Reciprocal Space Attention for Learning Long-Range Interactions

Anonymous Author(s)
Affiliation
Address
email

Abstract

Machine learning interatomic potentials (MLIPs) have revolutionized the modeling of materials and molecules by directly fitting to *ab initio* data. However, while these models excel at capturing local and semi-local interactions, they often prove insufficient when an explicit and efficient treatment of long-range interactions are required. To address this limitation, we introduce Reciprocal-Space Attention (RSA), designed to capture long-range interactions in the Fourier domain. RSA can be seamlessly integrated with any existing local or semi-local MLIP framework. Our key contribution is mapping the linear-scaling attention mechanism into Fourier space. This technique allows us to effectively capture long-range interactions, such as electrostatics and dispersion, without requiring predefined charges or other explicit empirical assumptions. We demonstrate the effectiveness of our method through a diverse set of benchmarks, including the dimer binding curve, dispersion interactions in layered phosphorene exfoliation, and molecular dynamics simulation of water. Our results show that RSA successfully captures long-range interactions in various such chemical environments.

1 Introduction

Machine learning interatomic potentials (MLIPs) are data-driven surrogates for the potential energy surface trained on *ab initio* energies and forces [9, 14]. By directly fitting electronic-structure datasets into high-dimensional models containing on the order of 10^4–10^6 parameters, MLIPs deliver accuracy comparable to density functional theory (DFT) at orders of magnitude lower computational cost per step. This efficiency enables molecular dynamics (MD) simulations of systems with thousands of atoms over nanosecond to microsecond timescales, extending the reach of atomistic modeling to mesoscopic regimes that are otherwise impractical with direct *ab initio* methods [39, 20, 44].

Two families of MLIPs have been especially influential. The first employs physics-inspired SE(3) invariant local descriptors—such as ACSFs, SOAP, and ACE—within kernels or neural networks [10, 4, 23]. The second relies on message-passing graph neural networks with explicit rotational equivariance, such as NequIP and MACE, which construct features on the fly while respecting symmetry constraints [5, 8, 7]. In both cases, representations are fundamentally local: atoms interact only within a cutoff radius r_{cut}, and for message passing with L layers, the receptive field extends to approximately $L \times r_{\mathrm{cut}}$, providing what is effectively a "semi-local" coverage. When trained on large and heterogeneous *ab initio* datasets, these models have demonstrated strong transferability across different chemistry and phases [6, 41].

Local and semi-local MLIPs work remarkably well in homogeneous bulk systems such as crystals and simple liquids [43, 35, 24], where slowly varying long-range contributions either cancel by symmetry or can be treated in a mean-field sense [47, 48, 29, 17]. However, many application-relevant chemical settings—surfaces, interfaces, nanostructures, molecular adsorption, charged or

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

polar media, and systems under external fields—are dominated by intrinsically nonlocal physics, including Coulombic interactions, polarization/induction, and dispersion [3, 22]. The nearsightedness of electronic matter (NEM) clarifies this distinction: the electronic density at a point is determined primarily by the *effective* potential in its vicinity, which itself contains contributions from long-range electric fields [28]. Thus, nearsightedness does not imply locality with respect to atomic coordinates. Incorporating explicit long-range interactions into MLIPs is therefore essential for the faithful modeling of heterogeneous environments.

Two broad strategies have been developed to move beyond the strict locality assumptions of MLIPs [3]. The first relies on charge-augmented schemes, as in PhysNet and its successors [50, 51, 15], which predict atomic charges (or charge-like) observables from local neighborhoods and then compute electrostatics using Ewald or PME-type solvers [21]. While straightforward to implement, these approaches face well-known challenges: atomic "charges" are not observables in *ab initio* theory; different density partitioning schemes yield inconsistent labels; locally predicted charges cannot capture physics beyond the cutoff (e.g., long-range charge transfer); and charge-equilibration fixes introduce additional, often ad hoc, parameters such as electronegativities [37]. A second line of work incorporates global interactions directly through fully long-range ML modules. Examples include long-range descriptors (e.g., LODE [33], reciprocal-space neural operators (e.g., Neural-P3M, Ewald message passing) [53, 38], attention mechanisms with efficient global reach (e.g., SpookyNet, Euclidean Fast Attention) [27, 51], and self-consistent field neural networks (SCFNN) [28].

From these developments, several key design considerations emerge for incorporating long-range physics into MLIPs: (i) end-to-end differentiability with strict energy–force consistency; (ii) a global receptive field with favorable scaling; (iii) natural compatibility with periodic boundary conditions; (iv) seamless integration with short-range MLIP backbones; and (v) avoidance of non-observables such as predefined atomic charges. In this work, we introduce Recriprocal Space Attention (RSA), a purely data-driven long-range framework that maps a linear-scaling attention mechanism into Fourier space, providing a global interaction channel while preserving end-to-end differentiability. Our RSA kernels integrate seamlessly with existing short-range MLIPs; in this study we pair it with MACE to construct a unified energy model. By operating directly in the Fourier domain, the module captures electrostatics and dispersion without relying on empirical charge partitioning, while remaining naturally compatible with periodic boundary conditions. We evaluate the approach on benchmarks designed to probe long-range physics - including S_N2 reaction system, dimer binding curves, gas of random charges, dispersion-controlled exfoliation in layered phosphorene, and molecular dynamics of liquid water — demonstrating systematic improvements over local and semi-local baselines in energies, forces, and physically relevant observables.

2 Methods

2.1 Real Space Attention

We begin by providing the necessary background for establishing a Reciprocal space–based attention mechanism that enables a computationally efficient and accurate treatment of long-range interactions in short-range MLIPs. In particular, our methodology extends naturally beyond real space and remains applicable to periodic systems.

The standard dot-product self-attention mechanism, widely used in transformers, processes inputs through alternating self-attention and feed-forward blocks, with positional information typically introduced via absolute positional embeddings [52]. In this formulation, let m index an item in a sequence of length N. The feature at position m is combined with a positional embedding and linearly projected to a query vector $\mathbf{q}_m$, while tokens at positions n are projected to key vectors $\mathbf{k}_n$ and value vectors $\mathbf{v}_n$. The attention output at position m is then given by

$$\text{Attention}_m(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \frac{\sum_{n=1}^{N} \exp\left(\langle \mathbf{q}_m, \mathbf{k}_n \rangle\right) \mathbf{v}_n}{\sum_{n=1}^{N} \exp\left(\langle \mathbf{q}_m, \mathbf{k}_n \rangle\right)} \tag{1}$$

where $\mathbf{Q} \in \mathbb{R}^{N \times d_k}$, $\mathbf{K} \in \mathbb{R}^{N \times d_k}$, and $\mathbf{V} \in \mathbb{R}^{N \times d_v}$ are the query, key, and value vectors respectively, with N tokens and d_k features; $\langle \mathbf{q}_m, \mathbf{k}_n \rangle = \mathbf{q}_m^\top \mathbf{k}_n$ denotes the scalar dot product.

A key limitation of Eq. 1 is its quadratic complexity, $\mathcal{O}(N^2)$, both in computation and in memory, as each of the N queries attends to all N key vectors. A further limitation is the reliance on absolute positional encoding which directly incorporates positional representations.

To address the latter issue, recent work has introduced relative positional encoding, which better reflect the pairwise relationships between tokens. Rotary positional embeddings (RoPE) [49], first introduced in RoFormer, implement this idea by applying position-dependent rotations $\mathbf{R}_m$ to the query and key vectors. This approach has been shown to improve performance in tasks where relative structure is essential. In parallel, efforts to reduce the quadratic computational cost have led to the development of linear attention mechanisms, which approximate the softmax kernel using feature maps that enable linearization [36].

To make RoPE compatible with such linear transformers, approximations have been proposed that combine position-dependent rotations with kernel feature maps, allowing relative positional information to be retained while preserving $\mathcal{O}(N)$ complexity. The RoPE integrated linear attention is then

$$\text{Attention}_m(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \frac{\sum_{n=1}^{N} (\mathbf{R}_m \phi(\mathbf{q}_m))^T \, \mathbf{R}_n \phi(\mathbf{k}_n) \mathbf{v}_n}{\sum_{n=1}^{N} \phi(\mathbf{q}_m)^T \, \phi(\mathbf{k}_n)} \tag{2}$$

where ϕ denotes a feature map (typically non-negative). For each query vector $\mathbf{q}_m$, the quantity $\sum_{n=1}^{N} \phi(\mathbf{k}_n) \, \mathbf{v}_n^\top$ is independent of m and can be precomputed. Consequently, Eq. 2 can be written as

$$\text{Attention}_m(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \frac{(\mathbf{R}_m \phi(\mathbf{q}_m))^T \sum_{n=1}^{N} \mathbf{R}_n \phi(\mathbf{k}_n) \mathbf{v}_n^T}{\phi(\mathbf{q}_m)^T \sum_{n=1}^{N} \phi(\mathbf{k}_n)} \tag{3}$$

which makes the overall computation linear via $\mathcal{O}(N)$. We encourage the interested reader to the original papers for more details on linear attention and rotary embeddings (RoPE) [49, 36]. In subsequent discussions, we focus on integrating Eq. 3 into atomic systems subject to periodic boundary conditions.

2.2 Reciprocal Space Periodic Attention

Building on the real-space attention framework discussed in Sec. 2.1, we now extend the formulation into reciprocal (Fourier) space, which naturally accommodates periodic boundary conditions and long-range interactions. In periodic molecular dynamics (MD) simulations, atoms interact with their neighbors not only through real-space short-range interactions but also via reciprocal-space long-range interactions, such as electrostatics, dipole–dipole couplings, dispersion, etc. Capturing these long-range interactions in a purely real-space setting necessarily incurs $\mathcal{O}(N^2)$ cost in both compute and memory, unless approximations such as multipole expansions are introduced [32, 30]. In contrast, reciprocal-space methods, such as Ewald summation and its generalizations, treat slowly decaying interactions more efficiently by decomposing the potential into short- and long-range contributions, making them naturally well suited for a Fourier-space attention mechanism. Following the classical Ewald partitioning, the total potential $V(r)$ can be decomposed into a rapidly varying Gaussian-truncated short-range and uniformly slowly varying long-range piece [31]

$$V(r) = v_{\text{short}}(r) + v_{\text{long}}(r) = \frac{\text{erfc}(\alpha r)}{r} + \frac{\text{erf}(\alpha r)}{r} \tag{4}$$

where erf and erfc denote the error and complementary error functions, respectively, and α is the screening parameter that defines the length scale separating short- and long-range contributions. We assume that $v_{\text{short}}(r)$ can be fully represented by any modern MLIP such as MACE, and therefore focus only on $v_{\text{long}}(r)$. Following Ewald summation, the long-range interaction energy of a charge-neutral system can then be written as

$$E_{\text{long}} = \frac{2\pi}{V} \sum_{\mathbf{k}\neq\mathbf{0}} \frac{e^{-k^2/(4\alpha^2)}}{k^2} \sum_{m}^{N} \sum_{n=1}^{N} \tilde{q}_m \tilde{q}_n \, e^{\,i\mathbf{k}\cdot(\mathbf{r}_m - \mathbf{r}_n)} \;=\; \frac{2\pi}{V} \sum_{\mathbf{k}\neq\mathbf{0}} \frac{e^{-\,k^2/(4\alpha^2)}}{k^2} \left|S(\mathbf{k})\right|^2 \tag{5}$$

where V indicates the volume of the simulation cell, α is the Ewald screening parameter which controls the length-scale of the separation of short and long-range terms, $\mathbf{k}$ are the reciprocal lattice vectors, and $\tilde{q}_m, r_m$ indicate the atomic charges and positions of the m-th atom respectively. By excluding the divergent $\mathbf{k} = 0$ term, the charge neutrality of the system is implicitly enforced. We express the Ewald sum in terms of the magnitude of the structure factor, $|S(\mathbf{k})|^2$, in Eq. 5. This reformulation is not merely cosmetic: for a fixed reciprocal-space grid, it reduces the complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$ where N is the total number of atoms. Looking closely at the expression for E_{long} in Eq. 5, one can see that the factor $S(\mathbf{k})S(-\mathbf{k})$ is the only term that globally couples all atoms through the exponential, $e^{i\mathbf{k}(\mathbf{r}_m - \mathbf{r}_n)}$, and thereby ensures correct treatment of long-range interactions. Moreover, the periodicity of the reciprocal lattice vectors ensures that direct position vectors in real-space can be used without costly construction of periodicity-aware edge lists. By integrating RoPE methodology with the Ewald-sum formalism, we develop a reciprocal space attention kernel that inherently satisfies periodic boundary conditions while preserving translational invariance. This is achieved through the encoding of pairwise interaction via Bloch-like phase factors. We refer to this as Fourier Positional Encoding (FPE), which is defined as

$$\text{FPE}_k(\mathbf{x}, \vec{\mathbf{r}}_m) = \mathbf{x} \cdot e^{i\vec{\mathbf{k}} \cdot \vec{\mathbf{r}}_m} \tag{6}$$

where m denotes the m-th atom in the simulation cell. A key advantage of using FPE is its phase invariance in the periodic lattice space, which gives the following

$$e^{i\mathbf{k}\cdot[(\mathbf{r}_m - \mathbf{r}_n) + \mathbf{T}]} = e^{i\mathbf{k}\cdot(\mathbf{r}_m - \mathbf{r}_n)} \tag{7}$$

where $\mathbf{T}$ is a lattice translational vector. This shows that the phase factor is *invariant* under the choice of periodic image $\mathbf{T}$, provided that $\mathbf{k}$ is a reciprocal lattice vector of the simulation cell. As a consequence, quantities built purely from such phase factors - such as the Fourier-transformed density are naturally fully periodic. We can formulate an attention kernel (see Fig. 1) using FPE for query and key vectors. Since both query and key vectors are complex, the scalar product is instead defined as $\langle \mathbf{q}, \mathbf{k} \rangle = \mathbf{q}^T \bar{\mathbf{k}}$, where $\bar{\mathbf{k}}$ indicates a complex conjugate. The inner product between the query and key vectors associated with atoms m and n in a standard attention format can then be written as,

$$\left\langle \mathbf{q}_m e^{i\mathbf{k}\cdot\mathbf{r}_m}, \mathbf{k}_n e^{i\mathbf{k}\cdot\mathbf{r}_n} \right\rangle = \langle \mathbf{q}_m,\ \mathbf{k}_n \rangle e^{i\mathbf{k}\cdot(\mathbf{r}_m - \mathbf{r}_n)} \tag{8}$$

Eq. 8 is one of the key advantages of FPE which enable us to write standard quadratic attention operation without normalization and a row-wise softmax as

$$\text{RSA}_m(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{V}_m = \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{n=1}^{N} \langle \mathbf{q}_m, \mathbf{k}_n \rangle e^{i\mathbf{k}\cdot(\mathbf{r}_m - \mathbf{r}_n)} \mathbf{V}_n \tag{9}$$

where we reduce over the $\mathbf{k}$ dimension to get back the final attention outputs $\mathbf{V}_m$. Except for the normalizing constants, the above expression strongly resembles the total long-range potential for an atom m given by Ewald sum [42, 31]

$$V_m^{\text{LR}} = \frac{2\pi}{V} \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{n=1}^{N} \frac{e^{-k^2/(4\alpha^2)}}{k^2} \tilde{q}_n e^{i\mathbf{k}(\mathbf{r}_m - \mathbf{r}_n)} \tag{10}$$

To ameliorate the cost associated with quadratic attention, we can approximate the above expression using a linear attention-like mechanism discussed in Eq.3, the inner summation within the value matrix computation (Eq.9) remains constant for a given atom *m*. This atom-independent property enables a significant reduction in computational complexity, reducing it from a quadratic to a linear operation. Thus, using a linear attention-like mechanism, the Reciprocal-Space Attention (RSA) formulation can be written as

$$\text{RSA}_m(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \simeq \sum_{\mathbf{k} \neq \mathbf{0}} \text{FPE}(\phi(\mathbf{q}_m), \mathbf{r_m})^T \sum_{n=1}^{N} \text{FPE}(\phi(\mathbf{k}_n), \mathbf{r_n}) \mathbf{v_n}^T \tag{11}$$

where $\text{FPE}(\phi(\mathbf{q}_m), \mathbf{r_m})$ and $\text{FPE}(\phi(\mathbf{k}_n), \mathbf{r_n})$ are feature-mapped FPE rotated query and key vector, respectively for atoms *m* and *n*, respectively.

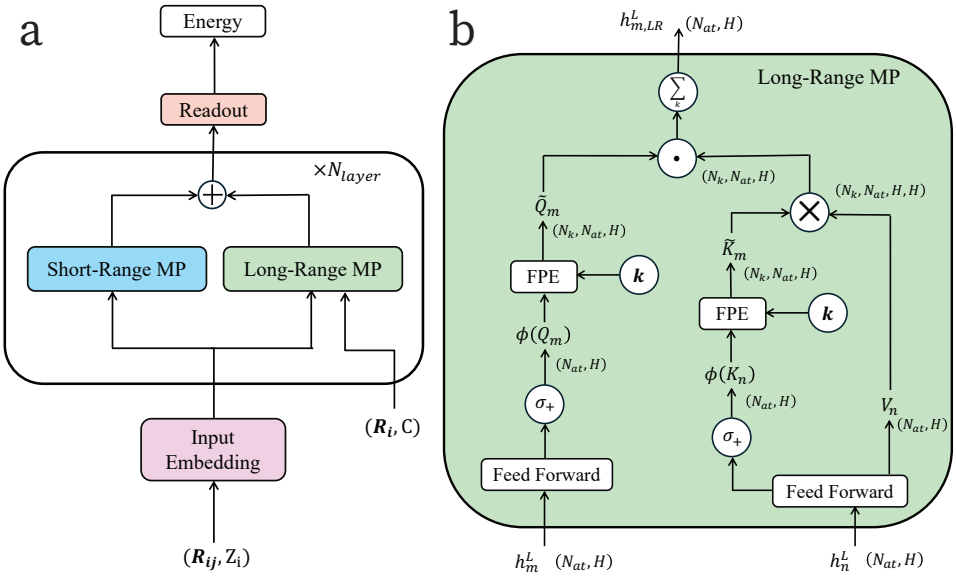

Figure 1: (a) Overview of the short-range and long-range GNN architecture. Here, Z_i denotes the atomic number of atom i, $\mathbf{C}$ are the crystal lattice vectors, $\mathbf{R}_i$ is the position of atom i, and $\mathbf{R}_{ij}$ is the displacement vector from atom j to atom i. Each interaction layer couples a short-range message-passing (SR-MP, blue) block with a long-range message-passing (LR-MP, green) block, resulting in a complete atomic representation that are passed through a readout head to predict energy and forces. (b) Schematic depiction of Reciprocal Space Attention (RSA) module, which provides the long-range message-passing channel. Atomic features h^L_m and h^L_n are first projected into query, key, and value vectors through feed-forward layers followed by nonlinearities σ_+ to obtain $\phi(Q_m)$ and $\phi(K_n)$. The transformed feature vectors are then rotated using Fourier Positional Encodings (FPE), parameterized by reciprocal lattice vectors $\mathbf{k}$, yielding the rotated $\tilde{\mathbf{Q}}_m$ and $\tilde{\mathbf{K}}_n$ vectors. The rotated keys, $\tilde{\mathbf{K}}_n$, are combined with the value vector $\mathbf{V}_n$ via an outer product, $\tilde{\mathbf{K}}_n \otimes \mathbf{V}_n$, to obtain a graph-level key-value cache. This cache is broadcast back to the node space and contracted (left-multiplied) with the rotated queries $\tilde{\mathbf{Q}}_m$ at the corresponding k-frequency, producing per-mode interactions for each node. Summation over reciprocal modes $\sum_k$ yields the long-range message.

3 Results

Several recent machine learning interatomic potentials [6, 46, 51, 54] replace or augment standard message passing updates with self-attention mechanisms over neighboring atoms or edges within local atomic environments. While message-passing neural networks (MPNNs) are effective at capturing semi-local interactions over a few hops, they are not expected to represent true long-range effects as discussed in [3]. In the absence of intermediary nodes connecting distant atoms, no information can propagate. In addition, MPNNs often suffer a loss of expressivity as the number of message-passing steps increases (e.g. over-smoothing), as discussed in [2, 13].

3.1 Disconnected Molecular Graphs with Long-range effects

3.1.1 $\mathrm{S_N2}$ Reaction Systems

The bimolecular nucleophilic substitution ($\mathrm{S_N2}$) is a canonical reaction mechanism in organic chemistry. In this process, a strong nucleophile attacks an sp^3-hybridized carbon from the backside, forming a new bond as the leaving group departs in a single, concerted step. As a representative case, we consider reactions involving fluoride and iodide, which are prototypical $\mathrm{S_N2}$ systems. These systems exhibit pronounced long-range electrostatic interactions between the methyl halide (with a large dipole moment) and the halide anion (bearing a negative charge), posing a significant challenge for strictly local models. We sampled the subset comprising F and I ions from the original $\mathrm{S_N2}$ data set constructed by [50]. We trained a short-ranged (SR) model using MACE [5] with a radial

cut-off of 5 Å and two message passing layers, leading to the total receptive field of 10 Å for each atom. We use the same architectural settings for the long-range (LR) model and with a smearing width ($\sigma = 5$ Å). Fig. 2 presents the one-dimensional potential energy surface along the reaction coordinate. The short-range model exhibits incorrect asymptotic behavior and unphysical artifacts at large ion–molecule separations; once the separation exceeds the cutoff, it predicts a constant energy. In contrast, the LR model accurately captures the potential energy surface across the full range of separations, including the long-distance tail. These results show that the LR model remains sensitive to interactions beyond the local receptive field, whereas purely local models saturate once the interatomic separations exceed their receptive cutoffs.

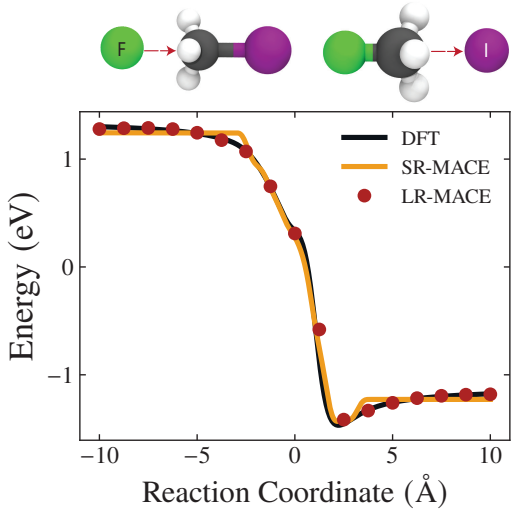

Figure 2: SN2 binding curves comparing DFT with LR-MACE and SR-MACE models.

3.1.2 Dimer Binding Curves

We next benchmark our LR method on binding curves for dimers of charged (C) and polar (P) molecules at varying separations in a periodic cubic box with an edge length of (30 Å). The dataset is originally derived from the BioFragment Database [12]. We additionally select a representative CP dimer class and recompute the curve with PBE0 plus many-body dispersion at coarser increments in interatomic distances of 0.1 Å. We also computed the potential energy curve of the water dimer (equivalent to the PP dimer case) as a function of the separation distance between the two molecular clusters. These calculations were performed using the SPC/E water model ([11]) with explicit long-range Coulomb interactions. Both the SR and LR models use the same architectural settings as in the S_N2 benchmark. Figs. 3(a) and (b) compare the binding curves for the water and CP dimers, respectively. Consistent with the S_N2 benchmark, the short-range model saturates once the inter-fragment separation exceeds its total receptive field, whereas the long-range model remains sensitive to interactions beyond this range and recovers the expected long-distance asymptotic behavior.

3.2 Periodic Graphs with Long-Range effects

3.2.1 A gas of Random Charges

As an initial validation, we consider a toy system comprising randomly distributed point charges within a cubic simulation cell with periodic boundary conditions, following the construction by [33]. Each configuration contains 128 atoms, of which 64 carry a positive charge of $+1e$ and the remaining 64 carry a negative charge of $-1e$. The atomic interactions are governed by the Coulomb potential, supplemented by the repulsive term of a Lennard-Jones potential. This benchmark is employed to assess the performance of our LR model in comparison to the SR model in an environment with strong electrostatic interactions. For the SR model, we used a cutoff of 5 Å with two message-passing layers. The LR component used the same receptive field and a Gaussian smearing of $\sigma = 5$ Å. With

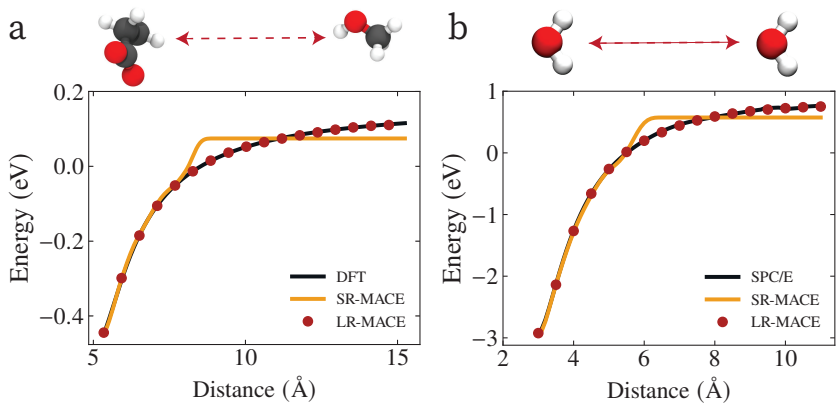

Figure 3: Binding curves comparing DFT with LR-MACE and SR-MACE models for (a) CP dimers (b) Water Dimer.

Table 1: MAEs of Energies and Forces for LR and SR MACE models with corresponding receptive fields.

Dataset	Model	Energy MAE (meV/atom)	Force MAE (meV/Å)
Random Charges	LR MACE (10 Å)	2.5	71.6
	SR MACE (10 Å)	3.0	97.1
Liquid NaCl	LR MACE (6 Å)	6.8	141.9
	SR MACE (6 Å)	8.7	175.1

these settings, the LR model consistently outperformed the SR model, with the improvement most pronounced in the Force MAEs, as shown in Table 1. This indicates that the LR formulation captures per-atom long-range electrostatics more effectively than its short-range counterpart.

3.2.2 Liquid sodium chloride

After validating the model on systems governed by $1/r$ interactions, we subsequently assessed its performance on a realistic molten salt system. Liquid sodium chloride (NaCl) was selected as the initial benchmark because of the pronounced electronegativity contrast between Na and Cl atoms, which is expected to induce substantial long-range Coulombic interactions. We employed the dataset of Ref. [26], comprising 1,014 configurations of 128 atoms (64 Na and 64 Cl) each, divided into 80% training and 20% validation splits. Table 1 compares the LR and SR MACE models evaluated with a 6 Å receptive field (single message passing layer with $r_{\text{cutoff}} = 6$ Å). The LR model consistently achieves lower MAEs for both energies and forces, reflecting clear relative reductions and indicating that explicit treatment of long-range interactions, absent in the SR formulation, improves predictive accuracy at the same cutoff.

3.2.3 Exfoliation of Phosphorene

We next consider the exfoliation behavior of black phosphorus, focusing on the interaction between pairs of phosphorene layers, building upon the work of [19]. In their study, the exfoliation curve was constructed by starting from the bulk crystal structure of black phosphorus and systematically varying the interlayer distance along the [010] crystallographic direction, while preserving the internal geometry of the puckered monolayers. The interlayer separation was defined as the distance between these layers, and the corresponding energies were computed using density functional theory with many-body dispersion (DFT+MBD). To develop a MLIP, [19] employed a Gaussian Approximation Potential (GAP) model augmented with an explicit long-range R^6 dispersion term. Their analysis revealed that a short-range GAP model lacking the R^6 correction failed to reproduce the exfoliation energy profile, especially at large interlayer separations, highlighting the necessity of explicitly including long-range interactions alongside local atomic descriptors. In this work, we leverage the original dataset from their study to evaluate our long-range machine learning framework. For

the short-range (SR) component, we used a cutoff of 6 Å with two message-passing layers. The long-range (LR) component employed a Gaussian smearing of $\sigma = 5$ Å. The exfoliation energy curve obtained using our method is shown in Fig. 4. This plot illustrates the energetic response as the interlayer distance increases, capturing the essential features of van der Waals interactions between phosphorene sheets. Notably, our approach achieves accuracy comparable to DFT+MBD without requiring empirical parameterization of long-range interactions, demonstrating its effectiveness for layered materials.

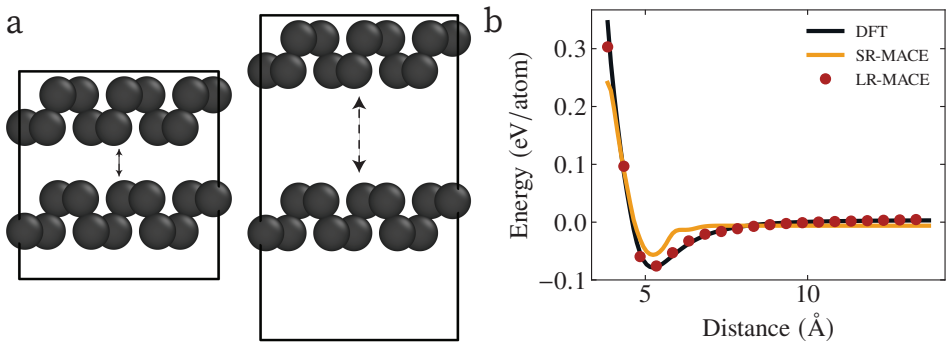

Figure 4: Inter-layer interaction dominated by dispersion. (a) Illustration of Phosphorene sheet exfoliation based on the original work of Deringer *et al.* [19] with the corresponding exfoliation curve shown in (b)

3.2.4 Effect of long-range beyond energies — The dynamics of the bulk water

To assess the capability of the LR MACE framework in representing complex molecular liquids, a data set comprising 1593 liquid water configurations [16], each containing 64 molecules, was used. Reference configurations were generated with the CP2K software [40] package at the revPBE0-D3 level [45] of the density functional theory. For the short-range (SR) component, we used a cutoff of 6 Å with two message-passing layers. The long-range (LR) component employed a Gaussian smearing of $\sigma = 5$ Å. Using these models, we performed MD simulations of bulk water with a density of 1 g/mL at 300 K. Consistent with previous observations[15, 55], Fig. 5a shows excellent agreement with the corresponding SR-MACE baseline, suggesting that our LR model provides stable continuous trajectories, crucial for MD simulations. Although the overall structure of liquid water is largely insensitive to long-range interactions, the dipolar correlations within the bulk are not [17, 15, 29]. To examine this, we evaluated $\chi_{zz}(k)$ for both the SR and LR water models. The longitudinal dipole–density correlations, shown in Fig. 5b, exhibit excellent agreement for both the models across most k values, except at long wavelengths (small k), where SR models diverge. We employed a semi-local two-layer model for bulk water, which exhibits a slightly delayed onset of divergence in the dipole–density correlations due to its larger receptive field compared to a fully local, non–message-passing model. However, as demonstrated previously [15, 28, 17], all SR models ultimately fail to capture the correct asymptotic screening behavior. In contrast, the LR model accurately reproduces the long-wavelength behavior of the dipole–density correlations, consistent with physical expectations for a properly screened polar liquid.

4 Discussion

The current RSA framework can be extended in several ways. In this work, we restrict the short-range MACE model to scalar features ($\ell_{max} = 0$), but it can also be naturally extended to higher-order tensors. The exponential factor $e^{i\mathbf{k}\cdot\mathbf{r}}$ is invariant with respect to rotations of $\mathbf{r}$ (where $\mathbf{r} \in \mathbb{R}^3$ is a real-space lattice vector), since $\mathbf{k} \cdot \mathbf{r}$ is a scalar dot product and $\mathbf{k}$ ($\mathbf{k} \in \mathbb{R}^3$ is a reciprocal-space wavevector) undergoes the same rotation as $\mathbf{r}$. So for a feature $T_m^{(\ell)}(\mathbf{r})$ one may write

$$\tilde{T}_m^{(\ell)}(\mathbf{r}, \mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{r}}\, T_m^{(\ell)}(\mathbf{r}), \tag{12}$$

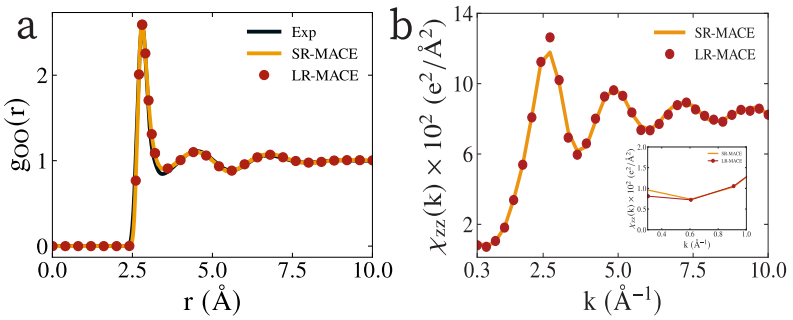

Figure 5: (a) Oxygen–oxygen radial distribution function ($g_{OO}(r)$) comparing the SR-MACE and LR-MACE models. (b) Longitudinal dipole–density correlation function, $\chi_{zz}(k)$, for bulk water at 300 K obtained from the two-layer short-range (SR-MACE) and long-range (LR-MACE) models. The inset highlights the low-k regime, which is most sensitive to long-range electrostatics.

where $T_m^{(\ell)}(\mathbf{r})$ and $\tilde{T}_m^{(\ell)}(\mathbf{r}, \mathbf{k})$ denotes the rank-ℓ spherical-tensor feature at $\mathbf{r}$ with component $m \in \{-\ell, \ldots, \ell\}$. As mentioned above, the exponential factor is invariant under rotation, so the tensor rank ℓ is preserved. This irrep-preserving form is the simplest extension to higher-order tensor features when rank-mixing is not required. Although the above method generalizes to equivariant features, [42] observed that invariant descriptors are sufficient to capture long-range pairwise interactions. Likewise, Kosmala *et. al.* [38] reported significant improvements when only the scalar embedding was upgraded within an equivariant architecture like PaiNN. For these reasons, in this work we focus on a MACE model with invariant (scalar) features only. Potentially, it is also interesting to explore the hybrid framework in which invariant long-range messages are passed through to the equivariant short-range messages. Another possible avenue for extending our framework is to adopt mesh-based Ewald techniques, such as Particle–Mesh Ewald (PME) [18], smooth PME (SPME) [25] or Particle–Particle Particle–Mesh (PPPM) [34, 21]. In the current formulation, despite the reciprocal-space attention being linear in the number of $\mathbf{k}$-vectors, summing over the full set of reciprocal lattice vectors incurs the classical $\mathcal{O}(N^{3/2})$ [1] scaling of direct Ewald methods when system size is increased at fixed density. In practice, restricting the sum to a fixed set of top-K $\mathbf{k}$-vectors (i.e. top-K low frequency wave-vectors) for relatively similar lattice sizes works well and reduces the cost to $\mathcal{O}(N)$, with minimal impact on the accuracy for many systems. The FPE combined with the attention kernel presents a promising method for data-driven learning of long-range interactions without invoking any ad hoc intermediate observables like classical electronegativities [37] or scalar/vector atomic charges [50, 15]. While accurate, the reciprocal space attention framework comes with limitations. The usage of reciprocal lattice vectors during training can cause the model to be dependent on the k-grid, and thus in turn, on the lattice boxes used during training. For most cases, we have tested we have found this dependence to be minimal, and the model can generalize to much larger simulation cells (as shown in the case of bulk water) and denser k-grids during inference.

5 Conclusion

We introduced Reciprocal Space Attention (RSA), a k-space attention framework for machine learning interatomic potentials that captures long-range interactions without relying on intermediate observables like charges or empirical corrections. A core component of RSA is Fourier Positional Encoding (FPE), which inherently encodes periodicity and relative atomic positions via Bloch phase factors. FPE when combined with linear attention-like mechanism arrives at the RSA method which respects translational invariance of the atomic system. Benchmarks on S_N2 reactions, pair of charged & polar dimers, random charge systems, and layered phosphorene demonstrate that RSA consistently recovers correct long-range asymptotics absent in local and semi-local models. These long-range effects are common in atomistic simulations of molecules and materials, particularly in organic electrolytes, aqueous solutions, and interfacial systems. Our results highlight RSA as a general strategy for integrating long-range interactions into MLIPs such as MACE, thereby extending their applicability to heterogeneous chemical and materials systems.

References

[1] Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.

[2] Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications, 2021. URL https://arxiv.org/abs/2006.05205.

[3] Dylan M. Anstine and Olexandr Isayev. Machine learning interatomic potentials and long-range physics. *The Journal of Physical Chemistry A*, 127(11):2417–2431, February 2023. ISSN 1520-5215. doi: 10.1021/acs.jpca.2c06778. URL http://dx.doi.org/10.1021/acs.jpca.2c06778.

[4] Albert P Bartók, Mike C Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical review letters*, 104(13):136403, 2010.

[5] Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in neural information processing systems*, 35:11423–11436, 2022.

[6] 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, Samuel M. Blau, Vlad Cărare, James P. Darby, Sandip De, Flaviano Della Pia, Volker L. Deringer, Rokas Elijošius, Zakariya El-Machachi, Fabio Falcioni, Edvin Fako, Andrea C. Ferrari, Annalena Genreith-Schriever, Janine George, Rhys E. A. Goodall, Clare P. Grey, Petr Grigorev, Shuang Han, Will Handley, Hendrik H. Heenen, Kersti Hermansson, Christian Holm, Jad Jaafar, Stephan Hofmann, Konstantin S. Jakob, Hyunwook Jung, Venkat Kapil, Aaron D. Kaplan, Nima Karimitari, James R. Kermode, Namu Kroupa, Jolla Kullgren, Matthew C. Kuner, Domantas Kuryla, Guoda Liepuoniute, 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, Lars L. Schaaf, Christoph Schran, Benjamin X. Shi, Eric Sivonxay, Tamás K. Stenczel, Viktor Svahn, Christopher Sutton, Thomas D. Swinburne, Jules Tilly, Cas van der Oord, Eszter Varga-Umbrich, Tejs Vegge, Martin Vondrák, Yangshuai Wang, William C. Witt, Fabian Zills, and Gábor Csányi. A foundation model for atomistic materials chemistry, 2024. URL https://arxiv.org/abs/2401.00096.

[7] Ilyes Batatia, Simon Batzner, Dávid Péter Kovács, Albert Musaelian, Gregor NC Simm, Ralf Drautz, Christoph Ortner, Boris Kozinsky, and Gábor Csányi. The design space of e (3)-equivariant atom-centred interatomic potentials. *Nature Machine Intelligence*, 7(1):56–67, 2025.

[8] 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): 1–11, 2022.

[9] Jörg Behler and Gábor Csányi. Machine learning potentials for extended systems: a perspective. *The European Physical Journal B*, 94(7):142, 2021.

[10] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters*, 98(14):146401, 2007.

[11] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271, November 1987. ISSN 1541-5740. doi: 10.1021/j100308a038. URL http://dx.doi.org/10.1021/j100308a038.

[12] Lori A. Burns, John C. Faver, Zheng Zheng, Michael S. Marshall, Daniel G. A. 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), September 2017. ISSN 1089-7690. doi: 10.1063/1.5001028. URL http://dx.doi.org/10.1063/1.5001028.

[13] Chen Cai and Yusu Wang. A note on over-smoothing for graph neural networks, 2020. URL https://arxiv.org/abs/2006.13318.

[14] Michele Ceriotti. Beyond potentials: Integrated machine learning models for materials. *Mrs Bulletin*, 47(10):1045–1053, 2022.

[15] Bingqing Cheng. Latent ewald summation for machine learning of long-range interactions. *npj Computational Materials*, 11(1), March 2025. ISSN 2057-3960. doi: 10.1038/s41524-025-01577-7. URL http://dx.doi.org/10.1038/s41524-025-01577-7.

[16] Bingqing Cheng, Edgar A. Engel, Jörg Behler, Christoph Dellago, and Michele Ceriotti. Ab initio thermodynamics of liquid and solid water. *Proceedings of the National Academy of Sciences*, 116(4):1110–1115, January 2019. ISSN 1091-6490. doi: 10.1073/pnas.1815117116. URL http://dx.doi.org/10.1073/pnas.1815117116.

[17] Stephen J Cox. Dielectric response with short-ranged electrostatics. *Proceedings of the National Academy of Sciences*, 117(33):19746–19752, 2020.

[18] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, June 1993. ISSN 1089-7690. doi: 10.1063/1.464397. URL http://dx.doi.org/10.1063/1.464397.

[19] Volker L. Deringer, Miguel A. Caro, and Gábor Csányi. A general-purpose machine-learning force field for bulk and nanostructured phosphorus. *Nature Communications*, 11(1), October 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-19168-z. URL http://dx.doi.org/10.1038/s41467-020-19168-z.

[20] Volker L Deringer, Albert P Bartók, Noam Bernstein, David M Wilkins, Michele Ceriotti, and Gábor Csányi. Gaussian process regression for materials and molecules. *Chemical reviews*, 121 (16):10073–10141, 2021.

[21] Markus Deserno and Christian Holm. How to mesh up ewald sums. i. a theoretical and numerical comparison of various particle mesh routines. *The Journal of Chemical Physics*, 109(18):7678–7693, November 1998. ISSN 1089-7690. doi: 10.1063/1.477414. URL http://dx.doi.org/10.1063/1.477414.

[22] Harender S Dhattarwal, Ang Gao, and Richard C Remsing. Dielectric saturation in water from a long-range machine learning model. *The Journal of Physical Chemistry B*, 127(16):3663–3671, 2023.

[23] Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B*, 99(1):014104, 2019.

[24] Hongwei Du, Jian Hui, Lanting Zhang, and Hong Wang. Universal machine learning interatomic potentials are ready for solid ion conductors. *arXiv preprint arXiv:2502.09970*, 2025.

[25] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103 (19):8577–8593, November 1995. ISSN 1089-7690. doi: 10.1063/1.470117. URL http://dx.doi.org/10.1063/1.470117.

[26] Carolin Faller, Merzuk Kaltak, and Georg Kresse. Density-based long-range electrostatic descriptors for machine learning force fields. *The Journal of Chemical Physics*, 161(21), December 2024. ISSN 1089-7690. doi: 10.1063/5.0245615. URL http://dx.doi.org/10.1063/5.0245615.

[27] J Thorben Frank, Stefan Chmiela, Klaus-Robert MÃžller, and Oliver T Unke. Euclidean fast attention: Machine learning global atomic representations at linear cost. *arXiv preprint arXiv:2412.08541*, 2024.

[28] Ang Gao and Richard C. Remsing. Self-consistent determination of long-range electrostatics in neural network potentials. *Nature Communications*, 13(1), March 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29243-2. URL http://dx.doi.org/10.1038/s41467-022-29243-2.

[29] Ang Gao, Richard C Remsing, and John D Weeks. Local molecular field theory for coulomb interactions in aqueous solutions. *The Journal of Physical Chemistry B*, 127(4):809–821, 2023.

[30] Paul Gibbon and Godehard Sutmann. Long-range interactions in many-particle simulation. *Quantum simulations of many-body systems: from theory to algorithm. Eds. J. Grotendorst, D. Marx and A. Muramatsu. NIC-series*, 10:467–506, 2002.

[31] Todd Gingrich. *Simulating surface charge effects in carbon nanotube templated ionic crystal growth*. PhD thesis, Master's thesis (Oxford University, Oxford), 2010.

[32] 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. URL `http://dx.doi.org/10.1016/0021-9991(87)90140-9`.

[33] Andrea Grisafi and Michele Ceriotti. Incorporating long-range physics in atomic-scale machine learning. *The Journal of Chemical Physics*, 151(20), November 2019. ISSN 1089-7690. doi: 10.1063/1.5128375. URL `http://dx.doi.org/10.1063/1.5128375`.

[34] Roger W Hockney and James W Eastwood. *Computer simulation using particles*. crc Press, 2021.

[35] Aaron D Kaplan, Runze Liu, Ji Qi, Tsz Wai Ko, Bowen Deng, Janosh Riebesell, Gerbrand Ceder, Kristin A Persson, and Shyue Ping Ong. A foundational potential energy surface dataset for materials. *arXiv preprint arXiv:2503.04070*, 2025.

[36] Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention. In *International conference on machine learning*, pages 5156–5165. PMLR, 2020.

[37] 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), January 2021. ISSN 2041-1723. doi: 10.1038/s41467-020-20427-2. URL `http://dx.doi.org/10.1038/s41467-020-20427-2`.

[38] Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. Ewald-based long-range message passing for molecular graphs, 2023. URL `https://arxiv.org/abs/2303.04791`.

[39] Dávid Péter Kovács, Cas van der Oord, Jiri Kucera, Alice EA Allen, Daniel J Cole, Christoph Ortner, and Gábor Csányi. Linear atomic cluster expansion force fields for organic molecules: beyond rmse. *Journal of chemical theory and computation*, 17(12):7696–7711, 2021.

[40] Thomas D. Kühne, Marcella Iannuzzi, Mauro Del Ben, Vladimir V. Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Z. Khaliullin, Ole Schütt, Florian Schiffmann, Dorothea Golze, Jan Wilhelm, Sergey Chulkov, Mohammad Hossein Bani-Hashemian, Valéry Weber, Urban Borštnik, Mathieu Taillefumier, Alice Shoshana Jakobovits, Alfio Lazzaro, Hans Pabst, Tiziano Müller, Robert Schade, Manuel Guidon, Samuel Andermatt, Nico Holmberg, Gregory K. Schenter, Anna Hehn, Augustin Bussy, Fabian Belleflamme, Gloria Tabacchi, Andreas Glöß, Michael Lass, Iain Bethune, Christopher J. Mundy, Christian Plessl, Matt Watkins, Joost VandeVondele, Matthias Krack, and Jürg Hutter. Cp2k: An electronic structure and molecular dynamics software package - quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics*, 152(19), May 2020. ISSN 1089-7690. doi: 10.1063/5.0007045. URL `http://dx.doi.org/10.1063/5.0007045`.

[41] Daniel S Levine, Muhammed Shuaibi, Evan Walter Clark Spotte-Smith, Michael G Taylor, Muhammad R Hasyim, Kyle Michel, Ilyes Batatia, Gábor Csányi, Misko Dzamba, Peter Eastman, et al. The open molecules 2025 (omol25) dataset, evaluations, and models. *arXiv preprint arXiv:2505.08762*, 2025.

[42] 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), April 2025. ISSN 1089-7690. doi: 10.1063/5.0251713. URL `http://dx.doi.org/10.1063/5.0251713`.

[43] Antoine Loew, Dewen Sun, Hai-Chen Wang, Silvana Botti, and Miguel AL Marques. Universal machine learning interatomic potentials are ready for phonons. *npj Computational Materials*, 11(1):178, 2025.

[44] Denghui Lu, Han Wang, Mohan Chen, Lin Lin, Roberto Car, Weile Jia, Linfeng Zhang, et al. 86 pflops deep potential molecular dynamics simulation of 100 million atoms with ab initio accuracy. *Computer Physics Communications*, 259:107624, 2021.

[45] Ondrej Marsalek and Thomas E. Markland. Quantum dynamics and spectroscopy of ab initio liquid water: The interplay of nuclear and electronic quantum effects. *The Journal of Physical Chemistry Letters*, 8(7):1545–1551, March 2017. ISSN 1948-7185. doi: 10.1021/acs.jpclett.7b00391. URL `http://dx.doi.org/10.1021/acs.jpclett.7b00391`.

[46] 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, 2025. URL `https://arxiv.org/abs/2503.14118`.

[47] Jocelyn M Rodgers and John D Weeks. Local molecular field theory for the treatment of electrostatics. *Journal of Physics: Condensed Matter*, 20(49):494206, 2008.

[48] Jocelyn M Rodgers, Zhonghan Hu, and John D Weeks. On the efficient and accurate short-ranged simulations of uniform polar molecular liquids. *Molecular Physics*, 109(7-10):1195–1211, 2011.

[49] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, February 2024. ISSN 0925-2312. doi: 10.1016/j.neucom.2023.127063. URL `http://dx.doi.org/10.1016/j.neucom.2023.127063`.

[50] 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, May 2019. ISSN 1549-9626. doi: 10.1021/acs.jctc.9b00181. URL `http://dx.doi.org/10.1021/acs.jctc.9b00181`.

[51] Oliver T. Unke, Stefan Chmiela, Michael Gastegger, Kristof T. Schütt, Huziel E. Sauceda, and Klaus-Robert Müller. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications*, 12(1), December 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-27504-0. URL `http://dx.doi.org/10.1038/s41467-021-27504-0`.

[52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, 2017. URL `https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf`.

[53] Yusong Wang, Chaoran Cheng, Shaoning Li, Yuxuan Ren, Bin Shao, Ge Liu, Pheng-Ann Heng, and Nanning Zheng. Neural p3 m: A long-range interaction modeling enhancer for geometric gnns. *Advances in Neural Information Processing Systems*, 37:120336–120365, 2024.

[54] 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, 2025. URL `https://arxiv.org/abs/2506.23971`.

[55] Shuwen Yue, Maria Carolina Muniz, Marcos F. Calegari Andrade, Linfeng Zhang, Roberto Car, and Athanassios Z. Panagiotopoulos. When do short-range atomistic machine-learning models fall short? *The Journal of Chemical Physics*, 154(3), January 2021. ISSN 1089-7690. doi: 10.1063/5.0031215. URL `http://dx.doi.org/10.1063/5.0031215`.