Equivariant message passing for the prediction of tensorial properties and molecular spectra

Kristof T. Schütt,1,2 Oliver T. Unke,1,2 and Michael Gastegger1,3

1) Machine Learning Group, Technische Universität Berlin, 10587 Berlin, Germany
2) Berlin Institute for the Foundations of Learning and Data, 10587 Berlin, Germany
3) BASLEARN – TU Berlin/BASF Joint Lab for Machine Learning, 10587 Berlin, Germany

(Dated: 8 June 2021)

Message passing neural networks have become a method of choice for learning on graphs, in particular the prediction of chemical properties and the acceleration of molecular dynamics studies. While they readily scale to large training data sets, previous approaches have proven to be less data efficient than kernel methods. We identify limitations of invariant representations as a major reason and extend the message passing formulation to rotationally equivariant representations. On this basis, we propose the polarizable atom interaction neural network (PaiNN) and improve on common molecule benchmarks over previous networks, while reducing model size and inference time. We leverage the equivariant atomwise representations obtained by PaiNN for the prediction of tensorial properties. Finally, we apply this to the simulation of molecular spectra, achieving speedups of 4-5 orders of magnitude compared to the electronic structure reference.

I. INTRODUCTION

Studying dynamics of chemical systems allows insight into processes such as reactions or the folding of proteins, and constitutes a fundamental challenge in computational chemistry. Since the motion of atoms is governed by the laws of quantum mechanics, accurate ab initio molecular dynamics (MD) simulations may require solving the Schrödinger equation for millions of time steps. While the exact solution is infeasible to compute for all but the smallest systems, even fast approximations such as density functional theory quickly become prohibitive for large systems and the prediction of accurate spectra.

Recently, machine learning potentials1–3 have gained popularity for studying systems ranging from small molecules at high levels of theory4,5 to systems with thousands or millions of atoms6–8. In particular, message-passing neural networks (MPNNs) yield accurate predictions for chemical properties across chemical compound space and can handle large amounts of training data. Albeit MPNNs have significantly increased in accuracy over the years (as well as in computational cost), kernel methods with manually crafted features9,10 have still proven to perform better when only small training sets are available.

While molecules are often represented as graphs, they are in fact interacting particles in a continuous 3d space. Consequently, SchNet11 modeled message passes as continuous-filter convolutions over that space, albeit with rotationally invariant filters. As Miller et al.12 pointed out, this leads to a loss of relevant directional, equivariant information. Klicpera, Groß, and Günne2222mann13 introduced directional message-passing, the angular information here is restricted to the messages while the representation of nodes (atoms) remains rotationally invariant. While equivariant convolutions have been successfully applied in computer vision14,15, previous approaches to molecular prediction16–18 have not reached the accuracy of their rotationally invariant counterparts. In this work, we propose rotationally equivariant message passing and the polarizable atom interaction neural network (PaiNN) architecture as one instance of it. We examine the limited capability of rotation-invariant representations to propagate directional information and show that equivariant representations do not suffer from this issue. PaiNN outperforms invariant message passing networks on common molecular benchmarks and performs at small sample sizes on par with kernel methods that have been deemed to be more data-efficient than neural networks. Beyond that, the rotationally equivariant representation of PaiNN enables the prediction of tensorial properties which we apply to the ring-polymer MD simulation of infrared and Raman spectra. By an acceleration of 4-5 orders of magnitude, PaiNN makes the simulation of these spectra feasible, reducing the runtime in one case from projected 25 years to one hour.

II. RELATED WORK

Behler and Parrinello21 introduced neural network potentials taking atom-centered symmetry functions based on distances and angles as features. Graph neural networks22–25 for molecular graphs26,27 and 3d geometries28 do not require such manually crafted features, but learn embeddings of atom types and use graph convolutions or, more general, message passing2 to model atom interactions based on interatomic distances. Klicpera, Groß, and Günne22 introduced directional message passing by including additional angular information in the message. Steerable CNN29 allow to build equivariant filter
banks according to known symmetries of associated feature types. While these approaches work on grids, architectures such as Tensor Field Network, Cormorant and NequiIP use equivariant convolutions based on spherical harmonics (SH) and Clebsch-Gordon (CG) transforms for point clouds. Kondor and Trivedi have described a general framework on equivariance and convolutions in neural networks focusing on irreducible representations. In contrast, PaiNN models equivariant interactions in Cartesian space which is conceptually simpler and does not require tensor contractions with CG coefficients. A similar approach was proposed by Jing et al. with the GVP-GNN. Both approaches are designed for distinct environments, neurons do not have to be scalar, but can be geometric objects such as vectors and tensors. For the purpose of this work, we restrict ourselves to scalar and vectorial representations \( s^t_i \) and \( \vec{v}^t_i \), respectively, such that a corresponding message pass can be written as

\[
\vec{m}^{t+1}_i = \sum_{j \in \mathcal{N}(i)} \vec{M}_i(s^t_i, s^t_j, \vec{r}_{ij})
\]

Message and update functions for scalar features can be defined analogously. Rotational equivariance of vector features \( \vec{v}^{t+1}_i \) can be ensured by employing rotationally equivariant functions \( \vec{U}_i \) and \( \vec{M}_i \), fulfilling

\[
R \vec{f}(\vec{x}) = \vec{f}(R \vec{x})
\]

for any rotation matrix \( R \in \mathbb{R}^{3\times3} \).

### B. Building equivariant MPNNs

To obtain more expressive representations of local environments, neurons do not have to be scalar, but can be geometric objects such as vectors and tensors. Therefore, any nonlinearities of the model need to be applied to scalar features. In particular, equivariant MPNNs may use the following operations:

- Any (nonlinear) function of scalars: \( f(s) \)
- Scaling of vectors: \( s \circ \vec{v} \)
- Linear combinations of equivariant vectors: \( \vec{W} \vec{v} \)
- Scalar products: \( s = \| \vec{v} \|^2 \), \( s = \langle \vec{v}_1, \vec{v}_2 \rangle \)
- Vector products: \( \vec{v}_1 \times \vec{v}_2 \)

Thus, directional information is preserved during message passing while invariant representations can be recovered from scalar products to obtain rotationally invariant predictions.

### C. Limits of rotationally invariant representations

In the following, we examine the expressiveness of invariant and equivariant representations at the example of simple 2d molecular structures. To ensure rotational invariance of a predicted property, the message function is often restricted to depend only on rotationally invariant inputs such as distances or angles. While a
single pass is limited to interactions within a local environment, successive message passes are supposed to construct more complex representations and propagate local information beyond the neighborhood. This raises the question whether rotationally invariant representations of atomic environments \( h_i \) are sufficient here.

Tab. I compares three simplified message functions for the example of a water molecule. Using a distance-based message function, the representation of the atom 1 (oxygen) is able to resolve changing bond lengths to atoms 2 and 3 (hydrogens), however it is not sensitive to changes of the bond angle. On the other hand, using the angle directly as part of the message function can not resolve the distances. Therefore, a combination of distances and angles is required to obtain a more expressive representation and design a message function that does not only propagate directions to neighboring atoms, but those of equivariant representations as well. This enables the efficient propagation of directional information by scaling linearly with the number of neighbors and keeping the required cutoff small.

Note that neither many-body representations, using angles, dihedral angles etc., nor equivariant representations corresponding to a multipole expansion are complete. It has been shown that even when including up to 4-body invariants, there are structures with as few as eight atoms which cannot be distinguished and that a many-body expansion up to \( n \)-body contributions is necessary to guarantee convergence for a system consisting of \( n \) atoms. The same holds for multipole expansions, where scalars and vectors correspond to the 0th and 1st order. Thus, even spherical harmonics expansions with less than infinite degree are unable to represent arbitrary equivariant \( n \)-body functions. Instead, a practically sufficient and computationally efficient approach for the problem at hand is desirable.

In the following, we propose a neural network architecture that takes advantage of these properties. It overcomes the limitations of invariant representations discussed above, which we will demonstrate at the example of an organometallic compound in Section V C 2.

IV. POLARIZABLE ATOM INTERACTION NEURAL NETWORK (PAINN)

The potential energy surface \( E(Z_1, \ldots, Z_N, \mathbf{r}_1, \ldots, \mathbf{r}_N) \), with nuclear charges \( Z_i \in \mathbb{N} \) and atom positions \( \mathbf{r}_i \in \mathbb{R}^3 \), exhibits certain symmetries. Generally, they include in-
variance of the energy towards the permutation of atom indices $i$, as well as rotations and translations of the molecule. A neural network potential should encode these constraints to ensure the symmetries of the predicted energy surface and increase data efficiency. A common inductive bias of the neural network potential is a decomposition of the energy into atomwise contributions

$$E = \sum_{i=1}^{N} \epsilon(s_i),$$

where an output network $\epsilon$ predicts energy contributions from atoms embedded within their chemical environment, represented by $s_i \in \mathbb{R}^{E \times 1}$. We keep the number of features $E$ constant throughout the network. The equivariant representations are set to $\vec{v}_i = \vec{0} \in \mathbb{R}^{E \times 3}$, since there is no directional information available initially.

The inputs to PaiNN are the nuclear charges $Z_i \in \mathbb{N}$ and positions $\vec{r}_i \in \mathbb{R}^3$ for each atom $i$. Similar to previous approaches, the invariant atom representations are initialized to learned embeddings of the atom type $s_i^0 = a_{Z_i} \in \mathbb{R}^{F \times 1}$. We keep the number of features $F$ constant throughout the network. The equivariant representations are set to $\vec{v}_i^0 = \vec{0} \in \mathbb{R}^{F \times 3}$, since there is no directional information available initially.

Next, we define message and update functions as introduced in Sec. (11). We use a residual structure introduced in Sec. III. We use a residual structure of inter-

The rotational-invariant filters $\mathcal{W}_s$ are linear combinations of radial basis functions $\sin(m \pi \| \vec{r}_{ij} \|)/\| \vec{r}_{ij} \|$ as proposed by Klicpera, Groß, and Günemann (15) with $1 \leq n \leq 20$. Additionally, we apply a cosine cutoff to the filters (19).

Analogously, we use continuous-filter convolutions for the residual of the equivariant message function

$$\Delta \bar{v}_i^m = \sum_j \vec{v}_j \circ \phi_{cv}(s_j) \circ \mathcal{W}_{cv}(|\vec{r}_{ij}|)$$

where the first term is a convolution of an invariant filter with scaled, equivariant features $\sum_j \vec{v}_j \circ \phi_{s}(s_j)$, related to the gating proposed by Weiler et al. (20) as an equivariant nonlinearity. This propagates directional information obtained in previous message passes to neighboring atoms. The second term is a convolution of invariant features with an equivariant filter. This can be derived as the gradient of an invariant filter

$$\nabla \mathcal{W}_{cv}(|\vec{r}_{ij}|) = \mathcal{W}_{cv}(|\vec{r}_{ij}|) \frac{\vec{r}_{ij}}{\| \vec{r}_{ij} \|^3}.$$
Since $W_{ij}^a(\|\vec{r}_{ij}\|)$ is another invariant filter, we can model it directly without taking the derivative. Furthermore, we use a shared network $\phi$ to perform the transform of all message functions and split the features afterwards (see Fig. 2b).

After the features-wise message blocks, the update blocks are applied atomwise across features. The residual of the scalar update function is given by

$$\Delta s_i^u = a_{ss}(s_i, \|\vec{V}_i\|) + a_{sv}(s_i, \|\vec{V}_i\|) (\vec{U}_i, \vec{V}_i).$$

Again, we use a scaling functions computed by a shared network $a(s, \|\vec{V}_i\|)$ as nonlinearity. In this case, the norm of a linear combination of features is also used to obtain the scaling, thus, coupling the scalar representations with contracted equivariant features. In a second term, we use the scalar product of two linear combinations of equivariant features. Similarly, we define the residual for the equivariant features

$$\Delta \vec{v}_i^u = a_{sv}(s_i, \|\vec{V}_i\|) \vec{U}_i,$$

which is again a nonlinear scaling of linearly combined equivariant features.

### B. Prediction of tensorial properties

PaINN is able to predict scalar properties using atomwise output layers of rotationally invariant representations $s_i$, in the same manner as previous approaches. Beyond that, the additional equivariant representations enable the prediction of equivariant tensors of order $M$ that can be constructed using a rank-1 tensor decomposition

$$T = \sum_{i=1}^{N} \sum_{k=1}^{R} \lambda(s_i) \vec{\nu}(\vec{v}_i)_{k,1} \otimes \cdots \otimes \vec{\nu}(\vec{v}_i)_{k,M}. \quad (11)$$

The scalar and vectorial components of this decomposition can be obtained from an output network of gated equivariant blocks, as shown in Fig. 3. These components again make use gated equivariant nonlinearities, and show similarities to geometric vector perceptrons (GVP)\(^{32}\). However, while GVP blocks are used as message functions, PaINN keeps the pairwise message functions comparatively lightweight. Instead complex transformations are restricted to the atomwise update function and tensor output network.

Depending on the chemical property of interest, one may want to include additional constraints or replace $\bar{r}_{k,m}$ with the atom positions to include the global geometry. In the following, we demonstrate how this can be employed for two tensorial properties that are important for the simulation of molecular spectra.

The molecular dipole moment $\vec{\mu}$ is the response of the molecular energy to an electric field $\nabla \vec{E}$ and, at the same time, the first moment of the electron density (see Eq. 6). It is often predicted using latent atomic charges $^{39}$

$$\vec{\mu} = \sum_{i=1}^{N} q_{\text{atom}}(s_i) \vec{r}_i. \quad (12)$$

This assumes the center of mass at $\vec{r} = \vec{0}$ for brevity. While this only leverages invariant representations, we can take advantage of the equivariant features of PaINN. In this picture, the molecular dipole is constructed from polarized atoms, i.e., both atomic dipoles and charges $^{41}$, yielding

$$\vec{\mu} = \sum_{i=1}^{N} \bar{\mu}_{\text{atom}}(\vec{v}_i) + q_{\text{atom}}(s_i) \vec{r}_i. \quad (13)$$

In terms of Eq. 11, this is a simple case with $M = 1, R = 2$, where the first term predicts the latent local dipoles and the second term computes the dipole generated by the local charges.

As an example with order $M = 2$, we consider the polarizability tensor, which describes the response of the molecular dipole to an electric field $J_{\vec{E}}(\vec{\mu}) = H_{\vec{E}}(\vec{E})$. We construct polarizability tensors using

$$\alpha = \sum_{i=1}^{N} \alpha_0(s_i) I_3 + \vec{\nu}(\vec{v}_i) \otimes \vec{r}_i + \vec{r}_i \otimes \vec{\nu}(\vec{v}_i), \quad (14)$$

where the first term of the sum models isotropic, atomwise polarizabilities. The other two terms add the anisotropic components of the polarizability tensor. Similar to the charges in Eq. 13, here the atom positions are used to incorporate the global structure of the molecule.

### V. RESULTS

PaINN has been implemented using PyTorch $^{42}$ and SCHNetPack$^{43}$. All models use $F = 128$ and two output layers as shown in Fig. 2 (blue) with the type of output layers depending on the property of interest. They were trained using the Adam optimizer $^{44}$ the squared loss and weight decay $\lambda = 0.01^2$, if not stated otherwise. We decay the learning rate by a factor of 0.5 if the validation loss plateaus. We apply exponential smoothing with factor 0.9
to the validation loss to reduce the impact of fluctuations which are particularly common when training with both energies and forces. Please refer to the supplement for further details on training parameters.

A. Chemical compound space

We use the QM9 dataset of \(\approx130k\) small organic molecules\(^{33}\) with up to nine heavy atoms to evaluate the performance of PaiNN for the prediction of scalar properties across chemical compound space. We predict the magnitude of the dipole moment using Eq. 13 and the electronic spatial extent by

\[
\langle R^2 \rangle = \sum_{i=1}^{N} q_{\text{atom}}(s_i) \| \vec{r}_i \|^2,
\]

as implemented by SCHNetPack. The remaining properties are predicted as sums over atomic contributions. PaiNN is trained on 110k examples while 10k molecules are used as a validation set for decaying the learning rate and early stopping. The remaining data is used as test set and results are averaged over three random splits. For the isotropic polarizability \(\alpha\), we first observed validation MAEs of 0.054 \(a_0\). Upon closer inspection, we notice that for this property both the squared loss as well as the MAE can be reduced when minimizing the MAE directly (as done by Klipper, Groß, and Günnemann\(^{33}\)). This yields both validation and test MAEs of 0.045\(a_0\) that are comparable to those of DimeNet++.

Tab. II shows the mean absolute error (MAE) of PaiNN for 12 target properties of QM9 in comparison with previous approaches. SCHNet\(^{33}\) and PhysNet\(^{27}\) are MPNNs with distance-based interactions, DimeNet++\(^{34}\) includes additional angular information and is an improved variant of DimeNet\(^{15}\). L1Net\(^{15}\) and Cormorant\(^{22}\) are equivariant neural networks based on spherical harmonics and Clebsch-Gordon coefficients.

PaiNN achieves state-of-the-art results in six target properties and yields comparable results to DimeNet++ on another two targets. On the remaining properties, PaiNN achieves the second best results after DimeNet++. Note that PaiNN using about 600k parameters is significantly smaller than DimeNet++ with about 1.8M parameters. For random batches of 50 molecules from QM9, the inference time is reduced from 45 ms to 13 ms, i.e. an improvement of more than 70\%, when comparing PaiNN to the reference implementation of DimeNet++\(^{26}\) using an NVIDIA V100.

B. Molecular dynamics trajectories

We evaluate the ability to predict combined energies and forces on the MD17 benchmark\(^{10}\) including molecular dynamics trajectories of small organic molecules. While the atomic forces could be predicted directly from vectorial features, we employ the gradients of the energy model \(\vec{F}_i = -\partial E/\partial \vec{r}_i\) to ensure conservation of energy. This property is crucial to run stable molecular dynamics simulations. To demonstrate the data efficiency of PaiNN, we use the more challenging setting with 1k known structures of which we use 950 for training and 50 for validation, where a separate model is trained for each trajectory. Tab. V shows the comparison with sGDMIL\(^{11}\) and NequiP\(^{19}\), which were trained on forces only, as well as SCHNet, PhysNet, DimeNet and FCHL19\(^{11}\) that were trained on a combined loss of energies and forces. Christensen and von Lilienfeld\(^{22}\) have found that the energies of MD17 are noisy. Thus, depending on the molecule and chosen tradeoff, using energies for training is not always beneficial. For this reason, we train two PaiNN models per trajectory: only on forces and on a combined loss including energy with a force error weight of \(\rho = 0.95\). PaiNN achieves the lowest mean absolute errors for 12 out 14 targets on models trained only on forces and exhibits errors in a similar range as Gaussian regression with the FCHL19 kernel. Overall, PaiNN performs best or equal to FCHL19 on 9 out of 14 targets. This demonstrates that equivariant neural networks approaches such as PaiNN are able to compete with kernel methods in the small data regime, while being able to scale to large data sets at the same time.

C. Advantages of equivariant features

1. Ablation studies

We evaluate the impact of equivariant vector features at the example of the aspirin MD trajectory from the previous section. Compared to the full model, we remove the scalar product of vector features in Eq. 9 from the update block and the convolution over vector features in Eq. 8 (i.e., \(W_{vv} = 0\)). Table IV show the results for the various ablations. The number of parameters is kept approximately constant by raising the number of node features \(F\) accordingly. We observe that all ablated components contribute to the final accuracy of the model, where the convolution over equivariant features in the message function has a slightly larger impact. This component also enables the propagation of directional information, which will be examined in Sec. V C 2. Finally, we remove all vector features from the model, resulting in an invariant model. Despite keeping the number of parameters constant by increasing the number of atoms features to \(F = 174\), the mean absolute error of the forces increases beyond 1 kcal/mol/Å.
TABLE II: Mean absolute errors on QM9 dataset for various chemical properties. Results for PaiNN are averaged over three random splits. Best in bold.

| Target Unit | SchNet | PhysNet | DimeNet++ | Cormorant | L1Net | PaiNN |
|-------------|--------|---------|-----------|-----------|-------|-------|
| μ           | 0.033  | 0.053  | 0.030     | 0.038     | 0.043 | 0.012 |
| α           | 0.235  | 0.062  | 0.444     | 0.085     | 0.088 | 0.045 |
| ε_{HOMO}    | 41     | 32.9   | 24.6      | 34        | 46.0  | 27.6  |
| ε_{LUMO}    | 34     | 24.7   | 19.5      | 38        | 34.6  | 20.4  |
| Δε          | 3     | 42.5   | 32.6      | 38        | 67.5  | 45.7  |
| ⟨R²⟩        | 0.073  | 0.765  | 0.331     | 0.961     | 0.354 | 0.066 |
| ZPVE        | 1.7    | 1.39   | 1.21      | 2.03      | 1.56  | 1.28  |
| U₀          | 14     | 8.15   | 6.32      | 22        | 13.46 | 5.85  |
| U            | 19     | 8.34   | 6.28      | 21        | 13.83 | 5.83  |
| H            | 14     | 8.42   | 6.53      | 21        | 14.36 | 5.98  |
| G            | 14     | 9.40   | 7.56      | 20        | 13.99 | 7.35  |
| ϵ_{HOMO}    |       |        |           |           |       |       |
| ϵ_{LUMO}    |       |        |           |           |       |       |
| Δε          |       |        |           |           |       |       |
| ⟨R²⟩        |       |        |           |           |       |       |
| ZPVE        |       |        |           |           |       |       |
| U₀          |       |        |           |           |       |       |
| U            |       |        |           |           |       |       |
| H            |       |        |           |           |       |       |
| G            |       |        |           |           |       |       |
| ϵ_{HOMO}    |       |        |           |           |       |       |
| ϵ_{LUMO}    |       |        |           |           |       |       |
| Δε          |       |        |           |           |       |       |
| ⟨R²⟩        |       |        |           |           |       |       |
| ZPVE        |       |        |           |           |       |       |
| U₀          |       |        |           |           |       |       |
| U            |       |        |           |           |       |       |
| H            |       |        |           |           |       |       |
| G            |       |        |           |           |       |       |

TABLE III: Mean absolute errors on MD17 dataset for energy and force predictions in kcal/mol and kcal/mol/Å, respectively. Batzner et al. only reported force errors for NequIP. Results for PaiNN are averaged over three random splits. Best in bold.

| Ablation | # params | F energy MAE | F force MAE | E energy MAE | E force MAE | F energy MAE | F force MAE |
|----------|----------|--------------|-------------|--------------|-------------|--------------|-------------|
| no ablation | 588.3k | 128 | 0.159 | 0.371 | 0.167 | 0.338 | 0.37 | 0.230 | 0.204 | 0.182 | 0.159 |
| no scalar product of vector features in Eq. 8 | 589.1k | 134 | 0.173 | 0.420 | 0.338 | 0.371 | 0.37 | 0.230 | 0.204 | 0.182 | 0.159 |
| no vector propagation (Wv_v = 0 in Eq. 7) | 589.2k | 135 | 0.183 | 0.441 | 0.338 | 0.371 | 0.37 | 0.230 | 0.204 | 0.182 | 0.159 |
| remove both | 590.1k | 142 | 0.200 | 0.507 | 0.338 | 0.371 | 0.37 | 0.230 | 0.204 | 0.182 | 0.159 |
| no vector features | 590.3k | 174 | 0.449 | 1.194 | 0.338 | 0.371 | 0.37 | 0.230 | 0.204 | 0.182 | 0.159 |

2. Propagation of directional information in substituted ferrocene

To demonstrate the advantages of equivariant over invariant representations in practice, we consider a ferrocene derivative where one hydrogen atom in each cyclopentadienyl ring is substituted by fluorine (see Fig. 4). This molecule has been chosen as it features small energy fluctuations (<1 kcal/mol) when the rings rotate relative to each other. Since the torsional energy profile depends mainly on the orientation of the distant fluorine atoms (measured by the rotation angle θ), it is a challenging prediction target for models without equivariant representations. Fig. 4 shows the predicted energy profiles for a full rotation of the cyclopentadienyl ring using cutoffs r_{cut} ∈ \{2.5, 3.0, 4.0\} Å. All models are trained on energies of 10k structures sampling thermal fluctuations (300 K) and ring rotations of substituted ferrocene. An-
other 1k structures are used for validation. SchNet\cite{Schutt2017}, which uses scalar features, predicts a flat energy profile for $r_{\text{cut}} \leq 3.0$ Å, because it is unable to resolve $\theta$. Although it is possible to encode the relevant information when using a larger cutoff, the learned representation seems to be susceptible to noise, as indicated by the deviations of the predicted energy profile. Since DimeNet\cite{DimeNet} includes terms that explicitly depend on angles between triplets of atoms, $\theta$ can be resolved for smaller cutoffs, but still fails for $r_{\text{cut}} = 2.5$ Å. In contrast, the equivariant environment representation of PaiNN allows to faithfully reproduce the torsional energy profile even for very small cutoffs.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Rotational energy profile of substituted ferrocene obtained by varying the rotation angle $\theta$ while keeping all bond distances fixed at their equilibrium values (H: white, C: black, F: purple, Fe: grey). For small cutoff $r_{\text{cut}}$ (dashed grey circles), MPNNs with scalar feature representations are unable to represent information about the rotation angle $\theta$.
}
\end{figure}

\subsection{D. Molecular Spectra}

We apply PaiNN to the efficient computation of infrared and Raman spectra of ethanol and aspirin. Although these spectra can in principle be obtained from a single molecular structure via the harmonic oscillator approximation, such an approach tends to neglect important effects, e.g. different molecular conformations.\cite{DimeNet} In order to obtain high quality spectra, molecular dynamics simulations have to be performed, requiring the prediction of forces at each time step. Additionally, one needs to compute dipole moments and polarizabilities along the generated trajectory. The infrared and Raman spectra are obtained as the Fourier transform of the time autocorrelation functions of the respective property. Using electronic structure methods, such simulations become prohibitively expensive due to the large number of successive computations required. Since nuclear quantum effects (NQE) need to be considered to obtain high quality spectra\cite{Bartlett2016}, ring-polymer molecular dynamics (RPMD) simulations need to be performed, which treat multiple copies of a molecule simultaneously and further increase computational cost.

We train a joined model for energies, forces, dipole moments and polarizability tensors on 8k ethanol conformations using additional 1k molecules each as validation and test set. A second model is trained for aspirin with 16k training examples and 2k molecules for validation and testing. Energies and forces are predicted as described above for MD\cite{Schutt2017}. The dipole moments and polarizability tensors are obtained as described in Section IV B employing two gated equivariant blocks each yielding atomwise scalars and vectors, that are used to compute the outputs as described in Eqs.\cite{DimeNet} and\cite{Schutt2017}. The joined model exhibits accurate predictions for energy, forces, dipole moments and polarizabilities, as shown in the inset table of Fig.\ref{fig:5}.

To evaluate the importance of equivariant features for the dipole prediction, we have trained an alternative model using only latent partial charges, but no local dipoles. The results show that predicting molecular dipoles using vector features (Eq.\cite{DimeNet}) reduces the mean abs. error by more than 50\%, reaching 0.009 D compared to 0.020 D using only scalars and positions (Eq.\cite{Schutt2017}). Simulated infrared and Raman spectra using classical MD and RPMD with 64 replicas and the respective PaiNN models are shown for both molecules in Fig.\ref{fig:5} alongside harmonic spectra obtained with the electronic structure reference and, for ethanol, experimental spectra recorded in the gas phase\cite{Sanz2005}.

The peak positions and intensities of ethanol infrared spectra (Fig.\ref{fig:5}a, top) computed with classical MD and PaiNN agree closely with the static electronic structure reference, e.g. in the C-H and O-H stretching regions at 3000 cm$^{-1}$ and 3900 cm$^{-1}$. This indicates, that PaiNN faithfully reproduces the original electronic structure method in general. However, when compared to experiment, both spectra are shifted towards higher frequencies. The PaiNN RPMD spectrum, on the other hand shows excellent agreement with experiment, illustrating the importance of NQEs. Similar observations can be made for the ethanol Raman spectrum (Fig.\ref{fig:5}a, bottom), where the PaiNN RMPD spectrum once again offers the most faithful reproduction of the experiment. We observe similar trends for aspirin (Fig.\ref{fig:5}b). RPMD spectra are once again shifted to lower frequencies while the classical MD closely resembles the static electronic structure spectrum.

PaiNN reduces the time required to obtain these spectra by several orders of magnitude. For ethanol, an individual reference computation takes $\approx 140$ seconds compared to 14 ms for a PaiNN evaluation on a V100 GPU. This corresponds to a reduction of the overall simulation time from 400 days to approximately one hour. The speedup offered by PaiNN is even more pronounced for aspirin, where a simulation that would have taken 25
FIG. 5: IR (top) and Raman (bottom) spectra of ethanol and aspirin. Spectra calculated with the reference method using the harmonic oscillator approximation are shown in black (QM harmonic). The inset table shows the mean absolute errors on the respective test set.

years (3140 seconds / step) now takes one hour (15 ms / step).

VI. CONCLUSIONS

We have given general guidelines to design equivariant MPNNs and discussed the advantages of equivariant representations over angular features in terms of computational efficiency as well as their ability to propagate directional information. On this basis, we have proposed PaiNN that yields fast and accurate predictions of scalar and tensorial molecular properties. Thereby, equivariant message passing allows us to significantly reduce both model size and inference time compared to directional message-passing while retaining accuracy. Finally, we have demonstrated that PaiNN can be applied to the prediction of tensorial properties, which we leverage to accelerate the simulation of molecular spectra by 4-5 orders of magnitude – from years to hours.

In future work, the equivariant representation of PaiNN as well as the ability to predict tensorial properties may be leveraged in generative models of 3d geometries or the prediction of wavefunctions. We see further applications of equivariant message passing in 3d shape recognition and graph embedding.

Many challenges remain for the fast and accurate prediction of molecular properties, e.g. modeling of enzymatic active sites or surface reactions. To describe such phenomena, highly accurate reference methods are required. Due to their computational cost, reference data generation can become a bottleneck, making data-efficient MPNNs such as PaiNN invaluable for future chemistry research.

ACKNOWLEDGEMENTS

KTS acknowledges support by the Federal Ministry of Education and Research (BMBF) for the Berlin Center for Machine Learning / BIFOLD (01IS18037A). MG works at the BASLEARN – TU Berlin/BASF Joint Lab for Machine Learning, co-financed by TU Berlin and BASF SE. OTU acknowledges funding from the Swiss National Science Foundation (Grant No. P2BSP2_188147).

1. J. Behler, “Perspective: Machine learning potentials for atomistic simulations,” J. Chem. Phys. 145, 170901 (2016).
2. O. T. Unke, S. Chmiela, H. E. Sauerca, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, “Machine Learning Force Fields,” arXiv preprint arXiv:2010.07067 (2020).
3. O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, “Exploring chemical compound space with quantum-based machine learning,” Nat. Rev. Chem. 4, 347–358 (2020).
4. S. Chmiela, H. E. Sauerca, K.-R. Müller, and A. Tkatchenko, “Towards exact molecular dynamics simulations with machine-learned force fields,” Nat. Commun. 9, 3887 (2018).
5. J. Westermayr, M. Gastegger, and P. Marquetand, “Combining SchNet and SHARC: The SchNarc machine learning approach for excited-state dynamics,” J. Phys. Chem. Lett. 11, 3828–3834 (2020).
6. T. Morawietz, A. Singraber, C. Dellago, and J. Behler, “How van der Waals interactions determine the unique properties of water,” Proc. Natl. Acad. Sci. 113, 8368–8373 (2016).
7. A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, “Machine learning a general-purpose interatomic potential for silicon,” Phys. Rev. X 8, 041048 (2018).
8. D. Lu, H. Wang, M. Chen, J. Liu, L. Liu, R. Caz, W. Jia, L. Zhang, et al., “86 PFLOPS Deep Potential Molecular Dynamics simulation of 100 million atoms with ab initio accuracy,” arXiv preprint arXiv:2004.11658 (2020).
9. J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, “Neural message passing for quantum chemistry,” International Conference on Machine Learning (2017).
"3D Steerable CNNs: Learning Rotationally Equivariant Features without the Electrons," Phys. Rev. Lett. 104, 136403 (2010).

K. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, “SchNet: A continuous-filter convolutional neural network for modeling quantum interactions,” Advances in Neural Information Processing Systems, 991–1001 (2017).

R. K. Müller, M. Geiger, T. E. Smidt, and F. Noé, “Relevance of rotationally equivariant convolutions for predicting molecular properties,” arXiv preprint arXiv:2008.08461 (2020).

J. Klicpera, J. Groß, and S. Günnemann, “Directional message passing for molecular graphs,” International Conference on Learning Representations (2020).

S. Cohen and M. Welling, “Steerable CNNs,” International Conference on Learning Representations (2017).

M. Weiler, F. A. Hamprecht, and M. Storath, “Learning Steerable Filters for Rotation-Equivariant CNNs,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 849–858 (2018).

D. Worrall and G. Brostow, “CubeNet: Equivarience to 3D Rotation and Translation,” Proceedings of the European Conference on Computer Vision (ECCV), 567–584 (2018).

N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, “Tensor field networks: Rotation-and-translation-equivariant neural networks for 3D point clouds,” arXiv preprint arXiv:1809.08219 (2018).

B. Anderson, T. S. Hy, and R. Kondor, “Cormorant: Covariant Molecular Neural Networks,” Advances in Neural Information Processing Systems 32, 14537–14546 (2019).

J. Behler and M. Parrinello, “Generalized neural-network representation of high-dimensional potential-energy surfaces,” Phys. Rev. Lett. 98, 146401 (2007).

F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, “The graph neural network model,” IEEE Trans. Neural Netw. 20, 61–80 (2008).

D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, “Convolutional networks on graphs for learning molecular fingerprints,” Advances in Neural Information Processing Systems 28, 2224–2232 (2015).

S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, “Molecular graph convolutions: Moving beyond fingerprints,” J. Comput. Aided Mol. Des. 30, 595–608 (2016).

K. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, “Quantum-chemical insights from deep tensor neural networks,” Nat. Commun. 8 (2017), 10.1038/ncomms13890.

K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, “SchNet—A deep learning architecture for molecules and materials,” J. Chem. Phys. 148, 241722 (2018).

Q. T. Unke and M. Meuwly, “PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges,” J. Chem. Theory Comput. 15, 3678–3690 (2019).

S. Lubbers, J. S. Smith, and K. Barros, “Hierarchical modeling of molecular energies using a deep neural network,” J. Chem. Phys. 148, 241715 (2018).

M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T. S. Cohen, “3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data,” Advances in Neural Information Processing Systems 31, 10381–10392 (2018).

S. Batzner, T. E. Smidt, L. Sun, J. P. Mailoa, M. Kornbluth, N. Molnari, and B. Kozinsky, “SE (3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials,” arXiv preprint arXiv:2101.03164 (2021).
J. Köhler, L. Klein, and F. Noé, “Equivariant Flows: sampling
configurations for multi-body systems with symmetric energies,”
Proceedings of the 37th International Conference on Machine
Learning (2019).

G. N. Simm, R. Pinsler, G. Csányi, and J. M. Hernández-Lobato,
“Symmetry-Aware Actor-Critic for 3D Molecular Design,” arXiv
preprint arXiv:2011.12747 (2020).

G. Hegde and R. C. Bowen, “Machine-learned approximations
to Density Functional Theory Hamiltonians,” Sci. Rep. 7, 1–11
(2017).

K. Schütz, M. Gasteiger, A. Tkatchenko, K.-R. Müller, and
R. J. Maurer, “Unifying machine learning and quantum chemistry
with a deep neural network for molecular wavefunctions,” Nat.
Commun. 10, 1–10 (2019).

J. Hermann, Z. Schätzle, and F. Noé, “Deep-neural-network so-
lution of the electronic Schrödinger equation,” Nat. Chem. 12,
891–897 (2020).

P. Goyal and E. Ferrara, “Graph embedding techniques, applica-
tions, and performance: A survey,” Knowledge-Based Systems
151, 78–94 (2018).

C. Bannwarth, S. Ehlert, and S. Grimme, “GFN2-xTB – An accu-
rate and broadly parametrized self-consistent tight-binding quan-
tum chemical method with multipole electrostatics and density-
dependent dispersion contributions,” J. Chem. Theory Comput.
15, 1652–1671 (2019).

https://github.com/grimme-lab/xtb

T. S. Smith, O. Isayev, and A. E. Roitberg, “ANI-1: an exten-
sible neural network potential with DFT accuracy at force field
computational cost,” Chem. Sci. 8, 3192–3203 (2017).

C. Adamo and V. Barone, “Toward reliable density functional
methods without adjustable parameters: The PBE0 model,” J.
Chem. Phys. 110, 6158–6170 (1999).

F. Neese, “The ORCA program system,” WIREs Comput. Mol
Sci. 2, 73–78 (2012).

F. Weigend, “A fully direct RI-HF algorithm: Implementation,
onoptised auxiliary basis sets, demonstration of accuracy and
efficiency,” Phys. Chem. Chem. Phys. 4, 4285–4291 (2002).

G. J. Martyna, M. L. Klein, and M. Tuckerman, “Nosé–Hoover
chains: The canonical ensemble via continuous dynamics,” J
Chem. Phys. 97, 2635–2643 (1992).

P. Goyal and E. Ferrara, “Graph embedding techniques, applica-
tions, and performance: A survey,” Knowledge-Based Systems
151, 78–94 (2018).

M. Ceriotti, M. Parrinello, T. E. Markland, and D. E. Manolopoulos,
“Efficient stochastic thermostatting of path integral molecular
dynamics,” J. Chem. Phys. 133, 124104 (2010).

N. Wiener, “Generalized harmonic analysis,” Acta Math. 55,
117–258 (1930).

R. B. Blackman and J. W. Tukey, “The measurement of power
spectra from the point of view of communications engineering –
Part I,” Bell Syst. Tech. J. 37, 185–282 (1958)
Appendix A: Data

1. Reference data for substituted ferrocene

Reference computations for substituted ferrocene were carried out with the semi-empirical GFN2-xTB method\cite{GFN2} using the xtb package. Training structures were generated by normal mode sampling at 300 K starting from the global minimum structure and randomly rotating the cyclopentadienyl (Cp) moieties relative to each other around the Cp-Fe-Cp axis by angles uniformly sampled from $[0, 2\pi]$. The rotational potential energy profile was generated by sampling a full rotation around the Cp-Fe-Cp axis using 1k steps, starting from the global minimum structure and keeping all other degrees of freedom fixed at their equilibrium positions.

2. Reference data for infrared and Raman spectra

Electronic structure reference computations for aspirin were carried out at the PBE0/def2-TZVP level of theory using the ORCA quantum chemistry package. SCF convergence was set to tight and integration grid levels of 4 and 5 were employed during SCF iterations and the final computation of properties, respectively. Computations were accelerated using the RIJK approximation. The reference data for aspirin was generated by selecting 20000 random configurations from the MD17 database and recomputing them at the above level of theory. Reference data for the ethanol molecule was taken from reference, which employed the same level of theory bar the RIJK approximation.

Appendix B: Training details

All models were trained using the Adam optimizer. The learning rate is decayed by a factor of 0.5 if the validation loss plateaus, starting with the largest learning rate that does not diverge in steps 1e-4, 5e-4, 1e-3, etc. We apply exponential smoothing with factor 0.9 to the validation loss to reduce the impact of fluctuations which are particularly common when training with both energies and forces. We use smaller batches for datasets where we train on both energies and forces, since it is commonly observed that larger batch sizes converge to larger errors in this setting.

Appendix C: Computation of infrared and Raman spectra

All simulations were carried out with the molecular dynamics module implemented in SchNetPack. Classical molecular dynamics simulations for ethanol and aspirin were carried out for 50 ps at a temperature of 300 K controlled via Nose-Hoover chain thermostat with a chain length of 3 and time constant of 100 fs. The first 10 ps of these trajectories were then discarded. Ring polymer molecular dynamics using 64 replicas were performed using the same simulation and equilibration periods and temperature, but employed a specially adapted global Nose-Hoover chain as introduced in Ref. for thermostatting instead. The overall chain settings were kept the same as above. In all cases, the velocity Verlet algorithm and a time step of 0.2 fs were used to integrate the equations of motion. Infrared and Raman spectra were computed from the time-autocorrelation functions of the dipole moment and polarizability time derivatives (see Ref). All autocorrelation functions were computed with the Wiener-Khinchin theorem using an autocorrelation depth of 2048 fs. A Hann window function and zero-padding were applied to the autocorrelation functions in order to enhance the spectra. Raman spectra were calculated using a laser frequency of 514 nm and temperature of 300 K.
Supplementary Table 5: Training parameters for all experiments.

| Data set | Batch size | Learning rate | Decay patience | Stopping patience | r_{cut} [Å] |
|----------|------------|---------------|----------------|-------------------|-------------|
| QM9      | 100        | 5 · 10^{-4}   | 5              | 30                | 5.0         |
| MD17     | 10         | 1 · 10^{-3}   | 50             | 150               | 5.0         |
| Ferrocene| 10         | 1 · 10^{-3}   | 10             | 30                | 2.5-4.0     |
| Spectra  | 10         | 5 · 10^{-4}   | 15             | 50                | 2.7         |