
A Hybrid Equivariant–Scalar Message Passing Framework for Geometric Molecular Graphs

Anna Katharina Picha

Department of Computational Biological Chemistry
Faculty of Chemistry
University of Vienna
Vienna 1090, Austria
anna.katharina.picha@univie.ac.at

Stefan Boresch

Department of Computational Biological Chemistry
Faculty of Chemistry
University of Vienna
Vienna 1090, Austria

Nils Morten Kriege

Research Group Data Mining and Machine Learning
Faculty of Computer Science
University of Vienna
Vienna 1090, Austria

Esther Heid

Institute for Materials Chemistry
Faculty of Technical Chemistry
Technical University of Vienna
Vienna 1040, Austria

Abstract

Geometric graph neural networks are widely applied to molecular systems, typically relying on local neighborhoods and a common message passing scheme across all distances. However, the distinct physical nature of short-range many-body effects and long-range behavior suggests using separate, distance-dependent representations. We propose a hybrid message passing framework combining equivariant local updates with scalar longer-range interactions. We evaluate the approach on molecular energy and force prediction tasks, analyzing the trade-off between accuracy and computational cost across cutoff configurations. Our results show that combining equivariant and scalar updates can achieve accuracy comparable to fully equivariant models while substantially reducing computational cost.

1 Introduction and Related Work

Neural networks, and in particular graph neural networks (GNNs), have become a central tool in computational chemistry and physics [1, 2, 3]. They are routinely applied to tasks ranging from conformer generation [4] and chemical property prediction [5, 6] to transition-state [7] and barrier-height prediction [8], as well as the construction of machine-learned interatomic potentials (MLIPs) for molecular dynamics (MD) simulations [9, 10]. Since molecular systems are embedded in three-dimensional space, these models often take the form of geometric GNNs, which represent atoms as nodes and their interactions as edges and incorporate spatial information such as interatomic distances and orientations [11]. Through iterative message passing, atom-wise representations are constructed and mapped to molecular properties such as total energies and atomic forces.

MLIPs learn atomic representations and aim to approximate potential energy surfaces that obey fundamental physical symmetries such as *translational*, *permutational*, and *rotational* invariance.

These properties ensure that the predicted energy depends only on the intrinsic geometry and composition of the system and not on arbitrary choices of coordinate system or atom ordering. Instead of restricting features to scalars, higher-order geometric features such as scalars, vectors, and higher-order tensors can be used to obtain more expressive atomic representations through equivariant representations [12, 13, 14, 15, 16], see also Section A in the Appendix. Recent MLIPs range from descriptor-based approaches such as ANI [17, 18] to geometric and equivariant message passing architectures including SchNet [19], PaiNN [20], DimeNet [21], MACE [22], and Equiformer [23, 24, 25].

MLIPs typically rely on a single distance cutoff to define the interaction graph, with interacting atom pairs subsequently processed by the same message passing scheme. While long-range effects such as electrostatics and dispersion can be incorporated through explicit corrections or auxiliary mechanisms [26, 27, 28, 29], a uniform treatment of interactions across distances remains common. This practice overlooks an important physical distinction between short- and long-range interactions. At short distances, interactions strongly rely on the specific spatial arrangement of atoms, requiring directional information such as angles and dihedrals. Such information can be captured through triplet-based representations [17, 21] or explicitly equivariant message passing [20, 22], but both introduce additional computational cost compared to purely distance-based scalar representations. In contrast, mid- and long-range interactions, including electrostatics and dispersion, can often be described effectively by pairwise, distance-dependent terms.

Several previous approaches have combined different representations or interaction mechanisms to describe molecular systems. For example, the FENNIX architecture [26] combines local equivariant representations with explicit electrostatic and dispersion energy terms, while SpookyNet [30] incorporates nonlocal interactions by using a global self-attention mechanism alongside local geometric representations. More recently, FeNNix-Bio1 [31] introduced a multi-resolution architecture combining short-range geometric representations with physics-inspired longer-range descriptors that account for relative orientations. These studies highlight the potential of combining different representations across interaction ranges. However, while scalar message passing offers computational advantages, it may not fully capture orientation-dependent effects in longer-range interactions.

In this context, our work investigates the accuracy–computational cost trade-off of combining short-range equivariant message passing with longer-range scalar message passing, without introducing explicit analytical electrostatic or dispersion energy terms. We propose a hybrid message passing scheme by using a hybrid geometric GNN that combines short-range equivariant message passing with computationally cheaper scalar interactions at longer distances. We restrict expressive equivariant message passing to a small local cutoff, while modeling mid-range interactions using scalar message passing based solely on interatomic distances, with a smooth transition between the two regimes. For this purpose, we propose PEANUT (Predicting Potential Energies with Artificial NeUral neTworks), a simple proof-of-concept GNN architecture using the `e3nn` package [32, 33] for the equivariant updates. We evaluate PEANUT on molecular energy and force prediction, varying the short-range cutoff and including or omitting longer-range scalar interactions. Since accurate forces are essential for stable MD simulations, we focus our analysis on force prediction using a subset of the chemically diverse SPICE v2 dataset [34] and the widely used MD17 dataset [35]. We show that our hybrid approach provides a favorable trade-off between computational efficiency and predictive performance.

2 Model Architecture

The central design principle of PEANUT is to separate local geometric and distance-dependent interactions by combining equivariant and scalar message passing at different cutoff radii. We introduce two cutoff radii, $0 \leq r_{\text{short}} < r_{\text{cut}}$, and corresponding neighborhoods. The short-range neighborhood $\mathcal{N}_{\text{short}}(i)$ comprises neighbors within r_{short} , while the larger neighborhood $\mathcal{N}_{\text{long}}(i)$ comprises neighbors within r_{cut} . Equivariant interactions are computed over $\mathcal{N}_{\text{short}}(i)$, whereas scalar interactions are computed over $\mathcal{N}_{\text{long}}(i)$. We apply equivariant updates before scalar updates to first construct a geometry-aware local representation, which is subsequently extended using scalar updates over the larger neighborhood. This sequential design may also allow future integration with existing equivariant architectures, whose learned representations could be augmented by an additional scalar message-passing stage. Alternative orderings, including alternating equivariant and scalar updates, are not investigated here.

An atomistic system consisting of N atoms is described by Cartesian coordinates $\{r_i\}_{i=1}^N$ and atomic species $\{Z_i\}_{i=1}^N$. Each atom is initially assigned a scalar embedding based on its element type:

$$h_i^{(0)} = e(Z_i) \in \mathbb{R}^D,$$

where i denotes the atom index, e is a learnable embedding lookup table and $D \in \mathbb{N}$ is the embedding dimension. Pairwise distances r_{ij} are encoded as G -dimensional radial feature vectors $\phi_{ij} \in \mathbb{R}^G$, which serve as edge-wise embeddings. Further details are provided in Appendix B.1. In the equivariant block, node embeddings $h_i^{(e,t)}$ contain F_s scalar and F_v vector features, corresponding to $F_s + 3F_v$ channels. Scalar features are initialized using $h_i^{(0)}$, while vector features are initialized to zero. For each edge (i, j) , radial feature vectors are used to compute learnable weights $w_{ij} = \text{MLP}(\phi_{ij}) \cdot c_s(r_{ij})$, where MLP denotes a multilayer perceptron and c_s denotes a smoothing function. Equivariant messages are computed using the `FullyConnectedTensorProduct` from `e3nn`, here denoted as TP , and aggregated over the short-range neighborhood:

$$h_i^{(e,t+1)} = h_i^{(e,t)} + \sum_{j \in \mathcal{N}_{\text{short}}(i)} \underbrace{\text{TP}(h_j^{(e,t)}, Y_{ij}; w_{ij})}_{m_{ij}^{(e,t)}},$$

where Y_{ij} denotes spherical harmonics evaluated on the normalized distance vector $\vec{r}_{ij}/r_{ij}$ up to degree l_{max} . The cutoff function c_s that is used to define the pairwise weights w_{ij} smoothly decays to zero at r_{short} . To ensure smooth energy and force predictions, all cutoff functions are at least twice continuously differentiable ($\mathcal{C}^2$). Since r_{short} is comparatively small, c_s preserves the signal over most of the inner region and decays only near the cutoff. Details on the construction of the smoothing functions are provided in the Appendix B.2. The equivariant block is iterated for N_e updates.

The resulting equivariant features are then mapped to a purely scalar representation,

$$h_i^{(s,N_e)} = \text{TP}(h_i^{(e,N_e)}, h_i^{(e,N_e)}),$$

where the `FullyConnectedTensorProduct` TP produces a scalar output. Scalar message passing is subsequently performed over the full cutoff range:

$$h_i^{(s,N_e+t+1)} = h_i^{(s,N_e+t)} + \sum_{j \in \mathcal{N}_{\text{long}}(i)} \underbrace{f_{\text{scupdate}}(h_j^{(s,N_e+t)} \parallel \phi_{ij}) \cdot c_l(r_{ij})}_{m_{ij}^{(s,N_e+t)}},$$

where f_{scupdate} is an MLP and $c_l(r_{ij})$ smoothly decays to zero at r_{cut} (for more details on the smoothing function, see also Appendix B.2). This process is repeated for N_s updates. After $T = N_e + N_s$ message passing iterations, atom-wise energy contributions are predicted as $E_i = f_{\text{energy}}(h_i^{(s,T)}) \in \mathbb{R}$ and the total potential energy is obtained by summation:

$$E = \sum_{i=1}^N E_i.$$

Atomic forces are obtained as the negative gradient of the predicted energy with respect to the atomic positions using automatic differentiation. The training loss is a weighted sum of energy and force losses; details are provided in the Appendix B.3. The model architecture is illustrated in Figure 4 in Appendix B.4.

3 Experiments

To investigate model performance across a chemically diverse set of configurations, we consider a subset of the SPICE v2 dataset [34], referred to as WPS (water, PubChem, solvated). The WPS dataset contains systems with 63–110 atoms and covers 10 different atom species. A detailed description of the subset selection and data splits is provided in Appendix C.1. Most contemporary MLIP architectures are evaluated on established benchmark datasets such as QM9 [36] and MD17 [35]. We do not consider QM9 because it does not provide force labels, which are essential for our analysis. Instead, we evaluate PEANUT on MD17, which provides both energy and force labels and is therefore well suited for assessing the effect of varying the short-range equivariant cutoff on force predictions.

For the WPS dataset, energy MAEs are reported per atom, whereas for MD17, energy MAEs refer to total molecular energies. Force MAEs are averaged over Cartesian force components for both datasets. As our primary goal is to investigate the interplay between equivariant and scalar message passing and the effect of the equivariant cutoff, we use a fixed set of hyperparameters across all experiments rather than performing extensive task-specific optimization. This allows for a controlled comparison of the different model configurations. The full set of hyperparameters is provided in Table 1 in Appendix B.5.

3.1 Baseline

As baselines, we study the purely equivariant PEANUT model without scalar updates as well as a scalar-only model obtained by setting the inner cutoff r_{short} to zero. We compare these baseline models with the corresponding hybrid PEANUT configurations in terms of both prediction accuracy and inference time. This allows us to investigate the effect of combining equivariant short-range and scalar mid-range interactions and to assess the resulting trade-off between accuracy and computational efficiency. All hybrid configurations share the same architectural hyperparameters and number of trainable parameters. The equivariant-only and scalar-only baselines are obtained by removing the respective message-passing stages, resulting in reduced network depth and parameter counts, as well as differences in the effective interaction range.

3.2 WPS Set

The WPS dataset contains diverse molecular and multi-molecular systems. We fix $r_{cut} = 8.0$ Å and vary the short-range cutoff r_{short} . Three independent 0.8/0.1/0.1 train/validation/test splits are used for each configuration. We report mean and standard deviation of the energy and force MAEs, together with inference time in ms (Figure 1 and Table 4 in Appendix D.1). Note that, as both energy and force evaluation are highly relevant, inference here corresponds to one forward and backward pass. While energy errors vary only moderately across cutoff configurations, force prediction accuracy improves substantially when introducing a non-zero equivariant cutoff, with further increases providing diminishing improvements. This suggests that force prediction is particularly sensitive to the inclusion of local equivariant representations, making it an informative metric for assessing the proposed accuracy–cost trade-off. Meanwhile, inference time increases substantially with increasing equivariant cutoff. A cutoff around 3.0 Å therefore provides a favorable compromise between accuracy and computational cost.

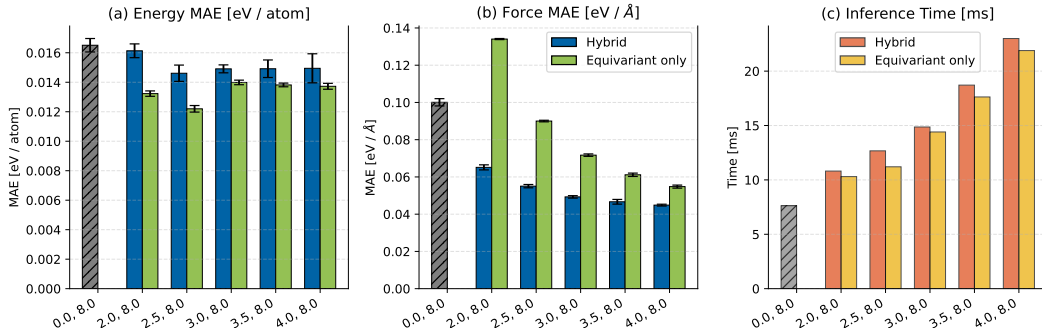


Figure 1: Mean absolute error (MAE) and inference time for different cutoff combinations tested on the WPS dataset. Each pair specifies the cutoffs for the equivariant block (first) and scalar updates (second). The [0.0, 8.0] configuration corresponds to the purely scalar model and is shown in gray. (a) Energy MAE in [eV/atom], (b) force MAE in [eV/Å], both shown as mean $\pm$ standard deviation, and (c) average inference time in ms, computed as mean over 10,000 model evaluations.

The WPS systems contain multiple molecules and interactions can extend beyond the 8 Å cutoff. As a baseline study, we compare the hybrid model with several equivariant model versions in which scalar updates are disabled (see Figure 1). The equivariant-only model with a 3.5 Å cutoff is the first to outperform the mixed configuration with a 2.0 Å equivariant cutoff, but requires substantially

higher computational cost. This comparison further supports the benefit of combining scalar and equivariant message passing to achieve a favorable accuracy–efficiency trade-off.

3.3 MD17 Set

The MD17 dataset [35] consists of structures, energies, and forces from ab initio molecular dynamics simulations of eight small organic molecules (for more details, see Appendix C.2). For each molecule, we train models with a fixed long-range cutoff of $r_{cut} = 8.0$ Å and short-range equivariant cutoffs $r_{short} \in \{0.0, 2.0, 2.5, 3.0\}$ Å. The $[0.0, 8.0]$ configuration corresponds to a purely scalar model. For each molecule, we randomly select 1250 structures, using 1000 for training and 125 each for validation and testing. To account for variations due to data selection, we generate three independent splits and report the mean MAE across them. Results are reported in Table 5, with corresponding standard deviations provided in Table 6 (Appendix D.2).

The MD17 experiments are intended to assess the influence of the equivariant cutoff on energy and force prediction within the PEANUT architecture, rather than to establish a direct performance ranking against existing MLIPs. For context, we also include published results for SchNet, DimeNet, PaiNN, PhysNet, MACE, and Equiformer [21, 20, 22, 23]. Energy errors are reported as molecular MAEs (kcal/mol), while force errors are reported as MAEs over Cartesian force components (kcal/mol/Å). Note that the literature values are not directly comparable due to differences in data splits and, in some cases, dataset versions. In particular, MACE was evaluated on the revised MD17 dataset (rMD17 [37]), whereas our experiments use the original MD17 dataset.

Overall, increasing the equivariant cutoff generally improves force prediction accuracy, although the magnitude of the improvement depends on the molecular system. In particular, introducing even a relatively small equivariant neighborhood substantially reduces force errors compared with the purely scalar configuration. Energy errors also tend to decrease with increasing equivariant cutoff, but show a less consistent dependence. For some molecules, such as benzene, relatively small equivariant cutoffs already yield comparable accuracy to larger cutoffs.

4 Conclusion, Limitations, and Outlook

In this work, we introduced a machine-learned interatomic potential that combines equivariant and scalar message passing through a dual-cutoff design. We showed that a short cutoff for equivariant updates combined with a longer cutoff for scalar interactions achieves accuracy comparable to models relying exclusively on equivariant updates over larger neighborhoods, while substantially reducing computational cost. This design is motivated by the idea that local geometric information can be captured through equivariant representations, while computationally cheaper scalar representations may be sufficient to describe additional interactions over larger neighborhoods. These results suggest that this separation of interaction regimes provides a promising route toward more computationally efficient machine-learned potentials while maintaining competitive predictive accuracy.

It should be noted that the use of a fixed global cutoff restricts the explicit treatment of interactions beyond this distance, and no hyperparameter tuning was performed, meaning that the reported results may not reflect the full potential of the architecture. Future work could systematically explore cutoff radii, including element-dependent inner cutoffs, to further optimize the balance between accuracy and efficiency. Training with optimized hyperparameters on larger and more diverse datasets, including evaluations on chemically distinct test systems, would enable a more comprehensive assessment of the model’s generalization and transferability. Furthermore, while the baseline comparisons provide insight into the performance of the individual and combined message-passing approaches, differences in model capacity may also contribute to the observed results. A more comprehensive ablation study, controlling for network depth and parameter count, would therefore help to better isolate the contributions of the different message-passing stages. Finally, as our approach may also serve as an extension to existing MLIP architectures, we plan to investigate whether the hybrid message-passing and dual-cutoff approach can be applied to state-of-the-art equivariant models, such as MACE, to assess its general applicability and potential computational benefits. Evaluating the computational benefits in full molecular dynamics simulations remains an important direction for future work.

References

- [1] Dylan M. Anstine and Olexandr Isayev. Machine learning interatomic potentials and long-range physics. *J. Phys. Chem. A*, 127(11):2417–2431, feb 2023.
- [2] Silvan Käser, Luis Itza Vazquez-Salazar, Markus Meuwly, and Kai Töpfer. Neural network potentials for chemistry: concepts, applications and prospects. *Digital Discovery*, 2(1):28–58, 2023.
- [3] Yuanqing Wang, Kenichiro Takaba, Michael S. Chen, Marcus Wieder, Yuzhi Xu, Tong Zhu, John Z. H. Zhang, Arnav Nagle, Kuang Yu, Xinyan Wang, Daniel J. Cole, Joshua A. Rackers, Kyunghyun Cho, Joe G. Greener, Peter Eastman, Stefano Martiniani, and Mark E. Tuckerman. On the design space between molecular mechanics and machine learning force fields. *Applied Physics Reviews*, 12(2):021304, April 2025.
- [4] Zhen Liu, Tetiana Zubatiuk, Adrian Roitberg, and Olexandr Isayev. Auto3d: Automatic generation of the low-energy 3d structures with ani neural network potentials. *Journal of Chemical Information and Modeling*, 62(22):5373–5382, 2022. PMID: 36112860.
- [5] Roman Zubatyuk, Justin S. Smith, Jerzy Leszczynski, and Olexandr Isayev. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Science Advances*, 5(8):eaav6490, 2019.
- [6] Esther Heid, Kevin P. Greenman, Yunsie Chung, Shih-Cheng Li, David E. Graff, Florence H. Vermeire, Haoyang Wu, William H. Green, and Charles J. McGill. Chemprop: A machine learning package for chemical property prediction. *Journal of Chemical Information and Modeling*, 64(1):9–17, December 2023.
- [7] Leonard Galustian, Konstantin Mark, Johannes Karwounopoulos, Maximilian P.-P. Kovar, and Esther Heid. Goflow: efficient transition state geometry prediction with flow matching and e(3)-equivariant neural networks. *Digital Discovery*, 4(12):3492–3501, 2025.
- [8] Johannes Karwounopoulos, Jasper De Landsheere, Leonard Galustian, Tobias Jechtl, and Esther Heid. Graph-based prediction of reaction barrier heights with on-the-fly prediction of transition states. *Digital Discovery*, 4:3208–3216, 2025.
- [9] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, Apr 2007.
- [10] Alea Miako Tokita and Jörg Behler. How to train a neural network potential. *The Journal of Chemical Physics*, 159(12):121501, 09 2023.
- [11] Alexandre Duval, Simon V. Mathis, Chaitanya K. Joshi, Victor Schmidt, Santiago Miret, Fragkiskos D. Malliaros, Taco Cohen, Pietro Liò, Yoshua Bengio, and Michael Bronstein. A hitchhiker’s guide to geometric gnns for 3d atomic systems, 2024.
- [12] Taco S. Cohen and Max Welling. Steerable cnns. *CoRR*, abs/1612.08498, 2016.
- [13] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds, 2018.
- [14] Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch-gordan nets: a fully fourier space spherical convolutional neural network, 2018.
- [15] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data, 2018.
- [16] Brandon Anderson, Truong Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- [17] J. S. Smith, O. Isayev, and A. E. Roitberg. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.*, 8:3192–3203, 2017.
- [18] Christian Devereux, Justin S. Smith, Kate K. Huddleston, Kipton Barros, Roman Zubatyuk, Olexandr Isayev, and Adrian E. Roitberg. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.*, 16(7):4192–4202, jun 2020.
- [19] Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *arXiv*, 2017.

- [20] Kristof T. Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra, 2021.
- [21] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs, 2022.
- [22] Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields, 2023.
- [23] Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs, 2022.
- [24] Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations, 2023.
- [25] Yi-Lun Liao, Alexander J. Hoffman, Sabrina C. Shen, Alexandre Duval, Sam Walton Norwood, and Tess Smidt. Equiformerv3: Scaling efficient, expressive, and general $se(3)$ -equivariant graph attention transformers, 2026.
- [26] Thomas Plé, Louis Lagardère, and Jean-Philip Piquemal. FENNIX: force-field-enhanced neural network interactions: from local equivariant embedding to atom-in-molecule properties and long-range effects. *Chem. Sci.*, 14(44):12554–12569, 2023.
- [27] Kevin K. Huguenin-Dumittan, Philip Loche, Ni Haoran, and Michele Ceriotti. Physics-inspired equivariant descriptors of nonbonded interactions. *The Journal of Physical Chemistry Letters*, 14(43):9612–9618, 2023. PMID: 37862712.
- [28] Tsz Wai Ko, Jonas A. Finkler, Stefan Goedecker, and Jörg Behler. Accurate fourth-generation machine learning potentials by electrostatic embedding. *Journal of Chemical Theory and Computation*, 19(12):3567–3579, June 2023.
- [29] Dongjin Kim, Xiaoyu Wang, Santiago Vargas, Peichen Zhong, Daniel S. King, Theo Jaffrelot Inizan, and Bingqing Cheng. A Universal Augmentation Framework for Long-Range Electrostatics in Machine Learning Interatomic Potentials. *J. Chem. Theory Comput.*, 21(24):12709–12724, December 2025.
- [30] Oliver T. Unke, Stefan Chmiela, Michael Gastegger, Kristof T. Schütt, Huziel E. Sauceda, and Klaus-Robert Müller. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications*, 12(1), December 2021.
- [31] Thomas Plé, Olivier Adjoua, Léa El Khoury, Lina Alferkh, Zhifeng Jing, Corentin Villot, Anouar Benali, Evgeny Posenitskiy, Louis Lagardère, and Jean-Philip Piquemal. A foundation model for accurate atomistic simulations in drug design. 2026.
- [32] Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks, 2022.
- [33] Mario Geiger, Tess Smidt, Alby M., Benjamin Kurt Miller, Wouter Boomsma, Bradley Dice, Kostiantyn Lapchevskyi, Maurice Weiler, Michał Tyszkiewicz, Simon Batzner, Dylan Madisetti, Martin Uhrin, Jes Frellsen, Nuri Jung, Sophia Sanborn, Mingjian Wen, Josh Rackers, Marcel Rød, and Michael Bailey. Euclidean neural networks: e3nn, April 2022.
- [34] Peter Eastman, Benjamin P. Pritchard, John D. Chodera, and Thomas E. Markland. Nutmeg and spice: Models and data for biomolecular machine learning. *Journal of Chemical Theory and Computation*, 20(19):8583–8593, September 2024.
- [35] Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017.
- [36] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1), August 2014.
- [37] Anders S. Christensen and O. Anatole von Lilienfeld. On the role of gradients for machine learning of molecular energies and forces, 2020.
- [38] Oliver T. Unke and Markus Meuwly. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of Chemical Theory and Computation*, 15(6):3678–3693, 2019. PMID: 31042390.

- [39] Alea Miako Tokita and Jörg Behler. How to train a neural network potential. *The Journal of Chemical Physics*, 159(12), 2023.
- [40] Peter Eastman, Pavan Behara, David Dotson, Raimondas Galvelis, John Herr, Josh Horton, Yuezhi Mao, John Chodera, Benjamin Pritchard, Yuanqing Wang, Gianni Fabritiis, and Thomas Markland. Spice, a dataset of drug-like molecules and peptides for training machine learning potentials. *Scientific Data*, 10, 01 2023.
- [41] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, October 1996.
- [42] Alexandre Tkatchenko and Matthias Scheffler. Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data. *Physical Review Letters*, 102(7):073005, February 2009.

A Desired Physical Properties

A.1 Invariance

Machine-learned interatomic potentials aim to approximate potential energy surfaces that obey fundamental physical symmetries. Independently of the specific model architecture, several invariance properties are therefore desirable for atomistic energy prediction models. The predicted total energy is a function of the Cartesian coordinates and atomic species, i.e. $E = E(\{r_i, Z_i\}_{i=1}^N)$, and should ideally satisfy the following invariances:

- *Translational invariance:*

$$E(\{r_i + t\}) = E(\{r_i\}), \quad \forall t \in \mathbb{R}^3, \quad (1)$$

- *Permutation invariance:*

$$E(\{r_{\pi(i)}, Z_{\pi(i)}\}) = E(\{r_i, Z_i\}), \quad \forall \text{ permutations } \pi, \quad (2)$$

- *Rotational invariance:*

$$E(R(\{r_i\})) = E(\{r_i\}), \quad \forall R \in SO(3). \quad (3)$$

These properties ensure that the predicted energy depends only on the intrinsic geometry and composition of the system and not on arbitrary choices of coordinate system or atom ordering.

A.2 Equivariance

To obtain more expressive representations of atomic environments, many modern architectures incorporate equivariant message passing. A function f is said to be equivariant with respect to rotations if

$$f(Rx) = Rf(x),$$

for any rotation matrix $R \in SO(3)$. This property is particularly important for vector-valued quantities such as interatomic forces, as it ensures that predictions transform consistently under rotations (and reflections) of the input geometry. In machine-learned interatomic potentials, equivariance is commonly achieved by constructing representations that transform according to irreducible representations of the rotation group. Instead of restricting features to scalars, higher-order geometric features transforming as scalars, vectors, and higher-order tensors can be used to obtain more expressive atomic representations [12, 13, 14, 15, 16].

In this work, equivariant operations are implemented using the e3nn framework [32, 33], which provides the necessary building blocks for equivariant neural networks based on spherical harmonics and Clebsch–Gordan coefficients. In particular, we employ the `FullyConnectedTensorProduct` available in e3nn, which combines features using learnable weights while preserving equivariance.

B Model Details and Training

B.1 Distance Representation – Gauss Functions

We use a set of Gauss functions f_k to encode pairwise interatomic distances r_{ij} :

$$f_k(r_{ij}) = e^{-\eta_k(r_{ij}-\psi_k)^2}, \quad k = 1, \dots, G,$$

where η_k and ψ_k denote the widths and centers of the Gaussians, respectively. These functions are then used to construct one radial feature vector ϕ_{ij} per atom-pair (i, j) , projecting the pairwise distance r_{ij} to the full range $[0, r_{cut}]$:

$$\phi_{ij} = (f_1(r_{ij}), f_2(r_{ij}), \dots, f_G(r_{ij}))^T \in \mathbb{R}^G.$$

Figure 2 shows representative plots for (a) Gauss functions and (b) two example radial feature vectors.

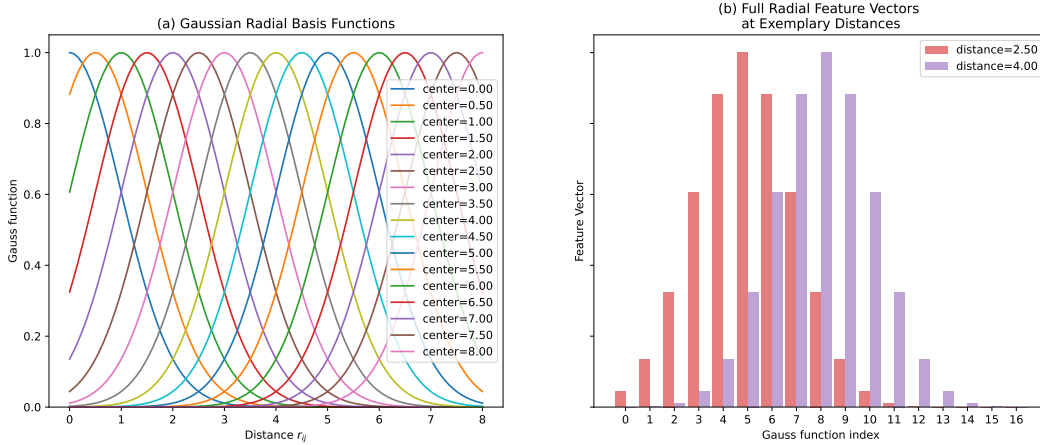


Figure 2: Gaussian Radial Basis functions with $G = 17$. Subplot (a) shows Gaussian functions $f_k(r_{ij}), k = 1, \dots, G$ with equidistant centers and with parameter $\eta_k = 1$, for all k , subplot (b) shows full exemplary radial feature vectors for two different pairwise atom distances.

B.2 Smoothing Functions

The key building block is a *quintic* function $q_a^b : [a, b] \rightarrow [0, 1]$, as also employed in [38]:

$$q_a^b(x) = 1 - 6 \left(\frac{x-a}{b-a} \right)^5 + 15 \left(\frac{x-a}{b-a} \right)^4 - 10 \left(\frac{x-a}{b-a} \right)^3,$$

which satisfies

$$\begin{aligned} q_a^b(a) &= 1, \\ q_a^b(b) &= 0, \\ (q_a^b)'(a) &= (q_a^b)'(b) = 0, \\ (q_a^b)''(a) &= (q_a^b)''(b) = 0. \end{aligned} \tag{4}$$

This function is applied to the final part of an interaction sphere to ensure smoothness at the cutoff, while the signal may initially decrease linearly to reduce early scaling. Hence, we additionally define a linear scaling function $l : [0, r_{cut}] \rightarrow [0, 1]$:

$$l(x) = 1 - \frac{x}{2r_{cut}}.$$

We combine these two functions to define a smoothing function $c(\cdot; r, B) : [0, \infty) \rightarrow [0, 1]$:

$$c(x; r, B) = \begin{cases} l(x), & 0 \leq x \leq B \cdot r, \\ l(x) \cdot q_{B \cdot r, r}^r(x), & B \cdot r \leq x \leq r, \\ 0, & x \geq r, \end{cases} \tag{5}$$

where $B \in (0, 1)$ controls the steepness of the smoothing function as well as the starting point for the quintic smooth scaling. We then define c_s and c_l for the short- and long-range parts, respectively, as $c_s(x) = c(x; r_{short}, B)$ and $c_l(x) = c(x; r_{cut}, B)$. The initial linear decay gradually reduces the contributions of more distant neighbors without excessively suppressing interactions within small cutoff radii, while the quintic transition ensures a smooth decay to zero near the cutoff. The smoothing functions are shown in Figure 3.

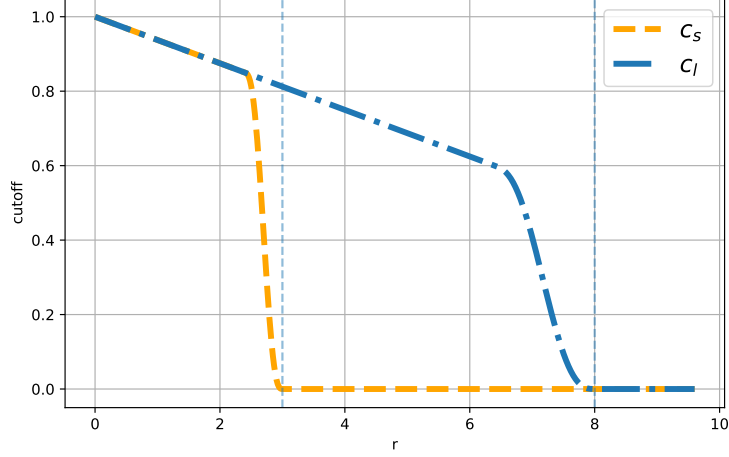


Figure 3: Smoothing functions used in the model, all in $\mathcal{C}^2(\mathbb{R})$. Parameters for illustration: $r_{cut} = 8.0$, $r_{short} = 3.0$, $B = 0.8$ for both c_s and c_l .

B.3 Loss Function

All models are trained in a supervised manner to predict total (system-wise) energies (see [39]). The predicted total energy is obtained by summing atom-wise energy contributions,

$$E^{pred} = \sum_{i=1}^{N_{atoms}} E_i.$$

The energy loss is normalized by the number of atoms,

$$L_E = \frac{1}{N_{atoms}} \mathcal{L}_E(E^{pred}, E^{true}),$$

where $\mathcal{L}_E$ denotes the chosen energy loss function (either the mean absolute error (MAE) or mean squared error (MSE)). Note that both E^{pred} and E^{true} are scalar values. Atomic forces are obtained as the negative gradient of the predicted energy with respect to atomic positions using automatic differentiation,

$$F_i^{pred} = -\nabla_{r_i} E^{pred},$$

ensuring energy-force consistency and energy conservation by construction. The force loss is averaged over all atoms and Cartesian components:

$$L_F = \frac{1}{3 \cdot N_{atoms}} \sum_{i=1}^{N_{atoms}} \sum_{d=1}^3 \mathcal{L}_F(F_{i,d}^{pred}, F_{i,d}^{true}).$$

Here, $\mathcal{L}_F$ denotes the force loss function, also chosen as either MAE or MSE. The total loss function used for training is defined as a weighted sum of the energy and force losses:

$$\mathcal{L} = \lambda_E L_E + \lambda_F L_F, \quad (6)$$

where $\lambda_E, \lambda_F \in \mathbb{R}_+$ control the relative importance of each contribution. The weights λ_E and λ_F are treated as hyperparameters that can be adjusted during training to balance the contributions of the energy and force losses.

B.4 Architecture Illustration

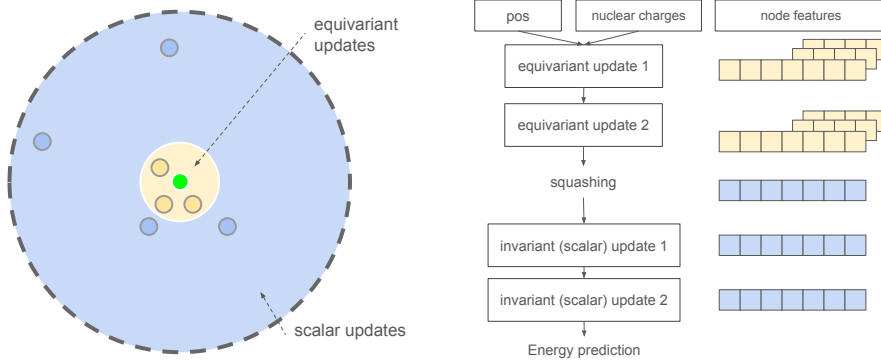


Figure 4: Schematic illustration of the model architecture with two equivariant and two scalar message-passing updates.

B.5 Hyperparameters and Hardware

All models were trained on NVIDIA GeForce RTX 4090. Training was performed at most for 1000 epochs for all experiments with the MD17 dataset, and at most for 500 epochs for the WPS dataset. We implemented early stopping based on validation loss. The model parameters corresponding to the lowest validation loss are retained. The number of trainable parameters is identical across hybrid configurations with different cutoff radii (348,801), whereas the equivariant- and scalar-only configurations have fewer parameters. The most important parameters are reported in Table 1.

Table 1: Model parameters were not optimized. We used the same set of hyperparameters across all experiments.

Parameter	Value
Number of equivariant message passes N_e	2
Number of scalar message passes N_s	2
Number of radial distribution functions G	32
Maximum degree of spherical harmonics l_{max}	1
Number of scalar features in equivariant block	32
Number of vector features in equivariant block	32
Total number of channels	128
Energy loss function	Mean Squared Error (MSE)
Force loss function	Mean Absolute Error (MAE)
(λ_E, λ_F) energy and force weights for the first x% epochs	[10, 0.1]
(λ_E, λ_F) energy and force weights for the final (1-x)% epochs	[5, 0.4]
Change weights at x% of the training epochs	80%
Batchsize	8
Learning rate	0.001
Number of parameters in embedding block	320
Number of parameters in equivariant block	268,288
Number of parameters in scalar block	12,480
Number of parameters in TP (equivariant to scalar mapping)	65,536
Number of parameters in readout MLP	2,177
Total number of parameters	348,801

C Datasets

C.1 WPS Dataset Construction

The WPS dataset is derived from the SPICE collection [40, 34]¹, specifically targeting the Solvated PubChem and Water Clusters subsets (see Table 2). After filtering for samples with complete energy and force labels, the final dataset comprises 13,909 configurations.

Table 2: Dataset – subset overview.

Subset	Number of Molecules / Clusters	Number of Conformations	Number of Atoms	Elements
Solvated PubChem	1397	13,934	63–110	H, C, N, O, F, P, S, Cl, Br, I
Water Clusters	1	1000	90	H, O

We further characterized the dataset in terms of atom-type co-occurrence. Table 3 reports the number of samples containing at least one atom pair corresponding to the specified row and column atom types. We emphasize that this analysis does not account for interatomic distances. Consequently, although all atom-type combinations are present in the dataset at the sample level, some pairs may lie beyond the interaction cutoff used during training.

Table 3: Single pairwise occurrence per sample. Each entry in the matrix O_{ij} corresponds to the number of samples in the data set that contain at least one atom of species i and one atom of species j .

	Atomic number	H	C	N	O	F	P	S	Cl	Br	I
H	1	13909	12566	10168	13909	2215	1042	5834	2359	1157	364
C	6	12566	12566	10038	12566	2128	899	5635	2339	1117	334
N	7	10168	10038	10168	10168	1698	599	4726	1640	827	236
O	8	13909	12566	10168	13909	2215	1042	5834	2359	1157	364
F	9	2215	2128	1698	2215	2215	147	898	350	160	40
P	15	1042	899	599	1042	147	1042	389	250	70	5
S	16	5834	5635	4726	5834	898	389	5834	940	470	97
Cl	17	2359	2339	1640	2359	350	250	940	2359	150	59
Br	35	1157	1117	827	1157	160	70	470	150	1157	7
I	53	364	334	236	364	40	5	97	59	7	364

C.2 MD17

The MD17 dataset [35] consists of structures, energies, and forces from ab initio molecular dynamics simulations of eight small organic molecules. All trajectories were generated at 500 K with a 0.5 fs timestep using the PBE+vdW-TS [41, 42] electronic structure method. The datasets range from 150k to nearly 1M conformations per molecule and cover energy ranges of 20–48 kcal / mol and force components of 266–570 kcal / mol / Å. The task is to predict energies (and forces) using separate models for each molecule. Since each system contains only a single molecule, MD17 probes conformational rather than chemical degrees of freedom.

D Additional Results for WPS and MD17

D.1 WPS

Table 4 reports mean values and standard deviations for energy and force MAE across three different data splits of the WPS dataset of equal size.

¹see also <https://github.com/openmm/spice-dataset>

Table 4: Mean absolute errors for the WPS dataset. All models were trained on energies and forces. Energy errors are given in [eV/atom], force errors are given in [eV / Å]. Training times are given in minutes.

equivariant and scalar updates, double cutoff	[0.0, 8.0]	[2.0, 8.0]	[2.5, 8.0]	[3.0, 8.0]	[3.5, 8.0]	[4.0, 8.0]
Energy MAE Mean	0.017	0.016	0.015	0.015	0.015	0.015
Energy MAE Std	0.000	0.000	0.001	0.000	0.001	0.001
Force MAE mean	0.100	0.065	0.055	0.049	0.047	0.045
Force MAE Std	0.002	0.001	0.001	0.001	0.001	0.000
Training time	121	300	348	425	538	702
equivariant updates only, single cutoff	NA	2.0	2.5	3.0	3.5	4.0
Energy MAE Mean	NA	0.013	0.012	0.014	0.014	0.014
Energy MAE Std	NA	0.000	0.001	0.000	0.001	0.000
Force MAE mean	NA	0.134	0.090	0.072	0.061	0.055
Force MAE Std	NA	0.000	0.000	0.001	0.001	0.001
Training time	NA	237	296	378	485	618

D.2 MD17

The results reported in Table 5 correspond to the mean MAE over the three data splits. Table 5 also includes results for SchNet and DimeNet, as reported in [21] as well as PaiNN and PhysNet, as reported in [20], MACE as reported in [22] and Equiformer as reported in [23]. Note that MACE was trained on the revised MD17 dataset (rMD17 [37]).

As we did not use the same data samples for the MD17 benchmark analysis as those reported in the literature, we constructed three independent datasets of equal size by randomly sampling 1,250 configurations per solute. Each dataset was split into 1,000, 125, and 125 samples for training, validation, and testing, respectively. Table 6 reports the mean and standard deviation of the MAE loss obtained by training the models independently on each of the three dataset variants.

Table 5: Mean absolute errors for the MD17 dataset. For each solute, PEANUT was trained and evaluated using three independent random splits, each comprising 1000 training, 125 validation, and 125 test configurations. PEANUT errors are reported as means over the three splits. Energy errors refer to total molecular energies and are given in [kcal/mol], while force errors are averaged over Cartesian force components and given in [kcal/mol/Å]. NA denotes unavailable data in the respective primary source. Published results for other MLIPs are included for context. In particular, MACE was evaluated on the revised MD17 dataset (rMD17 [37]), whereas the other reference results were obtained using the original MD17 dataset. Note that different data splits were used, and comparisons between models should therefore be made with caution.

solute	type	SchNet	DimeNet	PaiNN	PhysNet	Equiformer ($L_{max} = 3$)	MACE	[0.0, 8.0]	[2.0, 8.0]	[2.5, 8.0]	[3.0, 8.0]
Aspirin	energy	0.37	0.20	0.159	0.230	0.122	0.051	0.676	0.372	0.350	0.323
	force	1.35	0.50	0.371	0.605	0.152	0.152	1.453	0.789	0.717	0.702
Benzene	energy	0.08	0.08	NA	NA	0.058	0.009	0.093	0.071	0.069	0.070
	force	0.31	0.19	NA	NA	0.187	0.007	0.514	0.213	0.191	0.211
Ethanol	energy	0.08	0.06	0.063	0.059	0.051	0.009	0.190	0.099	0.101	0.079
	force	0.39	0.23	0.230	0.160	0.067	0.048	0.726	0.397	0.416	0.316
Malonaldehyde	energy	0.13	0.10	0.091	0.094	0.074	0.018	0.424	0.232	0.179	0.155
	force	0.66	0.38	0.319	0.319	0.125	0.095	1.232	0.730	0.570	0.521
Naphthalene	energy	0.16	0.12	0.117	0.142	0.101	0.012	0.344	0.182	0.175	0.146
	force	0.58	0.22	0.083	0.310	0.046	0.037	0.968	0.441	0.419	0.346
Salicylic acid	energy	0.20	0.13	0.114	0.126	0.099	0.021	0.353	0.236	0.199	0.170
	force	0.85	0.37	0.209	0.337	0.090	0.071	1.143	0.668	0.592	0.541
Toluene	energy	0.12	0.10	0.097	0.100	0.085	0.012	0.296	0.137	0.125	0.126
	force	0.57	0.22	0.102	0.191	0.048	0.035	0.969	0.391	0.380	0.377
Uracil	energy	0.14	0.12	0.104	0.108	0.099	0.012	0.434	0.142	0.143	0.132
	force	0.56	0.30	0.140	0.218	0.078	0.048	1.467	0.436	0.396	0.357

Table 6: Mean and standard deviation computed from all three independent dataset variants of the MD17 dataset. All splits contain 1000 training, 125 validation and 125 test samples.

		[0.0, 8.0]		[2.0, 8.0]		[2.5, 8.0]		[3.0, 8.0]	
solute	type	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Aspirin	energy	0.676	0.035	0.372	0.025	0.350	0.024	0.323	0.039
	force	1.453	0.046	0.789	0.020	0.717	0.030	0.702	0.014
Benzene	energy	0.093	0.007	0.071	0.003	0.069	0.005	0.070	0.005
	force	0.514	0.044	0.213	0.005	0.191	0.013	0.211	0.007
Ethanol	energy	0.190	0.048	0.099	0.026	0.101	0.013	0.079	0.004
	force	0.726	0.095	0.397	0.048	0.416	0.025	0.316	0.017
Malonaldehyde	energy	0.424	0.056	0.232	0.036	0.179	0.030	0.155	0.005
	force	1.232	0.100	0.730	0.068	0.570	0.053	0.521	0.013
Naphthalene	energy	0.344	0.038	0.182	0.010	0.175	0.012	0.146	0.003
	force	0.968	0.096	0.441	0.019	0.419	0.019	0.346	0.013
Salicylic acid	energy	0.353	0.016	0.236	0.013	0.199	0.008	0.170	0.010
	force	1.143	0.011	0.668	0.014	0.592	0.015	0.541	0.020
Toluene	energy	0.296	0.038	0.137	0.027	0.125	0.013	0.126	0.012
	force	0.969	0.046	0.391	0.038	0.380	0.027	0.377	0.017
Uracil	energy	0.434	0.017	0.142	0.006	0.143	0.006	0.132	0.007
	force	1.467	0.078	0.436	0.027	0.396	0.032	0.357	0.029