Learning Atomic Multipoles: Prediction of the Electrostatic Potential with Equivariant Graph Neural Networks

Moritz Thürlemann, Lennard Böselt, and Sereina Riniker*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The accurate description of electrostatic interactions remains a challenging problem for classical potential-energy functions. The commonly used fixed partial-charge approximation fails to reproduce the electrostatic potential at short range due to its insensitivity to conformational changes and anisotropic effects. At the same time, possibly more accurate machine-learned (ML) potentials struggle with the long-range behavior due to their inherent locality ansatz. Employing a multipole expansion offers in principle an exact treatment of the electrostatic potential such that the long-range and short-range electrostatic interactions can be treated simultaneously with high accuracy. However, such an expansion requires the calculation of the electron density using computationally expensive quantum-mechanical (QM) methods. Here, we introduce an equivariant graph neural network (GNN) to address this issue. The proposed model predicts atomic multipoles up to the quadrupole, circumventing the need for expensive QM computations. By using an equivariant architecture, the model enforces the correct symmetry by design without relying on local reference frames. The GNN reproduces the electrostatic potential of various systems with high fidelity. Possible uses for such an approach include the separate treatment of long-range interactions in ML potentials, the analysis of electrostatic potential surfaces, and static multipoles in polarizable force fields.
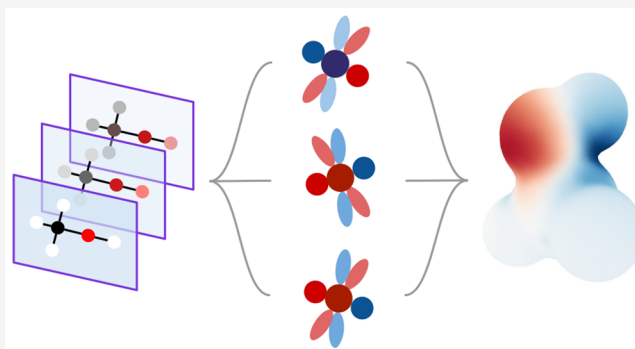
1. INTRODUCTION

The accurate treatment of the electrostatic potential (ESP) is crucial for the description of molecular interactions.[1] In classical biomolecular force fields (FFs), the electrostatic interaction is approximated using fixed, i.e., time- and conformation-invariant, partial charges.[2] Higher-order multipoles are not treated explicitly, thus neglecting the anisotropic part of the electrostatic interaction. Short-range interactions are therefore not accurately described for most systems of biological and pharmaceutical interest.[3,4] The desire for a more accurate description of the ESP motivated the development of polarizable FFs such as AMOEBA,[5] which describes the electrostatic interaction with higher moments up to quadrupoles, as well as several FFs used in crystal structure prediction.[6,7]

In recent years, there has been considerable effort to replace expensive ab initio and DFT calculations with machine-learned (ML) potentials.[8−17] In many cases, these models moved away from the historically established decomposition of the system energy into separate interactions. Instead, the total potential energy is decomposed into atomic contributions. While generally providing promising results at short range, these ML models often fail to model long-range interactions with sufficient accuracy.[18] In principle, the treatment of the ESP based on multipoles is able to resolve not only the long-range deficiency of ML models but also the inaccuracy of the partial-charge approximation in fixed-charge FFs. However, an explicit treatment of electrons is required to obtain multipoles, which is often cost prohibitive. Predicting multipoles with ML could remove this hurdle.

Previous efforts to use ML for the prediction of electrostatic interactions have mostly focused on partial charges, with either models that solely predict partial charges,[19−21] models that combine a local treatment with an explicit treatment of long-range interactions based on partial charges,[8,22,23] or models that predict partial charges as an auxiliary variable for properties such as the molecular dipole.[24,25] Due to the smaller influence of higher-order atomic multipoles on the ESP and their orientational dependence, these terms have found much less attention. Particularly, the orientational dependence poses a challenge for commonly deployed invariant descriptors. In previous work with ML models and in several polarizable FFs, this problem was circumvented by mapping multipole components onto a local reference frame spanned by

neighboring atoms.[26−28] However, local reference frames may lack transferability. In an alternative approach, Glick et al.[29] proposed Cartesian message-passing neural networks by introducing vectors between neighboring atoms as features. The resulting model was used to successfully predict higher-order multipoles. However, the proposed model architecture does not enforce the proper symmetry by design. Equivariance is hence only approximately achieved through data augmentation.

Recently, ML methods that respect such symmetries were introduced.[30−36] While not in all cases used to predict equivariant properties, several models applied to chemical problems showed that introducing directional information through equi- or covariant layers improves accuracy while often requiring fewer parameters and less training data than comparable models that do not take advantage of directional information.[9,37,38] Schütt et al.[39] introduced a model based on equivariant message passing, which includes directional information and is used to predict equivariant properties. For instance, their work showed that, similarly to ref 25, the introduction of atomic dipoles improves the prediction of infrared and Raman spectra of ethanol and aspirin in vacuum.

In this work, an ML model for the prediction of multipoles based on a standard message-passing graph neural network (GNN) architecture is introduced.[40,41] GNNs can be considered a natural choice for problems in (computational) chemistry, where nodes represent atoms and edges describe bonds or interactions between atoms. In this setting, GNNs have been applied to a wide range of problems, often delivering state-of-the-art results.[9,16,42,43] The graph construction and the manner in which information flows through the model are the main features discriminating between different GNN flavors. Due to its wide applicability, a standard message-passing architecture was chosen as introduced by Battaglia et al.[40,41] Most importantly, we employ a model based on the concepts of $E(n)$-equivariant GNN introduced by Satorras et al.[34] The proposed method provides an exceedingly simple and elegant approach, which introduces equivariance for standard message-passing GNNs without adding significant model complexity or computational overhead. While this approach was only applied to the prediction of positions and velocities in the original work, we show that it can be extended to the prediction of atomic multipoles. Due to its structure, the developed ML model enforces the proper symmetry by design, i.e., equivariance with respect to rotations.

2. THEORY

2.1. Electrostatic Potential (ESP) and Multipoles.
The ESP of a charge distribution $\rho(r')$ at a point r is defined as

$$V_{ES}(r) = \int_{V'} \frac{\rho(r')}{|r - r'|} dV' \tag{1}$$

Taylor expansion of $V_{ES}(r)$ around the center of the charge distribution gives rise to the following well-known multipole expansion[1]

$$V_{ES}(r) = \frac{M^{(0)}}{|r|} + \frac{M_\alpha^{(1)} r_\alpha}{|r|^3} + \frac{M_{\alpha\beta}^{(2)}(3r_\alpha r_\beta - r^2 \delta_{\alpha\beta})}{2|r|^5} + \dots \tag{2}$$

where $M^{(0)}$, $M^{(1)}$, and $M^{(2)}$ correspond to the monopole (q), the dipole (μ), and the quadrupole (θ), respectively. Greek indices run over the components of the coordinate system.

Note that the prefactor $(4\pi\epsilon_0)^{-1}$ is omitted throughout this work for simplicity.

Following the same argumentation and assuming that a molecular charge density can be decomposed into atomic charge densities,

$$\rho(r) = \sum_i \rho_i(r) \tag{3}$$

the same concept can be applied to each atomic density, giving rise to atomic or distributed multipoles.[1,44]

2.2. Atomic Multipoles.
Several methods for the calculation of atomic multipoles have been introduced, including distributed multipole analysis,[45] the transferable atom equivalents method,[46] and Gaussian multipole models[47] to name a few. In addition, several approaches that fit multipoles directly to the ESP have been proposed.[28,48] In a recent work, Verstraelen et al.[49−51] proposed the minimal basis iterative Stockholder (MBIS) method. MBIS was chosen for its low computational cost and fast convergence with respect to the multipole order. MBIS can be considered an atoms-in-molecule method, which attempts to partition the total electron density into atomic contributions. The method is based on the minimization of the Kullback−Leibler (KL) divergence between a pro-density $\rho_A^0(r)$ based on a minimal expansion in atom-centered s-type Slater functions and a target molecular density $\rho_A(r)$.[52]

$$\rho_A^0(r) = \sum_{i=1}^{m_A} \rho_{Ai}^0(r) \tag{4}$$

where $\rho_A^0(r)$ is the pro-atomic density constructed from Slater functions, i.e.

$$\rho_{Ai}^0(r) = N_{Ai} f_{Ai}(r) = \frac{N_{Ai}}{\sigma_{Ai}^3 8\pi} \exp\left(-\frac{|r - R_A|}{\sigma_{Ai}}\right) \tag{5}$$

with a fitting parameter N_{Ai} for the electron population, σ_{Ai} for the width, and R_A as the atomic center. The divergence between the pro-density and the target density is minimized by iteratively adjusting the population and width of each atomic shell, optimizing the KL divergence for every atom

$$\Delta S[\{\rho_A\}; \{\rho_A^0\}] = \sum_{A=1}^{N_{atoms}} \int \rho_A(r) \log\left(\frac{\rho_A(r)}{\rho_A^0(r)}\right) dr \tag{6}$$

The converged pro-density can then be processed to obtain desired properties, for instance, partial charges or atomic multipoles.[51]

2.3. Graph Neural Networks.
In the following sections, we follow the notation used by refs 16 and 34. GNNs are artificial NN (ANN) parametrized ML models, which process graph-structured data. In their commonly used form, node, edge, and/or global features are iteratively refined based on the current features. GNN models differ mainly by the features used, the way the underlying graph is constructed, as well as the updating or feature-refinement process applied.[16,40]

Considering a graph $G = (V, E)$ with nodes $v_i \in V$ and edges $e_{ij} \in E$, message passing can be defined as

$$m_{ij} = \phi_e(h_i^l, h_j^l, a_{ij})$$

$$m_i = \sum_{j \in N(i)} m_{ij}$$

$$h_i^{l+1} = \phi_h(h_i^l, m_i) \tag{7}$$

where $h_i^l \in \mathbf{R}^n$ describes the hidden-feature vector of node ν_i after l steps, $a_{ij} \in \mathbf{R}^n$ is the edge feature of edge e_{ij} between node i and j, $N(i)$ denotes the set of neighbors of ν_i, and ϕ_e and ϕ_h describe update functions, which are commonly parametrized by an ANN. The superscript l is the current layer or in its recurrent realization the current iteration. In the present work, atoms are represented by nodes ν and interactions between atoms as edges e.

2.4. Equivariance. Following the formalism used in ref 34, a function $\phi: X \rightarrow Y$ mapping from an input space X to an output space Y is called equivariant to a group action $g \in G$ if

$$\phi(\hat{T}_g(x)) = \hat{S}_g(\phi(x)) \tag{8}$$

for a transformation $\hat{T}_g$ on X, an equivalent transformation $\hat{S}_g$: $Y \rightarrow Y$ on the output space Y, and a group G.[34] In other words, equivariance describes maps that commute under the action of a group.[36,53]

For the practical case of a dipole $\mu(X)$ of an atom or molecule positioned at X, we require that $\hat{R}(\mu(X)) = \mu(\hat{R}(X))$ for an arbitrary rotation or reflection $\hat{R} \in \mathbf{O}(3)$. In this specific example, $\mathbf{O}(3)$ would be the group G, an arbitrary rotation or reflection would be the transformation $\hat{T}_g$ applied to the dipole $\mu(X)$ or the coordinates X, and μ would be the equivariant function ϕ. In other words, the dipole should rotate in the same way as the coordinates rotate. Any function computing the dipole of a system X must fulfill this condition. Note that it is not possible to fulfill this condition with ML models based on commonly employed invariant descriptors. We therefore introduce equivariance for GNNs to predict atomic multipoles following the ideas reported by Satorras et al. in ref 34. The authors implemented an $E(n)$ equivariance by introducing an auxiliary variable $x \in \mathbf{R}^n$ for each node in the graph. In their work, x represents a position or velocity in Cartesian coordinates. Further, in order to maintain equivariance, the following message-passing function was proposed

$$m_{ij} = \phi_e(h_i^l, h_j^l, \lVert x_i^l - x_j^l \rVert^2, a_{ij})$$

$$x_i^{l+1} = x_i^l + C \sum_{j \neq i} (x_i^l - x_j^l) \phi_x(m_{ij})$$

$$m_i = \sum_{j \in N(i)} m_{ij}$$

$$h_i^{l+1} = \phi_h(h_i^l, m_i) \tag{9}$$

where C is a normalization factor and x_i denotes the equivariant vector of atom i. This rule effectively separates equivariant from invariant features. In the subsequent sections we adapt this concept to the prediction of atomic multipoles of molecular systems.

2.5. Equivariance for Multipoles. The notation for multipoles follows the work by Burnham and English.[44] The present approach relies on the assumption that every multipole can be decomposed into a linear combination of coefficients depending on the local environment of an atom (invariant

features) and a spatial component (equivariant feature) depending on the relative orientation between atoms. Analogous to the model described in the previous section, the coefficients are predicted based on invariant features while the spatial (i.e., equivariant) part is taken as the k-times outer product of R_{ij}

$$R_{ij}^{(k)} = R_{ij} \otimes R_{ij} \otimes \dots \tag{10}$$

where k denotes the order of the multipole and R_{ij} labels the vector $R_{ij} = R_j - R_i$ between two atoms. Accordingly, the following composition for $M_i^{(k)}$, the kth multipole of atom i, is proposed

$$M_i^{(k)} = \sum_{j \in N(i)} c_{ij}^{(k)} R_{ij}^{(k)} \tag{11}$$

with a scalar coefficient $c_{ij}^{(k)} \in \mathbf{R}$ depending on the atom pair ij and $R_{ij}^{(k)}$ as the kth outer product of the vector pointing from atom i to atom j. Generally, $c_{ij}^{(k)}$ is predicted based on the features of atom i and atom j as

$$c_{ij}^{(k)} = \phi_{M^{(k)}}(h_i^n, h_j^n, a_{ij}) \tag{12}$$

where $\phi_{M^{(k)}}$ denotes an ANN parametrized function, mapping two atomic feature vectors h_i^n and h_j^n and a distance-based edge feature a_{ij} to a scalar coefficient.

Initially, atomic features are iteratively updated using a standard message-passing GNN. Since nodes and edges depend only on atom types and distances, these features contain purely rotation- and translation-invariant information. After n message-passing iterations, the atomic features h_i^n serve as the descriptor for the prediction of the scalar coefficients in the subsequent steps. We note that h_i^n could be replaced by any other invariant feature proposed in the literature. In particular, even atomic features which are purely molecular graph-based can be used while still retaining some degree of conformational dependence for atomic dipoles and quadrupoles.

Beginning with the atomic monopole, $M_i^{(0)}$ can be predicted as

$$M_i^{(0)} = \sum_{j \in N(i)} R_{ij}^{(0)} \phi_{M^{(0)}}(h_i^n, h_j^n, a_{ij}) \tag{13}$$

where $R_{ij}^{(0)}$ is the identity. As proposed by Metcalf et al.,[20] charge conservation can be enforced elegantly by imposing antisymmetry on $\phi_{M^{(0)}}$ with respect to the hidden features h_i^n and h_j^n. Alternatively, an atom-based scheme can be applied where each atomic monopole $M_i^{(0)}$ is predicted based on the hidden feature h_i^n of the respective atom, i.e.

$$M_i^{(0)} = \phi_{M^{(0)}}(h_i^n) \tag{14}$$

In this case, charge conservation can be enforced by subtracting the mean total charge of a given system (Z) by the mean predicted charge $\langle q \rangle = \frac{Z}{N} - \frac{1}{N}\sum_i^N q_i$ from each atomic monopole. In preliminary tests, the atom-based approach of eq 14 resulted in a better performance and is thus used in the final model.

In congruence with the equivariant updating rule, higher-order multipoles, beginning with the atomic dipole $M_i^{(1)}$, are predicted as

$$M_i^{(1)} = \sum_{j \in N(i)} R_{ij}^{(1)} \phi_{M^{(1)}}(h_i^n, h_j^n, a_{ij}) \tag{15}$$

For the atomic quadrupole $M_i^{(2)}$, which is the highest-order multipole considered in this work, the following rule is used

$$M_i^{(2)} = \sum_{j \in N(i)} R_{ij}^{(2)} \phi_{M^{(2)}}(h_i^n, h_j^n, a_{ij}) \tag{16}$$

If $M_i^{(2)}$ is taken as the traceless quadrupole, $R_{ij}^{(2)}$ should be detraced in the same manner. In the present work, the detraced quadrupole $M_i^{(2)}$ and outer product $R_{ij}^{(2)}$ are defined as

$$\hat{D} A_{\alpha\beta}^{(2)} = A_{\alpha\beta} - \frac{1}{3} \sum_\chi A_{\chi\chi} \delta_{\alpha\beta} \tag{17}$$

following the notation in ref 44. Greek indices run over the three Cartesian dimensions, $\hat{D}$ refers to the detracing operator, and A stands for a quadrupole $M_i^{(2)}$ or outer product of a vector with itself, $R_{ij}^{(2)}$.

2.6. ESP Derived from Multipoles. Using the same notation, the ESP V_{ES} at a point R_j arising from point multipoles at R_i is given as

$$V_{ES}(R_{ij}) = \sum_{n=0}^\infty \frac{(2n-1)!!}{n!|R_{ij}|^{2n+1}} \langle M_i^{(n)}, R_{ij}^{(n)} \rangle \tag{18}$$

where both $M_i^{(n)}$ and $R_{ij}^{(n)}$ are traceless tensors and $\langle ... \rangle$ denotes the inner product contracted over n dimensions.[44] Specifically, contributions up to the quadrupole moment are calculated as

$$V_{M^{(0)}} = \frac{1}{N} \sum_j^N \sum_i \frac{M_i^{(0)}}{|R_{ij}|}$$

$$V_{M^{(1)}} = V_{M^{(0)}} + \frac{1}{N} \sum_j^N \sum_i \frac{M_{i,\alpha}^{(1)} R_{ij,\alpha}}{|R_{ij}|^3}$$

$$V_{M^{(2)}} = V_{M^{(1)}} + \frac{1}{N} \sum_j^N \sum_i \frac{M_{i,\alpha\beta}^{(2)}(3R_{ij,\alpha}R_{ij,\beta} - R^2\delta_{\alpha\beta})}{2|R_{ij}|^5} \tag{19}$$

with i iterating over multipole sites and j iterating over a set of points at which the ESP is probed. The averaged ESP obtained in this manner is used to validate the ESP derived from multipoles.

2.7. Loss Functions. We define the mean-squared error for the multipole moments of order k, which was used during training and validation, as

$$\mathcal{L}_2(M_{ref}^{(k)}, M_{pred}^{(k)}) = \frac{1}{AN} \sum_i^N \sum_\alpha^A (M_{i\alpha,ref}^{(k)} - M_{i\alpha,pred}^{(k)})^2 \tag{20}$$

with N running over every atom of the molecule and α over each unique component with $A = \frac{(k+1)(k+2)}{2}$. The mean absolute error (MAE) used for model validation is defined analogously by replacing the $\mathcal{L}_2$ with the $\mathcal{L}_1$ norm.

2.8. Data Set. A data set of multipoles was constructed for a set of molecules taken from three sources. First, neutral molecules with up to 20 heavy atoms consisting of elements $\in$ {H, C, N, O, F, S, Cl} were selected from the ChEMBL[54−56] database. To improve coverage of very small molecules, systems from GDB11[57] with up to eight heavy atoms and systems from a recently published data set of dimer dissociation curves[58] were added. Due to their uniqueness, molecules from the GDB11 and dimer data sets with up to four heavy atoms (135 in total) were exclusively assigned to the

training set. All the other molecules were randomly and exclusively assigned to the training, validation, or test set. The final training set consisted of 272 924 molecules, the validation set contained 32 755 molecules, and the test set contained 993 molecules.

The ETKDG conformer generator,[59] version 2, as implemented in the RDKit[60] was used to generate 3D conformations. For the molecules from the GDB11 and dimer data sets, up to four conformations were generated using an RMS pruning threshold of 0.1 Å. For the very small systems (up to four heavy atoms), up to 16 additional conformations were generated by randomly perturbing the generated conformation with Gaussian noise ($\sigma = 0.05$ Å). For all other compounds, up to two conformations were generated for the molecules in the training set and validation set and up to 16 for the test set using a RMS pruning threshold of 0.5 Å. A larger number of conformations was produced for the test set to evaluate conformational dependence, resulting in a total of 737 280 conformations for the training set, 117 312 for the validation set, and 13 344 for the test set.

In addition to the fixed split, a five-fold cross validation was performed by randomly selecting one conformation for each unique molecule resulting in a total of 307 200 data points. The data points were split into five folds. For each validation cycle, one fold was used for validation while the remaining four were used to train the model.

DFT single-point calculations were performed using the PSI4 (version 1.4) package[61,62] on a PBE0-D3BJ/def2-TZVP level of theory.[63−69] If not noted otherwise, default PSI4 settings were used. While not having any effect on the multipoles derived from single-point calculations, the D3BJ dispersion correction was included for future use of gradients and potential energies. Distributed multipoles up to the octupole were extracted for all converged DFT single-point calculations with MBIS[51] as implemented in the recent PSI4 version.[70] Calculations were performed on the Euler HPC cluster at ETH Zürich.

2.9. Model Parameters and Training. h_i^0 was initialized with a one-hot encoding of the respective element type while each edge a_{ij} contained the kernel-expanded distance between i and j using evenly spaced triweight kernels

$$f(x) = \max(0, k(m - x^2)^3) \tag{21}$$

with a shape factor k and the center m. Thirty-two evenly spaced kernels covering the range (0.5 Å, 4 Å) were used. The triweight kernel was chosen because its function value and first derivative take on 0 at the cutoff. Edges were added for all atom pairs within a cutoff radius of 4.0 Å. No strong dependence on the cutoff distance was observed when models were tested based on cutoffs between 3 and 5 Å (data not shown). Molecular topology-based graphs were constructed by adding an edge for each chemical bond. Node features include the one-hot-encoded element type as well as the scalar-encoded hybridization state. Edges features include the two one-hot-encoded element types of the bonded atoms and four scalars encoding the bond type, aromaticity, conjugacy, and ring membership. RDKit was used to obtain this information.[60] All other parameters were kept consistent with geometry-based graphs. Node- and edge-feature dimensions were set to 128. Each GNN layer consisted of two fully connected feed-forward layers with 128 units each ([128, Mila, 128, Mila]) combined with the Mila nonlinearity with $\beta = -1$.[71] The same architecture was used for the output layers ϕ_x ([128, Mila,

128, Mila, 1, Linear]). Layer weights were initialized with the method introduced by He et al.[72] The graph was updated over four message-passing iterations. Further, a residual architecture was used for atom feature updates:[73]

$$h_i^{l+1} = h_i^l + \phi_h(h_i^l, m_i) \tag{22}$$

The model was trained over 2048 epochs, presenting 2048 randomly drawn batches of size 64 during each epoch. Weights were updated using the ADAM optimizer[74] with an exponentially decaying learning rate $[5 \times 10^{-4}, 10^{-5}]$, minimizing the mean-squared error between reference multipole components and predicted multipole components as defined in eq 20. Models for monopoles, dipoles, and quadrupoles were optimized independently. Shared weight models were found to provide slightly worse results at lower computational costs. For the fivefold cross validation, a model was randomly initialized for each fold and trained over 512 epochs with 2048 randomly drawn batches of size 16. All other parameters remained unchanged. Batch size and number of epochs were decreased to accelerate the training, resulting in a slightly lower accuracy for the cross validation models compared to the main model.

The models were implemented with TensorFlow (2.6.2)[75,76] and the GraphNets library (1.1.0)[40] using the InteractionNetwork model.[41] General code was written with Python (3.9.7)[77] and Numpy (1.19.5).[78] Plots and visualizations were made with Open3D (0.13.0),[79] Matplotlib (3.4.2),[80] Seaborn (0.11.0),[81] and VMD (1.9.3).[82]

3. RESULTS AND DISCUSSION

The results are structured in two parts: First, the prediction accuracy of the ML model is assessed for the explicitly trained quantity, i.e., the multipole components. Second, properties derived from predicted multipoles are used to validate the model. In particular, the ESP reconstructed from predicted multipoles is compared with the reference ESP obtained from DFT or reconstructed from reference multipoles.

3.1. Prediction Accuracy on Multipoles. Table 1 lists the mean absolute errors (MAEs) of the predicted multipoles

Table 1. Mean Absolute Errors (MAEs) for the Predicted Multipoles of the Training, Validation, and Test Sets[a]

MAE $M^{(0)}$ (me)			MAE $M^{(1)}$ (me·Å)			MAE $M^{(2)}$ (me·Å²)		
train	val	test	train	val	test	train	val	test
2.05	2.20	2.19	0.64	0.65	0.64	0.63	0.64	0.63

[a]Monopoles in 10^{-3} elementary charge (me), dipoles in (me·Å), and quadrupoles in (me·Å²).

with respect to the reference multipoles for the specified subsets. Errors are given in 10^{-3} elementary charge. The results indicate that the proposed model is able to predict multipoles up to the quadrupole with high accuracy. While a direct comparison is difficult due to different data set sizes and employed reference methods, there exists, to our knowledge, currently no model in the literature which is able to predict atomic multipoles up to the quadrupole with better or comparable accuracy.

When separated into element types, it is evident that the errors are not distributed equally for each element (Figure 1). While the frequency of each element in the respective data set (Table S2 in the Supporting Information) plays a role, the

largest factor seems to be the range of the respective reference multipole components. This effect can be seen most clearly for sulfur whose monopoles cover a range of more than two elementary charges (Figures S1−S3 in the Supporting Information) and for which the observed error is largest. On the other hand, the monopoles of fluorine atoms cover only a fraction of this range resulting in the smallest observed error. In addition, molecule size is a contributing factor. Particularly for very small molecules, which are unique by nature and less abundant in the data set, larger error fluctuations can be observed (Figures S4−S10 in the Supporting Information).

The robustness of the model performance and its ability to generalize were assessed with a five-fold cross validation. For this purpose, one conformation was randomly sampled for each unique molecule such that each molecule was exclusively present in the training or the validation set but not both. These selected data points were randomly assigned to one of five equal sized tranches resulting in 61 440 data points for each tranche and a total of 307 200 samples. For each fold, one tranche was used as the validation data set while the other four were used to train the model. The average MAE and standard deviation over the five folds is listed in Table 2. The small standard deviations observed indicate a robust performance of the model. Note that models were trained less extensively for the cross validation compared to the main model, resulting in slightly larger errors.

3.2. Accuracy of the Electrostatic Potential. To validate the fidelity with which multipoles reproduce the ESP, electron densities from PBE0-D3BJ/def2-TZVP calculations were used to obtain the ESP (ESP$_{ref}$) on a surface with a distance of two vdW radii around the molecule. For this case, up to 16 conformations were generated for each molecule in the test set. On average, each surface contained around 4000 points. For the calculation of the ESP from multipoles, a probe charge of 1e was placed on each point on the vdW surface.

Table 3 lists the MAEs with respect to the reference potential ESP$_{ref}$ for systems in the test set. The errors are averaged over 13 344 conformations of 997 unique molecules from the test set. Three molecules had to be excluded due to convergence issues. In addition to the analytically derived reference multipoles, which serve as the baseline, predicted multipoles and partial charges are compared. Further, truncations after monopoles, dipoles, and quadrupoles ($M^{(0)}$, $M^{(1)}$, and $M^{(2)}$) are provided. For comparison, we also provide errors obtained from a model which uses molecular graphs without distance information as input. These entries are labeled with the subscript "topo" in Table 3. The prediction accuracy of this model is provided in Table S1 in the Supporting Information. In order to better understand the error composition, the error obtained by only using predicted multipoles for the highest considered moment is listed using a Δ symbol; i.e., $\Delta M^{(1)}$ denotes the error of the ESP based on reference monopoles and predicted dipoles, while $\Delta M^{(2)}$ denotes the error obtained by using reference monopoles and dipoles in combination with predicted quadrupoles.

When using the analytically derived reference multipoles, a clear reduction of the MAE can be observed when including dipoles ($M_{ref}^{(1)}$) and quadrupoles ($M_{ref}^{(2)}$). The same trend is seen for the predicted multipoles. The results in Table 3 show an accuracy hierarchy and accumulation of the error when including higher multipoles. In absolute terms, ignoring the conformational dependence of monopoles (entries labeled with "topo") results clearly in the largest error, followed by an
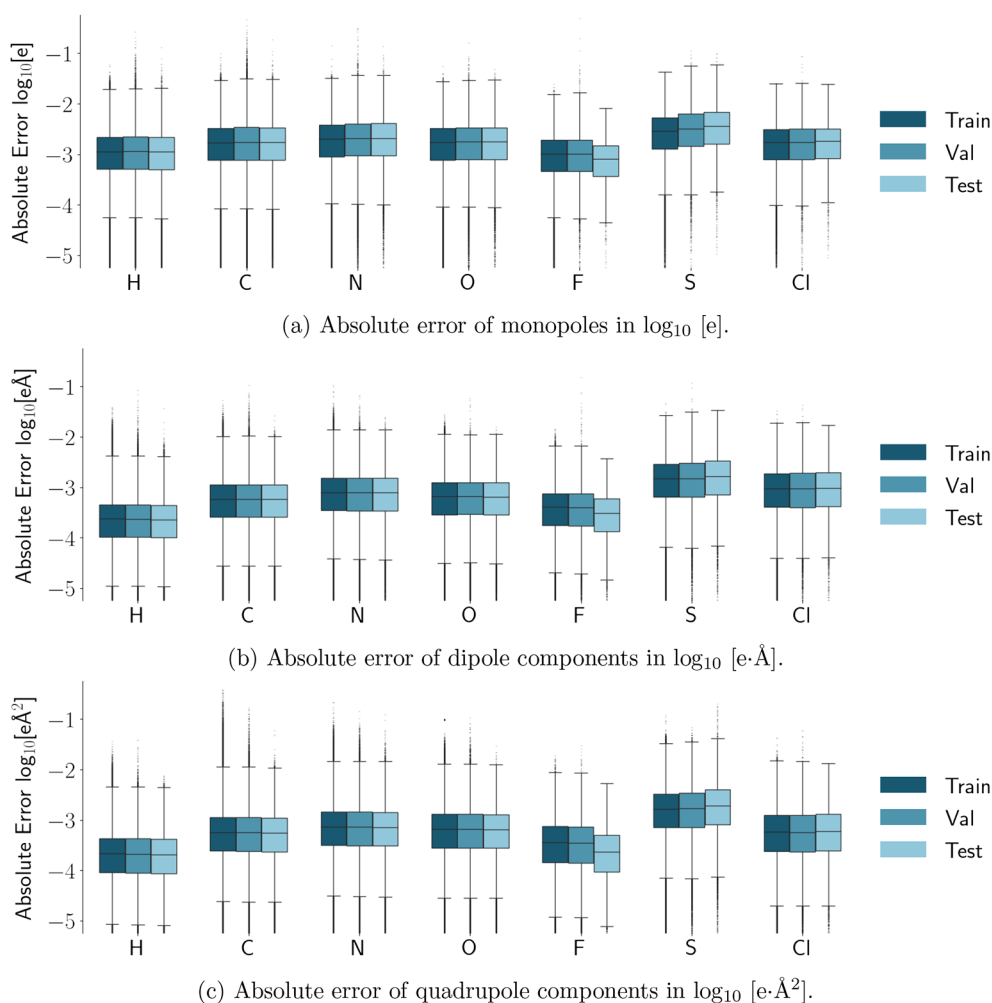
(a) Absolute error of monopoles in $\log_{10}$ [e].

(b) Absolute error of dipole components in $\log_{10}$ [e·Å].

(c) Absolute error of quadrupole components in $\log_{10}$ [e·Å^2].

Figure 1. Distribution of the absolute error of the multipole components by element for the training, validation, and test sets.

Table 2. Mean Absolute Errors (MAEs) for the Predicted Multipoles from a Five-fold Cross Validation[a]

	MAE cross validation	
	train (four-fold)	val (one-fold)
$M^{(0)}$	2.684 ± 0.017	2.918 ± 0.014
$M^{(1)}$	0.848 ± 0.005	0.863 ± 0.003
$M^{(2)}$	0.826 ± 0.008	0.845 ± 0.008

[a]Mean and standard deviation are taken over the MAE of the five folds. The data set consisted of all data points in the training and validation sets from the fixed split: monopoles are given in 10^{-3} elementary charge (me), dipoles in (me·Å), and quadrupoles in (me·Å^2).

early truncation of the multipole series. Replacing analytically derived multipoles with ML-predicted multipoles introduces an error of 0.15, 0.21, and 0.53 kJ ·mol^{-1}, respectively. $\Delta M^{(1)}$ and $\Delta M^{(2)}$ indicate that the error introduced by each predicted component itself is comparatively small; however, errors accumulate with each term of the multipole series. In cases where the ESP is approximated using the interaction between more than one multipole source, the error might be amplified additionally since each moment interacts with all other moments. This is not the case in the validation presented here where each multipole source interacts only with a monopole probe charge. Nevertheless, the results also clearly indicate that the error incurred by the ML model is in all cases smaller than the error caused by a premature truncation. These

Table 3. Mean Absolute Errors (MAEs) of the ESP on the vdW Surface in kJ ·mol^{-1} Based on Reference and Predicted Multipoles with Respect to the DFT ESP[a]

						MAE V_{ES} (kJ ·mol^{-1})					
$M^{(0)}_{ref}$	$M^{(1)}_{ref}$	$M^{(2)}_{ref}$	$M^{(0)}_{pred}$	$M^{(1)}_{pred}$	$M^{(2)}_{pred}$	$\Delta M^{(1)}$	$\Delta M^{(2)}$	$M^{(0)}_{topo}$	$M^{(1)}_{topo}$	$M^{(2)}_{topo}$
4.23	2.74	0.77	4.38	2.95	1.30	2.76	0.80	6.58	6.03	5.32

[a]Absolute errors are averaged over 13 344 data points from the test set. The order k of $M^{(k)}$ refers to the highest multipole term considered. Subscript "pred" refers to predicted multipoles, while "ref" denotes the reference multipoles based on the DFT density. For the entries labeled with Δ, only the highest moment was predicted by the model while lower multipoles were analytically derived. Entries denoted by "topo" were obtained from a model which uses only molecular graphs as input without distance information. The vdW surface is two times the vdW radii.

results suggest that the use of ML-predicted multipoles up to quadrupoles in the treatment of electrostatic interactions is preferable over calculations with only analytically derived monopoles (or dipoles).

To visually illustrate the results, the DFT ESP and predicted ESP on the vdW surface are compared for an example molecule (5-((4-fluorobenzyl)oxy)benzofuran-3(2H)-one) randomly selected from the test set (Figure 2). As can be
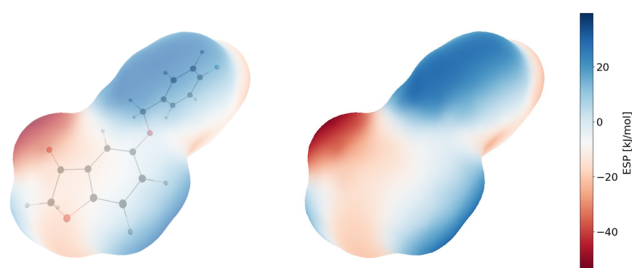

Figure 2. DFT ESP (left, $V_{\mathrm{ES,DFT}}$) and ESP derived from predicted multipoles (right, $V_{\mathrm{ES},M_{\mathrm{pred}}^{(2)}}$) using the same scale with different transparencies.
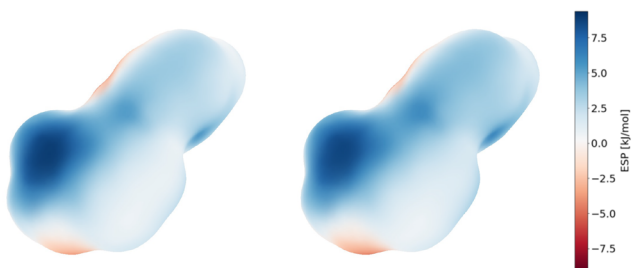

Figure 3. Delta ESP of the ESP based on analytically derived reference monopoles (left, $V_{\mathrm{ES,DFT}} - V_{\mathrm{ES},M_{\mathrm{ref}}^{(0)}}$) and predicted monopoles (right, $V_{\mathrm{ES,DFT}} - V_{\mathrm{ES},M_{\mathrm{pred}}^{(0)}}$) with respect to the DFT ESP.

seen in Figure 3, monopoles are insufficient to reproduce the more intricate features of the ESP. Including dipoles and quadrupoles reduces the error substantially (Figure 4). While it is only one illustrative case, this example shows that even though predicted multipoles result in a slightly less accurate ESP compared to analytically derived multipoles, predicted
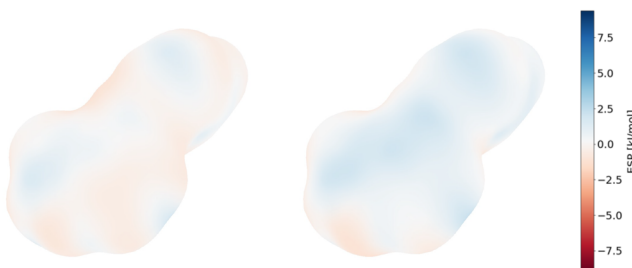

Figure 4. Delta ESP of the ESP based on analytically derived reference multipoles up to the quadrupole (left, $V_{\mathrm{ES,DFT}} - V_{\mathrm{ES},M_{\mathrm{ref}}^{(2)}}$) and predicted multipoles up to the quadrupole (right, $V_{\mathrm{ES,DFT}} - V_{\mathrm{ES},M_{\mathrm{pred}}^{(2)}}$) with respect to the DFT ESP. The same scale as in Figure 3 is used.

multipoles correctly reproduce qualitative features of the ESP and improve the accuracy (Figure 4).

For derived properties, such as the ESP or the molecular dipole, error cancellation when including higher-order terms is desirable. For specific future applications, this might be achieved by introducing additional loss terms. Particularly, a loss term for the molecular dipole could be introduced based on the error between a reference molecular dipole and a predicted molecular dipole constructed from predicted atomic monopoles and predicted atomic dipoles. In addition, a term for the error between the ESP based on predicted multipoles and reference multipoles could be added by randomly sampling the ESP around a given molecule.

3.3. Timing. Model speed was tested on a NVIDIA Titan V using the same batch with 50 064 atoms for all models (Table 4). In all cases, single precision was used. Single precision did

Table 4. Time Required to Predict Multipoles for a Single Atom Given a Preconstructed Graph[a]

	execution time (prediction) (μs/atom)	
n	main model	shared layers
3	6.53 $\pm$ 0.04	3.26 $\pm$ 0.08
4	8.21 $\pm$ 0.05	3.86 $\pm$ 0.08

[a]Numbers in brackets refer to the cutoff radii used for the graph construction (r_c). n refers to the number of graph-updating steps used.

not lower the model accuracy but resulted in a twofold speed-up on a GPU compared to double precision. The execution time was averaged over 100 runs with four repeats each and is given in μs/atom. To explore the dependence on model complexity and graph size, we also included a model that uses shared layers (for the embedding step and shared layers for the dipole and quadrupole coefficients, i.e., one module with two outputs). The same model parameters and model architecture were used for all models, consistent with the main model used for the results reported in the previous section. The time needed to construct the graph is not included as it can be sped up drastically in repeated calculations. Without specifically taking advantage of redundant calculations, the construction of a graph takes around 0.256 μs/atom for a molecule with 25 atoms. At 0.075 and 0.026 μs/atom, the construction of the edge features as well as distance calculations and pair extraction are the only steps, which cannot be circumvented in repeated calculations. In addition, execution speed can be improved further by reducing the number of graph-updating steps n or the cutoff radius r_c without strongly hampering prediction accuracy.

4. CONCLUSION

In the present work, we have introduced an equivariant GNN for the prediction of atomic multipoles. The proposed model is able to reproduce the ESP on a hybrid-DFT level of theory with an accuracy close to that of analytically derived multipoles. The uncertainty introduced by the ML prediction is consistently smaller than the error caused by the omission of higher-order terms (i.e., dipoles, quadrupoles). Thus, compared to calculations involving only analytically derived monopoles, predicted multipoles enable a much more precise treatment of the electrostatic interaction at a fraction of the cost of QM reference calculations. The equivariant architecture used in the GNN is not only able to predict multipole components with high precision but also strictly enforces the

proper behavior under rotations of the system without relying on local reference frames. Further, the sparse representation, which constructs atomic multipoles as a weighted combination of vectors and tensor products of vectors to an atom's neighbors, performs well for unseen structures in the test set.

Multipoles predicted by the proposed model may be used to describe the static multipoles in polarizable FFs, for the analysis of ESP surfaces, or for the treatment of long-range interactions in ML models. For future work, handling of charged species, extended coverage of the chemical and conformational space, as well as including an explicit treatment of polarization from the environment, which is important for condensed-phase systems, could be of value. In addition, directional message passing or various other atomic environment descriptors may further improve accuracy.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c01021.

> Assessment of the conformational dependence of multipoles (Table S1), distributions of reference multipoles (Figures S1−S3 and Table S2), and error dependence on molecule size (Figures S4−S10) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Sereina Riniker − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland;* Ⓞ orcid.org/0000-0003-1893-4031; Email: sriniker@ethz.ch

Authors

Moritz Thürlemann − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland*

Lennard Böselt − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.1c01021

Notes

The authors declare no competing financial interest.

The full data set covering 311 781 unique molecules and 1 013 949 conformations on a PBE0-D3BJ/def2-TZVP level of theory, including potential energies, gradients, MBIS multipoles up to the octupole, as well as ESP surfaces for the test set, is publicly available on the ETH Research Collection (https://www.research-collection.ethz.ch/handle/20.500.11\,850/509\,052). A repository with the ML model used to generate the results in this work can be found at https://github.com/rinikerlab/EquivariantMultipoleGNN.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 2013.

(2) Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *J. Chem. Inf. Model.* **2018**, *58*, 565−578.

(3) Kramer, C.; Spinn, A.; Liedl, K. R. Charge Anisotropy: Where Atomic Multipoles Matter Most. *J. Chem. Theory Comput.* **2014**, *10*, 4488−4496.

(4) Day, G. M.; Motherwell, W. D. S.; Jones, W. Beyond the Isotropic Atom Model in Crystal Structure Prediction of Rigid Molecules: Atomic Multipoles versus Point Charges. *Cryst. Growth Des.* **2005**, *5*, 1023−1033.

(5) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(6) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures Using Distributed Multipole and Polarizability-based Model Intermolecular Potentials. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478−8490.

(7) Halgren, T. A.; Damm, W. Polarizable Force Fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.

(8) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678−3693.

(9) Klicpera, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. *arXiv* **2020**, arXiv:2003.03123.

(10) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3*, No. e170816.

(11) Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2011**, *134*, 074106.

(12) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3*, No. e1603015.

(13) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *arXiv* **2020**, arXiv:2010.07067.

(14) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. *arXiv* **2017**, arXiv:1706.08566.

(15) Christensen, A. S.; Bratholm, L. A.; Faber, F. A.; von Lilienfeld, O. A. FCHL Revisited: Faster and More Accurate Quantum Machine Learning. *J. Chem. Phys.* **2020**, *152*, 044107.

(16) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *arXiv* **2017**, arXiv:1704.01212.

(17) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192−3203.

(18) Böselt, L.; Thürlemann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641−2658.

(19) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations. *J. Chem. Inf. Model.* **2018**, *58*, 579−590.

(20) Metcalf, D. P.; Jiang, A.; Spronk, S. A.; Cheney, D. L.; Sherrill, C. D. Electron-Passing Neural Networks for Atomic Charge Prediction in Systems with Arbitrary Molecular Charge. *J. Chem. Inf. Model.* **2021**, *61*, 115−122.

(21) Sifain, A. E.; Lubbers, N.; Nebgen, B. T.; Smith, J. S.; Lokhov, A. Y.; Isayev, O.; Roitberg, A. E.; Barros, K.; Tretiak, S. Discovering a Transferable Charge Assignment Model Using Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 4495−4501.

(22) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. General-Purpose Machine Learning Potentials Capturing Nonlocal Charge Transfer. *Acc. Chem. Res.* **2021**, *54*, 808−817.

(23) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A Fourth-Generation High-Dimensional Neural Network Potential with Accurate Electrostatics Including Non-Local Charge Transfer. *Nat. Commun.* **2021**, *12*, 398.

(24) Gastegger, M.; Behler, J.; Marquetand, P. Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra. *Chem. Sci.* **2017**, *8*, 6924−6935.

(25) Veit, M.; Wilkins, D. M.; Yang, Y.; DiStasio, R. A.; Ceriotti, M. Predicting Molecular Dipole Moments by Combining Atomic Partial Charges and Atomic Dipoles. *J. Chem. Phys.* **2020**, *153*, 024113.

(26) Bereau, T.; Andrienko, D.; von Lilienfeld, O. A. Transferable Atomic Multipole Machine Learning Models for Small Organic Molecules. *J. Chem. Theory Comput.* **2015**, *11*, 3225−3233.

(27) Handley, C. M.; Hawe, G. I.; Kell, D. B.; Popelier, P. L. A. Optimal Construction of a Fast and Accurate Polarisable Water Potential Based on Multipole Moments Trained by Machine Learning. *Phys. Chem. Chem. Phys.* **2009**, *11*, 6365−6376.

(28) Kramer, C.; Bereau, T.; Spinn, A.; Liedl, K. R.; Gedeck, P.; Meuwly, M. Deriving Static Atomic Multipoles from the Electrostatic Potential. *J. Chem. Inf. Model.* **2013**, *53*, 3410−3417.

(29) Glick, Z. L.; Koutsoukas, A.; Cheney, D. L.; Sherrill, C. D. Cartesian Message Passing Neural Networks for Directional Properties: Fast and Transferable Atomic Multipoles. *J. Chem. Phys.* **2021**, *154*, 224103.

(30) Thomas, N.; Smidt, T.; Kearnes, S. M.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds. *arXiv* **2018**, arXiv:1802.08219.

(31) Fuchs, F. B.; Worrall, D. E.; Fischer, V.; Welling, M. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. *arXiv* **2020**, arXiv:2006.10503.

(32) Esteves, C. Theoretical Aspects of Group Equivariant Neural Networks. *arXiv* **2020**, arXiv:2004.05154.

(33) Finzi, M.; Stanton, S.; Izmailov, P.; Wilson, A. G. Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data. *arXiv* **2020**, arXiv:2002.12880.

(34) Satorras, V. G.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. *arXiv* **2021**, arXiv:2102.09844.

(35) Anderson, B.; Hy, T.-S.; Kondor, R. Cormorant: Covariant Molecular Neural Networks. *arXiv* **2019**, arXiv:1906.04015.

(36) Cohen, T. S.; Welling, M. Group Equivariant Convolutional Networks. *arXiv* **2016**, arXiv:1602.07576.

(37) Batzner, S.; Smidt, T. E.; Sun, L.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Kozinsky, B. SE(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *arXiv* **2021**, arXiv:2101.03164.

(38) Miller, B. K.; Geiger, M.; Smidt, T. E.; Noé, F. Relevance of Rotationally Equivariant Convolutions for Predicting Molecular Properties. *arXiv* **2020**, arXiv:2008.08461.

(39) Schütt, K. T.; Unke, O. T.; Gastegger, M. Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra. *arXiv* **2021**, arXiv:2102.03150.

(40) Battaglia, P. W.; Hamrick, J. B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V. F.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R.; Gülçehre, Ç.; Song, H. F.; Ballard, A. J.; Gilmer, J.; Dahl, G. E.; Vaswani, A.; Allen, K. R.; Nash, C.; Langston, V.; Dyer, C.; Heess, N.; Wierstra, D.; Kohli, P.; Botvinick, M.; Vinyals, O.; Li, Y.; Pascanu, R. Relational Inductive Biases, Deep Learning, and Graph Networks. *arXiv* **2018**, arXiv:1806.01261.

(41) Battaglia, P. W.; Pascanu, R.; Lai, M.; Rezende, D. J.; Kavukcuoglu, K. Interaction Networks for Learning about Objects, Relations and Physics. *arXiv* **2016**, arXiv:1612.00222.

(42) Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *arXiv* **2015**, arXiv:1509.09292.

(43) Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chem. Mater.* **2019**, *31*, 3564−3572.

(44) Burnham, C. J.; English, N. J. A New Relatively Simple Approach to Multipole Interactions in Either Spherical Harmonics or Cartesians, Suitable for Implementation into Ewald Sums. *Int. J. Mol. Sci.* **2020**, *21*, 277.

(45) Stone, A. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chem. Phys. Lett.* **1981**, *83*, 233−239.

(46) Whitehead, C. E.; Breneman, C. M.; Sukumar, N.; Ryan, M. D. Transferable Atom Equivalent Multicentered Multipole Expansion Method. *J. Comput. Chem.* **2003**, *24*, 512−529.

(47) Elking, D. M.; Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A.; Pedersen, L. G. Gaussian Multipole Model (GMM). *J. Chem. Theory Comput.* **2010**, *6*, 190−202.

(48) Williams, D. E. Representation of the Molecular Electrostatic Potential by Atomic Multipole and Bond Dipole Models. *J. Comput. Chem.* **1988**, *9*, 745−763.

(49) Lillestolen, T. C.; Wheatley, R. J. Redefining the Atom: Atomic Charge Densities Produced by an Iterative Stockholder Approach. *Chem. Commun.* **2008**, 5909−5911.

(50) Lillestolen, T. C.; Wheatley, R. J. Atomic Charge Densities Generated Using an Iterative Stockholder Procedure. *J. Chem. Phys.* **2009**, *131*, 144101.

(51) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *J. Chem. Theory Comput.* **2016**, *12*, 3894−3912.

(52) Kullback, S.; Leiber, R. A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79−86.

(53) Zee, A. *Group Theory in a Nutshell for Physicists*; Princeton University Press: Princeton, NJ, 2016; Vol. 17.

(54) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.

(55) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*, D1083−D1090.

(56) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magarinos, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R. The ChEMBL Database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945−D954.

(57) Nguyen, K. T.; Blum, L. C.; van Deursen, R.; Reymond, J.-L. Classification of Organic Molecules by Molecular Quantum Numbers. *ChemMedChem.* **2009**, *4*, 1803−1805.

(58) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K.-H.; Gregersen, B. A.; Li, J.-L.; Palmo, K.; Siva, K.; Bergdorf, M.; Klepeis, J. L.; Shaw, D. E. Quantum Chemical Benchmark Databases of Gold-Standard Dimer Interaction Energies. *Sci. Data* **2021**, *8*, 55.

(59) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562−2574.

(60) Landrum, G.; Tosco, P.; Kelley, B.; Sriniker; Gedeck; Vianello, R.; Ric; NadineSchneider; Dalke, A.; Kawashima, E.; Cole, B.; Turk, S.; Swain, M.; Savelyev, A.; Cosgrove, D.; Vaucher, A.; Wójcikowski, M.; Probst, D.; Godin, G.; Jones, G.; Scalfani, V. F.; Pahl, A.; Berenger, F.; Varjo, J. L.; strets123, J. P.; DoliathGavid; Sforna, G.; Jensen, J. H. *rdkit/rdkit: 2020_09_5 (Q3 2020) Release*; 2021.

(61) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: An Open-Source Ab Initio Electronic Structure Program. *WIREs Comput. Mol. Sci.* **2012**, *2*, 556−565.

(62) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; Gonthier, J. F.; James, A. M.;

McAlexander, H. R.; Kumar, A.; Saitow, M.; Wang, X.; Pritchard, B. P.; Verma, P.; Schaefer, H. F.; Patkowski, K.; King, R. A.; Valeev, E. F.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185−3197.

(63) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew−Burke−Ernzerhof Exchange-Correlation Functional. *J. Chem. Phys.* **1999**, *110*, 5029−5036.

(64) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0Model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.

(65) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.

(66) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(67) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(68) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−B871.

(69) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(70) Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; James, A. M.; Lehtola, S.; Misiewicz, J. P.; Scheurer, M.; Shaw, R. A.; Schriber, J. B.; Xie, Y.; Glick, Z. L.; Sirianni, D. A.; O'Brien, J. S.; Waldrop, J. M.; Kumar, A.; Hohenstein, E. G.; Pritchard, B. P.; Brooks, B. R.; Schaefer, H. F.; Sokolov, A. Y.; Patkowski, K.; DePrince, A. E.; Bozkaya, U.; King, R. A.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. PSI4 1.4: Open-Source Software for High-Throughput Quantum Chemistry. *J. Chem. Phys.* **2020**, *152*, 184108.

(71) Misra, D. Mish: A Self Regularized Non-Monotonic Neural Activation Function. *arXiv* **2019**, arXiv:1908.08681.

(72) He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*; 2015.

(73) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.

(74) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.

(75) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*; 2015; https://www.tensorflow.org/.

(76) Developers, T. *TensorFlow*; 2021.

(77) Van Rossum, G.; Drake, F. L. *Python 3 Reference Manual*; 2009.

(78) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. *Nature* **2020**, *585*, 357−362.

(79) Zhou, Q.-Y.; Park, J.; Koltun, V. Open3D: A Modern Library for 3D Data Processing. *arXiv* **2018**, arXiv:1801.09847.

(80) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9*, 90−95.

(81) Waskom, M. L. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021.

(82) Humphrey, W.; Dalke, A.; Schulten, K. VMD − Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.