Microsoft

NEURAL INFORMATION PROCESSING SYSTEMS

QuinNet: Efficiently Incorporating Quintuple Interactions into Geometric Deep Learning Force Fields

Zun Wang, Guoqing Liu, Yichi Zhou, Tong Wang, Bin Shao
Microsoft Research AI4Science

Code at https://github.com/Zun-Wang/QuinNet

Outline

Introduction

Method

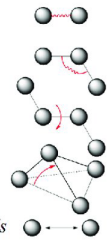
Results

Conclusions

Preliminary

Empirical force field

$$
\begin{aligned}
U(R) = & \sum_{bonds} k_r (r - r_{eq})^2 && \text{bond} \\
& + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 && \text{angle} \\
& + \sum_{dihedrals} k_\phi (1 + \cos[n\phi - \gamma]) && \text{dihedral} \\
& + \sum_{impropers} k_\omega (\omega - \omega_{eq})^2 && \text{improper} \\
& + \sum_{i<j}^{atoms} \varepsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^{6} \right] && \text{van der Waals} \\
& + \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} && \text{electrostatic}
\end{aligned}
$$



Two-body interactions:

1924 Lennard-Jones potential
1929 Morse potential
1938 Buckingham potential
...

2018 SchNet
...

Three-body interactions:

1988 Tersoff potential
1992 Modified embedded atom method (MEAM)
2005 Angular dependent potential (ADP)
...

2020 DimeNet
2020 DimeNet++
2021 PaiNN
...

Four-body interactions:

2000 Adaptive Intermolecular Reactive Empirical Bond Order (AIREBO) Potential
...

2021 GemNet
2022 ViSNet
...

Five-body interactions:

2010 AMOEBA and HIPPO
...

???

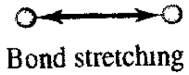
Group representation:

2022 NequIP
2022 Allegro
2022 MACE
...

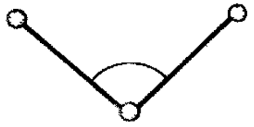
Machine Learning Force Field

Schütt K, et al. NeurIPS, 2017, 30.
Gasteiger J, et al. ICLR. 2019.
Gasteiger J, et al. arXiv:2011.14115, 2020.
Schütt K, et al. ICML, 2021: 9377-9388.
Gasteiger J, et al. NeurIPS, 2021, 34: 6790-6802.
Wang Y, et al. arXiv:2210.16518, 2022.
Batzner S, et al. Nature communications, 2022, 13(1): 2453.
Musaelian A, et al. Nature Communications, 2023, 14(1): 579.
Batatia I, et al. NeurIPS, 2022, 35: 11423-11436.

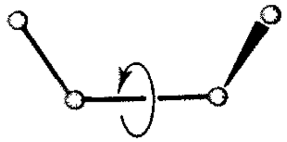
Empirical force fields



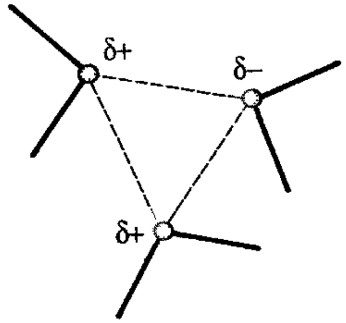
Bond stretching



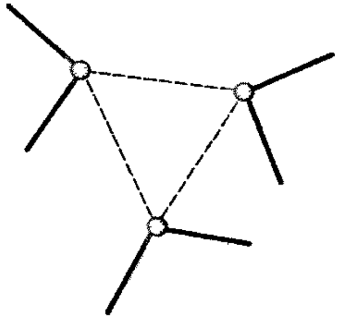
Angle bending



Bond rotation
(torsion)

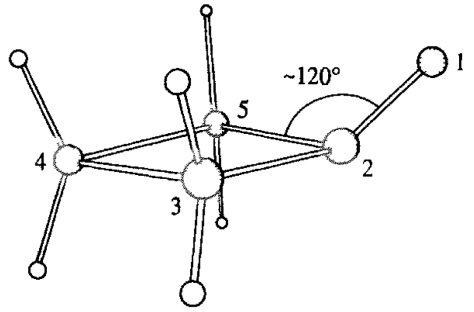


Non-bonded interactions
(electrostatic)

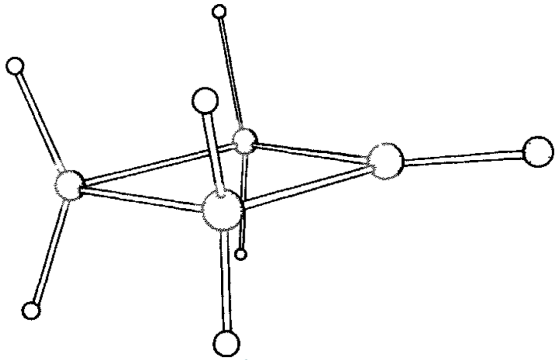


Non-bonded interactions
(van der Waals)

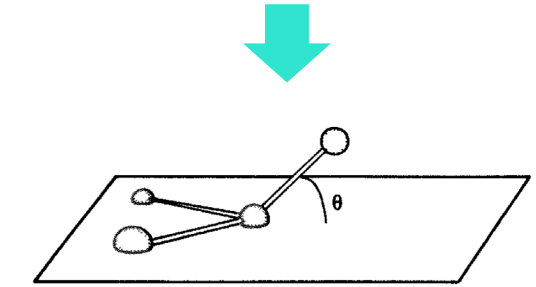
Oxygen in cyclobutane



Simulation



Experiment



Improper torsion

Challenges

- Selection of appropriate physical quantities
- Calculating such a physical quantity increases the computational complexity as the order of interactions increases

Methods

☑Selection of appropriate physical quantities ➢ Topology of many-body interactions

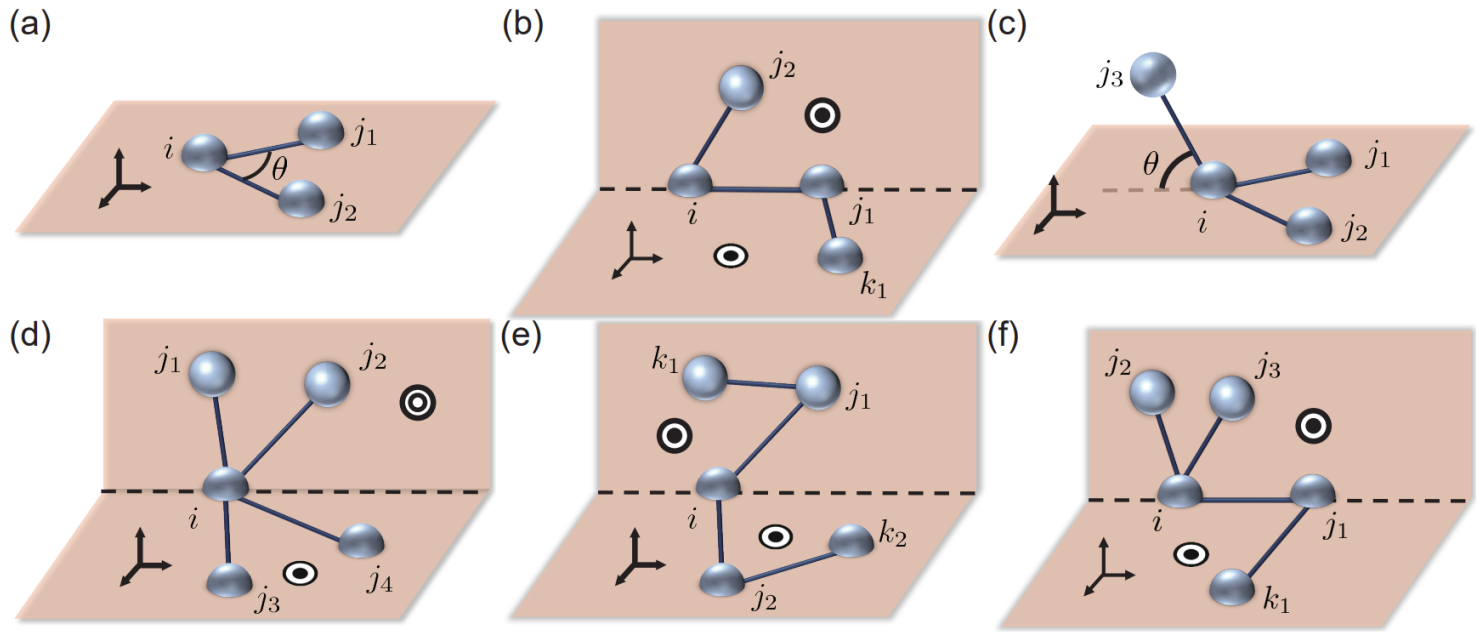


Figure 2: A schematic diagram that describes the topology of (a) three-body interaction (angles), (b) four-body interaction (torsions), (c) four-body interaction (improper torsions), and (d-f) five-body interactions. The marker ⊙ on a plane represents the normal vector of this plane.

Methods

☑ Calculating physical quantities efficiently ➢ Architecture of QuinNet

Three-body interactions

$$\left\| \sum_{j \in \mathcal{N}_i} \hat{r}_{ij} \right\|^2 = \sum_{j,k \in \mathcal{N}_i} \langle \hat{r}_{ij}, \hat{r}_{ij} \rangle = \sum_{j,k \in \mathcal{N}_i} \cos \alpha_{jik},$$

Four-body interactions (torsion)

$$\left(\sum_{k_1 \in \mathcal{N}_i} \hat{r}_{ik_1} \times \hat{r}_{ij} \right) \cdot \left(\sum_{k_2 \in \mathcal{N}_j} \hat{r}_{jk_2} \times (-\hat{r}_{ij}) \right) = \sum_{\substack{k_1 \in \mathcal{N}_i, \\ k_2 \in \mathcal{N}_j}} \langle \vec{n}_{ijk_1}, \vec{n}_{ijk2} \rangle$$

Four-body interactions (improper torsion)

$$\sum_{j \in \mathcal{N}_i} \hat{r}_{ij} \cdot \left[\left(\sum_{j \in \mathcal{N}_i} \alpha_j \hat{r}_{ij} \right) \times \left(\sum_{j \in \mathcal{N}_i} \beta_j \hat{r}_{ij} \right) \right] = \sum_{j_1, j_2, j_3 \in \mathcal{N}_i} \gamma_{ij_1j_2} \langle \vec{r}_{ij_3}, \vec{n}_{ij_1j_2} \rangle$$

Five-body interactions@I

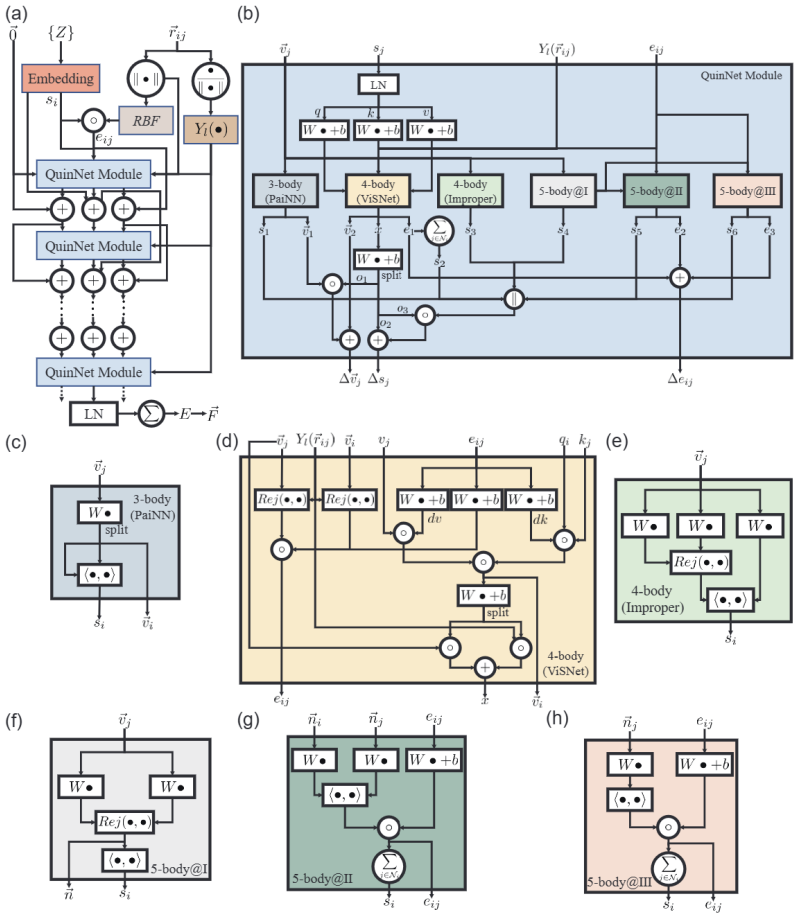
$$\left\| \left(\sum_{j \in \mathcal{N}_i} \alpha_j \hat{r}_{ij} \right) \times \left(\sum_{j \in \mathcal{N}_i} \beta_j \hat{r}_{ij} \right) \right\|^2 = \sum_{j_1, j_2, j_3, j_4 \in \mathcal{N}_i} \gamma_{ij_1j_2j_3j_4} \langle \vec{n}_{ij_1j_2}, \vec{n}_{ij_3j_4} \rangle$$

Five-body interactions@II

$$\left\| \left(\sum_{k \in \mathcal{N}_{j_1}} \alpha_k \hat{r}_{kj_1} \times \sum_{k \in \mathcal{N}_{j_1}} \beta_k \hat{r}_{kj_1} \right) \Big|_{j_1 \in \mathcal{N}_i} \right\|^2 = \sum_{\substack{k_1 \in \mathcal{N}_{j_1}, \\ k_2 \in \mathcal{N}_{j_2}}} \gamma_{j_1j_2k_1k_2} \langle \vec{n}_{ij_1k_1}, \vec{n}_{ij_2k_2} \rangle \Big|_{j_1, j_2 \in \mathcal{N}_i}$$

Five-body interactions@III

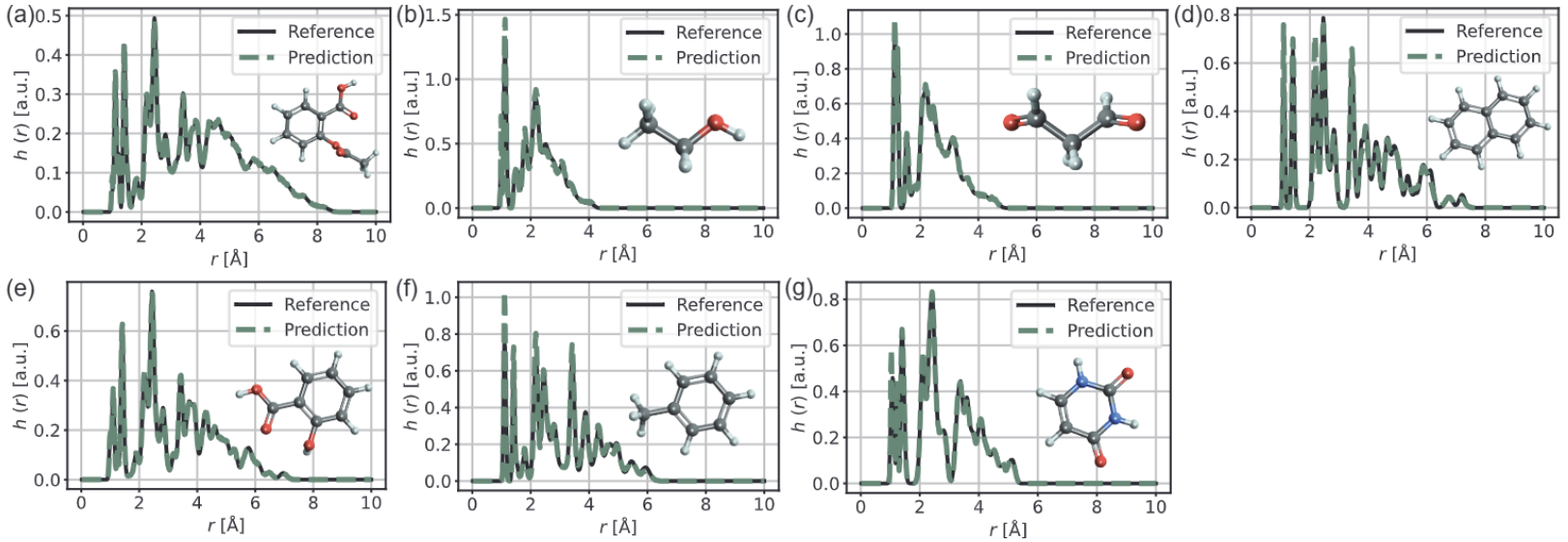
$$\left(\sum_{j \in \mathcal{N}_i} \alpha_j \hat{r}_{ij} \times \sum_{j \in \mathcal{N}_i} \beta_j \hat{r}_{ij} \right) \cdot \left(\sum_{k \in \mathcal{N}_j} \alpha_k \hat{r}_{jk} \times \sum_{k \in \mathcal{N}_j} \beta_k \hat{r}_{jk} \right) \Big|_{j \in \mathcal{N}_i} = \sum_{\substack{j1, j2 \in \mathcal{N}_i, \\ k_1, k_2 \in \mathcal{N}_j}} \gamma_{ij_1j_2k_1k_2} \langle \vec{n}_{ij_1j_2}, \vec{n}_{jk_1k_2} \rangle \Big|_{j \in \mathcal{N}_i}$$



Results

Benchmark on MD17 dataset

		SchNet [41, 42]	DimeNet [43]	PaiNN [15]	SpookyNet [45]	ET [16]	GemNet [17]	NequIP (l=3) [7]	SO3KRATES [46]	ViSNet [20]	QuinNet
Aspirin	Energy	0.37	0.204	0.167	0.151	0.123	-	0.131	0.139	**0.116**	0.119
	Force	1.35	0.499	0.338	0.258	0.253	0.217	0.184	0.236	0.155	**0.145**
Ethanol	Energy	0.08	0.064	0.064	0.052	0.052	-	0.051	0.052	0.051	**0.050**
	Force	0.39	0.230	0.224	0.094	0.109	0.085	0.071	0.096	**0.060**	**0.060**
Malonaldehyde	Energy	0.13	0.104	0.091	0.079	0.077	-	0.076	0.077	**0.075**	0.078
	Force	0.66	0.383	0.319	0.167	0.169	0.155	0.129	0.147	0.100	**0.097**
Naphthalene	Energy	0.16	0.122	0.116	0.116	**0.085**	-	0.113	0.115	**0.085**	0.101
	Force	0.58	0.215	0.077	0.089	0.061	0.051	**0.039**	0.074	**0.039**	**0.039**
Salicylic acid	Energy	0.20	0.134	0.116	0.114	0.093	-	0.106	0.016	**0.092**	0.101
	Force	0.85	0.374	0.195	0.180	0.129	0.125	0.090	0.145	0.084	**0.080**
Toluene	Energy	0.12	0.102	0.095	0.094	**0.074**	-	0.092	0.095	**0.074**	0.080
	Force	0.57	0.216	0.094	0.087	0.067	0.060	0.046	0.073	**0.039**	**0.039**
Uracil	Energy	0.14	0.115	0.106	0.105	**0.095**	-	0.104	0.103	**0.095**	0.096
	Force	0.56	0.301	0.139	0.119	0.095	0.097	0.076	0.111	**0.062**	**0.062**



Results

Benchmark on rMD17 dataset

		UNiTE [50]	GemNet (T/Q) [17]	NequIP (l=3) [7]	MACE [36]	Allegro [35]	BOTNet	ViSNet [20]	QuinNet
Aspirin	Energy	0.055	-	0.0530	0.0507	0.0530	0.0530	**0.0445**	0.0486
	Force	0.175	0.2191	0.1891	0.1522	0.1684	0.1960	0.1520	**0.1429**
Azobenzene	Energy	0.025	-	0.0161	0.0277	0.0277	0.0161	**0.0156**	0.0394
	Force	0.097	-	0.0669	0.0692	0.0600	0.0761	0.0585	**0.0513**
Benzene	Energy	0.002	-	0.0009	0.0092	0.0069	**0.0007**	**0.0007**	0.0096
	Force	0.017	0.0115	0.0069	0.0069	**0.0046**	0.0069	0.0056	0.0047
Ethanol	Energy	0.014	-	0.0092	**0.0032**	0.0092	0.0092	0.0078	0.0096
	Force	0.085	0.083	0.0646	**0.0484**	**0.0484**	0.0738	0.0522	0.0516
Malonaldehyde	Energy	0.025	-	0.0184	0.0185	0.0138	0.0185	**0.0132**	0.0168
	Force	0.152	0.1522	0.01176	0.0946	**0.0830**	0.1338	0.0893	0.0875
Naphthalene	Energy	0.011	-	**0.0046**	0.1153	**0.0046**	**0.0046**	0.0057	0.0174
	Force	0.060	0.0438	0.0300	0.0369	**0.0208**	0.0415	0.0291	0.0242
Paracetamol	Energy	0.044	-	0.0323	0.0300	0.0346	0.0300	**0.0258**	0.0362
	Force	0.164	-	0.1361	0.1107	0.1130	0.1338	0.1029	**0.0979**
Salicylic acid	Energy	0.017	-	**0.0161**	0.0208	0.0208	0.0185	**0.0161**	0.033
	Force	0.088	0.1222	0.0922	0.0715	**0.0669**	0.0992	0.0795	0.0771
Toluene	Energy	0.010	-	0.0069	0.0115	0.0092	0.0069	**0.0059**	0.0139
	Force	0.058	0.0507	0.0369	0.0350	0.0415	0.0438	0.0264	**0.0244**
Uracil	Energy	0.013	-	0.0092	0.0115	0.0138	0.0092	**0.0069**	0.0149
	Force	0.088	0.0876	0.0669	0.0484	**0.0415**	0.0738	0.0495	0.0487

QuinNet model shows comparable accuracy with the state-of-the-art models in **small** molecular datasets.

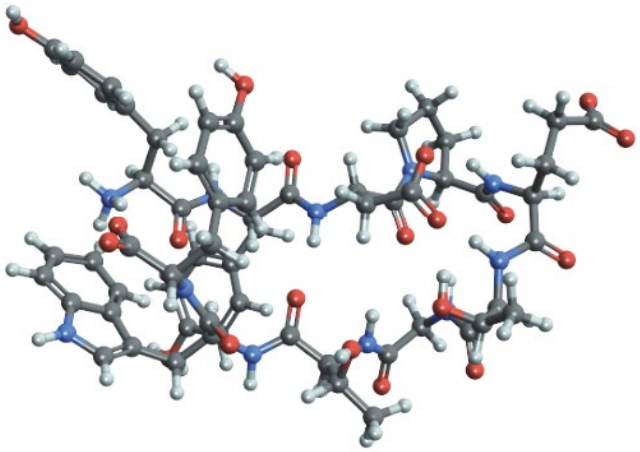
Results

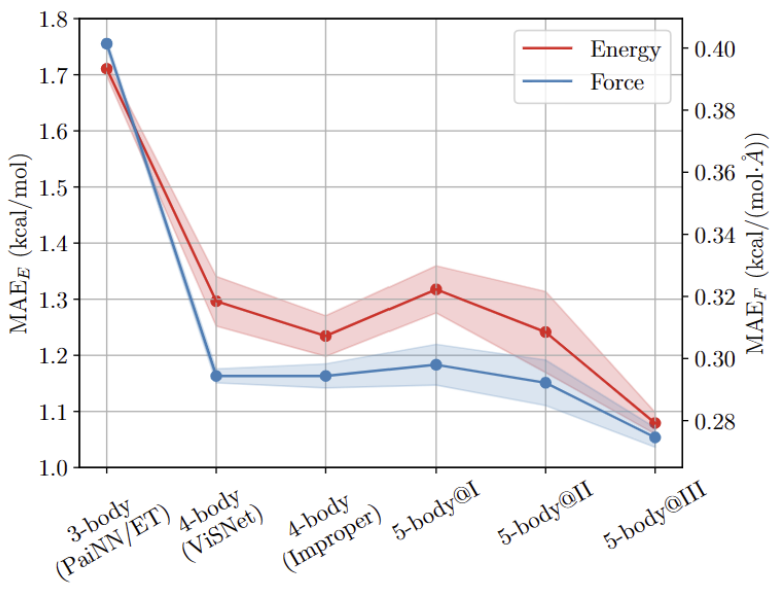
Benchmark on MD22 dataset

	# Train/Val		sGDML [51]	ViSNet-LSRM [52]	ViSNet [20,52]	MACE (3Å) [53]	MACE (6Å) [53]	MACE (5Å) [53]	QuinNet
Ac-Ala3-NHMe	5500/500	Energy	0.0093	**0.0016**	0.0019	0.0140	0.0080	0.0015	0.0020
		Force	0.79	0.0942	0.0972	0.1753	0.3920	0.0876	**0.0681**
DHA (docosahexaenoic acid)	7500/500	Energy	0.023	**0.0016**	0.0027	0.0103	0.0092	0.0024	0.0021
		Force	0.75	0.0598	0.0668	0.1430	0.5419	0.0646	**0.0515**
Stachyose	7500/500	Energy	0.046	**0.0012**	0.0015	0.0058	0.0082	0.0014	0.0026
		Force	0.68	0.0767	0.0869	0.1568	0.6226	0.0876	**0.0543**
AT-AT	2500/500	Energy	0.012	**0.0013**	0.0028	0.0208	0.0036	0.0018	0.0024
		Force	0.69	0.0781	0.1070	0.3067	0.3436	0.0992	**0.0687**
AT-AT-CG-CG	1500/500	Energy	0.012	**0.0010**	0.0017	0.0139	0.0038	0.0013	0.0032
		Force	0.70	**0.1064**	0.1563	0.3759	0.4635	0.1153	0.1273
Buckyball catcher	550/50	Energy	0.0079	**0.0029**	0.0030	0.0110	0.0039	0.0033	0.0038
		Force	0.68	0.1026	0.1335	0.3021	0.5120	**0.0853**	0.1091
Double-walled nanotube	750/50	Energy	0.0108	0.0049	**0.0028**	0.0048	0.0053	0.0045	0.0049
		Force	0.52	0.3391	0.3959	0.4128	0.9132	0.2767	**0.2473**

Results

Ablation study on Chignolin





		LSRM	3-body (ET)	4-body (ViSNet)	4-body (improper)	5-body@I	5-body@II	5-body (QuinNet)	QuinNet (6 Layer)
Chignolin	Energy	1.148	1.711± 0.012	1.296± 0.044	1.234± 0.036	1.317± 0.042	1.241± 0.072	1.079± 0.019	**1.036**
	Force	0.3638	0.4014± 0.0015	0.2944± 0.0022	0.2944± 0.0039	0.2980± 0.0066	0.2922± 0.0073	0.2747± 0.0030	**0.2665**

QuinNet accurately models these interactions and achieves higher accuracy in energy and force prediction compared to other models on the **larger** molecular systems.

Conclusions

- In this work, we propose the QuinNet architecture, which efficiently incorporates many-body interactions up to **whole five-body** in graph neural networks for molecular dynamics simulations.

- Our experiments on several public datasets, including MD17, revised MD17, MD22, and Chignolin, demonstrate that QuinNet achieves high accuracy **without significantly increasing computational complexity**.

- Notably, our ablation study on Chignolin highlights **the significance of five-body interactions** in accurately modeling complex bio-molecular systems.

Thank You!