum?id=BlEWbxStPH).
- 610 Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural
611 message passing for quantum chemistry. In *International conference on machine learning*, pp.
612 1263–1272. PMLR, 2017.
613
- 614 Jiaqi Han, Jiacheng Cen, Liming Wu, Zongzhao Li, Xiangzhe Kong, Rui Jiao, Ziyang Yu, Tingyang
615 Xu, Fandi Wu, Ziheng Wang, Hongteng Xu, Zhewei Wei, Yang Liu, Yu Rong, and Wenbing Huang.
616 A survey of geometric graph neural networks: Data structures, models and applications, 2024.
617 URL <https://arxiv.org/abs/2403.00485>.
- 618 Kexin Huang, Tianfan Fu, Lucas M Glass, Marinka Zitnik, Cao Xiao, and Jimeng Sun. Deeppurpose:
619 a deep learning library for drug–target interaction prediction. *Bioinformatics*, 36(22-23):5545–
620 5547, 2020.
621
- 622 Rui Jiao, Jiaqi Han, Wenbing Huang, Yu Rong, and Yang Liu. Energy-motivated equivariant pre-
623 training for 3d molecular graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*,
624 volume 37, pp. 8096–8104, 2023.
- 625 Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron
626 Dror. Learning from protein structure with geometric vector perceptrons. In *International
627 Conference on Learning Representations*, 2021. URL [https://openreview.net/forum?
628 id=1YLJDvSx6J4](https://openreview.net/forum?id=1YLJDvSx6J4).
629
- 630 Johannes Klicpera, Florian Becker, and Stephan Günnemann. GemNet: Universal directional graph
631 neural networks for molecules. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan
632 (eds.), *Advances in Neural Information Processing Systems*. URL [https://openreview.
633 net/forum?id=HS_sOaxS9K-](https://openreview.net/forum?id=HS_sOaxS9K-).
- 634 Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–gordan nets: a fully fourier space spherical
635 convolutional neural network. In *Advances in Neural Information Processing Systems* 32, pp.
636 10117–10126, 2018.
637
- 638 Tuan Le, Frank Noé, and Djork-Arné Clevert. Equivariant graph attention networks for molecular
639 property prediction. *arXiv preprint arXiv:2202.09891*, 2022.
- 640 Yunyang Li, Yusong Wang, Lin Huang, Han Yang, Xinran Wei, Jia Zhang, Tong Wang, Zun Wang,
641 Bin Shao, and Tie-Yan Liu. Long-short-range message-passing: A physics-informed framework
642 to capture non-local interaction for scalable molecular dynamics simulation. In *The Twelfth
643 International Conference on Learning Representations*, 2024. URL [https://openreview.
644 net/forum?id=rVDQtdMnO1](https://openreview.net/forum?id=rVDQtdMnO1).
645
- 646 Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic
647 graphs. In *International Conference on Learning Representations*, 2023. URL [https://
openreview.net/forum?id=KwmPfARgOTD](https://openreview.net/forum?id=KwmPfARgOTD).

- 648 Yi-Lun Liao, Brandon M Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant
649 transformer for scaling to higher-degree representations. In *The Twelfth International Confer-*
650 *ence on Learning Representations*, 2024. URL [https://openreview.net/forum?id=](https://openreview.net/forum?id=mCOBKZmrzD)
651 [mCOBKZmrzD](https://openreview.net/forum?id=mCOBKZmrzD).
- 652 Shengchao Liu, Hongyu Guo, and Jian Tang. Molecular geometry pretraining with SE(3)-invariant de-
653 noising distance matching. In *The Eleventh International Conference on Learning Representations*,
654 2023. URL <https://openreview.net/forum?id=CjTHVoldvR>.
- 655 Yi Liu, Limei Wang, Meng Liu, Yuchao Lin, Xuan Zhang, Bora Oztekin, and Shuiwang Ji. Spherical
656 message passing for 3D molecular graphs. In *International Conference on Learning Representa-*
657 *tions*. URL <https://openreview.net/forum?id=givsRXsOt9r>.
- 658 Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J Owen, Mordechai
659 Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic
660 dynamics. *Nature Communications*, 14(1):579, 2023.
- 661 Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh.
662 Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37
663 (8):1140–1147, 2021.
- 664 Yuyan Ni, Shikun Feng, Xin Hong, Yuancheng Sun, Wei-Ying Ma, Zhi-Ming Ma, Qiwei Ye, and
665 Yanyan Lan. Pre-training with fractional denoising to enhance molecular property prediction.
666 *Nature Machine Intelligence*, pp. 1–10, 2024.
- 667 Sergio Pablo-García, Santiago Morandi, Rodrigo A Vargas-Hernández, Kjell Jorner, Žarko Ivković,
668 Núria López, and Alán Aspuru-Guzik. Fast evaluation of the adsorption energy of organic
669 molecules on metals via graph neural networks. *Nature Computational Science*, 3(5):433–442,
670 2023.
- 671 Saro Passaro and C Lawrence Zitnick. Reducing so (3) convolutions to so (2) for efficient equivariant
672 gnns. In *International Conference on Machine Learning*, pp. 27420–27438. PMLR, 2023.
- 673 Zhuoran Qiao, Anders S Christensen, Matthew Welborn, Frederick R Manby, Anima Anandkumar,
674 and Thomas F Miller III. Informing geometric deep learning with electronic interactions to acceler-
675 ate quantum chemistry. *Proceedings of the National Academy of Sciences*, 119(31):e2205221119,
676 2022.
- 677 Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. *arXiv preprint*
678 *arXiv:1710.05941*, 2017.
- 679 Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum
680 chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.
- 681 Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam
682 Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, et al. Graph neural networks for
683 materials science and chemistry. *Communications Materials*, 3(1):93, 2022.
- 684 Lars Ruddigkeit, Ruud van Deursen, Lorenz C. Blum, and Jean-Louis Reymond. Enumeration
685 of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of*
686 *Chemical Information and Modeling*, 52(11):2864–2875, 2012. doi: 10.1021/ci300415d. URL
687 <https://doi.org/10.1021/ci300415d>. PMID: 23088335.
- 688 Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) Equivariant Graph Neural
689 Networks. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 9323–
690 9332. PMLR. URL <https://proceedings.mlr.press/v139/satorras21a.html>.
- 691 Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction
692 of tensorial properties and molecular spectra. In *International Conference on Machine Learning*,
693 pp. 9377–9388. PMLR, 2021.
- 694 Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Saucedo Felix, Stefan Chmiela, Alexandre
695 Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network
696 for modeling quantum interactions. 30.

- 702 Philipp Thölke and Gianni De Fabritiis. Equivariant transformers for neural network based molecular
703 potentials. In *International Conference on Learning Representations*, 2022.
704
- 705 Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley.
706 Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds.
707 *arXiv preprint arXiv:1802.08219*, 2018.
- 708 Oliver T Unke and Markus Meuwly. PhysNet: A neural network for predicting energies, forces,
709 dipole moments, and partial charges. 15(6):3678–3693.
710
- 711 Limei Wang, Yi Liu, Yuchao Lin, Haoran Liu, and Shuiwang Ji. ComENet: Towards complete and
712 efficient message passing for 3D molecular graphs. In Alice H. Oh, Alekh Agarwal, Danielle
713 Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*. URL
714 <https://openreview.net/forum?id=mCzMqeWSFJ>.
- 715 Yusong Wang, Shaoning Li, Tong Wang, Bin Shao, Nanning Zheng, and Tie-Yan Liu. Geometric trans-
716 former with interatomic positional encoding. In *Thirty-seventh Conference on Neural Information*
717 *Processing Systems*, 2023a. URL <https://openreview.net/forum?id=9o6KQrklrE>.
- 718 Yusong Wang, Tong Wang, Shaoning Li, Xinheng He, Mingyu Li, Zun Wang, Nanning Zheng,
719 Bin Shao, and Tie-Yan Liu. Enhancing geometric representations for molecules with equivariant
720 vector-scalar interactive message passing. *Nature Communications*, 15(1):313, 2024.
721
- 722 Zun Wang, Guoqing Liu, Yichi Zhou, Tong Wang, and Bin Shao. Efficiently incorporating quintuple
723 interactions into geometric deep learning force fields. In *Thirty-seventh Conference on Neural*
724 *Information Processing Systems*, 2023b. URL [https://openreview.net/forum?id=](https://openreview.net/forum?id=bPJm1PbZD)
725 [bPJm1PbZD](https://openreview.net/forum?id=bPJm1PbZD).
- 726 Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and
727 interpretable prediction of material properties. *Physical Review Letters*, 2018.
728
- 729 Zhao Xu, Youzhi Luo, Xuan Zhang, Xinyi Xu, Yaochen Xie, Meng Liu, Kaleb Dickerson, Cheng
730 Deng, Maho Nakata, and Shuiwang Ji. Molecule3d: A benchmark for predicting 3d geometries
731 from molecular graphs. *CoRR*, abs/2110.01717, 2021. URL [https://arxiv.org/abs/](https://arxiv.org/abs/2110.01717)
732 [2110.01717](https://arxiv.org/abs/2110.01717).
- 733 Ziduo Yang, Weihe Zhong, Lu Zhao, and Calvin Yu-Chian Chen. Mgraphdta: deep multiscale graph
734 neural network for explainable drug–target binding affinity prediction. *Chemical science*, 13(3):
735 816–833, 2022.
- 736 Sheheryar Zaidi, Michael Schaarschmidt, James Martens, Hyunjik Kim, Yee Whye Teh, Alvaro
737 Sanchez-Gonzalez, Peter Battaglia, Razvan Pascanu, and Jonathan Godwin. Pre-training via
738 denoising for molecular property prediction. In *The Eleventh International Conference on Learning*
739 *Representations*, 2023. URL <https://openreview.net/forum?id=tYIMtogyee>.
740
- 741 Xiao-Meng Zhang, Li Liang, Lin Liu, and Ming-Jing Tang. Graph neural networks and their current
742 applications in bioinformatics. *Frontiers in genetics*, 12:690049, 2021.
- 743 Zehong Zhang, Lifan Chen, Feisheng Zhong, Dingyan Wang, Jiabin Jiang, Sulin Zhang, Hualiang
744 Jiang, Mingyue Zheng, and Xutong Li. Graph neural network approaches for drug-target interac-
745 tions. *Current Opinion in Structural Biology*, 73:102327, 2022.
746
747
748
749
750
751
752
753
754
755

APPENDIX

A PROOF: EQUIVARIANCE OF GEOMETRIC TENSOR

Equivariance of Geometric Tensor: Let $T : M \rightarrow \mathcal{T}$ be a geometric tensor field, and $g \in E(3)$ be an element of the Euclidean group. Then T is equivariant under g if:

$$(g \cdot T)(p) = \rho(g)T(g^{-1} \cdot p)$$

where ρ is a representation of $E(3)$ on the space of tensors \mathcal{T} .

Proof. To prove the equivariance of the geometric tensor field, we start by recalling the definition of equivariance. A tensor field $T : M \rightarrow \mathcal{T}$ is equivariant under the group action of $g \in E(3)$ if, for all points $p \in M$,

$$T(g \cdot p) = \rho(g)T(p),$$

where $\rho(g)$ is a representation of the group element g on the space \mathcal{T} .

Now, consider the action of the group element $g \in E(3)$ on T . By the definition of the group action of g on the tensor field T , we have:

$$(g \cdot T)(p) = T(g^{-1} \cdot p).$$

Next, we apply the representation $\rho(g)$ to the transformed tensor $T(g^{-1} \cdot p)$. By the equivariance condition, we require that:

$$(g \cdot T)(p) = \rho(g)T(g^{-1} \cdot p).$$

This completes the proof, as we have shown that the transformed tensor field $g \cdot T$ is related to the original tensor field T by the representation $\rho(g)$, satisfying the equivariance condition. \square

B PROOF: GOTENNET PRESERVE THE EQUIVARIANCE PROPERTY

Proof. We will prove that layer operations in GotenNet preserve the equivariance property of geometric tensor fields under the action of the Euclidean group $E(3)$. To this end we will prove that each component used in GotenNet preserves the invariance/equivariance properties.

First we will prove that the initial node and edge representations are invariant under the action of the Euclidean group $E(3)$. We'll consider each component separately. The node embedding process consists of two main steps: message passing and representation update. We'll show that both steps are invariant under $E(3)$. The message passing equation in node embedding is:

$$\mathbf{m}_i = \sum_{j \in \mathcal{N}(i)} z_j \mathbf{A}_{\text{nbr}} \circ \left(\varphi(\tilde{\mathbf{r}}_{ij}^{(0)}) \mathbf{W}_{\text{ndp}} \circ \phi(\tilde{\mathbf{r}}_{ij}^{(0)}) \right).$$

Under the action of $g \in E(3)$ z_j is invariant as it's an atomic number. $\tilde{\mathbf{r}}_{ij,\text{init}}$ is invariant as it's the distance between nodes i and j . $\mathcal{N}(i)$ is invariant as the set of neighbors doesn't change under rigid transformations. Therefore, \mathbf{m}_i is invariant under $E(3)$. The node representation update is defined as:

$$\mathbf{h}_{i,\text{init}} = \left(\sigma \left(\text{LN} \left((z_i \mathbf{A}_{\text{na}}, \mathbf{m}_i) \mathbf{W}_{\text{nr}} \right) \right) \right) \mathbf{W}_{\text{nr}}.$$

Here, z_i is invariant, \mathbf{m}_i is invariant (as shown above), and all other operations (concatenation, linear transformations, layer normalization, and activation) are invariant. Thus, \mathbf{h}_i is invariant under $E(3)$. The node embedding is followed by edge embedding and the initial edge representation is computed as:

$$\mathbf{t}_{ij,\text{init}} = (\mathbf{h}_{i,\text{init}} + \mathbf{h}_{j,\text{init}}) \circ \left(\sigma \left(\text{LN} \left(\varphi(\tilde{\mathbf{r}}_{ij}^{(0)}) \mathbf{W}_{\text{erd}} \right) \right) \right) \mathbf{W}_{\text{eru}}.$$

We've already shown that h_{init} is invariant. The distance $r_{ij,\text{init}}$ is also invariant under $E(3)$. The operations φ , LN, σ , and linear transformations are all invariant. Therefore, $\mathbf{t}_{ij,\text{init}}$ is invariant under

$E(3)$. Thus, we have shown that both the initial node representations $h_{i,\text{init}}$ and edge representations $t_{ij,\text{init}}$ are invariant under the action of the Euclidean group $E(3)$.

After the initialization step the next module is GATA. Therefore, we will prove that the GATA module and the subsequent operations in the interaction layer preserve equivariance under the action of the Euclidean group $E(3)$. The query, key, and value Computation as defined in Equation (5), we have:

$$\mathbf{q}_i = \mathbf{h}_i \mathbf{W}_q, \quad \mathbf{k}_j = \mathbf{h}_j \mathbf{W}_k, \quad \mathbf{v}_j = \gamma_v(\mathbf{h}_j),$$

\mathbf{h}_i and \mathbf{h}_j are scalar (0-degree steerable features) representations, which are invariant under $E(3)$. The linear transformations \mathbf{W}_q and \mathbf{W}_k , and the function γ_v , preserve this invariance. Thus, \mathbf{q}_i , \mathbf{k}_j , and \mathbf{v}_j are invariant under $E(3)$. The attention coefficients are defined in Equation (6), the attention coefficient α_{ij} is computed as:

$$\alpha_{ij} = \mathbf{q}_i (\mathbf{k}_j \circ \sigma_k(\mathbf{t}_{ij} \mathbf{W}_{re}))^T.$$

We've shown that \mathbf{q}_i and \mathbf{k}_j are invariant. \mathbf{t}_{ij} is an edge embedding, which is invariant under $E(3)$. The operations σ_k , \circ , and matrix multiplication preserve invariance. Therefore, α_{ij} is invariant under $E(3)$. The self-attention operation in Equation (6) defined as:

$$\mathbf{sea}_{ij} = \frac{\exp(\alpha_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(\alpha_{ik})} \mathbf{v}_j.$$

Since α_{ij} and \mathbf{v}_j are invariant, and the softmax operation preserves invariance, \mathbf{sea}_{ij} is invariant under $E(3)$. Geometric Encoding defined in Equation (7), we have:

$$\mathbf{o}_{ij}^s, \{\mathbf{o}_{ij}^{d,(l)}\}_{l=1}^{L_{\max}}, \{\mathbf{o}_{ij}^{t,(l)}\}_{l=1}^{L_{\max}} = \text{split}(\mathbf{sea}_{ij} + (\mathbf{t}_{ij} \mathbf{W}_{rs}) \circ \gamma_s(\mathbf{h}_j) \circ \phi(\tilde{\mathbf{r}}_{ij}^{(0)}, d_{ne})).$$

\mathbf{sea}_{ij} , \mathbf{t}_{ij} , \mathbf{h}_j , and $\tilde{\mathbf{r}}_{ij}^{(0)}$ are all invariant under $E(3)$. The operations \mathbf{W}_{rs} , γ_s , ϕ , \circ , and split preserve this invariance. Therefore, \mathbf{o}_{ij}^s , $\{\mathbf{o}_{ij}^{d,(l)}\}_{l=1}^{L_{\max}}$ and $\{\mathbf{o}_{ij}^{t,(l)}\}_{l=1}^{L_{\max}}$ are invariant under $E(3)$. Finally, node tensor representation updated with Equation (8), we have:

$$\Delta \mathbf{h}_i = \bigoplus_{j \in \mathcal{N}(i)} (\mathbf{o}_{ij}^s), \quad \Delta \tilde{\mathbf{X}}_i^{(l)} = \bigoplus_{j \in \mathcal{N}(i)} \left(\mathbf{o}_{ij}^{d,(l)} \circ \tilde{\mathbf{r}}_{ij}^{(l)} + \mathbf{o}_{ij}^{t,(l)} \circ \tilde{\mathbf{X}}_j^{(l)} \right).$$

Here, \mathbf{h}_i , \mathbf{o}_{ij}^s , $\{\mathbf{o}_{ij}^{d,(l)}\}_{l=1}^{L_{\max}}$ and $\{\mathbf{o}_{ij}^{t,(l)}\}_{l=1}^{L_{\max}}$ are invariant under $E(3)$. Specifically, $\{\mathbf{o}_{ij}^{d,(l)}\}_{l=1}^{L_{\max}}$ and $\{\mathbf{o}_{ij}^{t,(l)}\}_{l=1}^{L_{\max}}$ are sets of invariant coefficients, where for each degree l of steerable features, we have distinct invariant scalars $\mathbf{o}_{ij}^{d,(l)}$ and $\mathbf{o}_{ij}^{t,(l)}$. $\tilde{\mathbf{r}}_{ij}^{(l)}$ and $\tilde{\mathbf{X}}_j^{(l)}$ are high-degree steerable features that transform equivariantly under $E(3)$. The \circ operation between invariant scalars and equivariant tensors preserves equivariance. Specifically, since \circ multiplies the same invariant values over the spatial dimension of the high-degree tensors, this operation preserves equivariance. To see this, let $g \in E(3)$ be a transformation, s be an invariant scalar, and T be an equivariant tensor. Then:

$$g(s \circ T) = g(sT) = sg(T) = s \circ g(T).$$

This shows that the \circ operation commutes with the group action, preserving equivariance. The permutation-invariant aggregation function \bigoplus (such as summation or averaging) preserves equivariance because it operates independently on each degree of the tensor, maintaining their transformation properties under $E(3)$. Therefore, the scalar $\Delta \mathbf{h}_i$ remains invariant, while the high-degree steerable features $\tilde{\mathbf{X}}_i^{(l)}$ transform equivariantly under $E(3)$. Thus, we have shown that the interaction layer preserves the equivariance of the input representations under the action of the Euclidean group $E(3)$. Thus, we have shown that the interaction layer preserves the equivariance of the input representations under the action of the Euclidean group $E(3)$.

Next, we will prove that the HTR component preserves equivariance under the action of the Euclidean group $E(3)$. We'll consider each operation in the HTR component. Tensor projections as defined in Equation (10), we have:

$$\widetilde{\mathbf{EQ}}_i^{(l)} = \tilde{\mathbf{X}}_i^{(l)} \mathbf{W}_{vq}, \quad \widetilde{\mathbf{EK}}_j^{(l)} = \tilde{\mathbf{X}}_j^{(l)} \mathbf{W}_{vk}^{(l)}, \quad \text{for } l \in \{1, \dots, L_{\max}\},$$

$\tilde{\mathbf{X}}_i^{(l)}$ and $\tilde{\mathbf{X}}_j^{(l)}$ are high-degree tensors that transform equivariantly under $E(3)$. The projection matrices \mathbf{W}_{vq} and $\mathbf{W}_{vk}^{(l)}$ apply uniform weights across the spatial dimensions for each representation

864 dimension. This uniform application preserves equivariance because it commutes with the action of
865 $E(3)$. Let $g \in E(3)$ be a transformation:

$$866 \quad g(\tilde{\mathbf{X}}_i^{(l)} \mathbf{W}_{vq}) = (g(\tilde{\mathbf{X}}_i^{(l)}))(\mathbf{W}_{vq}) = (g(\tilde{\mathbf{X}}_i^{(l)} \mathbf{W}_{vq})).$$

867 The same holds for $\widetilde{\mathbf{E}\mathbf{K}}_j^{(l)}$. Therefore, $\widetilde{\mathbf{E}\mathbf{Q}}_i^{(l)}$ and $\widetilde{\mathbf{E}\mathbf{K}}_j^{(l)}$ transform equivariantly under $E(3)$.
868 Aggregation of angular and magnitude information from Equation (11), we have:

$$869 \quad \mathbf{w}_{ij} = \text{Agg}_{l=1}^{L_{\max}} \left((\widetilde{\mathbf{E}\mathbf{Q}}_i^{(l)})^\top \widetilde{\mathbf{E}\mathbf{K}}_j^{(l)} \right).$$

870 The operation $(\widetilde{\mathbf{E}\mathbf{Q}}_i^{(l)})^\top \widetilde{\mathbf{E}\mathbf{K}}_j^{(l)}$ is an inner product between equivariant tensors. Crucially, both
871 $\widetilde{\mathbf{E}\mathbf{Q}}_i^{(l)}$ and $\widetilde{\mathbf{E}\mathbf{K}}_j^{(l)}$ are subject to the same global rotation and translation under the action of $E(3)$.
872 Let $g \in E(3)$ be a transformation. Then:

$$873 \quad \begin{aligned} 874 \quad (g(\widetilde{\mathbf{E}\mathbf{Q}}_i^{(l)}))^\top g(\widetilde{\mathbf{E}\mathbf{K}}_j^{(l)}) &= (\mathbf{D}^{(l)}(r) \cdot \widetilde{\mathbf{E}\mathbf{Q}}_i^{(l)})^\top (\mathbf{D}^{(l)}(r) \cdot \widetilde{\mathbf{E}\mathbf{K}}_j^{(l)}) \\ 875 &= (\widetilde{\mathbf{E}\mathbf{Q}}_i^{(l)})^\top (\mathbf{D}^{(l)}(r))^\top \mathbf{D}^{(l)}(r) \cdot \widetilde{\mathbf{E}\mathbf{K}}_j^{(l)} \\ 876 &= (\widetilde{\mathbf{E}\mathbf{Q}}_i^{(l)})^\top \widetilde{\mathbf{E}\mathbf{K}}_j^{(l)} \cdot \eta^l. \end{aligned}$$

877 Here, $\mathbf{D}^{(l)}(r)$ represents the Wigner D-matrix of degree l , which is the appropriate representation
878 for the transformation of spherical tensors under rotations. The matrix $\mathbf{D}^{(l)}(r)$ is unitary, meaning
879 $(\mathbf{D}^{(l)}(r))^\top \mathbf{D}^{(l)}(r) = \mathbf{I}$. The factor $\eta^l = \pm 1$ accounts for the parity transformation, where $\eta = -1$
880 for improper rotations and $\eta = 1$ for proper rotations, with the exponent l determining the overall
881 sign based on the spherical harmonic degree. This demonstrates that the inner product transforms
882 covariantly under the full $O(3)$ symmetry group, with the parity factor properly accounting for
883 improper rotations. The aggregation \oplus over these covariant scalars preserves the transformation
884 properties, ensuring that w_{ij} transforms appropriately under both rotations and inversions. The edge
885 representation update is given by:

$$886 \quad \mathbf{t}_{ij} = \mathbf{w}_{ij} \mathbf{W}_{edp} + \mathbf{t}_{ij} \mathbf{W}_{eru},$$

887 w_{ij} is invariant, as shown in the previous step, and \mathbf{t}_{ij} is an edge embedding, which is also invariant
888 under $E(3)$. Since both w_{ij} and \mathbf{t}_{ij} are invariant scalars, their multiplication with the weight matrices
889 \mathbf{W}_{edp} and \mathbf{W}_{eru} preserves invariance. The addition of two invariant quantities also results in an
890 invariant quantity. Therefore, we can conclude:

$$891 \quad g(\mathbf{t}_{ij}) = g(\mathbf{w}_{ij} \mathbf{W}_{edp} + \mathbf{t}_{ij} \mathbf{W}_{eru}) = \mathbf{w}_{ij} \mathbf{W}_{edp} + \mathbf{t}_{ij} \mathbf{W}_{eru} = \mathbf{t}_{ij}$$

892 Thus, \mathbf{t}_{ij} is invariant under $E(3)$. Thus, we have shown that the Hierarchical Tensor Refinement
893 component preserves the equivariance of the input representations under the action of the Euclidean
894 group $E(3)$. The high-degree tensors transform equivariantly, while the scalar quantities and edge
895 embeddings remain invariant.

896 Finally, we will prove that the EQFF (Equivariant Feed-Forward) blocks preserve equivariance under
897 the action of the Euclidean group $E(3)$. Consider each operation in the EQFF component as described
898 in Equation (13). The input scalars \mathbf{h} is invariant under $E(3)$ as it represents scalar features, while
899 $\tilde{\mathbf{X}}$ transforms equivariantly under $E(3)$ as it represents high-degree steerable features. Next, we
900 examine the computation of \mathbf{m}_1 and \mathbf{m}_2 :

$$901 \quad \mathbf{m}_1, \mathbf{m}_2 = \text{split}_2(\gamma_m(\|\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu}\|_2, \mathbf{h})).$$

902 The operation $\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu}$ preserves equivariance as it applies the same linear transformation across all
903 spatial dimensions. The L_2 norm of this equivariant tensor field, $\|\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu}\|_2$, is invariant under
904 $E(3)$. To see this, let $g \in E(3)$ be a transformation. Then $\|g(\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu})\|_2 = \|R \cdot (\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu})\|_2 =$
905 $\|\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu}\|_2$, where R is the rotation matrix corresponding to g . The translation component doesn't
906 affect the norm. The concatenation $(\|\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu}\|_2, \mathbf{h})$ is of two invariant quantities, resulting in
907 an invariant vector. As γ_m is applied to an invariant input, its output is also invariant. Finally,
908 splitting this invariant vector results in invariant components \mathbf{m}_1 and \mathbf{m}_2 . Now, we analyze the
909 EQFF operation:

$$910 \quad \text{EQFF}(\mathbf{h}, \tilde{\mathbf{X}}^{(l)}) = \left((\mathbf{h} + \mathbf{m}_1), (\tilde{\mathbf{X}}^{(l)} + (\mathbf{m}_2 \circ \tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu})) \right).$$

The term $\mathbf{h} + \mathbf{m}_1$ is a sum of two invariant quantities, resulting in an invariant scalar. The operation $\mathbf{m}_2 \circ \tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu}$ preserves equivariance because \mathbf{m}_2 is an invariant scalar and $\tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu}$ is equivariant. The element-wise product of an invariant scalar with an equivariant tensor is equivariant. Consequently, $\tilde{\mathbf{X}}^{(l)} + (\mathbf{m}_2 \circ \tilde{\mathbf{X}}^{(l)} \mathbf{W}_{vu})$ is a sum of two equivariant tensors, resulting in an equivariant tensor. Hence, EQFF returns tuple of updated representations for scalar and high-degree steerable features.

Thus, we have shown that the EQFF operation preserves the equivariance of the input representations under the action of the Euclidean group $E(3)$. The scalar part remains invariant, while the high-degree steerable features transform equivariantly.

Hence, we have shown that each major component of the GotenNet architecture preserves the equivariance property under the action of the Euclidean group $E(3)$. The initial node and edge embeddings are invariant under $E(3)$. The GATA module, including its self-attention mechanism and geometric encoding, maintains the invariance of scalar quantities and the equivariance of high-degree steerable tensors. The Hierarchical Tensor Refinement (HTR) component preserves equivariance in its tensor projections and ensures that edge updates remain invariant. Finally, the EQFF blocks maintain the overall equivariance structure by preserving the invariance of scalar parts and the equivariance of high-degree steerable features. In each of these components, we have demonstrated how the various operations interact with the group action of $E(3)$ to preserve the required invariance and equivariance properties. □

C PROOF: EQUIVARIANCE OF GOTENNET

Equivariance of GotenNet: If all layers of GotenNet are equivariant, then the entire network is equivariant.

Proof. We will prove this by induction on the number of layers in the network.

Base Case (Layer 1): Let $T^1 : M \rightarrow \mathcal{T}^1$ represent the output of the first layer of the network. Since the first layer is equivariant by assumption, we have that for all $g \in E(3)$,

$$(g \cdot T^1)(p) = \rho(g)T^1(g^{-1} \cdot p).$$

Thus, the first layer preserves equivariance.

Inductive Step: Assume that the output of layer l , denoted by T^l , is equivariant. That is, for all $g \in E(3)$,

$$(g \cdot T^l)(p) = \rho(g)T^l(g^{-1} \cdot p).$$

We need to show that layer $l + 1$, denoted by T^{l+1} , is also equivariant. Let Φ^{l+1} represent the operation of the $(l + 1)$ -th layer. Then $T^{l+1} = \Phi^{l+1} \circ T^l$. Since the $(l + 1)$ -th layer is also assumed to be equivariant, we have:

$$(g \cdot T^{l+1})(p) = \rho(g)T^{l+1}(g^{-1} \cdot p).$$

The equivariance of T^{l+1} follows from the equivariance of T^l and Φ^{l+1} , as the composition of equivariant functions is also equivariant. Explicitly:

$$\begin{aligned} (g \cdot T^{l+1})(p) &= (g \cdot (\Phi^{l+1} \circ T^l))(p) \\ &= \Phi^{l+1}((g \cdot T^l)(p)) \\ &= \Phi^{l+1}(\rho(g)T^l(g^{-1} \cdot p)) \\ &= \rho(g)\Phi^{l+1}(T^l(g^{-1} \cdot p)) \\ &= \rho(g)T^{l+1}(g^{-1} \cdot p) \end{aligned}$$

Hence, by the principle of mathematical induction, we have shown that if all individual layers in the GotenNet are equivariant, then for any number of layers, the final output of the network, being the composition of these equivariant layers, is also equivariant. This result relies on the property

that the composition of equivariant functions is itself equivariant. Specifically, we have shown that $T^{l+1} = \Phi^{l+1} \circ T^l$, and the equivariance of both Φ^{l+1} and T^l ensures the equivariance of T^{l+1} . Hence, the entire network is equivariant. This theorem, combined with the previous proof of component-wise equivariance (see Appendix B), establishes the overall equivariance of the GotenNet architecture, ensuring its consistency under Euclidean transformations of the input space. \square

D REPRODUCIBILITY STATEMENT

The details on components of the architecture, hyper-parameters, and model variations are outlined in Section E. The code used to reproduce the experiments will be available. All datasets used in this study are publicly available; the access instructions will be included in the source code. We have included information on the computational resources used for our experiments, including hardware specifications and software versions, to facilitate reproducibility of our results.

Table 6: Hyper-parameters for the datasets GotenNet compared against the baselines. The parameters are for GotenNet_B if multiple variations exists.

Hyper-parameters	QM9	Molecule3D	MD22	rMD17
Optimizer	AdamW	AdamW	AdamW	AdamW
Learning rate scheduling	Linear warmup with reduce on plateau			
Warmup steps	10,000	5,000	1,000	1,000
Maximum learning rate	[6e-5, 1e-4]	1e-4	[4e-5, 1e-4]	2e-4
Learning rate decay	0.8	0.8	0.8	0.8
Learning rate patience	15	5	30	30
Loss function	MSE	L_1	MSE	MSE
Gradient clipping	10	-	5	10
Batch size	32	256	4	4
Number of epochs	1,000	300	3,000	3,000
Weight decay	0.01	0.01	0.01	0.01
Dropout rate	0.1	0.1	0.1	0.1
Node dimension (d_{ne})	256	384	[256, 384]	192
Edge dimension (d_{ed})	256	384	[256, 384]	192
Edge refinement dimension (d_{xpd})	256	384	768	768
L_{\max}	2	2	2	2
Number of Layers	6	12	[6, 8]	12
Number of RBFs	64	32	32	64
Number of Attention Heads	8	8	8	8
Cutoff radius	5.0	5.0	[4.0, 5.0]	5.0

E TRAINING DETAILS AND HYPER-PARAMETERS

Table 6 presents the comprehensive set of hyper-parameters employed in our experiments across various datasets. These parameters were carefully selected to optimize model performance and ensure fair comparisons with baseline methods. For the GotenNet architecture, we primarily report the parameters for the base (GotenNet_B) variation where multiple model sizes exist. The optimization process utilized the AdamW optimizer across all datasets, coupled with a linear warmup strategy and learning rate reduction on plateau. This adaptive learning rate approach allows for more stable training and improved convergence.

The learning rates were fine-tuned for each dataset, with QM9 and MD22 employing a range of maximum learning rates to account for the diverse nature of their target properties. Molecule3D and rMD17 datasets, on the other hand, used fixed maximum learning rates of 1e-4 and 2e-4, respectively. To mitigate overfitting and promote generalization, we implemented weight decay (0.01) and dropout (0.1) consistently across all datasets. The choice of loss function varied, with Mean Squared Error (MSE) being the predominant choice, except for Molecule3D, which utilized the L1 loss.

The model architecture parameters, such as d_{ne} , d_{ed} , and d_{xpd} , were adjusted based on the complexity of the dataset and the specific prediction tasks. For instance, MD22 and rMD17 datasets, which involve more complex molecular dynamics simulations, employed larger edge refinement dimensions (768) compared to QM9 and Molecule3D (256 and 384, respectively). The number of layers in the model also varied, with Molecule3D and rMD17 using deeper architectures (12 layers) compared to QM9 and MD22 (6 layers for the base model).

It is worth noting that for QM9, we experimented with different model sizes by varying the number of layers (4, 6, and 12 for S, B, and L variations, respectively). Similarly, for MD22, we explored a more compact model variation by reducing the number of layers to 4 and halving the representation dimensions. These variations allow us to investigate the trade-offs between model complexity and performance across different molecular property prediction tasks.

The consistent use of 8 attention heads and a L_{\max} of 2 across all datasets suggests that these parameters provide a good balance between computational efficiency and model expressiveness for a wide range of molecular modeling tasks. The cutoff radius, predominantly set at 5.0 Å (with some variations in MD22), was chosen to capture relevant atomic interactions while maintaining computational feasibility.

F EFFICIENCY EXPERIMENT DETAILS

The efficiency experiments are conducted by sampling graphs from the Molecule3D dataset. Using a real-world dataset provides realistic neighborhoods for the nodes computed with a radius graph, which is more beneficial than using a synthetic dataset. The graph sampling process involves selecting graphs with a predetermined node count. If no graph is available for a given node count, we sample the larger graphs with the closest node count and subsample the nodes to satisfy the node limitation. If the number of graphs is less than desired, we oversample the graphs for experimentation.

After creating a dataset for each node count, we conduct experiments by first warming up the models without timing for 5,000 steps. Then, we start timing the forward and backward passes for the subsequent 5,000 steps. The final values are obtained by averaging the timings over the total number of steps. It is important to note that the timings are measured per batch. Each batch consists of a fixed number of graphs, and the reported timings represent the average time taken to process a single batch during the training process.

This experimental setup ensures a fair comparison of the scalability properties of GotenNet and the baseline models, providing insights into their efficiency in handling graphs of varying sizes. By measuring the training time per batch, we can accurately assess how the computational overhead scales with increasing graph sizes, independent of the number of epochs or the total dataset size.

G COMPUTATIONAL COMPLEXITY ANALYSIS

We present a comprehensive analysis of computational efficiency across state-of-the-art models, examining training requirements, inference speed, and model complexity. Table 7 compares these metrics across different architectures under standardized conditions.

Our analysis reveals several key insights about computational efficiency across models. First, training protocols show significant variation across architectures - from Equiformer’s relatively lightweight approach to Geformer’s more intensive requirements. Equiformer (Liao & Smidt, 2023) and EquiformerV2 (Liao et al., 2024) employ 300 epochs with batch sizes of 128 and 48/64 respectively, requiring 1.48 and 2.85/2.94 GPU days for completion. Geformer (Wang et al., 2023a) utilizes a batch size of 32 for up to 600 epochs with early stopping, theoretically requiring 5.05 GPU days for full training. This variation reflects different trade-offs between computational demands and model expressiveness.

GotenNet demonstrates superior efficiency across multiple metrics. The smallest variant, GotenNet_S, achieves competitive performance with just 6.1M parameters while requiring minimal computational resources (0.75 GPU days average training time). This efficiency extends to both training and inference latencies, with GotenNet_S achieving the lowest latencies in both categories (80ms and 37ms respectively).

Table 7: Computational complexity comparison of different methods. Training and inference latency are measured with batch size of 128 samples. Training time is reported in GPU days (min/avg/max/limit). Time per epoch is in seconds. Latency measurements are in milliseconds. Best results are in **bold**.

Model	Batch Size	Time per Epoch (s)	Training Time (GPU days)				Training Latency	Inference Latency	Trainable Parameters	Std. MAE	Log MAE
			min	avg	max	limit					
Equiformer	128	425	1.48	1.48	1.48	1.48	421	150	3.5M	0.70	-5.82
EquiformerV2	64	821	2.85	2.85	2.85	2.65	918	341	11.2M	0.67	-5.87
EquiformerV2	48	847	2.94	2.94	2.94	2.65	918	341	11.2M	0.67	-5.87
Geoformer	32	436	-	-	-	5.05	759	264	50.6M	0.75	-6.12
GotenNet _S	32	117	0.41	0.75	1.34	1.35	80	37	6.1M	0.67	-6.21
GotenNet _B	32	180	0.75	1.15	1.92	2.08	120	56	9.2M	0.61	-6.26
GotenNet _L	32	291	1.37	1.87	2.33	3.37	244	112	18.3M	0.56	-6.34

Notably, even as model capacity increases, GotenNet maintains its efficiency advantages. GotenNet_L (18.3M parameters) demonstrates remarkable scalability, requiring only 1.87 GPU days average training time while achieving 42% faster inference and 25% faster training compared to the closest competitor, Equiformer.

These results demonstrate that GotenNet’s architectural innovations - particularly its efficient handling of geometric tensor representations - translate to practical advantages across all operational metrics. The consistent performance improvements across model scales suggest that GotenNet’s approach to balancing expressiveness and efficiency is fundamentally sound, making it particularly suitable for real-world applications where computational resources are constrained.

H COMPREHENSIVE PERFORMANCE COMPARISON ON QM9 DATASET

Table 8 provides a comprehensive comparison of the performance of various baseline models against our proposed GotenNet on the QM9 dataset. This extended table includes a full list of baseline methods, offering a detailed assessment across all molecular property prediction tasks. The results showcase how GotenNet consistently surpasses existing models in both energy and force predictions, further highlighting its robustness and scalability. By including a wider range of baseline comparisons in this appendix, we aim to give a clearer view of GotenNet’s advantages in different metrics and provide a more exhaustive evaluation for the QM9 dataset.

I MD22 VISUALIZATIONS

Figure 4 visualizes the mean absolute errors (MAE) for energy and forces across molecules in the MD22 dataset. The x-axis represents the energy error (kcal/mol), and the y-axis denotes the force error (kcal/mol/Å). Each point corresponds to a model, with performance improving as the point approaches the origin (0,0), where lower values indicate better performance for both metrics. Our proposed GotenNet consistently outperforms the baseline models, achieving the best performance across all molecules, as evidenced by its closer proximity to the origin compared to competing methods.

J RMD17 VISUALIZATIONS

Figure 5 presents the MAE for energy and force predictions across nine molecules in the rMD17 dataset, including Aspirin, Azobenzene, Benzene, Ethanol, Malonaldehyde, Naphthalene, Paracetamol, Uracil, and Toluene. The x-axis represents energy error (kcal/mol), and the y-axis denotes force error (kcal/mol/Å). Each point corresponds to a model’s performance on a specific molecule, where better performance is indicated by proximity to the origin (0,0) — reflecting lower errors in both metrics. Our proposed GotenNet demonstrates consistent superiority over baseline models, achieving the lowest errors across all nine molecules, as evidenced by its closer alignment with the origin compared to other methods.

Table 8: Performance comparisons on QM9 dataset. † denotes using different data partitions.

Task	α	$\Delta\varepsilon$	$\varepsilon_{\text{HOMO}}$	$\varepsilon_{\text{LUMO}}$	μ	C_ν	G	H	R^2	U	U_0	ZPVE	std.	log
Units	ma_0^3	meV	meV	meV	mD	$\frac{\text{mcal}}{\text{mol}\cdot\text{K}}$	meV	meV	ma_0^2	meV	meV	meV	%	-
Invariant models														
Cormorant	85	61	34	38	38	26	20	21	961	21	22	2.03	2.14	-4.75
NMP	92	69	43	38	30	40	19	17	180	20	20	1.50	1.78	-5.08
DimeNet++†	44	32.6	24.6	19.5	29.7	23	7.56	6.53	331	6.28	6.32	1.21	0.98	-5.67
ComENet†	45	32.4	23.1	19.8	24.5	22	7.98	6.86	259	6.82	6.69	1.20	0.93	-5.69
SphereNet†	46	31.1	22.8	18.9	24.5	22	7.78	6.33	268	6.36	6.26	1.12	0.91	-5.73
Scalarization-based models														
ClofNet	63	53	33	25	40	27	9	9	610	9	8	1.23	1.37	-5.37
EGNN	71	48	29	25	29	31	12	12	106	12	11	1.55	1.23	-5.43
PaiNN†	45	45.7	27.6	20.4	12.0	24	7.35	5.98	66	5.83	5.85	1.28	1.01	-5.85
LEFTNet	48	40	24	18	12	23	7	6	109	7	6	1.33	0.91	-5.82
EQGAT	53	32	20	16	11	24	23	24	382	25	25	2.00	0.86	-5.28
ET	59	36.1	20.3	17.5	11	26	7.62	6.16	33	6.38	6.15	1.84	0.84	-5.90
Geoformer	40	33.8	18.4	15.4	10	22	6.13	4.39	28	4.41	4.43	1.28	0.75	-6.12
High-order steerable models														
SEGNN	60	42	24	21	23	31	15	16	660	13	15	1.62	1.08	-5.27
HDGNN†	46	32	18	16	17	23	11	10	342	8.12	8.34	1.21	0.80	-5.64
Equiformer	46	30	15.4	14.7	12	23	7.63	6.63	251	6.74	6.59	1.26	0.70	-5.82
Equiformer _{v2}	47	29.0	14.4	13.3	9.9	23	7.57	6.22	186	6.49	6.17	1.47	0.67	-5.87
Pre-trained models														
Transformer-M	37	27.4	17.5	16.2	37	22	9.63	9.39	75	9.41	9.37	1.18	0.86	-5.74
SE(3)-DDM	46	40.2	23.5	19.5	15	24	7.65	7.09	122	6.99	6.92	1.31	0.93	-5.76
3D-EMGP	57	37.1	21.3	18.2	20	26	9.30	8.70	92	8.60	8.60	1.38	0.92	-5.68
Coord	52	31.8	17.7	14.3	12	20	6.91	6.45	450	6.11	6.57	1.71	0.76	-5.75
Frad(VRN)	42	27.7	17.9	13.8	11	21	6.03	6.01	354	5.35	5.41	1.63	0.71	-5.85
Frad(RN)	37	27.8	15.3	13.7	10	20	6.19	5.55	342	5.62	5.33	1.42	0.66	-5.91
Hybrid geometric models														
GotenNet _S	35	23.2	16.3	14.7	7.5	20	5.51	3.86	26	3.76	3.82	1.15	0.62	-6.27
GotenNet _B	33	21.3	15.2	13.5	7.3	20	5.33	3.52	25	3.49	3.49	1.10	0.58	-6.33
GotenNet _L	30	19.9	13.7	12.2	7.7	19	4.98	3.36	21	3.33	3.37	1.08	0.54	-6.39

K FUTURE WORK AND EXTENSIONS

While GotenNet has demonstrated strong performance in molecular property prediction, several promising directions exist for future research and extensions. The architecture can naturally extend to other spatial data where geometric relationships significantly influence node interactions, such as point cloud processing, protein structure analysis, and dynamic molecular simulations. These applications share the fundamental requirement of processing geometric relationships while preserving symmetries, making them natural candidates for our framework. From an architectural perspective, the model could be enhanced through the incorporation of scale equivariance, exploration of higher-order features beyond second degree, and development of sparse implementations for larger systems. Memory efficiency improvements could also enable applications to even larger-scale systems. On the theoretical front, future work could focus on developing formal analyses of the expressiveness-efficiency trade-off, understanding generalization properties of geometric tensor representations. These potential extensions maintain GotenNet’s core principle of balancing expressiveness and efficiency while broadening its applicability across different domains of geometric deep learning.

1188
 1189
 1190
 1191
 1192
 1193
 1194
 1195
 1196
 1197
 1198
 1199
 1200
 1201
 1202
 1203
 1204
 1205
 1206
 1207
 1208
 1209
 1210
 1211
 1212
 1213
 1214
 1215
 1216
 1217
 1218
 1219
 1220
 1221
 1222
 1223
 1224
 1225
 1226
 1227
 1228
 1229
 1230
 1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241

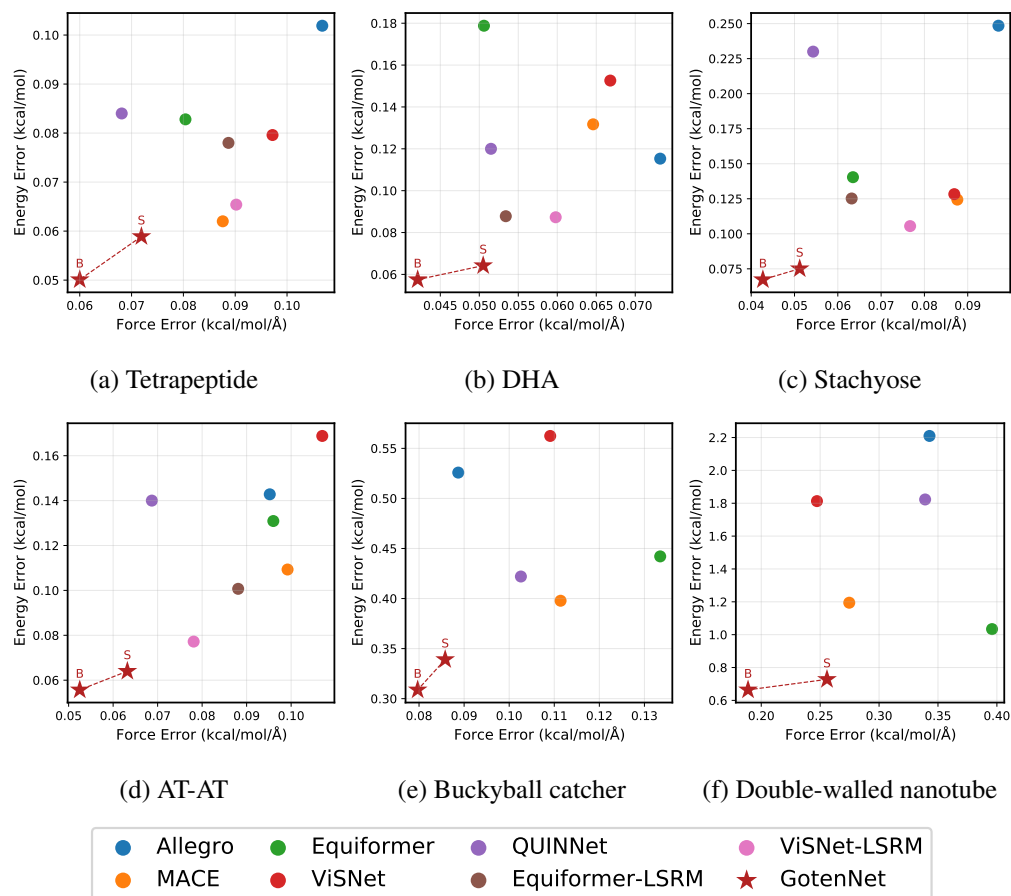


Figure 4: Mean absolute error of the molecules for energy and forces.

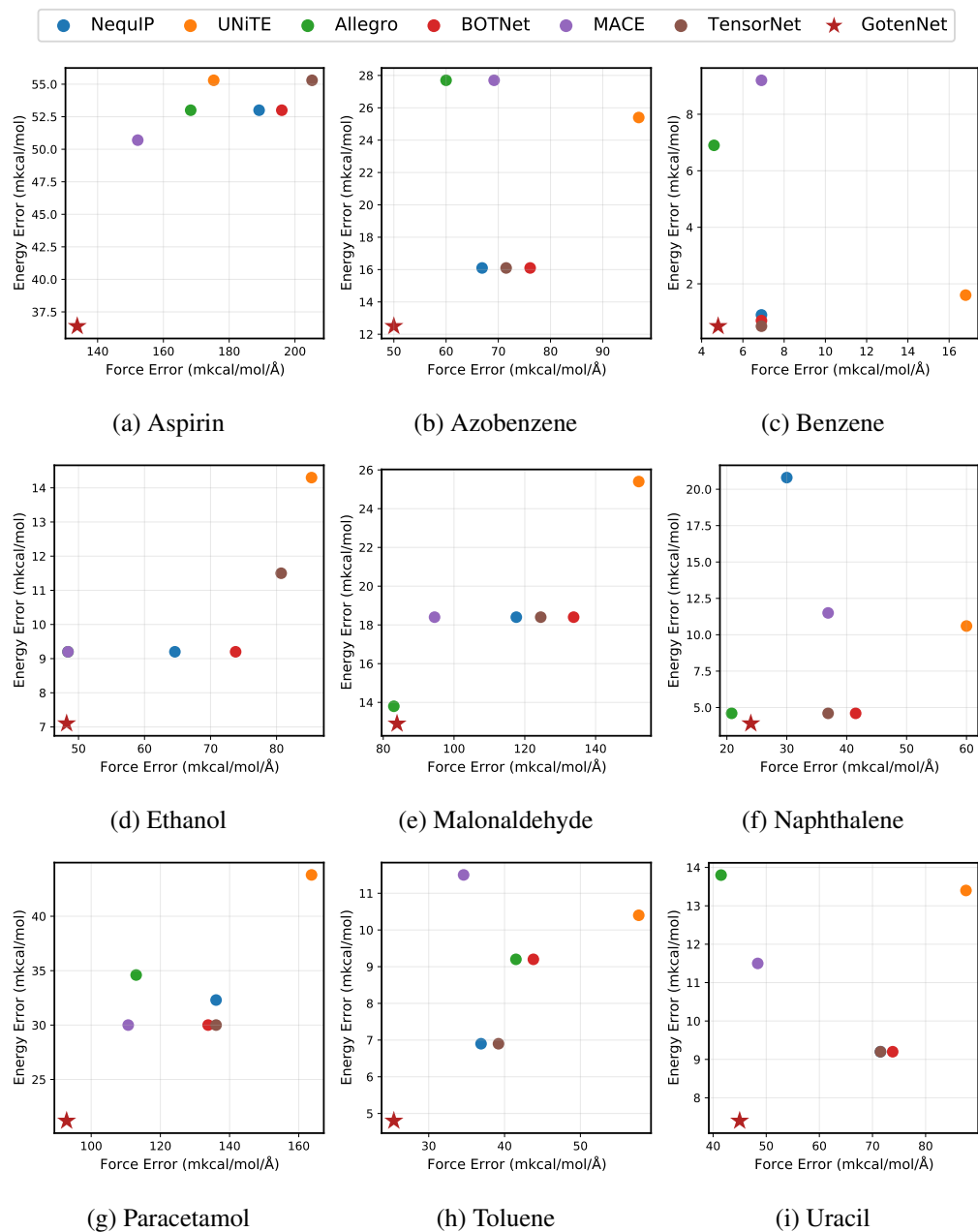


Figure 5: Mean absolute error of the molecules on rMD17 dataset for energy and forces.