

ViSNet: an equivariant geometry-enhanced graph neural network with vector-scalar interactive message passing for molecules

Yusong Wang^{1,2,3†}, Shaoning Li^{2,3†}, Xinheng He^{4,5,2,3}, Mingyu Li^{6,2,3}, Zun Wang², Nanning Zheng¹, Bin Shao^{2*}, Tie-Yan Liu² and Tong Wang^{2*}

¹Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, 710049, China.

²Microsoft Research AI4Science, Beijing, 100080, China.

³Work done during an internship at Microsoft Research AI4Science, Beijing, 100080, China.

⁴The CAS Key Laboratory of Receptor Research and State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China.

⁵University of Chinese Academy of Sciences, Beijing, 100049, China.

⁶Medicinal Chemistry and Bioinformatics Center, School of Medicine, Shanghai Jiaotong University, Shanghai, 200025, China.

*Corresponding author(s). E-mail(s): binshao@microsoft.com (B. S.);
watong@microsoft.com (T. W., Lead Contact);

†These authors contributed equally to this work.

Abstract

Geometric deep learning has been revolutionizing the molecular modeling field. Despite the state-of-the-art neural network models are approaching *ab initio* accuracy for molecular property prediction, their applications, such as drug discovery and molecular dynamics (MD) simulation, have been hindered by insufficient utilization of geometric information and high computational costs. Here we propose an equivariant geometry-enhanced graph neural network called ViSNet, which elegantly extracts geometric features and efficiently models molecular structures with low computational costs. Our proposed ViSNet outperforms state-of-the-art approaches on multiple MD benchmarks, including MD17, revised MD17 and MD22, and achieves excellent chemical property prediction on QM9 and Molecule3D datasets. Additionally, ViSNet achieved the top winners of PCQM4Mv2 track in the OGB-LCS@NeurIPS2022 competition. Furthermore, through a series of simulations and case studies, ViSNet can efficiently explore the conformational space and provide reasonable interpretability to map geometric representations to molecular structures.

Keywords: Geometric Deep Learning Potential; Equivariant Graph Neural Network; Molecular Modeling

1 Introduction

Molecular modeling plays a crucial role in modern scientific and engineering fields, aiding in the understanding of chemical reactions, facilitating new drug development, and driving scientific and technological advancements [1–4]. One commonly used method in molecular modeling is density functional theory (DFT). DFT enables accurate calculations of energy, forces, and other chemical properties of molecules [5, 6]. However, due to the large computational requirements, DFT calculations often demand significant computational resources and time, particularly for large molecular systems or high-precision calculations. Machine learning (ML) offers an alternative solution by learning from reference data with *ab initio* accuracy and high computational efficiency [7, 8]. Behler and Parrinello [9] were the first to introduce descriptors for characterizing atomic local environments combined with a shallow multi-layer perceptron to learn the potential energy of molecules. In recent years, deep learning (DL) has demonstrated its powerful ability to learn from raw data without any hand-crafted features in many fields and thus attracted more and more attention. However, the inherent drawback of deep learning, which requires large amounts of data, has become a bottleneck for its application to more scenarios [10]. To alleviate the dependency on data for DL potentials, recent works have incorporated the inductive bias of symmetry into neural network design, known as geometric deep learning (GDL). Symmetry describes the conservation of physical laws, i.e., the unchanged physical properties with any transformations such as translations or rotations. It allows GDL to be extended to limited data scenarios without any data augmentation.

Equivariant graph neural network (EGNN) is one of the representative approaches in GDL, which has extensive capability to model molecular geometry [10–19]. A popular kind of EGNN conducts equivariance from directional information and involves geometric features to predict molecular properties. GemNet [18] extends the invariant DimeNet/DimeNet+ [14, 15] with dihedral information. They explicitly extract geometric information in the Euclidean space with 1st-order geometric tensor, i.e., setting $l_{max} = 1$. PaiNN [16] and Equivariant Transformer [17] further adopt

vector embedding and scalarize the angular representation implicitly via the inner product of the vector embedding itself. They reduce the complexity of explicit geometry extraction by taking the angular information into consideration. Another mainstream approach to achieving equivariance is through group representation theory, which can achieve higher accuracy but comes with large computational costs. NequIP, Allegro, and MACE [10, 20, 21] achieve state-of-the-art performance on several molecular dynamics simulation datasets leveraging high-order geometric tensors. On the one hand, algorithms based on group representation theory have strong mathematical foundations and are able to fully utilize geometric information using high-order geometric tensors. On the other hand, these algorithms often require computationally expensive operations such as the Clebsch-Gordan product (CG-product) [22], making them possibly suitable for periodic systems with elaborate model design but impractical for large molecular systems such as chemical and biological molecules without periodic boundary conditions.

In this study, we propose ViSNet (short for “Vector-Scalar interactive graph neural Network”), which alleviates the dilemma between computational costs and sufficient utilization of geometric information. By incorporating an elaborate Runtime Geometry Calculation (RGC) strategy, ViSNet implicitly extracts various geometric features, i.e., angles, dihedral torsion angles, and improper angles in accordance with the force field of classical MD with linear time complexity, thus significantly accelerating model training and inference while reducing the memory consumption. To extend the vector representation, we introduce spherical harmonics and simplify the computationally expensive Clebsch-Gordan product with the inner product. Furthermore, we present a well-designed Vector-Scalar interactive equivariant Message Passing (ViS-MP) mechanism, which fully utilizes the geometric features by interacting vector hidden representations with scalar ones. When comprehensively evaluated on some benchmark datasets, ViSNet outperforms all state-of-the-art algorithms on all molecules in MD17, revised MD17 and MD22 datasets and shows superior performance on QM9, Molecule3D dataset indicating the powerful capability of molecular geometric

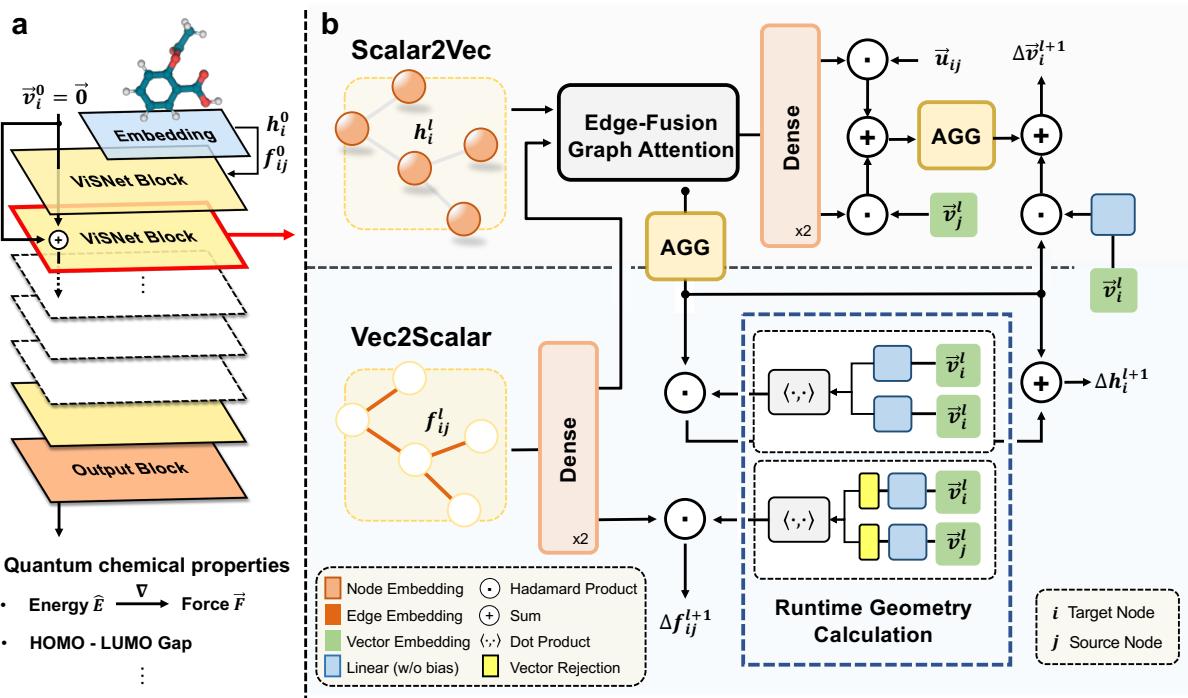


Fig. 1 The overall architecture of ViSNet. (a) Model sketch of ViSNet. ViSNet embeds the 3D structures of molecules and extracts the geometric information through a series of ViSNet blocks and outputs the molecule properties such as energy, forces, and HOMO-LUMO gap through an output block. (b) Flowchart of one ViSNet Block. One ViSNet block consists of two modules: i) *Scalar2Vec*, responsible for attaching scalar embeddings to vectors; ii) *Vec2Scalar*, renovates scalar embeddings built on RGC strategy. The inputs of *Scalar2Vec* are the node embedding h_i , edge embedding f_{ij} , direction unit \vec{v}_i and the relative positions between two atoms. The edge-fusion graph attention module (serves as ϕ_m^s) takes as input h_i and the output of the dense layer following f_{ij} , and outputs scalar messages. Before aggregation, each scalar message is transformed through a dense layer, then fused with the unit of the relative position \vec{u}_{ij} and its own direction unit \vec{v}_j . We further compute the vector messages and aggregate them all among the neighborhood. Through a gated residual connection, the final residual $\Delta\vec{v}_i$ is produced. In *Vec2Scalar* module, by Hadamard production of aggregated scalar messages and the output of RGC-Angle calculation and adding a gated residual connection, the final Δh_i is figured out. Likewise, combining the projected f_{ij} and the output of RGC-Dihedral calculation, the final Δf_{ij}^{l+1} is determined.

representation. ViSNet also has won PCQM4Mv2 track in the OGB-LCS@NeurIPS2022 competition (<https://ogb.stanford.edu/neurips2022/results/>). We then performed molecular dynamics simulations for each molecule on MD17 driven by ViSNet trained only with limited data (950 samples). The highly consistent interatomic distance distributions and the explored potential energy surfaces between ViSNet and quantum simulation illustrate that ViSNet is genuinely data-efficient and can perform simulations with high fidelity. To further explore the usefulness of ViSNet to real-world applications, we used an in-house dataset that consists of about 10,000 different conformations of the 166-atom protein Chignolin derived from replica exchange molecular dynamics and calculated at DFT-level. When evaluated on the

dataset, ViSNet also achieved significantly better performance than empirical force fields, and the simulations performed by ViSNet exhibited very close force calculation to DFT. In addition, ViSNet exhibits reasonable interpretability to map geometric representation to molecular structures. The contributions of ViSNet can be summarized as follows:

- Proposing RGC module that utilizes high-order geometric tensors to implicitly extract various geometric features, including angles, dihedral torsion angles, and improper angles, with linear time complexity.
- Introducing ViS-MP mechanism to enable efficient interaction between vector hidden representations and scalar ones and fully exploit the geometric information.

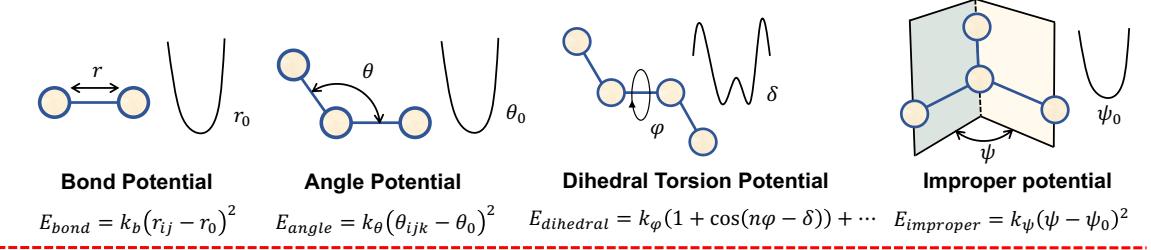
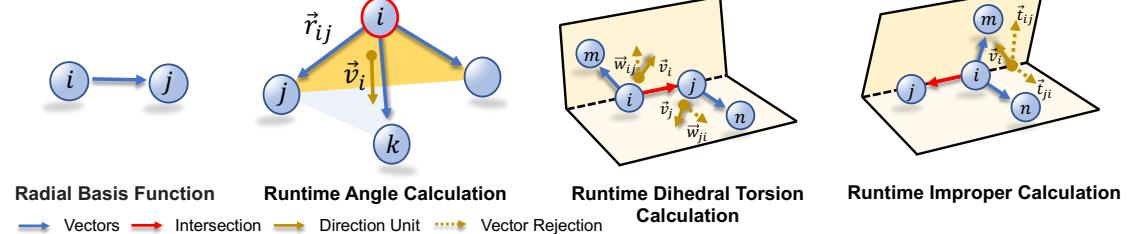
Bonded Term in Classical MD**RGC in ViSNet**

Fig. 2 Illustration of Runtime Geometry Calculation (RGC) module and its relevance to the potential of bonded terms in classical molecular dynamics. The bonded terms consist of bond length, bond angle, dihedral torsion and improper angle. The RGC module depicts all bonded terms of classical MD as model operations in linear time complexity. Yellow arrow \vec{v}_i denotes the direction unit in Eq. 1.

- Achieving the state-of-the-art performance in 6 benchmarks for predicting energy, forces, HOMO-LUMO gap, and other quantum properties of molecules.
- Performing molecular dynamics simulations driven by ViSNet on both small molecules and 166-atom Chignolin with high fidelity.
- Demonstrating reasonable model interpretability between geometric features and molecular structures.

2 Results

2.1 Overview of ViSNet

ViSNet is a versatile EGNN which predicts potential energy, atomic forces as well as various quantum chemical properties by taking atomic coordinates and numbers as inputs. As shown in Fig. 1(a), the model is composed of an embedding block and multiple stacked ViSNet blocks, followed by an output block. The atomic number and coordinates are fed into the embedding block followed by ViSNet blocks to extract and encode geometric representations. The geometric representations are then used to predict molecular properties through the output block. It is worth

noting that ViSNet is an energy-conserving potential, i.e., the predicted atomic forces are derived from the negative gradients of the potential energy with respect to the coordinates [23].

RGC: Runtime Geometry Calculation The success of classical force fields shows that geometric features such as interatomic distances, angles, and dihedral torsion angles, and improper angles in Fig. 2 are essential to determine the total potential energy of molecules. The explicit extraction of invariant geometric representations in previous studies often suffer from a large amount of time or memory consumption during model training and inference. Given an atom, the calculation of angular information scales $\mathcal{O}(N^2)$ with the number of neighboring atoms, while the computational complexity is even $\mathcal{O}(N^3)$ for dihedrals [18]. To alleviate this problem, inspired by [16], we propose runtime geometry calculation (RGC), which uses an equivariant vector representation (termed as “direction unit”) for each node to preserve its geometric information. RGC directly calculates the geometric information from the direction unit which only sums the vectors from the target node to its neighbors once. Therefore, the computational complexity can be reduced to $\mathcal{O}(N)$.

Considering the sub-structure of a toy molecule with four atoms shown in Fig. 2, the

angular information of the target node i could be conducted from the vector \vec{r}_{ij} as follows:

$$\vec{u}_{ij} = \frac{\vec{r}_{ij}}{\|\vec{r}_{ij}\|}, \quad \vec{v}_i = \sum_{j=1}^{N_i} \vec{u}_{ij} \quad (1)$$

$$\|\vec{v}_i\|^2 = \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \langle \vec{u}_{ij}, \vec{u}_{ik} \rangle = \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \cos \theta_{jik} \quad (2)$$

where \vec{r}_{ij} is the vector from node i to its neighboring node j , \vec{u}_{ij} is the unit vector of \vec{r}_{ij} . Here, we define the *direction unit* \vec{v}_i as the sum of all unit vectors from node i to its all neighboring nodes j , where node i is the intersection of all unit vectors. As shown in Eq. 2, we calculate the inner product of direction unit \vec{v}_i which represents the sum of inner products of unit vectors from node i to all its neighboring nodes. Combining with Eq. 1, the inner product of direction \vec{v}_i finally stands for the sum of cosine values of all angles formed by node i and any two of its neighboring nodes.

Similar to runtime angle calculation, we also calculate the vector rejection [24] of the direction unit \vec{v}_i of node i and \vec{v}_j of node j on the vector \vec{u}_{ij} and \vec{u}_{ji} , respectively.

$$\begin{aligned} \vec{w}_{ij} &= \text{Rej}_{\vec{u}_{ij}}(\vec{v}_i) = \vec{v}_i - \langle \vec{v}_i, \vec{u}_{ij} \rangle \cdot \vec{u}_{ij} \\ &= \sum_{m=1}^{N_i} \text{Rej}_{\vec{u}_{ij}}(\vec{u}_{im}) \\ \vec{w}_{ji} &= \text{Rej}_{\vec{u}_{ji}}(\vec{v}_j) = \vec{v}_j - \langle \vec{v}_j, \vec{u}_{ji} \rangle \cdot \vec{u}_{ji} \quad (3) \\ &= \sum_{n=1}^{N_j} \text{Rej}_{\vec{u}_{ji}}(\vec{u}_{jn}) \end{aligned}$$

where $\text{Rej}_{\vec{b}}(\vec{a})$ represents the vector component of \vec{a} perpendicular to \vec{b} , termed as the vector rejection. \vec{u}_{ij} and \vec{v}_i are defined in Eq. 1. \vec{w}_{ij} represents the sum of the vector rejection $\text{Rej}_{\vec{u}_{ij}}(\vec{u}_{im})$ and \vec{w}_{ji} represents the sum of the vector rejection $\text{Rej}_{\vec{u}_{ji}}(\vec{u}_{jn})$. The inner product between \vec{w}_{ij} and \vec{w}_{ji} is then calculated to conduct dihedral torsion angle information of the intersecting edge e_{ij} as

follows:

$$\begin{aligned} \langle \vec{w}_{ij}, \vec{w}_{ji} \rangle &= \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \left\langle \text{Rej}_{\vec{u}_{ij}}(\vec{u}_{im}), \text{Rej}_{\vec{u}_{ji}}(\vec{u}_{jn}) \right\rangle \\ &= \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \cos \varphi_{mijn} \end{aligned} \quad (4)$$

The improper angle is derived from a pyramid structure forming by 4 nodes. As the last toy molecule shown in Fig. 2, node i is the vertex of the pyramid, and the improper torsion angle is formed by two adjacent planes with an intersecting edge e_{ij} . We can also calculate the improper angle by vector rejection:

$$\begin{aligned} \vec{t}_{ij} &= \text{Rej}_{\vec{u}_{ij}}(\vec{v}_i) = \sum_{m=1}^{N_i} \text{Rej}_{\vec{u}_{ij}}(\vec{u}_{im}) \\ \vec{t}_{ji} &= \text{Rej}_{\vec{u}_{ji}}(\vec{v}_i) = \sum_{n=1}^{N_j} \text{Rej}_{\vec{u}_{ji}}(\vec{u}_{in}) \end{aligned} \quad (5)$$

In the same way, the inner product between \vec{t}_{ij} and \vec{t}_{ji} indicates the summation of improper angle information formed by e_{ij} :

$$\begin{aligned} \langle \vec{t}_{ij}, \vec{t}_{ji} \rangle &= \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \left\langle \text{Rej}_{\vec{u}_{ij}}(\vec{u}_{im}), \text{Rej}_{\vec{u}_{ji}}(\vec{u}_{in}) \right\rangle \\ &= \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \cos \psi_{mijn} \end{aligned} \quad (6)$$

Multiple works have shown the effectiveness of high-order geometric tensors for molecular modeling [10, 20, 25, 26]. However, the computational overheads of these approaches are generally expansive due to the CG-product, impeding their further application for large systems. In this work, we convert the vectors to high-order representation with *spherical harmonics* but discard CG-product with the inner product following the idea of RGC. We find that the extended high-order geometric tensors can still represent the above angular information in the form of Legendre polynomials

according to the *addition theorem*:

$$\begin{aligned} P_l(\cos \theta_{jik}) &= P_l(\vec{u}_{ij} \cdot \vec{u}_{ik}) \\ &\propto \sum_{m=-l}^l Y_{l,m}(\vec{u}_{ij}) Y_{l,m}^*(\vec{u}_{ik}) \end{aligned} \quad (7)$$

where the P_l is the Legendre polynomial of degree l , $Y_{l,m}$ denotes the spherical harmonics function and $Y_{l,m}^*$ denotes its complex conjugation. We sum the product of different order l to obtain the scalar angular representation, which is the same operation as inner product. It is worth noting that such an extension doesn't increase the model size and keeps the model architecture unchanged.

We also provide a proof about the rotational invariance of RGC strategy in the Section 4.2.

ViS-MP: Vector-Scalar interactive Message Passing In order to make full use of geometric information and enhance the interaction between scalars and vectors, we designed an effective vector-scalar interactive message passing mechanism with respect to the intersecting nodes and edges for angles and dihedrals, respectively. The key operations in ViS-MP are given as follows:

$$m_i^l = \sum_{j \in \mathcal{N}(i)} \phi_m^s(h_i^l, h_j^l, f_{ij}^l) \quad (8)$$

$$\vec{m}_i^l = \sum_{j \in \mathcal{N}(i)} \phi_m^v(m_{ij}^l, \vec{r}_{ij}, \vec{v}_j^l) \quad (9)$$

$$h_i^{l+1} = \phi_{un}^s(h_i^l, m_i^l, \langle \vec{v}_i^l, \vec{v}_i^l \rangle) \quad (10)$$

$$f_{ij}^{l+1} = \phi_{ue}^s(f_{ij}^l, \langle \text{Rej}_{\vec{r}_{ij}}(\vec{v}_i^l), \text{Rej}_{\vec{r}_{ji}}(\vec{v}_j^l) \rangle) \quad (11)$$

$$\vec{v}_i^{l+1} = \phi_{un}^v(\vec{v}_i^l, m_i^l, \vec{m}_i^l) \quad (12)$$

where h_i denotes the scalar embedding of node i , f_{ij} stands for the edge feature between node i and node j . \vec{v}_i represents the embedding of the direction unit mentioned in RGC. The superscript of variables indicates the index of the block that the variables belong to. We omit the improper angle here for brevity. A comprehensive version is depicted in Supplementary. ViS-MP extends the conventional message passing, aggregation, and update processes with vector-scalar interactions. Eq. 8 and Eq. 9 depict our message passing and aggregation processes. To be concrete, scalar messages m_{ij} incorporating scalar embedding h_j , h_i , and f_{ij} are passed and then aggregated to node

i through a message function ϕ_m^s (Eq. 8). Similar operations are applied for vector messages \vec{m}_i^l of node i that incorporates scalar message m_{ij} , vector \vec{r}_{ij} and vector embedding \vec{v}_j (Eq. 9). Eq. 10 and Eq. 11 demonstrate the update processes. h_i is updated by the aggregated scalar message output m_i while the inner product of \vec{v}_i is updated through an update function ϕ_{un}^s . Then f_{ij} is updated by the inner product of the rejection of the vector embedding \vec{v}_i and \vec{v}_j through an update function ϕ_{ue}^s . Finally, the vector embedding \vec{v}_i is updated by both scalar and vector messages through an update function ϕ_{un}^v . Notably, the vectors update function, i.e., ϕ^v require to be equivariant. The detailed message and update functions can be found in the Methods section. A proof about the equivariance of ViS-MP can be found in Supplementary Methods.

In summary, the geometric features are extracted by inner products in the RGC strategy and the scalar and vector embeddings are cyclically updating each other in ViS-MP so as to learn a comprehensive geometric representation from molecular structures.

2.2 Accurate quantum chemical property predictions

We evaluated ViSNet on several prevailing benchmark datasets including MD17 [23, 27, 28], revised MD17 [29], MD22 [30], QM9 [31], Molecule3D [32] and OGB-LSC PCQM4Mv2 [33] for energy, force, and other molecular property prediction. MD17 consists of the MD trajectories of 7 small organic molecules; the number of conformations in each molecule dataset ranges from 133,700 to 993,237. The dataset rMD17 is a reproduced version of MD17 with higher accuracy. MD22 is a newly proposed MD trajectories dataset that presents new challenges with respect to larger system sizes (42 to 370 atoms). Large molecules such as proteins, lipids, carbohydrates, nucleic acids, and supramolecules are included in MD22. QM9 consists of 12 kinds of quantum chemical properties of 133,385 small organic molecules with up to 9 heavy atoms. Molecule3D is a recently proposed dataset including 3,899,647 molecules collected from PubChemQC with their ground-state structures and corresponding properties calculated by DFT. We focus on the prediction of the HOMO-LUMO gap following ComENet [34]. OGB-LSC

Table 1 Mean absolute errors (MAE) of energy (kcal/mol) and force (kcal/mol/Å) for 7 small organic molecules on MD17 compared with state-of-the-art algorithms. The best one in each category is highlighted in bold. The last column indicates the percentage of improvements compared to the second-best approach, NequIP.

Molecule		SchNet	DimeNet	PaiNN	SpookyNet	ET	GemNet ¹	NequIP ²	SO3KRATES	ViSNet	Improvements
Aspirin	energy	0.37	0.204	0.167	0.151	0.123	-	0.131	0.139	0.116	11.45%
	forces	1.35	0.499	0.338	0.258	0.253	0.217	0.184	0.236	0.155	15.76%
Ethanol	energy	0.08	0.064	0.064	0.052	0.052	-	0.051	0.061	0.051	00.00%
	forces	0.39	0.230	0.224	0.094	0.109	0.085	0.071	0.096	0.060	15.49%
Malondialdehyde	energy	0.13	0.104	0.091	0.079	0.077	-	0.076	0.077	0.075	01.32%
	forces	0.66	0.383	0.319	0.167	0.169	0.155	0.129	0.147	0.100	22.48%
Naphthalene	energy	0.16	0.122	0.116	0.116	0.085	-	0.113	0.115	0.085	24.78%
	forces	0.58	0.215	0.077	0.089	0.061	0.051	0.039	0.074	0.039	00.00%
Salicylic Acid	energy	0.20	0.134	0.116	0.114	0.093	-	0.106	0.106	0.092	13.21%
	forces	0.85	0.374	0.195	0.180	0.129	0.125	0.090	0.145	0.084	06.67%
Toluene	energy	0.12	0.102	0.095	0.094	0.074	-	0.092	0.095	0.074	19.57%
	forces	0.57	0.216	0.094	0.087	0.067	0.060	0.046	0.073	0.039	15.22%
Uracil	energy	0.14	0.115	0.106	0.105	0.095	-	0.104	0.103	0.095	08.65%
	forces	0.56	0.301	0.139	0.119	0.095	0.097	0.076	0.111	0.062	18.42%

¹ The best results are reported among four variants of GemNet.

² NequIP only shows the results with $l = 3$.

PCQM4Mv2 is a quantum chemistry dataset originally curated under the PubChemQC including DFT-calculated HOMO-LUMO gap of 3,746,619 molecules. The 3D conformations are provided for 3,378,606 training molecules but not for the validation and test sets. The training details of ViSNet on each benchmark are described in the Methods section.

Energy and force for MD simulation. We compared ViSNet with the state-of-the-art algorithms, including DimeNet [14], PaiNN [16], SpookyNet [19], ET [17], GemNet [18], UNiTE [35], NequIP [10], SO3KRATES [36], Allegro [20], MACE [21] and so on. As shown in Table 1 (MD17) and Table 2 (rMD17) and Supplementary Table 2 (MD22), it is remarkable that ViSNet outperformed the compared algorithms for both small (MD17 and rMD17) and large molecules (MD22) with the lowest mean absolute errors (MAE) of predicted energy and forces. On the one hand, compared with PaiNN, ET and GemNet, ViSNet incorporated more geometric information and made full use of geometric information in ViS-MP, which contributes to the performance gains. On the other hand, compared with NequIP, Allegro, SO3KRATES, MACE et al, ViSNet testified the effect of introducing spherical harmonics in RGC module.

Quantum chemical properties. As shown in Table 3, ViSNet also achieved the superior performance for chemical property predictions on QM9. It outperformed the compared algorithms for 9 of 12 chemical properties and achieved comparable results on the remaining properties. Elaborated evaluations on Molecule3D confirmed the high prediction accuracy of ViSNet as shown in Table 4. ViSNet achieved 33.6% and 6.51% improvements than the second-best for random split and scaffold split, respectively. Furthermore, ViSNet exhibited good portability to other multimodality methods, e.g., Transformer-M [37] and outperformed other approaches on OGB-LSC PCQM4Mv2 (see Supplementary Fig. 1). ViSNet also achieved the top winners of PCQM4Mv2 track in the OGB-LCS@NeurIPS2022 competition when testing on unseen molecules [38] (<https://ogb.stanford.edu/neurips2022/results/>).

Computational Efficiency To evaluate the computational efficiency of our ViSNet, following [21], we compare the time latency of ViSNet with prevailing models in Fig. 3. The latency is defined as the time it takes to compute forces on a structure (i.e., the gradient calculation for a set of input coordinates through the whole deep neural network). As shown in Fig. 3, ViSNet (L=2) saved 42.8% time latency compared with MACE (L=2). Notably, despite the use of CG-product,

Table 2 Mean absolute errors (MAE) of energy (kcal/mol) and force (kcal/mol/Å) for 10 small organic molecules on rMD17 compared with state-of-the-art algorithms. The best one in each category is highlighted in bold.

Molecule		UNiTE ¹	ACE	GemNet ²	NequIP ²	BOTNet	Allegro	MACE	ViSNet ³
Aspirin	energy	0.055	0.141	-	0.0530	0.0530	0.0530	0.0507	0.0445
	forces	0.175	0.413	0.2191	0.1891	0.1960	0.1683	0.1522	0.1520
Azobenzene	energy	0.025	0.083	-	0.0161	0.0161	0.0277	0.0277	0.0156
	forces	0.097	0.251	-	0.0669	0.0761	0.0600	0.0692	0.0585
Benzene	energy	0.002	0.0009	-	0.0009	0.0007	0.0069	0.0092	0.0007
	forces	0.017	0.012	0.0115	0.0069	0.0069	0.0046	0.0069	0.0056
Ethanol	energy	0.014	0.028	-	0.0092	0.0092	0.0092	0.0092	0.0078
	forces	0.085	0.168	0.083	0.0646	0.0738	0.0484	0.0484	0.0522
Malonaldehyde	energy	0.025	0.039	-	0.0184	0.0184	0.0138	0.0184	0.0132
	forces	0.152	0.256	0.1522	0.1176	0.1338	0.0830	0.0945	0.0893
Naphthalene	energy	0.011	0.021	-	0.0046	0.0046	0.0046	0.0115	0.0057
	forces	0.060	0.118	0.0438	0.0300	0.0415	0.0208	0.0369	0.0291
Paracetamol	energy	0.044	0.092	-	0.0323	0.0300	0.0346	0.0300	0.0258
	forces	0.164	0.293	-	0.1361	0.1338	0.1130	0.1107	0.1029
Salicylic acid	energy	0.017	0.042	-	0.0161	0.0184	0.0208	0.0208	0.0161
	forces	0.088	0.214	0.1222	0.0922	0.0992	0.0669	0.0715	0.0795
Toluene	energy	0.010	0.025	-	0.0069	0.0069	0.0092	0.0115	0.0059
	forces	0.058	0.150	0.0507	0.0369	0.0438	0.0415	0.0346	0.0264
Uracil	energy	0.013	0.025	-	0.0092	0.0092	0.0138	0.0115	0.0069
	forces	0.088	0.152	0.0876	0.0715	0.0738	0.0415	0.0484	0.0495

¹ For a fair comparison, the “direct learning” results without any extra input are compared.

² The best results are reported among four variants of GemNet and four orders $l \in \{0, 1, 2, 3\}$ of NequIP.

³ ViSNet can achieve better results with longer convergence time.

Table 3 Mean absolute errors (MAE) of 12 kinds of molecular properties on QM9 compared with state-of-the-art algorithms. The best one in each category is highlighted in bold.

Target	Unit	SchNet	EGNN	DimeNet++	PaiNN	SphereNet	PaxNet	ET	ComENet	ViSNet
μ	mD	33	29	29.7	12	24.5	10.8	11	24.5	9.5
α	ma_0^3	235	71	43.5	45	44.9	44.7	59	45.2	41.1
ϵ_{HOMO}	meV	41	29	24.6	27.6	22.8	22.8	20.3	23.1	17.3
ϵ_{LUMO}	meV	34	25	19.5	20.4	18.9	19.2	17.5	19.8	14.8
$\Delta\epsilon$	meV	63	48	32.6	45.7	31.1	31	36.1	32.4	31.7
$\langle R^2 \rangle$	ma_0^2	73	106	331	66	268	93	33	259	29.8
$ZPVE$	meV	1.7	1.55	1.21	1.28	1.12	1.17	1.84	1.2	1.56
U_0	meV	14	11	6.32	5.85	6.26	5.9	6.15	6.59	4.23
U	meV	19	12	6.28	5.83	6.36	5.92	6.38	6.82	4.25
H	meV	14	12	6.53	5.98	6.33	6.04	6.16	6.86	4.52
G	meV	14	12	7.56	7.35	7.78	7.14	7.62	7.98	5.86
C_v	$\frac{\text{mcal}}{\text{mol K}}$	33	31	23	24	22	23.1	26	24	23

Table 4 Mean absolute errors (MAE) of HOMO-LUMO gap (eV) on Molecule3D test set for both random and scaffold splits compared with state-of-the-art algorithms.

Model	Random	Scaffold
GIN-Virtual	0.1036	0.2371
SchNet	0.0428	0.1511
DimeNet++	0.0306	0.1214
SphereNet	0.0301	0.1182
ComENet	0.0326	0.1273
ViSNet	0.0200	0.1105

Allegro had a significant speed improvement compared to NequIP and BOTNet. However, ViSNet still saved 6.1%, 4.1% and 61% time latency compared to Allegro with L=1, 2 and 3, respectively.

2.3 Efficient molecular dynamics simulations on MD17

Most of the recently proposed methods are quite accurate in predicting potential energy and atomic forces for the conformations in a given test set. Molecular dynamics simulation is one of the important applications of the predicted potential energy and atomic forces. To evaluate ViSNet as the potential for molecular dynamics simulation, we incorporated ViSNet that trained only with 950 samples on MD17 into the ASE simulation framework [39] to perform MD simulations for all 7 kinds of organic molecules. All simulations are run with a time step $\tau = 0.5$ fs under Berendsen thermostat with the other settings the same as those of the MD17 dataset. As shown in Fig. 4, we analyzed the interatomic distance distributions derived from both AIMD simulations with ViSNet as the potential and *ab initio* molecular dynamics simulations at DFT level for all 7 molecules, respectively. As shown in Fig. 4(a), the interatomic distance distribution $h(r)$ is defined as the ensemble average of atomic density at a radius r [28]. Fig. 4(b-h) illustrate the distributions derived from ViSNet are very close to those generated by DFT. We also compared the potential energy surfaces sampled by ViSNet and DFT for these molecules, respectively (Supplementary Fig. 2). The consistent potential energy surfaces suggest that ViSNet can well recover the kinetic properties and the conformational space from the simulation trajectories, indicating the usefulness of ViSNet for real molecular dynamics simulation. Furthermore, compared with the prohibitive

computational cost of DFT, ViSNet dramatically saves the computational time by 2-3 orders of magnitude (Supplementary Fig. 3 and Supplementary Table 3). These results demonstrate that with only a few of training samples, ViSNet can act as the potential to perform high-fidelity molecular dynamics simulations with much less computational cost.

2.4 Applications for real-world full-atom proteins

To examine the usefulness of ViSNet in real-world applications, we made evaluations on the 166-atom protein Chignolin. Based on a Chignolin dataset consisting of about 10,000 conformations that sampled by replica exchange MD and calculated at DFT level by Gaussian 16 in our another study, we split it as training, validation, and test sets by the ratio of 8:1:1. We trained ViSNet and compared it with molecular mechanics (MM). The DFT results were used as the ground truth. Fig. 5(a) shows the free energy landscape of *Chignolin* and depicted by d_{Y2-G6} (the distance between mainchain O on Y2 and mainchain N on G6) and d_{E4-T7} (the distance between mainchain O on E4 and mainchain N on T7). The concentrated energy basin on the left shows the folded state and the scattered energy basin on the right shows unfolded state. We picked six representative structures in the low potential energy regions with both folded and unfolded states and selected some intermediate states with high potential energy colored cyan or blue. We visualized the energy predictions for the six representative structures, and ViSNet produced a significantly better estimation of the potential energy than MM with empirical force fields did. Fig. 5(b) and (c) show the correlations between the predicted energies by ViSNet and MM, and the ground truth values calculated by DFT for all conformations in the test set. ViSNet achieved the lower MAE and the higher R^2 score. Similar results can be seen in the force correlations shown in the Supplementary Fig. 4. Furthermore, we performed MD simulations for Chignolin driven by ViSNet. 10 conformations were randomly selected as initial structures, and 10,000 simulation steps were run for each. As shown in the Fig. 5(d), the RMSF for 10 simulation trajectories are shown against simulation steps. In Fig. 5(e), we compared the

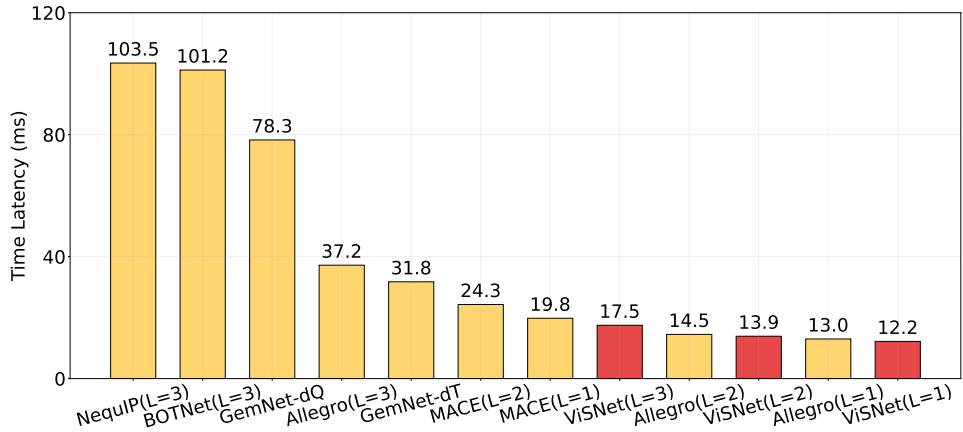


Fig. 3 Comparison of time latency with current methods following MACE [21]. Time latency is defined as the time the model takes to compute forces on a structure. Experiments are conducted on a Nvidia A100 GPU.

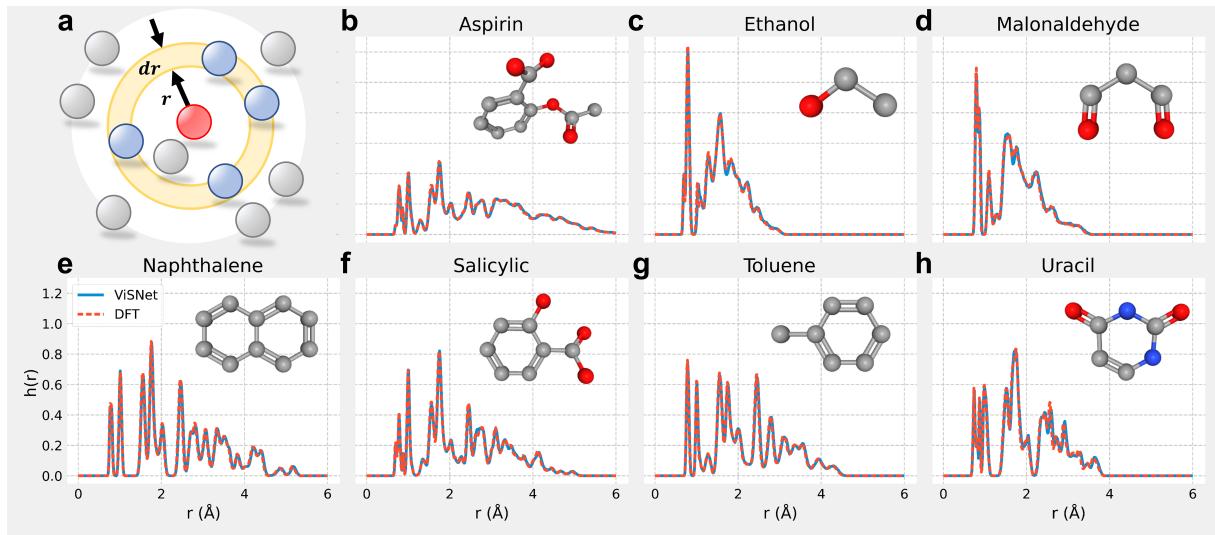


Fig. 4 The interatomic distance distributions of MD simulations driven by ViSNet and DFT. (a) An illustration about the atomic density at a radius r with the arbitrary atom as the center. The interatomic distance distribution is defined as the ensemble average of atomic density. (b) to (h) The interatomic distance distributions comparison between simulations by ViSNet and DFT for all seven organic molecules in MD17. The curve of ViSNet is shown using a solid blue line, while the dashed orange line is used for DFT curve. The structures of the corresponding molecules are shown in the upper right corner.

force differences between ViSNet and those calculated by Gaussian 16 at DFT level. The simulation trajectory driven by ViSNet exhibited small force difference to quantum mechanics, which implies that the accuracy and potential usefulness for real-world applications.

2.5 Interpretability of ViSNet on molecular structures

Prior works have shown the effectiveness of incorporating geometric features, such as angles. The primary method of geometry extraction utilized by ViSNet is the distinct inner product in its runtime geometry calculation. To this end, we illustrate a reasonable model interpretability of ViSNet by mapping the angle representations derived from inner product of direction units in

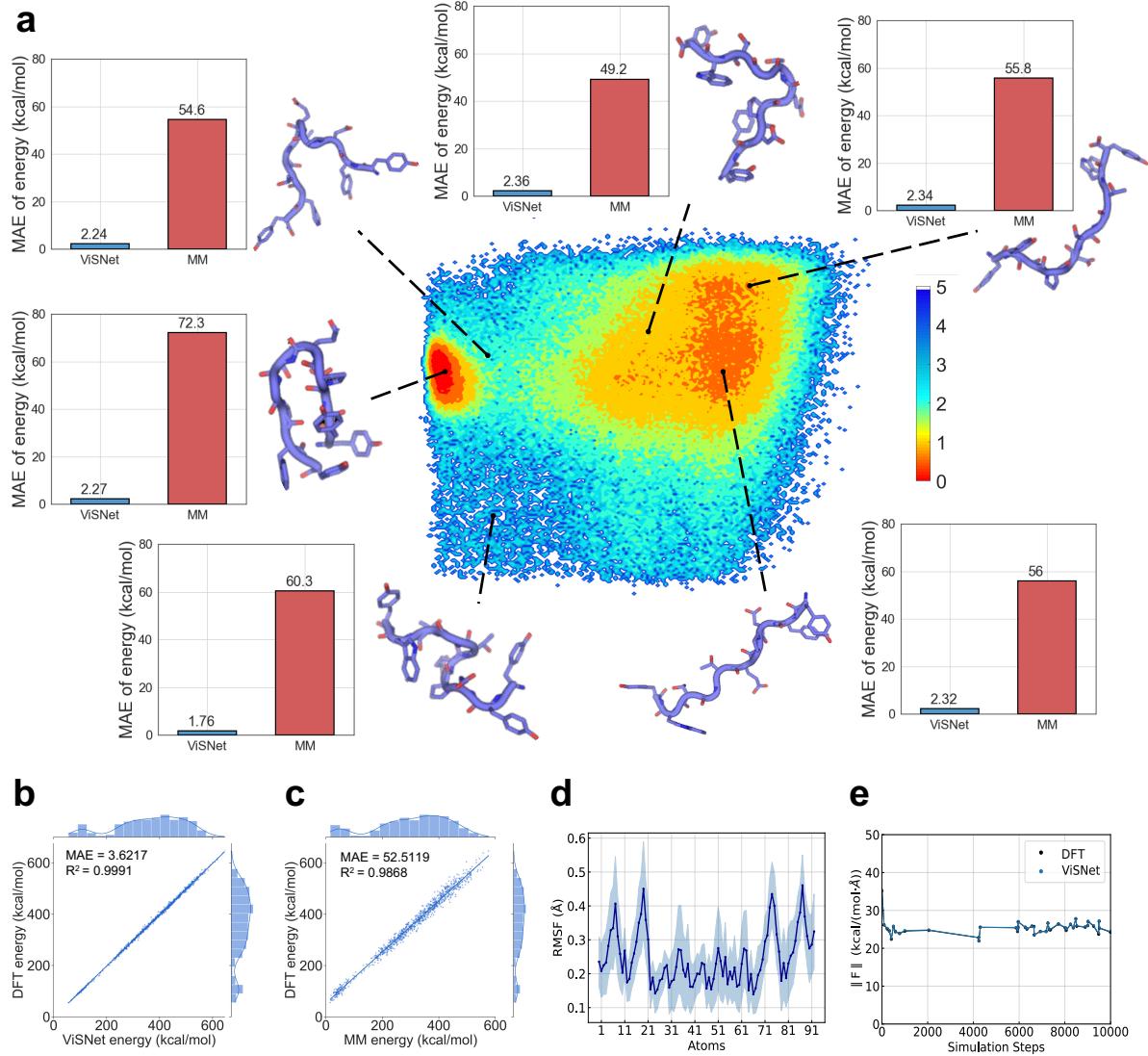


Fig. 5 Applications of ViSNet for Chignolin conformational space evaluation and MD simulations. (a) The energy landscape of *Chignolin* sampled by REMD. The x-axis of the landscape is the distance between mainchain O on Y2 and mainchain N on G6, while the y-axis is the distance between mainchain O on E4 and mainchain N on T7. 6 representative structures were then selected for visualization. Each structure is shown as cartoon and residues are depicted in sticks. The histograms show the mean absolute error (MAE) between the energy difference predicted/calculated by ViSNet or MM, and the ground truth calculated by DFT on the corresponding structure. (b) The energy correlations on the test dataset between the ground truth calculated by DFT and the predictions made by ViSNet. The corresponding distributions of energy predictions or calculations as well as the ground truth are shown. (c) The energy correlations on the test dataset between the ground truth calculated by DFT and the predictions made by molecular mechanics. (d) The average root mean square fluctuations (RMSF) of the Chignolin trajectories simulated by ViSNet were calculated from 10 different trajectories. The shaded areas indicate the standard deviation range. (e) The variation of the force norm during the ViSNet-driven simulation is shown in blue. Multiple frames were randomly selected from the simulation and the ground truth energies and forces were calculated using Gaussian, which are represented by black points.

the model to the atoms in the molecular structure. We aim to bridge the gap between geometric representation in ViSNet and molecular structures. We visualized the embeddings after the inner product of direction units $\langle \vec{v}_i, \vec{v}_i \rangle$ extracted from 50 aspirin samples on the validation set. The high-dimensional embeddings were reduced to 2-dimensional space using T-SNE [40] and then clustered using DBSCAN [41] without the prior of the number of clusters.

Fig. 6 exhibits the clustering results of nodes' embeddings after the inner product of their corresponding direction units. We further map the clustered nodes to the atoms of aspirin chemical structure. Interestingly, the embeddings for these nodes could be distinctly gathered into several clusters shown in different colors. For example, although carbon atom C_{11} and carbon atom C_{12} possess different positions and connect with different atoms, their inner product $\langle \vec{v}_i, \vec{v}_i \rangle$ are clustered into the same class for holding similar substructures ($\{C_{11} - O_2O_3C_6\}$ and $\{C_{12} - O_1O_4C_{13}\}$). To summarize, ViSNet can discriminate different molecular substructures in the embedding space.

2.6 Ablation study

To further explore where the performance gains of ViSNet come from, we conducted a comprehensive ablation study. Specifically, we excluded the runtime angle calculation (w/o A), runtime dihedral calculation (w/o D), and both of them (w/o A&D) in ViSNet, in order to evaluate the usefulness of each part. ViSNet-improper denotes the additional improper angles and ViSNet _{$l=1$} uses the 1st order spherical harmonics.

We designed some model variants with different message passing mechanisms based on ViS-MP for scalar and vector interaction. ViSNet-N directly aggregates the dihedral information to intersecting nodes, and ViSNet-T leverages another form of dihedral calculation. The details of these model variants are elaborated in Supplementary. The results of the ablation study are shown in Supplementary Table 5 and Supplementary Fig. 5. Based on the results, we can see that both kinds of directional geometric information are useful and the dihedral information contributes a little bit more to the final performance. The significant performance drop from

ViSNet-N and ViSNet-T further validate the effectiveness of ViS-MP mechanism. ViSNet-improper achieves similar performance to ViSNet for small molecules, but the contribution of improper angles is more obvious for large molecules (see Supplementary Table 2). Furthermore, ViSNet using higher order spherical harmonics achieves better performance.

3 Discussion and conclusion

We propose ViSNet, a novel geometric deep learning potential for molecular dynamics simulation. The group representation theory based methods and the directional information based methods are two mainstream classes of geometric deep learning potentials to enforce SE(3) equivariance [18]. ViSNet takes the advantages from both sides in designing RGC strategy and ViS-MP mechanism. On the one hand, the RGC strategy explicitly extracts and exploits the directional geometric information with computationally lightweight operations, making the model training and inference fast. On the other hand, ViS-MP employs a series of effective and efficient vector-scalar interactive operations, leading to the full use of the geometric information. Furthermore, according to the many-body expansion theory [42–44], the potential energy of the whole system equals to the potential of each single atom plus the energy corrections from two-bodies to many-bodies. Most of the previous studies model the truncated energy correction terms hierarchically with k -hop information via stacking k message passing blocks. Different from these approaches, ViSNet encodes the triplet and quadruplet interactions in a single block, which empowers the model to have much more powerful representation ability. In addition, considering that angle and dihedral are important potential terms in empirical force fields, the interpretability of the operations in the RGC strategy provides some insights in constructing hybrid force fields by combining empirical terms with deep learning.

Besides predicting energy, force, and chemical properties with high accuracy, performing molecular dynamics simulations with *ab initio* accuracy at the cost of empirical force field is a grand challenge. ViSNet proves its usefulness in real-world *ab initio* molecular dynamics simulations with less computational costs and the ability of scaling to

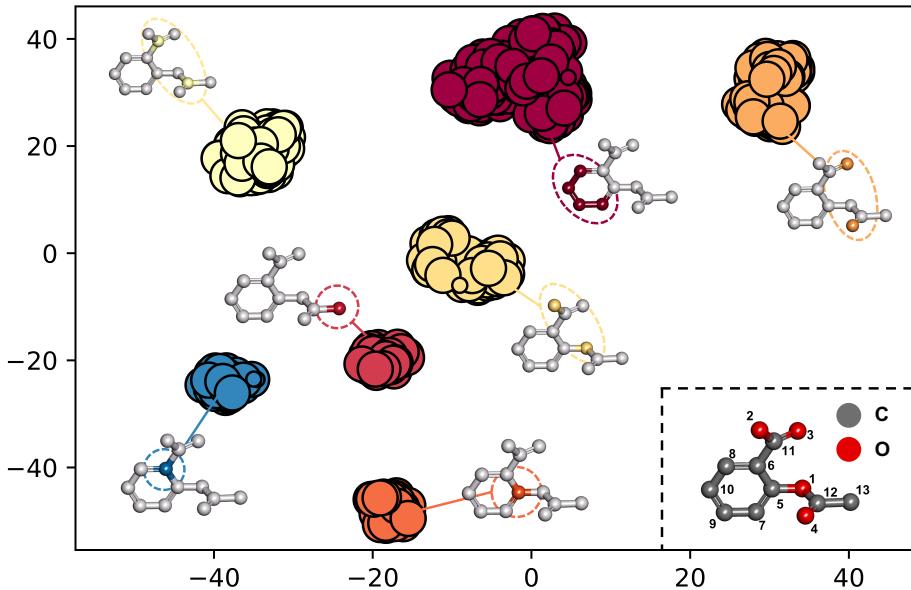


Fig. 6 Visualization and model interpretability of ViSNet. Clusters of nodes' embeddings after the inner product of the direction units $\langle \vec{v}_i, \vec{v}_i \rangle$. The $\langle \vec{v}_i, \vec{v}_i \rangle$ represents angle representations with the intersecting node i as the vertex. The atoms in the chemical structure of aspirin corresponding to each cluster are colored with the same color of the cluster, while the remaining atoms are colored light gray. A chemical structure of Aspirin and the indices of atoms are illustrated in the bottom right region. Carbon and oxygen atoms are colored dark grey and red, respectively. The hydrogen atoms are omitted in both the clustering results and the chemical structure of aspirin for simplification.

large molecules such as proteins. Extending ViSNet to support larger and more complex molecular systems will be our future research direction.

4 Methods

4.1 Equivariance

In the context of machine learning for atomic systems, *Equivariance* is a pervasive concept. Specifically, the atomic vectors such as dipoles or forces must rotate in a manner consistent with the conformation coordinates. In molecular dynamics, such equivariance can be ensured by computing gradients based on a predicted conservative scalar energy. Formally, a function $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y}$ is equivariant if it satisfies:

$$\mathcal{F}(\rho_{\mathcal{X}}(g) \circ x) = \rho_{\mathcal{Y}}(g) \circ \mathcal{F}(x), \quad (13)$$

where $\rho_{\mathcal{X}}(g)$ and $\rho_{\mathcal{Y}}(g)$ are group representations in input and output spaces. The integration of equivariance into model parameterization has been shown to be effective, as seen in the implementation of *shift-equivariance* in CNNs, which is critical for enhancing the generalization capacity.

4.2 Proofs of the rotational invariance of RGC

Assume that the molecule rotates in 3D space, i.e.,

$$\vec{r}'_{ij} = R\vec{r}_{ij} \quad (14)$$

where, $R \in SO(3)$ is an arbitrary rotation matrix that satisfies:

$$\det |R| = 1, R^T R = I \quad (15)$$

The angular information after rotation is calculated as follows:

$$\vec{u}'_{ij} = \frac{\vec{r}'_{ij}}{\|\vec{r}'_{ij}\|} = \frac{R\vec{r}_{ij}}{\det |R| \cdot \|\vec{r}_{ij}\|} = R\vec{u}_{ij} \quad (16)$$

$$\vec{v}'_i = \sum_{j=1}^{N_i} \vec{u}'_{ij} = R \sum_{j=1}^{N_i} \vec{u}_{ij} = R\vec{v}_i \quad (17)$$

$$\begin{aligned} \|\vec{v}'_i\|^2 &= \langle \vec{v}'_i, \vec{v}'_i \rangle = (\vec{v}'_i)^T \vec{v}'_i \\ &= \vec{v}_i^T R^T R \vec{v}_i = \langle \vec{v}_i, \vec{v}_i \rangle = \|\vec{v}_i\|^2 \end{aligned} \quad (18)$$

As shown in Eq. 18, the angle information does not change after rotation. The dihedral angular and

improper information is also rotationally invariant since:

$$\vec{w}'_{ij} = \vec{v}'_i - \langle \vec{v}'_i, \vec{u}'_{ij} \rangle \vec{u}'_{ij} = R\vec{v}_i - \langle R\vec{v}_i, R\vec{u}_{ij} \rangle R\vec{u}_{ij} \quad (19)$$

As Eq. 18 proved, the inner product has rotational invariance. Then, Eq. 19 can be further simplified as:

$$\vec{w}'_{ij} = R(\vec{v}_i - \langle \vec{v}_i, \vec{u}_{ij} \rangle \vec{u}_{ij}) = R\vec{w}_{ij} \quad (20)$$

The dihedral or improper angular information after rotation is calculated as:

$$\langle \vec{w}'_{ij}, \vec{w}'_{ji} \rangle = \langle R\vec{w}_{ij}, R\vec{w}_{ji} \rangle = \langle \vec{w}_{ij}, \vec{w}_{ji} \rangle \quad (21)$$

As a result, Eq. 18 and Eq. 21 have proved the rotational invariance of our proposed runtime geometry calculation (RGC).

We also provide a proof of the equivariance of our ViS-MP in Supplementary Methods.

4.3 Detailed operations and modules in ViSNet

ViSNet predicts the molecular properties (e.g., energy \hat{E} , forces $\vec{F} \in \mathbb{R}^{N \times 3}$, dipole moment μ) from the current states of atoms, including the atomic positions $X \in \mathbb{R}^{N \times 3}$ and atomic numbers $Z \in \mathbb{N}^N$. The architecture of the proposed ViSNet is shown in Fig. 1. The overall design of ViSNet follows the vector-scalar interactive message passing as illustrated from Eq. 8 - Eq. 11. First, an embedding block encodes the atom numbers and edge distances into the embedding space. Then, a series of ViSNet blocks update the node-wise scalar and vector representations based on their interactions. A residual connection is placed between two ViSNet blocks. Finally, stacked corresponding gated equivariant blocks proposed by [16] are attached to the output block for specific molecular property prediction.

The Embedding block ViSNet expands the direct node and edge embedding with their neighbors. It first embeds atomic chemical symbol z_i , and calculates the edge representation whose distances within the cutoff through radial basis functions (RBF). Then the initial embedding of the atom i , its 1-hop neighbors j and the directly connected edge e_{ij} within cutoff are fused together

as the initial node embedding h_i^0 and edge embedding f_{ij}^0 . In summary, the embedding block is given by:

$$h_i^0, f_{ij}^0 = \text{Embedding Block}(z_i, z_j, e_{ij}), \quad j \in \mathcal{N}(i) \quad (22)$$

$\mathcal{N}(i)$ denotes the set of 1-hop neighboring nodes of node i , and j is one of its neighbors. The embedding process is elaborated in Supplementary. The initial vector embedding \vec{v}_i is set to $\vec{0}$. The vector embeddings \vec{v} are projected into the embedding space by following [16]; $\vec{v} \in \mathbb{R}^{N \times 3 \times F}$ and F is the size of hidden dimension. The advantage of such projection is to assign a unique high-dimensional representation for each embedding to discriminate from each other. Further discussions on its effectiveness and interpretability are given in the Results section.

The Scalar2Vec module In the Scalar2Vec module, the vector embedding \vec{v} is updated by both the scalar messages derived from node and edge scalar embeddings (Eq. 8) and the vector messages with inherent geometric information (Eq. 9). The message of each atom is calculated through an Edge-Fusion Graph Attention module, which fuses the node and edge embeddings and computes the attention scores. The fusion of the node and edge embeddings could be the concatenation operation, Hadamard product, or adding a learnable bias [45]. We leverage the Hadamard product and the *vanilla* multi-head attention mechanism borrowed from Transformer [46] for edge-node fusion.

Following [17], we pass the fused representations through a nonlinear activation function as shown in Eq. 23. The value (V) in the attention mechanism is also fused by edge features before being multiplied by attention scores weighted by a cosine cutoff as shown in Eq. 24,

$$\alpha_{ij}^l = \sigma \left((W_Q^l h_i^l) \left(W_K^l h_j^l \odot \text{Dense}_K^l(f_{ij}^l) \right)^T \right) \quad (23)$$

$$m_{ij}^l = \alpha_{ij}^l \cdot \phi(\|\vec{r}_{ij}\|) \cdot \left(W_V^l h_j^l \odot \text{Dense}_V^l(f_{ij}^l) \right) \quad (24)$$

where $l \in \{0, 1, 2, \dots, L\}$ is the index of block, σ denotes the activation function (SiLU in this paper), W is the learnable weight matrix, \odot represents the Hadamard product, $\phi(\cdot)$ denotes the

cosine cutoff and Dense(\cdot) refers to one learnable weight matrix with activation function. For brevity, we omit the learnable bias for linear transformation on scalar embedding in equations, and there is no bias for vector embedding to ensure the equivariance.

Then, the computed m_{ij}^l is used to produce the geometric messages \vec{m}_{ij}^l for vectors:

$$\vec{m}_{ij}^l = \left(\text{Dense}_u^l(m_{ij}^l) \odot \vec{u}_{ij} \right) + \left(\text{Dense}_v^l(m_{ij}^l) \odot \vec{v}_j^l \right) \quad (25)$$

And the vector embedding \vec{v}^l is updated by:

$$m_i^l = \sum_{j \in \mathcal{N}(i)} m_{ij}^l, \quad \vec{m}_i^l = \sum_{j \in \mathcal{N}(i)} \vec{m}_{ij}^l \quad (26)$$

$$\Delta \vec{v}_i^{l+1} = \vec{m}_i^l + W_{\text{vm}}^l m_i^l \odot W_{\text{v}}^l \vec{v}_i^l \quad (27)$$

The Vec2Scalar module In the Vec2Scalar module, the node embedding h_i^l and edge embedding f_{ij}^l are updated by the geometric information extracted by the RGC strategy, i.e., angles (Eq. 10) and dihedrals (Eq. 11), respectively. The residual node embedding Δh_i^{l+1} , is calculated by a Hadamard product between the runtime angle information and the aggregated scalar messages with a gated residual connection:

$$\Delta h_i^{l+1} = \langle W_t^l \vec{v}_i^l, W_s^l \vec{v}_i^l \rangle \odot W_{\text{Angle}}^l m_i^l + W_{\text{res}}^l m_i^l \quad (28)$$

To compute the residual edge embedding Δf_{ij}^{l+1} , we perform the Hadamard product of the runtime dihedral information with the transformed edge embedding:

$$\Delta f_{ij}^{l+1} = \left\langle \text{Rej}_{\vec{r}_{ij}}(W_{Rt}^l \vec{v}_i^l), \text{Rej}_{\vec{r}_{ji}}(W_{Rs}^l \vec{v}_j^l) \right\rangle \odot \text{Dense}_{\text{Dihedral}}^l(f_{ij}^l) \quad (29)$$

After the residual hidden representations are calculated, we add them to the original input of block l and feed them to the next block.

A comprehensive version which includes improper angles is depicted in Supplementary Methods.

The output block Following PaiNN [16], we update the scalar embedding and vector embedding of nodes with multiple gated equivariant blocks:

$$t_i^l = \text{Dense}_{o_2}^l(\|W_{o_1}^l \vec{v}_i^l\|, h_i^l) \quad (30)$$

$$h_i^{l+1} = W_{o_3}^l t_i^l \quad (31)$$

$$\vec{v}_i^{l+1} = W_{o_4}^l \vec{v}_i^l \odot W_{o_5}^l t_i^l \quad (32)$$

where $[\cdot, \cdot]$ is the tensor concatenation operation. The final scalar embedding $h_i^L \in \mathbb{R}^{N \times 1}$ and vector embedding $\vec{v}_i^L \in \mathbb{R}^{N \times 3 \times 1}$ are used to predict various molecular properties.

On QM9, the molecular dipole is calculated as follows:

$$\mu = \left\| \sum_{i=1}^N \vec{v}_i^L + h_i^L (\vec{r}_i - \vec{r}_c) \right\| \quad (33)$$

where \vec{r}_c denotes the center of mass. Similarly, for the prediction of electronic spatial extent $\langle R^2 \rangle$, we use the following equation:

$$\langle R^2 \rangle = \sum_{i=1}^N h_i^L \|\vec{r}_i - \vec{r}_c\|^2 \quad (34)$$

For the remaining 10 properties y , we simply aggregate the final scalar embedding of nodes as follows:

$$y = \sum_{i=1}^N h_i^L \quad (35)$$

For models trained on the molecular dynamics datasets including MD17, revised MD17, and *Chignolin*, the total potential energy is obtained as the sum of the final scalar embedding of the nodes. As an energy-conserving potential, the forces are then calculated using the negative gradients of the predicted total potential energy with respect to the atomic coordinates:

$$E = \sum_{i=1}^N h_i^L \quad (36)$$

$$\vec{F}_i = -\nabla_i E \quad (37)$$

4.4 Dataset splitting schemes

For the QM9 dataset, we randomly split it into 110,000 samples as the train set, 10,000 samples as the validation set, and the rest as the test set by following the previous studies [16, 17]. For the Molecule3D and OGB-LSC PCQM4Mv2 dataset, the splitting has been provided in their paper [32, 33].

To evaluate the effectiveness of ViSNet to simulation data, ViSNet was trained on MD17 and

rMD17 with a limited data setting, which consists of only 950 uniformly sampled conformations for model training and 50 conformations for validation for each molecule. For MD22 dataset, we uses the same number of molecules as [30] for training and validation, and the rest as the test set.

Furthermore, the whole *Chignolin* dataset was randomly split into 80%, 10%, and 10% as the training, validation, and test datasets. Six representative conformations were picked from the test set for illustration.

4.5 Experimental settings

For the QM9 dataset, we adopted a batch size of 32 and a learning rate of 1e-4 for all the properties. For the Molecule3D dataset, we adopted a larger batch size of 512 and a learning rate of 2e-4. For the OGB-LSC PCQM4Mv2 dataset, we trained our model in a mixed 2D/3D mode with a batch size of 256 and a learning rate of 2e-4. The mean squared error (MSE) loss was used for model training. For the molecular dynamic dataset including MD17, rMD17, MD22 and *Chignolin*, we leveraged a combined MSE loss for energy and force prediction. The weight of energy loss was set to 0.05. The weight of forces loss was set to 0.95. The batch size was chosen from 2, 4, 8 due to the GPU memory and the learning rate was chosen from 1e-4 to 4e-4 for different molecules. The cutoff was set to 5 for small molecules in QM9, MD17, rMD17 and Molecule3D, and changed to 4 for *Chignolin* in order to reduce the number of edges in the molecular graphs. For MD22 dataset, the cutoff of relatively small molecules was set to 5, that of bigger molecules was set to 4. Cutoff was not used in OGB-LSC PCQM4Mv2 dataset. We used the learning rate decay if the validation loss stopped decreasing. The patience was set to 5 epochs for Molecule3D, 15 epochs for QM9, and 30 epochs for MD17, rMD17, MD22 and *Chignolin*. The learning rate decay factor was set to 0.8 for these models. We also adopted the early stopping strategy to prevent over-fitting. The ViSNet model trained on the molecular dynamic datasets and Molecule3D had 9 hidden layers and the embedding dimension was set to 256. We used a larger model for QM9 dataset, i.e., the embedding dimension changed to 512. For OGB-LSC PCQM4Mv2 dataset, we use the 12-layer and 768-dimension Transformer-M [37] as backbone. More

details about the hyperparameters of ViSNet can be found in Supplementary Table 5. Experiments were conducted on NVIDIA 32G-V100 GPUs.

Author contributions

T. W. led, conceived and designed the study. T. W. is the lead contact. Y. W., S. L., X. H. and M. L. conducted the work when they were visiting Microsoft Research. S. L., Y. W. and T. W. carried out algorithm design. Y. W., S. L. and X. H. carried out experiments, evaluations, analysis and visualization. Y. W. and S. L. wrote the original manuscript. T. W., X. H., M. L., Z. W. and B. S revised the manuscript. N. Z. and T. L. contributed to writing. All authors reviewed the final manuscript.

Data availability

The MD17, rMD17, MD22, QM9, Molecule3D and OGB-LSC PCQM4Mv2 dataset are available at their official website. The Chignolin dataset used in this study will be publicly available once the manuscript is published online.

Code availability

The code used to produce our results will be publicly available once the manuscript is published online.

References

- [1] Chow, E., Klepeis, J., Rendleman, C., Dror, R. & Shaw, D. 9.6 new technologies for molecular dynamics simulations. *Edward H. Egelman, editor. Comprehensive Biophysics*. Amsterdam: Elsevier 86–104 (2012).
- [2] Singh, S. & Singh, V. K. Molecular dynamics simulation: methods and application. In *Frontiers in protein structure, function, and dynamics*, 213–238 (Springer, 2020).
- [3] Lu, S. *et al.* Activation pathway of a g protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nature Communications* **12**, 1–15 (2021).

- [4] Li, Y. *et al.* Exploring the regulatory function of the n-terminal domain of sars-cov-2 spike protein through molecular dynamics simulation. *Advanced theory and simulations* **4**, 2100152 (2021).
- [5] Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **140**, A1133 (1965).
- [6] Marx, D. & Hutter, J. *Ab initio molecular dynamics: basic theory and advanced methods* (Cambridge University Press, 2009).
- [7] Christensen, A. S., Bratholm, L. A., Faber, F. A. & Anatole von Lilienfeld, O. Fchl revisited: Faster and more accurate quantum machine learning. *The Journal of chemical physics* **152**, 044107 (2020).
- [8] Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical review letters* **104**, 136403 (2010).
- [9] Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *Journal of Physics: Condensed Matter* **26**, 183001 (2014).
- [10] Batzner, S. *et al.* E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications* **13**, 1–11 (2022).
- [11] Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E. & Welling, M. Geometric and physical quantities improve e (3) equivariant message passing. *International Conference on Learning Representations* (2022).
- [12] Hutchinson, M. J. *et al.* Lietransformer: Equivariant self-attention for lie groups. In *International Conference on Machine Learning*, 4533–4543 (PMLR, 2021).
- [13] Fuchs, F., Worrall, D., Fischer, V. & Welling, M. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems* **33**, 1970–1981 (2020).
- [14] Gasteiger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations* (2019).
- [15] Gasteiger, J., Giri, S., Margraf, J. T. & Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *Advances in Neural Information Processing Systems* (2020).
- [16] Schütt, K., Unke, O. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, 9377–9388 (PMLR, 2021).
- [17] Thölke, P. & De Fabritiis, G. Torchmd-net: Equivariant transformers for neural network based molecular potentials. *The International Conference on Learning Representations* (2022).
- [18] Gasteiger, J., Becker, F. & Günnemann, S. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems* **34**, 6790–6802 (2021).
- [19] Unke, O. T. *et al.* Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature communications* **12**, 1–14 (2021).
- [20] Musaelian, A. *et al.* Learning local equivariant representations for large-scale atomistic dynamics. *arXiv preprint arXiv:2204.05249* (2022).
- [21] Batatia, I., Kovács, D. P., Simm, G. N., Ortner, C. & Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *arXiv preprint arXiv:2206.07697* (2022).
- [22] Han, J., Rong, Y., Xu, T. & Huang, W. Geometrically equivariant graph neural networks: A survey. *arXiv preprint arXiv:2202.07230* (2022).
- [23] Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular

- dynamics simulations with machine-learned force fields. *Nature communications* **9**, 1–10 (2018).
- [24] Perwass, C., Edelsbrunner, H., Kobbelt, L. & Polthier, K. *Geometric algebra with applications in engineering*, vol. 4 (Springer, 2009).
- [25] Zitnick, C. L. *et al.* Spherical channels for modeling atomic interactions. *arXiv preprint arXiv:2206.14331* (2022).
- [26] Liao, Y.-L. & Smidt, T. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990* (2022).
- [27] Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nature communications* **8**, 1–8 (2017).
- [28] Chmiela, S. *et al.* Machine learning of accurate energy-conserving molecular force fields. *Science advances* **3**, e1603015 (2017).
- [29] Christensen, A. S. & Von Lilienfeld, O. A. On the role of gradients for machine learning of molecular energies and forces. *Machine Learning: Science and Technology* **1**, 045018 (2020).
- [30] Chmiela, S. *et al.* Accurate global machine learning force fields for molecules with hundreds of atoms. *arXiv preprint arXiv:2209.14865* (2022).
- [31] Ramakrishnan, R., Dral, P. O., Rupp, M. & Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data* **1**, 1–7 (2014).
- [32] Xu, Z. *et al.* Molecule3d: A benchmark for predicting 3d geometries from molecular graphs. *arXiv preprint arXiv:2110.01717* (2021).
- [33] Hu, W. *et al.* OGB-LSC: A large-scale challenge for machine learning on graphs. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)* (2021). URL <https://openreview.net/forum?id=qkcLxoC52kL>.
- [34] Wang, L., Liu, Y., Lin, Y., Liu, H. & Ji, S. Comenet: Towards complete and efficient message passing for 3d molecular graphs. *Advances in Neural Information Processing Systems* (2022).
- [35] Qiao, Z. *et al.* Informing geometric deep learning with electronic interactions to accelerate quantum chemistry. *Proceedings of the National Academy of Sciences* **119**, e2205221119 (2022).
- [36] Frank, T., Unke, O. T. & Muller, K. R. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. In *Advances in Neural Information Processing Systems* (2022).
- [37] Luo, S. *et al.* One transformer can understand both 2d & 3d molecular data. *arXiv preprint arXiv:2210.01765* (2022).
- [38] Wang, Y. *et al.* An ensemble of visnet, transformer-m, and pretraining models for molecular property prediction in ogb large-scale challenge@ neurips 2022. *arXiv preprint arXiv:2211.12791* (2022).
- [39] Larsen, A. H. *et al.* The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
- [40] Van der Maaten, L. & Hinton, G. Visualizing data using t-sne. *Journal of machine learning research* **9** (2008).
- [41] Ester, M., Kriegel, H.-P., Sander, J., Xu, X. *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, vol. 96, 226–231 (1996).
- [42] Nesbet, R. Atomic Bethe-Goldstone equations. III. correlation energies of ground states of Be, B, C, N, O, F, and Ne. *Physical Review* **175**, 2 (1968).

- [43] Hankins, D., Moskowitz, J. & Stillinger, F. Water molecule interactions. *The Journal of Chemical Physics* **53**, 4544–4554 (1970).
- [44] Gordon, M. S., Fedorov, D. G., Pruitt, S. R. & Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chemical reviews* **112**, 632–672 (2012).
- [45] Ying, C. *et al.* Do transformers really perform badly for graph representation? *Advances in Neural Information Processing Systems* **34** (2021).
- [46] Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).