ANISOTROPIC MESSAGE PASSING: GRAPH NEURAL NETWORKS WITH DIRECTIONAL AND LONG-RANGE INTERACTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph neural networks have shown great potential for the description of a variety of chemical systems. However, standard message passing does not explicitly account for long-range and directional interactions, for instance due to electrostatics. In this work, an anisotropic state based on Cartesian multipoles is proposed as an addition to the existing hidden features. With the anisotropic state, message passing can be modified to explicitly account for directional interactions. Compared to existing models, this modification results in relatively little additional computational cost. Most importantly, the proposed formalism offers as a distinct advantage the seamless integration of (1) anisotropic long-range interactions, (2) interactions with surrounding fields and particles that are not part of the graph, and (3) the fast multipole method. As an exemplary use case, the application to quantum mechanics/molecular mechanics (QM/MM) systems is demonstrated.

1 INTRODUCTION

Message passing graph neural networks (GNN) have shown great potential for the description of a wide range of chemical systems (Scarselli et al., 2009; Battaglia et al., 2016; 2018). Particularly the description of quantum molecular (QM) systems with machine learning (ML) potentials has received a lot of interest (Gilmer et al., 2017; Unke et al., 2021b). However, in its general form, message passing does not explicitly account for directionality, which plays an important role in many physical interactions (Glotzer & Solomon, 2007; Kramer et al., 2014). In recent years, a range of models which include directional information have been proposed (Anderson et al., 2019; Klicpera et al., 2020; Miller et al., 2020; Schütt et al., 2021). Especially models based on Clebsch-Gordan tensor products have shown superior data efficiency (Batzner et al., 2022; Batatia et al., 2022b; Musaelian et al., 2022). However, the relatively high cost of these operations might hinder the application to larger systems such as biomolecules in solution. This difficulty is compounded by growing evidence that message passing cannot accurately resolve long-range interactions (Alon & Yahav, 2020; Dwivedi et al., 2022).

In light of these challenges, which are exemplary illustrated in Figure 1, we propose a model with the aim to (1) include directional information while (2) retaining high computational efficiency and (3) incorporating (anisotropic) long-range interactions, including polarization. Specifically, the addition of an anisotropic state to the existing hidden features is proposed. This anisotropic state is based on Cartesian multipoles and expressed as a linear combination of local frames. As a result, these multipoles are equivariant under rotations and can be used to describe the anisotropy of interactions. The formalism is developed analogously to the concept of atomic multipoles commonly used in computational chemistry (Stone, 2013). The proposed modification results in relatively little computational overhead compared to standard message passing models. Most importantly, the formulation based on multipoles allows for the hybrid treatment of particles and fields. This is of particular interest for two cases: (1) Systems of particles embedded in an external field, and (2) systems with large numbers of particles where long-range interactions may be treated with the fast multipole method (Rokhlin, 1985).

As a practical use case, we consider quantum mechanics/molecular mechanics (QM/MM) simulations (Warshel & Levitt, 1976). In QM/MM simulations, a QM system is embedded in an external

electrostatic field generated by MM particles, allowing for an efficient description of large systems, for instance protein-ligand complexes or enzymatic reactions in solution. As such, the QM/MM formalism might be ideally suited to apply ML potentials to extended systems. Consequently, a ML/MM formalism within the message passing framework is formulated analogous to QM/MM.

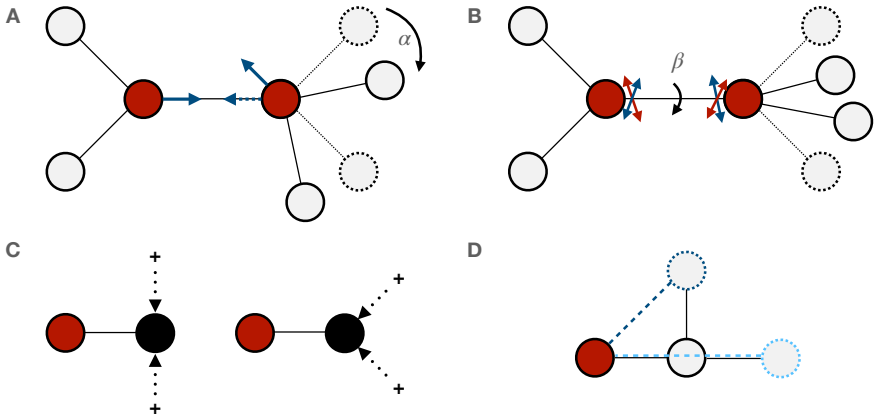

Figure 1: Exemplary situations arising in molecular systems for which message passing with distance labelled edges fails. (**A**): Rotation of one molecule by an angle α cannot be recognized without adding more edges. However, this change can be resolved through the addition of a dipole (blue arrow). (**B**): Similarly, a rotation by an angle β around the axis described by the aligned dipoles ('torsion') can be resolved by including a quadrupole (blue/red crossed arrows). (**C**): An electric field caused by charges (+) surrounding the molecule. While the electrostatic interaction is additive, the polarization on the black atom due to the opposing charges cancels out in the left case but adds up in the right case. (**D**): 'Long-range' interactions (dashed lines): Message passing can neither distinguish distance nor direction for the interactions between the red and the two blue-dotted particles.

2 Related Work

ML potentials have sparked interest as a possible solution to the steeply scaling computational cost of quantum chemical methods (Schuch & Verstraete, 2009; Unke et al., 2021b), with high-dimensional neural network potentials being some of the earliest proposed methods (Behler, 2011). With the introduction of GNN (Scarselli et al., 2009; Battaglia et al., 2016; 2018), the development of ML potentials has shifted away from handcrafted features to end-to-end differentiable models which learn features solely from distances and element types (Gilmer et al., 2017; Unke et al., 2021b). Recognizing that GNN may fail to distinguish certain graphs has spurred the development of modified message passing schemes (Morris et al., 2019; Maron et al., 2019; Pozdnyakov & Ceriotti, 2022). As a solution, the explicit integration of many-body interactions in the message passing (Kondor, 2018; Klicpera et al., 2020; Shui & Karypis, 2020; Zhao et al., 2021), directional messages (Anderson et al., 2019; Schütt et al., 2021), and combinations thereof (Gasteiger et al., 2021; Batatia et al., 2022b) have been proposed. Building on the concept of equivariant CNNs (Cohen & Welling, 2016; Weiler et al., 2018), equivariant message passing models based on Clebsch-Gordan tensor products have received increased attention (Anderson et al., 2019; Miller et al., 2020; Satorras et al., 2021; Brandstetter et al., 2021; Batzner et al., 2022; Musaelian et al., 2022). A categorization based on employed architectural features was recently proposed by Batatia et al. (2022a), offering an overview of essential ingredients of message passing models for ML potentials. Deficiencies in the description of long-range interactions have led to models that include explicit interaction terms (Unke & Meuwly, 2019; Ko et al., 2021), as well as models that integrate this information in the representation itself (Grisafi & Ceriotti, 2019; Grisafi et al., 2021; Unke et al., 2021a). As an alternative to the ML potentials mentioned so far, ML within the QM/MM formalism might facilitate the description of condensed phase systems. Recent work has highlighted the challenges and potential of ML for the simulation of QM/MM systems (Zhang et al., 2018; Böselt et al., 2021; Pan et al., 2021; Hofstetter et al., 2022; Giese et al., 2022).

3 METHODS

3.1 MESSAGE PASSING GRAPH NEURAL NETWORKS

Message passing GNN process graph-structured data. In its commonly used form, node, edge, and/or global features are iteratively refined. Various models have been proposed in recent years with differences found in the construction of the graph and the employed message passing mechanism (Scarselli et al., 2009; Battaglia et al., 2016; Gilmer et al., 2017; Battaglia et al., 2018).

Considering a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes $\boldsymbol{v}_i \in \mathcal{V}$ and edges $\boldsymbol{e}_{ij} \in \mathcal{E}$, message passing can be defined as

$$\boldsymbol{h}_i^{l+1} = \phi_h(\boldsymbol{h}_i^l, \sum_{j \in N(i)} \phi_e(\boldsymbol{h}_i^l, \boldsymbol{h}_j^l, \boldsymbol{u}_{ij})) \tag{1}$$

with $\boldsymbol{h}_i^l \in \mathbb{R}^n$ describing the hidden-feature vector of node $\boldsymbol{v}_i$ after l iterations, $\boldsymbol{u}_{ij} \in \mathbb{R}^n$ the edge feature of the edge $\boldsymbol{e}_{ij}$ between node i and j, $N(i)$ denoting the set of neighbours of $\boldsymbol{v}_i$, and ϕ_e and ϕ_h describing edge and node update functions. The superscript l denotes the current message passing iteration with n being the total number of message passing layers.

In the following paragraphs, we will refer to QM and MM particles. In the context of this work, QM particles are nodes $\boldsymbol{v}_i$ of a graph $\mathcal{G}$, which are labelled with a hidden feature $\boldsymbol{h}_i$. MM particles, on the other hand, are not part of the graph, and therefore do not hold a hidden feature. Instead, only a scalar partial charge q_i (monopole) is attached to MM particles. More generally, MM particles can be understood as a specific example of a source of an external field in which the graph is embedded. In addition, both QM and MM particles hold a position $\boldsymbol{r}_i \in \mathbb{R}^3$ in Cartesian space.

3.2 THE ANISOTROPIC STATE

The present work proposes the introduction of an internal anisotropic state $\mathbf{M}_i$ composed of Cartesian multipoles of order k to the existing hidden features. Cartesian multipoles are ideally suited for this task as they result in relatively little additional computational costs and conceptual complexity. Specifically, $\mathbf{M}_i^k \in \mathbb{R}^{3^n}$ refers to the k-th moment of the anisotropic state of node i composed of the monopole $\mathbf{M}_i^0$, the dipole $\mathbf{M}_i^1$ and so on. For clarity, the iteration superscript l will be omitted. However, in practice, the anisotropic state can be updated during each iteration much like the hidden feature $\boldsymbol{h}_i^l$.

The anisotropic state is expressed as a linear combination of a local basis (Thürlemann et al., 2022). This formulation preserves equivariance under rotations and invariance to translations. As the local basis, the traceless tensor product of order k of the unit vector $\hat{\boldsymbol{r}}_{ij}$ pointing from the position of node i to the position of its neighbour j is used

$$\mathbf{R}_{ij}^k = \hat{\boldsymbol{r}}_{ij} \otimes \hat{\boldsymbol{r}}_{ij} \otimes ... \tag{2}$$

with k referring to the order of the multipole and $\boldsymbol{r}_{ij}^0 = 1$. Each component of order k of the anisotropic state $\mathbf{M}_i^k$ is then expressed as

$$\mathbf{M}_i^k = \sum_{j \in N(i)} c_{ij}^k \mathbf{R}_{ij}^k \tag{3}$$

with coefficients c_{ij}^k predicted for each interaction as

$$c_{ij}^k = \phi_{M(k)}(\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{u}_{ij}) \tag{4}$$

where $\phi_{M(k)}$ refers to a learnable function. The scalar coefficient c_{ij}^k is based on the hidden features of the interacting nodes i and j as well as an edge feature $\boldsymbol{u}_{ij}$. The coefficient c_{ij}^k can be interpreted as the contribution of the node j to the polarization of order k of node i analogous to a (hyper-)polarizability. We note that the multipoles discussed here serve as a tool to introduce anisotropy. Albeit possible, they do not necessarily carry physical meaning.

3.3 ANISOTROPIC INTERACTION FEATURES

Given the anisotropic state $\mathbf{M}_i$ proposed in the previous section, multipole-multipole interactions can be defined. Formulation of the multipole interaction follows the notation and formalism described in (Smith, 1998; Lin, 2015; Burnham & English, 2020) which decomposes the interaction between multipoles on two distinct sites into a product of a radial component $b_k(||r_{ij}||)$ and coefficients of the multipole interactions $g^k_{ij}(\boldsymbol{r}_{ij})$, resulting in the following terms

$$u_{ij} = \sum_{k=0}^{\infty} b_k(||\boldsymbol{r}_{ij}||) g^k_{ij}(\boldsymbol{r}_{ij}) \tag{5}$$

with $b_k(||\boldsymbol{r}_{ij}||) = (2k-1)!!/||\boldsymbol{r}_{ij}||^{2k+1}$. In the context of machine learning, b_k can be reinterpreted as filters analogous to the continuous-filter convolutions introduced by Schütt et al. (2018). Consequently, the coefficients g^k_{ij} encode the angular information based on the relative orientation of the multipoles centered on the interacting nodes. In the notation of Burnham & English (2020), the coefficients can be generated with

$$\boldsymbol{g}^k_{ij} = \bigoplus_{d_i+d_c+d_j=k} \langle (\mathbf{M}_i^{(d_i+d_c)}.d_i.\mathbf{R}_{ij}^{d_i}), (\mathbf{M}_j^{(d_j+d_c)}.d_j.\mathbf{R}_{ij}^{d_j}) \rangle \tag{6}$$

given multipoles $\mathbf{M}_i^{(d_i+d_c)}$ of order $(d_j + d_c)$ on site i and the previously defined traceless tensor products of the vector $\hat{\boldsymbol{r}}_{ij}$. $\bigoplus$ refers to the concatenation of the scalar coefficients. All d_i, d_j, d_c where $d_i+d_j+d_c = k$ and $d_i, d_j, d_c \geq 0$ are included. d denotes the number of contractions, with d_i being the number of contractions over the indices of the first bracket, d_j the number of contractions in the second bracket and d_c the number of contractions between the two brackets, indicated by $\langle ., . \rangle$. Analogous to the multipole interaction defined in Eq. 5, the anisotropic feature $\boldsymbol{a}_{ij}$ is defined as

$$\boldsymbol{a}_{ij} = \phi_b(||\boldsymbol{r}_{ij}||) \circ G(\hat{\boldsymbol{r}}_{ij}, \mathbf{M}_i, \mathbf{M}_j) \tag{7}$$

where we use $\circ$ to denote the aforementioned element-wise multiplication with radial weights generated by ϕ_b and $G(\hat{\boldsymbol{r}}_{ij}, \mathbf{M}_i, \mathbf{M}_j)$ for the concatenated multipole interaction coefficients $\boldsymbol{g}^k_{ij}$ of all considered orders k.

3.4 ANISOTROPIC MESSAGE PASSING (AMP)

Based on the anisotropic feature defined in the previous section, the modified message passing is obtained as

$$\begin{aligned} \mathbf{M}_i^k &= \sum_{j \in N(i)} \phi_{M(k)}(\boldsymbol{h}_i^l, \boldsymbol{h}_j^l, \boldsymbol{a}_{ij}) \mathbf{R}_{ij}^k \\ \boldsymbol{a}_{ij} &= \phi_b(||\boldsymbol{r}_{ij}||) \circ G_{ij}(\hat{\boldsymbol{r}}_{ij}, \mathbf{M}_i, \mathbf{M}_j) \\ \boldsymbol{h}_i^{l+1} &= \phi_h(\boldsymbol{h}_i^l, \sum_{j \in N(i)} \phi_e(\boldsymbol{h}_i^l, \boldsymbol{h}_j^l, \boldsymbol{a}_{ij})) \end{aligned} \tag{8}$$

Thus, two additional steps are added to the standard message passing formalism: (1) Expansion of the multipoles on each node, and (2) contraction of the multipoles for each pair of interacting nodes. The resulting feature is used to incorporate directional information in the message $\boldsymbol{m}_{ij}$. For the use as a ML potential, the total potential energy is expressed as a sum of atom-based contributions of each QM particle, i.e. $V_{QM} = \sum_i^{QM} \phi_V(\boldsymbol{h}_i^n)$. We will refer to this model as AMP(k) where k denotes the highest order of the multipoles used and n is the number of graph layers.

3.5 QUANTUM MECHANICS/MOLECULAR MECHANICS (QM/MM)

QM/MM has become an integral part of computational chemistry since its inception in 1976 (Warshel & Levitt, 1976). QM/MM offers an efficient way to incorporate interactions with an environment, for instance solvent molecules and/or proteins, through combination of a QM and a classical (MM) description of a system. As such it retains the fidelity of QM-based methods for the region of interest, e.g. the solute, and the computational efficiency of classical force fields for the surroundings. Here, a brief description of the electrostatic embedding, one of several possible QM/MM schemes, is given. We refer the interested reader to Senn & Thiel (2007) for further details.

Within an electrostatic embedding, a designated QM zone interacts with the partial charges of the surrounding MM particles. For this purpose, the molecular Hamiltonian $\hat{H}_{QM}$ is extended by an interaction with the MM zone

$$\hat{H}_{QM/MM} = -\sum_{i}^{MM}\sum_{j}^{EL}\frac{q_i}{||\boldsymbol{r}_i - \boldsymbol{r}_j||} + \sum_{i}^{QM}\sum_{j}^{MM}\frac{Z_i q_j}{||\boldsymbol{r}_i - \boldsymbol{r}_j||} \tag{9}$$

where q_i are the partial charges of the MM zone and Z_i refer to the core charges of the nuclei. MM refers to the MM particles, EL to the electrons in the QM zone and QM to the QM nuclei. While the first term incorporates the interaction of the surrounding partial charges analogous to the core-electron interaction, the second term describes the Coulomb interaction between the nuclei within the QM zone and the partial charges. In other words, the QM system interacts with point charges representing the MM particles. The point charges polarize the QM system and interact with the resulting charge distribution. Remaining interactions between QM and MM particles, i.e. exchange-repulsion and dispersion, are described with classical force-field terms such as the Lennard-Jones potential (Jones & Chapman, 1924).

3.6 QM/MM for Message Passing: Direct Polarization

As a synthesis of the two previous sections, we propose a further modification to anisotropic message passing, which incorporates interactions with a surrounding field or particles that are not part of the graph.

$$\mathbf{M}_i^k = \sum_{j\in N_{QM}(i)} \phi_{M_{QM}(k)}(\boldsymbol{h}_i^l, \boldsymbol{h}_j^l, \boldsymbol{a}_{ij})\mathbf{R}_{ij}^k + \sum_{j\in N_{MM}(i)} \phi_{M_{MM}(k)}(\boldsymbol{h}_i^l, \boldsymbol{a}_{ij})\mathbf{R}_{ij}^k \tag{10}$$

Here, $N_{QM}(i)$ refers to the neighbours of i in the QM zone and $N_{MM}(i)$ to the neighbours of i in the MM zone. In addition to the previously described polarization due to neighbouring atoms that are part of the graph, a second term is introduced, which adds a contribution caused by the MM particles. The second term is conditioned on the hidden feature of the QM particle and the anisotropic feature $\boldsymbol{a}_{ij}$, which is based on the multipoles on the QM particle i and the partial charge on the MM particle. Evidently, this formulation permits also the application to polarizable embeddings schemes, use with the fast multipole method, and use of higher-order multipoles. We note that MM particles contribute only to the polarization of multipoles with order $k \geq 1$, analogous to the definition of (hyper-)polarizabilities (Stone, 2013). Since MM particles are only labelled with a charge and not with a hidden feature $\boldsymbol{h}_i^l$, the interaction depends solely on the feature of the QM particle and the anisotropic interaction feature $\boldsymbol{a}_{ij}$, which encodes the strength of the interaction and the relative orientation of the MM particle with respect to the polarized QM particle.

The advantages of the proposed interaction mechanism are twofold: First, the evaluation of messages between MM particles and between QM and MM particles is avoided. Instead, only the cheaper polarization term is evaluated for the more numerous QM–MM interactions. Second, directional and long-range interactions are efficiently incorporated in addition to the proper cancellation of opposing polarization terms. The total potential energy is then obtained as the atom-based contribution due to the QM particles described previously and a contribution due to the Coulomb interaction between QM and MM particles,

$$V_{QM/MM,ESP} = \sum_{i}^{QM}\sum_{j}^{MM}\frac{\phi_\rho(\boldsymbol{h}_i^n)\cdot q_j}{||\boldsymbol{r}_{ij}||} \tag{11}$$

where q_j denotes the partial charge of the respective MM particle and $\phi_\rho(\boldsymbol{h}_i^n)$ refers to a scalar charge density localized on the respective QM particle, analogous to a discretized Coulomb interaction. In other words, a pairwise interaction between each QM atom and its neighbours in the MM zone is added. In principle, $\phi_\rho(\boldsymbol{h}_i^n, \boldsymbol{a}_{ij})$ may be used to incorporate anisotropy. Here, only the hidden feature $\boldsymbol{h}_i^n$ is used. Thus, asymptotically correct long-range behaviour can be guaranteed by enforcing that the resulting scalar charges conserve total charge within the QM zone by subtraction of the mean excess charge, i.e. $\sum_i^{QM}\phi_\rho(\boldsymbol{h}_i^n) = 0$ for neutral systems. Expansion of $\phi_\rho(\boldsymbol{h}_i^n)$ itself may be an alternative route to obtain anisotropic QM/MM electrostatic interactions, as shown by Grisafi & Ceriotti (2019). The forces F_i acting on the QM and MM particles are obtained as the derivative of the negative total energy $(V_{QM/MM,ESP} + V_{QM})$ with respect to the positions $\boldsymbol{r}_i$ of the QM and MM particles, respectively. We will refer to models that include this direct polarization terms as AMP(k)-D. Use of the pairwise long-range potential in Eq. 11 will be denoted by '-P'.

4 RESULTS

Results are split into three parts: First, two model systems are investigated to explore the proposed modifications. Second, the model is applied to an existing dataset of QM/MM systems. Finally, results for QM9 are shown to allow comparison with other models.

4.1 WATER DIMERS: DIRECTIONAL FEATURES

As an illustrative example of the discussed concepts, a water dimer as shown in Figure 1B is investigated. Configurations were generated by rotating one molecule in steps of one degree around the axis described by the aligned dipoles, resulting in a total of 360 configurations. The molecules were separated by a distance of 4 Å. For such a system, three cutoff regimes are significant: (1) The cutoff r_{cut} used to construct the graph is larger than the maximal distance r_{max} in the system, resulting in a fully-connected graph ($r_{\text{max}} < r_{\text{cut}}$). (2) The cutoff is larger than the distance r_{OO} between the two opposing oxygen atoms but smaller than the maximal distance ($r_{\text{OO}} < r_{\text{cut}} < r_{\text{max}}$). (3) The cutoff is smaller than the distance between the oxygen atoms, resulting in two disconnected graphs ($r_{\text{cut}} < r_{\text{OO}}$). Particularly case (2) is interesting as it can be related to torsional degrees of freedom, which play an important role in molecular dynamics.

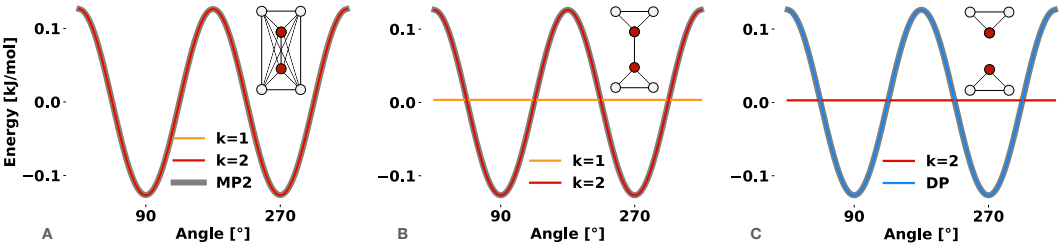

Figure 2: Potential energy (y-axis) for the quadrupole-quadrupole interaction between two water molecules for the proposed cutoff regimes with r_{cut} = 8.0 Å (A), r_{cut} = 4.25 Å (B), and r_{cut} = 2.0 Å (C). The relative orientation is indicated by the angle on the x-axis. The resulting graph topology for each cutoff is shown as inset. The reference (MP2) is shown in grey. Results for AMP(1), AMP(2), and AMP(2)-DP are shown as orange, red, and blue lines, respectively.

For all results shown in this section, a minimal 1-layer model was used. While the $k = 1$ model is still able to describe the interaction using a cutoff that results in a fully connected graph (i.e. case (1) shown in Figure 2A), the learning task seems to be more challenging for the AMP(1) compared to the AMP(2) model, indicated by marginally larger errors (MAE of 0.32 versus 0.05 J mol^{-1}). For case (2) shown in Figure 2B, the $k = 1$ model (orange line) fails to describe the interaction, as the dipole-dipole interaction remains invariant under the considered rotation. This issue can be traced back to the quadrupole–quadrupole interaction of the opposing oxygen atoms, motivating the need for $k \geq 2$. We note that this case is also interesting as it exemplary demonstrates the limitations of two existing directional message passing architectures, i.e. Dime and PaiNN Klicpera et al. (2020); Schütt et al. (2021). These two models seem to fail to describe this quadrupole-quadrupole interaction despite the use of directional information. Our interpretation is that for Dime this is due to the invariance of bond angles under the considered rotation, and for PaiNN because it only uses vectorial information (i.e. $k = 1$). Finally for case (3) shown in Figure 2C, all message passing models without additional long-range interactions fail (red line), motivating the proposed pairwise interaction and contributions to the polarization (models with suffix '-D' and '-P', shown in blue). We note that the long-range interaction would in principle also include dispersion, which we do not address here due to the availability of robust and efficient classical models.

4.2 WATER CLUSTERS: EXTENSIVITY

For this task, clusters of water molecules, ranging in size from six to 20 molecules, were investigated. For each cluster, one molecule was randomly selected as the QM zone while all other molecules were treated as MM particles. The goal of this task is the prediction of the difference in energy and gradients caused by placing the QM system in the external field generated by the MM particles. A

successful model must be able to discriminate between (1) the polarization of the QM zone by the external field, which does not directly scale with cluster size, and (2) the electrostatic interaction, which increases with cluster size. All models were trained on $1'289$ samples from clusters of size 10. All other cluster sizes were used for testing. For clarity, the cluster sizes are grouped in the results (Table 1). In the following paragraphs, the role of the proposed model modifications is probed.

Baseline: Message Passing To establish a baseline, a standard message passing model based on Eq. 1 was trained on the data. In this case, the MM particles are part of the graph. We also consider a second baseline model that uses the pairwise Coulomb interaction between QM and MM particles as described in Eq. 11. We refer to these two models as GNN and GNN-P.

Ablation Studies: Multipole Order and Polarization In addition, the role of the multipole order k was investigated. To compensate for different sizes of $\boldsymbol{a}_{ij}$ due to the multipole order, the number of radial weights was adjusted such that the total number of trainable parameters ($\sim$ 1M) remained constant. The resulting models are labelled as AMP(k)-D, where k is the highest multipole order used. The additional suffix '-P' indicates the use of the pairwise Coulomb interaction between QM and MM particles given in Eq. 11. In addition, the role of the contribution of the MM particles to the polarization defined in Eq. 10 was considered. The model without this additional polarization term is referred to as AMP(k)-P.

Table 1: Mean absolute errors (MAE) with respect to the reference (MP2) for the energy (E_{QM}) and gradients (F_{QM}) of water clusters of different size: 6-9, 10, 11-15, and 16-20. Note that the water clusters of size 10 were used for training. Energies are reported in kJ mol^{-1} and forces in kJ mol^{-1} Å^{-1}. % indicates the absolute relative error for E_{QM}.

	6-9			10 (Train)			11-15			16-20		
Model	E_{QM}	%	F_{QM}	E_{QM}	%	F_{QM}	E_{QM}	%	F_{QM}	E_{QM}	%	F_{QM}
GNN	3.74	4.3	2.39	**0.07**	**0.1**	0.07	3.13	2.6	1.97	6.59	5.1	3.54
GNN-P	0.71	0.8	1.21	0.09	**0.1**	**0.05**	0.88	0.8	1.44	2.66	2.3	3.16
AMP(0)-DP	6.05	7.4	8.08	5.11	5.3	8.40	5.24	4.8	8.20	4.91	3.8	8.36
AMP(1)-DP	1.69	1.9	1.10	0.25	0.2	0.18	0.95	0.9	1.78	4.79	4.0	6.61
AMP(2)-DP	**0.27**	**0.3**	**0.42**	0.16	0.2	0.12	**0.28**	**0.3**	**0.47**	**0.48**	**0.4**	**1.05**
AMP(3)-DP	0.30	0.4	0.48	0.16	0.2	0.10	0.42	0.4	0.72	0.82	0.7	1.67
AMP(3)-P	6.10	7.4	8.13	5.14	5.4	8.44	5.27	4.8	8.24	4.93	3.8	8.37

As expected, both models that do not incorporate polarization, i.e. AMP(3)-P and AMP(0)-DP, show the largest errors (Table 1). Adding the direct polarization, i.e., AMP(0)-DP versus AMP(1)-DP, results in the largest relative improvement. This is no surprise as there is no way for the information to flow from the MM particles to the QM particles in these models. This result demonstrates that the proposed modification allows the model to incorporate information about the surrounding particles without explicit message passing. Surprisingly, increasing k to 3, i.e. AMP(2)-DP versus AMP(3)-DP, seems to diminish the extrapolation capabilities, most notably for the largest clusters. As previously mentioned, a diminished ability to generalize to different cluster sizes is observed for both GNN models compared to the AMP-DP models. The fact that GNN-P and AMP(2)-DP perform comparable on the training set adds further evidence that this issue is, at least for this specific case, not caused by the lack of explicit directional information. Instead, this observation may indicate that the GNN model cannot distinguish between the two involved interactions. Assuming that '-P' captures the electrostatic interaction for both models equally well, results for the GNN-P model add further evidence that the error is mostly caused by a misrepresentation of polarization. Finally, it is worth mentioning that even for these relatively small systems, one training step with the GNN-P model takes approximately three times longer than the AMP(3)-DP model.

4.3 QM/MM Systems

As a more realistic example, the model was tested on a previously published datasets of QM/MM systems introduced by Böselt et al. (2021). These datasets include QM energies, QM gradients, and MM gradients of five molecules of varying size solvated in water. In this work, we focus on the three largest and most challenging systems: uracil, retinoic acid (RA) and the S-adenosyl methionine/cysteine dimer (SAM) with 12, 50 and $49 + 14$ QM atoms, respectively. On average, each molecule is surrounded by approximately $1'491$ (uracil), $2'391$ (RA) and $2'553$ (SAM) MM

particles. Each water molecule is represented by three partial charges centered on each atom. These datasets are challenging due to the large number of surrounding MM particles that polarize the QM system, the resulting electrostatic long-range interactions, as well as the size and flexibility of RA and SAM. A Δ-ML approach, which learned a correction on top of a semi-empirical tight-binding method (DFTB (Elstner et al., 1998)), was used in previous work (Böselt et al., 2021; Hofstetter et al., 2022).

Table 2: MAE for the QM energy (E_{QM}), QM gradients (F_{QM}), and MM gradients (F_{MM}) of the three QM/MM systems uracil, RA and SAM in water as described in Böselt et al. (2021). Energies are reported in kJ mol^{-1} and forces in kJ mol^{-1} Å^{-1}. The terms "unshuffled" and "shuffled" refer to the dataset with and without temporal ordering, respectively.

	AMP(3)-DP 50 Å (unshuffled)			AMP(3)-DP 50 Å (shuffled)		
System	E_{QM}	F_{QM}	F_{MM}	E_{QM}	F_{QM}	F_{MM}
Uracil	0.2/3.1/2.5	0.4/7.3/7.1	0.04/0.08/0.09	0.3/0.3/0.3	0.6/0.6/0.6	0.03/0.03/0.03
RA	0.5/9.1/12.0	1.2/5.0/9.6	0.06/0.16/0.19	0.6/0.6/0.6	1.4/1.5/1.5	0.06/0.06/0.06
SAM	2.2/29.1/44.3	3.3/16.1/16.0	0.20/0.39/0.34	3.1/3.2/3.3	3.2/3.5/3.5	0.15/0.15/0.15

No Δ-learning scheme was used in this work. In Table 2, results for a 3-layer AMP(3)-DP model are reported. The performance is compared for two pre-processing steps: While (unshuffled) uses the temporally sequential ordering used in the initial work Böselt et al. (2021); Hofstetter et al. (2022), (shuffled) refers to a shuffled (i.e. randomized) version of the same data. Comparing the same dataset (unshuffled), we observe in most cases larger errors compared to the previously reported Δ-models but smaller errors than previously reported results without Δ-learning (Böselt et al., 2021; Hofstetter et al., 2022). These results suggest that the proposed model can in principle successfully incorporate long-range polarization and electrostatic interactions. However, the relatively large drop in accuracy for the (unshuffled) dataset, which was previously observed and discussed in Hofstetter et al. (2022), indicates a limited capability to generalise to unseen data points. The difference between random splits or time splits might be important for commonly used benchmark datasets such as MD17 (Chmiela et al., 2017), as random splits tend to give a too optimistic performance assessment. This issue is already well known in biological activity prediction Sheridan (2013). A single call of a 3-layer AMP(3)-DP with a long-range cutoff of 50 Å (including QM energy, QM gradients, and MM gradients) for uracil, with 12 QM atoms and an average of 630 MM atoms, requires around 9.5 ms on a NVIDIA Titan V, averaged over a batch of size 40. For an AMP(2)-DP model, this cost is reduced to 7.5 ms. Thus, a (QM)ML/MM simulation based on the AMP model could in principle remove the bottleneck of current QM/MM molecular dynamics (MD) simulations almost entirely. However, further improvements are required to bring the cost down to a level comparable to the calculation of classical force fields, which is on the order of 1 ms/step for a system with around $14'000$ atoms (Eastman et al., 2017). Such improvements are clearly within reach, for instance through the use of the fast multipole method (Rokhlin, 1985). Nevertheless, it is important to mention that a low MAE does not guarantee a stable MD simulation (Stocker et al., 2022). In practice, a Δ-learning scheme might still be required to obtain stable simulations over long timescales. The robustness of the AMP(k)-DP model for the propagation of (QM)ML/MM MD simulations will therefore be investigated in the future.

Table 3: Comparison of long-range cutoffs. Energies are reported in kJ mol^{-1} and forces in kJ mol^{-1} Å^{-1}. The models used here were trained for fewer epochs than those in Table 2, resulting in slightly larger errors. The shuffled dataset was used. Results are given for a training/test split ($7'000$/$3'000$ data points).

	AMP(3)-DP 30 Å			AMP(3)-DP 50 Å		
System	E_{QM}	F_{QM}	F_{MM}	E_{QM}	F_{QM}	F_{MM}
Uracil	0.3/0.4	0.8/0.9	0.06/0.06	0.3/0.3	0.8/0.9	0.03/0.03
RA	5.3/5.5	2.9/3.1	0.14/0.14	1.5/1.6	2.5/2.7	0.13/0.13
SAM	6.0/6.4	6.1/6.4	0.45/0.45	5.5/5.7	5.3/5.5	0.33/0.33

In addition, two long-range cutoffs, 30 Å and 50 Å, for the interactions between QM and MM particles are investigated (Table 3). Notable performance differences are observed for the two investigated long-range cutoffs, which reaffirms the importance of long-range interactions. In both cases, a smooth cutoff function is used, thus the differences cannot simply be explained by cutoff artifacts.

In this context, it is further important to mention that an increased long-range cutoff does not change the number of parameters and QM–QM interactions but only the number of QM–MM interactions. A $50\,\text{Å}$ cutoff covers all observed QM–MM interactions for all systems. For a $30\,\text{Å}$ cutoff, the same is only true for uracil. Particularly for retinoic acid, for which the solvent-shell is elongated, the difference of the MAE for E_{QM} between the two cutoffs is considerable. Hence, even if a standard message passing architecture with six layers and a cutoff of $5\,\text{Å}$, could accurately resolve long-range interactions, these interactions would still fall outside the perceptive radius. However, given that there is little evidence that message passing models can faithfully resolve and incorporate such long-range interactions in the first place, this observation might be crucial for the future development of ML potentials.

4.4 QM9 Dataset

Results for a 4-layer AMP(3) model with a cutoff of $5\,\text{Å}$ for the 12 properties of the QM9 dataset are given in Table 4 and compared with existing models reported in the literature. Inference of a 4-layer AMP(3) model used here requires approximately $24\,\text{ms}$ for a batch of 100 samples with a total of $1'833$ atoms (NVIDIA Titan V). As can be seen in Table 4, both the computational cost and the reported results compare favourably with alternative models proposed in the literature. For several molecular properties, considerable improvements are observed. However, the error for targets related to the potential energy (i.e., U, U_0, H and G) is higher than for some other models (Allegro, PaiNN, PaxNet).

Table 4: Mean absolute error (MAE) for the 12 properties of the QM9 dataset (Ramakrishnan et al., 2014). AMP refers to the model proposed in this work. Values for alternative models in the literature are taken from: SchNet (Schütt et al., 2018), DimeNet (Klicpera et al., 2020), PaiNN (Schütt et al., 2021), Allegro (Musaelian et al., 2022), SEGNN (Brandstetter et al., 2021), and PaxNet (Zhang et al., 2022).

Target	Unit	SchNet	DimeNet	PaiNN	Allegro	SEGNN	PaxNet	AMP
U_0	meV	14	8.02	5.85	**4.7**	15	5.90	6.5
U	meV	19	7.89	5.83	**4.4**	13	5.92	7.1
H	meV	14	8.11	5.98	**4.4**	16	6.04	7.4
G	meV	14	8.98	7.35	**5.7**	15	7.14	6.9
ZPVE	meV	1.7	1.29	1.28	-	1.62	1.17	**0.88**
ϵ_{HOMO}	meV	41	27.8	27.6	-	24	22.8	**13.3**
ϵ_{LUMO}	meV	34	19.7	20.4	-	21	19.2	**11.8**
$\Delta\epsilon$	meV	63	34.8	45.7	-	42	31.0	**23.1**
μ	mD	33	28.6	12.0	-	23.0	10.8	**7.3**
α	ma_0^3	235	46.9	45	-	60	44.7	**32.5**
$\langle \text{R}^2 \rangle$	ma_0^2	73	**33.1**	66	-	660	93	167
c_v	$\frac{\text{mcal}}{\text{mol}\cdot K}$	33	24.9	24	-	31	23.1	**15.1**

5 Outlook

In this work, a modified message passing formalism that allows for the efficient incorporation of directional and long-range information was proposed. For future work, an implementation, which makes use of the fast multipole method, as well as the explicit inclusion of many-body interactions, as demonstrated by Klicpera et al. (2020) and Batatia et al. (2022b), and additional weights, as shown by Schütt et al. (2021), could be promising avenues. As a particular use case, we envision the application in (QM)ML/MM MD simulations, which could facilitate the study of large biomolecular systems and enzymatic reactions.

6 REPRODUCIBILITY STATEMENT

Code and model parameters used to produce the results in this work are made available at https://anonymous.4open.science/r/AMP-BC40/. All datasets used in this work are either already publicly available or made available under the same URL.

REFERENCES

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. *arXiv*, pp. arXiv:2006.05205, 2020.

Brandon Anderson, Truong-Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. *Adv. Neural Inf. Process. Syst.*, 32, 2019.

Ilyes Batatia, Simon Batzner, Dávid Péter Kovács, Albert Musaelian, Gregor NC Simm, Ralf Drautz, Christoph Ortner, Boris Kozinsky, and Gábor Csányi. The design space of e(3)-equivariant atom-centered interatomic potentials. *arXiv*, pp. arXiv:2205.06643, 2022a.

Ilyes Batatia, Dávid Péter Kovács, Gregor NC Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *arXiv*, pp. arXiv:2206.07697, 2022b.

Peter W. Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, and Koray Kavukcuoglu. Interaction networks for learning about objects, relations and physics. *Adv. Neural Inf. Process. Syst.*, 29, 2016.

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinícius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Çaglar Gülçehre, H. Francis Song, Andrew J. Ballard, Justin Gilmer, George E. Dahl, Ashish Vaswani, Kelsey R. Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matthew Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. *arXiv*, pp. arXiv:1806.01261, 2018.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.*, 13:2453, 2022.

Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys*, 134:074106, 2011.

Lennard Böselt, Moritz Thürlemann, and Sereina Riniker. Machine learning in QM/MM molecular dynamics simulations of condensed-phase systems. *J. Chem. Theory Comput.*, 17:2641–2658, 2021.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik Bekkers, and Max Welling. Geometric and physical quantities improve e(3) equivariant message passing. *arXiv*, pp. arXiv:2110.02905, 2021.

Christian J. Burnham and Niall J. English. A new relatively simple approach to multipole interactions in either spherical harmonics or cartesians, suitable for implementation into Ewald sums. *Int. J. Mol. Sci.*, 21:277, 2020.

Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.*, 3:e1603015, 2017.

Taco S. Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pp. 2990–2999, 2016.

TensorFlow Developers. Tensorflow, May 2021.

Robert A. Distasio JR., Ryan P. Steele, Young Min Rhee, Yihan Shao, and Martin Head-Gordon. An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order møller-plesset perturbation theory: Application to alanine tetrapeptide conformational analysis. *J. Comput. Chem.*, 28:839–856, 2007.

Brett I. Dunlap. Robust and variational fitting. *Phys. Chem. Chem. Phys.*, pp. 2113–2116, 2000.

Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007–1023, 1989.

Vijay Prakash Dwivedi, Ladislav Rampášek, Mikhail Galkin, Ali Parviz, Guy Wolf, Anh Tuan Luu, and Dominique Beaini. Long range graph benchmark. *arXiv*, pp. arXiv:2206.08164, 2022.

Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.*, 13:e1005659, 2017.

M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, 1998.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *Adv. in Neural Inf. Processing Syst.*, 34:6790–6802, 2021.

Timothy J. Giese, Jinzhe Zeng, Şölen Ekesan, and Darrin M. York. Combined qm/mm, machine learning path integral approach to compute free energy profiles and kinetic isotope effects in rna cleavage reactions. *J. Chem. Theory and Comput.*, 18:4304–4317, 2022.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pp. 1263–1272, 2017.

Sharon C. Glotzer and Michael J. Solomon. Anisotropy of building blocks and their assembly into complex structures. *Nature Materials*, 6:557–562, 2007.

Andrea Grisafi and Michele Ceriotti. Incorporating long-range physics in atomic-scale machine learning. *J. Chem. Phys.*, 151(20):204105, 2019.

Andrea Grisafi, Jigyasa Nigam, and Michele Ceriotti. Multi-scale approach for the prediction of atomic scale properties. *Chem. Sci.*, 12:2078–2090, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 770–778, 2016.

Albert Hofstetter, Lennard Böselt, and Sereina Riniker. Graph convolutional neural networks for (qm)ml/mm molecular dynamics simulations. *arXiv*, pp. arXiv:2206.14422, 2022.

J. E. Jones and Sydney Chapman. On the determination of molecular fields. I. From the variation of the viscosity of a gas with temperature. *Proc. R. Soc. Lond.*, 106:441–462, 1924.

Rick A. Kendall, Thom H. Dunning, and Robert J. Harrison. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *J. Chem. Phys.*, 96:6796–6806, 1992.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv*, pp. arXiv:1412.6980, 2017.

Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv*, pp. arXiv:2003.03123, 2020.

Tsz Wai Ko, Jonas A. Finkler, Stefan Goedecker, and Jörg Behler. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.*, 12:398, 2021.

Risi Kondor. N-body networks: A covariant hierarchical neural network architecture for learning atomic potentials. *arXiv*, pp. arXiv:1803.01588, 2018.

Christian Kramer, Alexander Spinn, and Klaus R. Liedl. Charge anisotropy: Where atomic multipoles matter most. *J. Chem. Theory Comput.*, 10:4488–4496, 2014.

Dejun Lin. Generalized and efficient algorithm for computing multipole energies and gradients based on cartesian tensors. *J. Chem. Phys.*, 143:114115, 2015.

Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Adv. Neural Inf. Process Syst.*, volume 32, 2019.

Benjamin Kurt Miller, Mario Geiger, Tess E Smidt, and Frank Noé. Relevance of rotationally equivariant convolutions for predicting molecular properties. *arXiv*, pp. arXiv:2008.08461, 2020.

Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, 1934.

Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 4602–4609, 2019.

Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *arXiv*, pp. arXiv:2204.05249, 2022.

Xiaoliang Pan, Junjie Yang, Richard Van, Evgeny Epifanovsky, Junming Ho, Jing Huang, Jingzhi Pu, Ye Mei, Kwangho Nam, and Yihan Shao. Machine-learning-assisted free energy simulation of solution-phase and enzyme reactions. *J. Chem. Theory and Comput.*, 17:5745–5758, 2021.

Robert M. Parrish, Lori A. Burns, Daniel G. A. Smith, Andrew C. Simmonett, A. Eugene DePrince, Edward G. Hohenstein, U. Bozkaya, Alexander Yu. Sokolov, Roberto Di Remigio, Ryan M. Richard, J. F. Gonthier, Andrew M. James, Harley R. McAlexander, Ashutosh Kumar, Masaaki Saitow, Xiao Wang, Benjamin P. Pritchard, Prakash Verma, Henry F. Schaefer, Konrad Patkowski, Rollin A. King, Edward F. Valeev, Francesco A. Evangelista, Justin M. Turney, T. Daniel Crawford, and C. David Sherrill. Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. *J. Chem. Theory Comput.*, 13:3185–3197, 2017.

Razvan Pascanu, Tomás Mikolov, and Yoshua Bengio. Understanding the exploding gradient problem. *arXiv*, pp. arXiv:1211.5063, 2012.

Sergey N Pozdnyakov and Michele Ceriotti. Incompleteness of graph convolutional neural networks for points clouds in three dimensions. *arXiv*, pp. arXiv:2201.07136, 2022.

Avijit Rakshit, Pradipta Bandyopadhyay, Joseph P. Heindel, and Sotiris S. Xantheas. Atlas of putative minima and low-lying energy networks of water clusters n = 3—25. *J. Chem. Phys.*, 151: 214307, 2019.

Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions. *arXiv*, 2017.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1:140022, 2014.

V Rokhlin. Rapid solution of integral equations of classical potential theory. *J. Comput. Phys.*, 60: 187–207, 1985.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 9323–9332, 2021.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Trans. Neural Netw. Learn. Syst.*, 20:61–80, 2009.

Norbert Schuch and Frank Verstraete. Computational complexity of interacting electrons and fundamental limitations of density functional theory. *Nat. Phys.*, 5:732–735, 2009.

Kristof T. Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. Schnet – a deep learning architecture for molecules and materials. *Journal of Chem. Phys.*, 148:241722, 2018.

Hans Martin Senn and Walter Thiel. *QM/MM Methods for Biological Systems*, pp. 173–290. Springer, 2007.

Robert P Sheridan. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *J. Chem. Inf. Model.*, 53:783–790, 2013.

Zeren Shui and George Karypis. Heterogeneous molecular graph neural networks for predicting molecule properties. In *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 492–500, 2020.

Daniel G. A. Smith, Lori A. Burns, Andrew C. Simmonett, Robert M. Parrish, Matthew C. Schieber, Raimondas Galvelis, Peter Kraus, Holger Kruse, Roberto Di Remigio, Asem Alenaizan, Andrew M. James, Susi Lehtola, Jonathon P. Misiewicz, Maximilian Scheurer, Robert A. Shaw, Jeffrey B. Schriber, Yi Xie, Zachary L. Glick, Dominic A. Sirianni, Joseph Senan O'Brien, Jonathan M. Waldrop, Ashutosh Kumar, Edward G. Hohenstein, Benjamin P. Pritchard, Bernard R. Brooks, Henry F. Schaefer, Alexander Yu. Sokolov, Konrad Patkowski, A. Eugene DePrince, U. Bozkaya, Rollin A. King, Francesco A. Evangelista, Justin M. Turney, T. Daniel Crawford, and C. David Sherrill. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.*, 152:184108, 2020.

William Smith. Point multipoles in the ewald summation (revisited). *Information Newsletter for Computer Simulation of Condensed Phases*, pp. 15–25, 1998.

Sina Stocker, Johannes Gasteiger, Florian Becker, Stephan Günnemann, and Johannes Margraf. How robust are modern graph neural network potentials in long and hot molecular dynamics simulations? *ChemRxiv*, 2022.

Anthony Stone. *The Theory of Intermolecular Forces*. Oxford University Press, 2013.

Moritz Thürlemann, Lennard Böselt, and Sereina Riniker. Learning atomic multipoles: Prediction of the electrostatic potential with equivariant graph neural networks. *J. Chem. Theory Comput.*, 18:1701–1710, 2022.

Justin M. Turney, Andrew C. Simmonett, Robert M. Parrish, Edward G. Hohenstein, Francesco A. Evangelista, Justin T. Fermann, Benjamin J. Mintz, Lori A. Burns, Jeremiah J. Wilke, Micah L. Abrams, Nicholas J. Russ, Matthew L. Leininger, Curtis L. Janssen, Edward T. Seidl, Wesley D. Allen, Henry F. Schaefer, Rollin A. King, Edward F. Valeev, C. David Sherrill, and T. Daniel Crawford. Psi4: An open-source ab initio electronic structure program. *WIREs Comput. Mol. Sci.*, 2:556–565, 2012.

Oliver T. Unke and Markus Meuwly. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.*, 15:3678–3693, 2019.

Oliver T. Unke, Stefan Chmiela, Michael Gastegger, Kristof T. Schütt, Huziel E. Sauceda, and Klaus-Robert Müller. Spookynet: Learning force ffields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.*, 12:7273, 2021a.

Oliver T. Unke, Stefan Chmiela, Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine learning force fields. *Chem. Rev.*, 121:10142–10186, 2021b.

Max Veit, David M. Wilkins, Yang Yang, Robert A. DiStasio, and Michele Ceriotti. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *J. Chem. Phys*, 153:024113, 2020.

Toon Verstraelen, Steven Vandenbrande, Farnaz Heidar-Zadeh, Louis Vanduyfhuys, Veronique Van Speybroeck, Michel Waroquier, and Paul W. Ayers. Minimal basis iterative stockholder: Atoms in molecules for force-field development. *J. Chem. Theory Comput.*, 12:3894–3912, 2016.

A. Warshel and M. Levitt. Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.

Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S Cohen. 3d steerable cnns: earning rotationally equivariant features in volumetric data. *Adv. Neural Inf. Process Syst.*, 31, 2018.

Shuo Zhang, Yang Liu, and Lei Xie. Efficient and accurate physics-aware multiplex graph neural networks for 3d small molecules and macromolecule complexes. *arXiv*, pp. arXiv:2206.02789, 2022.

Yin-Jia Zhang, Alireza Khorshidi, Georg Kastlunger, and Andrew A. Peterson. The potential for machine learning in hybrid qm/mm calculations. *J. Chem. Phys.*, 148(24):241740, 2018.

Lingxiao Zhao, Wei Jin, Leman Akoglu, and Neil Shah. From stars to subgraphs: Uplifting any gnn with local structure awareness. *arXiv*, pp. arXiv:2110.03753, 2021.

A Appendix

A.1 General Implementation Details

If not noted otherwise, the following architecture and hyperparameters were used. Three message passing layers were used except for the QM9 dataset, where we used four message passing layers. The architecture looked as follows:

1. Generation of edge weights with [Linear(16)], which parametrized ϕ_b. As input, the onehot features of the two connecting atoms and Bessel function ($n = 20$) expanded distance features were used.
2. Embedding of node features: [Linear(128)]
3. Message layers with [Linear(128), Swish, Linear(128), Swish] for ϕ_h and [Linear(128), Swish, Linear(128), Swish, Linear(k+1)] for the prediction of the coefficients c_{ij}, i.e., $\phi_{M(k)}$, $\phi_{M_{QM}(k)}$ and $\phi_{M_{MM}(k)}$. The coefficients for all orders k were predicted with the same module.
4. A readout layer [Linear(128), Swish, Linear(128), Swish, Linear(1)].

Linear(n) refers to a standard fully connected network layer with bias. No bias was used for the last layer of the output. For each message passing iteration, independent modules were used. Swish was employed as the non-linearity (Ramachandran et al., 2017). Layer weights were initialized based on the method proposed by He et al. (2015). Coefficients c_{ij} were scaled by 10^{-4} to support training stability during early epochs.

Node features were initialized as onehot vectors of the element. Norms of the multipoles were concatenated with the hidden feature. Edges were added for all atom pairs within a distance of 5 Å. For interactions between QM and MM particles, an independent long-range cutoff of 50 Å was used. Radial weights were conditioned on the initial onehot features, i.e., $\phi_b(\boldsymbol{h}_i^0, \boldsymbol{h}_j^0, ||\boldsymbol{r}_{ij}||)$. Distances $||\boldsymbol{r}_{ij}||$ were encoded with the enveloped Bessel functions proposed by Klicpera et al. (2020) with 20 functions and $p = 5$ for the envelope. The frequencies were not optimized. A ResNet-like update of the hidden feature was used, i.e., $\boldsymbol{h}_i^{l+1} = \boldsymbol{h}_i^l + \phi_h(\boldsymbol{h}_i^l, \boldsymbol{m}_i)$ (He et al., 2016). Model parameters were optimized using ADAM (Kingma & Ba, 2017) with default parameters ($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-7}$) and an exponentially decaying learning rate ($5 \cdot 10^{-4}$, 10^{-5}). Gradients were clipped by their global norm with a clip norm of 1 (Pascanu et al., 2012). Mean-squared errors were optimized. If the model was jointly trained on energies and gradients, the following loss function was used

$$\mathcal{L} = \frac{(1-\lambda)}{B} \sum_b^B (\hat{V}_b - V_b)^2 + \frac{\lambda}{3BN} \sum_i^{BN} \sum_{\alpha=0}^{2} \Big(-\frac{\partial \hat{E}_b}{\partial r_{i,\alpha}} - F_{i,\alpha} \Big)^2 \tag{12}$$

with the potential energy V and the force component $F_{i,\alpha}$ for each Cartesian dimension α. $\lambda = 0.8$ was used to balance the contributions of the energies and gradients to the loss. The loss term for the MM gradients was obtained in the same way but scaled by an additional factor 100 as proposed in (Hofstetter et al., 2022). Models were implemented with TensorFlow (2.6.2 and 2.9.1) (Abadi et al., 2015; Developers, 2021). Single precision `float32` was used.

A.2 Water Dimer

Calculations were performed with PSI4 (1.6) (Turney et al., 2012; Parrish et al., 2017; Smith et al., 2020) on a MP2/aug-cc-pVTZ level of theory using density fitted MP2 (Dunlap, 2000; Distasio JR. et al., 2007). The molecules were placed such that the two oxygen atoms were separated by 4 Å. Structures were assigned to training and test sets in an alternating manner. Single layer models were used for all experiments. Models were trained for $10'000$ steps using a batch size of five. 64-dimensional features were used as node features and four edge weights. Since the whole system was simulated with MP2, no partial charge embedding was used. Instead, the pairwise interaction for the model '-DP' was implemented as

$$V_{ESP} = \sum_i \sum_{j>i} \frac{\phi_\rho(\boldsymbol{h}_i) \cdot \phi_\rho(\boldsymbol{h}_j)}{||\boldsymbol{r}_{ij}||} \tag{13}$$

for all long-range pairs, i.e., $||\boldsymbol{r}_{ij}|| \geq r_{\text{cut}}$. Double precision `float64` was used for this system.

A.3 Water Clusters

The dataset with the water clusters was created based on the geometries presented in (Rakshit et al., 2019) For each set of clusters ranging from 6 to 20 water molecules, up to $2'000$ clusters were randomly sampled. For each cluster, one water molecule was randomly designated as the QM zone with the remaining molecules forming the MM zone. The long-range cutoff was set to 30 Å , including all QM–MM interactions. All accessible clusters were used, i.e., all known clusters within a 5 kcal/mol energy window above the putative minimum. All clusters of size $n = 10$ were used for training while the other clusters were used for testing. For the GNN models, the energy was predicted for the QM particles. Furthermore, in addition to the existing atom types (oxygen-QM and hydrogen-QM), two new types were introduced: oxygen-MM and hydrogen-MM. Partial charges of the MM zone were assigned based on MBIS monopoles (Verstraelen et al., 2016), i.e., $-0.914197e$ for oxygen and $0.457098e$ for hydrogen. Calculations were performed on a MP2/aug-cc-pVTZ level of theory (Møller & Plesset, 1934; Dunning, 1989; Kendall et al., 1992) using density fitted MP2 (Dunlap, 2000; Distasio JR. et al., 2007) as implemented in PSI4 (1.6) (Turney et al., 2012; Parrish et al., 2017; Smith et al., 2020). Gradients and energies of the isolated QM zone were subtracted. The models were trained over $50'000$ steps, presenting each time a batch of 16 randomly drawn samples.

A.4 QM/MM Systems

The datasets were taken from Böselt et al. (2021) and randomly split into training/validation/test sets of $7'000/2'000/1'000$ samples, respectively. Polarization due to the MM particles described in

Eq. 10 was only incorporated during the last message passing step. In addition, only eight distance weights were used for interactions between QM and MM particles while the default of 16 was used for edges within the graph, i.e., between QM particles. The models were trained over $400'000$ steps using batches of size 5 for uracil and batches of size 1 for retinoic acid and SAM/cysteine. For the comparison between the long-range cutoffs, only $200'000$ training steps were used.

A.5 QM9 DATASET

The dataset was taken from Ramakrishnan et al. (2014). An independent 4-layer AMP(3) model was trained for each property over $2'000$ epochs with an exponentially decaying learning rate ($5 \cdot 10^{-4}$, $5 \cdot 10^{-6}$). A 4-layer AMP(3) results in $1'153'793$ trainable parameters. Properties in units of Eh were converted and trained in units of kcal mol^{-1}. Each property was predicted as the sum of atomic contributions except for the dipole moment μ. For all properties, except for the dipole moment μ, the mean of the training set was subtracted. In addition, the energy of the isolated atoms was subtracted for U_0, U, H, and G. $3'054$ molecules for which the geometry consistency checks failed were removed from the dataset Ramakrishnan et al. (2014). The remaining data points were randomly split into a training/validation/test sets ($110'000$, $10'000$, $10'697$) using a batch size of 100. As proposed by Veit et al. (2020), the molecular dipole moment $\boldsymbol{\mu}$ was modeled as the sum of the contribution from atomic monopoles q_i and atomic dipoles $\boldsymbol{\mu}_i$

$$\boldsymbol{\mu} = \sum_i \phi_q(\boldsymbol{h}_i^n)\boldsymbol{r}_i + \boldsymbol{\mu}_i \tag{14}$$

with each atomic dipole $\boldsymbol{\mu}_i$ being predicted according to Eq. 3. The mean excess charge was subtracted from each q_i to enforce charge conservation.