Learning Long-Range Representations with Equivariant Messages

Egor Rumiantsev[*] *egor.rumiantsev@epfl.ch*
Laboratory of Computational Science and Modeling, École Polytechnique Fédérale de Lausanne

Marcel F. Langer[*†] *marcel.langer@epfl.ch*
Laboratory of Computational Science and Modeling, École Polytechnique Fédérale de Lausanne

Tulga-Erdene Sodjargal *tulga-erdene.s@kaist.ac.kr*
Laboratory of Computational Science and Modeling, École Polytechnique Fédérale de Lausanne
Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST)

Michele Ceriotti *michele.ceriotti@epfl.ch*
Laboratory of Computational Science and Modeling, École Polytechnique Fédérale de Lausanne

Philip Loche[†] *philip.loche@tum.de*
Laboratory of Computational Science and Modeling, École Polytechnique Fédérale de Lausanne
Department of Physics, Technical University of Munich
Atomistic Modelling Center, Munich Data Science Institute, Technical University of Munich
Munich Center for Machine Learning (MCML)

[*] *Contributed equally*
[†] *Corresponding authors*

Reviewed on OpenReview: *https://openreview.net/forum?id=pZI9e4SW9P*

Abstract

Machine learning interatomic potentials trained on first-principles reference data are becoming valuable tools for computational physics, biology, and chemistry. Equivariant message-passing neural networks, including transformers, achieve state-of-the-art accuracy but rely on cutoff-based graphs, limiting their ability to capture long-range effects such as electrostatics or dispersion, as well as electron delocalization. While long-range correction schemes based on inverse power laws of interatomic distances have been proposed, they are unable to communicate higher-order geometric information and are thus limited in applicability. To address this shortcoming, we propose the use of equivariant, rather than scalar, charges for long-range interactions, and design a graph neural network architecture, Lorem, around this long-range message passing mechanism. We consider several datasets specifically designed to highlight non-local physical effects, and compare short-range message passing with different receptive fields to invariant and equivariant long-range message passing. Even though most approaches work for careful dataset-specific choices of their model hyperparameters, Lorem works consistently without such changes, with excellent benchmark performance.

1 Introduction

Machine learning interatomic potentials (MLIPs) for atomistic simulations are trained on quantum-mechanical simulations to predict energies and forces of new atomic structures. Most MLIPs assume locality: The energy of each atom depends only on neighbors within a cutoff radius. This leads to linear scaling with respect to the number of atoms but fails for systems with long-range interactions, such as electrostatics, dispersion, or electron delocalization. (Behler & Parrinello, 2007; Ko et al., 2021; Tkatchenko & Scheffler, 2009; Grisafi & Ceriotti, 2019; Unke et al., 2021b; Huguenin-Dumittan et al., 2023) Several approaches aim to overcome this limitation. Message-passing graph neural networks (MPNNs, Gilmer et al. (2017)) overcome locality by iteratively exchanging information between neighboring atoms, but are still constrained by the number of message-passing steps, graph connectivity, and reduced information flow with increasing number of iterations (Cai & Wang, 2020; Alon & Yahav, 2020; Nigam et al., 2022).

An alternative approach is to use physics-inspired corrections to the total energy, written as an inverse power law of interatomic distances, $1/r^p$, with $p = 1$ for charge–charge, $p = 3$ for dipole–dipole, and $p = 6$ for dispersion. Some models predict partial atomic charges, either using explicit charge labels and equilibration schemes, or learning them implicitly from energies and forces to include electrostatic terms (Unke et al., 2021a; Ko et al., 2021; Fedik et al., 2022; Pellegrini et al., 2023; Maruf et al., 2025; Cheng, 2025). Physical long-range interactions can also serve as building block: For instance, Kosmala et al. (2023) propose Ewald message passing, i.e., the use of electrostatic interactions for message passing, and Grisafi & Ceriotti (2019) propose the long-distance equivariant (LODE) framework that uses inverse power-law interactions to compute equivariant features from the scalar potential around each atom.

In this work, we combine the strengths of physical inverse power law $1/r^p$ interactions with the ability of equivariant message passing to communicate higher-order geometric information. Inverse power-law interactions are well-defined for periodic systems, have physically meaningful asymptotic behavior, and can be computed efficiently using established techniques from computational physics. Extending the idea of Ewald message passing, we treat charges as equivariant objects and use inverse power-law potentials as a mechanism for long-range communication. Based on this mechanism, we design Lorem, an MLIP architecture that combines short- and long-range message passing. Lorem offers consistently high accuracy across a series of long-range benchmark tasks, outperforming other short- and long-range message passing models.

Our contributions are:

- We introduce an equivariant, global, long-range message passing mechanism that can leverage efficient methods from computational physics for asymptotic $O(N \log N)$ scaling in periodic systems, and potentially $O(N)$ scaling in non-periodic ones,
- We design a novel MLIP architecture, Lorem, around this mechanism,
- We conduct a series of experiments to probe the limits of equivariant short-range message passing to model long-range physics.

2 Background

Machine learning interatomic potentials Under the Born & Oppenheimer (1927) approximation, which decouples the nuclear and the electronic degrees of freedom, the atoms in a molecule or material move on a potential energy surface (PES) $E = E\left(\{\,(\boldsymbol{r}_i, Z_i) \,|\, i = 1...N\,\}\right)$ where $\boldsymbol{r}_i$ and Z_i are positions and atomic numbers for the N atoms; in a periodic system, i.e., materials or liquids, the arrangement of N atoms is repeated periodically in space, described by three cell vectors $\{\,\boldsymbol{c}_a \,|\, a = 1, 2, 3\,\}$ and corresponding integer offsets $\{\,n_a \,|\, a = 1, 2, 3\,\}$ so that each position $\boldsymbol{r}_i$ is associated with the replicas $\{\,\boldsymbol{r}_i + \sum_{a=1}^{3} n_a \boldsymbol{c}_a \,|\, \boldsymbol{n} \in \mathbb{Z}^3\,\}$. The energy is then computed for the positions in the unit cell ($\boldsymbol{n} = 0$) considering their interactions with the infinite 'crystal' system. The forces, which drive the dynamics of the atoms, are defined as derivatives of the energy $\boldsymbol{f}_i = -\nabla_{\boldsymbol{r}_i} E$. Traditionally, this PES has been approximated by physics-inspired analytical expressions, called force fields, that are parametrized manually or through global optimization. In the past

decades, in tandem with the increasing availability of large datasets of quantum mechanical reference data, MLIPs have emerged as a less computationally efficient, but more accurate, data-driven alternative.

Atomistic graph neural networks Most MLIPs can be seen as graph neural networks (GNNs, Battaglia et al. (2018)) acting on a description of an arrangement of atoms in space as a geometric graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ with edges $\mathcal{E}$ corresponding to interatomic relative-position vectors $\boldsymbol{r}_{ij}$ and nodes (or vertices) $\mathcal{V}$ corresponding to atoms. Edges connect all nodes that lie within a cutoff radius r_c of each other. In periodic systems, $\boldsymbol{r}_{ij}$ are constructed to respect periodic boundary conditions, i.e., if a replica lies closer than an original atom, $\boldsymbol{r}_{ij}$ points to the replica. If more than one replica lies within the cutoff, multiple edges with different labels are drawn between nodes. By restricting the range of interactions to neighbors on this graph, linear scaling with the number of atoms N (at constant density) can be achieved; efficient scaling with system size is required to make MLIPs practical for large-scale simulations. MLIPs typically predict E as a sum over atomic contributions, $E = \sum_{i=1}^{N} E_i$, predicted from node features.

Invariance and equivariance The potential energy E is invariant under permutations, i.e., reordering, of atomic positions, as well as global translations and rotations of the coordinate system.[1] In other words, E is invariant under actions of the Euclidean symmetry group E(3) applied to all positions (including replicas). MLIPs must respect these symmetries, at least approximately. Translation invariance is respected by construction in atomistic GNNs through the use of relative-position vectors as edge labels. Permutation invariance is typically ensured through commutative aggregation functions, for instance sums. Finally, rotation invariance can either be learned through data augmentation (Pozdnyakov & Ceriotti (2023); Langer et al. (2024); Domina et al. (2026)), or ensured by requiring that internal features remain aware of their geometric meaning, i.e., that they are *equivariant* to rotations. While in principle, MLIPs could be constructed from invariant features only, equivariant internal features have been found to improve accuracy and data efficiency by allowing the model to access orientation information (Thomas et al. (2018); Batzner et al. (2022); Batatia et al. (2022)). A thorough discussion of equivariance for MLIPs can be found in other works (Smidt, 2021; Unke & Maennel, 2024); essentially, we consider the transformation of internal features of the models under rotations $g \in \mathrm{SO}(3)$ applied equally to all geometric inputs (positions and cell vectors), leading to a joint rotation of all bulk positions. Consider the MLIP up to some hidden layer as a function $f : \mathbb{X} \to \mathbb{Y}$, where $\mathbb{X} = \mathbb{R}^{3\times(N+3)}$ for periodic systems and $\mathbb{X} = \mathbb{R}^{3\times N}$ for molecules. Equivariance is defined by $f \circ g = g \circ f$ for all $g \in \mathrm{SO}(3)$: rotations can be equivalently applied before or after f. To ensure that this is the case, internal features must be constructed as direct sum of different irreducible representations of SO(3), indexed by l, combined with a feature (channel) dimension c: $\mathbb{Y} = \left(\oplus_{l=0}^{l_\mathrm{max}} \mathbb{R}^{2l+1}\right) \otimes \mathbb{R}^c$. Rotations act as linear transformations on these irreducible representations. We write spherical features as tensor $\mathbf{S}$ with the last three indices corresponding to the representation order $l = 0, ..., l_\mathrm{max}$, the component $m = -l, ..., l$, and the channel index c. Such tensors are therefore ragged: Different l are associated with different numbers of components m. The $l = 0$ components are called scalars and are invariant under rotations. Collections of purely scalar features are also denoted as matrices $\boldsymbol{P}$. In Appendix A, we describe a number of operations that can be applied to spherical features without disrupting equivariance.

Long-range interactions The potential energy E arises from the many-body Schrödinger equation, which involves only Coulomb interactions between electrons and nuclei without any range separation. Efficient approximations, such as empirical force fields or MLIPs, typically restrict interactions to local environments, motivated by the nearsightedness principle of electronic matter (Prodan & Kohn (2005)), which states that electronic properties are insensitive to distant perturbations. In the long-range regime, interactions reduce to inverse power laws $1/r^p$ of the interatomic distance. Since in nature no fixed nearsightedness length scale exists, MLIPs must capture both local many-body quantum effects, possibly extending beyond the model's cutoff r_c, and formally infinite-range electrostatic interactions. Additional complexity arises from charge distributions that depend on distant atoms and from electron wavefunctions that may delocalize over large distances; see Section 6.

[1] In fact, the energy is invariant under global inversions as well, i.e., under the action of the full group O(3). To simplify notation in this manuscript, we focus on proper scalars and tensors, i.e., irreducible representations of SO(3) (which does not contain reflections). All arguments and operations can be readily generalized to O(3).

Ewald summation The evaluation of inverse power-law potentials $1/r^p$ has been the object of much study in computational physics and chemistry. The general task is to compute the potential $V_i = V(\boldsymbol{r}_i)$ at a given atomic position $\boldsymbol{r}_i$, induced by (generalized) point charges[2] q_j placed at the position of other atoms j:

$$V_i = \sum_{j=1}^{N} \sum_{\boldsymbol{n}\in\mathbb{Z}^3} \frac{q_j}{|\boldsymbol{r}_i - (\boldsymbol{r}_j + n_1\boldsymbol{c}_1 + n_2\boldsymbol{c}_2 + n_3\boldsymbol{c}_3)|^p} \,. \tag{1}$$

Where $|\cdot|$ denotes the vector norm. In periodic systems, Eq. (1) implies an infinite sum over replicas, denoted by $\sum_{\boldsymbol{n}\in\mathbb{Z}^3}$ and defined by the cell vectors $\boldsymbol{c}_a$; for non-periodic systems (e.g., molecules), this sum and the extra term in the denominator are omitted. For $p \leq 3$, the periodic, infinite, sum converges conditionally, i.e., convergence or divergence depends on the summation order. Ewald (1921) summation was developed to tackle this problem for electrostatics ($p = 1$) and later extended to other exponents; its basic concept is to split $1/r$ into a short-ranged part, which converges fast in real space, and a long-range part, which is smooth, and therefore converges well, in reciprocal space. A naive implementation of Ewald summation scales $O(N^2)$, which can be brought down to $O(N^{3/2})$ by choosing the cutoffs to be proportional to the size of the simulation cell. This is, however, undesirable for MLIPs, which are typically constructed and trained for a fixed cutoff radius. Particle–mesh Ewald (PME, P3M) algorithms reduce this to $O(N \log N)$ by interpolating charges onto a grid and employing the fast Fourier transform for the reciprocal-space part (Darden et al., 1993; Hockney & Eastwood, 2021). While such methods are standard in force fields, implementations in popular machine learning frameworks, PyTorch (Paszke et al., 2019) and JAX (Bradbury et al., 2018), have only become available recently (Loche et al., 2025). In non-periodic systems, the naive evaluation of Eq. (1) scales $O(N^2)$. However, methods like the fast multipole expansion can reduce this cost to $O(N)$ (Greengard & Rokhlin, 1987). Another approach is multi-level summation (Hardy et al., 2015), which scales $O(N)$ for non-periodic and periodic systems and has recently been implemented in JAX (Buchner et al., 2025). For small systems on the order of a hundred atoms, naive implementations—Ewald summation for periodic systems and direct summation for non-periodic systems—are typically faster than more complex, but better-scaling, methods; this is benchmarked, for instance, in Appendix E and Loche et al. (2025). As we largely focus on small systems in this work, we use naive implementations throughout.

3 Related work

Equivariant message passing Bond-order potentials first introduced the idea of repeatedly updating atomic environments to extend interactions beyond the cutoff radius r_c (Tersoff, 1988; Brenner, 1990), a principle now central to modern MLIPs. In MPNNs (Gilmer et al., 2017), atoms exchange messages over M steps, so features and energies depend on neighbors within $M \cdot r_\mathrm{c}$. Letting ${}^k\boldsymbol{P}_i$ denote the features at atom i and message-passing step k:

$${}^{k+1}\boldsymbol{M}_i = \sum_j {}^k m({}^k\boldsymbol{P}_i, {}^k\boldsymbol{P}_j, \boldsymbol{r}_{ij}) \tag{2}$$

$${}^{k+1}\boldsymbol{P}_i = {}^k u({}^k\boldsymbol{P}_i, {}^{k+1}\boldsymbol{M}_i)\,, \tag{3}$$

with learned message and update functions m and u. Early models used invariant updates (Schütt et al., 2017; Xie & Grossman, 2018), later extended to equivariant ones (Gasteiger et al., 2019; Schütt et al., 2021; Batatia et al., 2022; Batzner et al., 2022; Frank et al., 2024b). Equivariant MPNNs are restricted in the choice of operations within the network, as they must retain equivariance throughout (see Appendix A). Recent universal MLIPs trained on big datasets (Batatia et al., 2025; Wood et al., 2025; Mazitov et al., 2025; Rhodes et al., 2025) sometimes replace message passing with local self-attention. While effective at capturing semi-local interactions, MPNNs cannot model true long-range effects, since distant atoms without intermediates never interact, and many steps reduce expressivity (Cai & Wang, 2020; Alon & Yahav, 2020; Nigam et al., 2022).

[2]For $p = 1$, i.e., electrostatics, it is common to speak of charges. For $p > 1$, for example dispersion ($p = 6$), 'coefficients' is more common.

Physics-based long-range models Many long-range models explicitly add physics-inspired terms to E that capture interactions decaying more slowly with distance. A common example of such a form is given by Eq. (1): $E^{\text{long}} = \sum_i q_i V_i$ where q_i are latent per-atom descriptors (e.g., partial charges or polarizabilities) predicted by the model and V_i is computed as an inverse power law of interatomic distances. Usually, these q_i are predicted directly as scalar functions of local atomic features (Unke et al., 2021a; Staacke et al., 2021; Loche et al., 2025; Kim et al., 2024; Ji et al., 2025; Kabylda et al., 2025; Cheng, 2025). A modification of this approach is to include the q_i in an equilibration scheme, thus allowing these descriptors to capture information otherwise missed due to their initial dependence on local environments (Ko et al., 2021; Pellegrini et al., 2023; Maruf et al., 2025). Grisafi & Ceriotti (2019); Huguenin-Dumittan et al. (2023) propose to use physics-inspired kernels to compute long-range features, mathematically equivalent to a multipole expansion, but find that higher-order features contain little additional information.

Other long-range models Some models avoid handcrafted corrections and instead learn long-range interactions directly. Ewald message passing augments GNNs with Fourier-space invariants (Kosmala et al., 2023), while fully connected approaches use all-to-all distances (Chmiela et al., 2018) or global attention (Unke et al., 2021a), though these lose efficiency or distance information. Linear-scaling attention with geometric embeddings (Frank et al., 2024a) enables global orientation exchange but needs symmetrization and has not yet been adapted to periodic systems. Alternatives include virtual nodes for global aggregation (Caruso et al., 2025) or message passing in spherical harmonics space (Frank et al., 2022). Despite approximations, most methods still struggle to bridge periodic and non-periodic systems. Some methods outside atomistic modeling also aim to capture long-range effects (Dwivedi et al., 2022; Bamberger et al., 2025; Moskalev et al., 2025; Zhdanov et al., 2025), but require further adaptation to include geometric information or periodicity.

4 Equivariant long-range message passing

Ewald summation, as introduced in Section 2 and Eq. (1), allows the efficient, and convergent in periodic systems, evaluation of a potential V_i at each atomic position, based on coefficients q_j associated with all other atoms j and a power of the inverse distances $1/r_{ij}^p$. This computation can be seen as a physics-inspired form of message passing (Eqs. (2) and (3)), with radial filter $1/r_{ij}^p$, neighbor feature q_i, and resulting message V_i; this correspondence was used by Grisafi & Ceriotti (2019) to compute general long-range features and cast in the language of MPNNs by Kosmala et al. (2023).

To bring it more in line with standard message passing, the operation can be carried out in parallel across an extra feature dimension, promoting q_i to an array $\boldsymbol{q}_i$. This allows to communicate more information, but in this form, this information is restricted to geometric invariants. We propose to promote q_i to an *equivariant* tensor instead: $\mathsf{Q}_{i,l,m}$. This results in equivariant messages, or potentials:

$$\mathsf{V}_{i,l,m} = \sum_{j=1}^{N} \sum_{\boldsymbol{n}\in\mathbb{Z}^3} \frac{\mathsf{Q}_{j,l,m}}{|\boldsymbol{r}_i - (\boldsymbol{r}_j + n_1\boldsymbol{c}_1 + n_2\boldsymbol{c}_2 + n_3\boldsymbol{c}_3)|^p}\,, \tag{4}$$

carrying out the Ewald summation over each spherical order l and component m in parallel.

To see that the result is equivariant, recall (see Appendix A) that adding two equivariant objects yields another equivariant object, and that multiplication with a prefactor, provided that the factor is shared across all entries for a given l, also retains equivariance. The prefactor $1/r_{ij}^p$ depends on the pair (i, j) and therefore varies across pairs; however, for any given pair, it is a single scalar that multiplies all m components equally within each l. Since the action of rotations on spherical features is a linear map along the m index, this scalar multiplication commutes with it. It is then easy to see that multiplying $\mathbf{Q}_i$ with a factor $1/r_{ij}^p$ yields an equivariant quantity, and that the sum in Eq. (4) also yields an equivariant, since all summands are equivariant objects. We argue and numerically confirm in Appendix B that a compensating background charge correction preserves equivariance.

Overall, this approach allows the global aggregation of equivariant messages in a way that is amenable to efficient implementation, scaling $O(N \log N)$ in periodic systems (see Appendix E) and—at least in principle, see Section 2—$O(N)$ in finite systems. Even more advanced computational methods can bring $O(N)$ scaling

to both. The method also naturally accommodates the long-range effects in physical systems, which decay asymptotically following a power law.

5 LOREM

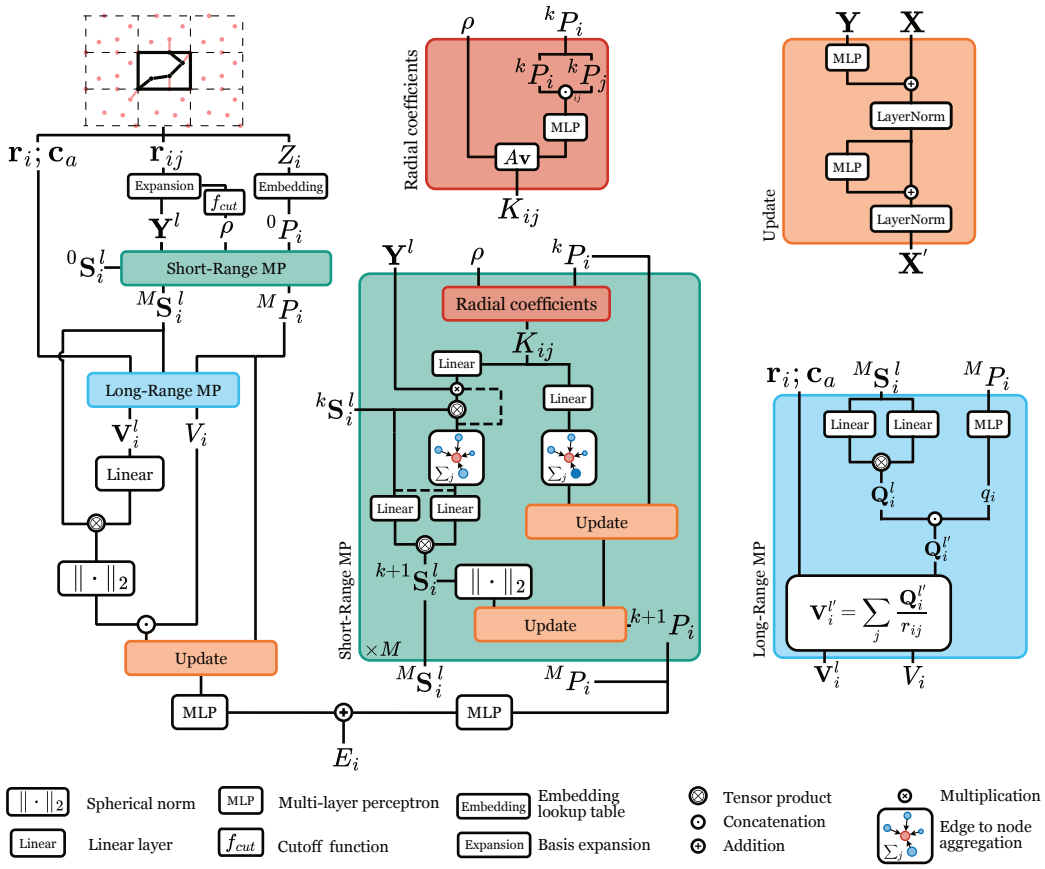

Figure 1: Sketch of the LOREM architecture. See main text for a detailed description.

LOREM follows the blueprint of equivariant MPNNs, making a few modifications: Scalar and spherical features are handled separately, following Frank et al. (2022; 2024b), and mixed using a variant of the power spectrum (Bartók et al., 2013) and a residual update block (He et al., 2015).

Overview LOREM, illustrated in Fig. 1 maps a point cloud of N atomic positions $\{\boldsymbol{r}_i\}$, potentially contained in a periodic unit cell $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$, and labeled with chemical species $\{Z_i\}$, to atomic energies $\{E_i\}$ which, when summed, yield the total potential energy E. Internally, the point cloud is processed as a graph, with edges between nodes, the atoms, defined by a cutoff radius r_c. Accordingly, edges correspond to vectors $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$ connecting atoms i and j. Node features are updated through either **short-range message passing** or **long-range message passing**. Spherical information is used to update scalar features by computing their **spherical norm** and passing it through an **update** block. Updates of scalar features are followed by a residual prediction of atomic energy contributions. We discuss the main components of LOREM below; specialized equivariant operations are described in more detail in Appendix A.

Short-range message passing Initial scalar features ${}^0\boldsymbol{P}_i$ are a learned embedding of chemical species. At each short-range message passing step k, edge features $\boldsymbol{K}_{ij}$ are obtained from distances r_{ij} and scalar features ${}^k\boldsymbol{P}_i$ and ${}^k\boldsymbol{P}_j$ through a **radial expansion**. These edge features are then linearly transformed twice, with different learned weight matrices: Once to yield scalar messages that are aggregated into scalar updates, and

again to yield pre-factors for the spherical harmonics $\mathbf{Y}_{ij}$ (**angular expansion**), which are combined with neighboring spherical features $\mathbf{S}_j$ in a tensor product.[3] The results are aggregated into a spherical message, which updates the spherical features via another **tensor product**, resulting in ${}^{k+1}\mathbf{S}_i$. The spherical norm of the updated spherical node features is then used to further update the scalar features, finally yielding ${}^{k+1}\boldsymbol{P}_i$. This process is repeated M times (the total number of short-range message passing steps).

Long-range message passing We use the method discussed in Section 4 to communicate equivariant information beyond the effective interaction cutoff of short-range message passing. To minimize computational cost, which is proportional to the number of channels over which Ewald summation is carried out, node features are transformed into low-dimensional charges: The scalar features ${}^{M}\boldsymbol{P}_i$ are transformed into a single per-atom charge q_i, and spherical features ${}^{M}\mathbf{S}_i$ are likewise, using a linear transformation followed by a self tensor product, transformed into $Q_{i,l,m}$ with a low $l_{\text{max, LR}}$ and a singular feature dimension. Unless otherwise noted, we use $l_{\text{max, LR}} = 2$. The scalar charge is concatenated with the $l = 0$ spherical charge, which yields a total of 10 charge channels for $l_{\text{max, LR}} = 2$.[4] After Ewald summation, which is carried out in parallel across l and m, the potentials are split back into the purely scalar V_i and the spherical $\mathbf{V}_i$. The spherical potentials are combined with spherical node features through a tensor product; the spherical norm of the result is concatenated with the scalar potential to update scalar representations. We find that using only p=1, i.e., the Coulomb interaction, rather than a set of different exponents, is sufficient in practice. Experiments for $l_{\text{max, LR}} = 0, 1, 2$ can be found in Appendix F.

Update block Updates $\boldsymbol{Y}$ to scalar features $\boldsymbol{X}$ use an update block consisting of multi-layer perceptrons (MLP) and layer normalization (LayerNorm) (Ba et al., 2016), following a residual structure (He et al., 2015):

$$\begin{aligned}
\boldsymbol{X} &\leftarrow \boldsymbol{X} + \text{MLP}(\boldsymbol{Y}) \\
\boldsymbol{X} &\leftarrow \text{LayerNorm}(\boldsymbol{X}) \\
\boldsymbol{X} &\leftarrow \boldsymbol{X} + \text{MLP}(\boldsymbol{X}) \\
\boldsymbol{X} &\leftarrow \text{LayerNorm}(\boldsymbol{X})\,.
\end{aligned}$$

Radial expansion LOREM processes information about the distance between two atoms r_{ij} through learned coefficients for linear transformations of an initial radial basis expansion of r_{ij}. Distances r_{ij} are first expanded in Bernstein polynomials multiplied with f_{cut}, the cosine cutoff function (extending from 0 to r_{c}), yielding initial radial features $\rho_{ij,c}$. These features are multiplied with a weight matrix $\boldsymbol{A}$, which is in turn obtained through a MLP (and subsequent reshaping operation) applied to the concatenated scalar atom features $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$. This allows the model to learn a radial basis based on the features of both atoms i and j. The resulting edge features are called $\boldsymbol{K}_{ij}$.

Angular expansion The vectors $\boldsymbol{r}_{ij}$ are expanded in spherical harmonics $\mathbf{Y}_{ij} = \mathbf{Y}(\hat{\boldsymbol{r}}_{ij})$, which are polynomials of vector components that produce outputs in different irreducible representations of SO(3) (Unke & Maennel, 2024). During message passing, they are multiplied with per-l coefficients: $Y'_{ij,l,m,c} = A_{ij,l,c} Y_{ij,l,m}$, where $\mathbf{A}_{ij}$ are obtained as linear transformation of $\boldsymbol{K}_{ij}$, followed by reshaping operation. Since the factors are shared across m, this preserves equivariance.

6 Experiments

We perform experiments on a number of existing benchmark tasks designed to probe the ability of MLIPs to model long-range interactions, comparing LOREM to short and long-range MLIPs. The experiments are divided into two parts: In Section 6.1, we compare LOREM with other models using standardized settings. We find that LOREM performs well, but also observe that most tasks can be solved by models that do not consider long-range interactions at all—the effective interaction range of typical message passing models is

[3] In the initial message passing step, no spherical node features are available, and therefore the tensor product with $\mathbf{S}_j$ is omitted. Initial spherical node features are obtained via a self tensor product (dotted lines) rather than a tensor product of the updates with the previous features.

[4] Two entries with l=0: One from the scalar features and one from the scalar component of the spherical features.

Table 1: Root mean squared errors (RMSE) for energy E and forces $\boldsymbol{f}$ for datasets and models used in Section 6.1. Where available, a held-out test set was used; otherwise, the validation set was used instead and indicated in the table. The lowest error is indicated in bold, the second lowest is underlined. The second line after each model name indicates the number of short-range (SR) message-passing steps and whether some form of long-range (LR) interactions are included in the model.[5] A mean absolute error (MAE) version of this table is provided in Appendix I.

Dataset		Lorem 1×SR+LR	Cace-Les 1×SR+LR	Mace 2×SR	Pet 2× SR	4G-NN 1×SR+LR	SpookyNet 6×SR+LR
MgO surface	E (meV/at)	**0.064**	<u>0.071</u>	0.376	0.210	0.219	0.107
(Validation)	$\boldsymbol{f}$ (meV/Å)	**4.076**	7.913	5.971	6.261	66.000	<u>5.337</u>
NaCl cluster	E (meV/at)	**0.112**	0.210	1.681	1.517	0.481	<u>0.135</u>
(Validation)	$\boldsymbol{f}$ (meV/Å)	<u>1.155</u>	9.784	40.219	42.438	32.780	**1.052**
Biodimers	E (meV/at)	**0.222**	<u>2.259</u>	7.793	6.758	–	–
	$\boldsymbol{f}$ (meV/Å)	**1.646**	<u>3.163</u>	16.150	16.470	–	–
Cumulene	E (meV/at)	<u>3.309</u>	17.803	12.592	**3.205**	–	–
	$\boldsymbol{f}$ (meV/Å)	<u>50.084</u>	147.616	104.318	**46.905**	–	–

sufficient. However, predictions break down beyond this interaction range. In Section 6.2, we probe this limit of short-range message passing. We find that while careful consideration of the number of message-passing steps and the cutoff radius is required to resolve long-range interactions with short-range models, Lorem can solve these benchmarks without adaptation. Additional experiments can be found in the appendix: Runtime benchmarks comparing Ewald and particle-mesh Ewald implementations of Lorem, demonstrating near-linear scaling up to 30 k atoms in Appendix E, ablations of $l_{\text{max, LR}}$ and the long-range block in Appendix F, as well as competitive performance on the larger ADAPT dataset (Dramko et al., 2025) in Appendix J.

Models We compare Lorem with a number of purely short-ranged MLIPs, Mace and Pet, as well as the recently introduced Cace-Les model that combines short-range message passing with a scalar long-range part, and 4G-NN, which includes a physics-based long-range energy contribution and charge equilibration. Additionally we include metrics for SpookyNet (Unke et al., 2021a), which predicts scalar partial charges and nuclear dipoles for long-range corrections, together with global attention. Full details on models can be found in Appendix D; approximate parameter counts are: Lorem ∼1 M, Cace-Les ∼70 k for MgO and ∼13 k for others, Mace ∼800 k, Pet ∼1 M, 4G-NN ∼5 k (estimated), SpookyNet ∼3 M (from (Blücher et al., 2023)).

The datasets and associated benchmark tasks used in our experiments are briefly introduced below; additional details are given in Appendix C. An overview of validation or test set (where available) metrics for all datasets is given in Table 1. Lorem performs very well across datasets, and is competitive with, or more accurate than, other models.

MgO surface This benchmark task is the first in a series designed by Ko et al. (2021) to highlight the need for long-range information in MLIPs. Illustrated in Fig. 2A, it consists of a magnesium oxide (MgO) surface on which a gold (Au_2) dimer is placed. Depending on the presence of an aluminum (Al) dopant deep inside the surface, the lowest-energy position of the gold dimer is either in the 'wetting' (flat), or 'non-wetting' (standing up), position. The benchmark consists of two parts: Correctly identifying the ordering between upright and flat geometries in the doped and undoped case, and reproducing the energy-distance curve for the non-wetting geometry, in particular the local minimum corresponding to the equilibrium distance.

NaCl cluster Similar to the presence of a dopant modifying the potential-energy surface for the gold dimer in the MgO surface task, this benchmark by Ko et al. (2021) relies on the presence or absence of a sodium atom at one end of a charged sodium chloride (NaCl) cluster changing the behavior of a sodium atom at the

[5] 4G-NN training requires DFT-computed charges, which are not available for the biodimers and cumulene datasets.

opposite end: Since one atom is removed while the charge remains constant, the charge must redistribute over the remaining atoms. The benchmark task consists of reproducing the location of the local minimum and the energy profile when moving the sodium atom farthest from the removed one, indicated in Fig. 2C.

Cumulene The cumulene benchmark task was proposed by Unke et al. (2021b) as an example of a long-range problem that is due to a non-local effect of the electronic structure of a molecule. This molecule consists of a chain of nine carbon atoms, with a pair of hydrogen atoms, the rotors, at the opposite ends. The orientation of the rotors determines the shape of the atomic orbitals for each carbon, which propagates along the chain to the other end; in the absence of bending or stretching, the energy is therefore fully determined by the relative orientation of the rotors. The task is to recover the energy profile of this rotation for an idealized fully extended geometry, illustrated in Fig. 3.

Biodimers This benchmark dataset by Huguenin-Dumittan et al. (2023) consists of pairs of relaxed organic molecules placed at distances of 4 Å to 15 Å from each other, illustrated in Fig. 4A. Depending on the chemical nature of the molecules, interactions between the molecules take different asymptotic power-law forms, ranging from charge-charge $1/r$ interactions to apolar-apolar $1/r^6$ interactions. Here, we simply compare energy and force prediction errors on test sets stratified by the dominating power-law interaction.

S_N2 reactions Finally, the S_N2 reactions benchmark was introduced by Frank et al. (2024a) to probe the ability of MLIPs to model the long-range interactions required to mediate gas-phase chemical reactions. It consists of the nucleophilic substitution of methyl halides by another halide ion: $X^- + H_3C{-}Y \rightarrow X{-}CH_3 + Y^-$ where X,Y = F, Cl, Br, or I. The benchmark task in this case is predicting the energy along the reaction coordinate, i.e., correctly modeling the potential energy profile as the two reactants approach one another, react, and separate again.

6.1 Standardized settings

We compare Lorem to other models using standardized settings: Two message-passing steps for short-range models, one message-passing step for long-range models, and $r_c = 5$ Å for most models (see Appendix D).

MgO surface The results for this experiment can be seen in Fig. 2B: Despite being designed to require long-range interactions, this benchmark task can be solved by both purely short-range and long-range message-passing models. In this case, the success of short-range message passing is due to the small size of this benchmark system: With an effective cutoff radius above 6 Å, centrally located atoms can 'see' the full system and hence, Mace and Pet can solve this task. All models also resolve the orientation preference for the Au_2 dimer between the doped and undoped surfaces.

NaCl cluster The results of this experiment, seen in Fig. 2D, are drastically different from the previous one: Here, only models with a long-range component are able to resolve the difference between $Na_9Cl_8^+$ and $Na_8Cl_8^+$. All such models show excellent agreement with reference values. The failure of short-range message-passing is due to the larger system size compared to the MgO surface: Here, effective cutoff radii exceeding 10.5 Å are required to solve the benchmark task.

Cumulene Resolving the cumulene energy profile, seen in Fig. 3, requires simultaneous knowledge of the orientation of both rotors at opposite ends of the molecule. This can be achieved in two ways: Through equivariant short-range message passing (Mace, Pet), provided that the chain is not too long (see Section 6.2), or through *equivariant* long-range message passing (Lorem). For this reason, Cace-Les cannot solve this benchmark: Scalar charges are not sufficiently expressive to communicate relative orientation. All equivariant models that are able to access this information can solve this benchmark, achieving good agreement with the reference data. Pet resolves the dihedral angle as well, but requires long training and an adaptation in the number of transformer layers to succeed at this benchmark (see Appendix D). We note that the sharp cusp at 180° is an artifact of the underlying reference method; it is a desirable behavior of MLIPs to smooth it out.

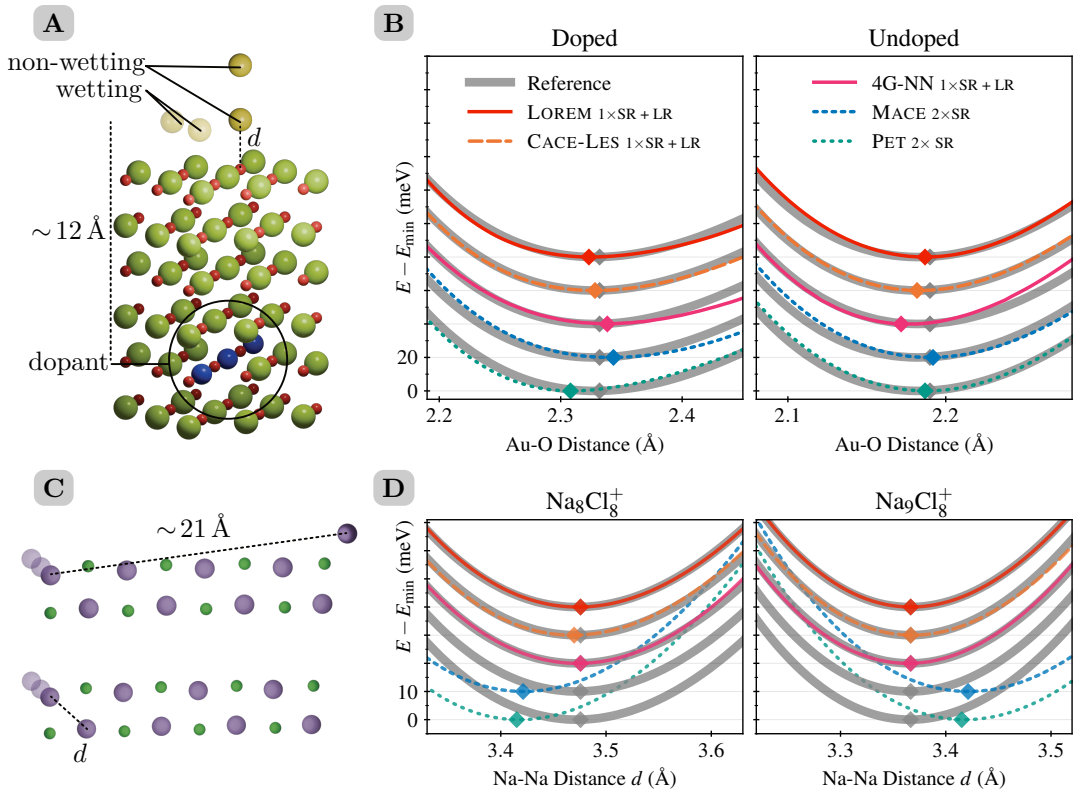

Figure 2: (**A**) Au_2 dimer on MgO surface, showing both wetting and non-wetting geometries, as well as the Al dopant. (**B**) Energy over distance d for the non-wetting geometry for the doped and undoped surface. The minima are indicated with a diamond symbol; the reference energy curve is drawn in grey. Offsets are added to distinguish the curves and the value at the minimum is subtracted. (**C**) $Na_9Cl_8^+$ (top) and $Na_8Cl_8^+$ (bottom) cluster, the moving atom is marked with transparent copies of itself, and the distance of interest is labeled with d. (**D**) Energy over distance for both clusters.

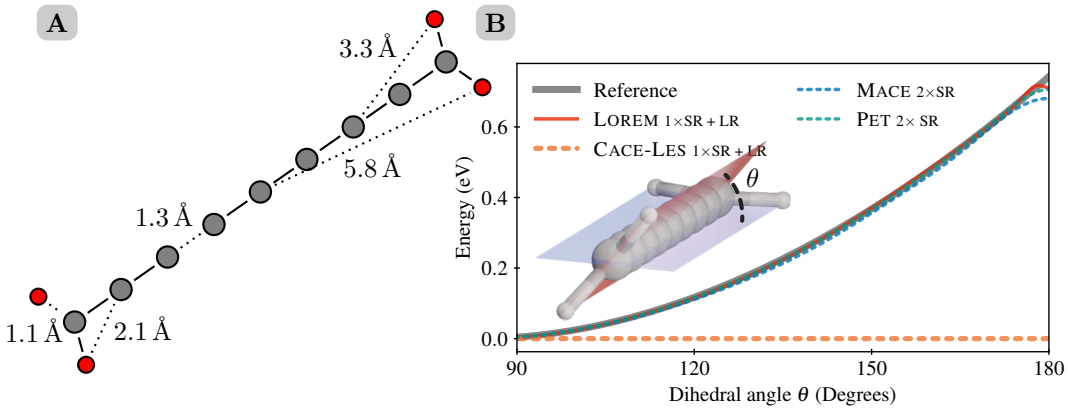

Figure 3: (**A**) Illustration of cumulene laid flat, indicating relevant distances between atoms. (**B**) Energy profile over a 90° rotation of one rotor. The minimum value of each curve is subtracted before plotting. The inset shows a 3D representation of cumulene, defining the dihedral angle θ.

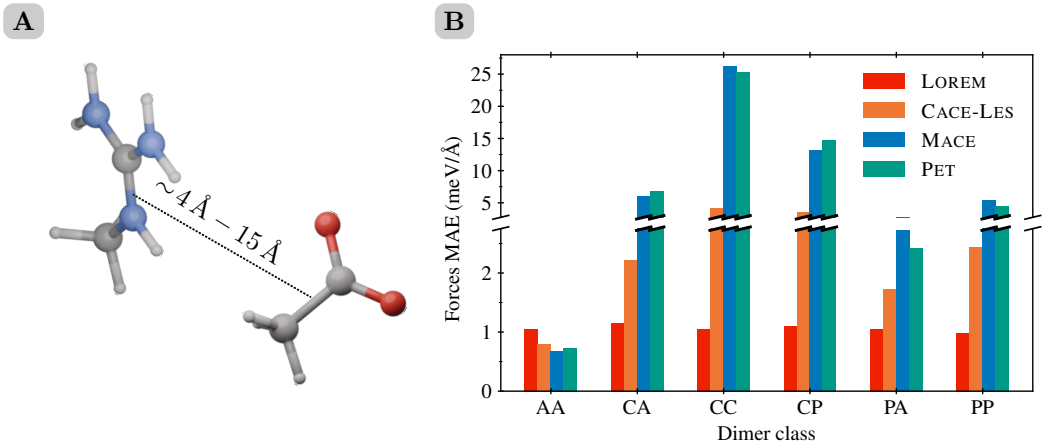

Figure 4: (**A**) Charge-charge pair from the biodimers dataset. (**B**) Mean absolute error on forces for different models on the different dimer classes: Apolar-apolar (AA), charge-apolar (CA), charge-charge (CC), charge-polar (CP), polar-apolar (PA), polar-polar (PP). Note that the vertical axis has been split at 2.75 meV/Å.

Biodimers Since the pairs of molecules in the biodimers benchmark are placed at separations up to 15 Å, much beyond the cutoff used for graph construction, high accuracy requires a long-range component. This is confirmed by the results of this experiment, seen in Fig. 4: The long range Lorem and Cace-Les models yield lower error than Mace and Pet. Predictive error varies between dimer classes, i.e., the expected type of inverse power-law interaction, for all models, with the exception for Lorem, which yields consistent, and in all but one class the highest, accuracy. It is important to note that Lorem uses $p = 1$ (Coulomb) for long-range message passing while different exponents describe the *underlying physics* of the dimer classes: The nonlinearity after the long-range message passing block may allow the model to correct this mismatch between exponents; see also the supplement of (Huguenin-Dumittan et al., 2023).

6.2 Limits of short-range message passing

In the previous section, we observed that the performance of models with short-range message passing strongly depends on a problem-specific choice of the effective interaction cutoff. The biodimers experiment also demonstrated that message passing fails when no intermediate atoms are present. To study these failure modes, we perform additional experiments using Lorem, varying cutoff and long-range message passing.

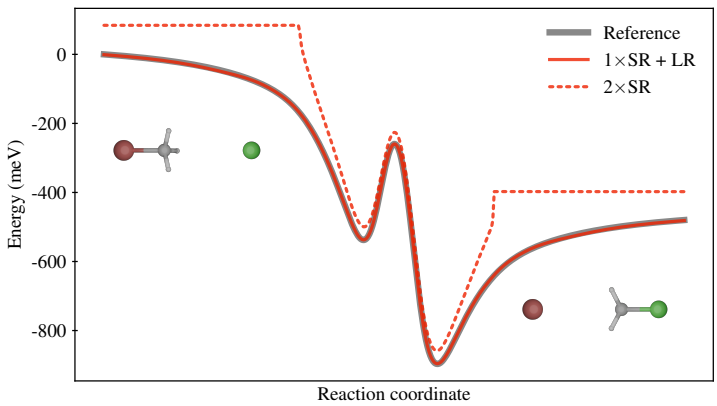

Figure 5: Energy over the reaction coordinate for the nucleophilic substitution reaction $Cl^- + H_3C{-}Br \rightarrow Cl{-}CH_3 + Br^-$; snapshots are shown as insets.

S_N2 reactions Solving the benchmark task for the S_N2 reactions dataset requires long-range interactions, since it involves modeling intra-molecular interactions over distances exceeding the typical cutoffs used for MLIPs. This is confirmed by Fig. 5, which compares the performance of Lorem with and without long-range message passing. The former can reproduce the energy over the course of the reaction with excellent accuracy, including the tails as the reactants approach and separate. The latter, which does not include long-range interactions, cannot account for the tails and consequently predicts a constant once the molecules separate more than 5 Å.

Table 2: Ability of different LOREM models, with and without long-range message passing and with different numbers of short-range message passing steps, to solve the cumulene benchmark task. A tick (✓) indicates yes, a cross (×) indicates no. Figure 10 shows the full curves.

Cutoff	1×SR + LR	2×SR + LR	1×SR	2×SR	3×SR	4×SR
2.5 Å	✓	✓	×	×	×	✓
3.0 Å	✓	✓	×	×	✓	✓
3.5 Å	✓	✓	×	✓	✓	✓

Cumulene with different cutoffs We probe the dependence of the performance of message-passing models on their hyperparameters by training a set of LOREM models with and without long-range message passing at different cutoffs (2.5 Å, 3 Å and 3.5 Å) and numbers of short-range message passing steps (1, 2, 3 and 4). Small cutoffs are chosen to simulate the longer chain lengths of real biomolecular systems. The results are shown in Table 2 with additional plots in Appendix G: In all cases, models that include long-range message passing are able to resolve the angle. Short-range models, on the other hand, can only resolve the angle in certain combinations of hyperparameters: One message passing step is never sufficient. For two, a minimum cutoff of 3.5 Å is required. For three message passing steps, 3.0 Å is required. Therefore, an effective cutoff substantially larger than the 5.8 Å distance between the central carbon atom in the chain and the hydrogen rotors is needed. This is because MPNNs rely on the input graph structure (see Fig. 3A) for information flow: Below $r_c = 2.6$ Å, connections in the graph only extend to nearest neighbors, with the exception of the rotors and the second-to-last carbon atoms. Consequently, at least four message passing steps are required. This example illustrates that short-range message passing is difficult to apply to this class of problems: Parameters have to be adapted to the chain length, with careful consideration of graph structure. In contrast, long-range message passing is robust, requiring no change in model hyper-parameters to solve this task.

7 Discussion

We introduced a simple yet effective equivariant long-range message passing scheme: Latent equivariant charges are predicted from local node features, and well-established techniques for evaluating inverse power-law potentials are used to efficiently compute long-range messages in a way that is convergent and well-defined for periodic systems. Building on this message-passing mechanism, we developed LOREM, an MLIP architecture that achieves consistently strong performance across benchmarks that require accurate long-range modeling.

By construction, our model assumes that interactions decay asymptotically with distance. While it is therefore not suited for learning truly global representations that have no notion of locality, this limitation is largely theoretical for physical systems: Electrostatics dominates most long-range behavior in realistic systems, and even other effects typically do not extend over arbitrary distances.

A more practical limitation arises in non-periodic systems, where the cost of a naive long-range message evaluation scales as $O(N^2)$. Although this is acceptable for small molecules and unit cells, this is a bottleneck for larger systems. However, linear-scaling methods for the evaluation of inverse power-law potentials are available: Fast multipole methods (Greengard & Rokhlin, 1987; Andy L Jones, 2020) or multi-level summation (Hardy et al., 2015; Buchner et al., 2025). Being able to leverage such methods is a key advantage of our proposed physics-inspired message-passing scheme.

We also investigated the capabilities of purely short-range message passing and found that, in many cases, it performs very well—even on datasets explicitly designed to require long-range interactions or charge equilibration. However, its success depends on matching the cutoff radius and number of message passing steps to the specific task, a process that can be both tedious and error-prone. More fundamentally, short-range methods cannot resolve interactions between distant atoms without intermediaries. Our augmentation with long-range message passing overcomes these limitations, providing robust results across different tasks without requiring changes to model architecture or its hyperparameters (cutoff radius, number of message-passing

steps, representation order). Training hyperparameters such as learning rate, optimizer, and loss weights are tuned per dataset, as is standard practice for MLIPs.

Finally, our results underscore a broader issue: the lack of challenging long-range benchmarks. Many current datasets are based on simplified, small-scale systems and fail to capture the complexity of real-world applications where long-range interactions are essential. Our deliberate focus on these benchmarks reflects the proof-of-concept nature of this work: On large, heterogeneous datasets, improvements in aggregate loss metrics can often be achieved by increasing the capacity of the short-range model, making it difficult to isolate and attribute gains to improved modeling of long-range physics (Huguenin-Dumittan et al., 2023). The controlled benchmarks we consider allow a clearer mechanistic validation of the proposed equivariant Ewald message passing. Nevertheless, as a step towards larger-scale evaluation, we present results on the ADAPT silicon point-defect dataset (Dramko et al., 2025) in Appendix J, where Lorem achieves competitive force accuracy and substantially better energy predictions compared to all baselines, using default model hyperparameters. Addressing the gap of challenging large long-range datasets is a key direction for future work, which requires the development of application-oriented benchmarks.

Reproducibility statement

Data, code, configuration files, scripts, and trained models are available at doi:10.5281/zenodo.19421697. Scripts include data pre-processing, model training, model evaluation, and the creation of figures and tables in this work. Hyperparameters are also described in Appendix D. The reference implementation of Lorem in JAX is available at `https://github.com/lab-cosmo/lorem-jax`, a PyTorch implementation as part of metatrain (Bigi et al., 2026) at `https://github.com/metatensor/metatrain`. An overview of code and data repositories can also be found at `https://marcel.science/lorem`.

Acknowledgements

MFL acknowledges funding from the German Research Foundation (DFG) under project number 544947822. MFL, MC, and PL acknowledge funding from the European Research Council (ERC) under the research and innovation programme (Grant Agreement No. 101001890-FIAMMA). ER, MC, and PL acknowledge funding from the NCCR MARVEL, funded by the Swiss National Science Foundation (SNSF, grant number 182892). The authors would like to thank Matthias Kellner, Filippo Bigi, J. Thorben Frank, Bingqing Cheng, and Niklas Schmitz for insightful discussions, and also thank Romain Greub, Luis Falke, and Harsh Talathi for contributing initial experiments on the ADAPT dataset.

References

Joseph W. Abbott, Carlos Mera Acosta, Alaa Akkoush, Alberto Ambrosetti, Viktor Atalla, Alexej Bagrets, Jörg Behler, Daniel Berger, Björn Bieniek, Jonas Björk, Volker Blum, Saeed Bohloul, Connor L. Box, Nicholas Boyer, Danilo Simoes Brambila, Gabriel A. Bramley, Kyle R. Bryenton, María Camarasa-Gómez, Christian Carbogno, Fabio Caruso, Sucismita Chutia, Michele Ceriotti, Gábor Csányi, William Dawson, Francisco A. Delesma, Fabio Della Sala, Bernard Delley, Robert A. DiStasio Jr, Maria Dragoumi, Sander Driessen, Marc Dvorak, Simon Erker, Ferdinand Evers, Eduardo Fabiano, Matthew R. Farrow, Florian Fiebig, Jakob Filser, Lucas Foppa, Lukas Gallandi, Alberto Garcia, Ralf Gehrke, Simiam Ghan, Luca M. Ghiringhelli, Mark Glass, Stefan Goedecker, Dorothea Golze, Matthias Gramzow, James A. Green, Andrea Grisafi, Andreas Grüneis, Jan Günzl, Stefan Gutzeit, Samuel J. Hall, Felix Hanke, Ville Havu, Xingtao He, Joscha Hekele, Olle Hellman, Uthpala Herath, Jan Hermann, Daniel Hernangómez-Pérez, Oliver T. Hofmann, Johannes Hoja, Simon Hollweger, Lukas Hörmann, Ben Hourahine, Wei Bin How, William P. Huhn, Marcel Hülsberg, Timo Jacob, Sara Panahian Jand, Hong Jiang, Erin R. Johnson, Werner Jürgens, J. Matthias Kahk, Yosuke Kanai, Kisung Kang, Petr Karpov, Elisabeth Keller, Roman Kempt, Danish Khan, Matthias Kick, Benedikt P. Klein, Jan Kloppenburg, Alexander Knoll, Florian Knoop, Franz Knuth, Simone S. Köcher, Jannis Kockläuner, Sebastian Kokott, Thomas Körzdörfer, Hagen-Henrik Kowalski, Peter Kratzer, Pavel Kůs, Raul Laasner, Bruno Lang, Björn Lange, Marcel F. Langer, Ask Hjorth Larsen, Hermann Lederer, Susi Lehtola, Maja-Olivia Lenz-Himmer, Moritz Leucke, Sergey Levchenko, Alan Lewis, O. Anatole von Lilienfeld, Konstantin Lion, Werner Lipsunen, Johannes Lischner, Yair Litman, Chi

Liu, Qing-Long Liu, Andrew J. Logsdail, Michael Lorke, Zekun Lou, Iuliia Mandzhieva, Andreas Marek, Johannes T. Margraf, Reinhard J. Maurer, Tobias Melson, Florian Merz, Jörg Meyer, Georg S. Michelitsch, Teruyasu Mizoguchi, Evgeny Moerman, Dylan Morgan, Jack Morgenstein, Jonathan Moussa, Akhil S. Nair, Lydia Nemec, Harald Oberhofer, Alberto Otero-de-la-Roza, Ramón L. Panadés-Barrueta, Thanush Patlolla, Mariia Pogodaeva, Alexander Pöppl, Alastair J. A. Price, Thomas A. R. Purcell, Jingkai Quan, Nathaniel Raimbault, Markus Rampp, Karsten Rasim, Ronald Redmer, Xinguo Ren, Karsten Reuter, Norina A. Richter, Stefan Ringe, Patrick Rinke, Simon P. Rittmeyer, Herzain I. Rivera-Arrieta, Matti Ropo, Mariana Rossi, Victor Ruiz, Nikita Rybin, Andrea Sanfilippo, Matthias Scheffler, Christoph Scheurer, Christoph Schober, Franziska Schubert, Tonghao Shen, Christopher Shepard, Honghui Shang, Kiyou Shibata, Andrei Sobolev, Ruyi Song, Aloysius Soon, Daniel T. Speckhard, Pavel V. Stishenko, Muhammad Tahir, Izumi Takahara, Jun Tang, Zechen Tang, Thomas Theis, Franziska Theiss, Alexandre Tkatchenko, Milica Todorović, George Trenins, Oliver T. Unke, Álvaro Vázquez-Mayagoitia, Oscar van Vuren, Daniel Waldschmidt, Han Wang, Yanyong Wang, Jürgen Wieferink, Jan Wilhelm, Scott Woodley, Jianhang Xu, Yong Xu, Yi Yao, Yingyu Yao, Mina Yoon, Victor Wen-zhe Yu, Zhenkun Yuan, Marios Zacharias, Igor Ying Zhang, Min-Ye Zhang, Wentao Zhang, Rundong Zhao, Shuo Zhao, Ruiyi Zhou, Yuanyuan Zhou, and Tong Zhu. Roadmap on Advancements of the FHI-aims Software Package, June 2025.

Uri Alon and Eran Yahav. On the Bottleneck of Graph Neural Networks and its Practical Implications. In *International Conference on Learning Representations*, October 2020. URL `https://openreview.net/forum?id=i80OPhOCVH2`.

Andy L Jones. pybbfmm, 2020. URL `https://www.github.com/andyljones/pybbfmm`.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer Normalization, July 2016.

Jacob Bamberger, Benjamin Gutteridge, Scott le Roux, Michael M. Bronstein, and Xiaowen Dong. On Measuring Long-Range Interactions in Graph Neural Networks. In *Forty-Second International Conference on Machine Learning*, June 2025. URL `https://openreview.net/forum?id=2fBcAOi8lO`.

Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Physical Review B*, 87(18):184115, May 2013. doi: 10.1103/PhysRevB.87.184115.

Ilyes Batatia, David P. Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Advances in Neural Information Processing Systems*, 35:11423–11436, December 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/hash/4a36c3c51af11ed9f34615b81edb5bbc-Abstract-Conference.html`.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, Janosh Riebesell, Xavier R. Advincula, Mark Asta, Matthew Avaylon, William J. Baldwin, Fabian Berger, Noam Bernstein, Arghya Bhowmik, Filippo Bigi, Samuel M. Blau, Vlad Cărare, Michele Ceriotti, Sanggyu Chong, James P. Darby, Sandip De, Flaviano Della Pia, Volker L. Deringer, Rokas Elijošius, Zakariya El-Machachi, Edvin Fako, Fabio Falcioni, Andrea C. Ferrari, John L. A. Gardner, Mikołaj J. Gawkowski, Annalena Genreith-Schriever, Janine George, Rhys E. A. Goodall, Jonas Grandel, Clare P. Grey, Petr Grigorev, Shuang Han, Will Handley, Hendrik H. Heenen, Kersti Hermansson, Cheuk Hin Ho, Stephan Hofmann, Christian Holm, Jad Jaafar, Konstantin S. Jakob, Hyunwook Jung, Venkat Kapil, Aaron D. Kaplan, Nima Karimitari, James R. Kermode, Panagiotis Kourtis, Namu Kroupa, Jolla Kullgren, Matthew C. Kuner, Domantas Kuryla, Guoda Liepuoniute, Chen Lin, Johannes T. Margraf, Ioan-Bogdan Magdău, Angelos Michaelides, J. Harry Moore, Aakash A. Naik, Samuel P. Niblett, Sam Walton Norwood, Niamh O'Neill, Christoph Ortner, Kristin A. Persson, Karsten Reuter, Andrew S. Rosen, Louise A. M. Rosset, Lars L. Schaaf, Christoph Schran, Benjamin X. Shi, Eric Sivonxay, Tamás K. Stenczel, Christopher Sutton, Viktor Svahn, Thomas D. Swinburne, Jules Tilly, Cas van der Oord, Santiago Vargas, Eszter Varga-Umbrich, Tejs Vegge, Martin Vondrák, Yangshuai Wang, William C. Witt, Thomas Wolf, Fabian Zills, and Gábor Csányi. A foundation model for atomistic materials chemistry. *The Journal of Chemical Physics*, 163(18):184110, November 2025. ISSN 0021-9606. doi: 10.1063/5.0297006.

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis

Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks, October 2018.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1), May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5.

Jörg Behler and Michele Parrinello. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters*, 98(14):146401, April 2007. ISSN 0031-9007, 1079-7114. doi: 10.1103/physrevlett.98.146401.

Filippo Bigi, Joseph W. Abbott, Philip Loche, Arslan Mazitov, Davide Tisi, Marcel F. Langer, Alexander Goscinski, Paolo Pegolo, Sanggyu Chong, Rohit Goswami, Pol Febrer, Sofiia Chorna, Matthias Kellner, Michele Ceriotti, and Guillaume Fraux. Metatensor and metatomic: Foundational libraries for interoperable atomistic machine learning. *The Journal of Chemical Physics*, 164(6):064113, February 2026. ISSN 0021-9606. doi: 10.1063/5.0304911.

Stefan Blücher, Klaus-Robert Müller, and Stefan Chmiela. Reconstructing Kernel-Based Machine Learning Force Fields with Superlinear Convergence. *Journal of Chemical Theory and Computation*, 19(14):4619–4630, July 2023. ISSN 1549-9618. doi: 10.1021/acs.jctc.2c01304.

M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–484, January 1927. ISSN 0003-3804, 1521-3889. doi: 10.1002/andp.19273892002.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: Composable transformations of Python+NumPy programs, 2018. URL `http://github.com/jax-ml/jax`.

Donald W. Brenner. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. *Physical Review B*, 42(15):9458–9471, November 1990. doi: 10.1103/PhysRevB.42.9458.

Florian Buchner, Johannes Schörghuber, Nico Unglert, Jesús Carrete, and Georg K. H. Madsen. msmJAX: Fast and Differentiable Electrostatics on the GPU in Python, October 2025.

Lori A Burns, John C Faver, Zheng Zheng, Michael S Marshall, Daniel GA Smith, Kenno Vanommeslaeghe, Alexander D MacKerell, Kenneth M Merz, and C David Sherrill. The biofragment database (bfdb): An open-data platform for computational chemistry analysis of noncovalent interactions. *The Journal of chemical physics*, 147(16), 2017.

Chen Cai and Yusu Wang. A Note on Over-Smoothing for Graph Neural Networks, June 2020.

Alessandro Caruso, Jacopo Venturin, Lorenzo Giambagli, Edoardo Rolando, Frank Noé, and Cecilia Clementi. Extending the RANGE of Graph Neural Networks: Relaying Attention Nodes for Global Encoding, February 2025.

Bingqing Cheng. Cartesian atomic cluster expansion for machine learning interatomic potentials. *npj Computational Materials*, 10(1):157, July 2024. ISSN 2057-3960. doi: 10.1038/s41524-024-01332-4.

Bingqing Cheng. Latent Ewald summation for machine learning of long-range interactions. *npj Computational Materials*, 11(1):80, March 2025. ISSN 2057-3960. doi: 10.1038/s41524-025-01577-7.

Stefan Chmiela, Huziel E. Sauceda, Klaus-Robert Müller, and Alexandre Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, 9(1):3887, September 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06169-2.

Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, June 1993. ISSN 0021-9606. doi: 10.1063/1.464397.

Michelangelo Domina, Joseph William Abbott, Paolo Pegolo, Filippo Bigi, and Michele Ceriotti. How unconstrained machine-learning models learn physical symmetries, March 2026.

Evan Dramko, Yihuang Xiong, Yizhi Zhu, Geoffroy Hautier, Thomas Reps, Christopher Jermaine, and Anastasios Kyrillidis. ADAPT: Lightweight, Long-Range Machine Learning Force Fields Without Graphs, September 2025.

Geneviève Dusson, Markus Bachmayr, Gábor Csányi, Ralf Drautz, Simon Etter, Cas van der Oord, and Christoph Ortner. Atomic cluster expansion: Completeness, efficiency and stability. *Journal of Computational Physics*, 454:110946, April 2022. ISSN 0021-9991. doi: 10.1016/j.jcp.2022.110946.

Vijay Prakash Dwivedi, Ladislav Rampášek, Mikhail Galkin, Ali Parviz, Guy Wolf, Anh Tuan Luu, and Dominique Beaini. Long Range Graph Benchmark. In *Thirty-Sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, June 2022. URL `https://openreview.net/forum?id=in7XC5RcjEn`.

P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3): 253–287, 1921. ISSN 1521-3889. doi: 10.1002/andp.19213690304.

Nikita Fedik, Roman Zubatyuk, Maksim Kulichenko, Nicholas Lubbers, Justin S. Smith, Benjamin Nebgen, Richard Messerly, Ying Wai Li, Alexander I. Boldyrev, Kipton Barros, Olexandr Isayev, and Sergei Tretiak. Extending machine learning beyond interatomic potentials for predicting molecular properties. *Nature Reviews Chemistry*, 6(9):653–672, September 2022. ISSN 2397-3358. doi: 10.1038/s41570-022-00416-3.

J. Thorben Frank, Stefan Chmiela, Klaus-Robert Müller, and Oliver T. Unke. Euclidean Fast Attention: Machine Learning Global Atomic Representations at Linear Cost, December 2024a.

J. Thorben Frank, Oliver T. Unke, Klaus-Robert Müller, and Stefan Chmiela. A Euclidean transformer for fast and stable machine learned force fields. *Nature Communications*, 15(1):6539, August 2024b. ISSN 2041-1723. doi: 10.1038/s41467-024-50620-6.

Thorben Frank, Oliver Unke, and Klaus-Robert Müller. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. *Advances in Neural Information Processing Systems*, 35:29400–29413, December 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/hash/bcf4ca90a8d405201d29dd47d75ac896-Abstract-Conference.html`.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional Message Passing for Molecular Graphs. In *International Conference on Learning Representations*, September 2019. URL `https://openreview.net/forum?id=B1eWbxStPH`.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 1263–1272. PMLR, July 2017. URL `https://proceedings.mlr.press/v70/gilmer17a.html`.

L Greengard and V Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73 (2):325–348, December 1987. ISSN 0021-9991. doi: 10.1016/0021-9991(87)90140-9.

Andrea Grisafi and Michele Ceriotti. Incorporating long-range physics in atomic-scale machine learning. *The Journal of Chemical Physics*, 151(20):204105, November 2019. ISSN 0021-9606. doi: 10.1063/1.5128375.

David J. Hardy, Zhe Wu, James C. Phillips, John E. Stone, Robert D. Skeel, and Klaus Schulten. Multilevel Summation Method for Electrostatic Force Evaluation. *Journal of Chemical Theory and Computation*, 11 (2):766–779, February 2015. ISSN 1549-9618. doi: 10.1021/ct5009075.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015.

R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. CRC Press, Boca Raton, March 2021. ISBN 978-0-367-80693-4. doi: 10.1201/9780367806934.

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, November 2023. doi: 10.1021/acs.jpclett.3c02375.

Yajie Ji, Jiuyang Liang, and Zhenli Xu. Machine-Learning Interatomic Potentials for Long-Range Systems, February 2025.

Keller Jordan, Yuchen Jin, Vlado Boza, You Jiacheng, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. Muon: An optimizer for hidden layers in neural networks, 2024. URL `https://kellerjordan.github.io/posts/muon/`.

Adil Kabylda, J. Thorben Frank, Sergio Suarez Dou, Almaz Khabibrakhmanov, Leonardo Medrano Sandonas, Oliver T. Unke, Stefan Chmiela, Klaus-Robert Müller, and Alexandre Tkatchenko. Molecular Simulations with a Pretrained Neural Network and Universal Pairwise Force Fields, June 2025.

Dongjin Kim, Daniel S. King, Peichen Zhong, and Bingqing Cheng. Learning charges and long-range interactions from energies and forces, December 2024.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017.

Tsz Wai Ko, Jonas A. Finkler, Stefan Goedecker, and Jörg Behler. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nature Communications*, 12(1):398, January 2021. ISSN 2041-1723. doi: 10.1038/s41467-020-20427-2.

Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. Ewald-based Long-Range Message Passing for Molecular Graphs. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 17544–17563. PMLR, July 2023. URL `https://proceedings.mlr.press/v202/kosmala23a.html`.

Marcel F Langer, Sergey N Pozdnyakov, and Michele Ceriotti. Probing the effects of broken symmetries in machine learning. *Machine Learning: Science and Technology*, 5(4):04LT01, October 2024. ISSN 2632-2153. doi: 10.1088/2632-2153/ad86a0.

Philip Loche, Kevin K. Huguenin-Dumittan, Melika Honarmand, Qianjun Xu, Egor Rumiantsev, Wei Bin How, Marcel F. Langer, and Michele Ceriotti. Fast and flexible long-range models for atomistic machine learning. *The Journal of Chemical Physics*, 162(14):142501, April 2025. ISSN 0021-9606. doi: 10.1063/5.0251713.

Moin Uddin Maruf, Sungmin Kim, and Zeeshan Ahmad. Equivariant Machine Learning Interatomic Potentials with Global Charge Redistribution, March 2025.

Arslan Mazitov, Filippo Bigi, Matthias Kellner, Paolo Pegolo, Davide Tisi, Guillaume Fraux, Sergey Pozdnyakov, Philip Loche, and Michele Ceriotti. PET-MAD, a lightweight universal interatomic potential for advanced materials modeling, August 2025.

Artem Moskalev, Mangal Prakash, Junjie Xu, Tianyu Cui, Rui Liao, and Tommaso Mansi. Geometric Hyena Networks for Large-scale Equivariant Learning. In *Forty-Second International Conference on Machine Learning*, June 2025. URL `https://openreview.net/forum?id=jJRkkPr474`.

Jigyasa Nigam, Sergey Pozdnyakov, Guillaume Fraux, and Michele Ceriotti. Unified theory of atom-centered representations and message-passing machine-learning schemes. *The Journal of Chemical Physics*, 156(20): 204115, May 2022. ISSN 0021-9606. doi: 10.1063/5.0087042.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://papers.nips.cc/paper_files/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html`.

Franco Pellegrini, Ruggero Lot, Yusuf Shaidu, and Emine Küçükbenli. PANNA 2.0: Efficient neural network interatomic potentials and new architectures. *The Journal of Chemical Physics*, 159(8):084117, August 2023. ISSN 0021-9606. doi: 10.1063/5.0158075.

John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 77(18):3865–3868, October 1996. doi: 10.1103/PhysRevLett.77.3865.

Sergey Pozdnyakov and Michele Ceriotti. Smooth, exact rotational symmetrization for deep learning on point clouds. *Advances in Neural Information Processing Systems*, 36:79469–79501, December 2023. URL `https://proceedings.neurips.cc/paper_files/paper/2023/hash/fb4a7e3522363907b26a86cc5be627ac-Abstract-Conference.html`.

E. Prodan and W. Kohn. Nearsightedness of electronic matter. *Proceedings of the National Academy of Sciences*, 102(33):11635–11638, August 2005. doi: 10.1073/pnas.0505436102.

Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: Atomistic simulation at scale, April 2025.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://papers.nips.cc/paper_files/paper/2017/hash/303ed4c69846ab36c2904d3ba8573050-Abstract.html`.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 9377–9388. PMLR, July 2021. URL `https://proceedings.mlr.press/v139/schutt21a.html`.

Tess E. Smidt. Euclidean Symmetry and Equivariance in Machine Learning. *Trends in Chemistry*, 3(2):82–85, February 2021. ISSN 2589-5974. doi: 10.1016/j.trechm.2020.10.006.

Carsten G. Staacke, Hendrik H. Heenen, Christoph Scheurer, Gábor Csányi, Karsten Reuter, and Johannes T. Margraf. On the Role of Long-Range Electrostatics in Machine-Learned Interatomic Potentials for Complex Battery Materials. *ACS Applied Energy Materials*, 4(11):12562–12569, November 2021. doi: 10.1021/acsaem.1c02363.

J. Tersoff. New empirical approach for the structure and energy of covalent systems. *Physical Review B*, 37 (12):6991–7000, April 1988. doi: 10.1103/PhysRevB.37.6991.

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, February 2018. URL `https://arxiv.org/abs/1802.08219v3`.

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. ISSN 0031-9007, 1079-7114. doi: 10.1103/physrevlett.102.073005.

Oliver T. Unke and Hartmut Maennel. E3x: E(3)-Equivariant Deep Learning Made Easy, November 2024.

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, June 2019. ISSN 1549-9618. doi: 10.1021/acs.jctc.9b00181.

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 2021a. ISSN 2041-1723. doi: 10.1038/s41467-021-27504-0.

Oliver T. Unke, Stefan Chmiela, Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine Learning Force Fields. *Chemical Reviews*, 121 (16):10142–10186, August 2021b. ISSN 0009-2665. doi: 10.1021/acs.chemrev.0c01111.

Brandon M. Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R. Kitchin, Daniel S. Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Ammar Rizvi, Sushree Jagriti Sahoo, Zachary W. Ulissi, and C. Lawrence Zitnick. UMA: A Family of Universal Models for Atoms, June 2025.

Tian Xie and Jeffrey C. Grossman. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters*, 120(14):145301, April 2018. doi: 10.1103/PhysRevLett.120.145301.

Yang You, Jing Li, Sashank Reddi, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, and Cho-Jui Hsieh. Large Batch Optimization for Deep Learning: Training BERT in 76 minutes. In *International Conference on Learning Representations*, September 2019. URL `https://openreview.net/forum?id=Syx4wnEtvH`.

Maksim Zhdanov, Max Welling, and Jan-Willem van de Meent. Erwin: A Tree-based Hierarchical Transformer for Large-scale Physical Systems. In *Forty-Second International Conference on Machine Learning*, June 2025. URL `https://openreview.net/forum?id=MrphqqwnKv`.

A Equivariant modules in Lorem

As discussed in Section 2, only certain operations can be applied to equivariant features $\mathbf{S}$ without disrupting equivariance. We briefly discuss the operations used in Lorem; a more thorough discussion can be found in Smidt (2021); Unke & Maennel (2024). All operations are implemented using the `e3x` library (Unke & Maennel, 2024). We work with spherical features $\mathbf{S}$, dropping the atom index i, as operations are typically broadcast across i or pairs of atoms.

$\mathbf{S}$ is a three-dimensional tensor $S_{l,m,c}$ in which l enumerates the order of the irreducible representation of SO(3), m the $2l+1$ components of that irreducible representation, and c the channel (feature) dimension. For a fixed l and c, $\mathbf{S}_{l,:,c}$ can be thought of as a generalised Cartesian vector. Indeed, for $l = 1$ it is a vector in three-dimensional space.

Prerequisites A rotation $g \in \mathrm{SO}(3)$ applied to all inputs causes a corresponding rotation of the spherical features $\mathbf{S}$ expressed in terms of irreducible representations. For a set of features with order l, this rotation is represented by a matrix $\boldsymbol{R}(g) \in \mathbb{R}^{2l+1\times 2l+1}$ and acts on these features through matrix multiplication along the m index,

$$\mathbf{S}_{l,m,:} \longrightarrow_g \sum_{m'=-l}^{l} R_{m,m'}(g) S_{l,m',:}\,. \tag{5}$$

In other word, the action of rotations on spherical features is a linear map.

Addition, Multiplication, Linear layers Since the action of rotations is linear, adding two spherical features of equal l together does not disrupt equivariance. By the same logic, any multiplication of spherical features that is broadcast along the m index, i.e., that scales spherical features of order l equally, is permissible. Therefore, a linear layer can be applied to spherical features, provided it is only applied to the channel index, and broadcast along m. We therefore define a learned linear transformation as

$$\mathrm{Linear}(\mathbf{S}_{l,:,c}) = \sum_{c'} W^{l}_{c,c'} \mathbf{S}_{l,:,c'} \tag{6}$$

with per-l learned weights $\boldsymbol{W}^l$. A bias term could be applied to $l = 0$, i.e., the scalar part of the spherical features, but we choose not to.

Tensor products Two spherical features of order l_1 and l_2 can be combined into a new spherical feature with l_3 using a specialized tensor product

$$\mathrm{Tensor}(\mathbf{S}_{l_1,m_1,:}, \mathbf{Q}_{l_2,m_2,:}) = \mathbf{U}_{l_3,m_3,:} = w^{l_1,l_2,l_3}_{:} \sum_{m_1,m_2} C^{l_1,l_2,l_3}_{m_1,m_2,m_3} \mathbf{S}_{l_1,m_1,:} \mathbf{Q}_{l_2,m_2,:} \tag{7}$$

$\mathbf{C}^{l_1,l_2,l_3}$ are the Clebsch-Gordan coefficients, and $\boldsymbol{w}^{l_1,l_2,l_3} \in \mathbb{R}^c$ is a per-channel weight vector for a given combination of l_1, l_2, l_3. This tensor product can be carried out for every valid combination $|l_1 - l_2| \leq l_3 \leq |l_1 + l_2|$. In Lorem, unless specified otherwise, tensor products are only carried out to a fixed maximum dimension $l_{\max}$, which all spherical features have in common, not the maximum possible one. Some operations, for example the preparation of the equivariant message-passing block, perform a tensor product only to a specified target order. If both inputs of the tensor product are the same feature, we call the operation a self-tensor product. This operation increases the body-order of the representation (Dusson et al., 2022; Batatia et al., 2022).

Spherical Norm To predict invariant quantities, we require a way to extract invariant information from spherical, equivariant, features. One way to do achieve this is a tensor product to target order $l = 0$. Another, which is used in Lorem, is to take the norm of each irreducible representation

$$\mathrm{Norm}(\mathbf{S}) = \sqrt{(2l+1)^{1/2} \sum_{m} S^2_{l,m,c}}\,. \tag{8}$$

We found empirically that the prefactor $(2l+1)^{1/2}$ helps reduce variance across l. As opposed to a tensor product to $l=0$, this operation, inspired by the power spectrum (Bartók et al., 2013), keeps the norms per l separate; a tensor product would linearly combine all l into one.

B Invariance of Lorem

As explained in Section 4 and Appendix A, the potential at each atom, $\mathbf{V}_i$ is equivariant because the long-range message passing step consists of equivariant operations: Addition and multiplication with a factor shared per l. In practical implementations of Ewald summation in periodic systems, there is one extra step: To prevent divergences, the total charge must be zero, which can be done either by simply subtracting the sum of the total charge from each charge, or equivalently, by subtracting an analytical correction from the potentials. Both approaches are equivalent, and since summation is equivariant, also do not break invariance. For this reason, the total procedure retains equivariance.

To numerically confirm these considerations, we compute the mean absolute error of energy predictions over a degree $L=3$ Lebedev grid of rotations (Domina et al., 2026), plus inversions, with respect to the unrotated case, for the first 5 structures of the MgO surface validation set. The errors are in line with precision expectations: For single precision, $1.031 \cdot 10^{-6}$ meV and for double precision $5.913 \cdot 10^{-15}$ meV.

C Dataset construction

MgO surface The MgO surface dataset from Ko et al. (2021) is used without modification. A representative snapshot of the unit cell is shown in Fig. 2A. The dataset is built from four configuration types, each derived from a distinct initial structure:

1. A pure MgO surface with the Au_2 dimer oriented perpendicular to the surface,
2. A pure MgO surface with the Au_2 dimer oriented parallel to the surface,
3. An Al-doped MgO surface with the Au_2 dimer perpendicular to the surface, and
4. An Al-doped MgO surface with the Au_2 dimer parallel to the surface.

Each of these initial configurations was first geometry optimized. For the two perpendicular, 'non-wetting', cases (1 and 3), the distance between the lower Au atom and the O atom directly beneath it was systematically varied, displacing the Au_2 dimer as a whole. From these distance-dependent samples, a subset was randomly selected. To introduce structural diversity, Gaussian noise was applied to each configuration: a standard deviation of 0.02 Å for atoms in the MgO substrate and 0.1 Å for the gold cluster. After perturbation, 1250 structures were selected from each of the four configuration types, yielding a total of 5000 samples. For these, energies and forces were computed using the FHI-aims code (Abbott et al., 2025) and the Perdew–Burke–Ernzerhof (PBE) functional (Perdew et al., 1996). A random 90/10 train–validation split was used, as reported in Ko et al. (2021). Additional energy–distance curves with equal spacing were constructed for the perpendicular cases and are shown in Fig. 2B.

NaCl cluster The NaCl cluster dataset from Ko et al. (2021) is used without modification. First, the $Na_9Cl_8^+$ cluster was optimized in vacuum. A snapshot is shown in the top panel of Fig. 2C. From this geometry, the Na atom farthest from all other Na atoms was removed, yielding the $Na_8Cl_8^+$ cluster shown in the bottom panel of Fig. 2C. From these two structures, additional configurations were created by varying the distance between a selected pair of Na atoms along the line connecting them. In Fig. 2C, this moving atom is illustrated by transparent copies along its trajectory. Training datasets for both $Na_9Cl_8^+$ and $Na_8Cl_8^+$ were constructed by randomly sampling configurations along these trajectories. Gaussian noise with a standard deviation of 0.05 Å was applied to the atomic coordinates, resulting in 2500 perturbed structures for each molecule and a total of 5000 instances. Energies and forces were computed using the FHI-aims code and the PBE functional. A random 90/10 train–validation split was used, as reported by Ko et al. (2021). Additional energy–distance curves with equal spacing were constructed and are shown in Fig. 2D.

Cumulenes The cumulene dataset from Unke et al. (2021b) is used without modification. The geometry of the linear molecule is shown in Fig. 3A. The dataset contains 4500 randomly sampled cumulene structures with nine carbon atoms, divided into training, validation, and test sets with 2000/500/2000 instances, respectively. The energy curves shown in Fig. 3B are based on a controlled subset in which only one terminal carbon atom is rotated, while the rest of the molecule remains fixed. A visualization of this rotational motion is provided in the inset of Fig. 3B.

Biodimers The Biodimers dataset from Huguenin-Dumittan et al. (2023); Burns et al. (2017) is used without modification. It consists of 2291 relaxed organic sidechain–sidechain fragments, including small molecules such as ethanol, acetamide, and others. Based on molecular properties, the dataset is divided into six categories: each molecule is classified as either polar, apolar, or charged, resulting in the following dimer types—polar–polar (PP), polar–apolar (PA), charged-polar (CP), apolar–apolar (AA), charged-apolar (CA), and charged–charged (CC). A representative CC dimer is shown in Fig. 4A. The initial separation between molecules reflects their positions in protein structures. From this configuration, the intermolecular distance is incrementally increased up to 15 Å, resulting in a total of 29 783 dimer instances. Instances where the final separation exceeds the initial distance by more than 4 Å (13 743 samples) are designated as the test set. The remaining 16 040 instances form the training set. Energies and forces were computed using the FHI-aims code and the HSE06 hybrid functional. For each of the six dimer types, a random 80/20 train–validation split was applied. The resulting subsets were then merged into a single training set and a single validation set.

S_N2 reactions The S_N2 reactions dataset from Unke & Meuwly (2019); Frank et al. (2024a) is used without modification. It contains molecular structures in vacuum that model nucleophilic substitution (S_N2) reactions. The dataset includes molecules of the types XCH_3Y^-, CH_3X, HX, CHX, CH_2X, XY, X, and Y, for all possible combinations of $X, Y \in \{F, Cl, Br, I\}$. A representative reaction coordinate with corresponding snapshots is shown in Fig. 5. Additional species such as H_2, CH_2, and CH_3 are also included in the dataset. Training configurations were generated using ab initio molecular dynamics simulations at 5000 K, with a time step of 0.1 fs. Full computational details, including the level of theory, are provided in the original publications. The dataset is randomly split into 405 000 training, 5000 validation, and 42 708 test samples.

D Model description and training setup

For all models, training parameters, and in some cases, model parameters, vary slightly between datasets. Where we were unable to use models from previous work, we tuned parameters to minimize the error on the validation set,[6] aiming to perform a similar number of experiments, on the order of ten, per model and experiment. With this, we aim to report results that reflect a practical degree of hyper-parameter tuning.

Lorem For all experiments, except the ones where variations were tested, the same LOREM model architecture (Section 5) was used: A cutoff radius of 5 Å, $l_{\max} = 6$ for spherical features and $l_{\max,\ \mathrm{LR}} = 2$ for the long-range message passing, 128 scalar features, 8 channels for spherical features, and 32 radial basis functions. We perform only the initial short-range message passing step; performing additional steps typically increases accuracy but hinders comparison of long-range expressivity. This model has 1 021 198 learnable parameters. It was implemented using the `e3x` library (Unke & Maennel, 2024). Training is performed entirely in `float32` precision; while we find that reduced precision has only a minor effect on training dynamics, it can significantly alter validation and test results.

Training parameters vary slightly between datasets; the exact parameters can be found in Table 3. We report errors for the better out of two runs with different seeds; the presented conclusions hold for both models.

In all cases, except NaCl[7], the learning rate was decayed linearly after 10 epochs, from the indicated starting value to $1 \cdot 10^{-6}$; the ADAM (Kingma & Ba, 2017) and LAMB (You et al., 2019) optimizers were used. The loss function was a simple mean squared error, with the energy residuals normalized by number of atoms.

[6] We note that minimizing the validation error does not always yield the best qualitative agreement: In the cumulene dataset, it does not readily correlate with ability to predict the dihedral curve in Fig. 3—while all models below a certain error threshold are able to resolve the task, the qualitative agreement varies.

[7] Here, exponential learning rate decay was employed.

Dataset	Optimizer	Initial LR	Epochs	Batch size	E weight	$\boldsymbol{f}$ weight	Time
MgO surface	ADAM	$4 \cdot 10^{-4}$	4000	32	1000	1.0	17 h
Biodimers	ADAM	$1 \cdot 10^{-4}$	4000	32	0.5	0.5	42 h
Cumulene	LAMB	$1 \cdot 10^{-3}$	2000	32	0.5	0.5	2 h
NaCl cluster	ADAM	$1 \cdot 10^{-3}$	8000	64	0.5	0.5	7 h
S_N2 reactions	ADAM	$1 \cdot 10^{-4}$	500	32	0.5	0.5	17 h

Table 3: Training settings for LOREM for different datasets. Training times are given for a single Nvidia H100 SXM5 GPU.

The squared residuals were averaged over the whole batch, including over atoms and components in the case of forces, and then summed and weighted with a factor. The checkpoint with the lowest summed R^2 of energy and forces, evaluated on the validation set, was used for experiments. Training times are given for the entire run, not the time until the best checkpoint.

For NaCl and AuMgO, prior work (Kim et al., 2024; Ko et al., 2021) reports error directly on the validation set, as the benchmark was originally intended as an overfitting exercise. We follow this practice, but verify in Appendix H that, due to the statistical uniformity of the dataset, there is no significant difference to tuning hyper-parameters systematically on an inner train/validation split, keeping the data used for Table 1 held out.

Mace The MACE MLIP (Batatia et al. (2022)) is an equivariant message-passing neural network as described in Section 2. It is used with the standard setting of two message-passing steps, i.e., an effective cutoff of 10 Å. Training generally used default hyperparameters, with some adjustments made on hyperparameters related to training dynamics. Specifically, the cumulene system was trained using the default energy-to-forces loss weight ratio of 1:100 and '128x0e + 128x1o + 128x2e' hidden irreps. All other systems were trained using a 1:10 ratio and '128x0e + 128x1o' hidden irreps. The MACE model has 848 528 parameters.

All experiments, except those involving cumulenes and the NaCl cluster, used the SWA protocol, which swaps the loss weights between energy and forces at a specified epoch. This epoch was set where the energy RMSE plateaued, with the subsequent training continued until loss saturation. All models were trained with $l_{\max} = 2$, a batch size of 32, and utilized cuEquivariance acceleration.

Pet PET (Pozdnyakov & Ceriotti (2023)) is an unconstrained transformer model, consisting of multiple edge-to-edge transformer layers within local neighborhoods followed by message passing. Similar to MACE, it is also used with two message-passing steps and an effective cutoff of 10 Å. To ensure approximate rotational invariance, it is trained with data augmentation. No inference-time symmetrization is used in our experiments.

Unless otherwise specified, we used default hyperparameters (cutoff radius of 5 Å with cosine smoothing over the outermost 0.2 Å, $d_{\text{PET}} = 128$, $d_{\text{head}} = 128$, $d_{\text{feedforward}} = 512$, with 8 heads per attention layer, and 2 attention layers per GNN layer), with some adjustments made related to training dynamics. Models were trained using an epoch-based scheduler, which halved the learning rate after 250 epochs. This applies to all datasets except biodimers. For biodimers, a `ReduceLROnPlateau` scheduler was used instead, reducing the learning rate by 20% if the loss did not improve for 100 consecutive epochs. To improve training stability, biodimers training also employed gradient clipping with a maximum gradient norm of 5. Every training run included 10 warmup epochs, during which the learning rate was linearly increased from zero to the preset learning rate of 1×10^{-4}.

For the biodimers and NaCl cluster datasets, the targets were normalized by their standard deviation in the training set. The energy-to-forces loss weight ratio was set to 1:1 for biodimers and MgO surface datasets, while the NaCl cluster dataset used a ratio of 1:10.

For cumulene, PET was trained for 20 100 epochs with a batch size of 16, equal energy and forces weight, and a maximum learning rate of $2 \cdot 10^{-4}$. The learning rate was increased linearly from 0 over 100 epochs, and then reduced by 10% every 500 epochs over the entire training run. Architecturally, 4 attention layers

and cosine cutoff function was employed to resolve the cumulenes. We found that increasing the number of attention layers, and adapting the cutoff function, were critical to resolve this benchmark task. This PET model consequently has 1 836 804 parameters, while other models have 1 043 716.

Cace-Les The CACE-LES model was originally presented in Cheng (2024), with an additional modification introducing a long-range component in Cheng (2025). Similar to MACE, CACE-LES is an equivariant message-passing neural network. After short-range message passing, invariant scalar pseudo-charges are predicted and passed to the long-range part of Ewald summation; the resulting energy contribution is added to the energy predictions of the short-range message-passing model. We use this model with the recommended setting of one message-passing step and, depending on the system, the following corresponding effective cutoff radii: 5.5 Å for the MgO surface, 5.29 Å for the NaCl cluster—these are the settings used in the original paper Kim et al. (2024)—and we chose a cutoff of 5 Å for biodimers and cumulene.

For the MgO surface and NaCl cluster datasets, models were taken from Kim et al. (2024). For the cumulenes and biodimers datasets, the hyperparameters were as follows: 6 Bessel radial functions, $c = 8$, $l_{\max} = 3$, $\nu_{\max}$, $N_{\text{embedding}} = 2$, one message-passing layer, one-dimensional hidden variable, $\sigma = 1$, and $dl = 2$. Training followed the example in the original paper: the first 200 epochs used an energy loss weight of 0.1 and a forces loss weight of 1000, after which the energy loss weight was changed to 1, 10, and 1000 every 100 epochs, yielding a total of 500 epochs. The learning rate was 5×10^{-3}, with a step learning rate schedule decreasing it by a factor of 2 every 20 steps. The model for the MgO surface has 72 340 parameters, all others 12 808.

4G-NN The 4G-NN model was designed to tackle the benchmarks introduced by Ko et al. (2021), and introduced in that work. It consists of a shallow neural network acting on rotationally invariant features, predicting both a local energy contribution and an electronegativity, which is then used in a charge equilibration procedure that globally redistributes charges to minimize an energy expression. This process can be seen as a physics-inspired long-range message passing scheme iterated until a fixed point is reached. As the model requires charge labels to train, which are not available for all datasets, we did not train this model for our experiments but instead include results from Ko et al. (2021), which are only available for the NaCl cluster and MgO surface datasets. The cutoffs used in the original paper are as follows: 4.23 Å for the MgO surface and 5.29 Å for the NaCl cluster. We estimate around 5 k parameters for these models.

SpookyNet SpookyNet (Unke et al., 2021a) is an equivariant MLIP that predicts scalar partial charges and nuclear dipoles, which are used to compute long-range electrostatic and dispersion corrections via Ewald summation. Additionally, the model includes global attention without geometric information. The model has approximately 3 M parameters (Blücher et al., 2023). Results for the MgO surface and NaCl cluster datasets are taken from Ko et al. (2021).

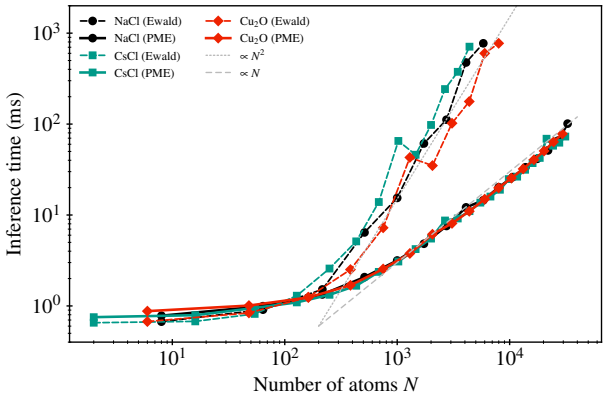

Figure 6: Runtime scaling of LOREM with Ewald (dashed) and PME (solid) long-range implementations for NaCl, CsCl, and Cu_2O supercells. Reference scaling lines ($\propto N^2$ and $\propto N$) are shown for comparison.

Table 4: Runtime of LOREM with Ewald and PME long-range implementations for NaCl supercells of varying size N. LR denotes the time spent in the long-range block alone.

	Ewald			PME		
N	SR+LR (ms)	LR (ms)	LR (%)	SR+LR (ms)	LR (ms)	LR (%)
8	0.7	0.2	32.4	0.8	0.3	41.9
64	0.9	0.4	45.4	1.0	0.5	50.0
512	6.4	5.3	82.6	2.1	1.0	46.0
1728	61.2	58.4	95.5	4.8	2.1	42.6
4096	475.8	469.8	98.7	12.1	6.1	50.5
8000	–	–	–	20.2	8.9	43.8
17576	–	–	–	41.5	16.9	40.7
32768	–	–	–	101.2	51.6	51.0

E Runtime benchmark

To estimate the runtime overhead of the long-range block and compare the Ewald and particle-mesh Ewald (PME) implementations, we benchmarked LOREM on supercells of three physical crystal structures: NaCl, CsCl, and Cu_2O. For each crystal, we constructed supercells of increasing size and ran predictions (energy and forces) for LOREM models with and without the long-range block, using both Ewald and PME for the long-range evaluation. All benchmarks were repeated ten times and averaged, and executed on a single NVIDIA H100 GPU.

The scaling behavior is shown in Fig. 6. Ewald summation scales quadratically with system size and runs out of memory beyond approximately 8000 atoms; PME scales roughly linearly and extends to over 30 000 atoms. At 4096 atoms, PME is approximately 40× faster than Ewald. Detailed timings, including the fraction of total runtime spent in the long-range block, are reported in Tables 4 to 6. With Ewald, the LR fraction grows from around 30 % at small sizes to over 95 % at large sizes. With PME, the LR fraction remains roughly constant and depends on the density of the crystal: Cu_2O (the densest, ~ 35 neighbors per atom) shows around 25 %, while CsCl (the least dense, ~ 14 neighbors) shows around 60 %. Overall, performance with PME on the order of single-digit µs/atom is in line with expectations for modern MLIPs.

Table 5: Runtime of LOREM with Ewald and PME long-range implementations for CsCl supercells of varying size N.

	Ewald			PME		
N	SR+LR (ms)	LR (ms)	LR (%)	SR+LR (ms)	LR (ms)	LR (%)
2	0.7	0.3	42.7	0.8	0.4	50.2
54	0.8	0.4	48.1	0.9	0.5	55.4
432	5.1	4.3	84.5	1.7	0.9	52.4
1458	46.0	44.3	96.2	4.2	2.5	58.8
4394	710.1	705.7	99.4	11.0	6.7	60.6
8192	–	–	–	19.0	11.2	59.2
18522	–	–	–	42.5	25.5	60.0
31250	–	–	–	72.9	44.2	60.7

Table 6: Runtime of LOREM with Ewald and PME long-range implementations for Cu_2O supercells of varying size N.

	Ewald			PME		
N	SR+LR (ms)	LR (ms)	LR (%)	SR+LR (ms)	LR (ms)	LR (%)
6	0.7	0.3	42.2	0.9	0.5	55.6
48	0.8	0.4	42.8	1.0	0.5	52.5
384	2.5	1.5	58.6	1.7	0.6	38.4
1296	43.0	40.3	93.9	3.8	1.1	30.4
4374	177.1	169.1	95.4	11.0	2.9	26.5
7986	773.1	759.0	98.2	19.9	5.7	28.7
16464	–	–	–	40.5	11.3	27.8
29478	–	–	–	78.1	19.6	25.1

Table 7: Root mean squared errors for energy E and forces $\boldsymbol{f}$ for datasets used in Sec. 6.1 of the main text, for different LOREM variants. See Tab. 1 of the main text for details.

Dataset		LOREM No LR	LOREM LR $l = 0$	LOREM LR $l = 1$	LOREM LR $l = 2$
MgO surface	E (meV/at)	2.234	0.063	**0.062**	0.064
(Validation set)	$\boldsymbol{f}$ (meV/Å)	61.487	5.870	5.284	**4.076**
NaCl cluster	E (meV/at)	1.582	0.113	**0.111**	0.112
(Validation set)	$\boldsymbol{f}$ (meV/Å)	50.110	1.243	**1.062**	1.155
Biodimers	E (meV/at)	8.259	0.302	0.329	**0.222**
	$\boldsymbol{f}$ (meV/Å)	16.452	1.677	1.725	**1.646**
Cumulene	E (meV/at)	16.203	15.793	5.576	**3.309**
	$\boldsymbol{f}$ (meV/Å)	126.000	133.092	85.956	**50.084**

F Ablations of LR block and l

We repeated the experiments on the MgO surface, NaCl cluster, cumulene, and biodimers with LOREM models with $l = 0$ and $l = 1$, as well as without the long-range message-passing block. The parameter counts for no LR and $l = 0, 1, 2$ (in that order) are 839129, 1020300, 1020749, and 1021198.

Results can be seen in Table 7 and Figs. 7 to 9. Generally, higher l improve error metrics. For experiments where only scalar long-range interactions are required (MgO surface, NaCl cluster), higher l do not improve qualitative agreement. The cumulene example, which requires access to relative orientation between rotors, requires equivariant long-range interactions with $l = 2$ to be resolved. Removing the long-range message-passing block significantly increases error and renders the model unable to solve all the benchmark tasks.

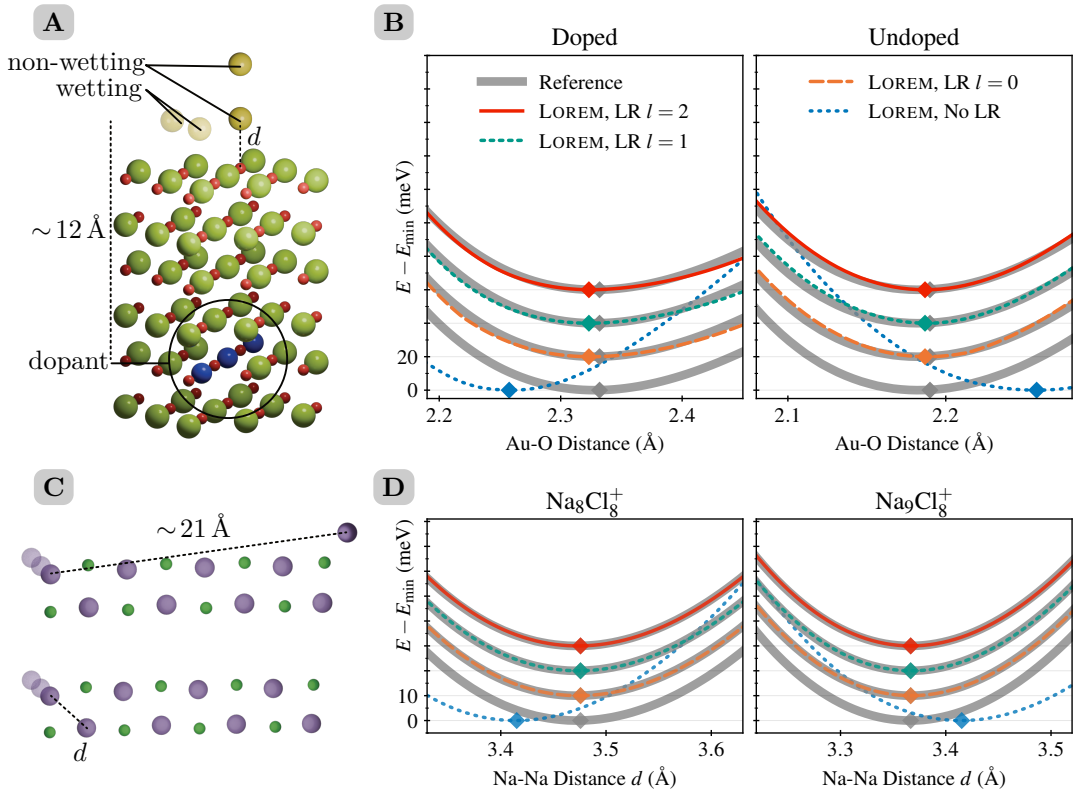

Figure 7: (**A**) Au_2 dimer on MgO surface, showing both wetting and non-wetting geometries, as well as the Al dopant. (**B**) Energy over distance d for the non-wetting geometry for the doped and undoped surface. The minima are indicated with a diamond symbol; the reference energy curve is drawn in grey. Offsets are added to distinguish the curves and the value at the minimum is subtracted. (**C**) $Na_9Cl_8^+$ (top) and $Na_8Cl_8^+$ (bottom) cluster, the moving atom is marked with transparent copies of itself, and the distance of interest is labeled with d. (**D**) Energy over distance for both clusters.

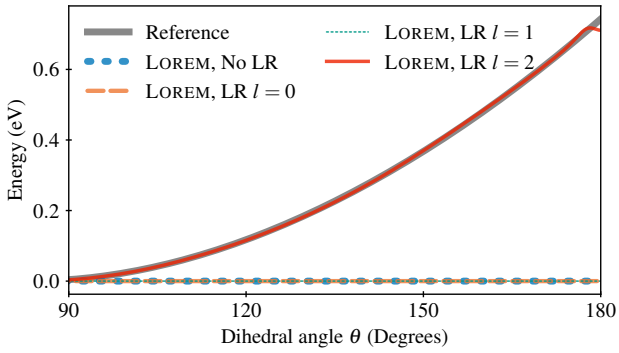

Figure 8: Energy profile over a 90° rotation of one rotor. The minimum value of each curve is subtracted before plotting.

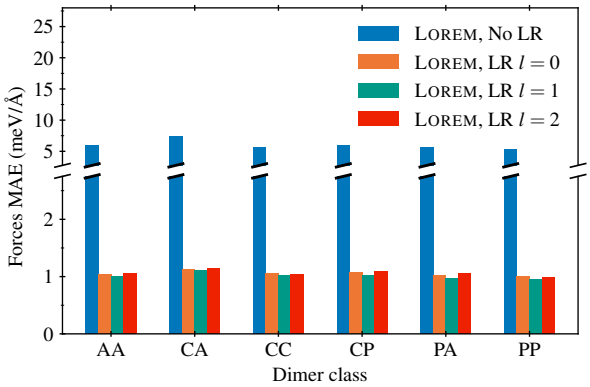

Figure 9: Mean absolute error on forces for different models on the different dimer classes: Apolar-apolar (AA), charge-apolar (CA), charge-charge (CC), charge-polar (CP), polar-apolar (PA), polar-polar (PP).

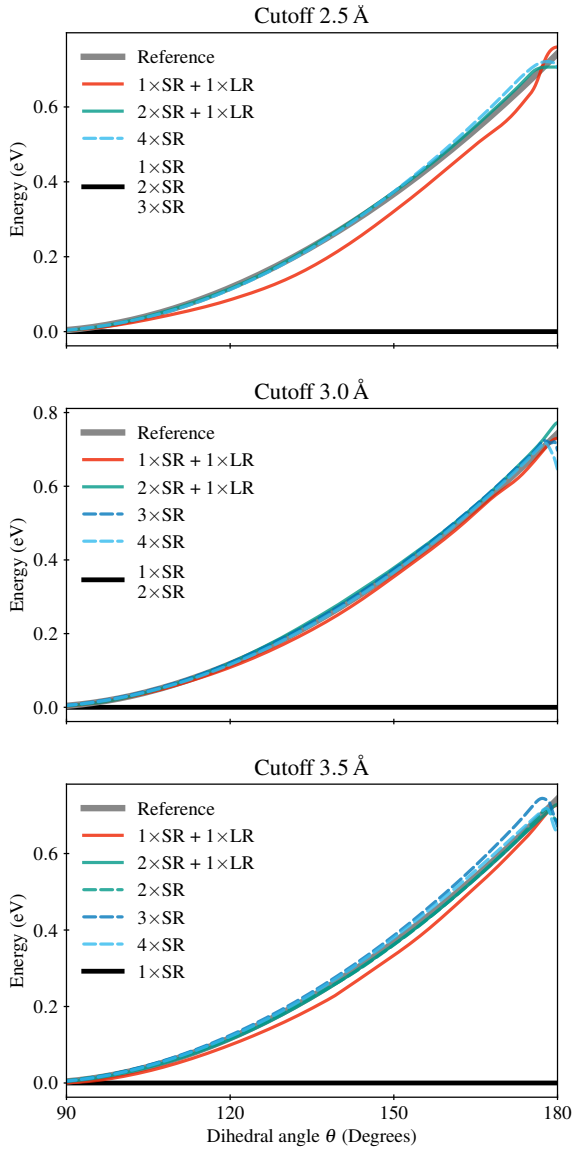

Figure 10: Rotational profile of cumulene for variations of the LOREM model with different r_c, different numbers of short-range message passing steps, and with and without long-range message passing. All curves that are identical to zero have been collapsed into a single one for readability.

G Additional results for cumulene

In Fig. 10, we present the curves related to Table 2. All models where the dihedral information is accessible can resolve the benchmark, but vary in match to the ground truth. In particular, we note that models with two short-range message passing steps perform better than a single one. This may be due to the high degree of degeneracy in atomic environments at short cutoffs, which is alleviated by message passing.

H Hyper-parameter sweep for the NaCl cluster and MgO surface datasets

We report the results of a hyper-parameter sweep for the NaCl cluster and MgO surface datasets on a 'inner' train/validation split of the training set used for the experiments in Section 6, aiming to gauge the impact of hyper-parameter tuning and early stopping on the validation set used for Table 1.

We used random train/validation split (4000 samples / 500 samples) of the NaCl cluster and MgO surface training datasets, sweeping the following combinations of parameters: Initial learning rate $(0.001, 0.0001)$, learning rate decay (linear, exponential), batch size $(32, 64)$, and energy/force loss weights $(0.5/0.5, 100/1, 1000/1)$ over 4000 epochs of training with the ADAM optimizer. From the resulting models, we selected those with the best 'inner' validation RMSE on energy and forces and evaluated the models on the 'outer' validation split used for Table 1. The results are presented in Table 8 and Table 9. The model selected by the best energy error on the validation set performs best in both cases, similar to the model used in the main text. The model with the best force error ranks second for the MgO surface task. We also verified that both models are able to reproduce the curves in Fig. 2. In summary, explicitly tuning the model on an inner train/validation split, including performing early stopping on this validation set, does not significantly impact benchmark results as reported in Table 1.

Table 8: Errors on the validation set for the hyperparameter set with the best inner validation metrics for the MgO task.

Model	Test RMSE E (meV/atom)	Test RMSE $\boldsymbol{f}$ (meV/Å)
Lorem	0.064	4.076
Nearest other	0.071 (Cace-Les)	5.971 (Mace)
Best E model	0.064	5.229
Best F model	0.078	3.630

Table 9: Errors on the validation set for the hyperparameter set with the best inner validation metrics for the NaCl task.

Model	Test RMSE E (meV/atom)	Test RMSE $\boldsymbol{f}$ (meV/Å)
Lorem	0.112	1.155
Nearest other	0.210 (Cace-Les)	9.784 (Cace-Les)
Best E model	0.076	2.613
Best F model	0.101	1.473

I Result variants: MAE metrics, RMSE biodimers forces, different seeds

Table 1 in the main text reports root mean squared errors (RMSE). For completeness, we provide the corresponding mean absolute errors (MAE) in Table 10. Additionally, Fig. 11 shows the biodimers force errors using RMSE (the main text uses MAE). The conclusions are unchanged: LOREM achieves the lowest or competitive errors across all datasets and dimer classes.

To assess sensitivity to random initialization and dataset shuffling, we also report results for an alternative seed for LOREM (†) in Table 11 (RMSE) and Table 12 (MAE). This represents the worse of two seeds trained per task. The results are consistent with the main table, confirming that LOREM's performance is robust to the choice of random seed. We also confirmed that qualitative performance on the benchmark tasks is similar.

Table 10: Mean absolute errors for energy E and forces $\boldsymbol{f}$. Models for which MAE metrics are not available (4G-NN, SPOOKYNET) are omitted. Best in bold, second best underlined. See Table 1 for details.

	Dataset	LOREM 1×SR+LR	CACE-LES 1×SR+LR	MACE 2×SR	PET 2× SR
MgO surface	E (meV/at)	<u>0.041</u>	**0.036**	0.353	0.188
(Validation set)	$\boldsymbol{f}$ (meV/Å)	**2.869**	5.584	<u>3.620</u>	4.191
NaCl cluster	E (meV/at)	**0.090**	<u>0.161</u>	1.345	1.249
(Validation set)	$\boldsymbol{f}$ (meV/Å)	**0.778**	<u>6.388</u>	23.598	20.011
Biodimers	E (meV/at)	**0.155**	<u>0.566</u>	3.220	2.657
	$\boldsymbol{f}$ (meV/Å)	**1.023**	<u>1.787</u>	5.278	5.352
Cumulene	E (meV/at)	**1.096**	14.567	8.961	<u>1.363</u>
	$\boldsymbol{f}$ (meV/Å)	<u>22.078</u>	107.946	70.467	**13.610**

Table 11: RMSE for energy E and forces $\boldsymbol{f}$, with LOREM results for an alternative seed (†), representing the worse of two seeds. Best in bold, second best underlined. See Table 1 for details.

	Dataset	LOREM† 1×SR+LR	CACE-LES 1×SR+LR	MACE 2×SR	PET 2× SR	4G-NN 1×SR+LR	SPOOKYNET 6×SR+LR
MgO surface	E (meV/at)	**0.065**	<u>0.071</u>	0.376	0.210	0.219	0.107
(Validation)	$\boldsymbol{f}$ (meV/Å)	**4.381**	7.913	5.971	6.261	66.000	<u>5.337</u>
NaCl cluster	E (meV/at)	**0.112**	0.210	1.681	1.517	0.481	<u>0.135</u>
(Validation)	$\boldsymbol{f}$ (meV/Å)	<u>1.275</u>	9.784	40.219	42.438	32.780	**1.052**
Biodimers	E (meV/at)	**0.370**	<u>2.259</u>	7.793	6.758	–	–
	$\boldsymbol{f}$ (meV/Å)	**1.985**	<u>3.163</u>	16.150	16.470	–	–
Cumulene	E (meV/at)	<u>3.307</u>	17.803	12.592	**3.205**	–	–
	$\boldsymbol{f}$ (meV/Å)	<u>55.412</u>	147.616	104.318	**46.905**	–	–

Table 12: MAE for energy E and forces $\boldsymbol{f}$, with LOREM results for an alternative seed (†), representing the worse of two seeds. Models for which MAE metrics are not available (4G-NN, SPOOKYNET) are omitted. Best in bold, second best underlined. See Table 1 for details.

	Dataset	LOREM† 1×SR+LR	CACE-LES 1×SR+LR	MACE 2×SR	PET 2× SR
MgO surface	E (meV/at)	<u>0.040</u>	**0.036**	0.353	0.188
(Validation set)	$\boldsymbol{f}$ (meV/Å)	**3.133**	5.584	<u>3.620</u>	4.191
NaCl cluster	E (meV/at)	**0.090**	<u>0.161</u>	1.345	1.249
(Validation set)	$\boldsymbol{f}$ (meV/Å)	**0.882**	<u>6.388</u>	23.598	20.011
Biodimers	E (meV/at)	**0.322**	<u>0.566</u>	3.220	2.657
	$\boldsymbol{f}$ (meV/Å)	**1.110**	<u>1.787</u>	5.278	5.352
Cumulene	E (meV/at)	**1.212**	14.567	8.961	<u>1.363</u>
	$\boldsymbol{f}$ (meV/Å)	<u>28.448</u>	107.946	70.467	**13.610**

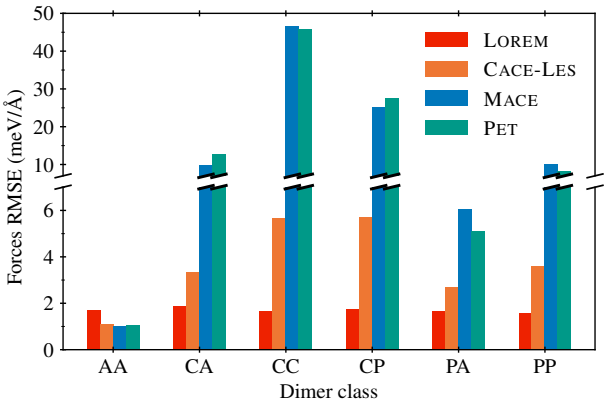

Figure 11: Root mean squared error on forces for different models on the different dimer classes. Compare with Fig. 4 in the main text, which shows MAE.

J ADAPT benchmark

To evaluate LOREM on a larger-scale dataset, we trained it on the ADAPT silicon point-defect dataset (Dramko et al., 2025). This dataset contains 206 973 training, 51 743 validation, and 4425 test structures, each consisting of 217 atoms (215 Si atoms plus 2 dopant atoms drawn from 24 elements, forming 52 unique dopant pairs). The test set consists of the first frames from 100 held-out relaxation trajectories, following the evaluation protocol of the original work.

We used LOREM with default model hyperparameters ($r_c = 5$ Å, $l_{max} = 6$, $l_{max,\,LR} = 2$, 128 scalar features, 8 spherical channels, 32 radial basis functions). Training hyperparameters were selected via a sweep of 24 combinations on a 9500-structure debug subset (100 epochs): optimizer (Muon Jordan et al. (2024), ADAM), initial learning rate ($1 \cdot 10^{-3}, 1 \cdot 10^{-4}$), learning rate decay (linear, exponential), and energy/force loss weights ($0.5/0.5, 100/1, 1000/1$). Four configurations from the Pareto frontier were selected for production training on the full dataset, all using the Muon optimizer with initial learning rate $1 \cdot 10^{-3}$: energy/force loss weights 0.5/0.5 with linear and exponential decay, and 100/1 with linear and exponential decay. These were trained for 300 epochs (65 h on a single H100 GPU).

Results are shown in Table 13, reporting the best force MAE and best energy MAE across the four production runs, each evaluated at the checkpoint with the highest summed R^2 of energy and forces on the validation set. LOREM nearly matches the force accuracy of the ADAPT model (which uses an all-to-all transformer architecture with ∼4 M parameters) while achieving substantially better energy accuracy—despite using a single model for both properties, compared to ADAPT's separate energy-only model. LOREM also significantly outperforms all other baselines, including MACE (both retrained and foundation models) and MatterSim.

Table 13: Mean absolute errors on the 100-structure ADAPT benchmark test set. Baseline results are from Dramko et al. (2025). The ADAPT model uses a separate energy-only model for energies; all other models predict both properties jointly. Best in bold, second best underlined. Parameter counts: ADAPT Small ∼4 M, ADAPT Large ∼18 M.

Model	$\boldsymbol{f}$ MAE (eV/Å)	E MAE (eV)
LOREM (best $\boldsymbol{f}$)	<u>0.0128</u>	<u>0.097</u>
LOREM (best E)	0.0136	**0.076**
ADAPT Small	**0.0126**	0.578
ADAPT Large	0.0136	–
MACE Retrained	0.0217	1.313
MACE MP0a Large	0.0439	6.101
MACE MPA-0 Medium	0.0349	2.048
MACE OMAT-0 Medium	0.0283	3.223
MatterSim 1M	0.0323	1.743
MatterSim 5M	0.0335	0.829