Ewald-based Long-Range Message Passing for Molecular Graphs

Arthur Kosmala [1 2 3] **Johannes Gasteiger** [4 2] **Nicholas Gao** [2] **Stephan Günnemann** [2]

Abstract

Neural architectures that learn potential energy surfaces from molecular data have undergone fast improvement in recent years. A key driver of this success is the Message Passing Neural Network (MPNN) paradigm. Its favorable scaling with system size partly relies upon a spatial distance limit on messages. While this focus on locality is a useful inductive bias, it also impedes the learning of long-range interactions such as electrostatics and van der Waals forces. To address this drawback, we propose Ewald message passing: a nonlocal Fourier space scheme which limits interactions via a cutoff on frequency instead of distance, and is theoretically well-founded in the Ewald summation method. It can serve as an augmentation on top of existing MPNN architectures as it is computationally inexpensive and agnostic to architectural details. We test the approach with four baseline models and two datasets containing diverse periodic (OC20) and aperiodic structures (OE62). We observe robust improvements in energy mean absolute errors across all models and datasets, averaging $10\,\%$ on OC20 and $16\,\%$ on OE62. Our analysis shows an outsize impact of these improvements on structures with high long-range contributions to the ground truth energy.

1. Introduction

Graph neural networks (GNNs) have shown great promise in learning molecular properties from quantum chemical reference data, speeding up inference times by several orders of magnitude (Gilmer et al., 2017) while often preserving the accuracy of their ground truth method. Unfortunately, despite recent success in understanding theoretically universal models (Gasteiger et al., 2021; Dym & Maron, 2021), computational constraints still hamper GNNs in learning long-ranged physical interactions between atoms: the standard notion of a molecular graph (if no hand-crafted bond information is used) treats atoms as vertices and connects them by an edge if their 3D distance is within a cutoff hyperparameter (Schütt et al., 2018). This can be a problem if more remote atom interactions contribute to a target: they are long-ranged also on the graph, potentially relating vertices many edges apart. There is ample empirical and theoretical evidence of limited GNN performance in this case, e.g., due to over-squashing (Alon & Yahav, 2021).

The GNN *status quo* is in sharp contrast to the reality in molecular dynamics, where long-range interactions are an elementary ingredient. A task at the heart of the field (and of this work in particular) is to predict the potential energy $E(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N)$ based on the atom positions $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ in a molecule. Hand-crafted parametrizations of this map, known as empirical force fields, distinguish between bonded and non-bonded energy terms, $E = E^{\text{b}} + E^{\text{nb}}$ (Leach, 2001). The bonded terms model the short-ranged nature of covalent bonding with its complex, but highly local interactions. The non-bonded terms, however, capture interactions with heavy-tailed (typically power-law) decay in distance. An example of such a term is the electrostatic interaction,

$$E^{\text{es}} = \sum_{i<j} q_i q_j \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{-1}, \tag{1}$$

where the partial charges q_i quantify net electric charge around atom i. Their attraction or repulsion decays inversely with distance, and the energy is the sum of all atom pair interactions. Due to its scale-free decay, simply truncating such a power law with a distance cutoff can induce severe artifacts in measurable thermochemical predictions, even at outsized cutoffs (Wells & Chaffee, 2015). Instead, the interactions of all atom pairs have to be included in the sum. The technique of *Ewald summation* achieves this efficiently. It decomposes the actual physical interaction into a short-range and long-range part. The short-range part can be summed directly with a distance cutoff; the long-range part, while slowly decaying in distance, has a quickly decaying Fourier transform and can be efficiently evaluated as a Fourier space sum with a frequency cutoff (cf. Fig. 1).

[1]Max Planck Institute for the Science of Light & Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany [2]Department of Computer Science & Munich Data Science Institute, Technical University of Munich, Germany [3]Department of Physics, Ludwig-Maximilians-Universität München, Germany [4]Google Research. Correspondence to: Arthur Kosmala <arthur.kosmala@tum.de>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

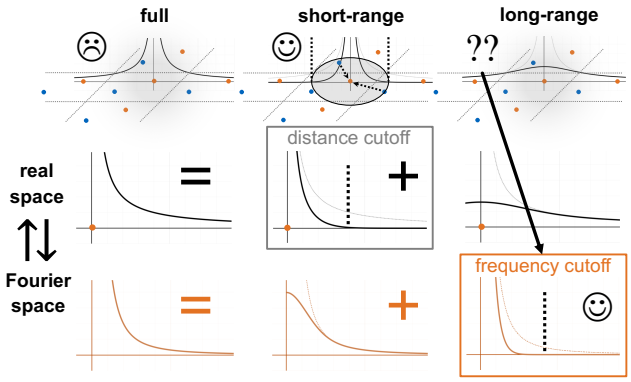

Figure 1. Ewald decomposition of an interaction, exhibiting $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{-1}$ decay in this example, into short- and long-range parts.

This work lays out a bridge between Ewald summation and the current landscape of GNN models. Our proposed method of *Ewald message passing* (Ewald MP) is a general framework that complements existing GNN layers in analogy to how the frequency-truncated long-range part complements the distance-truncated short-range part in Ewald summation. It is architecture-agnostic and computationally efficient, which we demonstrate by implementing and testing it as a modification on top of existing GNN models. Our experiments on the periodic structures of the OC20 dataset and the aperiodic structures of the OE62 dataset indicate robust improvements in energy mean absolute errors averaging 10 % and 16 % OC20 and OE62, respectively. Moreover, we study the tradeoff of these improvements against runtime and parameter count and find that it cannot be consistently replicated by any tested baseline setting. In a closer examination of the OE62 results, we observe that Ewald message passing recovers the effects of DFT-D3, a hand-crafted long-range correction term. This supports our claim that the method targets the learning of long-range interactions in particular. In summary, our contributions are:

- The new framework of **Ewald message passing**, which is readily integrated with existing GNN models.
- Insights abouts its **cost-performance** tradeoff in comparison to the wider GNN design landscape.
- A study underscoring the utility of Ewald message passing as a **long-range correction scheme**.

2. Background

Message passing neural networks. MPNNs were introduced by Gilmer et al. (2017) as a conceptual framework covering many molecular GNN architectures. Molecules are represented as graphs with atoms i as vertices. Atoms are adjacent, $j \in \mathcal{N}(i)$, if $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| < c_x$, where c_x is the distance cutoff. An MPNN first featurizes atoms as embeddings $h_i^0 \in \mathbb{R}^F$ based on local atom properties alone. Next, it models the complex atomic interactions by iteratively updating the embeddings in *message* and *update* steps,

$$M_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} f_{\text{int}}\left(h_i^{(l)}, h_j^{(l)}, e_{(ij)}\right), \tag{2}$$

$$h_i^{(l+1)} = f_{\text{upd}}\left(h_i^{(l)}, M_i^{(l+1)}\right), \tag{3}$$

where the *message sum* $M_i^{(l+1)}$ gathers information from the neighborhood $\mathcal{N}(i)$, using a function f_{int} that depends on current embeddings and edge features $e_{(ij)}$ (e.g., pairwise distances). Next, a learnable function f_{update} processes the message sums. A *readout step* then generates outputs (e.g., atom-wise energy contributions) from the embeddings. We develop the core framework of Ewald message passing around a particular type of message step, the continuous-filter convolution (Schütt et al., 2018):

$$M_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} h_j^{(l)} \cdot \Phi^{(l)}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \tag{4}$$

where $\cdot$ acts component-wise in the feature dimension, and $\Phi^{(l)}(\|\cdot\|) \colon \mathbb{R}^3 \longrightarrow \mathbb{R}^F$ is a stack of F learned, radial filters. In every channel $1 \leq d \leq F$, $\Phi_d^{(l)}(\|\cdot\|)$ learns attending to fuzzy distance ranges around an atom, allowing regression tasks on continuously varying molecular geometries.

Periodic systems. The inadequacy of a distance cutoff for generic interactions becomes particularly apparent in materials, where interatomic distances are unbounded.

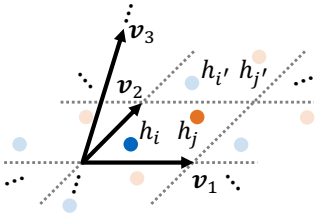

Figure 2. Schematic illustration of periodic boundary conditions.

Periodic boundary conditions (PBC) can be used to approximate an infinite system as repeating copies of a finite atom collection (cf. Fig. 2), tiling up space within *supercells*[1] spanned by the lattice vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) \in \mathbb{R}^{3\times 3}$. Let $\mathcal{S} = \{(h_1, \boldsymbol{x}_1), \ldots (h_N, \boldsymbol{x}_N)\}$ denote the atom embeddings and positions of one chosen supercell and let $\boldsymbol{t} \in \Lambda$ be any integer shift on the supercell lattice $\Lambda := \{\lambda_1 \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 + \lambda_3 \boldsymbol{v}_3 \mid \boldsymbol{\lambda} \in \mathbb{Z}^3\}$. Atoms i' at $\boldsymbol{x}_{i'} := \boldsymbol{x}_i + \boldsymbol{t}$ are called *images* of $i \in \mathcal{S}$, with the PBC fixing $h_{i'} \equiv h_i$.

Under PBC, generic message sums become infinite series as every image may contribute a message. Previous MPNN work on materials mitigates this issue with a distance cutoff, alongside a multigraph representation that avoids storing redundant $h_{i'}$ on separate vertices (Xie & Grossman, 2018; Chanussot et al., 2021; Cheng et al., 2021; Xie et al., 2021).

[1] A supercell should not be confused with the unit cell of a crystalline material. Unit cell translation symmetry is broken in non-equilibrium structures, but PBC are still applied, with a typical supercell containing multiple unit cells.

3. Ewald Message Passing

Analogy to electrostatics. The reason why Ewald summation can be applied to MPNNs is a formal correspondence between electrostatics-type interactions and continuous-filter convolutions. Previous work has highlighted this analogy as a useful source of inductive bias (Anderson et al., 2019). The connection becomes clear if we take Eq. (1) for the electrostatic energy E^{es} and express it in a new notation alluding to the continuous-filter convolution of Eq. (4):

$$E^{\text{es}} = \frac{1}{2}\sum_{i=1}^{N} q_i \cdot V_i^{\text{es}}(\boldsymbol{x}_i), \tag{5}$$

$$V_i^{\text{es}}(\boldsymbol{x}_i) = \sum_{j\neq i} q_j \cdot \Phi^{\text{es}}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \quad \Phi^{\text{es}}(r) = \frac{1}{r}, \tag{6}$$

where N is the total number of atoms ($N := \infty$ in materials) and we introduced the *electric potentials* $\{V_i^{\text{es}} \mid i = 1, \ldots, N\}$ and the *interaction kernel* Φ^{es}. Compare the expression for V_i^{es} to the message sum $M_i^{(l+1)}$ in Eq. (4): setting the cutoff $c_x \to \infty$ (such that $\mathcal{N}(i) = \{1, \ldots, N\}\backslash\{i\}$) and substituting

$$q_j \leftrightarrow h_j^{(l)}, \quad \Phi^{\text{es}} \leftrightarrow \Phi^{(l)}, \quad V_i^{\text{es}}(\boldsymbol{x}_i) \longleftrightarrow M_i^{(l+1)} \tag{7}$$

puts F-dimensional embedding vectors in place of scalar-valued partial charges as atom descriptors, and replaces the simple, domain-informed $1/r$ kernel by a stack of F learned convolution filters formally acting as interaction kernels. Any continuous-filter message sum M_i (we drop the layer superscript (l) from now on) can be *canonically identified* with a potential function $V_i(\boldsymbol{x}_i)$ evaluated at source atom location $\boldsymbol{x}_i$. Aggregating atomic output after just one message-passing step would inductively bias the model towards two-body interaction energies of similar form to E^{es}. Several message and update steps in principle enable the modeling of n-body terms (Batatia et al., 2022).

Conceptual outline. We now give an overview of Ewald summation and show how it connects to our learning approach, Ewald message passing, based on the above analogy.

Ewald summation evaluates potentials (equivalently, message sums) $V_i(\boldsymbol{x}_i) = M_i$ given a PBC lattice, a generic interaction kernel $\Phi\colon \mathbb{R}^+ \to \mathbb{R}^F$ and generic atom "charges" $h_i \in \mathbb{R}^F$. We apply Ewald summation if Φ is long-ranged.

In this case, the naive distance cutoff approximation $M_i \approx \sum_{j\neq i, \|\boldsymbol{x}_i - \boldsymbol{x}_j\| < c_x} h_j \cdot \Phi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|)$ becomes inaccurate for practically-sized values of c_x. A slow distance decay of Φ means that a significant part of its tail remains outside of c_x, even for large c_x. This long-range information is subsequently lost by truncation. Ewald summation presumes a known decomposition $\Phi(r) = \Phi^{\text{sr}}(r) + \Phi^{\text{lr}}(r)$ into a *short-range* and *long-range* kernel, where

- Φ^{sr} has fast decay in distance,
- $\Phi^{\text{lr}}(0)$ is well-defined, and $\hat{\Phi}^{\text{lr}}(\|\cdot\|)$, the 3D Fourier transform of $\Phi^{\text{lr}}(\|\cdot\|)$, has fast decay in frequency.

By construction, atom sums over distance-truncated short-range kernels $\Phi^{\text{sr}}_{[0,c_x]} := \Phi^{\text{sr}} \cdot \mathbf{1}_{[0,c_x]}$ (where $\mathbf{1}_{[0,c_x]}$ is the indicator function on $[0, c_x]$) converge quickly for growing c_x, meaning that a distance cutoff can be efficiently applied. This cannot be the case for Φ^{lr}, the leftover fat tail of Φ. Ewald summation deals with this *long-range sum* over Φ^{lr} by rewriting it as a sum over the Fourier transform $\hat{\Phi}^{\text{lr}}$. Replacing $\hat{\Phi}^{\text{lr}}$ by $\hat{\Phi}^{\text{lr}}_{[0,c_k]} := \hat{\Phi}^{\text{lr}} \cdot \mathbf{1}_{[0,c_k]}$, i.e., truncation by a *frequency cutoff* c_k, now admits efficient evaluation of the long-range sum. Adding the short-range and long-range sums recovers the sum over the full kernel Φ. Fig. 1 summarizes the approach. Convenient decompositions satisfying the above properties for all power law kernels are provided by Nijboer & De Wette (1957).

Ewald message passing essentially flips the rationale behind Ewald summation: instead of starting with a known physical kernel Φ and searching for a decomposition with the above properties, we seek to *parametrize* a filter Φ that is not short-ranged. However, we impose that it can be fitted by an Ewald sum over cutoff-supported short-range and long-range kernels. The kernels are learned independently of each other, $\hat{\Phi}^{\text{lr}}_{[0,c_k]}$ is learned directly in Fourier space.

This section proceeds by rewriting the crucial long-range sum as a sum over Fourier frequencies, bringing it into a form that admits a frequency cutoff. This prepares a subsequent in-depth discussion of Ewald message passing and its implementation as part of existing GNN models.

Frequency truncation. From here on, we assume that the long-range sum runs over *all* atoms, *including* target atom i, and thereby denote it by $M^{\text{lr}}(\boldsymbol{x}_i)$ (without subscript-i, but highlighting the dependence on source atom position $\boldsymbol{x}_i$). Note that this minor redefinition turns the map $\boldsymbol{x} \mapsto M^{\text{lr}}(\boldsymbol{x})$ into a periodic function. We check this by phrasing it as a double sum over the atoms of one supercell $\mathcal{S}$ and the supercell lattice Λ:

Figure 4. Sketched surfaces of constant phase for Fourier modes $\exp(i\boldsymbol{k}^T\boldsymbol{x})$ at small $\boldsymbol{k} \in \Lambda'$.

$$\begin{aligned} M^{\text{lr}}(\boldsymbol{x}) &= \sum_{\boldsymbol{t}\in\Lambda}\sum_{j\in\mathcal{S}} h_j \cdot \Phi^{\text{lr}}(\|\boldsymbol{x} - (\boldsymbol{x}_j + \boldsymbol{t})\|) \\ &= M^{\text{lr}}(\boldsymbol{x} + \tilde{\boldsymbol{t}}), \qquad \forall \tilde{\boldsymbol{t}} \in \Lambda. \end{aligned} \tag{8}$$

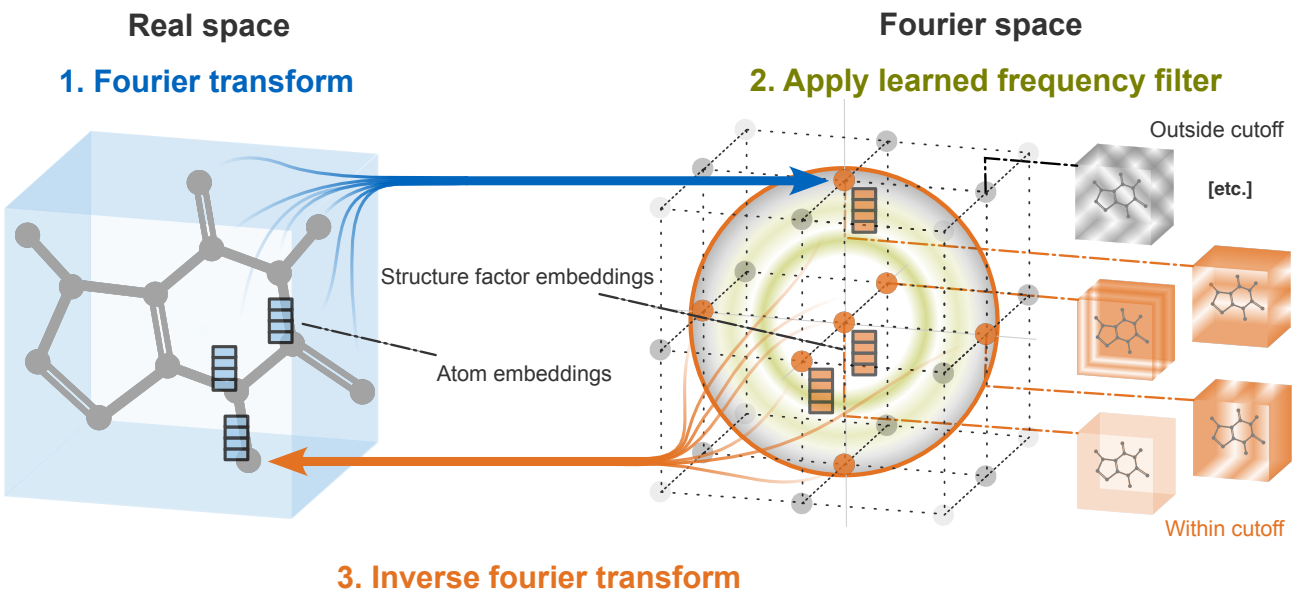

Figure 3. Three-part structure of long-range message passing: first, structure factor embeddings $\{s_{\boldsymbol{k}}\}$ are evaluated as a Fourier sum over atoms $\sum_{j\in\mathcal{S}} \exp(-i\boldsymbol{k}\boldsymbol{x}_j)\,[\,\cdot\,]$ and stored. Next, these are multiplied $\boldsymbol{k}$-wise with learned filter values. Finally, an inverse Fourier sum $\sum_{\boldsymbol{k}} \exp(i\boldsymbol{k}\boldsymbol{x}_i)[\,\cdot\,]$ scatters messages back to each atom i. Note the nonlocal *all-to-one* character of this scheme. If PBC are explicitly prescribed, the Fourier domain is given by $\{\boldsymbol{k}\in\Lambda' \mid \|\boldsymbol{k}\| < c_k\}$ with frequency cutoff c_k. If not, it needs to be suitably discretized.

Thus, M^{lr} has a Fourier series expansion (we use complex exponentials as Fourier modes for notational compactness):

$$M^{\text{lr}}(\boldsymbol{x}_i) = \sum_{\boldsymbol{k}\in\Lambda'} \hat{M}^{\text{lr}}_{\boldsymbol{k}} \exp\left(i\boldsymbol{k}^T\boldsymbol{x}_i\right) \tag{9}$$

with Fourier coefficients $\hat{M}^{\text{lr}}_{\boldsymbol{k}} \in \mathbb{C}^F$ (determined next). Just as frequencies of a 1D Fourier series must be integer multiples of the basis frequency (i.e., form a 1D lattice), the 3D spatial frequencies $\boldsymbol{k} \in \Lambda' \subset \mathbb{R}^3$ must be integer combinations of *three* spatial basis frequencies $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3) \in \mathbb{R}^{3\times 3}$ spanning the *reciprocal lattice* $\Lambda' := \{\lambda'_1\boldsymbol{w}_1 + \lambda'_2\boldsymbol{w}_2 + \lambda'_3\boldsymbol{w}_3 \mid \boldsymbol{\lambda'} \in \mathbb{Z}^3\}$. Each basis frequency amounts to a single-period phase shift along one supercell direction and constant phase along the others, i.e., $\boldsymbol{w}_i^T\boldsymbol{v}_j = 1$ if $i = j$ and 0 otherwise (cf. Fig. 4). The supercell lattice Λ therefore fixes the reciprocal lattice Λ', and the relations $\boldsymbol{w}_1 = 2\pi(\boldsymbol{v}_2\times\boldsymbol{v}_3)\Omega^{-1}$, $\boldsymbol{w}_2 = 2\pi(\boldsymbol{v}_3\times\boldsymbol{v}_1)\Omega^{-1}$, $\boldsymbol{w}_3 = 2\pi(\boldsymbol{v}_1\times\boldsymbol{v}_2)\Omega^{-1}$ (with the cross product $\times$ and supercell volume $\Omega = \boldsymbol{v}_1^T(\boldsymbol{v}_2\times\boldsymbol{v}_3)$) are easily verified.

As our final ingredient leading up to a frequency cutoff, we relate the Fourier *coefficients* $\hat{M}^{\text{lr}}_{\boldsymbol{k}}$ to the 3D Fourier *transform* $\hat{\Phi}^{\text{lr}}(\|\cdot\|)\colon \mathbb{R}^3 \to \mathbb{R}^F$ of $\Phi^{\text{lr}}(\|\cdot\|)$ – the kernel is aperiodic unlike $M^{\text{lr}}(\cdot)$, implying that its frequency spectrum is a continuous, radial function of $\boldsymbol{k}\in\mathbb{R}^3$ rather than a discrete set of lattice coefficients. In App. A.5, we derive the identity $\hat{M}^{\text{lr}}_{\boldsymbol{k}} = \frac{1}{\Omega}\sum_{j\in\mathcal{S}} h_j \exp(-i\boldsymbol{k}^T\boldsymbol{x}_j)\hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|)$. Essentially, Eq. (8) periodizes $\sum_{j\in\mathcal{S}} h_j\cdot\Phi^{\text{lr}}(\|\cdot - \boldsymbol{x}_j\|)$ by adding all of its lattice shifts on Λ; in Fourier space, this periodization maps $\hat{\Phi}^{\text{lr}}(\|\cdot\|)$ onto a discrete spectrum by recording its values only at PBC-compatible frequencies in Λ'. Inserting this back into Eq. (9), and absorbing the prefactor $\frac{1}{\Omega}$ in our *learned* filter by redefining $\frac{1}{\Omega}\hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|) \mapsto \hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|)$, yields the desired long-range component of Ewald message passing, a message sum over Fourier frequencies:

$$M^{\text{lr}}(\boldsymbol{x}_i) = \sum_{\boldsymbol{k}\in\Lambda'} \exp(i\boldsymbol{k}^T\boldsymbol{x}_i)\cdot s_{\boldsymbol{k}}\cdot\hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|), \tag{10}$$

$$s_{\boldsymbol{k}} = \sum_{j\in\mathcal{S}} h_j \exp(-i\boldsymbol{k}^T\boldsymbol{x}_j), \tag{11}$$

with $\{s_{\boldsymbol{k}} \in \mathbb{C}^F \mid \boldsymbol{k}\in\Lambda'\}$, which we call *structure factor embeddings*, as inspired by the analogous crystallographic quantity. If we replace $\hat{\Phi}^{\text{lr}}$ by a cutoff-truncated Fourier space kernel $\hat{\Phi}^{\text{lr}}_{[0,c_k]}$, we can limit the outer sum $\sum_{\boldsymbol{k}\in\Lambda'} \mapsto \sum_{\boldsymbol{k}\in\Lambda',\|\boldsymbol{k}\|<c_k}$, while the inner sum computing $s_{\boldsymbol{k}}$ only extends over the atoms of a single supercell. The resulting cutoff-truncated kernel is complementary to regular continuous filters (Eq. (4)): While they have no limit on distance, the frequency cutoff does limit how fast the filters can change in space. More specifically, their spatial resolution is proportional to the inverse frequency cutoff.

Structure factor embeddings. Without a k-cutoff, the set of tuples $\{(s_{\boldsymbol{k}}, \boldsymbol{k}) \mid \boldsymbol{k}\in\Lambda'\}$ would hold equivalent information to the set $\{(h_i, \boldsymbol{x}_i) \mid i\in\mathcal{S}\}$, i.e., fully quantify the hidden state of the model on a given input structure. This is seen by viewing $\{(h_i, \boldsymbol{x}_i) \mid i\in\mathcal{S}\}$ equivalently as an "embed-

ding density distribution", $\rho(\boldsymbol{x}) = \sum_{\boldsymbol{t}\in\Lambda}\sum_{i\in\mathcal{S}} h_i \delta(\boldsymbol{x} - \boldsymbol{x}_i - \boldsymbol{t})$, with δ denoting the Dirac delta distribution. Its Fourier transform (in the distributional sense) is given by $\hat{\rho}(\boldsymbol{k}) = \sum_{\boldsymbol{k}'\in\Lambda'}\left[\sum_{j\in\mathcal{S}} h_i \exp(-i\boldsymbol{k}'^T\boldsymbol{x}_j)\right]\delta(\boldsymbol{k}-\boldsymbol{k}') = \sum_{\boldsymbol{k}'\in\Lambda'} s_{\boldsymbol{k}'}\delta(\boldsymbol{k}-\boldsymbol{k}')$ which corresponds to the pairs $\{(s_{\boldsymbol{k}}, \boldsymbol{k})\}$. Disregarding all $s_{\boldsymbol{k}}$ outside a k-cutoff removes fine-grained resolution information about the atom encoding distribution below a scale $\sim c_k^{-1}$. In this sense, the structure factor representation introduced in Ewald message passing coarse-grains the hidden model state. Moreover, the embedding information is represented non-locally since each $s_{\boldsymbol{k}}$ is a weighted sum over *all* supercell embeddings. This all-to-one relationship is another way of understanding the long-range character of Ewald message passing.

Ewald message passing. The combined usage of regular short-range message passing via *any* GNN layer and the long-range message passing in Eqs. (10) and (11) is called *Ewald message passing*. This combination enables Ewald message passing to express interactions outside of either limited class. We combine the two message passing schemes simply by adding up both embedding updates during each message-passing step, for details see Eq. (12). This way of integration works nearly out-of-the-box in our experiments.

Aperiodic systems. Many structural datasets containing important long-range interactions do not feature periodic boundary conditions. In this case, a reciprocal lattice that would discretize the Fourier domain is not provided, which in turn means the structure factor embeddings $s_{\boldsymbol{k}}$ turn into a function $s(\boldsymbol{k})\colon \mathbb{R}^3 \mapsto \mathbb{C}^F$ defined on a continuous domain. The aperiodic case can equally well be understood as the limit of infinite supercell size, corresponding to a continuum limit of the reciprocal lattice as the basis frequency vectors $\boldsymbol{w}_i$ approach zero. As we can only represent $s(\cdot)$ and $\hat{\Phi}(\|\cdot\|)$ by a finite set of numbers, we need to impose a suitable discretization scheme ourselves. In our implementation, we tile the cutoff region by a grid of cubic voxels and replace all continuous functions by the set of their average values within each voxel. The discretization resolution is controlled by the voxel sidelength Δ, a new hyperparameter that is relevant for the aperiodic case. In our main experiments, we use equal resolution $\Delta = 0.2\,\text{Å}^{-1}$ for each tested model. Ablation studies with other resolutions are reported in App. A.3. Note that this approach preserves rotation invariance only if the voxel grid sits in a coordinate frame that rotates *with* the input structure. To this end, we use the orthonormal basis obtained from a singular value decomposition, similar to Gao & Günnemann (2022). In App. A.6 we outline a more general mathematical framework for discretizing the structure factor, provide implementation-relevant details about our voxel approach and discuss its relation to other possible discretization schemes.

Frequency filters: aperiodic case. Thus far, we left open how the frequency filters $\hat{\Phi}(\|\cdot\|)$ are learned. In the aperiodic case, we parametrize filters as linear combinations of S radial basis functions $\{\hat{\psi}_i(\|\cdot\|)\colon \mathbb{R}^3 \to \mathbb{R} \mid i = 1, \ldots, S\}$: $\hat{\Phi}(\|\boldsymbol{k}\|) = \boldsymbol{W}\hat{\Psi}(\boldsymbol{k})$ with $\hat{\Psi}(\boldsymbol{k}) = (\psi_1(\|\boldsymbol{k}\|), \ldots \psi_S(\|\boldsymbol{k}\|))$ and the learned weight matrix $\boldsymbol{W} \in \mathbb{R}^{F\times S}$. All basis functions are supported within the k-cutoff radius c_k. We use Gaussian radial basis functions (Schütt et al., 2018).

Frequency filters: periodic case. On OC20 we pursue a non-radial filtering strategy that allows for a more liberal use of the information provided by the PBC frame $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) \in \mathbb{R}^{3\times 3}$ of a given input structure. This information can be meaningful: given a material surface slab, for example, the model should clearly "know" which direction is normal to the surface. This does not break rotation invariance as the PBC frame rotates *with* the structure.

Since the reciprocal lattice Λ' varies depending on the input PBC frame, the number of frequencies inside any fixed cutoff volume would also be variable. Motivated by our empirical findings (App. A.3), we instead *fix* the number of frequencies in each reciprocal basis direction $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3) \in \mathbb{R}^{3\times 3}$, dropping the notion of a direct frequency cutoff. Formally, we define positive integers N_x, N_y, N_z and consider the product set $I = \bigtimes_{j=x,y,z}\{-N_j, \ldots, N_j\} \subset \mathbb{Z}^3$. We limit the scope of the Fourier message sum in Eq. (10) by only including frequencies with reciprocal lattice positions in I. The model then learns a filter weight $\hat{\Phi}_{\boldsymbol{\lambda}'} \in \mathbb{R}^F$ for each $\boldsymbol{\lambda}' \in I$. Since a general supercell lattice has at least point symmetry, we moreover set $\hat{\Phi}_{\boldsymbol{\lambda}'} \equiv \hat{\Phi}_{-\boldsymbol{\lambda}'}$. This differs from the strict notion of frequency filters discussed earlier: the filter weights $\hat{\Phi}_{\boldsymbol{\lambda}'}$ are not assigned to a fixed point $\boldsymbol{k}$ in Fourier space, but to a position on the reciprocal lattice. The frequency corresponding to this position $\boldsymbol{k}_{\boldsymbol{\lambda}'} = \lambda_1'\boldsymbol{w}_1 + \lambda_2'\boldsymbol{w}_2 + \lambda_3'\boldsymbol{w}_3$ varies with the basis frequencies of each input structure. Note that this filtering approach is not invariant under the "supercell symmetry" that maps any supercell to its multiple across one or several lattice directions while keeping atom positions unchanged. We found that this does not affect empirical performance on OC20. Still, a filtering approach invariant to this symmetry (e.g. radial basis filtering) might be advisable for datasets with large variations in supercell size.

Computational complexity. The three-part procedure of our long-range message passing scheme, illustrated in Fig. 3, has a computational complexity of $\mathcal{O}(N_{\text{at}}N_{\text{k}})$, where N_{at} is the number of atoms in the structure (or supercell in the case of PBC), and N_{k} is the number of frequencies contained in the long-range sum. As the density of atoms in space can be roughly assumed constant, the atom count increases linearly with the supercell volume Ω in the periodic case. By the definition of the reciprocal basis $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3) \in \mathbb{R}^{3\times 3}$,

the amount of frequencies $\boldsymbol{k} \in \Lambda'$ contained within a fixed frequency cutoff volume therefore scales linearly with the atom count N_{at}. Altogether, this would amount to $\mathcal{O}(N_{\text{at}}^2)$ scaling if both cutoffs are left fixed. In the aperiodic case, the same holds if we adapt the voxel resolution Δ to the finest features of s, which, by Fourier reciprocity, scale inversely to the structural length. However, in contrast to standard Ewald summation (Wells & Chaffee, 2015), Ewald MP does not put any fixed relationship between c_x and c_k. To achieve a computational complexity of $\mathcal{O}(N_{\text{at}})$, we can instead choose to fix N_{k}, which effectively varies $c_k \propto N_{\text{at}}^{-1}$. Note that this approach can leave a "gap" of medium-range interactions that are covered neither by the short-range nor the long-range component of Ewald MP. Another way to achieve sub-quadratic scaling would be adapting more efficient Ewald summation variants to MPNNs (cf. Sec. 4).

4. Related Work

Hand-crafted long-range corrections. Approaches augmenting a short-ranged model by a long-range correction make a frequent appearance in computational chemistry. This often concerns corrections to density functional theory (DFT), the method behind the OC20 (Chanussot et al., 2021), OE62 (Stuke et al., 2020) and many other machine learning benchmarks. DFT suffers from the same type of problem as MPNN potentials: its standard functionals like LDA (Jones & Gunnarsson, 1989) or PBE (Perdew et al., 1996) cannot correctly express long-range effects such as London dispersion (Jones, 2015). A zoo of interatomic dispersion corrections exists (Grimme, 2006; Tkatchenko & Scheffler, 2009; Grimme et al., 2010; Caldeweyher et al., 2019), which can be adapted for use alongside many DFT functionals. From this historic angle, Ewald MP fills the same demand for an inexpensive, out-of-the-box correction in the MPNN context. Others have combined long-range terms of various physics-based functional forms with short-range models, including fully hand-crafted terms (Staacke et al., 2021) and electrostatic terms with MPNN-predicted partial charges (Unke et al., 2021). In contrast, our approach remains entirely within the MPNN paradigm, focusing on a fully data-driven modeling of long-range effects.

Non-local learning schemes. Ewald message passing falls into the broader category of learning approaches targeting non-local interactions in a structure. An example of prominence well beyond quantum chemistry is attention (Vaswani et al., 2017). Whereas early approaches in the graph domain apply the attention mechanism to local node neighborhoods only (Veličković et al., 2018; Dwivedi & Bresson, 2021), others soon extended it to the fully-connected graph (Ying et al., 2021; Kreuzer et al., 2021; Mialon et al., 2021). A drawback of such standard, fully-connected graph attention is its $\mathcal{O}(N^2)$ scaling in structure size. However, models using subquadratic alterations of standard attention have been proposed, including an $\mathcal{O}(N \log(N))$ non-local aggregation scheme (Liu et al., 2022), as well as linearly-scaling modified attention layers for both chemistry (Unke et al., 2021; Frank et al., 2022) and generic graph applications (Rampášek et al., 2022). Shallow models such as Gaussian-process-based sGDML (Chmiela et al., 2018) also typically use locality approximations, motivating recent effort towards their nonlocal extension by Chmiela et al. (2023). Such approaches are mostly orthogonal to the MPNN line of research and not directly informed by a computational physics technique as with Ewald MP.

Use of long-range sums. Some MPNN-based works share our approach of computing multiple message sums, in analogy to the physical separation of hand-crafted energies into bonded and non-bonded terms. They likewise use a more complex short-range part which includes higher-order geometric information like bond (Zhang et al., 2020) or even dihedral angles (Gasteiger et al., 2022b), and a simpler, purely radial atom-to-atom long-range part. However, the long-range part in both works is a standard distance-truncated message sum in a larger atomic neighborhood, thereby precluding the scaling advantage and physical inductive bias achievable with Ewald summation. Apart from this, two concurrent works exist which combine forms of Ewald summation with MPNN models. Lin et al. (2023) use it in a traditional sense to compute *fixed-form* physical potentials for downstream use as physics-informed edge features, rather than as a *free-form* ansatz for the learning of continuous-filter messages. Yu et al. (2022) share our notion of learnable Fourier space filters, but compute a scalar-valued structure factor instead of structure factor embeddings. More generally, Yu et al. (2022) do not use the Ewald part as a GNN message passing step with feedback onto the base GNN embeddings, but rather apply it once and then combine it with the short-range GNN output after its final layer. Both Ewald-related works exclusively consider the periodic case.

Ewald summation. Ewald summation is a technique for the summation of long-range interactions. The concept dates back to work in theoretical physics from over a century ago (Ewald, 1921) and has been adapted various times to encompass more general interactions (Nijboer & De Wette, 1957; Laino & Hutter, 2008; Mazars, 2010). Further development has focused on the numerically efficient evaluation of the long-range sum. While direct Ewald summation with optimized distance and frequency cutoffs has $N^{\frac{3}{2}}$ scaling in the atom number (Wells & Chaffee, 2015), approaches using the Fast Fourier Transform (Darden et al., 1993; York & Yang, 1994; Essmann et al., 1995) bring it down to $N \log(N)$. In this work, we focus on introducing the general method of Ewald message passing, and consider numerical improvements an interesting route for future work.

5. Experiments

Datasets. The OC20 dataset (Chanussot et al., 2021) features adsorption energies and atom forces for roughly 265 million structures, each one a snapshot from a DFT-computed relaxation trajectory of an adsorbate-catalyst combination under 3D periodic boundary conditions. We train our models on the OC20-2M subsplit, which has been empirically established as an efficient proxy for the full-size training set (Gasteiger et al., 2022b). The OE62 dataset (Stuke et al., 2020) features, among other targets, DFT-computed energies (in eV) for roughly 62,000 large organic molecules. The energies account for the long-ranged London dispersion interaction, computed via the DFT-D3 dispersion correction (Grimme et al., 2010). Both benchmarks contain structures with comparably large numbers of atoms ($N > 100$). OC20 features PBC, while structures in OE62 are aperiodic but can reach large spatial extent ($> 20\,\text{Å}$). This makes these two datasets a strong testing case for the long-range improvements achievable with Ewald message passing.

Baseline GNN models. An overview of training and baseline hyperparameters can be found in App. A.2. We test Ewald message passing by modifying four models: SchNet (Schütt et al., 2017), PaiNN (Schütt et al., 2021), DimeNet^{++} (Gasteiger et al., 2020) and GemNet-T (Gasteiger et al., 2021). These are intended as a representative, but by no means exhaustive sample from the modern GNN landscape. Guided by the settings used by Chanussot et al. (2021), most baseline atom embedding sizes and cutoffs are larger than the originally reported values to account for the size and chemical diversity of our structures.

Ewald modifications. We aim to upgrade each baseline architecture in a minimal way such that the Ewald additions can be studied in isolation. All models except DimeNet^{++} update atom embeddings in each interaction block using a skip connection. We simply add long-range message passing using M_i^{lr} from Eqs. (10) and (11) as another contribution,

$$\begin{aligned} h_i^{(l+1)} &= \frac{1}{\sqrt{2}}\left[h_i^{(l)} + f_{\text{upd}}^{\text{sr}}\left(M_i^{\text{sr}}\right)\right] \\ &\downarrow \\ h_i^{(l+1)} &= \frac{1}{\sqrt{3}}\left[h_i^{(l)} + f_{\text{upd}}^{\text{sr}}\left(M_i^{\text{sr}}\right) + f_{\text{upd}}^{\text{lr}}\left(M_i^{\text{lr}}\right)\right], \end{aligned} \tag{12}$$

where the short-range message sum M_i^{sr} and update function $f_{\text{upd}}^{\text{sr}}$ is specific to the short-range model, while $f_{\text{upd}}^{\text{lr}}$ is of the same kind across models (architecture details in App. A.1). For PaiNN, which also incorporates equivariant vector embeddings, we only upgrade messages between its scalar embeddings. As DimeNet^{++} does not pass messages between atom embeddings by default, we add them to the model with a default embedding block taken from GemNet-T and update them purely by long-range message passing with a skip connection, $h_i^{(l+1)} = \frac{1}{\sqrt{2}}\left[h_i^{(l)} + f_{\text{upd}}\left(M_i^{\text{lr}}\right)\right]$. We concatenate these atom embeddings with the atom-aggregated DimeNet^{++} edge embeddings along the feature dimension, prior to the atom-wise dense layer of each output block. Any remaining differences among our Ewald modifications are on the level of hyperparameters. Aiming for Pareto-efficient baseline improvements, we choose numbers of included k-vectors adequate to the cost of each baseline (see App. A.3).

Comparison studies. We test whether using Ewald message passing alongside our baselines leads to significant and robust improvements of the energy mean absolute errors (EMAEs) on OC20 and OE62. We also measure the average runtime per structure (details in App. A.8) for every model configuration and dataset to study how the benefits of Ewald message passing compare against its runtime overhead. Moreover, to establish a baseline for the long-range improvements achievable by a mere change in hyperparameters, we repeat the EMAE and runtime measurements in a configuration without Ewald MP, but with increased settings for the distance cutoff c_x and maximum number of message-passing neighbors N_{max} (details in A.4). An adequate comparison ideally means increasing these settings to a point that roughly matches the Ewald modification in cost. This is always possible for the periodic structures of OC20. In the case of OE62, a setting of $c_x = 12\,\text{Å}$ and $N_{\text{max}} = 100$ usually covers the entire molecule, meaning that no significant cost increase can be achieved beyond this point. We furthermore compare the Ewald improvements against a test setting where the Ewald block, $f_{\text{upd}}^{\text{lr}}\left(M_i^{\text{lr}}\right)$ in Eq. (12), is instead replaced by regular pairwise message passing with increased distance cutoff (similarly to related work discussed in Sec. 4). Concretely, we substitute it by a SchNet interaction block. In contrast to the increased-cutoff variant from before, we now extend the cutoff for this block *only*. As a final check, we study whether the benefits of Ewald MP are significant beyond what follows just from an increase in free parameters. On OE62, we measure the EMAE and runtime of baseline variants with more atom embedding channels to recover the parameter count of Ewald MP. For OC20, we point to the extensive hyperparameter studies by Gasteiger et al. (2022a), which suggest that changing embedding sizes by the small amounts relevant to this work has negligible impact on OC20 EMAEs.

First off, our comparison studies on OE62 (Fig. 5) and OC20 (Fig. 6) underscore the primary role of short-range model choice in the GNN design landscape (consider, e.g., how DimeNet^{++} is pareto-dominated on OC20). At the same time, the consistency of accuracy gains through Ewald MP strongly encourages its use alongside any particular short-range model. The associated runtime overheads are at least comparable to, if not smaller than the runtime differences between baselines. Meanwhile, the average Ewald

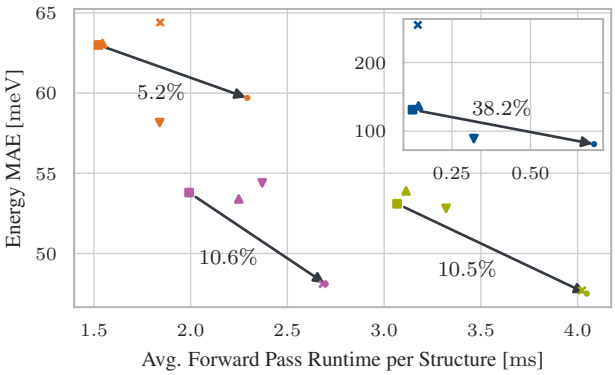

Figure 5. Cost vs. energy MAE on OE62. Ewald MP improves the models by 5 % to 38 %. Increasing the cutoff performs similarly for DimeNet^{++} and GemNet-T, but is even detrimental for SchNet and PaiNN. Using a SchNet in place of an Ewald LR block works well for SchNet and PaiNN, but leaves DimeNet^{++} and GemNet-T unaffected. Increasing the atom embedding size has little effect.

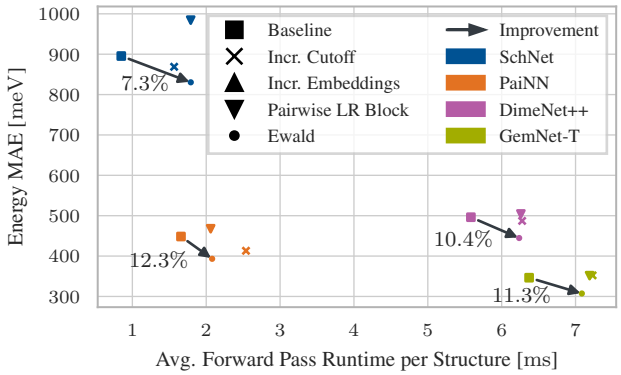

Figure 6. Cost vs. energy MAE on OC20, averaged across test splits. Ewald MP consistently improves all models and significantly outperforms the two baseline settings featuring an increased global cutoff as well as a standard pairwise (here: SchNet) block with increased distance cutoff in place of the Ewald LR block.

improvements across all models (and all test splits in the case of OC20) are 16.1 % for OE62, and 10.3 % for OC20. Ewald MP achieves a better cost-performance tradeoff than the large cutoff setting for all tested instances on OC20. Interestingly, the large cutoff setting on OE62 has a very comparable tradeoff to Ewald MP for the two baselines which use edge embeddings, but has a detrimental effect for the two models that only use atom embeddings. Quite the opposite happens on OE62 if the Ewald block is replaced by a SchNet block with increased distance cutoff: it is effective for PaiNN and SchNet, but achieves no improvement at all on the better DimeNet++ and GemNet-T baseline models. Its cost cannot be further increased as the added SchNet block already approaches full graph connectivity. On OC20, it always has a neutral or detrimental effect. In contrast to these two previous strategies, improvements through Ewald MP appear *robust* – no detrimental effect is observed in *any* tested model configuration. Finally, matching the free parameter count of Ewald MP with larger atom embeddings shows a negligible or even detrimental effect. Beyond these empirical comparisons, we note in general that pairwise message passing on a fully-connected graph (like our SchNet LR block setting on OE62) inevitably scales quadratically in structural size. Ewald MP, on the other hand, can propagate long-range information subquadratically while also providing a physically principled inductive bias (cf. Sec. 3).

Our results inform various design recommendations for future use. Ewald MP appears to be more efficient on OC20: similar relative improvements are consistently achieved at about half the relative cost. This hints at periodic structures as the more attractive domain of application. Given how we integrate Ewald MP with a short-range GNN, it is relatively cheaper alongside models featuring few but compute-heavy message passing blocks (in our case, DimeNet^{++} and GemNet-T). In such cases, adding a long-range message sum to each block results in less relative overhead. Besides energies, we also test for improvements of force MAEs on OC20, which are consistent but small. This is to be expected, since the frequency truncation scheme of long-range MP removes the fast-changing high-frequency contributions to the potential energy surface, limiting its impact on the energy gradients (forces). Further comments and numerical results are found in App. A.7 and A.11.

Impact on long-range interactions. To test our hypothesis that Ewald MP learns long-range effects in particular, we leverage the fact that the energy targets on OE62 contain an additive term, DFT-D3 (Grimme et al., 2010), to correct for the long-ranged London-dispersion interaction. This allows us to study the long-range component in isolation. To evaluate the DFT-D3 term, we use the coefficient set recommended for Becke-Johnson damping and the PBE0 DFT functional, as obtained from Reckien (2017). We proceed by assigning all 6000 molecules in the `OE62-val` dataset to one of 15 equally-sized bins according to the value of its DFT-D3 correction. We disregard the two outermost bins as they contain strong outliers. In each bin, we record the statistics (bootstrapped means and 68 % confidence intervals with 10000 resamples) of the energy MAE for the baseline and Ewald model variants. Fig. 7 shows that Ewald message passing gives an outsize relative improvement for structures with high long-range energy contributions. As an additional check, we collect the same statistics as previously for a model without Ewald modifications that is trained while the DFT-D3 correction is added explicitly to the model output. This provides a "cheating" model that only needs to learn the short-range component, while the long-range component

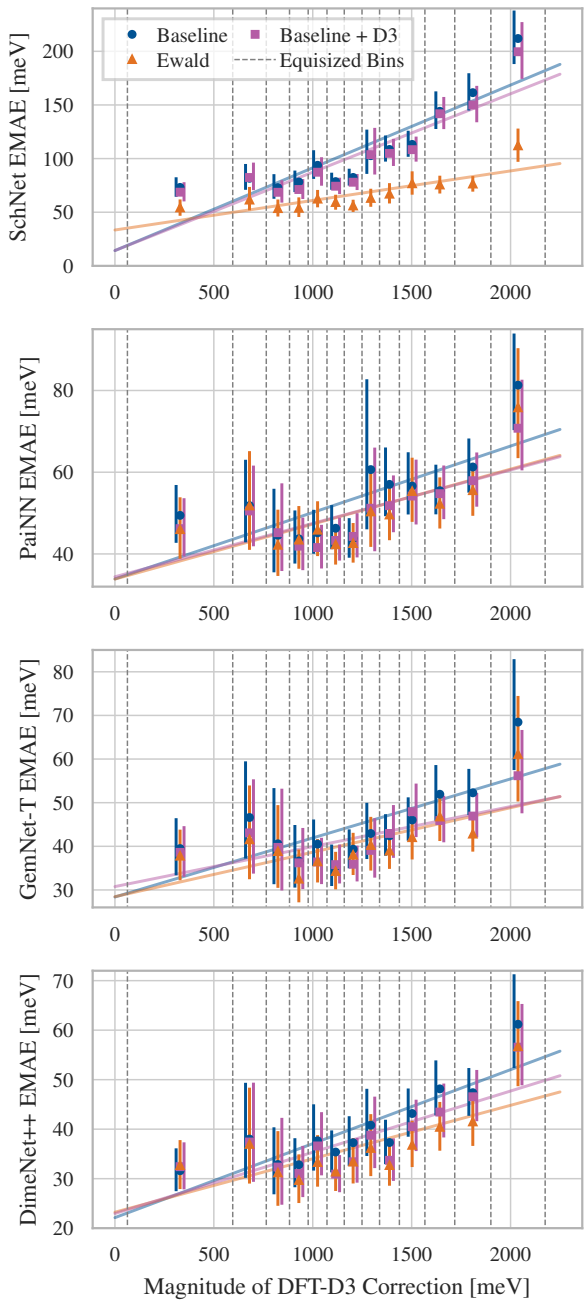

Figure 7. Energy MAE on `OE62-val` structures binned according to their DFT-D3 long-range term. Ewald MP has an outsize impact for molecules with larger long-range components, and recovers or surpasses a "cheating" baseline that just adds the D3 ground truth to the prediction ("Baseline + D3").

is perfectly reconstructed. Ewald message passing keeps up with this perfect baseline across all models, and even outperforms it for SchNet and DimeNet^{++}. This analysis shows that Ewald MP is able to recover the effects of a dispersion correction on OE62 without any need for hand-crafting.

6. Conclusion

In this work, we show how the analogy between sums of GNN messages and physical interactions leads to Ewald message passing, a general message-passing framework based on Ewald summation. Being computationally efficient and model-agnostic, it brings value as a plug-and-play addition to many existing GNN models. Our results demonstrate robust improvements in energy targets across models and datasets. They furthermore indicate an outsize impact on structures with high long-range energy contributions. We encourage using Ewald MP especially for large or periodic structures containing a diverse set of atoms.

Acknowledgements

We greatly thank Abhishek Das for his swift help in evaluating our models on the `OC20-test` split, as well as the anonymous reviewers for their instructive suggestions.

This work is funded by the Munich Quantum Valley by the Federal Ministry of Education and Research (BMBF), from funds of the Hightech Agenda Bayern Plus, and the Federal Ministry of Education and Research and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder.

References

Alon, U. and Yahav, E. On the Bottleneck of Graph Neural Networks and its Practical Implications. In *ICLR*, 2021. URL https://openreview.net/forum?id=i80OPhOCVH2.

Anderson, B. M., Hy, T.-S., and Kondor, R. Cormorant: Covariant Molecular Neural Networks. In *NeurIPS*, 2019. URL http://arxiv.org/abs/1906.04015.

Batatia, I., Batzner, S., Kovács, D. P., Musaelian, A., Simm, G. N. C., Drautz, R., Ortner, C., Kozinsky, B., and Csányi, G. The design space of e(3)-equivariant atom-centered interatomic potentials, 2022. URL https://arxiv.org/abs/2205.06643.

Caldeweyher, E., Ehlert, S., Hansen, A., Neugebauer, H., Spicher, S., Bannwarth, C., and Grimme, S. A generally applicable atomic-charge dependent london dispersion correction. *The Journal of Chemical Physics*, 150(15): 154122, 2019. doi: 10.1063/1.5090222. URL https://doi.org/10.1063/1.5090222.

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., Palizhati, A., Sriram, A., Wood, B., Yoon, J., Parikh, D., Zitnick, C. L., and Ulissi, Z. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catalysis*, 11(10):6059–6072, May 2021. doi:

10.1021/acscatal.0c04525. URL https://doi.org/10.1021/acscatal.0c04525. Publisher: American Chemical Society.

Cheng, J., Zhang, C., and Dong, L. A geometric-information-enhanced crystal graph network for predicting properties of materials. *Communications Materials*, 2 (1):92, Sep 2021. ISSN 2662-4443. doi: 10.1038/s43246-021-00194-3. URL https://doi.org/10.1038/s43246-021-00194-3.

Chmiela, S., Sauceda, H. E., Müller, K.-R., and Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, 9(1):1–10, September 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06169-2. URL https://www.nature.com/articles/s41467-018-06169-2.

Chmiela, S., Vassilev-Galindo, V., Unke, O. T., Kabylda, A., Sauceda, H. E., Tkatchenko, A., and Müller, K.-R. Accurate global machine learning force fields for molecules with hundreds of atoms. *Science Advances*, 9(2):eadf0873, 2023. doi: 10.1126/sciadv.adf0873. URL https://www.science.org/doi/abs/10.1126/sciadv.adf0873.

Darden, T., York, D., and Pedersen, L. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12): 10089–10092, 1993. doi: 10.1063/1.464397. URL https://doi.org/10.1063/1.464397.

Dwivedi, V. P. and Bresson, X. A generalization of transformer networks to graphs, 2021.

Dym, N. and Maron, H. On the Universality of Rotation Equivariant Point Cloud Networks. In *ICLR*, 2021. URL https://openreview.net/forum?id=6NFBvWlRXaG.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103 (19):8577–8593, 1995. doi: 10.1063/1.470117. URL https://doi.org/10.1063/1.470117.

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

Frank, T., Unke, O. T., and Muller, K. R. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=tlUnxtAmcJq.

Freysoldt, C., Neugebauer, J., Tan, A. M. Z., and Hennig, R. G. Limitations of empirical supercell extrapolation for calculations of point defects in bulk, at surfaces, and in two-dimensional materials. *Phys. Rev. B*, 105: 014103, Jan 2022. doi: 10.1103/PhysRevB.105.014103. URL https://link.aps.org/doi/10.1103/PhysRevB.105.014103.

Gao, N. and Günnemann, S. Ab-Initio Potential Energy Surfaces by Pairing GNNs with Neural Wave Functions. In *International Conference on Learning Representations*, April 2022.

Gasteiger, J., Giri, S., Margraf, J. T., and Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. In *Machine Learning for Molecules Workshop, NeurIPS*, December 2020. URL http://arxiv.org/abs/2011.14115. arXiv: 2011.14115.

Gasteiger, J., Becker, F., and Günnemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. In *NeurIPS*, November 2021. URL http://arxiv.org/abs/2106.08903. arXiv: 2106.08903.

Gasteiger, J., Shuaibi, M., Sriram, A., Günnemann, S., Ulissi, Z., Zitnick, C. L., and Das, A. How Do Graph Networks Generalize to Large and Diverse Molecular Systems? *arXiv*, 2204.02782, April 2022a. URL http://arxiv.org/abs/2204.02782. arXiv: 2204.02782.

Gasteiger, J., Shuaibi, M., Sriram, A., Günnemann, S., Ulissi, Z., Zitnick, C. L., and Das, A. Gemnet-oc: Developing graph neural networks for large and diverse molecular simulation datasets, 2022b. URL https://arxiv.org/abs/2204.02782.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural Message Passing for Quantum Chemistry. In Precup, D. and Teh, Y. W. (eds.), *ICML*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 2017. URL http://proceedings.mlr.press/v70/gilmer17a.html.

Grimme, S. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, 2006. doi: https://doi.org/10.1002/jcc.20495. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20495.

Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132 (15):154104, 2010. doi: 10.1063/1.3382344. URL https://doi.org/10.1063/1.3382344.

Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.*, 87: 897–923, Aug 2015. doi: 10.1103/RevModPhys.87.897. URL https://link.aps.org/doi/10.1103/RevModPhys.87.897.

Jones, R. O. and Gunnarsson, O. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61:689–746, Jul 1989. doi: 10.1103/RevModPhys.61.689. URL https://link.aps.org/doi/10.1103/RevModPhys.61.689.

Kreuzer, D., Beaini, D., Hamilton, W. L., Létourneau, V., and Tossou, P. Rethinking graph transformers with spectral attention. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=huAdB-Tj4yG.

Laino, T. and Hutter, J. Notes on "ewald summation of electrostatic multipole interactions up to quadrupolar level" [j. chem. phys. 119, 7471 (2003)]. *The Journal of Chemical Physics*, 129(7):074102, 2008. doi: 10.1063/1.2970887. URL https://doi.org/10.1063/1.2970887.

Leach, A. *Molecular Modelling: Principles and Applications*. Prentice Hall, Harlow, England ; New York, 2. edition edition, January 2001. ISBN 978-0-582-38210-7.

Lin, Y., Yan, K., Luo, Y., Liu, Y., Qian, X., and Ji, S. Efficient approximations of complete interatomic potentials for crystal property prediction, 2023. URL https://openreview.net/forum?id=KaKXygtEGK.

Liu, M., Wang, Z., and Ji, S. Non-local graph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10270–10276, 2022. doi: 10.1109/TPAMI.2021.3134200.

Mallat, S. Chapter 7 - wavelet bases. In Stéphane, M. (ed.), *A Wavelet Tour of Signal Processing (Third Edition)*, pp. 263–376. Academic Press, Boston, third edition edition, 2009. ISBN 978-0-12-374370-1. doi: https://doi.org/10.1016/B978-0-12-374370-1.00011-2. URL https://www.sciencedirect.com/science/article/pii/B9780123743701000112.

Mazars, M. Ewald methods for inverse power-law interactions in tridimensional and quasi-two-dimensional systems. *Journal of Physics A: Mathematical and Theoretical*, 43(42):425002, sep 2010. doi: 10.1088/1751-8113/43/42/425002. URL https://dx.doi.org/10.1088/1751-8113/43/42/425002.

Mialon, G., Chen, D., Selosse, M., and Mairal, J. Graphit: Encoding graph structure in transformers, 2021.

Nijboer, B. and De Wette, F. On the calculation of lattice sums. *Physica*, 23(1):309–321, 1957. ISSN 0031-8914. doi: https://doi.org/10.1016/S0031-8914(57)92124-9. URL https://www.sciencedirect.com/science/article/pii/S0031891457921249.

Perdew, J. P., Burke, K., and Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996. doi: 10.1103/PhysRevLett.77.3865. URL https://link.aps.org/doi/10.1103/PhysRevLett.77.3865.

Rampášek, L., Galkin, M., Dwivedi, V. P., Luu, A. T., Wolf, G., and Beaini, D. Recipe for a general, powerful, scalable graph transformer. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 14501–14515. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/5d4834a159f1547b267a05a4e2b7cf5e-Paper-Conference.pdf.

Reckien, W. List of functionals and coefficients for bj-damping, 2017. URL https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dft-d3/functionalsbj.

Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9377–9388. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/schutt21a.html.

Schütt, K., Kindermans, P.-J., Sauceda Felix, H. E., Chmiela, S., Tkatchenko, A., and Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *NeurIPS*, pp. 991–1001. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/6700-schnet-a-continuous-filter-convolutional-neural-network-for-modeling-quantum-interactions.pdf.

Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A., and Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24): 241722, March 2018. ISSN 0021-9606. doi: 10.1063/1.5019779. URL https://aip.scitation.org/doi/abs/10.1063/1.5019779.

Staacke, C. G., Heenen, H. H., Scheurer, C., Csányi, G., Reuter, K., and Margraf, J. T. On the role of long-range electrostatics in machine-learned interatomic potentials for complex battery materials. *ACS Applied Energy Materials*, 4(11):12562–12569, 2021. doi: 10.1021/acsaem.1c02363. URL https://doi.org/10.1021/acsaem.1c02363.

Stuke, A., Kunkel, C., Golze, D., Todorović, M., Margraf, J. T., Reuter, K., Rinke, P., and Oberhofer, H. Atomic structures and orbital energies of 61,489 crystal-forming organic molecules. *Scientific Data*, 7(1):58, 2020.

Tkatchenko, A. and Scheffler, M. Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.*, 102:073005, Feb 2009. doi: 10.1103/PhysRevLett.102.073005. URL https://link.aps.org/doi/10.1103/PhysRevLett.102.073005.

Unke, O. T., Chmiela, S., Gastegger, M., Schütt, K. T., Sauceda, H. E., and Müller, K.-R. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications*, 12(1):7273, December 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-27504-0. URL https://www.nature.com/articles/s41467-021-27504-0. Number: 1 Publisher: Nature Publishing Group.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is All you Need. In Guyon, I., Luxburg, U. v., Bengio, S., Wallach, H. M., Fergus, R., Vishwanathan, S. V. N., and Garnett, R. (eds.), *NeurIPS*, pp. 5998–6008, 2017. URL http://papers.nips.cc/paper/7181-attention-is-all-you-need.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. Graph Attention Networks. In *ICLR*, 2018. URL http://arxiv.org/abs/1710.10903. arXiv: 1710.10903.

Wells, B. A. and Chaffee, A. L. Ewald summation for molecular simulations. *Journal of Chemical Theory and Computation*, 11(8):3684–3695, 2015. doi: 10.1021/acs.jctc.5b00093. URL https://doi.org/10.1021/acs.jctc.5b00093. PMID: 26574452.

Xie, T. and Grossman, J. C. 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. URL https://link.aps.org/doi/10.1103/PhysRevLett.120.145301.

Xie, T., Fu, X., Ganea, O.-E., Barzilay, R., and Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation, 2021. URL https://arxiv.org/abs/2110.06197.

Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., and Liu, T.-Y. Do Transformers Really Perform Badly for Graph Representation? In *NeurIPS*, May 2021. URL https://openreview.net/forum?id=OeWooOxFwDa.

York, D. and Yang, W. The fast fourier poisson method for calculating ewald sums. *The Journal of Chemical Physics*, 101(4):3298–3300, 1994. doi: 10.1063/1.467576. URL https://doi.org/10.1063/1.467576.

Yu, H., Hong, L., Chen, S., Gong, X., and Xiang, H. Capturing long-range interaction with reciprocal space neural network, 2022.

Zhang, S., Liu, Y., and Xie, L. Molecular Mechanics-Driven Graph Neural Network with Multiplex Graph for Molecular Structures. In *Machine Learning for Molecules Workshop, NeurIPS*, November 2020. URL http://arxiv.org/abs/2011.07457. arXiv: 2011.07457.

A. Appendix

A.1. Architecture and Implementation Details

Fig. 8 shows detailed computing steps used to obtain the expression $f_{\text{upd}}^{\text{lr}}\left(M_i^{\text{lr}}\right)$ (as given in Eq. (12)), which we add to the GNN-specific short-range embedding update as part of a skip connection during every interaction step (Eq. (12)). The part highlighted in yellow computes the long-range sum M_i^{lr} that we discuss in Sec. 3. The index i denotes the target atom, while the source index j runs over all atoms in the structure (or the atoms of a single supercell in the case of periodic boundary conditions).

Real and imaginary parts. While we use complex exponential notation to simplify the conceptual part of this work, our implementation handles the real and imaginary parts of the structure factor individually. Eq. (11) gives $s_{\boldsymbol{k}} = \sum_{j\in\mathcal{S}} h_j \cos(\boldsymbol{k}^T\boldsymbol{x}_j)$, $\text{Im}(s_{\boldsymbol{k}}) = -\sum_{j\in\mathcal{S}} h_j \sin(\boldsymbol{k}^T\boldsymbol{x}_j)$. Plugging this into Eq. (10) and considering the real and imaginary parts of M_i^{lr}, we obtain

$$\begin{aligned}\text{Re}(M_i^{\text{lr}}) &= \sum_n \cos(\boldsymbol{k}_n^T\boldsymbol{x}_i)\sum_{j\in\mathcal{S}} h_j\cos(\boldsymbol{k}_n^T\boldsymbol{x}_j)\hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n)\\ &\quad+\sum_n \sin(\boldsymbol{k}_n^T\boldsymbol{x}_i)\sum_{j\in\mathcal{S}} h_j\sin(\boldsymbol{k}_n^T\boldsymbol{x}_j)\hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n),\\ \text{Im}(M_i^{\text{lr}}) &= 0.\end{aligned} \tag{13}$$

This prescription leads to the steps outlined in Fig. 8. In the aperiodic case in which we apply radial filtering with a direct cutoff, the index n enumerates all frequencies within the cutoff, $\{\boldsymbol{k}_n \in \Lambda', \|\boldsymbol{k}_n\| \leq c_k\}$, and we have $\hat{\Phi}(\boldsymbol{k}_n) \equiv \hat{\Phi}(\|\boldsymbol{k}_n\|)$. In the periodic case, we drop the radial symmetry of $\hat{\Phi}$ as well as the notion of a direct cutoff, and instead identify $n \equiv \lambda' \in I$, with the index set I enumerating reciprocal lattice sites close to the origin as defined in Sec. 3. To obtain the above result (particularly $\text{Im}(M_i^{\text{lr}}) = 0$), we implicitly use that the Fourier filter coefficients $\hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n)$ are real-valued, yet the real-valuedness of the *position-space* potential Φ^{lr} alone would *only* guarantee $\hat{\Phi}^{\text{lr}}(-\boldsymbol{k}_n) = \hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n)^*$, with $*$ denoting complex conjugation. The real-valuedness of all coefficients $\hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n)$ follows from the additional point symmetry constraint that we impose on the Fourier space filters. In the aperiodic filtering case, the point symmetry follows directly from radial symmetry, $\hat{\Phi}^{\text{lr}}(\|-\boldsymbol{k}_n\|) = \hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}_n\|)$. In the periodic case, we enforce the symmetry via $\hat{\Phi}_{\boldsymbol{\lambda}'} \equiv \hat{\Phi}_{-\boldsymbol{\lambda}'}$ for all $\boldsymbol{\lambda}' \in I$. The reason for this constraint is that an $\mathbb{R}^3$ lattice spanned by an arbitrary basis always has point symmetry around the origin – the model should therefore give rise to the same output on two structures that differ only by a point reflection of their respective PBC frames while their atom positions are equal.

Damping values. In the aperiodic case, all trigonometric functions in Eq. (13) are multiplied by a damping value $f_n(\boldsymbol{x})$, which depends on the position $\boldsymbol{x}$ as well as the frequency $\boldsymbol{k}_n$, and which can differ depending on the chosen discretization scheme for the structure factor (App. A.6). In our tested case of discretization by voxel averaging with voxel resolution Δ, we have

$$f_n(\boldsymbol{x}) = \prod_{c\,\in\,\{x,y,z\}} \text{sinc}\left(\frac{k_n^c x^c \Delta}{2}\right), \tag{14}$$

where the superscript c enumerates vector components.

Implementation of filtering schemes. Apart from damping values, the second, implementation-related difference between the periodic and aperiodic case comes from our choice of Fourier-space filtering scheme (Sec. 3). In the aperiodic case, the matrix product $\boldsymbol{W}_\uparrow\boldsymbol{W}_\downarrow$ gives a weight matrix $\boldsymbol{W}$, which we apply to the set of radial basis function values to get the filter values $\{\hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}_n\|),\ \boldsymbol{k}_n \in \Lambda',\ \boldsymbol{k}_n \leq c_k\}$ (as discussed in the aperiodic filtering paragraph of Sec. 3). In the periodic case, we do not apply the product $\boldsymbol{W}_\uparrow\boldsymbol{W}_\downarrow$ directly to a vector, but rather evaluate and store the resulting matrix given our current pair of weight matrices $(\boldsymbol{W}_\uparrow, \boldsymbol{W}_\downarrow)$. We identify the transpose $(\boldsymbol{W}_\uparrow\boldsymbol{W}_\downarrow)^T$ of this matrix with the filter coefficients $\{\hat{\Phi}_{\boldsymbol{\lambda}'} = \hat{\Phi}^{\text{lr}}(\boldsymbol{k}_{\boldsymbol{\lambda}'}) = \hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n),\ \boldsymbol{\lambda}' \in I\}$ (recall our identification of n with the multi-index $\boldsymbol{\lambda}'$). In both the periodic and aperiodic cases, we end up with a matrix of shape (N_{k}, F) containing F-dimensional stacks of filter values $\hat{\Phi}^{\text{lr}}(\boldsymbol{k}_n)$, that we multiply component-wise with the structure factor according to the prescription in Eq. (13). We use the downprojection $\boldsymbol{W}_\downarrow$ onto a bottleneck layer of dimension $N_\downarrow$ (with shared weight values among all interaction blocks l) to encourage generalization, as inspired by Gasteiger et al. (2021). Hence, $\boldsymbol{W}_\uparrow$ has shape $(F, N_\downarrow)$, while $\boldsymbol{W}_\downarrow$ has shape $(N_\downarrow, N_{\text{k}})$ (periodic case) or $(N_\downarrow, N_{\text{RBF}})$ (aperiodic case). We set $N_\downarrow = 16$ for GemNet-T and $N_\downarrow = 8$ otherwise.

Update function. As the update function $f_{\text{upd}}^{\text{lr}}$, we use among all models a dense layer followed by N_{hidden} residual layers, as specified by Gasteiger et al. (2021) in the atomic update block of the GemNet architecture. While the authors report using $N_{\text{hidden}} = 2$ residual blocks, we set $N_{\text{hidden}} = 3$ for all models except PaiNN, where we set $N_{\text{hidden}} = 0$, opting for less relative overhead compared the inexpensive short-range blocks of the PaiNN architecture.

A.2. Training and Hyperparameters

Baseline hyperparameters, unless explicitly listed here, are the same as the values used for the baseline models in (Chanussot et al., 2021). All differences from these original OCP hyperparameter settings either serve a fairer assessment of Ewald modifications by improving baseline performance (especially on the energy MAE metric primarily

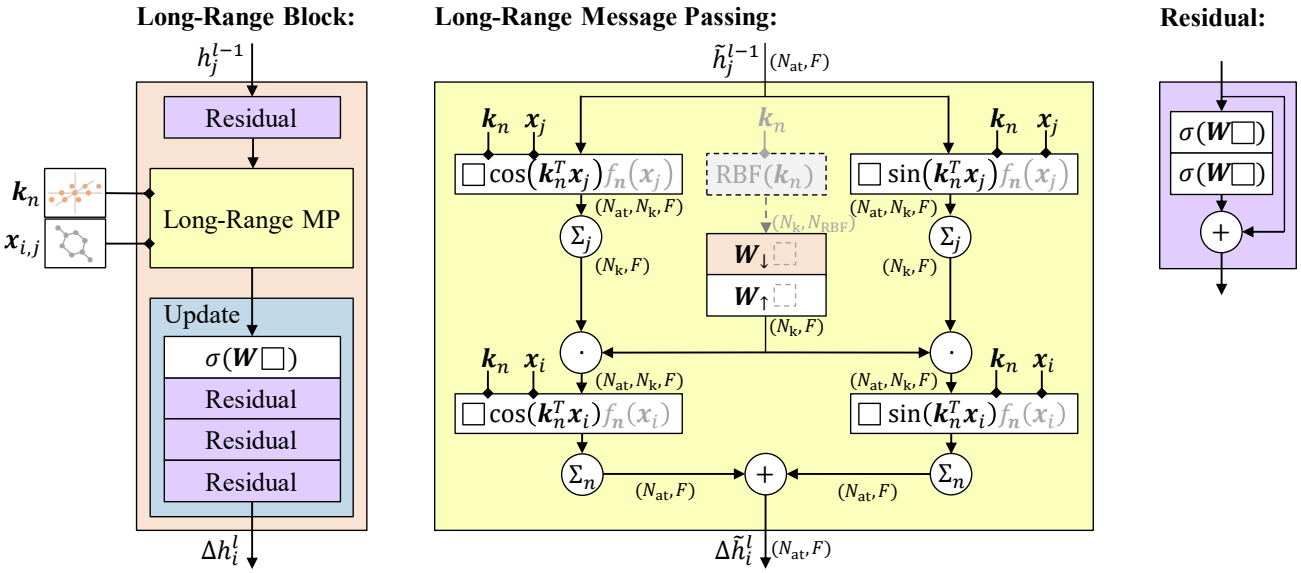

Figure 8. The architecture of the long-range part of Ewald message passing. The long-range block on the left corresponds to the expression $f_{\text{upd}}^{\text{lr}}\left(M_i^{\text{lr}}\right)$. □ denotes the input of a layer, encircled operations act component-wise. Triangular arrowheads define information flow through layers, diamond arrowheads denote parametric dependence. Annotated in brackets are the shapes of tensors handled throughout various parts of the computation. The steps marked by gray and dashed lines apply only to the aperiodic case.

addressed by Ewald message passing), or were introduced to achieve more practical training times for our experimental setup. Except for the new hyperparameters entering with Ewald message passing, we use exactly the same model settings for the baseline and Ewald versions. On OE62, we use embedding sizes of 512 for the PaiNN and SchNet models and 256 for the GemNet-T and DimeNet^{++} models. On OC20, we had to reduce the DimeNet^{++} filter size to 192 due to memory constraints. All baselines use distance cutoff values of 6 Å and neighbor thresholds N_{max} of 50. Our SchNet and PaiNN models use 4 interaction blocks, DimeNet^{++} and GemNet-T use 3 blocks.

Training hyperparameters on OE62. We find on OE62 that smaller batch sizes improve performance for all baselines except SchNet. Hence, we use OE62 batch sizes of 64 (SchNet) and 8 (else). All models have an initial learning rate of $1 \cdot 10^{-4}$ except for GemNet-T ($5 \cdot 10^{-4}$). As in the PaiNN reference (Schütt et al., 2021), we use the Adam optimizer with weight decay $\lambda = 0.01$, along with a plateau scheduler (patience 10 and decay factor 0.5). For DimeNet^{++} and GemNet-T, we use a milestone scheduler with a warmup period of 250000 steps (warmup factor 0.2) and decay factor 0.1 at 750000, 1125000 and 1500000 steps. For SchNet, we also use a milestone scheduler, with 30000 warmup steps (warmup factor 0.2) and decay factor 0.1 at 60000, 120000 and 180000 steps. Training on OE62 converges after 200 maximum epochs for PaiNN, 250 epochs for SchNet and DimeNet^{++}, and 300 epochs for GemNet-T.

Training hyperparameters on OC20. To reduce training times, we increase batch sizes on OC20. GemNet-T and DimeNet^{++} use batch size 16, PaiNN batch size 48 and SchNet batch size 64. We use a plateau scheduler with patience 3 and factor 0.8 for PaiNN, training for 25 maximum epochs. GemNet-T uses the same equal plateau scheduler. We train SchNet with a milestone scheduler (90000 warmup steps with warmup factor 0.2, milestones with factor 0.1 at 150000, 250000, 310000 steps) for 25 epochs, and DimeNet^{++} with 250000 warmup steps (warmup factor 0.2) and factor 0.1 milestones at 750000, 1125000 and 1500000 steps for 15 epochs. In the loss function, we use force multipliers of 33.3 for SchNet, 100 for PaiNN, 100 for GemNet-T, and 16.7 for DimeNet^{++}.

Loss function. Like Chanussot et al. (2021), we use a linear combination of energy and force mean absolute errors. Using the index $\mathcal{S}$ to run over the $|\mathcal{D}|$ atom structures in the set $\mathcal{D}$ (which becomes a mini-batch in practice), as well as i to run over the $|\mathcal{S}|$ individual atoms in a structure $\mathcal{S}$, our loss function has the form

$$\begin{aligned}&\frac{1}{|\mathcal{D}|}\sum_{\mathcal{S}\in\mathcal{D}}\left|E^{(\mathcal{S})}-E_{\text{t}}^{(\mathcal{S})}\right|\\&+\frac{\lambda}{|\mathcal{D}|}\sum_{\mathcal{S}\in\mathcal{D}}\frac{1}{|\mathcal{S}|}\left\|\boldsymbol{F}_i^{(\mathcal{S})}-\boldsymbol{F}_{i;\text{t}}^{(\mathcal{S})}\right\|_p,\end{aligned} \tag{15}$$

with the predicted/target truth energies $E^{(\mathcal{S})}/E_{\text{t}}^{(\mathcal{S})}$, the predicted/target forces $\boldsymbol{F}_i^{(\mathcal{S})}/\boldsymbol{F}_{i;\text{DFT}}^{(\mathcal{S})}$, the force multiplier

$\lambda \in \mathbb{R}_+$ defining the desired tradeoff between energy and force accuracy, and $\|\cdot\|_p$ denoting the Euclidean p-norm. According with model literature, we use $p = 1$ for SchNet and DimeNet^{++}, and $p = 2$ for PaiNN and GemNet-T. On OE62, where there are only energy targets, we set $\lambda = 0$.

A.3. Ewald Hyperparameters and Ablation Studies

Aperiodic case. Ewald message passing introduces two main new hyperparameters in the aperiodic case: the frequency cutoff c_k and the voxel resolution Δ. An additional source of hyperparameters is the parametrization of the frequency filters. In the case of our implementation, the number of Gaussian radial basis functions N_{RBF} needs to be fixed. In order to achieve pareto-efficient improvements through Ewald message passing, we adapt these hyperparameters to the comparative cost of each short-range baseline GNN. Our main experiment settings are listed in Tab. 1.

	$c_k[\text{Å}^{-1}]$	$\Delta[\text{Å}^{-1}]$	N_{RBF}
SchNet	0.4	0.2	48
PaiNN	0.6	0.2	128
DimeNet^{++}	0.8	0.2	128
GemNet-T	1.0	0.2	128

Table 1. Aperiodic Ewald hyperparameter settings.

Periodic case. In the periodic case, the main hyperparameters of Ewald message passing are the numbers N_x, N_y, N_z which define the frequencies included in the long-range sum. In principle, they can all be varied independently, but in our work we tie them into one independent effective hyperparameter by the prodedure explained below.

Recall that we keep the amount of included frequencies fixed across structures in our periodic filtering setting, which leads to better performance empirically. This rules out simply including all reciprocal lattice frequencies within a fixed cutoff volume, as their total number would vary across input structures with differing reciprocal lattice supercells. Still, we impose that the range of included frequencies does not extend out to significantly further distance in any of the three reciprocal lattice directions, just as it would be the case with a spherical cutoff volume. In our setting, this is not possible for an arbitrary input structure, as the set of included reciprocal lattice indices I (defined in Sec. 3, and fixed via our hyperparameters) is mapped onto different points of Fourier space for any structure via $\boldsymbol{k}_{\boldsymbol{\lambda}'} = \lambda_1' \boldsymbol{w}_1 + \lambda_2' \boldsymbol{w}_2 + \lambda_3' \boldsymbol{w}_3$. However, we can make sure that an even Fourier space distance is covered by included frequencies along all three lattice directions for the *average* structure. For instance, if the reciprocal lattice points are twice as dense on average in one direction as compared to another, we double N_i for this direction to have the same average Fourier space distance covered by frequencies along both axes. This single distance effectively acts like an "average cutoff", while the number of included frequencies stays fixed for all structures. On OC20, we observe that the average reciprocal space unit cell is roughly 3 times narrower along the z direction (the surface normal) than it is along the x and y directions. Applying this hyperparameter tying constraint and accounting for different baseline model runtimes leads to our main settings listed in Tab. 2.

	N_x	N_y	N_z
SchNet	1	1	3
PaiNN	1	1	3
DimeNet^{++}	2	2	5
GemNet-T	2	2	5

Table 2. Periodic Ewald hyperparameter settings.

Additional hyperparameters. Additional hyperparameters introduced by Ewald message passing in both the periodic and aperiodic case are the bottleneck dimension $N_{\downarrow}$ and the amount of residual layers N_{hidden} used as part of the update function (cf. App. A.1).

Ablation studies: aperiodic case. We test OE62 EMAEs of the DimeNet^{++}-Ewald model with alternative settings for c_k and Δ to probe the robustness of our approach in hyperparameter space. We obtain EMAEs of $46.9\,\text{meV}$ for ($c_k = 0.6\,\text{Å}^{-1}$, $\Delta = 0.2\,\text{Å}^{-1}$), $47.2\,\text{meV}$ for ($c_k = 0.4\,\text{Å}^{-1}$, $\Delta = 0.1\,\text{Å}^{-1}$), and $47.3\,\text{meV}$ for ($c_k = 1.6\,\text{Å}^{-1}$, $\Delta = 0.4\,\text{Å}^{-1}$), all of which is a consistent improvement over the $51.2\,\text{meV}$ baseline MAE.

We also test similar variations of the SchNet model, obtaining EMAEs of $83.0\,\text{meV}$ for ($c_k = 0.4\,\text{Å}^{-1}$, $\Delta = 0.4\,\text{Å}^{-1}$), $85.9\,\text{meV}$ for ($c_k = 0.2\,\text{Å}^{-1}$, $\Delta = 0.2\,\text{Å}^{-1}$), and $82.8\,\text{meV}$ for ($c_k = 0.8\,\text{Å}^{-1}$, $\Delta = 0.4\,\text{Å}^{-1}$) which are again very close to our main setting results and strongly improve the $133.5\,\text{meV}$ baseline MAE.

Ablation studies: periodic case. We test OC20 EMAEs (averaged across all four test splits) of the SchNet-Ewald model with different settings for $(N_x,\, N_y,\, N_z)$, probing whether our hyperparameter tying constraint (see above) is a sensitive choice. We obtain EMAEs of $836\,\text{meV}$ $(1, 1, 1)$ and $828\,\text{meV}$ $(1, 1, 2)$. Again, we find that our improvements over the $895\,\text{meV}$ baseline MAE are robust to different settings.

We do not use radial filters in the PBC setting as we find them to perform worse empirically. We train our GemNet-T-Ewald in a time-reduced setting of 30 maximum message-passing neighbors and for 12 epochs on OC20-2M. We record EMAEs on the four `OC20-val` splits. Our standard filtering approach with $(N_x,\, N_y,\, N_z) = (2, 2, 5)$ has 137 filter weights (due to mirror symmetrization as men-

Setting	Baseline	Standard	Radial	Cuboid
ID	305	263	264	278
OOD	455	416	447	436

Table 3. EMAEs [meV] of filtering variants on OC20.

Setting	SchNet	PaiNN	DimeNet^{++}	GemNet-T
Ewald	12.2M	15.7M	4.8M	16.1M
Emb.	14.4M	15.7M	5.4M	15.8M

Table 4. Parameter counts (in million parameters) for the Ewald and increased-embedding variants on OE62.

tioned in Sec. 3). We compare this to an Ewald variant using radial filtering with 128 Gaussian radial basis functions. Furthermore, we test another filtering approach which uses 3D Gaussian basis functions $\mathcal{G}(\Sigma, \mu_i)$ with covariance matrices $\Sigma = \mathrm{diag}(\Delta_x^2, \Delta_y^2, \Delta_z^2) \in \mathbb{R}^{3x3}$ centered on the points $\mu_i \in \Lambda'_c$ of a cuboid lattice $\Lambda'_c = \{\lambda'_x \Delta_x (1,0,0)^T + \lambda'_y \Delta_y (0,1,0)^T + \lambda'_z \Delta_z (0,0,1)^T \mid \boldsymbol{\lambda}' \in I\}$, with $I = \bigtimes_{j=x,y,z} \{-N_j, \ldots, N_j\} \subset \mathbb{Z}^3$. We test $(N_x, N_y, N_z) = (3,3,6)$ and $(\Delta_x, \Delta_y, \Delta_z) = (0.333, 0.333, 0.167)\text{Å}^{-1}$. As seen above, all variants improve the baseline, but the radial and cuboid filtering variants generalize worse on the OOD-both split, in which both the catalyst and adsorbate have out-of-distribution composition (Tab. 3).

A.4. Hyperparameters for Comparison Studies

Increased cutoff setting. For the spatially infinite structures of OC20, we assume the amount of message passing neighbors for any given cutoff roughly satisfies $N \sim c_x^3$ and scale it up accordingly. Given starting values of $c_x = 6\,\text{Å}$ and $N_{\text{max}} = 50$ for all baselines, we find the modified OC20 settings $(6.25\,\text{Å}, 55)$ for GemNet-T and DimeNet^{++} $(7\,\text{Å}, 80)$ for PaiNN and $(8.5\,\text{Å}, 140)$ for SchNet which roughly match the respective Ewald model variants in cost. On OE62, we double the neighbor threshold from 50 to 100 (exceeding the total size of most structures), use an increased 9 Å cutoff for GemNet-T and a 12 Å cutoff for all other baselines. On OE62, we double the neighbor threshold from 50 to 100 (exceeding the total size of most structures), use an increased 8 Å cutoff for GemNet-T and 12 Å for all other baselines.

Increased embedding size setting. In our OE62 increased embedding size study, we increase sizes to 1536 for SchNet, 576 for PaiNN, 384 for DimeNet^{++} and 320 for GemNet-T, which approximately matches the free parameter counts of the Ewald variants (Tab. 4).

Pairwise long-range block setting. We replace the Ewald block ($f_{\text{upd}}^{\text{lr}}\left(M_i^{\text{lr}}\right)$ in Eq. (12)) by a standard SchNet block with 200 Gaussians and 256 filters. On OE62, cost saturates at 12 Å and 100 maximum neighbors, where the molecular graph approaches full connectivity. On OC20, long-ranged SchNet blocks cannot compete with the Ewald cost: the Ewald overhead is already matched at 6.25 Å for PaiNN, 6.75 Å for SchNet and DimeNet++, and 7.25 Å for GemNet-T, where we scale up the neighbor threshold as in the increased cutoff setting. Increasing distance cutoffs on OC20 is more costly due to the higher atomic density.

A.5. Relationship between $\hat{M}_{\boldsymbol{k}}^{\mathbf{lr}}$ and $\hat{\Phi}^{\mathbf{lr}}(\|\cdot\|)$

To derive the desired identity as done in standard Ewald summation, we start with the Fourier coefficients $\hat{M}_{\boldsymbol{k}}^{\text{lr}}$,

$$\hat{M}_{\boldsymbol{k}}^{\text{lr}} = \frac{1}{\Omega} \int_{\Omega(\mathcal{S})} M^{\text{lr}}(\boldsymbol{x}) \exp(-i\boldsymbol{k}^T \boldsymbol{x}) d^3\boldsymbol{x}, \tag{16}$$

where $\Omega(\mathcal{S}) \subset \mathbb{R}^3$ denotes a single supercell parallelepiped. We now plug in Eq. (8) for $M^{\text{lr}}(\boldsymbol{x})$, resulting in

$$\hat{M}_{\boldsymbol{k}}^{\text{lr}} = \frac{1}{\Omega} \int_{\Omega(\mathcal{S})} \sum_{\boldsymbol{t}\in\Lambda} \sum_{j\in\mathcal{S}} h_j \Phi^{\text{lr}}\left(\|\boldsymbol{x} - \boldsymbol{x}_j - \boldsymbol{t}\|\right) \cdot \exp(-i\boldsymbol{k}^T \boldsymbol{x}) d^3\boldsymbol{x}. \tag{17}$$

Assuming the lattice sum over Λ is absolutely convergent, we can exchange it with the supercell integral such that

$$\begin{aligned} &\int_{\Omega(\mathcal{S})} \sum_{\boldsymbol{t}\in\Lambda} \Phi^{\text{lr}}\left(\|\boldsymbol{y}_j - \boldsymbol{t}\|\right) \exp(-i\boldsymbol{k}^T \boldsymbol{y}_j) d^3\boldsymbol{y}_j \\ &= \sum_{\boldsymbol{t}\in\Lambda} \int_{\Omega(\mathcal{S})} \Phi^{\text{lr}}\left(\|\boldsymbol{y}_j - \boldsymbol{t}\|\right) \exp(-i\boldsymbol{k}^T \boldsymbol{y}_j) d^3\boldsymbol{y}_j \\ &= \int_{\mathbb{R}^3} \Phi^{\text{lr}}\left(\|\boldsymbol{y}_j\|\right) \exp(-i\boldsymbol{k}^T \boldsymbol{y}_j) d^3\boldsymbol{y}_j \end{aligned} \tag{18}$$

holds (since the supercells form a tessellation of $\mathbb{R}^3$), where we have set $\boldsymbol{y}_j := \boldsymbol{x} - \boldsymbol{x}_j$. We now identify the obtained integral over $\mathbb{R}^3$ as the Fourier transform $\hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|)$:

$$\begin{aligned} \hat{M}_{\boldsymbol{k}}^{\text{lr}} &= \frac{1}{\Omega} \sum_{j\in\mathcal{S}} h_j \left[\int_{\mathbb{R}^3} \Phi^{\text{lr}}(\|\boldsymbol{y}_j\|) \exp(-i\boldsymbol{k}^T \boldsymbol{y}_j) d^3\boldsymbol{y}_j\right] \\ &\cdot \exp(-i\boldsymbol{k}^T \boldsymbol{x}_j) = \frac{1}{\Omega} \sum_{j\in\mathcal{S}} h_j \exp(-i\boldsymbol{k}^T \boldsymbol{x}_j) \hat{\Phi}^{\text{lr}}(\|\boldsymbol{k}\|). \end{aligned} \tag{19}$$

A.6. Discretization of the Structure Factor

There is no unique approach to discretization and we believe our application leaves future room for more sophisticated schemes, e.g., approaches based on wavelet theory (Mallat, 2009). To aid such efforts, we first present a quite general framework for discretizing the structure factor. Afterwards, we outline our concrete implementation and show how it arises as a special case. Assume a finite set of N_{k} square-integrable, complex-valued basis functions $\{f_i \in L^2(D) \mid 1 \leq i \leq N_{\text{k}}\}$ on a bounded domain D

containing the cutoff domain $\mathcal{B}_c = \{\boldsymbol{k} \in \mathbb{R}^3 \mid k \leq c_k\}$, i.e. $\mathcal{B}_c \subset D \subset \mathbb{R}^3$. Further assume orthonormality $\langle f_j, f_k \rangle_{L^2} = \int_D f_j(\boldsymbol{k})^* f_k(\boldsymbol{k}) d^3\boldsymbol{k} = \delta_{jk}$, where * denotes complex conjugation. The structure factor $s(\boldsymbol{k})$ can be projected onto this basis set, resulting in a discrete approximate representation (Eq. (20)) in terms of basis coefficients $\{s_i \mid 1 \leq i \leq N_\mathrm{k}\}$:

$$s(\boldsymbol{k}) \approx \sum_{i=1}^{N_\mathrm{k}} \langle f_i, s \rangle f_i(\boldsymbol{k}) =: \sum_{i=1}^{N_\mathrm{k}} s_i f_i(\boldsymbol{k}) \tag{20}$$

Integrating these scalar products numerically would require knowing $s(\boldsymbol{k})$ at high resolution. Instead, Eq. (21) takes advantage of the appearing Fourier transform:

$$\begin{aligned} s_i = \langle f_i, s \rangle &= \sum_{b \in \mathcal{S}} h_b \int_D f_i(\boldsymbol{k})^* \exp(-i\boldsymbol{k}\boldsymbol{x}_b) d^3\boldsymbol{k} \\ &= \sum_{b \in \mathcal{S}} h_b \widehat{[f_i^*]}(\boldsymbol{x}_b) = \sum_{b \in \mathcal{S}} h_b \hat{f}_i^*(-\boldsymbol{x}_b). \end{aligned} \tag{21}$$

Note the analogy to $s_{\boldsymbol{k}}$, with $\hat{f}_i^*(-\boldsymbol{x}_b)$ replacing $\exp(-i\boldsymbol{k}\boldsymbol{x}_b)$. We illustrate this with two examples.

(a) Auxiliary supercell. A standard literature approach (Freysoldt et al., 2022) is to approximate an aperiodic structure as part of a supercell that is assumed finite, but sufficiently large that spurious interactions can be ignored. This reintroduces a reciprocal lattice Λ' on which discrete $s_{\boldsymbol{k}_i} := s(\boldsymbol{k}_i)$ can be evaluated. Due to the defining property of the delta distribution, $\langle \delta(\cdot - \boldsymbol{k}_i), s \rangle = s(\boldsymbol{k}_i)$, this is a special (i.e., limiting) case of our formalism featuring a distributional 'basis' and Fourier transform,

$$\begin{aligned} &\{f_i^{\mathrm{SC}}(\boldsymbol{k}) = \delta(\boldsymbol{k} - \boldsymbol{k}_i) \mid \boldsymbol{k}_i \in \Lambda' \cap \mathcal{B}_c\}, \\ &\hat{f}_i^{\mathrm{SC}*}(-\boldsymbol{x}_b) = \exp(-i\boldsymbol{k}_i\boldsymbol{x}_b), \end{aligned} \tag{22}$$

As seen above from Eq. (22), the definition of coefficient s_i in Eq. (21) indeed reduces to the PBC definition of $s_{\boldsymbol{k}}$ in this special case.

(b) Voxel basis. In our implementation of aperiodic Ewald message passing, we discretize $s(\boldsymbol{k})$ by taking voxel averages s_i on the cutoff-covering portion $\Lambda'_\Delta \cap \mathcal{B}_c$ of a 3D voxel grid $\Lambda'_\Delta = \{\lambda'_1 \Delta(1,0,0)^T + \lambda'_2 \Delta(0,1,0)^T + \lambda'_3 \Delta(0,0,1)^T \mid \lambda' \in \mathbb{Z}^3\}$ with voxel sidelength Δ. The elements $\boldsymbol{k}_i \in \Lambda'_\Delta \cap \mathcal{B}_c$ are the voxel center locations. In our general framework, this corresponds to a basis set

$$\begin{aligned} &\{f_i^{\mathrm{V}} = \Delta^{-3} \mathcal{X}_{\boldsymbol{k}_i} \mid \boldsymbol{k}_i \in \Lambda'_\Delta \cap \mathcal{B}_c\}, \\ &\hat{f}_i^{\mathrm{V}*}(-\boldsymbol{x}_b) = \exp(-i\boldsymbol{k}_i\boldsymbol{x}_b) \prod_{j \in \{xyz\}} \operatorname{sinc}\left(\frac{k_i^j x_b^j \Delta}{2}\right), \end{aligned} \tag{23}$$

where $\mathcal{X}_{\boldsymbol{k}_i}$ takes on a value of 1 the $\boldsymbol{k}_i$-centered cube $\prod_{j \in \{x,y,z\}} \left[k_i^j - \frac{\Delta}{2}, k_i^j + \frac{\Delta}{2}\right]$ and 0 otherwise. The hyperparameter Δ controls the discretization resolution. Taking the scalar product,

$$\langle f_i^{\mathrm{V}}, s \rangle = \Delta^{-3} \int_{k_i^x - \frac{\Delta}{2}}^{k_i^x + \frac{\Delta}{2}} \int_{k_i^y - \frac{\Delta}{2}}^{k_i^y + \frac{\Delta}{2}} \int_{k_i^z - \frac{\Delta}{2}}^{k_i^z + \frac{\Delta}{2}} s(\boldsymbol{k}) d^3\boldsymbol{k}, \tag{24}$$

computes the average value of $s(\boldsymbol{k})$ inside the voxel cube around $\boldsymbol{k}_i$. Using the Fourier transform result from Eq. (23), we see that s_i only differs from the PBC case $s_{\boldsymbol{k}}$ by an additional sinc damping function. Taking $\Delta \to 0$ recovers the continuum limit: consistently, $\operatorname{sinc}\left(\frac{x_b^j \Delta}{2}\right)$ converges pointwise to 1 in this limit.

So far, we have outlined how to represent the structure factor by discrete basis coefficients s_i. To compute long-range messages in our voxel implementation, we proceed as follows. The continuous filter $\hat{\Phi}(k)$ is assigned the same type of voxel representation $\sum_{i=1}^{N_\mathrm{k}} \hat{\Phi}_i f_i^{\mathrm{V}}$ in terms of coefficients $\hat{\Phi}_i = \langle f_i^{\mathrm{V}}, \hat{\Phi} \rangle$. These coefficients are obtained by reading out the learned radial basis combinations $\hat{\Phi}(k) = \boldsymbol{W}\hat{\Psi}(k)$ on the discrete grid of voxel centerpoints $\Lambda'_\Delta \cap \mathcal{B}_c$. Now, the complete expression for the message sum assumes the approximate form

$$\begin{aligned} M_a^{\mathrm{lr}} &= \int_D \exp(i\boldsymbol{k}\boldsymbol{x}_a) \left[\sum_{i=1}^{N_\mathrm{k}} s_i \hat{\Phi}_i f_i^{\mathrm{V}}(\boldsymbol{k})\right] d^3\boldsymbol{k} \\ &= \sum_{i=1}^{N_\mathrm{k}} \hat{f}_i^{\mathrm{V}}(-\boldsymbol{x}_a) \left[\sum_b h_b \hat{f}_i^{\mathrm{V}*}(-\boldsymbol{x}_b)\right] \hat{\Phi}_i, \end{aligned} \tag{25}$$

where in the last step we substituted in the expression for s_i from Eq. (21). Note the analogy to the PBC message passing formula Eq. (10) with $\exp(i\boldsymbol{k}\boldsymbol{x}_a)$ and $\exp(-i\boldsymbol{k}\boldsymbol{x}_b)$ replaced by $f_i^{\mathrm{V}}(-\boldsymbol{x}_a)$ and $f_i^{\mathrm{V}*}(-\boldsymbol{x}_b)$, respectively. As these differ only by an additional sinc damping factor for the voxel case, our periodic and aperiodic implementations of Ewald message passing use just marginally different code.

A.7. Impact on Force Predictions

Ewald message passing is primarily a correction scheme for energy rather than force targets. The force on atom i is the negative energy gradient $\boldsymbol{F}(\boldsymbol{x}_i) = -\nabla_i E(\boldsymbol{x}_1, \ldots \boldsymbol{x}_N)$. Due to the wavelength truncation used in Eq. (10), Ewald message passing can only contribute terms to the energy prediction which vary more slowly in space than the scale of its inverse frequency cutoff. To see this quantitatively, consider the Fourier series ansatz for the function $V_i(\boldsymbol{x}) := E(\boldsymbol{x}_1, \ldots \boldsymbol{x}_{i-1}, \boldsymbol{x}, \boldsymbol{x}_{i+1}, \boldsymbol{x}_N)|_{\boldsymbol{x}_1, \boldsymbol{x}_{i-1}, \ldots, \boldsymbol{x}_{i+1}, \boldsymbol{x}_N}$, reading $V_i(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \Lambda'} \hat{V}_{\boldsymbol{k}} \exp(i\boldsymbol{k}^T\boldsymbol{x})$. In these terms,

$$\boldsymbol{F}(\boldsymbol{x}_i) = -\nabla V(\boldsymbol{x}_i) = -\sum_{\boldsymbol{k} \in \Lambda'} i\boldsymbol{k}\hat{V}_{\boldsymbol{k}} \exp(i\boldsymbol{k}^T\boldsymbol{x}_i). \tag{26}$$

Contributions at low frequencies $\|\boldsymbol{k}\| \leq c_k$ are therefore being suppressed compared to the gradient contributions of higher frequencies. Hence, we hypothesize that the *short*-range message sums of the model take over a major part of the force prediction.

A.8. Runtime Measurements

We run all models on `Nvidia A100` GPUs and evaluate the runtimes of all model configurations in one session and on the same machine. After 50 warmup batches, we average the runtime per structure over 500 batches and repeat this measurement three times. Afterwards, we take the minimum of these three measurements.

A.9. Further Long-Range Binning Results

As a supplement to our analysis in the main body of this work, we provide additional binning results similar to Fig. 7. They give additional insights about the correlation of various structural features with Ewald improvements.

Further OE62 analyses. We group structures by two structural size metrics: the maximum distance between any pair of atoms, and the standard deviation σ along the principal axis for the atomic point cloud. Furthermore, we bin according to electrostatic energy Eq. (6), obtained using DFT-Hirshfeld atomic partial charges from the OE62 data.

Further OC20 analyses. On OC20, no obvious size metric exists due to the periodic boundary conditions. However, we can still bin according to electrostatic energy as for OE62 on the `IS2RE-val-id` split of OC20, for which Bader partial charges are provided on all structures.

Discussion. Our additional analyses inform the following conclusions and usage recommendations:

- Ewald MP appears to address dispersion effects. We hope our method might also learn such effects for reference targets going beyond DFT-D3 in the cost hierarchy, such as energies obtained from a nonlocal DFT functional.
- We observe no increased relative improvement for higher electrostatic energy, only a constant correction across all bins. Apparently, standard GNNs can already learn electrostatics fairly well on our data. It could be valuable to repeat this study on upcoming benchmarks in fields like interface or colloid chemistry.
- We expect the usefulness of Ewald MP to grow with structural size due to its favorable scaling. However, our binning plots suggest that structural size is no direct proxy for expected model error or Ewald improvement.

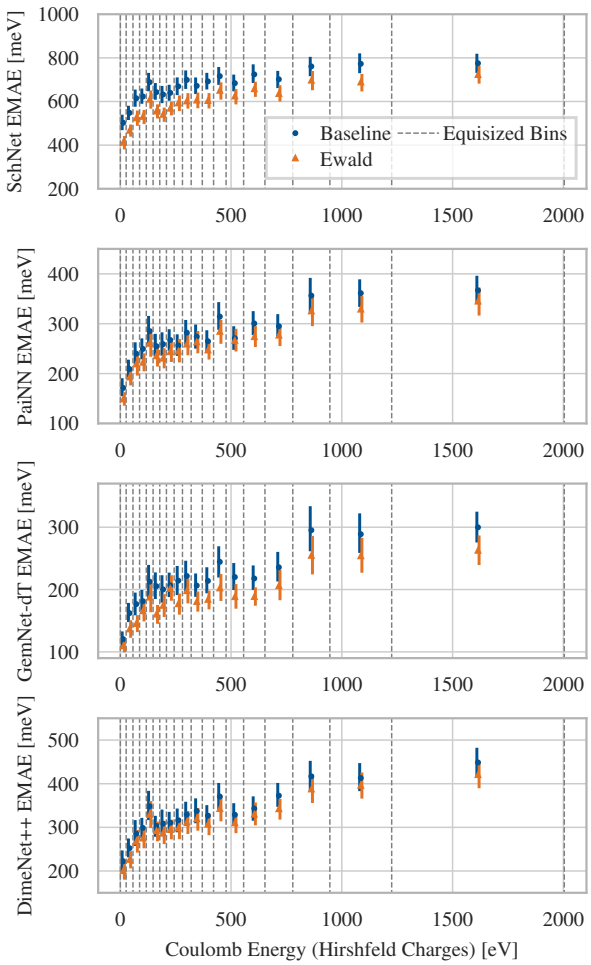

Figure 9. Energy MAE on `OC20-IS2RE-val-id` structures binned according to electrostatic energy. Ewald MP has *no* outsize impact on structures with a high electrostatic energy contribution.

A.10. OE62 Dataset Preprocessing

As the raw OE62 DFT energies have a large average offset from zero, we compute least-squares-fitted targets for all structures $\mathcal{S}$ of the form $E_{\mathrm{t}}^{(\mathcal{S})} = E_{\mathrm{raw}}^{(\mathcal{S})} - \sum_Z C_Z N_Z^{(\mathcal{S})} - C_0$, where $E_{\mathrm{t}}^{(\mathcal{S})}/E_{\mathrm{raw}}^{(\mathcal{S})}$ are the preprocessed target / raw DFT energies, $Z \in \mathbb{N}$ enumerates atom types with type-specific regression coefficients $C_Z \in \mathbb{R}$ and bias coefficient C_0, and $N_Z^{(\mathcal{S})} \in \mathbb{N}_0$ is the number of atoms belonging to type Z in the structure $\mathcal{S}$. The model only learns the residuals of this fit, which have close-to-zero mean by construction. We partition the data in ca. 50000 structures for `OE62-train`, and ca. 6000 structures for each of `OE62-val` and `OE62-test`.

A.11. Numerical Energy, Force and Runtime Results

Tabs. 5 and 6 provide the numerical OC20 and OE62 results on which we base our visualizations in Sec. 5.

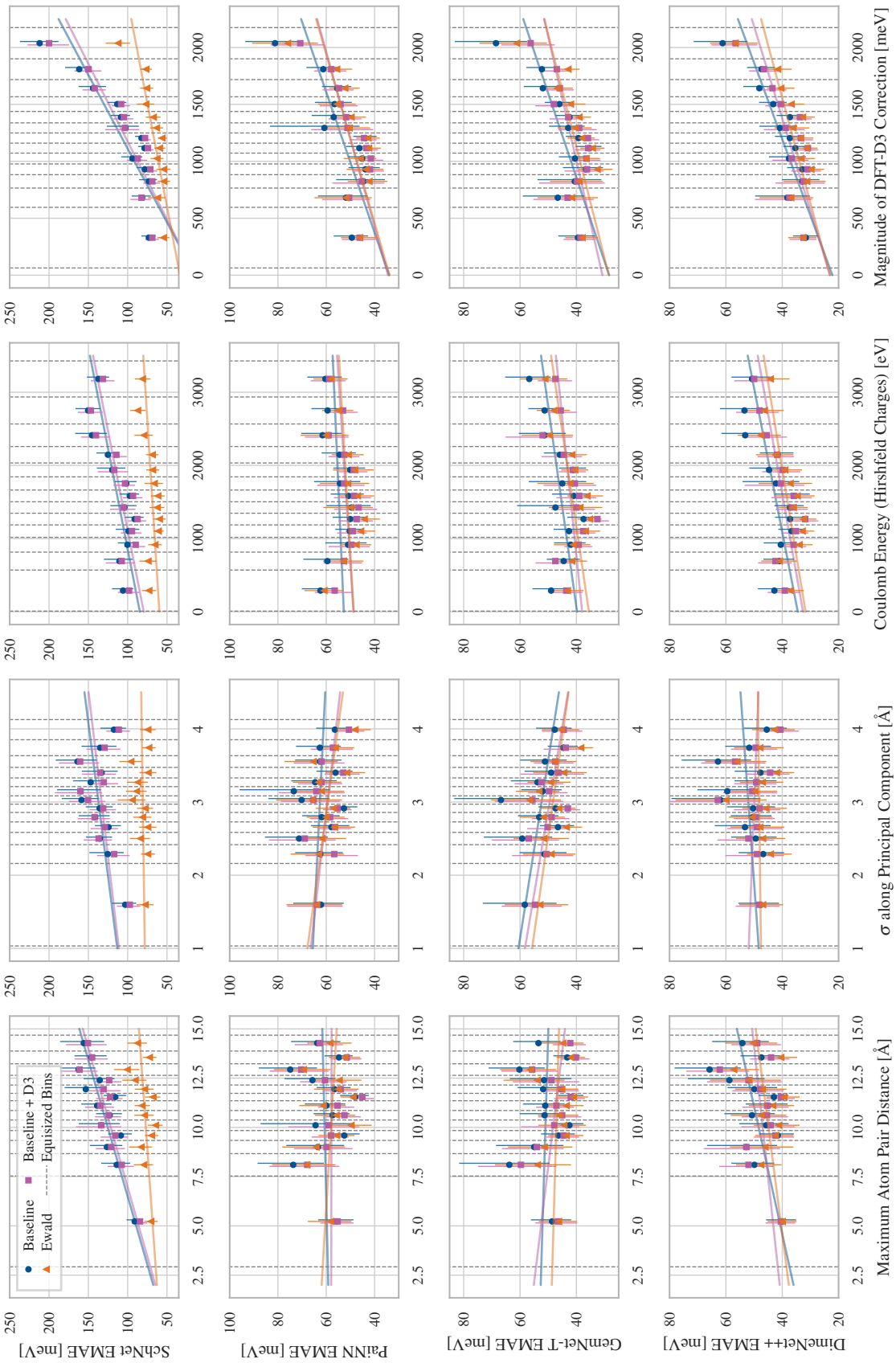

Figure 10. Binning of the energy EMAE on `OE62-val` by two structural size metrics, electrostatic and dispersion energy. Only dispersion is a good predictor for Ewald improvement.

Table 5. Energy MAEs and runtimes per input structure on OE62.

		OE62-val		OE62-test		Forward Pass		Forward & Backward Pass	
Model	Variant	MAE meV ↓	Rel. % ↑	MAE meV ↓	Rel. % ↑	Runtime ms/struct. ↓	Rel. % ↓	Runtime ms/struct. ↓	Rel. % ↓
SchNet	Baseline	133.5	-	131.3	-	0.13	-	0.28	-
	Embeddings	144.7	-8.4	136.7	-4.1	0.14	15.2	0.33	17.8
	Cutoff	257.4	-92.8	254.8	-94.1	0.14	13.6	0.31	11.6
	SchNet-LR	86.6	35.1	89.2	32.1	0.32	156.0	0.75	171.7
	Ewald	**79.2**	**40.7**	**81.1**	**38.2**	0.70	461.6	1.03	271.4
PaiNN	Baseline	61.4	-	63	-	1.52	-	3.16	-
	Embeddings	63.5	-3.4	63.1	-0.2	1.54	1.4	3.28	3.8
	Cutoff	65.1	-6.0	64.4	-2.2	1.84	20.9	3.91	23.6
	SchNet-LR	58.3	5.1	58.2	7.7	1.84	20.7	4.21	33.1
	Ewald	**57.9**	**5.7**	**59.7**	**5.2**	2.29	50.5	4.57	44.4
DimeNet^{++}	Baseline	51.2	-	53.8	-	1.99	-	4.26	-
	Embeddings	50.4	1.6	53.4	0.7	2.25	12.9	4.93	15.8
	Cutoff	48.3	5.7	**48.1**	**10.6**	2.68	34.7	6.10	43.4
	SchNet-LR	51.4	-0.5	54.4	-1.1	2.37	19.0	4.73	11.2
	Ewald	**46.5**	**9.2**	**48.1**	**10.6**	2.70	35.5	5.93	39.5
GemNet-T	Baseline	51.5	-	53.1	-	3.07	-	6.96	-
	Embeddings	52.7	-2.3	53.9	-1.5	3.11	1.5	6.98	0.4
	Cutoff	47.8	7.2	47.7	10.2	4.02	31.2	8.88	27.7
	SchNet-LR	51.2	0.6	52.8	0.5	3.32	8.3	7.73	11.1
	Ewald	**47.4**	**8.0**	**47.5**	**10.5**	4.05	32.0	8.86	27.4

Table 6. Energy and force mean absolute errors, as well as runtimes per input structure, averaged across all four splits of OC20-test.

		Energy		Forces		Forward Pass		Forward & Backward Pass	
Model	Variant	MAE meV ↓	Rel. % ↑	MAE meV Å^{-1} ↓	Rel. % ↑	Runtime ms/struct. ↓	Rel. % ↓	Runtime ms/struct. ↓	Rel. % ↓
SchNet	Baseline	895	-	61.1	-	0.85	-	1.83	-
	Cutoff	869	3.0	60.3	1.3	1.57	84.2	3.92	114.2
	SchNet-LR	984	-9.9	65.3	-6.8	1.79	110.6	5.24	186.0
	Ewald	**830**	**7.3**	**56.7**	**7.2**	1.79	110.7	4.68	155.3
PaiNN	Baseline	448	-	42.0	-	1.66	-	3.44	-
	Cutoff	413	7.9	40.9	2.6	2.54	53.1	5.06	46.9
	SchNet-LR	468	-4.3	43.3	-3.2	2.06	24.2	5.24	44.3
	Ewald	**393**	**12.3**	**40.5**	**3.5**	2.08	25.4	4.24	23.1
DimeNet^{++}	Baseline	496	-	47.1	-	5.58	-	15.68	-
	Cutoff	487	1.8	46.4	1.4	6.28	12.4	17.23	9.9
	SchNet-LR	504	-1.6	48.3	-2.6	6.26	12.1	16.14	2.9
	Ewald	**445**	**10.4**	**45.5**	**3.3**	6.24	11.7	17.00	8.4
GemNet-dT	Baseline	346	-	29.9	-	6.37	-	14.77	-
	Cutoff	352	-1.7	30.2	-1.0	7.23	13.5	16.40	11.0
	SchNet-LR	351	-1.5	29.8	0.3	7.19	12.1	16.35	10.7
	Ewald	**307**	**11.3**	**29.3**	**1.8**	7.09	11.2	16.23	9.9