NEMORA: NEURAL EQUIVARIANT MULTIPOLE OPERATORS FOR LONG-RANGE ATOMISTIC LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Equivariant graph neural networks have emerged as foundational architectures for machine-learned interatomic potentials, approaching quantum-chemical accuracy at a fraction of the computational cost. These models describe local atomic environments accurately, but finite spatial cutoffs truncate long-range information flow, and stacking message-passing layers can lead to over-smoothing and over-squashing. Existing long-range extensions either prescribe a fixed analytical propagation kernel, restrict long-range communication to scalars or degree-preserving channels, or incur super-linear computational cost. Combining learnable long-range equivariant transport with multiscale many-body expressivity and efficient scaling remains a central challenge. We introduce Neural Equivariant Multipole Operators (NEMORA), a neural equivariant extension of the Fast Multipole Method (FMM) for learning long-range tensorial representations. NEMORA generalizes the FMM's analytical multipole expansion and translation operators to learned equivariant counterparts on an adaptive spatial hierarchy. Its operators couple angular degrees and form many-body interactions across length scales, retaining the FMM's hierarchical organization and analytical radial factors as physical inductive biases while learning data-dependent long-range couplings. NEMORA is linear in time and memory, allowing it to treat larger systems than other long-range methods, and it augments both symmetry-constrained and unconstrained short-range backbones. On non-local benchmarks, it reduces energy and force errors relative to the short-range backbones by up to three orders of magnitude, and it is competitive with or better than existing long-range extensions in accuracy.

1 INTRODUCTION

Modeling atomic interactions underpins central tasks in computational science, from finding low-energy conformations to computing thermodynamic and kinetic properties by molecular simulation. Classical force fields enable large simulations at modest cost, but their prescribed functional forms limit the chemistry they can represent. Machine-learned interatomic potentials (MLIPs) learn potential energy surfaces from quantum-chemical data, approaching this same accuracy at a fraction of the cost (Behler & Parrinello, 2007; Bartók et al., 2010; Smith et al., 2017; Unke et al., 2021b). Equivariant architectures are especially accurate and data-efficient because they build the symmetries of atomic environments into their representations (Schütt et al., 2021; Batzner et al., 2022; Batatia et al., 2023; Musaelian et al., 2023; Liao & Smidt, 2023; Frank et al., 2024; Wang et al., 2024c).

However, these architectures rely on finite spatial cutoffs for efficiency, which limits their treatment of long-range interactions such as electrostatics, charge transfer, dispersion, polarization and other non-local electronic effects. These interactions govern interfaces, charged systems, polar condensed phases and biomolecular assemblies, among others (Jackson et al., 2019; Zhou & Pang, 2018; Yue et al., 2021; Niblett et al., 2021; Parsaeifard et al., 2021; Wang et al., 2024a; Parker et al., 2026). Stacking message-passing layers extends the receptive field on connected graphs, but increases cost, causes over-smoothing and over-squashing (Oono & Suzuki, 2020; Alon & Yahav, 2021; Di Giovanni et al., 2023), and cannot connect fragments that are disconnected in the cutoff graph. These limitations motivate dedicated long-range channels with tractable scaling.

Long-range extensions include latent scalar methods (Kosmala et al., 2023; Cheng, 2025), charge equilibration (Ko et al., 2021; Fuchs et al., 2025; Gubler et al., 2024), latent tensor fields (Rumiantsev

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al., 2026; Zhang et al., 2026), polarizable multipoles (Batatia et al., 2026; Park et al., 2026), global attention and aggregation (Qu et al., 2026; Frank et al., 2026; Caruso et al., 2026), matrix-function representations (Batatia et al., 2024), and multiscale representations (Han et al., 2022; Zhdanov et al., 2025; Solé et al., 2026) (Table 16). These advances extend non-local modeling, but none jointly learns the long-range propagation kernel, cross-degree angular coupling, or explicit higher-body-order interactions among distant sources on an adaptive hierarchy with $\mathcal{O}(N)$ time and memory. The Fast Multipole Method (FMM) is a natural foundation for these requirements (Greengard & Rokhlin, 1987; Cheng et al., 1999). It represents well-separated source clusters by compact multipole expansions across a spatial hierarchy with linear complexity. For the Coulomb kernel, spherical harmonics and analytical radial factors give a physically grounded basis for equivariant transport, built from the same ingredients as equivariant message passing (Cheng et al., 1999; Batatia et al., 2025).

We introduce *Neural Equivariant Multipole Operators (NEMORA)*, a neural equivariant extension of the FMM for learning long-range many-body interactions (Figure 1). NEMORA replaces scalar source charges with learned equivariant tensor charges, and the analytical expansion and translation operators with learnable $E(3)$-equivariant counterparts on an adaptive spatial hierarchy. Its operators learn weights over Clebsch–Gordan paths, coupling angular degrees in the long-range channel. Learnable non-linear radial modulations and invariant gates adapt the analytical Green's-function kernels to the data. Symmetric tensor contractions (Drautz, 2019; Batatia et al., 2023) on the aggregated fields form explicit higher-body-order interactions among distant collections of particles. The long-range messages then update the short-range backbone features and are passed through a readout layer as a residual energy term, with forces obtained via automatic differentiation. Evaluation is $\mathcal{O}(N)$ in time and memory for both periodic and non-periodic systems. Our contributions are:

- **Learnable equivariant FMM.** Learned angular transport and symmetric tensor contractions build multiscale long-range many-body representations with $\mathcal{O}(N)$ time and memory cost over an adaptive spatial hierarchy that preserves equivariance up to eigenframe degeneracy.
- **Global reaction-field channel.** Inspired by continuum electrostatics (Born, 1920; Onsager, 1936; Kirkwood, 1939), a learnable Kelvin transform (Amaral et al., 2017) inverts globally pooled multipoles into a same-center local expansion, an equivariant continuum boundary response broadcast to atoms that complements hierarchical transport (Appendix I).
- **Accuracy, portability, and scaling.** Up to three orders of magnitude lower error than short-range models on non-local benchmarks, and competitive with or better than other long-range methods; portability across four short-range backbones, both symmetry-constrained and unconstrained; and linear runtime and memory through 65,536 atoms.

2 Background

2.1 Equivariant Machine-Learned Interatomic Potentials

Machine-learned Interatomic Potentials An MLIP maps atomic positions $\{\boldsymbol{r}_i\}_{i=1}^N \subset \mathbb{R}^3$ and species $\{Z_i\}_{i=1}^N$ to an invariant potential energy $E = \sum_i E_i$, with forces $\boldsymbol{F}_i = -\nabla_{\boldsymbol{r}_i} E$ (Behler & Parrinello, 2007; Batzner et al., 2022). Species embeddings initialize scalar node features $\boldsymbol{h}_i^{(0)}$, and a cutoff graph defines $\mathcal{N}(i) = \{j \neq i : \|\boldsymbol{r}_j - \boldsymbol{r}_i\| < r_{\text{cut}}\}$. With $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$, message-passing layer $t \in \{1, \dots, T\}$ updates these features as (Gilmer et al., 2017)

$$\boldsymbol{h}_i^{(t)} = U_t\Big(\boldsymbol{h}_i^{(t-1)}, \bigoplus_{j \in \mathcal{N}(i)} M_t\big(\boldsymbol{h}_i^{(t-1)}, \boldsymbol{h}_j^{(t-1)}, \boldsymbol{r}_{ij}\big)\Big), \tag{1}$$

where M_t and U_t are learned message and update functions, and $\bigoplus$ is permutation-invariant aggregation. Invariant readouts give $E_i = \sum_{t=1}^T \mathcal{R}_t(\boldsymbol{h}_i^{(t)})$.

Equivariant Graph Neural Networks Equivariant features are organized into blocks $h_{i,km}^{(\ell)}$ with channel k, angular degree ℓ, and magnetic component $m \in \{-\ell, \dots, \ell\}$. Under a rotation $R \in \mathrm{SO}(3)$, $h_{i,km}^{(\ell)} \mapsto \sum_{m'} D_{mm'}^{(\ell)}(R) h_{i,km'}^{(\ell)}$, where $D^{(\ell)}$ is a Wigner D-matrix. The Clebsch–Gordan (CG) tensor product couples feature blocks (Kondor, 2018):

$$\big(\boldsymbol{f}^{(\ell_1)} \otimes \boldsymbol{g}^{(\ell_2)}\big)_m^{(\ell)} = \sum_{m_1, m_2} C^{\ell m}_{\ell_1 m_1, \ell_2 m_2} f_{m_1}^{(\ell_1)} g_{m_2}^{(\ell_2)}, \qquad |\ell_1 - \ell_2| \le \ell \le \ell_1 + \ell_2. \tag{2}$$

Learned weights, which may depend on invariants such as pair distances, combine channels along the allowed coupling paths (Thomas et al., 2018; Geiger & Smidt, 2022). Pooling these tensor-product messages over neighbors gives the one-particle features $\boldsymbol{a}_i$. Atomic cluster expansions build higher-body-order features from $\boldsymbol{a}_i$ by symmetric tensor contractions, where correlation order ν gives terms of up to $\nu + 1$ atoms (Drautz, 2019; Dusson et al., 2022; Batatia et al., 2023).

2.2 Fast Multipole Method

For scalar charges q_j, the Coulomb potential $V(\boldsymbol{r}_i) = (4\pi\varepsilon_0)^{-1} \sum_{j\neq i} q_j/|\boldsymbol{r}_i - \boldsymbol{r}_j|$ costs $\mathcal{O}(N^2)$ to evaluate directly. The FMM reduces this to $\mathcal{O}(N)$ by grouping particles in a spatial hierarchy and summarizing each distant group by a truncated multipole expansion (Greengard & Rokhlin, 1987; Cheng et al., 1999). It rests on the Laplace expansion of $1/|\boldsymbol{r}_> - \boldsymbol{r}_<|$, where $\boldsymbol{r}_< = \boldsymbol{r}_j - \boldsymbol{c}_s$ is a particle's offset from its source center and $\boldsymbol{r}_> = \boldsymbol{c}_t - \boldsymbol{c}_s$ is the separation of source and target centers. This expansion splits each interaction into a regular factor $|\boldsymbol{r}_<|^\ell Y^\star_{\ell m}(\hat{\boldsymbol{r}}_<)$ and an irregular factor $Y_{\ell m}(\hat{\boldsymbol{r}}_>)/|\boldsymbol{r}_>|^{\ell+1}$. The regular factor builds and re-centers expansions. Particle-to-multipole (P2M) and multipole-to-multipole (M2M) aggregate source moments up the tree, and local-to-local (L2L) and local-to-particle (L2P) carry local expansions down and evaluate them. The irregular factor performs the single long-range step, multipole-to-local (M2L), which converts the multipoles of well-separated clusters into a local expansion. For the interaction list $\mathcal{N}_t$ of $\boldsymbol{c}_t$, the far field is

$$\Phi_t \approx \sum_{s\in\mathcal{N}_t} \sum_{\ell=0}^{p} \sum_{m=-\ell}^{\ell} \Phi_s \cdot \frac{4\pi}{2\ell+1} \frac{|\boldsymbol{r}_<|^\ell}{|\boldsymbol{r}_>|^{\ell+1}} Y^\star_{\ell m}(\hat{\boldsymbol{r}}_<)\, Y_{\ell m}(\hat{\boldsymbol{r}}_>). \tag{3}$$

where p is the expansion order and Φ_s is the source: the charge q_j at P2M, and thereafter the moments $\Phi_{s,\ell m} = \sum_{j\in s} q_j |\boldsymbol{r}_<|^\ell Y^\star_{\ell m}(\hat{\boldsymbol{r}}_<)$. The series converges when $(R_s + R_t)/|\boldsymbol{r}_>| < \theta < 1$ for bounding radii R_s, R_t. At fixed p, with bounded leaf occupancy and interaction lists, evaluation takes $\mathcal{O}(N)$ time and memory (Cheng et al., 1999). Equivariant models absorb exactly these explicit sums over ℓ and m into CG tensor products.

3 Related Work

Scalar Long-Range Channels and Charge Equilibration Physics-informed corrections predict per-atom scalars and couple them through prescribed Coulomb or damped dispersion potentials (Yao et al., 2018; Unke et al., 2021a; Plé et al., 2023; Kabylda et al., 2024). Reciprocal-space approaches communicate the same invariant channels through Ewald or particle-mesh sums (Kosmala et al., 2023; Wang et al., 2024b; Cheng, 2025; Loche et al., 2025). Sum-of-Gaussians networks instead learn the reciprocal-space decay tail at near-linear cost via non-uniform FFTs (Ji et al., 2025; 2026), and prolate spheroidal mollification represents prescribed inverse-power channels with fewer Fourier modes (Liang et al., 2026). Charge equilibration redistributes globally coupled scalar charges (Rappe & Goddard, 1991; Ghasemi et al., 2015; Khajehpasha et al., 2022; Ko et al., 2021; Fuchs et al., 2025; Maruf et al., 2025; Plé et al., 2025). Dense solves cost $\mathcal{O}(N^3)$ time and $\mathcal{O}(N^2)$ memory (Fuchs et al., 2025; Gubler et al., 2024), and particle-mesh, iterative, and variational formulations improve scaling (Gubler et al., 2024; Kocer et al., 2025; Shaidu et al., 2024; Vondrák et al., 2025; Kim & Cheng, 2026). NEMORA transmits scalar and higher-degree tensor charges through a learned kernel at $\mathcal{O}(N)$ cost, on a real-space hierarchy that requires neither a Fourier grid nor periodicity.

Tensorial Transport and Many-Body Representations Density-based descriptors encode non-local environments through atomic density and potential representations (Grisafi & Ceriotti, 2019; Faller et al., 2024). Multipolar corrections (Thürlemann & Riniker, 2023; Park et al., 2026; Batatia et al., 2026; Guo et al., 2026) and multipole embeddings of environmental charges (Barrett et al., 2026) capture directional electrostatics but keep prescribed interaction kernels. Tensorial long-range extensions use fixed componentwise radial kernels or degree-preserving spectral filters (Rumiantsev et al., 2026; Zhang et al., 2026). For equivariant linear maps conditioned only on invariants, Schur's lemma restricts learned weights to channel mixing within each degree (Geiger & Smidt, 2022; Rumiantsev et al., 2026). Global matrix-function approaches give non-local equivariant many-body representations, but at $\mathcal{O}(N^3)$ time and $\mathcal{O}(N^2)$ memory (Batatia et al., 2024). NEMORA's operators condition on directions through $\boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{r}})$, so learned CG paths couple degrees, and symmetric contractions of the aggregated tensors add many-body terms at each level, at linear cost.

Global and Hierarchical Communication Global all-to-all attention learns non-local coupling at $\mathcal{O}(N^2)$ cost (Qu et al., 2026). Scalar virtual-node aggregation is linear but transmits only invariants (Caruso et al., 2026). Tensorial global attention is linear at a fixed spherical quadrature, but its rotational equivariance is exact only up to the quadrature degree, which must grow with frequency and spatial extent (Frank et al., 2026). Equivariant atom–fragment attention transmits directional features over an enlarged finite cutoff (Wang et al., 2026). Graph hierarchies communicate across scales through coarsening, message passing and pooling, including equivariant constructions (Ying et al., 2019; Rampášek & Wolf, 2021; Ngo et al., 2023; Han et al., 2022; Hy & Kondor, 2023; Xu et al., 2025; Solé et al., 2026). Tree-based architectures combine spatial partitioning with local attention (Zhdanov et al., 2025), fragment-based corrections couple atoms and chemistry-defined fragments (Li et al., 2024), and hierarchical particle models canonicalize two-level local and global graphs in PCA frames, with predefined groups aggregated into averaged invariant nodes (Linot et al., 2025). NEMORA instead builds its hierarchy on the FMM well-separation criterion, so its analytical limit inherits multipole convergence guarantees.

FMM-Inspired Neural Architectures Multilevel graph kernels and $\mathcal{H}^2$-matrix networks learn non-local operators on graphs and point clouds (Li et al., 2020; Sushnikova et al., 2022). Neural FMMs learn hierarchical translation stages for PDE solution operators, including point-cloud variants (Fognini et al., 2025; 2026), and multipole-inspired attention cuts global attention cost (Colagrande et al., 2025; Kang et al., 2023). Differentiable FMM solvers give gradients for inverse problems with prescribed kernels (Stücker & Foldal, 2026). For atomistic potentials, NEMORA translates multipole features between well-separated regions of an adaptive hierarchy with equivariant tensor transport, learned angular coupling, modulated analytical radial factors, and many-body contractions.

4 NEMORA Architecture

NEMORA Framework NEMORA makes the FMM operators (P2M, M2M, M2L, L2L, L2P) learnable, passing equivariant tensor features up to a chosen $\ell_{\max}$ through an adaptive hierarchy. Unless stated otherwise, NEMORA is attached to a MACE backbone (Batatia et al., 2023).

4.1 Neural Equivariant Multipole Operators

Learnable equivariant tensor charges. We replace q_j with an equivariant tensor charge $\boldsymbol{Q}_j^{(\ell_Q)}$. As in equivariant message passing, the scalar product in Equation 3 becomes a CG tensor product that absorbs the sum over m, leaving a concatenated filter over degrees ℓ. The kernel is unchanged:

$$\boldsymbol{\Phi}_t^{(\ell_M)} \approx \sum_{s\in\mathcal{N}_t} \left[\boldsymbol{\Phi}_s \otimes \bigoplus_{\ell=0}^{p} \frac{4\pi}{2\ell+1} \frac{\|\boldsymbol{r}_<\|^\ell}{\|\boldsymbol{r}_>\|^{\ell+1}} \boldsymbol{Y}^{\star(\ell)}(\hat{\boldsymbol{r}}_<)\, \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{r}}_>)\right]^{(\ell_M)}. \tag{4}$$

Here $\boldsymbol{\Phi}_s$ is $\boldsymbol{Q}_j^{(\ell_Q)}$ at P2M and the tensor moments afterwards. The charge couples to $\boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{r}}_<)$ at an intermediate degree ℓ_1 and then to $\boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{r}}_>)$. With the fixed recoupling weights $c_{\ell_1} = (-1)^{\ell_Q+\ell_1}\sqrt{(2\ell_1+1)/(2\ell_Q+1)}$ (Appendix I), every output degree except $\ell_M = \ell_Q$ vanishes. The transport is then the component-wise Coulomb sum of Equation 14, which for $\ell_Q = 0$ is Equation 3. We call this analytical variant of NEMORA a *latent FMM*: analytical long-range operators that transmit learned charges, which are scalar monopoles at $\ell_Q = 0$ and, through kernel derivatives, atomic multipoles at higher ℓ_Q. These atomic multipoles add angular degrees of freedom that distributed scalar charges lack. The source degree ℓ_Q sets expressivity and the expansion order p sets accuracy. Latent Ewald summation (Cheng, 2025; Rumiantsev et al., 2026) evaluates learned charges with Ewald sums, whereas the latent FMM uses the adaptive, $\mathcal{O}(N)$ FMM algorithm.

Centered on atom i with backbone features as sources, the P2M stage of Equation 4 takes the form of an equivariant MLIP message:

$$\boldsymbol{a}_i^{(\ell_M)} = \sum_{j\in\mathcal{N}(i)} \left[\boldsymbol{h}_j \otimes_\theta \bigoplus_{\ell=0}^{\ell_{\max}} \boldsymbol{R}_\theta^{(\ell)}(\|\boldsymbol{r}_{ij}\|)\, \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{r}}_{ij})\right]^{(\ell_M)}. \tag{5}$$

This is the pooled tensor-product message of Section 2.1, read as a multipole expansion. The feature $\boldsymbol{h}_j$ plays the role of $\boldsymbol{Q}_j$, the radial weights $\boldsymbol{R}_\theta^{(\ell)}$ the role of $\|\boldsymbol{r}_{ij}\|^\ell$, and the path weights of $\otimes_\theta$ the

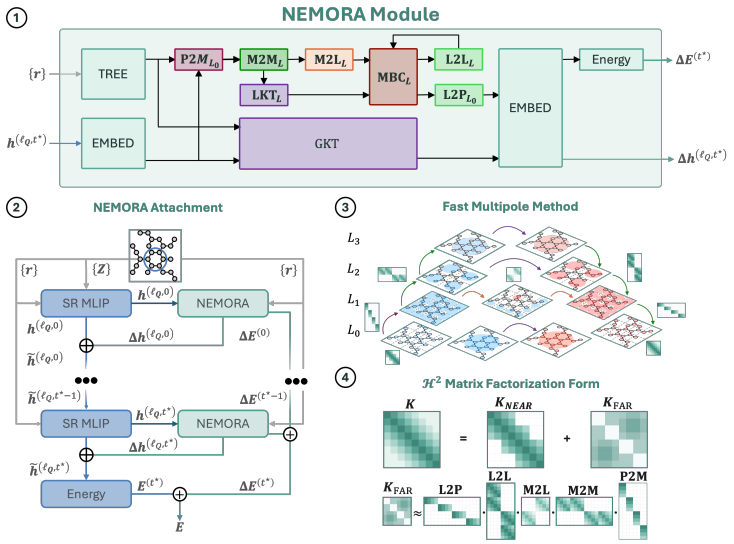

Figure 1: **NEMORA: a learnable Fast Multipole Method for atomistic ML. (1) NEMORA module (top).** Given backbone features $\boldsymbol{h}_i^{(\ell_Q,t^\star)}$ and positions $\{\boldsymbol{r}_i\}$, NEMORA returns equivariant corrections $\Delta\boldsymbol{h}_i^{(\ell_Q,t^\star)}$ and a scalar contribution $\Delta E^{(t^\star)}$. An upward sweep (P2M, M2M) builds cluster moments; at each downward level, the M2L far field is fused with a local Kelvin self-correction (LKT) and a many-body contraction (MBC) before L2L. Leaf L2P outputs are summed with a global Kelvin reaction-field branch (GKT). **(2) NEMORA–MLIP integration (bottom left).** NEMORA updates the backbone equivariants at iteration $t^\star$, and per-iteration readouts add $\Delta E^{(t^\star)}$ to the atomic energy. **(3, 4) Classical FMM (right).** Translation of multipole (blue) and local (red) expansions, and the corresponding $\mathcal{H}^2$ factorization $K = K_{\text{NEAR}} + K_{\text{FAR}}$ with $K_{\text{FAR}} \approx \text{L2P}\cdot\text{L2L}\cdot\text{M2L}\cdot\text{M2M}\cdot\text{P2M}$.

role of c_{ℓ_1}. Setting $h_j^{(0)} = q_j$ and $\boldsymbol{R}^{(\ell)} = \|\boldsymbol{r}_{ij}\|^\ell$ recovers the classical moments about atom i. We therefore read the pooled features of the one-particle basis (Batatia et al., 2025) as learnable local multipole expansions, whose symmetric contractions build body order, and let the FMM carry them across length scales. This correspondence is NEMORA's design principle: local message passing already makes P2M learnable, and NEMORA does the same for every other FMM stage.

NEMORA acts at chosen layers $t^\star$ of the backbone (Section 2.1) and takes its tensor charges from these features. An input projection mixes channels within each degree, $\boldsymbol{Q}_i^{(\ell_Q)} = \boldsymbol{W}_{\text{in}}^{(\ell_Q)}\boldsymbol{h}_i^{(\ell_Q,t^\star)}$, and the operators couple degrees along the hierarchy. A degree-preserving projection $\boldsymbol{W}_{\text{out}}^{(\ell_Q)}$ and an invariant gate return the far-field update $\Delta\boldsymbol{h}_i^{(\ell_Q,t^\star)}$. The corrected features replace $\boldsymbol{h}_i^{(\ell_Q,t^\star)}$ in the layer-$t^\star$ readout and all later layers, and a scalar readout of the update adds $\Delta E_i^{(t^\star)}$ to the energy:

$$\tilde{\boldsymbol{h}}_i^{(\ell_Q,t^\star)} = \boldsymbol{h}_i^{(\ell_Q,t^\star)} + \Delta\boldsymbol{h}_i^{(\ell_Q,t^\star)}, \qquad E_i = \sum_{t=1}^{T}\mathcal{R}_t(\boldsymbol{h}_i^{(t)}) + \underbrace{\text{MLP}\big(\Delta\boldsymbol{h}_i^{(0,t^\star)}\big)}_{\Delta E_i^{(t^\star)}}. \tag{6}$$

Learnable equivariant multipole operators. Following this correspondence, we replace each analytical operator with a learnable equivariant one:

$$\boldsymbol{\Psi}_t^{(\ell_M)} = \sum_{s\in\mathcal{N}_t}\alpha_{\theta,st}^{(\ell_M)}\cdot\Big[\boldsymbol{W}^V\boldsymbol{\Psi}_s\otimes_\theta\bigoplus_{\ell=0}^{p}\frac{\|\boldsymbol{r}_<\|^\ell}{\|\boldsymbol{r}_>\|^{\ell+1}}\boldsymbol{\rho}_\theta^{(\ell)}\big(\boldsymbol{R}(\|\boldsymbol{r}\|)\big)\,\boldsymbol{Y}^{\star(\ell)}(\hat{\boldsymbol{r}}_<)\,\boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{r}}_>)\Big]^{(\ell_M)}. \tag{7}$$

Here $\boldsymbol{\Psi}_s$ is the tensor charge at P2M and the learned moments afterwards. As in Section 2.2, each operator applies the factor for its own displacement $\boldsymbol{r}$: the regular factor for P2M, M2M, L2L and L2P, and the irregular factor for M2L. $\boldsymbol{W}^V$ mixes channels within each degree, and $\otimes_\theta$ learns a weight for every allowed CG path, replacing the fixed c_{ℓ_1} and $4\pi/(2\ell+1)$. $\boldsymbol{\rho}_\theta^{(\ell)}$ modulates the analytical radial factor per degree and channel, as a learned function of the same radius on a sinc–Bessel basis

$\boldsymbol{R}$ (Gasteiger et al., 2020). A gate $\alpha_{\theta,st}^{(\ell_M)} = \phi_\theta(\mathcal{G}_{st} \oplus \mathcal{M}_{st} \oplus \mathcal{A}_{st})^{(\ell_M)}$ weights each source by geometric invariants $\mathcal{G}_{st}$, normalized features $\mathcal{M}_{st}$ and alignment invariants $\mathcal{A}_{st}$. It acts as a learned analogue of damping and polarization factors (Tang & Toennies, 1984; Ponder et al., 2010; Stone, 2013) (Appendix I). Setting $\boldsymbol{W}^V = \boldsymbol{I}$, $\boldsymbol{\rho}_\theta = 1$, $\alpha_\theta = 1$ and the analytical path weights recovers the latent FMM. At each downward level, symmetric contractions of the pooled M2L field add explicit many-body features (MBC; Appendix I), and a learnable Kelvin transform adds a continuum boundary response (LKT/GKT; Appendix I). Under periodic boundary conditions, a finite-shell, degree-diagonal correction at the root accounts for periodic images (Equation 9).

4.2 Adaptive $E(3)$-Equivariant Hierarchy

The hierarchy is built by top-down adaptive subdivision. Each cluster splits along the PCA eigenframe of its centered atomic offsets, or the Gram–Schmidt lattice frame under PBC. An axis is active when its eigenvalue ratio $\lambda_k/\lambda_{\max}$ exceeds a threshold, and the fanout $2^{n_{\text{active}}}$ gives binary, quadtree, or octree splits for chain, slab, and bulk geometries. A cluster becomes a leaf when its predicted child diameter falls below the backbone receptive field $n_{\text{SR}} r_{\text{cut}}$ or its atom count drops below a threshold. Equivariance requires frame selection and cluster assignment to transform with the geometry, including periodic cells, with a fixed convention for degenerate eigenframes and sign or partition ties. At level L, the near set is $\text{Near}_L(B) = \{A : \|\boldsymbol{c}_A - \boldsymbol{c}_B\|_{\text{MIC}} \le \rho_{\text{near}}(R_A + R_B)\}$ with $\rho_{\text{near}} \ge 1$, so every far pair meets the convergence criterion of Section 2.2 with $\theta = 1/\rho_{\text{near}}$. The M2L list $\mathcal{I}_L(B)$ holds the children of the parent's near clusters that are not near B. Near leaf pairs also use M2L, weighted by an envelope of their gap distance $d_{\min} = \max(\|\boldsymbol{c}_A - \boldsymbol{c}_B\|_{\text{MIC}} - R_A - R_B, 0)$ so that the short- and long-range channels hand off smoothly (Gasteiger et al., 2020). Evaluation is linear in N when the hierarchy has $\mathcal{O}(N)$ nodes and bounded interaction lists (Appendix I).

5 Results

We evaluate NEMORA on molecular and periodic systems that probe different forms of nonlocal interaction. Unless otherwise stated, tables report means over three independent training runs, with energy in meV/atom and force in meV/Å. Training details are in Appendix A.

5.1 Cumulene

Two cumulene experiments probe long-range electronic delocalization and its dependence on geometry (Batatia et al., 2024). The first trains on $n_C = 3$–$10, 13, 14$ and tests on these lengths (ID), on the unseen lengths 11 and 12 (semi-out-of-distribution, SOD), and on the longer chains 15 and 16 (OOD). The second covers the torsional profile and chain-length dependence on a separate 3–64 carbon set.

Table 1: Cumulene chain-length, charged-system (Ko) and MD22 benchmark errors. Each entry is E RMSE (meV/atom) / F RMSE (meV/Å). Cumulene splits are defined in Section 5.1. In each row and metric, lowest means are bold and second-lowest underlined (ties retained).

System	MACE E RMSE / F RMSE	MACE–LES E RMSE / F RMSE	MACE–RANGE E RMSE / F RMSE	LOREM E RMSE / F RMSE	EFA E RMSE / F RMSE	MACE–MFN E RMSE / F RMSE	**NEMORA (Ours)** E RMSE / F RMSE
Cumulene							
ID	16.898 / 85.07	17.235 / 84.82	19.505 / 95.60	6.756 / 29.36	5.918 / 30.28	**3.231 / 13.83**	3.286 / 28.18
SOD	16.709 / 102.16	16.560 / 102.17	18.796 / 112.98	27.251 / 166.70	27.884 / 158.17	**0.643 / 14.25**	1.877 / 43.46
OOD	11.994 / 72.26	11.180 / 72.36	16.411 / 94.90	25.008 / 133.06	20.810 / 93.92	**2.812 / 15.28**	3.563 / 51.55
Charged							
Carbon chain	0.425 / 15.00	0.524 / 20.15	0.932 / 12.38	0.640 / 4.62	0.502 / 5.64	**0.156** / 1.90	0.175 / **1.88**
NaCl	1.600 / 43.89	0.380 / 7.21	0.622 / 2.93	**0.243** / 1.76	0.861 / 18.13	0.361 / 2.75	0.412 / **1.75**
Au_2/MgO	2.385 / 110.07	1.696 / 21.85	0.207 / 12.18	0.235 / **4.64**	1.890 / 13.03	—	**0.197** / 5.26
MD22							
Ac-Ala3-NHMe	0.405 / 7.44	0.348 / 7.08	—	—	—	—	**0.323 / 6.93**
DHA	0.575 / **5.64**	0.586 / 5.87	—	—	—	—	**0.455** / 5.70
Stachyose	**0.375** / 6.64	0.463 / **6.24**	—	—	—	—	0.401 / 6.45
AT-AT	1.168 / 10.78	1.005 / 8.04	—	—	—	—	**0.340 / 7.85**
AT-AT-CG-CG	0.766 / 10.26	0.927 / **8.12**	—	—	—	—	**0.341** / 8.60
buckyball-catcher	1.099 / 8.22	1.087 / 7.96	—	—	—	—	**1.081 / 7.54**
Nanotube	1.406 / 24.75	1.806 / 28.91	—	—	—	—	**1.362 / 22.32**

Table 1 shows that NEMORA substantially improves energy and force predictions over short-range MACE and scalar long-range models, and is second only to MACE–MFN on every split and metric.

LOREM, EFA and NEMORA have similar ID force errors, but on unseen chain lengths LOREM and EFA degrade below short-range MACE, whereas NEMORA remains well below it. Together with the scalar NEMORA ablations (Appendix A.3), these results indicate that transmitting angular information is important for cumulene. MACE–MFN remains the most accurate on these splits, at the cost of cubic time and quadratic memory scaling (Figure 4).

Figure 2(a,b) presents the separate angular and chain-length experiments. Rotating the terminal groups changes the energy of the entire conjugated chain, even when the terminal groups lie beyond the short-range receptive field. Short-range MACE and the scalar long-range models MACE–LES and MACE–RANGE predict flat torsional profiles and miss this angular dependence. NEMORA closely follows the DFT profile and reproduces the even–odd energy variation with chain length. The corresponding energy and force errors are in Table 6.

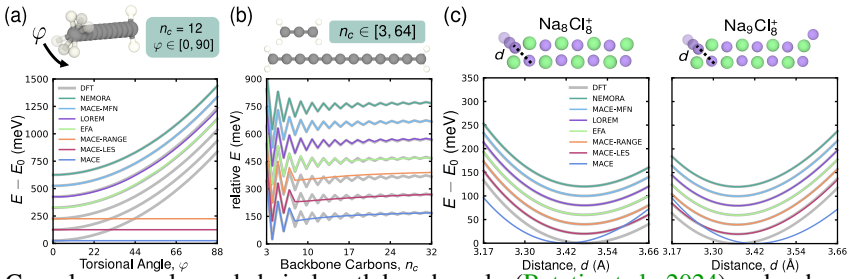

Figure 2: Cumulene angular and chain-length benchmarks (Batatia et al., 2024) and a charged sodium chloride cluster benchmark (Ko et al., 2021). (a) Torsional energy profile for a 12-carbon chain. (b) Energy as a function of chain length at $\varphi = 5°$. (c) Energy versus Na displacement in $Na_8Cl_8^+$ and $Na_9Cl_8^+$, where an extra neutral Na atom about 21 Å away changes the binding profile. Gray curves are DFT references, and model curves are shifted vertically for clarity.

5.2 Nonlocal charge transfer

The benchmark datasets introduced by Ko et al. (2021) test the effect of nonlocal charge transfer in different contexts. In the $C_{10}H_2/C_{10}H_3^+$ system, protonation changes the electronic structure along a conjugated carbon chain. In the $Na_8Cl_8^+/Na_9Cl_8^+$ (NaCl) cluster, an additional neutral sodium atom alters the binding of an atom on the opposite side. The Au_2/MgO dataset tests the effect of substrate doping on an adsorbed gold dimer under periodic boundary conditions.

NEMORA substantially improves on MACE across all three datasets and, among long-range models, ranks first or second on five of six metrics, with the lowest force errors on the carbon chain and NaCl and the lowest energy error on Au_2/MgO. It does so while retaining $\mathcal{O}(N)$ scaling in both periodic and non-periodic systems. Figure 2(c) illustrates the nonlocal energy response in the NaCl cluster.

Table 1 excludes charge-equilibration and explicitly electrostatic models (CELLI, 4G–HDNNP and SpookyNet) (Fuchs et al., 2025; Ko et al., 2021; Unke et al., 2021a), which add reference-charge terms to the loss or take total charge and spin as inputs. Using these extra inputs and loss terms, CELLI and SpookyNet reach lower NaCl and Au_2/MgO energy errors (Table 4, Appendix A).

5.3 MD22

To test whether NEMORA degrades accuracy when long-range interactions are weak, we train and evaluate it on all seven MD22 systems (Chmiela et al., 2023), comparing MACE, MACE–LES and NEMORA with the same short-range backbone and data splits. As shown in Table 1, NEMORA lowers the energy and force errors relative to MACE on six of seven systems each, and has the lowest energy error of the three models on six systems. The largest gains are on the DNA base-pair systems, where energy errors fall from 1.168 to 0.340 meV/atom (AT-AT) and from 0.766 to 0.341 meV/atom (AT-AT-CG-CG). On the nanotube, NEMORA reduces force error from 24.75 to 22.32 meV/Å, while LES increases it to 28.91 meV/Å. The two regressions relative to MACE are small: 0.06 meV/Å in force on DHA and 0.026 meV/atom in energy on Stachyose. NEMORA therefore preserves or improves the accuracy of the short-range backbone even when long-range interactions are weak.

5.4 PORTABILITY ACROSS BACKBONES AND THE ROLE OF ANGULAR FEATURES

Although we attach NEMORA to MACE throughout for simplicity, it can be attached to short-range architectures of many styles. To demonstrate this, we test MACE, NequIP, ViSNet, and PET (Batatia et al., 2023; Batzner et al., 2022; Wang et al., 2024c; Pozdnyakov & Ceriotti, 2023), varying the degree of the short-range features and the NEMORA sources (Table 2). For MACE, NequIP, and ViSNet, we compare the native equivariant backbone with a scalar-only variant. Each is trained without NEMORA, with scalar NEMORA ($\ell_Q = 0$), and with degree-two NEMORA ($\ell_Q = 2$). When the backbone supplies only scalar features, a geometry-based lift constructs the tensorial sources. The same interface allows attachment to the rotationally unconstrained PET architecture.

Table 2: Backbone and source-degree ablations, mean validation E / F RMSE (meV/atom / meV/Å). SR is the short-range backbone, and $\ell_Q = 0$ and $\ell_Q = 2$ are scalar and degree-two NEMORA sources, lifted for scalar and untyped hosts. Bold and underline mark the lowest and second-lowest of the three settings per host, dataset and metric (Appendix A.2).

		NaCl			Cumulene		
Backbone	Host	SR	$\ell_Q = 0$	$\ell_Q = 2$	SR	$\ell_Q = 0$	$\ell_Q = 2$
MACE (Batatia et al., 2023)	Native	1.597 / 43.92	0.457 / 3.45	**0.406 / 1.80**	13.672 / 109.82	13.767 / 105.93	**3.776 / 65.93**
	Scalar	1.600 / 44.51	0.475 / 7.56	**0.437 / 1.88**	52.466 / 196.28	53.357 / 195.93	**4.134 / 77.32**
NequIP (Batzner et al., 2022)	Native	1.565 / 43.89	0.664 / 6.39	**0.258 / 1.85**	54.699 / 203.77	55.097 / 206.56	**3.916 / 46.06**
	Scalar	1.774 / 62.70	0.561 / 11.40	**0.242 / 2.23**	55.939 / 239.69	56.752 / 239.29	**5.567 / 81.59**
ViSNet (Wang et al., 2024c)	Native	1.533 / 42.02	0.230 / 4.04	**0.145 / 1.67**	14.630 / 120.86	25.020 / 150.24	**3.723 / 46.35**
	Scalar	1.711 / 59.46	0.876 / 8.12	**0.302 / 2.30**	56.132 / 240.12	56.871 / 239.85	**4.150 / 70.53**
PET (Pozdnyakov & Ceriotti, 2023)	Untyped	1.524 / 38.91	**0.260** / 5.64	0.485 / **4.55**	54.814 / 241.43	45.183 / 221.56	**9.363 / 112.11**

On NaCl, where the nonlocal interaction is predominantly electrostatic, scalar NEMORA already captures most of the improvement over the short-range models. Degree-two sources further reduce force errors for every backbone, and energy errors for all but PET. On cumulene, scalar long-range communication gives little or no improvement, while degree-two sources reduce errors substantially for every backbone. The geometry-based lift recovers much of this improvement when the short-range backbone contains only scalar features. These results show that the NEMORA channel can supply angular information missing from the backbone, across all four backbones tested.

5.5 PERIODIC IONS AND WATER

We next consider a periodic binary ionic melt with Lennard-Jones interactions and fixed charges $\pm 0.6e$, with Ewald/PPPM reference energies and forces. For this pairwise system we use the latent FMM form of NEMORA. MACE–LES and NEMORA both reduce the errors of short-range MACE by two to three orders of magnitude in energy and by more than an order of magnitude in force, and extrapolate to larger unseen cells without losing accuracy (Table 3), with NEMORA slightly more accurate on nearly every size. Both recover the $\pm 0.6e$ charge in their latent scalar features (Table 17).

Table 3: Periodic-ion and water test errors, E RMSE (meV/atom) / F RMSE (meV/Å). Ion models are trained on $N = 256$–1024. Water models are trained on 128 and 256 molecules. NEMORA is the latent FMM with scalar (ions) or dipolar (water, $\ell_Q = 1$) sources (Appendix E). Except for the ion MACE rows, which are three-run means, entries are single training runs.

System	MACE	MACE–LES	NEMORA
Periodic ions			
$N = 256$	43.43 / 168.6	0.308 / 9.80	**0.284 / 9.64**
$N = 512$	24.59 / 195.8	**0.203** / 10.01	0.211 / **9.83**
$N = 1024$	65.90 / 265.2	0.155 / 9.88	**0.133 / 9.69**
$N = 2048$ (OOD)	97.24 / 329.8	0.142 / 9.95	**0.109 / 9.78**
$N = 4096$ (OOD)	110.63 / 363.5	0.120 / 9.95	**0.090 / 9.79**
Water	0.415 / 20.83	0.400 / 14.35	**0.305 / 11.86**

On periodic water, dipolar NEMORA ($\ell_Q = 1$) has the lowest energy and force errors (Table 3). With scalar sources it matches MACE–LES (Table 18), indicating that the gain comes from the dipoles.

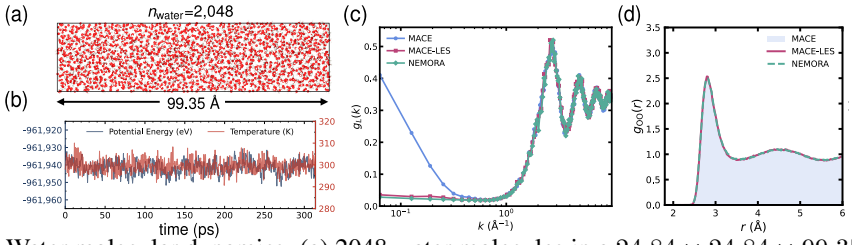

Figure 3: Water molecular dynamics. (a) 2048 water molecules in a $24.84 \times 24.84 \times 99.35$ Å box at 300 K. (b) NEMORA potential energy and temperature along the NVT trajectory. (c) Longitudinal dipole correlation $g_L(k)$. (d) O–O radial distribution $g_{\mathrm{OO}}(r)$. Models are those of Table 3.

Figure 3 shows NVT molecular dynamics of 2048 water molecules, eight times the largest training system, in an elongated periodic box whose long axis gives access to small wavevectors (Appendix A.5). Potential energy and temperature remain stable along the trajectory (b). NEMORA and MACE–LES reproduce the long-wavelength behavior of $g_L(k)$ (c), whereas short-range MACE rises sharply as $k \to 0$. All three models give nearly identical O–O radial distributions (d), so standard structural observables cannot distinguish them, which highlights the need for dedicated long-range benchmarks.

5.6 Time and Memory Scaling

NEMORA scales linearly in time and memory from 64 to 65,536 atoms (Figure 4), with fitted exponents of 0.82–0.91 for the long-range component. Values below one reflect fixed per-call overhead, such as kernel launches and GPU under-utilization at small N, rather than a sublinear algorithm. NEMORA is the only long-range model evaluated to reach 65,536 atoms within GPU memory for both periodic and non-periodic systems. Dense MFN runs out of memory between 512 and 1024 atoms. The evaluated LES and LOREM implementations use direct Ewald summation and scale roughly as $\mathcal{O}(N^2)$; particle-mesh variants would give $\mathcal{O}(N \log N)$, but we are not aware of public particle-mesh implementations of either model. RANGE and EFA also scale linearly but use more memory than NEMORA at a similar parameter count.

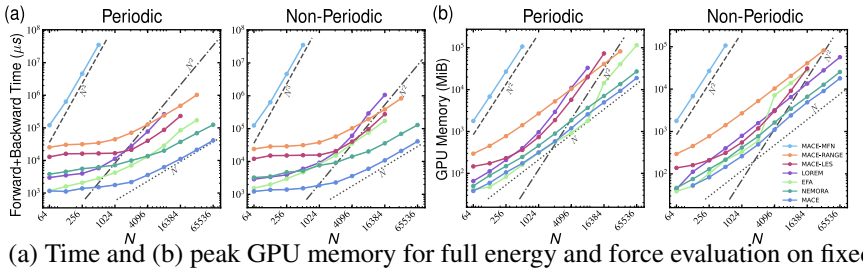

Figure 4: (a) Time and (b) peak GPU memory for full energy and force evaluation on fixed-density periodic and non-periodic argon, $N = 64$–$65{,}536$, on one H200. All models share the MACE backbone except EFA and LOREM, whose short-range parameter counts are closely matched.

6 Discussion and Conclusion

NEMORA extends short-range interatomic potentials with an equivariant multipole hierarchy that learns to transmit and combine angular information over long distances. On cumulene, scalar long-range models miss the response to terminal-group rotation, while NEMORA captures the torsional profile and, on unseen chain lengths, is second only to MACE–MFN at linear rather than cubic cost. The backbone and component ablations support this across architectures. On ionic clusters, where scalar charges describe much of the interaction, scalar NEMORA already gives a large improvement.

Fixing the transport operators gives a latent FMM with learned atomic charges and dipoles. The scalar model is comparable to the LES energy functional, and atomic dipoles improve water predictions. NEMORA combines these capabilities with linear runtime and memory, preserves or improves short-range accuracy on MD22, and supports stable periodic-water dynamics at eight times the training size. Its hierarchy is equivariant only up to eigenframe degeneracy, and learned operators forgo the convergence guarantees of the analytical FMM. Next steps are broader chemical tests.

Reproducibility statement

The reported experiments use both PyTorch and JAX implementations. Benchmark run counts, exclusions and differences in training schedules are recorded in Appendix A; the backbone, optimizer and NEMORA settings for the original benchmarks are given in Tables 20, 21 and 28. Additional experiment protocols accompany their results. Across-run uncertainties are SEMs unless explicitly identified as block estimates for molecular dynamics. Single-run measurements and selected-seed profiles are identified separately. Datasets originate from the cited primary sources, except the periodic water dataset generated in this work. Code and that dataset will be released upon publication.

AI use statement

In this work, we used generative AI tools under the authors' direction to help implement methods in code, run experiments, and produce figures for results. We also used them to draft and edit parts of the paper and to search and summarize literature. We have reviewed all AI-assisted work: code was verified against analytical references and unit tests, reported numbers were checked against the runs that produced them, and cited works were checked against their sources. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

This work develops machine-learning interatomic potentials for molecular and materials simulation. It involves no human subjects, personal data or crowd-sourced annotation. The benchmarks use published datasets of atomic configurations with reference energies and forces, cited in the text, and the periodic water dataset generated in this work contains only configurations of this kind. Faster and more accurate atomistic simulation could in principle support harmful applications, but the methods presented here add no capability specific to such uses beyond that of existing interatomic potentials. The authors have read and adhere to the ICLR Code of Ethics.

References

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2006.05205.

R. L. P. G. Amaral, O. S. Ventura, and N. A. Lemos. Kelvin transformation and inverse multipoles in electrostatics. *European Journal of Physics*, 38(2):025206, March 2017. ISSN 0143-0807, 1361-6404. doi: 10.1088/1361-6404/aa55a0. URL http://arxiv.org/abs/1611.05942. arXiv:1611.05942 [physics].

Dylan M. Anstine and Olexandr Isayev. Machine Learning Interatomic Potentials and Long-Range Physics. *The Journal of Physical Chemistry A*, 127(11):2417–2431, March 2023. ISSN 1089-5639. doi: 10.1021/acs.jpca.2c06778. URL https://doi.org/10.1021/acs.jpca.2c06778.

B. M. Axilrod and E. Teller. Interaction of the van der Waals Type Between Three Atoms. *The Journal of Chemical Physics*, 11(6):299–300, June 1943. ISSN 0021-9606. doi: 10.1063/1.1723844. URL https://pubs.aip.org/aip/jcp/article/11/6/299/182086/Interaction-of-the-van-der-Waals-Type-Between.

Rhyan Barrett, Johannes C. B. Dietschreit, and Julia Westermayr. Incorporating long-range interactions via the multipole expansion into ground and excited-state molecular simulations. *npj Computational Materials*, 12(1):135, March 2026. ISSN 2057-3960. doi: 10.1038/s41524-026-02048-3. URL https://www.nature.com/articles/s41524-026-02048-3.

Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Physical Review Letters*, 104(13): 136403, April 2010. doi: 10.1103/PhysRevLett.104.136403. URL https://link.aps.org/doi/10.1103/PhysRevLett.104.136403.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields, January 2023. URL http://arxiv.org/abs/2206.07697. arXiv:2206.07697 [cond-mat, physics:physics, stat].

Ilyes Batatia, Lars L. Schaaf, Gábor Csányi, Christoph Ortner, and Felix A. Faber. Equivariant matrix function neural networks. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/0d5bfa7cb4d570b23a0088e508020843-Abstract-Conference.html.

Ilyes Batatia, Simon Batzner, Dávid Péter Kovács, Albert Musaelian, Gregor N. C. Simm, Ralf Drautz, Christoph Ortner, Boris Kozinsky, and Gábor Csányi. The design space of E(3)-equivariant atom-centred interatomic potentials. *Nature Machine Intelligence*, 7:56–67, January 2025. ISSN 2522-5839. doi: 10.1038/s42256-024-00956-x. URL https://www.nature.com/articles/s42256-024-00956-x.

Ilyes Batatia, William J. Baldwin, Domantas Kuryla, Joseph Hart, Elliott Kasoar, Alin M. Elena, Harry Moore, Mikołaj J. Gawkowski, Benjamin X. Shi, Venkat Kapil, Panagiotis Kourtis, Ioan-Bogdan Magdău, and Gábor Csányi. MACE-POLAR-1: A Polarisable Electrostatic Foundation Model for Molecular Chemistry, February 2026. URL http://arxiv.org/abs/2602.19411. arXiv:2602.19411 [physics].

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5. URL https://www.nature.com/articles/s41467-022-29939-5.

Jörg Behler and Michele Parrinello. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters*, 98(14):146401, April 2007. doi: 10.1103/PhysRevLett.98.146401. URL https://link.aps.org/doi/10.1103/PhysRevLett.98.146401.

M. Born. Volumen und Hydratationswärme der Ionen. *Zeitschrift für Physik*, 1(1):45–48, February 1920. ISSN 1434-6001, 1434-601X. doi: 10.1007/BF01881023. URL http://link.springer.com/10.1007/BF01881023.

Alessandro Caruso, Jacopo Venturin, Lorenzo Giambagli, Edoardo Rolando, Zakariya El-Machachi, Frank Noé, and Cecilia Clementi. Extending the range of graph neural networks with global encodings. *Nature Communications*, 17(1):1855, February 2026. ISSN 2041-1723. doi: 10.1038/s41467-026-69715-3. URL https://www.nature.com/articles/s41467-026-69715-3.

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

H. Cheng, L. Greengard, and V. Rokhlin. A Fast Adaptive Multipole Algorithm in Three Dimensions. *Journal of Computational Physics*, 155(2):468–498, November 1999. ISSN 00219991. doi: 10.1006/jcph.1999.6355. URL https://linkinghub.elsevier.com/retrieve/pii/S0021999199963556.

Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T. Unke, Adil Kabylda, Huziel E. Sauceda, Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate global machine learning force fields for molecules with hundreds of atoms. *Science Advances*, 9(2):eadf0873, January 2023. doi: 10.1126/sciadv.adf0873. URL https://www.science.org/doi/10.1126/sciadv.adf0873.

Alex Colagrande, Paul Caillon, Eva Feillet, and Alexandre Allauzen. Linear Attention with Global Context: A Multipole Attention Mechanism for Vision and Physics, July 2025. URL http://arxiv.org/abs/2507.02748. arXiv:2507.02748 [cs].

S. W. de Leeuw, J. W. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 373(1752):27–56, October 1980. ISSN 0080-4630. doi: 10.1098/rspa.1980.0135. URL https://doi.org/10.1098/rspa.1980.0135.

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Lio', and Michael Bronstein. On Over-Squashing in Message Passing Neural Networks: The Impact of Width, Depth, and Topology, May 2023. URL http://arxiv.org/abs/2302.02941.

Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B*, 99(1):014104, January 2019. doi: 10.1103/PhysRevB.99.014104. URL https://link.aps.org/doi/10.1103/PhysRevB.99.014104.

Geneviève Dusson, Markus Bachmayr, Gábor Csányi, Ralf Drautz, Simon Etter, Cas van der Oord, and Christoph Ortner. Atomic cluster expansion: Completeness, efficiency and stability. *Journal of Computational Physics*, 454:110946, April 2022. ISSN 0021-9991. doi: 10.1016/j.jcp.2022.110946. URL https://www.sciencedirect.com/science/article/pii/S0021999122000080.

Carolin Faller, Merzuk Kaltak, and Georg Kresse. Density-Based Long-Range Electrostatic Descriptors for Machine Learning Force Fields, October 2024. URL http://arxiv.org/abs/2406.17595. arXiv:2406.17595 [cond-mat].

Emilio McAllister Fognini, Marta M. Betcke, and Ben T. Cox. Learning Greens Operators through Hierarchical Neural Networks Inspired by the Fast Multipole Method, September 2025. URL http://arxiv.org/abs/2509.20591. arXiv:2509.20591 [cs].

Emilio McAllister Fognini, Marta M. Betcke, and Ben T. Cox. CloudNFMM: A hybrid hierarchical and local neural operator inspired by the fast multipole method. In *5th Annual AAAI Workshop on AI to Accelerate Science and Engineering*, 2026. URL https://ai-2-ase.github.io/papers/CameraReadys%203-41/26/CameraReady/26.pdf.

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

J. Thorben Frank, Stefan Chmiela, Klaus-Robert Müller, and Oliver T. Unke. Machine learning global atomic representations with Euclidean fast attention. *Nature Machine Intelligence*, 8 (3):388–402, March 2026. ISSN 2522-5839. doi: 10.1038/s42256-026-01195-y. URL https://www.nature.com/articles/s42256-026-01195-y.

Paul Fuchs, Michał Sanocki, and Julija Zavadlav. Learning non-local molecular interactions via equivariant local representations and charge equilibration. *npj Computational Materials*, 11(1): 287, September 2025. ISSN 2057-3960. doi: 10.1038/s41524-025-01790-4. URL https://www.nature.com/articles/s41524-025-01790-4.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations (ICLR)*, 2020.

Mario Geiger and Tess Smidt. e3nn: Euclidean Neural Networks, July 2022. URL http://arxiv.org/abs/2207.09453. arXiv:2207.09453 [cs].

S. Alireza Ghasemi, Albert Hofstetter, Santanu Saha, and Stefan Goedecker. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Physical Review B*, 92(4):045131, July 2015. doi: 10.1103/PhysRevB.92.045131. URL https://link.aps.org/doi/10.1103/PhysRevB.92.045131.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for Quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1263–1272, Sydney, NSW, Australia, August 2017. JMLR.org.

Stefan Goedecker, Michael Teter, and Jürg Hutter. Separable dual-space Gaussian pseudopotentials. *Physical Review B*, 54(3):1703–1710, 1996. doi: 10.1103/PhysRevB.54.1703. URL https://doi.org/10.1103/PhysRevB.54.1703.

L Greengard and V Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73(2):325–348, December 1987. ISSN 0021-9991. doi: 10.1016/0021-9991(87)90140-9. URL https://www.sciencedirect.com/science/article/pii/0021999187901409.

Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. 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.

Andrea Grisafi and Michele Ceriotti. Incorporating long-range physics in atomic-scale machine learning. *The Journal of Chemical Physics*, 151(20):204105, November 2019. ISSN 0021-9606. doi: 10.1063/1.5128375. URL https://pubs.aip.org/aip/jcp/article/151/20/204105/197247/Incorporating-long-range-physics-in-atomic-scale.

Moritz Gubler, Jonas A. Finkler, Moritz R. Schäfer, Jörg Behler, and Stefan Goedecker. Accelerating Fourth-Generation Machine Learning Potentials Using Quasi-Linear Scaling Particle Mesh Charge Equilibration. *Journal of Chemical Theory and Computation*, 20(16):7264–7271, August 2024. ISSN 1549-9618. doi: 10.1021/acs.jctc.4c00334. URL https://doi.org/10.1021/acs.jctc.4c00334.

Zhiyue Guo, Junjie Wang, Haoting Zhang, Zhixin Liang, Ziyang Yang, Yujian Pan, and Jian Sun. Differentiable particle-mesh Ewald with Cartesian tensor message passing for learning long-range electrostatics and dipole response. *arXiv preprint arXiv:2606.01598*, June 2026. URL https://arxiv.org/abs/2606.01598.

Jiaqi Han, Wenbing Huang, Tingyang Xu, and Yu Rong. Equivariant Graph Hierarchy-Based Neural Networks, October 2022. URL http://arxiv.org/abs/2202.10643. arXiv:2202.10643 [cs].

Kevin K. Huguenin-Dumittan, Philip Loche, Ni Haoran, and Michele Ceriotti. Physics-Inspired Equivariant Descriptors of Nonbonded Interactions. *The Journal of Physical Chemistry Letters*, 14(43):9612–9618, November 2023. doi: 10.1021/acs.jpclett.3c02375. URL https://doi.org/10.1021/acs.jpclett.3c02375.

Truong Son Hy and Risi Kondor. Multiresolution equivariant graph variational autoencoder. *Machine Learning: Science and Technology*, 4(1):015031, March 2023. ISSN 2632-2153. doi: 10.1088/2632-2153/acc0d8. URL https://dx.doi.org/10.1088/2632-2153/acc0d8.

Nicholas E Jackson, Michael A Webb, and Juan J De Pablo. Recent advances in machine learning towards multiscale soft materials design. *Current Opinion in Chemical Engineering*, 23:106–114, March 2019. ISSN 22113398. doi: 10.1016/j.coche.2019.03.005. URL https://linkinghub.elsevier.com/retrieve/pii/S2211339819300036.

Yajie Ji, Jiuyang Liang, and Zhenli Xu. Machine-Learning Interatomic Potentials for Long-Range Systems. *Physical Review Letters*, 135(17):178001, October 2025. doi: 10.1103/ssp9-7s81. URL https://link.aps.org/doi/10.1103/ssp9-7s81.

Yajie Ji, Jiuyang Liang, and Zhenli Xu. Accurate learning of long-range interatomic potentials by coupling Cartesian atomic cluster expansion and sum-of-Gaussians neural networks. *The Journal of Chemical Physics*, 164(2):024109, January 2026. doi: 10.1063/5.0303312. URL https://doi.org/10.1063/5.0303312.

Adil Kabylda, J. Thorben Frank, Sergio Suarez Dou, Almaz Khabibrakhmanov, Leonardo Medrano Sandonas, Oliver T. Unke, Stefan Chmiela, Klaus-Robert Muller, and Alexandre Tkatchenko. Molecular Simulations with a Pretrained Neural Network and Universal Pairwise Force Fields, October 2024. URL https://doi.org/10.26434/chemrxiv-2024-bdfr0.

Yanming Kang, Giang Tran, and Hans De Sterck. Fast Multipole Attention: A Divide-and-Conquer Attention Mechanism for Long Sequences. *arXiv preprint arXiv:2310.11960*, October 2023. URL https://arxiv.org/abs/2310.11960v1.

Ehsan Rahmatizad Khajehpasha, Jonas A. Finkler, Thomas D. Kühne, and S. Alireza Ghasemi. CENT2: Improved charge equilibration via neural network technique. *Physical Review B*, 105(14): 144106, April 2022. doi: 10.1103/PhysRevB.105.144106. URL https://link.aps.org/doi/10.1103/PhysRevB.105.144106.

Dongjin Kim and Bingqing Cheng. Long-range electrostatics for machine learning interatomic potentials is easier than we thought. *The Journal of Chemical Physics*, 164(6):060901, February 2026. ISSN 0021-9606. doi: 10.1063/5.0316886. URL https://doi.org/10.1063/5.0316886.

Dongjin Kim, Xiaoyu Wang, Santiago Vargas, Peichen Zhong, Daniel S. King, Theo Jaffrelot Inizan, and Bingqing Cheng. A universal augmentation framework for long-range electrostatics in machine learning interatomic potentials. *Journal of Chemical Theory and Computation*, 21 (24):12709–12724, December 2025. ISSN 1549-9618. doi: 10.1021/acs.jctc.5c01400. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC12707604/.

John G. Kirkwood. The Dielectric Polarization of Polar Liquids. *The Journal of Chemical Physics*, 7(10):911–919, October 1939. ISSN 0021-9606. doi: 10.1063/1.1750343. URL https://doi.org/10.1063/1.1750343.

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

Emir Kocer, Andreas Singraber, Jonas A. Finkler, Philipp Misof, Tsz Wai Ko, Christoph Dellago, and Jörg Behler. Iterative charge equilibration for fourth-generation high-dimensional neural network potentials. *The Journal of Chemical Physics*, 162(12):124106, March 2025. ISSN 0021-9606, 1089-7690. doi: 10.1063/5.0252566. URL https://pubs.aip.org/jcp/article/162/12/124106/3340486/Iterative-charge-equilibration-for-fourth.

Risi Kondor. N-body Networks: a Covariant Hierarchical Neural Network Architecture for Learning Atomic Potentials, March 2018. URL http://arxiv.org/abs/1803.01588.

Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. Ewald-based Long-Range Message Passing for Molecular Graphs, June 2023. URL http://arxiv.org/abs/2303.04791. arXiv:2303.04791 [cond-mat, physics:physics].

Konstantin N Kudin and Gustavo E Scuseria. A fast multipole method for periodic systems with arbitrary unit cell geometries. *Chemical Physics Letters*, 283(1–2):61–68, January 1998. ISSN 0009-2614. doi: 10.1016/S0009-2614(97)01329-8. URL https://www.sciencedirect.com/science/article/pii/S0009261497013298.

Thomas D. Kühne, Marcella Iannuzzi, Mauro Del Ben, Vladimir V. Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Z. Khaliullin, Ole Schütt, Florian Schiffmann, et al. CP2K: An electronic structure and molecular dynamics software package – Quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics*, 152(19):194103, 2020. doi: 10.1063/5.0007045. URL https://doi.org/10.1063/5.0007045.

Yunyang Li, Yusong Wang, Lin Huang, Han Yang, Xinran Wei, Jia Zhang, Tong Wang, Zun Wang, Bin Shao, and Tie-Yan Liu. Long-Short-Range Message-Passing: A Physics-Informed Framework to Capture Non-Local Interaction for Scalable Molecular Dynamics Simulation, July 2024. URL http://arxiv.org/abs/2304.13542. arXiv:2304.13542 [physics].

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Multipole Graph Neural Operator for Parametric Partial Differential Equations, June 2020. URL https://arxiv.org/abs/2006.09535v2.

Jiuyang Liang, Libin Lu, Yajie Ji, and Shidong Jiang. Prolate spheroidal wave functions enable fast and exponent-aware long-range machine learning interatomic potentials. *arXiv preprint arXiv:2606.06617*, June 2026. URL https://arxiv.org/abs/2606.06617.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs, February 2023. URL http://arxiv.org/abs/2206.11990. arXiv:2206.11990 [cs].

Alec J. Linot, Haotian Hang, Eva Kanso, and Kunihiko Taira. Hierarchical equivariant graph neural networks for forecasting collective motion in vortex clusters and microswimmers. *Communications Physics*, 8:515, November 2025. doi: 10.1038/s42005-025-02417-2. URL https://www.nature.com/articles/s42005-025-02417-2.

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

Moin Uddin Maruf, Sungmin Kim, and Zeeshan Ahmad. Learning Long-Range Interactions in Equivariant Machine Learning Interatomic Potentials via Electronic Degrees of Freedom. *The Journal of Physical Chemistry Letters*, 16(35):9078–9087, September 2025. doi: 10.1021/acs.jpclett.5c02352. URL https://doi.org/10.1021/acs.jpclett.5c02352.

Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J. Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-36329-y. URL https://doi.org/10.1038/s41467-023-36329-y.

Nhat Khang Ngo, Truong Son Hy, and Risi Kondor. Multiresolution Graph Transformers and Wavelet Positional Encoding for Learning Long-Range and Hierarchical Structures. *The Journal of Chemical Physics*, 159(3):034109, July 2023. ISSN 0021-9606, 1089-7690. doi: 10.1063/5.0152833. URL http://arxiv.org/abs/2302.08647. arXiv:2302.08647 [cs].

Samuel P. Niblett, Mirza Galib, and David T. Limmer. Learning intermolecular forces at liquid–vapor interfaces. *The Journal of Chemical Physics*, 155(16):164101, October 2021. ISSN 0021-9606. doi: 10.1063/5.0067565. URL https://doi.org/10.1063/5.0067565.

Lars Onsager. Electric Moments of Molecules in Liquids. *Journal of the American Chemical Society*, 58(8):1486–1493, August 1936. ISSN 0002-7863. doi: 10.1021/ja01299a050. URL https://doi.org/10.1021/ja01299a050.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/1905.10947.

Yoonjae Park, Dongjin Kim, Daniel S. King, Nam H. Đào, Roya Savoj, Sebastien Hamel, Xiaoyu Wang, and Bingqing Cheng. Polarizable atomic multipoles for learning long-range electrostatics, September 2026. URL https://arxiv.org/abs/2605.05746v2. Version 2.

Isaac J. Parker, Mandy J. Hoffmann, William J. Baldwin, Shuang Han, Srishti Gupta, Kara D. Fong, Angelos Michaelides, Christoph Schran, Sandip De, and Gábor Csányi. False Metallization in Short-Ranged Machine Learned Interatomic Potentials, March 2026. URL http://arxiv.org/abs/2603.04228. arXiv:2603.04228 [physics].

Behnam Parsaeifard, Deb Sankar De, Jonas A. Finkler, and Stefan Goedecker. Fingerprint-Based Detection of Non-Local Effects in the Electronic Structure of a Simple Single Component Covalent System. *Condensed Matter*, 6(1):9, March 2021. ISSN 2410-3896. doi: 10.3390/condmat6010009. URL https://www.mdpi.com/2410-3896/6/1/9.

Thomas Plé, Louis Lagardère, and Jean-Philip Piquemal. Force-Field-Enhanced Neural Network Interactions: from Local Equivariant Embedding to Atom-in-Molecule properties and long-range effects. *Chemical Science*, 14(44):12554–12569, 2023. ISSN 2041-6520, 2041-6539. doi: 10.1039/D3SC02581K. URL http://arxiv.org/abs/2301.08734. arXiv:2301.08734 [physics].

Thomas Plé, Olivier Adjoua, Anouar Benali, Evgeny Posenitskiy, Corentin Villot, Louis Lagardère, and Jean-Philip Piquemal. A Foundation Model for Accurate Atomistic Simulations in Drug Design, May 2025. URL https://doi.org/10.26434/chemrxiv-2025-f1hgn-v3.

Jay W. Ponder, Chuanjie Wu, Pengyu Ren, Vijay S. Pande, John D. Chodera, Michael J. Schnieders, Imran Haque, David L. Mobley, Daniel S. Lambrecht, Robert A. Jr. DiStasio, Martin Head-Gordon, Gary N. I. Clark, Margaret E. Johnson, and Teresa Head-Gordon. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B*, 114(8):2549–2564, March 2010. ISSN 1520-6106. doi: 10.1021/jp910674d. URL https://doi.org/10.1021/jp910674d.

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

Eric Qu, Brandon M. Wood, Aditi S. Krishnapriyan, and Zachary W. Ulissi. A recipe for scalable attention-based MLIPs: unlocking long-range accuracy with all-to-all node attention, March 2026. URL http://arxiv.org/abs/2603.06567. arXiv:2603.06567 [cs].

Ladislav Rampášek and Guy Wolf. Hierarchical graph neural nets can capture long-range interactions, August 2021. URL http://arxiv.org/abs/2107.07432. arXiv:2107.07432 [cs].

Anthony K. Rappe and William A. III Goddard. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry*, 95(8):3358–3363, April 1991. ISSN 0022-3654. doi: 10.1021/j100161a070. URL https://doi.org/10.1021/j100161a070.

Egor Rumiantsev, Marcel F. Langer, Tulga-Erdene Sodjargal, Michele Ceriotti, and Philip Loche. Learning Long-Range Representations with Equivariant Messages. *Transactions on Machine Learning Research*, January 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=pZI9e4SW9P.

Kristof T. Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9377–9388. PMLR, July 2021. URL https://proceedings.mlr.press/v139/schutt21a.html.

Yusuf Shaidu, Franco Pellegrini, Emine Küçükbenli, Ruggero Lot, and Stefano de Gironcoli. Incorporating long-range electrostatics in neural network potentials via variational charge equilibration from shortsighted ingredients. *npj Computational Materials*, 10(1):47, March 2024. ISSN 2057-3960. doi: 10.1038/s41524-024-01225-6. URL https://www.nature.com/articles/s41524-024-01225-6.

J. S. Smith, O. Isayev, and A. E. Roitberg. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science*, 8(4):3192–3203, March 2017. ISSN 2041-6539. doi: 10.1039/C6SC05720A. URL https://pubs.rsc.org/en/content/articlelanding/2017/sc/c6sc05720a.

Àlex Solé, Sergio Suárez-Dou, Albert Mosella-Montoro, Silvia Gómez-Coca, Eliseo Ruiz, Alexandre Tkatchenko, and Javier Ruiz-Hidalgo. Machine learning multiscale interactions, May 2026. URL https://arxiv.org/abs/2605.25710.

A.J. Stone. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chemical Physics Letters*, 83(2):233–239, October 1981. ISSN 00092614. doi: 10.1016/0009-2614(81)85452-8. URL https://linkinghub.elsevier.com/retrieve/pii/0009261481854528.

Anthony J. Stone. *The Theory of Intermolecular Forces*. Oxford University Press, Oxford, 2nd edition, 2013. ISBN 978-0-19-967239-4. doi: 10.1093/acprof:oso/9780199672394.001.0001. URL https://academic.oup.com/book/9446.

Jens Stücker and Oskar Foldal. JZ-FMM: GPU-native differentiable N-body simulations with the Fast Multipole Method, September 2026. URL https://arxiv.org/abs/2609.09307.

Daria Sushnikova, Pavel Kharyuk, and Ivan Oseledets. FMM-Net: neural network architecture based on the Fast Multipole Method, December 2022. URL http://arxiv.org/abs/2212.12899. arXiv:2212.12899 [math].

K. T. Tang and J. Peter Toennies. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *The Journal of Chemical Physics*, 80(8):3726–3741, April 1984. ISSN 0021-9606. doi: 10.1063/1.447150. URL https://doi.org/10.1063/1.447150.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, May 2018. URL http://arxiv.org/abs/1802.08219. arXiv:1802.08219 [cs].

Aidan P. Thompson, H. Metin Aktulga, Richard Berger, Dan S. Bolintineanu, W. Michael Brown, Paul S. Crozier, Pieter J. in 't Veld, Axel Kohlmeyer, Stan G. Moore, Trung Dac Nguyen, Ray Shan, Mark J. Stevens, Julien Tranchida, Christian Trott, and Steven J. Plimpton. LAMMPS – a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, 271:108171, February 2022. doi: 10.1016/j.cpc.2021.108171. URL https://doi.org/10.1016/j.cpc.2021.108171.

Moritz Thürlemann and Sereina Riniker. Anisotropic Message Passing: Graph Neural Networks with Directional and Long-Range Interactions. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=socffUzSIlx.

Alexandre Tkatchenko, Robert A. DiStasio, Roberto Car, and Matthias Scheffler. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Physical Review Letters*, 108(23): 236402, June 2012. doi: 10.1103/PhysRevLett.108.236402. URL https://link.aps.org/doi/10.1103/PhysRevLett.108.236402.

Nguyen Thien Phuc Tu, Nazanin Rezajooei, Erin R. Johnson, and Christopher N. Rowley. A neural network potential with rigorous treatment of long-range dispersion. *Digital Discovery*, 2(3): 718–727, 2023. doi: 10.1039/D2DD00150K. URL https://pubs.rsc.org/en/content/articlelanding/2023/dd/d2dd00150k.

Nguyen Thien Phuc Tu, Siri Williamson, Erin R. Johnson, and Christopher N. Rowley. Modeling Intermolecular Interactions with Exchange-Hole Dipole Moment Dispersion Corrections to Neural Network Potentials. *The Journal of Physical Chemistry B*, 128(35):8290–8302, September 2024. ISSN 1520-6106. doi: 10.1021/acs.jpcb.4c02882. URL https://doi.org/10.1021/acs.jpcb.4c02882.

Oliver T. Unke, Stefan Chmiela, Michael Gastegger, Kristof T. Schütt, Huziel E. Sauceda, and Klaus-Robert Müller. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications*, 12(1):7273, December 2021a. ISSN 2041-1723. doi: 10.1038/s41467-021-27504-0. URL https://www.nature.com/articles/s41467-021-27504-0.

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

Joost VandeVondele and Jürg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics*, 127(11):114105, 2007. doi: 10.1063/1.2770708. URL https://doi.org/10.1063/1.2770708.

Martin Vondrák, Karsten Reuter, and Johannes T. Margraf. Pushing charge equilibration-based machine learning potentials to their limits. *npj Computational Materials*, 11(1):288, September 2025. ISSN 2057-3960. doi: 10.1038/s41524-025-01791-3. URL https://www.nature.com/articles/s41524-025-01791-3.

Chu Wang, Lin Huang, Xinran Wei, Tao Qin, Arthur Jiang, Lixue Cheng, and Jia Zhang. Scalable machine learning force fields for macromolecular systems through long-range aware message passing, January 2026. URL https://arxiv.org/abs/2601.03774.

Tong Wang, Xinheng He, Mingyu Li, Yatao Li, Ran Bi, Yusong Wang, Chaoran Cheng, Xiangzhen Shen, Jiawei Meng, He Zhang, Haiguang Liu, Zun Wang, Shaoning Li, Bin Shao, and Tie-Yan Liu. Ab initio characterization of protein molecular dynamics with AI2BMD. *Nature*, 635(8040): 1019–1027, November 2024a. ISSN 1476-4687. doi: 10.1038/s41586-024-08127-z. URL https://doi.org/10.1038/s41586-024-08127-z.

Yusong Wang, Chaoran Cheng, Shaoning Li, Yuxuan Ren, Bin Shao, Ge Liu, Pheng-Ann Heng, and Nanning Zheng. Neural P^3M: A Long-Range Interaction Modeling Enhancer for Geometric GNNs, September 2024b. URL http://arxiv.org/abs/2409.17622. arXiv:2409.17622 [cs].

Yusong Wang, Tong Wang, Shaoning Li, Xinheng He, Mingyu Li, Zun Wang, Nanning Zheng, Bin Shao, and Tie-Yan Liu. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nature Communications*, 15(1):313, 2024c. ISSN 2041-1723. doi: 10.1038/s41467-023-43720-2. URL https://doi.org/10.1038/s41467-023-43720-2.

Brandon M. Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R. Kitchin, Daniel S. Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Ammar Rizvi, Sushree Jagriti Sahoo, Zachary W. Ulissi, and C. Lawrence Zitnick. UMA: A Family of Universal Models for Atoms, March 2026. URL http://arxiv.org/abs/2506.23971. arXiv:2506.23971 [cs].

Junjie Xu, Jiahao Zhang, Mangal Prakash, Xiang Zhang, and Suhang Wang. DualEquiNet: A Dual-Space Hierarchical Equivariant Network for Large Biomolecules, June 2025. URL http://arxiv.org/abs/2506.19862. arXiv:2506.19862 [q-bio].

Kun Yao, John E. Herr, David W. Toth, Ryker Mckintyre, and John Parkhill. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chemical Science*, 9 (8):2261–2269, February 2018. ISSN 2041-6539. doi: 10.1039/C7SC04934J. URL https://pubs.rsc.org/en/content/articlelanding/2018/sc/c7sc04934j.

Rex Ying, Jiaxuan You, Christopher Morris, Xiang Ren, William L. Hamilton, and Jure Leskovec. Hierarchical Graph Representation Learning with Differentiable Pooling, February 2019. URL http://arxiv.org/abs/1806.08804. arXiv:1806.08804 [cs].

Hongyu Yu, Liangliang Hong, Shiyou Chen, Xingao Gong, and Hongjun Xiang. Capturing long-range interaction with reciprocal space neural network, November 2022. URL http://arxiv.org/abs/2211.16684. arXiv:2211.16684 [cond-mat].

Shuwen Yue, Maria Carolina Muniz, Marcos F. Calegari Andrade, Linfeng Zhang, Roberto Car, and Athanassios Z. Panagiotopoulos. When do short-range atomistic machine-learning models fall short? *The Journal of Chemical Physics*, 154(3):034111, January 2021. ISSN 0021-9606. doi: 10.1063/5.0031215. URL https://doi.org/10.1063/5.0031215.

Linfeng Zhang, Han Wang, Maria Carolina Muniz, Athanassios Z. Panagiotopoulos, Roberto Car, and Weinan E. A deep potential model with long-range electrostatic interactions. *The Journal of Chemical Physics*, 156(12):124107, March 2022. doi: 10.1063/5.0083669. URL https://doi.org/10.1063/5.0083669.

Lingfeng Zhang, Taoyong Cui, Dongzhan Zhou, Lei Bai, Sufei Zhang, Luca Rossi, Mao Su, Wanli Ouyang, and Pheng-Ann Heng. An SO(3)-equivariant reciprocal-space neural potential for long-range interactions, March 2026. URL http://arxiv.org/abs/2603.18389. arXiv:2603.18389 [physics].

Yingkai Zhang and Weitao Yang. Comment on "Generalized gradient approximation made simple". *Physical Review Letters*, 80(4):890, 1998. doi: 10.1103/PhysRevLett.80.890. URL https://doi.org/10.1103/PhysRevLett.80.890.

Maksim Zhdanov, Max Welling, and Jan-Willem van de Meent. Erwin: A Tree-based Hierarchical Transformer for Large-scale Physical Systems, February 2025. URL http://arxiv.org/abs/2502.17019. arXiv:2502.17019 [cs].

Huan-Xiang Zhou and Xiaodong Pang. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chemical Reviews*, 118(4):1691–1741, February 2018. ISSN 0009-2665. doi: 10.1021/acs.chemrev.7b00305. URL https://pubs.acs.org/doi/10.1021/acs.chemrev.7b00305.

A Additional experimental results and protocols

A.1 Reporting conventions and benchmark runs

For run-level errors $x_1, \ldots, x_n$, we report the arithmetic mean and $\mathrm{SEM} = s/\sqrt{n}$, where s is the sample standard deviation with denominator $n - 1$. The means are averages of per-run metrics, not metrics computed after pooling predictions from different models. Energy errors are calculated per atom in each frame before aggregation. Force RMSE pools Cartesian force components within a run. A reported SEM smaller than the displayed precision is written as an upper bound rather than zero.

Lowest means are bold and second-lowest underlined (ties retained). Rankings compare means rather than statistical significance; SEMs, where reported, describe variability across runs. Comparisons are within each dataset and metric, with narrower groups identified in the captions. Published Ko results are ranked together with our runs, although their supervision and evaluation protocols differ.

Charge-equilibration and explicitly electrostatic references. The charge-supervised models (*) and SpookyNet in Table 4 differ in supervision, computational cost and training protocol (Fuchs et al., 2025; Ko et al., 2021; Unke et al., 2021a). *Labels.* CELLI and 4G–HDNNP fit their electronegativities to DFT Hirshfeld partial charges; CELLI's charge-loss weight is 50 on the carbon chain, 10 on NaCl and 200 on Au_2/MgO. SpookyNet takes the total charge and spin as inputs and fits partial charges to reference dipoles or charges, omitting its electrostatic term where neither is available. Every other model in that table learns from energies and forces alone; MACE–LES and NEMORA infer their latent charges or multipoles without supervision. The NaCl and Au_2/MgO releases of Ko et al. (2021) include per-atom charges, but the cumulene release does not, so for these models we computed Hirshfeld charges with ORCA 6.1.1 at the dataset's ωB97X-D3/def2-TZVP level. *Cost.* Both charge-equilibration methods obtain charges from a direct constrained linear solve, which both papers describe as cubic in the atom count, with iterative or particle-mesh alternatives discussed but not benchmarked; SpookyNet's electrostatics are an explicit all-pairs sum, with standard Ewald summation for periodic cells. *Protocol.* SpookyNet uses six message-passing layers against our two, SpookyNet's carbon-chain and Au_2/MgO results use a larger cutoff than its original, and the published evaluation splits are not stated to coincide with ours. The published Ko columns are therefore marked with a dagger; their highlighting does not establish a controlled comparison. Published CELLI and Allegro values follow Fuchs et al. (2025), and the 2G–BPNN, 4G–HDNNP and SpookyNet values follow Frank et al. (2026). On $Na_8Cl_8^+$/$Na_9Cl_8^+$ and Au_2/MgO, the two systems built to probe monopolar charge transfer, charge equilibration is the stronger energy corrector. The gap persists for the two-interaction CELLI–MACE, which matches our backbone depth, and is not explained by units or label precision; we therefore attribute it to the atom-resolved charge labels and prescribed Coulomb functional. NEMORA nonetheless has the lowest carbon-chain energy error and the lowest or second-lowest force error on every Ko system in Table 4. On cumulene, every reference model in Table 4 has higher force errors than short-range MACE on all three splits.

The Ko MACE and MACE–LES models use the JAX implementation, the other models use their original implementations, and all models trained in this work use the benchmark splits. The EFA Au_2/MgO results average two runs. Energy RMSE is computed for each run as $10^3[\frac{1}{M}\sum_{i=1}^{M}((\widehat{E}_i - E_i)/N_i)^2]^{1/2}$, where M is the number of evaluation frames, N_i is the atom count and energies are in eV, and is then averaged across runs; no constant energy offset is removed. The SpookyNet cumulene values are the published RMSEs (Batatia et al., 2024). One of the three EFA cumulene runs uses a shorter training budget, because the full-budget run with that initialization developed large extrapolation errors.

A.2 Backbone interfaces and protocol exceptions

For a native equivariant host, NEMORA receives the host's typed scalar and tensor features. For scalar-only hosts, a geometry-based lifting map combines scalar features with relative-position angular information to construct tensor sources. PET has no native irreducible-representation decomposition, so its attachment uses an explicit lift as well. The lift preserves equivariance when its scalar inputs are invariant; it does not impose exact equivariance on the entire PET backbone.

Table 4: Charge-equilibration and explicitly electrostatic references on the cumulene and charged-system (Ko) benchmarks, energy and force RMSEs, with our MACE and NEMORA runs from Table 1. Our cumulene test and Ko validation results are means across runs. †Published Ko results, which use study-specific splits, losses and architectures; all other values are our runs, except the published SpookyNet cumulene values. *Additional reference-charge supervision; SR disables charge-equilibration electrostatics. SpookyNet also takes total charge and spin as inputs. In each row, lowest means are bold and second-lowest underlined, separately for energy and force (ties retained), with published results ranked together with our runs. Sources, run counts, charge-label generation and protocol details are given in this appendix.

		CELLI–MACE (2)†*	Allegro†	CELLI–Allegro†*	CELLI–Allegro (J^R)†*	CELLI–Allegro (J^R, SR)*	2G–BPNN†	4G–HDNNP†*	SpookyNet†	MACE	**NEMORA (Ours)**
Energy RMSE (meV/atom)											
Cumulene	ID	58.791	—	—	59.431	59.588	42.815	41.892	30.800	16.898	**3.286**
	SOD	17.089	—	—	16.853	16.798	16.733	16.768	31.400	16.709	**1.877**
	OOD	11.202	—	—	11.133	11.116	12.070	11.317	23.400	11.994	**3.563**
Charged	Carbon chain	0.398	0.772	0.599	0.609	—	1.619	1.194	0.364	0.425	**0.175**
	NaCl	**0.097**	1.612	0.127	0.114	—	1.692	0.481	0.135	1.600	0.412
	Au_2/MgO	**0.069**	2.329	0.077	0.077	—	2.287	0.219	0.107	2.385	0.197
Force RMSE (meV/Å)											
Cumulene	ID	204.30	—	—	208.78	212.85	221.06	210.82	162.30	85.07	**28.18**
	SOD	126.99	—	—	135.93	145.78	174.54	161.90	116.20	102.16	**43.46**
	OOD	104.26	—	—	115.81	123.74	126.15	127.46	87.60	72.26	**51.55**
Charged	Carbon chain	21.45	49.18	31.00	32.31	—	129.50	78.00	5.80	15.00	**1.88**
	NaCl	3.54	47.51	6.44	5.15	—	57.39	32.78	**1.05**	43.89	1.75
	Au_2/MgO	7.95	123.67	12.04	12.01	—	153.10	66.00	5.34	110.07	**5.26**

Table 5: Numbers of runs contributing to the benchmark means and SEMs. Counts apply to energy and force unless noted. The separate 3–64-carbon sweep uses three runs for LOREM, EFA and NEMORA and one for the other models. NR denotes that no run count or no result is reported. Additional cumulene controls have three-run means but no available SEM in the source summary; SpookyNet is a published reference with no run count given there.

Model	Cumulene	Carbon chain	NaCl	Au_2/MgO
MACE	3	3	3	3
MACE–LES	3	3	3	3
MACE–RANGE	3	3	3	2
LOREM	3	3	3	3
EFA	3	3	3	2
MACE–MFN	3	3	3	NR
NEMORA	3	3	3	3
CELLI–MACE and short-range control	3	NR	NR	NR
CELLI–Allegro and short-range control	3	NR	NR	NR
4G–HDNNP and 2G–BPNN	3	NR	NR	NR

Table 6: Additional cumulene energy and force errors. Validation and aggregate-test results are mean ± SEM. The 3–64-carbon length/angle experiment uses separately trained models; LOREM, EFA and NEMORA have three seeds on that corpus and are quoted as mean ± SEM, while MACE, MACE–LES, MACE–RANGE and MACE–MFN have one and are point estimates. Bold and underlined entries mark the lowest and second-lowest errors. LOREM has the lowest force error on that corpus; the lowest energy error is the single-run MACE–MFN result.

	Energy MAE (meV/atom)			Force RMSE (meV/Å)		
Model	Validation	Test	$n_C = 3$–64	Validation	Test	$n_C = 3$–64
MACE (Batatia et al., 2023)	8.915 ± 0.153	11.833 ± 0.432	10.373	109.71 ± 0.75	86.41 ± 0.30	81.19
MACE–LES (Cheng, 2025)	8.671 ± 0.200	11.471 ± 0.166	10.323	98.16 ± 0.50	86.40 ± 0.42	80.73
MACE–RANGE (Caruso et al., 2026)	11.445 ± 0.822	14.184 ± 1.013	12.977	121.53 ± 2.90	101.66 ± 3.18	81.97
LOREM (Rumiantsev et al., 2026)	4.678 ± 1.198	16.765 ± 1.678	5.068 ± 2.159	60.60 ± 5.87	132.89 ± 19.68	$\mathbf{23.03 \pm 0.97}$
EFA (Frank et al., 2026)	4.890 ± 1.789	15.912 ± 1.267	4.083 ± 0.734	56.61 ± 8.80	113.03 ± 10.81	31.07 ± 0.87
MACE–MFN (Batatia et al., 2024)	$\mathbf{1.211 \pm 0.109}$	$\mathbf{0.970 \pm 0.297}$	**0.984**	$\mathbf{22.08 \pm 1.69}$	$\mathbf{14.81 \pm 0.82}$	31.86
NEMORA	1.627 ± 0.054	1.691 ± 0.475	1.202 ± 0.055	52.87 ± 2.87	44.55 ± 4.10	32.05 ± 0.87

Table 7: Carbon-chain backbone and source-degree validation ablations. Each entry gives mean ± SEM over three runs. Energy and force RMSEs are in meV/atom and meV/Å. Bold and underlining mark the lowest and second-lowest means across SR, scalar and degree-two NEMORA for each host and metric. Degree-two sources are lifted for scalar and untyped hosts. See Appendix A.2 for the evaluation protocol.

Backbone	Host	Metric	SR	$\ell = 0$	$\ell = 2$
MACE (Batatia et al., 2023)	Native	Energy	$\underline{0.425 \pm 0.006}$	0.484 ± 0.024	$\mathbf{0.163 \pm 0.015}$
		Force	$\underline{15.15 \pm 0.51}$	16.70 ± 0.46	$\mathbf{2.21 \pm 0.11}$
	Scalar	Energy	$\underline{0.623 \pm 0.006}$	0.889 ± 0.037	$\mathbf{0.284 \pm 0.031}$
		Force	$\underline{25.91 \pm 0.51}$	27.39 ± 0.48	$\mathbf{7.85 \pm 2.08}$
NequIP (Batzner et al., 2022)	Native	Energy	$\underline{0.871 \pm 0.002}$	0.972 ± 0.009	$\mathbf{0.182 \pm 0.006}$
		Force	42.38 ± 0.15	$\underline{32.16 \pm 0.12}$	$\mathbf{2.85 \pm 0.18}$
	Scalar	Energy	$\underline{2.618 \pm 0.001}$	9.491 ± 0.320	$\mathbf{0.448 \pm 0.158}$
		Force	93.43 ± 0.13	$\underline{47.21 \pm 1.23}$	$\mathbf{5.97 \pm 0.92}$
ViSNet (Wang et al., 2024c)	Native	Energy	0.673 ± 0.039	$\underline{0.670 \pm 0.009}$	$\mathbf{0.500 \pm 0.016}$
		Force	15.67 ± 1.12	$\underline{11.56 \pm 0.13}$	$\mathbf{3.49 \pm 0.12}$
	Scalar	Energy	$\underline{2.615 \pm 0.016}$	7.452 ± 0.742	$\mathbf{0.820 \pm 0.121}$
		Force	85.78 ± 0.09	$\underline{48.67 \pm 0.17}$	$\mathbf{6.35 \pm 0.35}$
PET (Pozdnyakov & Ceriotti, 2023)	Untyped	Energy	$\underline{1.047 \pm 0.086}$	1.119 ± 0.128	$\mathbf{0.944 \pm 0.019}$
		Force	$\underline{39.58 \pm 0.46}$	39.82 ± 0.58	$\mathbf{35.06 \pm 1.07}$

Table 8: Definitions of the component-ablation arms. All arms are retrained. Joint changes are identified explicitly.

Arm	Change relative to FULL
NOHIER	Remove the hierarchy, retaining the standalone Kelvin branch. Periodic root injection is also removed.
NOKELVIN	Remove the standalone Kelvin branch.
NOKESL	Remove the per-box Kelvin contribution to the downward pass.
FIXEDTREE	Replace adaptive coarsening by a fixed top-down octree.
MAXLEV1/3	Set maximum hierarchy depth to one or three.
LMAX0/1/3	Set source-feature and operator degree together. On cumulene, degrees below two also disable the quadrupole-dependent Kelvin head.
MAXL0/1	Reduce only the source-feature degree, retaining degree-two operators.
ELL3	Increase only the operator degree to three.
NU0/2/3	Change the many-body contraction order; zero removes the contraction.
M2LPIN	Pin the radial shape to $(1+u^2)^{-(\ell+1)/2}$, retaining a learned gain.
M2LDECAY	Learn the radial decay exponent.
NOVSQRT	Remove V-list square-root normalization.
TRES1	Reduce the resolvent truncation order from two to one.
NOFUSE	Remove the additive long-range feature update.
NOLRHEAD	Remove the separate long-range energy head.
NOGATE	Remove the multipole alignment gate.
LAD2/3	Remove many-body contraction and use scalar degree; LAD3 also removes the Kelvin branch.
FMMCLASSIC	Remove contraction, Kelvin branch, alignment gate and three normalization modes, and pin the radial shape. Learned source/readout maps remain.
NOHIERPBC	Remove periodic root injection.
NOKELVPBC	Remove the Kelvin periodic image sum.
NOTINFOIL	Remove the surface-dipole correction.
NSHELLS2	Increase the periodic image-shell radius from one to two.
SLABMASK	Mask replica images along the slab vacuum direction.
MICSPLIT	Use a minimum-image near/far split with two image shells.

Table 9: Complete available component-ablation validation force RMSEs (meV/Å), mean ± SEM over three runs. Each arm is retrained, and FULL is the corresponding retrained control. These controls differ from the benchmark configurations in Table 1. A dash denotes an unavailable result. Lowest means are bold and second-lowest underlined (ties retained).

Arm	Carbon chain	NaCl	Cumulene	Au_2/MgO
FULL	2.07 ± 0.05	1.89 ± 0.05	57.15 ± 6.53	$\underline{8.50 \pm 0.17}$
NU3	$\mathbf{1.93 \pm 0.08}$	—	50.31 ± 1.22	—
FMMCLASSIC	$\underline{1.95 \pm 0.03}$	2.78 ± 0.23	68.86 ± 3.48	$\mathbf{4.86 \pm 0.10}$
NOKELVIN	2.12 ± 0.04	1.92 ± 0.04	61.21 ± 9.06	10.28 ± 0.13
LMAX3	2.15 ± 0.02	—	—	—
M2LDECAY	2.15 ± 0.03	—	—	—
NOLRHEAD	2.18 ± 0.06	$\underline{1.86 \pm 0.02}$	71.15 ± 0.98	—
M2LPIN	2.21 ± 0.03	1.97 ± 0.07	67.19 ± 3.43	8.86 ± 0.03
ELL3	2.23 ± 0.13	—	—	—
NU0	2.26 ± 0.06	2.33 ± 0.02	62.66 ± 4.04	8.95 ± 0.09
MAXLEV1	2.64 ± 0.16	2.61 ± 0.18	$\underline{49.18 \pm 5.24}$	—
MAXLEV3	2.75 ± 0.24	—	—	—
FIXEDTREE	3.21 ± 0.09	3.56 ± 0.22	88.37 ± 12.24	8.58 ± 0.06
LMAX1	3.52 ± 0.33	3.36 ± 0.40	106.03 ± 3.31	8.71 ± 0.27
MAXL1	4.27 ± 0.54	2.51 ± 0.09	107.21 ± 2.56	—
NOFUSE	14.94 ± 0.52	1.89 ± 0.09	58.03 ± 8.18	—
NOHIER	15.01 ± 0.25	44.17 ± 0.21	89.41 ± 7.53	112.97 ± 0.37
LAD2	16.86 ± 0.50	—	—	—
LMAX0	17.00 ± 0.55	3.50 ± 0.22	107.74 ± 1.02	—
LAD3	17.30 ± 0.59	—	—	—
MAXL0	17.53 ± 0.95	—	—	—
NOKESL	2.32 ± 0.07	2.40 ± 0.02	64.76 ± 5.17	—
NOGATE	2.39 ± 0.18	1.88 ± 0.04	$\mathbf{47.67 \pm 9.72}$	—
NOVSQRT	—	$\mathbf{1.83 \pm 0.05}$	—	—
TRES1	—	1.99 ± 0.06	—	—
NU2	—	2.00 ± 0.07	—	—
NOHIERPBC	—	—	—	8.65 ± 0.13
NOKELVPBC	—	—	—	8.77 ± 0.10
SLABMASK	—	—	—	8.83 ± 0.36
NSHELLS2	—	—	—	8.87 ± 0.14
MICSPLIT	—	—	—	8.92 ± 0.23
NOTINFOIL	—	—	—	8.92 ± 0.25
MACE	15.00 ± 0.32	43.89 ± 0.09	105.43 ± 2.19	111.56 ± 1.65

Table 10: Cumulene component-ablation test force RMSEs (meV/Å), mean ± SEM from the per-run split evaluations. FULL is the three-run reference configuration. Lowest means are bold and second-lowest underlined (ties retained).

Arm	Test-ID	Test-SOD	Test-OOD
FIXEDTREE	67.52 ± 20.60	127.97 ± 29.75	104.35 ± 8.19
FMMCLASSIC	39.04 ± 2.03	151.67 ± 15.65	106.64 ± 2.53
FULL	$\underline{28.53 \pm 1.77}$	58.39 ± 9.79	64.12 ± 8.91
LMAX0	89.19 ± 1.73	104.11 ± 0.66	75.79 ± 0.46
LMAX1	88.39 ± 1.64	130.28 ± 7.88	80.10 ± 2.19
M2LPIN	33.27 ± 1.95	96.72 ± 36.88	77.94 ± 14.54
MACE	86.00 ± 0.40	102.07 ± 0.05	71.91 ± 0.08
MAXL1	89.75 ± 2.48	107.57 ± 1.98	81.72 ± 1.62
MAXLEV1	278.23 ± 48.46	183.77 ± 25.81	116.91 ± 10.55
NOFUSE	30.92 ± 2.85	72.28 ± 6.19	101.14 ± 40.98
NOGATE	$\mathbf{26.45 \pm 6.12}$	$\underline{55.77 \pm 3.62}$	$\underline{62.83 \pm 1.40}$
NOHIER	77.23 ± 6.64	151.20 ± 28.10	121.50 ± 29.02
NOKELVIN	29.45 ± 2.30	64.11 ± 15.40	70.20 ± 9.44
NOKESL	33.18 ± 0.38	$\mathbf{44.96 \pm 3.06}$	$\mathbf{54.27 \pm 2.60}$
NOLRHEAD	29.23 ± 2.67	87.29 ± 35.32	85.92 ± 13.58
NU0	33.08 ± 4.32	77.38 ± 22.92	64.77 ± 2.23
NU3	30.10 ± 2.10	62.60 ± 9.87	70.03 ± 1.32

A.3 Component definitions and complete ablations

We retrain NEMORA with individual components removed or modified on the carbon chain, NaCl, cumulene and Au_2/MgO. Tables 9 and 10 report the complete results. The hierarchy accounts for most of the improvement on the carbon chain, NaCl and Au_2/MgO. Removing it returns force errors close to those of short-range MACE. On cumulene, the Kelvin branch alone retains some improvement, but the hierarchy is needed for the lowest errors.

The angular-degree ablations agree with the backbone experiments. Restricting NEMORA to scalar features and operators increases carbon-chain force error from 2.07 to 17.00 meV/Å, while on NaCl the scalar model still reaches 3.50 meV/Å, well below the short-range error of 43.89 meV/Å. Cumulene is especially sensitive to reducing the feature degree, with degree-zero and degree-one variants returning close to the short-range model's error. The joint feature/operator ablations also remove a quadrupole-dependent Kelvin head, as specified in Table 8.

The remaining components have more system-dependent effects. Removing the standalone Kelvin branch has little effect on the two small cluster systems, but increases Au_2/MgO force error from 8.50 to 10.28 meV/Å. Removing the per-box Kelvin contribution increases the NaCl error from 1.89 to 2.40 meV/Å. A reduced configuration with pinned radial functions, no many-body contraction and no alignment gate improves Au_2/MgO force error to 4.86 meV/Å, but performs worse on the cumulene test splits. This FMMCLASSIC configuration retains learned source and energy readouts. Its performance suggests that simpler transport is sufficient for some predominantly electrostatic systems, whereas cumulene benefits from the additional flexibility of the learned operators.

Each ablation is compared with its corresponding FULL control. Several smaller effects are comparable to the variation across runs, particularly for cumulene. Validation and test behavior also differ for MAXLEV1, which has lower validation error but substantially larger ID test error.

A.4 MD22 protocol

Each model is trained with three seeds, a 4 Å cutoff, 400 epochs and SWA starting at epoch 200. NEMORA uses an occupancy-based hierarchy with a maximum leaf occupancy of 16. The numbers of training/validation/test configurations are 6000/750/6000 for Ac–Ala3–NHMe, 8000/1000/8000 for DHA and stachyose, 3000/375/3000 for AT–AT, 2000/250/2000 for AT–AT–CG–CG, 600/75/600 for the buckyball catcher, and 800/100/800 for the double-walled nanotube. Batch sizes are 8, 6, 4, 6, 2, 2 and 2, respectively.

A.5 Elongated-water molecular dynamics and dipole correlations

The simulations in Figure 3 use the MACE, MACE–LES and NEMORA models of Table 3, which share the MACE backbone described in Appendix E. Each simulation contains 2048 water molecules (6144 atoms) in an orthorhombic periodic cell of $24.84 \times 24.84 \times 99.35$ Å. All models start from the same equilibrated configuration and are run in the NVT ensemble at 300 K with Langevin dynamics, a 0.5 fs timestep and deuterium masses for hydrogen, with frames saved every 50 fs.

Molecular dipoles $\boldsymbol{\mu}_i$ are computed from Wannier-centroid positions predicted by the Deep Wannier model of the DPLR framework, trained on maximally localized Wannier centers of revPBE-D3 water (Zhang et al., 2022), with point charges $+6$ on O, $+1$ on H and -8 on the centroid. For wavevectors along the long axis, $\boldsymbol{k} = k\hat{\boldsymbol{z}}$ with $k = 2\pi n/L_z$ and $n = 1, 2, \ldots$ up to $k = 10$ Å^{-1}, we compute

$$g_L(k) = \frac{\langle |\hat{m}_z(k)|^2 \rangle}{N \langle \mu^2 \rangle}, \qquad \hat{m}_z(k) = \sum_{i=1}^{N} \mu_{i,z}\, e^{ikz_i}, \tag{8}$$

where z_i is the oxygen coordinate and N the number of molecules. The elongated cell gives a smallest wavevector of 0.063 Å^{-1}. We discard the first 5 ps and average over frames separated by 100 fs, giving more than 2000 frames per model. As $k \to 0$, $g_L(k)$ approaches a finite value set by the static dielectric constant (Cheng, 2025); a model without long-range electrostatics instead shows a spurious rise.

A.6 SCALING PROTOCOL AND COMPLETE MEASUREMENTS

All scaling models use two MACE interactions, correlation order three, a 4 Å cutoff and $128 \times 0\mathrm{e} + 128 \times 1\mathrm{o}$ hidden irreps. Models are built with untrained weights. Configurations are fixed-density argon clusters and cubic periodic boxes. Timing includes energy evaluation and coordinate differentiation for forces. CUDA events are used for PyTorch and synchronized device timing for JAX. Peak allocated memory is measured by the corresponding framework allocator. The reported scaling does not include dataset preparation or compilation.

Long-range exponents are log–log fits after subtraction of the matched short-range cost. PyTorch models subtract the common MACE measurement; JAX models use paired evaluations with the long-range module enabled and disabled. This removes much of the framework-specific fixed overhead from the slope estimate without making absolute timings or allocator statistics directly comparable across frameworks. The finite-range exponents characterize these implementations and do not replace asymptotic analysis.

Table 11: Forward-and-force wall time (ms) for non-periodic clusters. Single H200, fixed-density argon, ten frames per size. α_{LR} is fitted after subtracting the backbone cost. Appendix A.6 gives the timing protocol and framework details. A dash denotes a size not reached within available GPU memory. Bold and underlining mark the lowest and second-lowest measurements at each size across the listed implementations, including the short-range baseline. Frameworks differ; these marks rank measured cost, not matched accuracy.

N	MACE	LES	RANGE	EFA	LOREM	MFN	NEMORA
64	**1.2**	12.0	24.0	1.5	2.9	125.0	3.2
128	**1.4**	15.1	28.6	2.0	3.4	635.0	3.5
256	**1.4**	15.2	29.0	3.0	3.9	4600.0	4.7
512	**1.5**	15.5	31.4	4.7	6.0	35100.0	5.2
1024	**1.9**	15.7	38.1	8.3	9.2	—	7.6
2048	**2.3**	20.7	57.4	15.2	19.0	—	9.2
4096	**3.6**	40.3	103.0	35.2	60.7	—	14.1
8192	**6.0**	96.4	194.0	75.1	290.0	—	20.1
16384	**10.8**	274.0	382.0	173.0	1060.0	—	35.7
32768	**20.9**	—	833.0	—	—	—	69.1
65536	**40.8**	—	—	—	—	—	129.0
α_{LR}	—	1.96	1.17	1.00	1.99	2.93	0.86

Table 12: Forward-and-force wall time (ms) under periodic boundaries. Single H200, fixed-density argon, ten frames per size. α_{LR} is fitted after subtracting the backbone cost. Appendix A.6 gives the timing protocol and framework details. A dash denotes a size not reached within available GPU memory. Bold and underlining mark the lowest and second-lowest measurements at each size across the listed implementations, including the short-range baseline. Frameworks differ; these marks rank measured cost, not matched accuracy.

N	MACE	LES	RANGE	EFA	LOREM	MFN	NEMORA
64	**1.2**	12.9	25.4	**1.2**	3.0	121.0	3.8
128	**1.1**	16.0	30.4	1.6	3.5	636.0	4.5
256	**1.4**	16.2	31.6	2.1	4.0	4610.0	5.6
512	**1.5**	16.4	34.6	2.8	5.8	35200.0	6.2
1024	**1.8**	16.7	44.7	4.2	11.2	—	7.2
2048	**2.2**	21.2	70.8	7.1	27.1	—	9.9
4096	**3.6**	40.2	127.0	13.1	77.2	—	13.7
8192	**6.2**	86.8	241.0	28.5	264.0	—	20.0
16384	**11.1**	231.0	477.0	83.8	—	—	37.1
32768	**21.3**	—	1030.0	171.0	—	—	68.7
65536	**41.4**	—	—	—	—	—	125.0
α_{LR}	—	1.93	1.14	1.05	1.97	2.93	0.82

Table 13: Peak allocated GPU memory (MiB) for non-periodic clusters. Single H200, fixed-density argon, ten frames per size. α_{LR} is fitted after subtracting the backbone cost. Appendix A.6 gives the timing protocol and framework details. A dash denotes a size not reached within available GPU memory. The EFA value at $N = 16{,}384$ repeats the $N = 8192$ measurement and was not re-measured. Bold and underlining mark the lowest and second-lowest measurements at each size across the listed implementations, including the short-range baseline. Frameworks differ; these marks rank measured cost, not matched accuracy.

N	MACE	LES	RANGE	EFA	LOREM	MFN	NEMORA
64	**10**	27	213	644	70	1580	67
128	66	**51**	371	1240	266	6230	67
256	**66**	101	689	2210	268	24700	132
512	**131**	199	1320	4180	526	98600	261
1024	**132**	439	2600	8340	842	—	263
2048	**263**	1070	5140	16600	1630	—	522
4096	**368**	2930	10200	33200	3220	—	1040
8192	**763**	8990	20400	66300	6530	—	1950
16384	**1540**	30500	40800	66300	13600	—	3810
32768	**3170**	—	81600	—	—	—	8000
65536	**6470**	—	—	—	—	—	15500
α_{LR}	—	2.10	0.99	0.98	2.01	2.00	0.91

Table 14: Peak allocated GPU memory (MiB) under periodic boundaries. Single H200, fixed-density argon, ten frames per size. α_{LR} is fitted after subtracting the backbone cost. Appendix A.6 gives the timing protocol and framework details. A dash denotes a size not reached within available GPU memory. Bold and underlining mark the lowest and second-lowest measurements at each size across the listed implementations, including the short-range baseline. Frameworks differ; these marks rank measured cost, not matched accuracy.

N	MACE	LES	RANGE	EFA	LOREM	MFN	NEMORA
64	66	**33**	214	134	134	1590	134
128	66	**63**	373	271	264	6240	134
256	**67**	118	691	532	266	24800	134
512	**131**	250	1330	905	523	98600	261
1024	**133**	624	2600	1860	1100	—	263
2048	**264**	1760	5150	3590	3490	—	525
4096	**526**	5600	10200	7260	13100	—	1040
8192	**825**	19600	20400	14700	51800	—	2080
16384	**1640**	72800	40800	29600	—	—	4130
32768	**3310**	—	81600	59900	—	—	8270
65536	**6710**	—	—	—	—	—	15700
α_{LR}	—	2.04	1.00	1.04	2.20	2.00	0.87

B NOTATION

Table 15: Symbol glossary used throughout the paper.

Symbol	Meaning
Atomic system / backbone	
$\boldsymbol{r}_i, Z_i, N$	atomic position, species, atom count
r_{cut}	short-range backbone cutoff radius
$\boldsymbol{h}_i^{(\ell_Q, t^\star)}$	backbone equivariant feature at atom i, irrep degree ℓ_Q, GNN iteration $t^\star$
T	total number of MACE backbone iterations
$t^\star$	GNN iteration index, $t^\star \in \{1, \dots, T\}$ (asterisk distinguishes from FMM target t)
Classical FMM	
$\boldsymbol{c}_t, \boldsymbol{c}_s$	target / source expansion centers
$\boldsymbol{r}_> = \boldsymbol{c}_t - \boldsymbol{c}_s$	separation of the source and target centers; the irregular argument
$\boldsymbol{r}_< = \boldsymbol{r}_j - \boldsymbol{c}_s$	offset of a particle from its source center; the regular argument
R_s, R_t	source / target cluster bounding radii
p	multipole expansion order (truncation degree); $p = \ell_{\max,\text{N}}$ in the learned models
ℓ, m	harmonic degree and magnetic index of an expansion term
Φ_s, Φ_t	source moments / target local coefficients; $\Phi_s = q_j$ at the first stage (Section 2.2)
$\tau, \boldsymbol{r}_{st}$	learned translation stage and its displacement; a regular stage sets $\boldsymbol{r}_< = \boldsymbol{r}_{st}$, M2L sets $\boldsymbol{r}_> = \boldsymbol{r}_{st}$ (Equation 7)
ℓ_1, c_{ℓ_1}	intermediate degree of the tensor moments and its fixed recoupling weight (Equation 16)
ζ_ℓ	fixed constants of the split-degree multipole expansion (Appendix I)
$\tilde{\boldsymbol{R}}^{(\ell)}, \tilde{\boldsymbol{I}}^{(\ell)}$	Racah-normalized regular / irregular solid harmonics (Appendix I)
$Y_{\ell m}$	orthonormal spherical harmonic; complex Condon–Shortley basis in Equation 3 and the appendix
$\theta, \rho_{\text{near}}$	far-field MAC threshold and near-field admissibility multiplier
Equivariant / NEMORA	
ℓ_Q, ℓ_M, ℓ_0	backbone irrep degree, NEMORA internal irrep degree, scalar irrep ($\ell_0 \equiv 0$)
$\ell_{\max,Q}, \ell_{\max,M}$	maximum backbone / NEMORA irrep degree
ν	ACE / SymContract correlation order
$\boldsymbol{Q}_i^{(\ell_Q)}$	equivariant tensor charge at atom i ($\boldsymbol{Q}_i^{(\ell_Q)} \equiv \boldsymbol{W}_{\text{in}} \boldsymbol{h}_i^{(\ell_Q, t^\star)}$)
$\boldsymbol{\Psi}_t^{(\ell_M, L)}$	NEMORA multipole feature at hierarchy node t, level L, irrep ℓ_M
$\boldsymbol{W}_{\text{in}}^{(\ell)}, \boldsymbol{W}_{\text{out}}^{(\ell)}$	degree-preserving input / output channel projections
$\mathcal{R}_{t^\star}$	invariant backbone energy readout at iteration $t^\star$
$\boldsymbol{W}^V$	per-stage equivariant linear channel mixer
$\otimes_\theta, \otimes$	learned-path / fixed Clebsch–Gordan tensor product
$\boldsymbol{Y}^{(\ell)}, \boldsymbol{R}$	real spherical harmonic block and sinc–Bessel radial basis
$\boldsymbol{R}_\theta^{(\ell)}$	learned radial and channel weights of the backbone message (Equation 5)
$\boldsymbol{m}_{st}$	per-source gated M2L message
$\boldsymbol{\rho}_\theta^{(\ell)}, \boldsymbol{\rho}_{\theta,\tau}^{(\ell)}$	invariant radial modulation of the degree-ℓ filter channels / its stage-τ instance
$\boldsymbol{\phi}_\theta$	invariant gate MLP feeding $\alpha_{\theta,st}^{(\ell_M)}$
$\boldsymbol{T}_{st}^{(\ell_M)}$	bare (un-gated) translation block (Appendix Eq. 19)
$\boldsymbol{\alpha}_{st}^{\tau}, \alpha_{\theta,st}^{(\ell_M)}$	bundled invariant stage gate / an output-irrep gate block
$\alpha_{\text{agg}}, \alpha_{\text{agg}}^{\text{up}}$	density-mean fusion / extensive upward aggregation parameters
$\mathcal{G}_{st}, \mathcal{M}_{st}, \mathcal{A}_{st}$	geometric, magnitude, and alignment invariants entering the gate

$\boldsymbol{q}_t^{(\ell)}, \boldsymbol{k}_s^{(\ell)}, \boldsymbol{W}_Q, \boldsymbol{W}_K$	query / key projections feeding $\mathcal{A}_{st}$
$\boldsymbol{\Phi}_s, \boldsymbol{\Phi}_t^{(\ell_M)}$	tensor moments / tensor local coefficients; $\boldsymbol{\Phi}_s = \boldsymbol{Q}_j^{(\ell_Q)}$ at the first stage (Equation 4)
Hierarchy / Kelvin / PBC	
Λ, L	set of hierarchy levels, generic level index
$L_{\max}$	hierarchy depth cap
$\mathcal{N}_t, \mathcal{I}_L(B)$	interaction list at target t / level-L M2L list of cluster B
λ_k	PCA eigenvalue along axis k; axis is active when $\lambda_k/\lambda_{\max} >$ threshold
$n_{\text{active}}, n_{\text{SR}}$	PCA active-axis count (fanout $2^{n_{\text{active}}}$) and SR-receptive-field multiplier
$\text{Near}_L(B), \rho_{\text{near}}$	near-set of cluster B at level L and admissibility multiplier
$d_{\min}$	cluster-gap distance $\max(\|\boldsymbol{c}_A - \boldsymbol{c}_B\|_{\text{MIC}} - R_A - R_B, 0)$
$\boldsymbol{c}_k, R_k, \tilde{r}_k, \boldsymbol{r}_{sk}, K$	Kelvin anchor center, radius, normalized radius R_k/r_{cut}, atom-to-anchor displacement, anchor count
$\boldsymbol{P}(\cdot)$	polynomial radial encoder for the Kelvin branch
$\kappa_\ell(R_k)$	Kelvin inversion scalar ($\tilde{r}_k^{-(2\ell+1)}$ with optional learned modulation)
$\boldsymbol{H}, \boldsymbol{a}_\alpha, V$	periodic cell matrix, lattice basis vectors, cell volume
$\boldsymbol{R}_{\boldsymbol{n}}, \mathcal{S}_{\text{lat}}^{(\ell_M)}, n_{\text{far}}, \gamma_{\text{pbc}}$	lattice displacement, finite-shell Madelung sum, image-shell radius, lattice-sum strength

C LONG-RANGE MODELS

Table 16: Long-range neural methods along six architectural axes. *Equiv.*: $\ell > 0$ tensorial long-range messages. *Learn.*: learnable propagation parameters, excluding source construction and readout. *Many-body*: pairwise (○), single-mode many-body via aggregation/path (◐), or both atom-level and coarse-level long-range coupling (●). *ℓ-mix*: angular channel mixing during long-range transport. *Hier.*: hierarchical multi-scale structure. *Scaling*: time cost of the long-range component at fixed model settings; linear costs require bounded interaction lists or graph degree. Symbols describe the long-range mechanism, not the total body order of the complete potential. ○ = no, ◐ = partial/approximate, ● = yes.

Method	Equiv.	Learn.	MB	ℓ-mix	Hier.	Scaling
Scalar long-range corrections						
SpookyNet (Unke et al., 2021a), TensorMol (Yao et al., 2018), AIMNet (Anstine & Isayev, 2023)	○	○	○	○	○	$\mathcal{O}(N^2)$
SO3LR (Kabylda et al., 2024), FENNOL (Plé et al., 2023)	○	○	○	○	○	$\mathcal{O}(N \cdot \langle N_{\mathrm{LR}} \rangle)$
4G–HDNNP (Ko et al., 2021), CENT/CENT2 (Ghasemi et al., 2015; Khajehpasha et al., 2022), CELLI (Fuchs et al., 2025), FeNNix-Bio1 (Plé et al., 2025), Maruf et al. (2025)	○	○	○	○	○	$\mathcal{O}(N \log N)$ †
LES (Cheng, 2025; Kim et al., 2025), Neural P^{3}M (Wang et al., 2024b), Ewald MP (Kosmala et al., 2023), Reciprocal Space NN (Yu et al., 2022)	○	○	○	○	○	$\mathcal{O}(N \log N)$
LODE (Grisafi & Ceriotti, 2019; Huguenin-Dumittan et al., 2023; Faller et al., 2024),	○	○	◐	○	○	$\mathcal{O}(N \log N)$
sGDML (Chmiela et al., 2023)	○	●	●	○	○	$\mathcal{O}(N^2)$
SOG-Net (Ji et al., 2025)	○	●	○	○	○	$\mathcal{O}(N \log N)$
MLXDM (Tu et al., 2023; 2024)	○	○	○	○	○	$\mathcal{O}(N \cdot \langle N_{\mathrm{LR}} \rangle)$
Equivariant long-range corrections						
LOREM (Rumiantsev et al., 2026)	●	○	◐	○	○	$\mathcal{O}(N \log N)$
EquiEwald (Zhang et al., 2026)	●	◐	○	○	○	$\mathcal{O}(N)$
FieldMACE (Barrett et al., 2026)	●	○	◐	○	○	$\mathcal{O}(N + N_{\mathrm{env}})$ ‡
MACE-POLAR-1 (Batatia et al., 2026)	●	○	◐	○	○	$\mathcal{O}(N \log N)$
AMP (Thürlemann & Riniker, 2023)	●	○	○	○	○	$\mathcal{O}(N^2)$
LSR-MP (Li et al., 2024)	●	●	◐	○	○	$\mathcal{O}(N \cdot N_{\mathrm{frag}})$
MFN (Batatia et al., 2024)	●	●	◐	◐	○	$\mathcal{O}(N^3)$ §
Hierarchical graph neural networks						
DiffPool (Ying et al., 2019), HGNet (Rampášek & Wolf, 2021), MGT (Ngo et al., 2023)	○	●	◐	○	●	$\mathcal{O}(N)$
EGHN (Han et al., 2022), DualEquiNet (Xu et al., 2025), Hy & Kondor (2023)	●	●	◐	○	●	$\mathcal{O}(N)$
Global attention						
RANGE (Caruso et al., 2026)	○	●	◐	○	○	$\mathcal{O}(N)$
EFA (Frank et al., 2026)	●	●	◐	●	○	$\mathcal{O}(N)$
AllScAIP (Qu et al., 2026)	◐	●	◐	○	○	$\mathcal{O}(N^2)$
UMA (Wood et al., 2026)	◐	●	◐	○	○	$\mathcal{O}(N^2)$
Erwin (Zhdanov et al., 2025)	○	●	◐	○	●	$\mathcal{O}(N)$
FMM-inspired neural architectures						
MGNO (Li et al., 2020), FMM-Net (Sushnikova et al., 2022)	○	●	●	○	●	$\mathcal{O}(N)$
MANO (Colagrande et al., 2025), FMA (Kang et al., 2023)	○	●	●	○	●	$\mathcal{O}(N)$
Neural FMM (Fognini et al., 2025)	○	●	●	○	●	Unverified
NEMORA (ours)	●	●	●	●	●	$\mathcal{O}(N)$

† Original 4G–HDNNP scales as $\mathcal{O}(N^3)$; reducible to $\mathcal{O}(N \log N)$ via PM-Qeq (Gubler et al., 2024), iQEq (Kocer et al., 2025), or variational Qeq (Shaidu et al., 2024); scaling limits in (Vondrák et al., 2025; Kim & Cheng, 2026). ‡ Global multipole embedding; direct per-atom embedding costs $\mathcal{O}(N N_{\mathrm{env}})$. § Dense matrix-function evaluation; approximations can change scaling. Neural FMM does not report end-to-end size-scaling measurements.

D Periodic Boundary Correction

Periodic cells of volume $V = |\det \boldsymbol{H}|$ require lattice-summed interactions for analytical periodic FMM (Kudin & Scuseria, 1998), with surface terms depending on the electrostatic boundary convention (de Leeuw et al., 1980). The learned NEMORA branch aggregates a cell multipole $\boldsymbol{\Psi}_{\square}^{(\ell_M)}$ from the top-level roots and uses the following finite-shell, degree-diagonal correction:

$$\tilde{\boldsymbol{\Psi}}_{\square}^{(\ell_M)} = \Big[(1-\delta_{\ell_M,0})\mathcal{S}_{\mathrm{lat}}^{(\ell_M)} + \delta_{\ell_M,1}\frac{4\pi r_{\mathrm{cut}}^3}{3V}\Big]\boldsymbol{\Psi}_{\square}^{(\ell_M)}, \qquad \mathcal{S}_{\mathrm{lat}}^{(\ell_M)} = \gamma_{\mathrm{pbc}} \sum_{\boldsymbol{n}\in\mathcal{N}_{\mathrm{lat}}} \left(\frac{r_{\mathrm{cut}}}{\|\boldsymbol{R}_{\boldsymbol{n}}\|}\right)^{2\ell_M+1} \tag{9}$$

with truncated image shell $\mathcal{N}_{\mathrm{lat}} = \{\boldsymbol{n} \in \mathbb{Z}^3 : \|\boldsymbol{n}\|_\infty \le n_{\mathrm{far}}\} \setminus \{\boldsymbol{0}\}$ and lattice displacements $\boldsymbol{R}_{\boldsymbol{n}} = \sum_\alpha n_\alpha \boldsymbol{a}_\alpha$. The deltas suppress the monopole contribution to this correction and add a volume-dependent dipole term; they do not enforce neutrality of the learned atomic sources. This modeled root response is not the full angular periodic multipole translation or an exact implementation of a general tinfoil lattice sum. Its explicit cell dependence can be differentiated for stress, and L2L reinjection costs $\mathcal{O}(N)$ at fixed image shell, truncation, and hierarchy assumptions. The analytical latent FMM uses the separate periodic operator described below.

E ANALYTICAL LATENT FMM

Table 17: Periodic and non-periodic ionic testing errors. Training includes systems with $N = 256, 512, 1024$, while OOD denotes larger systems not included in training. Within each system size and error metric, the lowest means are bold and second-lowest underlined (ties retained).

		Energy RMSE (meV/atom)		Force RMSE (meV/Å)	
N	Domain	MACE–LES	NEMORA	MACE–LES	NEMORA
Non-periodic clusters					
256	ID	**5.042**	5.272	**16.81**	17.05
512	ID	**0.496**	0.565	**16.93**	17.27
1024	ID	**4.538**	4.870	**16.82**	17.15
2048	OOD	**8.000**	8.481	**16.76**	17.16
4096	OOD	**10.792**	11.410	**16.86**	17.22
Periodic cells					
256	ID	0.308	**0.284**	9.80	**9.64**
512	ID	**0.203**	0.211	10.01	**9.83**
1024	ID	0.155	**0.133**	9.88	**9.69**
2048	OOD	0.142	**0.109**	9.95	**9.78**
4096	OOD	0.120	**0.090**	9.95	**9.79**
latent source *charges* $(q_{\mathrm{Cl}}, q_{\mathrm{Na}})$					
Non-periodic, LES		$(-0.5999, +0.5999)$			
Non-periodic, NEMORA		$(-0.5999, +0.5999)$			
Periodic, LES		$(-0.6002, +0.6003)$			
Periodic, NEMORA		$(-0.6000, +0.6000)$			

For periodic and non-periodic ionic clusters with Coulomb and truncated Lennard-Jones interactions, the scalar-source latent FMM reduction of NEMORA maintains similar force and energy accuracy to LES and is able to extrapolate to unseen larger system sizes (Table 17). This strongly indicates that the long-range interactions are correctly computed between the latent scalar charges and that those learned charges are consistent. Supporting this, the independently trained LES and latent FMM scalar charges both agree with the scalar charges assigned in the underlying coarse-grained model, ± 0.6 in reduced units.

Table 18: Periodic revPBE-D3 water test errors. Systems with $N = 384$ and 768 atoms are included in the training and testing sets. Energy error is in meV/atom and force error in meV/Å. Appendix E gives the numerical settings and evaluation details. $\langle q_{\mathrm{O}} \rangle$ and $\langle q_{\mathrm{H}} \rangle$ are test-set means of the latent scalar sources. Within the same system size and error metric, lowest means are bold and second-lowest underlined (ties retained).

Model	Energy RMSE	Force RMSE	$\langle q_{\mathrm{O}} \rangle$	$\langle q_{\mathrm{H}} \rangle$
$N = 384$				
MACE	0.340	21.86	—	—
MACE–LES	0.351	15.75	−0.1919	0.0959
NEMORA ($\ell_{\mathrm{source}} = 0$)	0.345	15.88	−0.2054	0.1027
NEMORA ($\ell_{\mathrm{source}} = 1$)	**0.259**	**13.07**	−0.4766	0.2383
$N = 768$				
MACE	0.499	20.09	—	—
MACE–LES	0.459	13.31	−0.1878	0.0939
NEMORA ($\ell_{\mathrm{source}} = 0$)	0.474	13.37	−0.2011	0.1006
NEMORA ($\ell_{\mathrm{source}} = 1$)	**0.358**	**10.95**	−0.4755	0.2377

We find similar results for periodic water systems labeled at the revPBE-D3 level. Specifically, both LES and NEMORA with learned scalar charges improve significantly upon bare MACE on force accuracy. The species-dependent learned scalar latent charges on O and H have averages reported in Table 18. When NEMORA is reduced to its simplest analytical form with scalar sources, the two methods differ only in the algorithm used to evaluate the long-range interactions. The most

efficient implementation of LES could scale as $\mathcal{O}(N \log N)$, whereas FMM scales as $\mathcal{O}(N)$. For this reason, NEMORA can enable the inclusion of long-range interactions in near-DFT accurate MLIP simulations at larger scales than direct summation. Including atomic dipoles ($\ell_{\text{source}} = 1$) improves the water results further for both energy and forces (Table 18), at an additional computational cost. Interestingly, inclusion of higher order sources also changes the magnitude of the learned scalar charges.

At fixed expansion order and bounded interaction lists, the analytical FMM translation chain has $\mathcal{O}(N)$ cost. For the relatively small system sizes tested here, the latent FMM method is more computationally demanding and slower than LES. However, FMM becomes advantageous for moderate and large systems because of its superior scaling.

Reduction from NEMORA. We construct an analytical latent FMM to isolate the numerical multipole approximation from the learned long-range operators. We restrict the atomic sources to scalar latent charges, fix the P2M, M2M, M2L, L2L, and L2P maps to their analytical coefficients, and use extensive summation throughout the hierarchy. The learned gates, radial deformations, channel mixing, many-body contractions, and Kelvin reaction branches are removed. Every pair is assigned either to an admissible multipole interaction or to direct P2P evaluation, including pairs inside the backbone cutoff. Finally, the long-range feature update and learned energy correction are replaced by a charge–potential contraction. Only the short-range backbone, its energy readouts, and the atomic charge readouts remain trainable.

Specifically, independent scalar heads on the MACE features give $q_i = \sum_{t^\star} \mathcal{Q}_{t^\star}(\boldsymbol{h}_i^{(t^\star)})$, and the energy is

$$E = E_{\text{SR}} + \frac{1}{2}\widetilde{\boldsymbol{q}}^\top K_\sigma^{\text{BC}} \widetilde{\boldsymbol{q}}, \qquad (K_\sigma^{\text{free}})_{ij} = c\,\frac{\operatorname{erf}(r_{ij}/(\sqrt{2}\sigma))}{r_{ij}} \quad (i \neq j), \tag{10}$$

with zero free-space diagonal, $r_{ij} = \|\boldsymbol{r}_i - \boldsymbol{r}_j\|$, $\sigma = 1$ Å, and the fixed LES normalization $c = 90.4756/(2\pi)$. We evaluate the Coulomb part by a Cartesian Taylor FMM and subtract the direct $\operatorname{erfc}(r/(\sqrt{2}\sigma))/r$ correction. The Cartesian translation tables contain no learned parameters. Under PBC, images with $\|\boldsymbol{n}\|_\infty \leq w$ enter the P2P/M2L lists; the remaining lattice sum is represented by a precomputed root operator with the zero reciprocal mode omitted and the same self-interaction convention as LES. Thus, for identical sources, smoothing, and boundary conventions, Equation 10 is the LES energy functional (Cheng, 2025) evaluated by a different numerical method. The reported periodic FMM training runs use $\widetilde{q}_i = q_i - N^{-1}\sum_j q_j$, whereas the LES controls use the raw charges; this neutrality projection is an additional modeling choice and must be matched in an exact operator comparison. Non-periodic runs use $\widetilde{q}_i = q_i$. Forces differentiate the full energy, including the dependence of the latent charges on the coordinates.

Ionic systems. We use equimolar binary Lennard-Jones systems with fixed reference charges $\pm 0.6\,e$: spherical clusters labeled by direct Coulomb summation and periodic cells labeled by Ewald/PPPM. These are separate benchmarks from the small NaCl clusters of Table 1. Training and validation contain $N = 256, 512, 1024$; $N = 2048, 4096$ test size extrapolation. Within each corpus, latent FMM and LES share the backbone, data splits, optimizer, and training schedule. The ionic backbone has two interactions, a 4.5 Å cutoff, and $64 \times 0\text{e} + 64 \times 1\text{o} + 64 \times 2\text{e}$ hidden irreps. Training uses 240 epochs with stochastic weight averaging (SWA) starting at epoch 120. Each model in this section is trained once, and the checkpoint is selected by validation force error. The analytical expansion order is $p = 3$, with $\theta = 0.5$ for clusters and $\theta = 0.35$, $w = 2$ for periodic interaction lists. The periodic LES controls use reciprocal-grid parameter $d_l = 2.0$ Å.

Force errors remain nearly constant across the tested sizes under both boundary conditions (Table 17). For clusters, latent FMM is within 2.4% of LES in force RMSE. These fixed-charge systems test recovery and size transfer of a known long-range functional.

Periodic water. The revPBE-D3 corpus contains 1,366 training and 547 validation configurations with 384 or 768 atoms. MACE, MACE–LES and each analytical model share one backbone: one MACE interaction, a 6 Å cutoff and $128 \times 0\text{e} + 128 \times 1\text{o}$ hidden irreps, trained for 800 epochs with SWA starting at epoch 400. Both source-degree arms use per-size batches of four 384-atom or two

768-atom frames. The analytical settings are $p = 3$, $\theta = 0.35$ for periodic lists and image-shell radius $w = 2$.

Numerical convergence. We hold charges fixed to separate solver error from fitting error. For an admissible pair, $\rho = (R_s + R_t)/\|\boldsymbol{c}_t - \boldsymbol{c}_s\| < \theta < 1$. Classical multipole estimates predict geometric decay of truncation error with p at fixed separation; a smaller θ tightens admissibility at the expense of additional direct work (Greengard & Rokhlin, 1987; Cheng et al., 1999). Table 19 tests both controls in float64 on 96 uniformly sampled points in $[-9, 9]^3$ with centered Gaussian charges, exercising all five translation stages. Increasing p reduces energy and force errors, while tightening θ reduces force error. At $\theta = 0.4$, every pair is routed to P2P, recovering direct summation to roundoff. The partition audit finds no missing or multiply counted pairs.

Table 19: Fixed-charge numerical errors. In the free-space Coulomb test, $\epsilon_E = |E - E_{\text{ref}}|/|E_{\text{ref}}|$ and $\epsilon_{F,\infty} = \|\boldsymbol{F} - \boldsymbol{F}_{\text{ref}}\|_\infty / \|\boldsymbol{F}_{\text{ref}}\|_\infty$ use direct summation as reference. The periodic test uses one 256-ion configuration with charges $\pm 0.6\, e$ and a converged Ewald reference (reciprocal-grid parameter $d_l = 0.6$ Å); its last column is the absolute force component RMS difference in meV/Å. Here w is the explicit image-shell radius. All entries use float64 and contain no prediction error against training labels.

Free-space Coulomb, $N = 96$				
p	θ	M2L rows	ϵ_E	$\epsilon_{F,\infty}$
2	0.5	4	3.01×10^{-4}	1.89×10^{-4}
4	0.5	4	1.28×10^{-5}	1.01×10^{-5}
6	0.5	4	5.66×10^{-7}	1.09×10^{-6}
4	0.6	10	8.95×10^{-6}	2.38×10^{-5}
4	0.4	0	2.56×10^{-16}	3.87×10^{-16}
Periodic smoothed Coulomb, $N = 256$				
p	θ	w	ϵ_E	Force RMS difference
3	0.50	1	8.77×10^{-4}	2.427
3	0.50	2	4.22×10^{-5}	2.369
3	0.35	1	1.09×10^{-3}	1.016
3	0.35	2	1.60×10^{-4}	0.531
4	0.35	2	8.04×10^{-5}	0.101
5	0.35	2	3.86×10^{-5}	0.031

For periodic ions at $p = 3, w = 2$, tightening θ from 0.5 to 0.35 reduces force error by a factor of 4.46, while direct interaction rows increase from 312,989 to 915,502. Individual energy errors need not decrease monotonically because signed contributions can cancel. The observed trends agree with classical FMM convergence; they do not establish error bounds for NEMORA's learned operators. Tree geometry also affects finite-order error: on a 384-atom water configuration with fixed charges $q_{\text{H}} = 0.2$, $q_{\text{O}} = -0.4$, the $p = 3, \theta = 0.35, w = 2$ evaluator with maximum leaf occupancy 32 has relative energy error 3.03×10^{-6} and force error 0.057 meV/Å against converged Ewald. Occupancy 8 and the fixed-depth training tree give 0.345 and 0.168 meV/Å, respectively. High-order force-loss derivatives require float64 in the current implementation. The reference smoothing correction enumerates all within-graph pairs, so these accuracy tests do not establish linear runtime for the complete analytical backend.

Higher-order latent atomic moments. We have implemented latent atomic dipoles (1o) and symmetric traceless quadrupoles (2e) predicted from equivariant interaction features (Stone, 1981). Their interactions use fixed kernel derivatives throughout the translation chain, including P2P and periodic terms. Zero initialization of these heads recovers the scalar model. Dipolar water results are reported in Table 18; no quadrupolar accuracy comparison is reported here. Increasing the atomic source degree ℓ_{source} adds independent anisotropy beyond the molecular and cluster moments already generated by spatially distributed scalar charges. It is distinct from increasing the box expansion order p, which controls numerical accuracy. The dipolar model improves both water test metrics in the reported single-seed experiment. This extension retains analytical translations; NEMORA additionally learns transport and many-body coupling.

F DATASETS

F.1 CUMULENE

The cumulene datasets employed here are the guaranteed nonlocal (GNL) datasets (Batatia et al., 2024) which serve as a benchmark for an MLIP to learn and extrapolate long-range higher-order irrep interactions. Cumulene is a double-bonded carbon chain with terminal CH_2 groups at each end. Even for long chains, the energy depends strongly on the relative angle between the terminal groups due to significant electron delocalization along the chain. For the comparisons in Table 1, the training and validation sets consist of optimized geometries of cumulene chains with lengths of $n_C \in [3-10, 13, 14]$ carbon atoms that have been perturbed by Gaussian noise and had the CH_2 terminal ends relatively rotated to generate additional angular distribution training data. This gives a total of 200 training and 50 validation data points. The test set contains 170 configurations consisting of 50 structures from the in-distribution (ID) training chain lengths, 60 structures from a semi-out-of-distribution (SOD) set, and 60 structures from an out-of-distribution (OOD) test set with lengths of $n_C \in [11, 12]$ and $n_C \in [15, 16]$ carbon atoms, respectively. Energies and forces were computed in ORCA at the *wB97X*-D3 level of theory with the *def2-TZVP* basis set.

For the length and angle sweep plots in Figure 2(a,b), and the $n_C \in 3$–64 columns of Table 6, a separate model was trained. The training data consisted of all of the frames from the previously described set, plus 145 additional configurations gathered by increasing the chain length at fixed dihedral angles ($\varphi = 5°$ and $85°$) and sweeping the dihedral angle at fixed chain length ($n_C = 12$). The set of 565 configurations was split randomly into 509 training and 56 validation frames.

F.2 CHARGED CARBON CHAIN ($C_{10}H_2$/$C_{10}H_3^+$)

We utilize the $C_{10}H_2$/$C_{10}H_3^+$ dataset introduced by Ko et al. (2021), which serves as a rigorous test for nonlocal interactions. The set comprises configurations of neutral decadiyne ($C_{10}H_2$) and its protonated, positively charged counterpart ($C_{10}H_3^+$). The addition of the proton induces global charge delocalization across the quasi-1D π-conjugated chain, resulting in a strongly nonlocal energy and force response. We retain the original splits, consisting of 9,017 training and 1,002 validation structures with a roughly equal distribution of neutral and charged species. To handle the long-range dependencies, we coarsen these non-periodic geometries using NEMORA's adaptive octree (Section 4.2).

F.3 NA CLUSTERS ($NA_8CL_8^+$, $NA_9CL_8^+$)

We further evaluate our model on the NaCl cluster benchmark (Ko et al., 2021), which consists of $Na_8Cl_8^+$ and $Na_9Cl_8^+$ clusters in vacuum under non-periodic boundary conditions. Both clusters carry a net charge of $+1$. In $Na_9Cl_8^+$, an additional neutral sodium atom is introduced on one side of the cluster. This changes how the excess charge is distributed throughout the lattice, significantly modulating the energy and forces across the entire system. Specifically, the benchmark measures the energy required to displace a sodium ion on the opposite side of the cluster (≈ 21 Å away). Capturing the dependence of this local binding energy on the distal charge state requires the model to explicitly account for long-range interactions that span the cluster. We utilize a stratified split of 4,500 training and 500 validation frames. Consistent with prior work (Ko et al., 2021; Rumiantsev et al., 2026), we report validation errors in Table 1.

F.4 AU_2/MGO

The Au_2/MgO benchmark (Ko et al., 2021) consists of an Au_2 dimer adsorbed on a periodic MgO(001) slab, with or without Al dopant atoms placed deep inside the slab. Each configuration contains 110 atoms in a $9.05 \times 9.05 \times 26.46$ Å cell. Although the dopants lie far from the surface, they change the charge distribution of the slab and, with it, the preferred adsorption geometry of the dimer. We use 4,500 training and 500 validation configurations and, as for NaCl, report validation errors in Table 1.

F.5 MD22

MD22 (Chmiela et al., 2023) contains *ab initio* molecular dynamics trajectories of seven biomolecular and supramolecular systems with 42 to 370 atoms: the tetrapeptide Ac–Ala3–NHMe, docosahexaenoic acid (DHA), the tetrasaccharide stachyose, the DNA base-pair stacks AT–AT and AT–AT–CG–CG, the buckyball catcher and a double-walled nanotube. Energies and forces are computed at the PBE+MBD level. The training, validation and test splits for each system are given in Appendix A.4.

F.6 Periodic Ionic Melt

The ionic benchmark is a binary melt of equal numbers of cations and anions with charges $\pm 0.6\, e$, interacting through a Lennard-Jones potential ($\epsilon = 0.1$ eV, $\sigma = 3.0$ Å, 6 Å cutoff) and the full periodic Coulomb interaction, which we evaluate with PPPM in LAMMPS (Thompson et al., 2022). Configurations are randomized arrangements drawn from molecular dynamics trajectories in cubic cells at fixed density, with edge lengths from 20.5 Å ($N = 256$) to 51.7 Å ($N = 4096$). Each trajectory draws its own random, charge-neutral assignment of cations and anions, which is held fixed along the trajectory. Because the charge distribution is quenched rather than equilibrated, it is not screened, and the long-range contribution cannot be absorbed into a short-range model. Splits are drawn from independent trajectories: at each training size ($N = 256, 512, 1024$) there are 500 training, 200 validation and 50 test configurations, and $N = 2048$ and 4096 contribute 50 test configurations each. A non-periodic counterpart uses spherical clusters of the same sizes and interactions, labeled by direct Coulomb summation, with 400 training, 100 validation and 50 test configurations per training size.

F.7 Periodic Water

The periodic water dataset was generated in this work. It contains liquid-water configurations in cubic periodic cells with 128 and 256 molecules (384 and 768 atoms), labeled with CP2K (Kühne et al., 2020) using the revPBE functional (Zhang & Yang, 1998) with D3 dispersion corrections (Grimme et al., 2010). The calculations use the Gaussian and plane-wave method with TZV2P MOLOPT basis sets (VandeVondele & Hutter, 2007), GTH pseudopotentials (Goedecker et al., 1996) and a 1200 Ry plane-wave cutoff. The dataset is split into 1,366 training, 547 validation and 820 test configurations (788, 328 and 474 at 384 atoms; 578, 219 and 346 at 768 atoms).

G MODEL SETTINGS

To ensure an equitable comparison across long-range architectures, we standardized the expressivity of all models. For the MACE-based variants (MACE, MACE–LES, MACE–RANGE, and NEMORA), which share a common short-range backbone, parameter matching was straightforward. For LOREM and EFA, we prioritized consistency in key physical hyperparameters: the receptive field ($T \cdot r_{\text{cut}}$), angular resolution (ℓ_Q), and maximum irrep degree (L_Q). Finally, we balanced the total learnable parameter count across all models by tuning the number of channels in their respective short-range irreps.

G.1 MACE

To isolate the effects of different long-range treatments, we utilize a consistent short-range MACE backbone across the MACE, MACE–LES, MACE–RANGE, and NEMORA architectures. For each benchmark, the hyperparameters of this backbone are standardized according to Table 20, while the optimization protocol follows the settings detailed in Table 21. Any parameters not explicitly specified are kept at their default values as defined in the MACE repository (v0.3.15) to ensure reproducibility.

Table 20: Shared parameters for short-range MACE backbone upon which all MACE-based long-range models are built.

Hyperparameter	**Value**
Interactions	2
ℓ_{max}	3
Hidden irreps	128x0e + 128x1o + 128x2e
ν	3
Cutoff basis functions	5
Radial basis functions	8
Precision	float32 (the Au_2/MgO NEMORA run: float64)

Table 21: Shared optimizer settings for all MACE-based models

Hyperparameter	**Value**
Optimizer	Adam
Learning rate	0.01
Energy:Force weight	10:100
SWA Energy:Force weight	1000:1000
EMA decay	0.99

All models are trained by minimizing a weighted energy + force regression loss

$$\mathcal{L} = \frac{w_E}{N} \sum_{n=1}^{N} \left(E_n^{\text{pred}} - E_n^{\text{ref}}\right)^2 + \frac{w_F}{3\,N_{\text{atoms}}} \sum_{n,i,\alpha} \left(F_{n,i,\alpha}^{\text{pred}} - F_{n,i,\alpha}^{\text{ref}}\right)^2, \tag{11}$$

with per-system (w_E, w_F) from Table 21: pre-SWA training uses $(10, 100)$ and the SWA stage switches to $(1000, 1000)$. Stress / virial / dipole supervision is not used. In addition to the shared parameters described above, there are a few parameters that vary on a per-system basis, which are specified in Table 22.

Table 22: System specific MACE and MACE–LES hyperparameters for the cumulene and Ko benchmark systems.

Hyperparameter	Cumulene	Cumulene (3-64)	NaCl	Carbon
Cutoff radius (Å)	3.0	3.0	3.0	3.0
Max epochs	2000	1000	600	600
SWA start epoch	1000	500	300	300
Batch size	16	16	16	32

G.2 LOREM

Table 23: LOREM hyperparameters for the cumulene and Ko benchmark systems.

Hyperparameter	Cumulene	Cumulene (3-64)	NaCl	Carbon
Cutoff radius (Å)	3.0	3.0	3.0	3.0
$\ell_{\text{max,lr}}$	2	2	2	2
ℓ_{max}	6	6	6	6
Feature channels	128	128	128	128
Message-passing layers	1	1	1	1
Radial basis functions	8	8	8	8
Spherical features	8	8	8	8
Optimizer	Adam	Adam	Adam	Adam
Learning rate	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Schedule	linear decay	linear decay	linear decay	linear decay
Minimum learning rate	10^{-6}	10^{-6}	10^{-6}	10^{-6}
Max epochs	4000	2000	1200	1200

G.3 EFA

Table 24: EFA hyperparameters for the cumulene and Ko benchmark systems.

Hyperparameter	Cumulene	Cumulene (3-64)	NaCl	Carbon
Cutoff radius (Å)	2.0	2.0	3.0	2.0
Feature channels	144	144	256	144
Message-passing layers	3	3	2	3
MP maximum degree	3	3	3	3
MP radial basis functions	32	32	32	32
ERA maximum degree	2	2	2	2
ERA maximum length (Å)	19.0	86.777	25.0	19.0
ERA query/key channels	32	32	32	32
ERA value channels	64	64	64	64
Lebedev points	50	50	50	50
Optimizer	Adam	Adam	Adam	Adam
Learning rate	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Schedule	exp decay	exp decay	exp decay	exp decay
Final learning rate	10^{-5}	10^{-5}	10^{-5}	10^{-5}
Loss weights (E/F)	0.01 / 0.99	0.01 / 0.99	0.01 / 0.99	0.01 / 0.99
Max epochs	200000	50000	10000	200000

G.4 MACE–RANGE

Table 25: MACE–RANGE hyperparameters for the cumulene and Ko benchmark systems.

Hyperparameter	Cumulene	Cumulene (3-64)	NaCl	Carbon
Cutoff radius (Å)	3.0	3.0	3.0	3.0
Hidden irreps	128x0e + 128x1o + 128x2e	128x0e + 128x1o + 128x2e	128x0e + 128x1o + 128x2e	128x0e + 128x1o + 128x2e
MLP irreps	16x0e	16x0e	16x0e	16x0e
Maximum degree	3	3	3	3
Correlation order	3	3	3	3
Interactions	2	2	2	2
Bessel functions	8	8	8	8
Polynomial cutoff basis	5	5	5	5
Virtual nodes	4	3	3	3
Virtual hidden channels	432	432	432	432
Virtual basis dimension	10	10	10	10
Virtual heads	8	8	8	8
Optimizer	AdamW	AdamW	AdamW	AdamW
Learning rate	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Weight decay	5×10^{-7}	5×10^{-7}	5×10^{-7}	5×10^{-7}
Scheduler	CustomStepLR	CustomStepLR	CustomStepLR	CustomStepLR
Scheduler steps	300, 200, 100	300, 200, 100	608, 192, 160	608, 192, 160
Scheduler gammas	0.8, 0.8, 0.8	0.8, 0.8, 0.8	0.8, 0.8, 0.8	0.8, 0.8, 0.8
Max epochs	2000	2000	600	600
Batch size	64	16	16	32

G.5 NEMORA

In addition to the short-range MACE parameters in Table 20, all NEMORA models share the parameters in Table 26. The defaults follow the production NEMORA configuration: density-mean fusion-side aggregation ($\alpha_l = 1$) paired with extensive upward moments ($\alpha_l^{\mathrm{up}} = 0$); a scale-invariant gate input that drops $\log R$ and retains only invariant ratios; bounded sigmoid per-channel gates with the recipe (HIER $0.03/0.10$, KELVIN/KELVIN_PBC/P2P $0.005/0.05$); the learned P2P channel disabled by default in favor of a smooth handoff to the short-range backbone (leaves promote to M2L); a single Kelvin anchor with the algebraic self-Kelvin scalar enabled; one Highway-gated Neumann iteration with cumulant modulation; and a fixed 1×10^{-4} tensor-product weight decay applied via a separate optimizer group. System-specific knobs ($r_{\max}$, hierarchy depth, MAC threshold θ, periodic-tail mode, batch size) for the cumulene, NaCl and carbon-chain benchmarks appear in Tables 27 and 28.

Table 26: Shared hyperparameters of the NEMORA module. The split optimizer routes the backbone group to the standard MACE LR / weight decay and the NEMORA group to the values above. Non-PBC systems use `kelvin_pbc_mode=off` and `hier_pbc_image_mode=off`; PBC systems use `madelung` for both. The cumulene and NaCl models use $\nu = 3$, and the NaCl model uses $T_{\mathrm{res}} = 2$.

Hyperparameter	Value
NEMORA irreps	32x0e + 32x1o + 32x2e
$\ell_{\max,\mathrm{N}}$	2
Correlation order ν (atom-level SymContract)	2
Attachment layer (MACE `products[i]`)	-2
Bessel basis size n_{rbf}	8
Gate-MLP hidden width	64
Use full Clebsch–Gordan TP	False (sparse uvu+linear)
P2P channel	disabled (`kill_p2p`=True)
Kelvin anchors K	1
Self-Kelvin $\kappa_\ell(R)$	enabled
Highway-gated Neumann iterations T_{res}	1
Cumulant modulation γ_{cum}	0.5
Density-mean aggregation α_l (downward / fusion)	1.0
Upward aggregation α_l^{up}	0.0
Use $\log R$ in gate input	False
Channel-gate recipe (init / max)	hier $0.03/0.10$; kelvin, kelvin_pbc, p2p $0.005/0.05$
NEMORA learning rate	1×10^{-3}
Warmup steps (linear, NEMORA group only)	2000
TP weight decay	1×10^{-4}
Gradient clipping	10.0
EMA decay	0.99

Table 27: System specific hyperparameters for model training.

System	$r_{\max}$	Batch	Epochs	Start SWA	Periodic	$r_{\max,\mathrm{EFA}}$
cumulene	3.0	16	2000	1000	no	25.000
NaCl	3.0	16	600	300	no	25.819
carbon-chain	3.0	32	600	300	no	18.858

EFA EPE max frequency is $\pi/4$ for periodic systems and π for nonperiodic systems.
Shared MACE–LOREM and MACE–EFA parameters are unchanged from Tables 23 and 24, respectively.

Table 28: Per-system NEMORA settings used in addition to the shared NEMORA hyperparameters of Table 26 and the shared model-training settings of Table 27. k_{target} (`target_k`) is the children-per-cluster target during coarsening, $L_{\max}$ (`max_levels`) the hierarchy depth cap, θ_{MAC} the multipole acceptance threshold, and K_{Kelvin} (`kelvin_K`) the number of Kelvin reaction-field anchors. Periodic systems route the lattice tail through both the Madelung hierarchical injection and the Madelung Kelvin image branch; non-periodic systems disable both. All systems use the adaptive octree coarsener, with the quasi-1D and small-cluster systems collapsing the fanout to binary or quadtree in their PCA frame.

System	k_{target}	$L_{\max}$	θ_{MAC}	K_{Kelvin}	PBC modes (`hier` / `kelvin`)	Coarsener
cumulene	3	2	0.5	1	off / off	adaptive_octree
NaCl	8	2	0.5	1	off / off	adaptive_octree
carbon-chain	4	2	0.5	1	off / off	adaptive_octree

H GREEN'S OPERATOR FAMILIES

Tier 1: Drop-in scalar radial families. These require only changing the scalar $f_\ell(r)$ function; modifications to the operator structure, TP, gating, or hierarchy are not necessary. Any combination of Tier 1 rows may be selected simultaneously.

Table 29: **Tier 1: drop-in scalar radial families.** For each kernel, the irregular radial enters M2L, P2P, and the PBC lattice sum; the regular radial enters P2M, M2M, L2L, L2P, and Kelvin. The ϵ_{rad}-smoothing ($r \to \sqrt{r^2 + \epsilon_{\mathrm{rad}}^2}$) is applied to all forms. Selecting *only* the kernel-free row yields the fully learned kernel-independent limit, where no analytical radial prior is imposed. Apart from the Coulomb default, these families are not benchmarked in this paper.

Kernel family	$G(r)$	**Irregular** f_ℓ^{irreg}	**Regular** f_ℓ^{reg}	**Status / use case**
Coulomb	$1/r$	$r^{-(\ell+1)}$	r^ℓ	**Default.** Electrostatics, Madelung energy, charge–charge and charge–dipole.
Yukawa / mod. Helmholtz	$e^{-\kappa r}/r$	$\kappa^{\ell+1} k_\ell(\kappa r)$	$\kappa^{-\ell} i_\ell(\kappa r)$	Screened electrostatics (Debye–Hückel), ionic solutions, Thomas–Fermi screening. i_ℓ, k_ℓ: modified spherical Bessel functions.
London C_6	$1/r^6$	$r^{-(6+\ell)}$	$r^{5+\ell}$	Leading van der Waals dispersion. Molecular crystals, layered materials, rare-gas solids.
London C_8	$1/r^8$	$r^{-(8+\ell)}$	$r^{7+\ell}$	Next-order dispersion for high-accuracy extended systems.
Dipolar	$1/r^3$	$r^{-(3+\ell)}$	$r^{2+\ell}$	Dipole–dipole interaction tensor. Magnetic dipolar interactions, polarization propagation.
Biharmonic	r	$r^{-\ell}$	$r^{\ell+1}$	Higher-order elliptic / plate-like models. Unusual: free-space kernel *grows* with r; FMM translation theory is established.
Kernel-free	**—**	**1**	**1**	**Learned channel with no physics prior.** Relies entirely on $Y_\ell^m \cdot R_n$ and the FCTP to discover effective radial behavior from data. Initialize near zero (see below).

I PROOFS

Analytical equivariant FMM operators. Equation 3 is the Laplace expansion of the Coulomb kernel, $\|\boldsymbol{r}_> - \boldsymbol{r}_<\|^{-1} = \sum_{\ell,m} \frac{4\pi}{2\ell+1} \frac{\|\boldsymbol{r}_<\|^\ell}{\|\boldsymbol{r}_>\|^{\ell+1}} Y^\star_{\ell m}(\hat{\boldsymbol{r}}_<) Y_{\ell m}(\hat{\boldsymbol{r}}_>)$ for $\|\boldsymbol{r}_<\| < \|\boldsymbol{r}_>\|$, summed over the charges of the source cluster and truncated at $\ell \le p$ (Greengard & Rokhlin, 1987; Cheng et al., 1999). The other stages follow from the addition theorems of the solid harmonics. To state them with their normalization, we use orthonormal complex spherical harmonics with the Condon–Shortley phase and the Racah-normalized regular and irregular solid harmonics

$$\tilde{\boldsymbol{R}}^{(\ell)}(\boldsymbol{x}) = \sqrt{\frac{4\pi}{2\ell+1}}\, \|\boldsymbol{x}\|^\ell\, \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{x}}), \qquad \tilde{\boldsymbol{I}}^{(\ell)}(\boldsymbol{x}) = \sqrt{\frac{4\pi}{2\ell+1}}\, \|\boldsymbol{x}\|^{-(\ell+1)}\, \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{x}}). \tag{12}$$

The moments of Section 2.2 are $\boldsymbol{\Phi}_s^{(\ell)} = [(2\ell+1)/4\pi]^{1/2} \sum_{j\in s} q_j \tilde{\boldsymbol{R}}^{(\ell)\star}(\boldsymbol{r}_j - \boldsymbol{c}_s)$, the target local coefficients are $\boldsymbol{\Phi}_t^{(\ell)} = \sum_j q_j \tilde{\boldsymbol{I}}^{(\ell)}(\boldsymbol{r}_j - \boldsymbol{c}_t)$, summed over the transferred sources, and they give $V_{\text{far}}(\boldsymbol{c}_t + \boldsymbol{x}) \approx (4\pi\varepsilon_0)^{-1} \sum_{\ell=0}^{p} \boldsymbol{\Phi}_t^{(\ell)} \cdot \tilde{\boldsymbol{R}}^{(\ell)}(\boldsymbol{x})$ with $\boldsymbol{a}\cdot\boldsymbol{b} = \sum_m a_m b^\star_m$. Regular solid harmonics at the origin are defined by continuity, so $\tilde{\boldsymbol{R}}^{(0)} = 1$ and $\tilde{\boldsymbol{R}}^{(\ell)}(\boldsymbol{0}) = \boldsymbol{0}$ for $\ell > 0$, and a particle expanded about its own position holds only q_j. A consistent unitary transformation of the harmonics, coefficient vectors, and CG tensors gives the real-harmonic convention of the main text, in which the conjugation in $\boldsymbol{a}\cdot\boldsymbol{b}$ drops.

The addition theorems in this normalization are

$$\begin{aligned} \tilde{\boldsymbol{R}}^{(\ell)}(\boldsymbol{a}+\boldsymbol{b}) &= \sum_{n=0}^{\ell} \binom{2\ell}{2n}^{1/2} \big[\tilde{\boldsymbol{R}}^{(\ell-n)}(\boldsymbol{a}) \otimes \tilde{\boldsymbol{R}}^{(n)}(\boldsymbol{b})\big]^{(\ell)}, \\ \tilde{\boldsymbol{I}}^{(\ell)}(\boldsymbol{a}-\boldsymbol{b}) &= \sum_{\ell'=0}^{\infty} (-1)^{\ell+\ell'} \left[\frac{2(\ell+\ell')+1}{2\ell+1} \binom{2(\ell+\ell')}{2\ell'}\right]^{1/2} \big[\tilde{\boldsymbol{R}}^{(\ell')}(\boldsymbol{a}) \otimes \tilde{\boldsymbol{I}}^{(\ell+\ell')}(\boldsymbol{b})\big]^{(\ell)}. \end{aligned} \tag{13}$$

The irregular theorem requires $\|\boldsymbol{a}\| < \|\boldsymbol{b}\|$ and is truncated at $\ell' \le p$; the regular theorem is a finite sum and exact at every order. With $\boldsymbol{a} = \boldsymbol{r}_j - \boldsymbol{c}_s$ and $\boldsymbol{b} = \boldsymbol{r}_>$, the irregular theorem at $\ell = 0$ is Equation 3, and at $\ell > 0$ it is M2L, which builds every local coefficient at $\boldsymbol{c}_t$ from the moments at $\boldsymbol{c}_s$, entering the complex-basis theorem through their conjugates $\overline{\boldsymbol{\Phi}}_s^{(\ell')}$. The regular theorem with $\boldsymbol{b}$ the shift between two centers re-centers a multipole expansion (M2M) and, through the transpose of the same coupling, a local expansion (L2L); L2P is the evaluation of V_{far} above. Every stage couples one harmonic to the transported state and never to the other harmonic, and approximation enters only through truncating the physical expansions. The source enclosure about $\boldsymbol{c}_s$ must contain its particles, and $(R_s + R_t)/\|\boldsymbol{r}_>\| < 1$ ensures separation from the target enclosure. The algorithm reuses these factors across the hierarchy.

The same construction applies componentwise to tensor-valued sources $\boldsymbol{Q}_j^{(\ell_Q)}$, transforming as $\boldsymbol{Q}_j^{(\ell_Q)} \mapsto D^{(\ell_Q)}(R)\boldsymbol{Q}_j^{(\ell_Q)}$ under $R \in \mathrm{SO}(3)$. Convolution with the scalar Coulomb Green's function gives

$$\boldsymbol{V}^{(\ell_Q)}(\boldsymbol{r}_i) = \frac{1}{4\pi\varepsilon_0} \sum_{j\neq i} \frac{\boldsymbol{Q}_j^{(\ell_Q)}}{\|\boldsymbol{r}_i - \boldsymbol{r}_j\|}. \tag{14}$$

Because pair distances are rotation invariant, simultaneous rotation of the positions and sources transforms the output by $D^{(\ell_Q)}(R)$. This is componentwise tensor transmission through a scalar kernel (Rumiantsev et al., 2026); physical dipole and quadrupole interactions instead require derivatives of the kernel (Stone, 2013).

Retain the source-tensor component $\sigma \in \{-\ell_Q, \ldots, \ell_Q\}$ separately from the spatial indices. Replacing q_j by $Q_{j,\sigma}^{(\ell_Q)}$ gives componentwise multipoles, and every stage above acts on the spatial indices only, leaving σ unchanged. Setting $\ell_Q = 0$ recovers scalar-source FMM at any retained order p, including spatial multipoles with $\ell > 0$.

Equation 4 writes the same componentwise transport with CG products, and two identities connect it to Equation 3. The contraction over m is an invariant CG pairing,

$$\sum_{m=-\ell}^{\ell} Y^{\star}_{\ell m}(\hat{\boldsymbol{a}})\, Y_{\ell m}(\hat{\boldsymbol{b}}) = (-1)^{\ell}\sqrt{2\ell+1}\,\big[\boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{a}}) \otimes \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{b}})\big]^{(0)}, \tag{15}$$

and multiplying a tensor source by this invariant equals coupling the source to the two harmonics one after the other, provided the intermediate degree ℓ_1 is summed with the fixed recoupling weights $c_{\ell_1} = (-1)^{\ell_Q+\ell_1}[(2\ell_1+1)/(2\ell_Q+1)]^{1/2}$:

$$\sum_{\ell_1} c_{\ell_1}\Big[\big[\boldsymbol{Q}^{(\ell_Q)} \otimes \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{a}})\big]^{(\ell_1)} \otimes \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{b}})\Big]^{(\ell_Q)} = \boldsymbol{Q}^{(\ell_Q)} \sum_{m=-\ell}^{\ell} Y^{\star}_{\ell m}(\hat{\boldsymbol{a}})\, Y_{\ell m}(\hat{\boldsymbol{b}}). \tag{16}$$

This is the Wigner $6j$ recoupling of the angular momenta ℓ_Q, ℓ, ℓ, specialized to a scalar coupling of the two harmonics. In Equation 4 the first product is P2M, the second is M2L, ℓ_1 is the total degree of the tensor moments, and the output degree is $\ell_M = \ell_Q$; for $\ell_Q = 0$ the only weight is $c_0 = 1$ and the equation reduces to Equation 3. Every other admissible path, with an output degree $\ell_M \neq \ell_Q$ or with different harmonic degrees at the two stages, carries zero weight in the analytical operator. Physical multipole interactions live on the second kind of path: the potential of a point multipole of degree ℓ_Q at $\boldsymbol{r}_<$, evaluated at $\boldsymbol{r}_>$, is $\sum_{\ell \geq \ell_Q} \zeta_\ell \, \|\boldsymbol{r}_<\|^{\ell-\ell_Q} \|\boldsymbol{r}_>\|^{-(\ell+1)} \big[\big[\boldsymbol{Q}^{(\ell_Q)} \otimes \boldsymbol{Y}^{(\ell-\ell_Q)}(\hat{\boldsymbol{r}}_<)\big]^{(\ell)} \otimes \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{r}}_>)\big]^{(0)}$ with fixed constants ζ_ℓ, a regular harmonic of degree $\ell - \ell_Q$ followed by an irregular harmonic of degree ℓ, and the same-degree paths of Equation 4 cannot represent it (by parity for odd ℓ_Q, and numerically for $\ell_Q = 2$). We verified all of this numerically at expansion order 16 against direct summation: the exact identities, namely Equations 15 and 16, the addition theorems, Equation 3, and Equation 4 against Equation 14 for $\ell_Q \leq 2$, agree to 10^{-15}; the truncated M2L and L2L coefficients to 2×10^{-10}; and the split-degree expansions of the dipole and quadrupole potentials, fitted on random geometries and checked on fresh ones, to 5×10^{-10}, while the best same-degree fit leaves a relative residual of 0.9. Because each coupling path admits a one-dimensional space of real equivariant bilinear maps, the analytical stage coefficients are particular real path weights, which is what separates the fixed CG coefficients and recoupling weights from the learned path weights in $\otimes_\theta$.

In a real harmonic basis, the spatial and source indices carry the product representation $D^{(\ell)} \otimes D^{(\ell_Q)}$, which decomposes into irreducible blocks by Clebsch–Gordan coupling (Geiger & Smidt, 2022). The classical addition theorems make the translation operators equivariant for fixed cluster memberships and centers that rotate with the geometry (Cheng et al., 1999). Retaining complete angular blocks preserves this property under truncation. Equivariance of the full hierarchical algorithm additionally requires the hierarchy construction to transform consistently with the input geometry.

For learned irrep features, equivariance follows from the Clebsch–Gordan intertwining identity,

$$\Big[D^{(\ell')}(R)\boldsymbol{a}^{(\ell')} \otimes D^{(\kappa)}(R)\boldsymbol{b}^{(\kappa)}\Big]^{(\ell)} = D^{(\ell)}(R)\Big[\boldsymbol{a}^{(\ell')} \otimes \boldsymbol{b}^{(\kappa)}\Big]^{(\ell)},$$

for $|\ell' - \kappa| \leq \ell \leq \ell' + \kappa$ (Geiger & Smidt, 2022). Learned channel weights and gates that depend only on invariant quantities preserve this transformation law, as does summation over sources and coupling paths. This establishes rotational equivariance of the learned translation blocks; exact classical FMM recovery additionally requires the analytical coefficients and stage-specific operators above. For $O(3)$ equivariance, the coupled blocks must also satisfy the parity rule stated in Section 2.1. Translation invariance follows because every operator acts only on relative displacements between atoms, cluster centers and Kelvin anchors.

Many-body contractions over multipole expansions. For M2L, let $\boldsymbol{m}_{st}$ denote the gated tensor-product message from source s, so that $\boldsymbol{\Psi}_t = \sum_{s\in\mathcal{N}_t} \boldsymbol{m}_{st}$. ACE/MACE-style symmetric contractions (Drautz, 2019; Batatia et al., 2023) act on this pooled field. Their degree-ν terms have the structure

$$\mathrm{SymContract}_\nu(\boldsymbol{\Psi}_t) \sim \Big[\Big(\sum_{s\in\mathcal{N}_t} \boldsymbol{m}_{st}\Big)^{\otimes\nu}\Big]_{\mathrm{CG}} = \sum_{s_1,\ldots,s_\nu\in\mathcal{N}_t} \big[\boldsymbol{m}_{s_1 t} \otimes \cdots \otimes \boldsymbol{m}_{s_\nu t}\big]_{\mathrm{CG}}, \tag{17}$$

where $[\cdot]_{\mathrm{CG}}$ denotes coupling to the retained output irreps; learned contraction weights and channel indices are suppressed. Cross-terms over distinct source clusters couple the target region to multiple

far-field environments. They provide explicit higher-order features relevant to non-additive long-range effects, including Axilrod–Teller–Muto dispersion (Axilrod & Teller, 1943) and cooperative polarization (Tkatchenko et al., 2012). NEMORA applies $\tilde{\mathbf{\Psi}}^{(\ell_M,L)} = \mathbf{\Psi}^{(\ell_M,L)} + [\mathrm{SymContract}_\nu(\mathbf{\Psi}^{(L)})]^{(\ell_M)}$ at each hierarchy level L, contracting all retained input irreps before projecting onto the output degree. Disabling this block recovers the gated M2L baseline without explicit symmetric contractions; its input features and gates may still carry many-body information. At fixed degrees, channels, and contraction order, the additional cost is linear in the number of hierarchy nodes.

Kelvin transform. Building on the Kelvin transform for spherical boundary-value problems (Amaral et al., 2017) and classical continuum solvation (Born, 1920; Onsager, 1936; Kirkwood, 1939), NEMORA introduces a learnable continuum boundary response over the support's bounding sphere of radius R_k centered at $\boldsymbol{c}_k$:

$$\mathbf{\Psi}_k^{(\ell_M)} = \sum_{s=1}^{N} \boldsymbol{\alpha}_{sk}^{(\ell_M)} \cdot \left(\tilde{r}_k^{-(2\ell_M+1)} \cdot \boldsymbol{\rho}_\theta^{(\ell_M)}\big(\boldsymbol{P}(\|\boldsymbol{r}_{sk}\|)\big) \cdot \boldsymbol{W}^V \mathbf{\Psi}_s^{(\ell_M)} \right) \tag{18}$$

Here $\boldsymbol{\alpha}_{sk}^{(\ell_M)}$ is the invariant gate, $\tilde{r}_k = R_k / r_{\mathrm{cut}}$ is the normalized support radius, $\boldsymbol{r}_{sk} = \boldsymbol{r}_s - \boldsymbol{c}_k$, and $\boldsymbol{P}$ is a polynomial radial encoder. A spherical reaction-field response is diagonal in angular degree. NEMORA retains its radius dependence through $\tilde{r}_k^{-(2\ell_M+1)}$, together with learned degree-preserving channel mixing, radial modulation, and gating. The same form is applied as a local Kelvin reaction at internal hierarchy nodes during the downward sweep, providing a learned boundary-response correction at each scale. The global branch costs $\mathcal{O}(NK)$ for K anchors, and the node-local updates cost $\mathcal{O}(N)$ on a linear-size hierarchy.

Equivariance of the Kelvin reaction operator. The Kelvin transform is the classical inversion symmetry of the Laplace equation: if u is harmonic away from the origin in $\mathbb{R}^3$, then

$$(\mathcal{K}_R u)(\boldsymbol{r}) = \frac{R}{\|\boldsymbol{r}\|} u\left(\frac{R^2 \boldsymbol{r}}{\|\boldsymbol{r}\|^2}\right)$$

is also harmonic away from the inverted singularities. On the spherical-harmonic basis it acts diagonally, mapping each regular solid harmonic to the corresponding irregular harmonic,

$$\mathcal{K}_R\left[r^\ell Y_m^{(\ell)}(\hat{\boldsymbol{r}})\right] = R^{2\ell+1} r^{-(\ell+1)} Y_m^{(\ell)}(\hat{\boldsymbol{r}}),$$

so the Kelvin image of a degree-ℓ multipole differs only by a scalar radial factor and preserves the angular irrep degree. This diagonal action is the mathematical origin of classical reaction-field models: Born, Onsager, and Kirkwood replace the exterior dielectric response by image or reaction multipoles whose amplitudes are degree-dependent scalar factors multiplying the original multipoles.

NEMORA's Kelvin branch is the learnable equivariant analogue of this reaction-field map. For each support k, the pooled anchor feature $\mathbf{\Psi}_k^{(\ell)}$ transforms as

$$\mathbf{\Psi}_k^{(\ell)} \mapsto \boldsymbol{D}^{(\ell)}(R) \mathbf{\Psi}_k^{(\ell)}.$$

The learned boundary response applies degree-preserving channel mixing and invariant factors to each source contribution,

$$\mathbf{\Psi}_{\mathrm{out},k}^{(\ell)} = \sum_s \beta_{sk}^{(\ell)} \boldsymbol{W}^V \mathbf{\Psi}_s^{(\ell)}, \qquad \beta_{sk}^{(\ell)} = \alpha_{sk}^{(\ell)} \tilde{r}_k^{-(2\ell+1)} \, \boldsymbol{\rho}_\theta^{(\ell)}\big(\boldsymbol{P}(\|\boldsymbol{r}_{sk}\|)\big),$$

where $\beta_{sk}^{(\ell)}$ is invariant because it depends only on radii, distances, and invariant features. Channelwise scalar multiplication is implicit, and $\boldsymbol{W}^V$ acts within the same irrep degree. Hence

$$\mathbf{\Psi}_{\mathrm{out},k}^{(\ell)} \mapsto \sum_s \beta_{sk}^{(\ell)} \boldsymbol{W}^V \boldsymbol{D}^{(\ell)}(R) \mathbf{\Psi}_s^{(\ell)} = \boldsymbol{D}^{(\ell)}(R) \mathbf{\Psi}_{\mathrm{out},k}^{(\ell)}.$$

The subsequent broadcast to atoms is an L2P translation built from spherical harmonics and Clebsch–Gordan products, which is equivariant by the same argument as the analytical equivariant FMM operator. Therefore the complete Kelvin reaction branch—pooling to translation-equivariant supports, multiplying each irrep by the Kelvin radial response, applying degree-preserving channel mixing, and translating back to particles—commutes with the $\mathrm{SO}(3)$ action. The node-local self-Kelvin update

$$\tilde{\mathbf{\Psi}}_u^{(\ell)} = \mathbf{\Psi}_u^{(\ell)} + \tilde{R}_u^{-(2\ell+1)} \mathbf{\Psi}_u^{(\ell)}, \qquad \tilde{R}_u = R_u / r_{\mathrm{cut}}$$

is the degenerate case of the same proof: it is a sum of a degree-ℓ tensor and an invariant scalar multiple of that tensor, and is therefore degree-ℓ equivariant.

Linear-time complexity. The linear evaluation bound assumes a hierarchy with $\mathcal{O}(N)$ nodes and edges, bounded far-field and near-leaf interaction lists, and fixed angular degrees, channel counts, radial basis size, and contraction order. These are the finite-capacity and geometric conditions underlying the FMM scaling argument (Cheng et al., 1999); a separation test alone does not establish bounded lists for every adaptive hierarchy. Under these assumptions, each retained translation or node-local contraction has a cost independent of N, giving $\mathcal{O}(N)$ work and feature storage. The dependence on angular truncation and channel count is absorbed into this constant; generic CG contractions are not assumed to cost $\mathcal{O}(p^2)$. The global Kelvin branch costs $\mathcal{O}(NK)$ for fixed anchor count K, while the root image correction costs $\mathcal{O}(|\mathcal{N}_{\mathrm{lat}}|)$ before linear-cost reinjection. These bounds concern operator evaluation and storage. The cost of constructing and rebuilding the hierarchy depends on the partitioning implementation and is assessed separately by runtime measurements.

Learnable multipole operator gate. Equation 7 gives one learned stage. Hierarchical evaluation applies these stages successively, with independent parameters and an independent gate at each stage τ. The stage displacement enters the regular numerator $\|\boldsymbol{r}_<\|^\ell$ for P2M/M2M/L2L/L2P and the irregular denominator $\|\boldsymbol{r}_>\|^{\ell+1}$ for M2L, with radial smoothing where specified, and the modulation $\boldsymbol{\rho}^{(\ell)}_{\theta,\tau}$ acts on the radial basis of the stage displacement. One output-irrep block of the ungated translation is

$$\boldsymbol{T}^{(\ell_M)}_{st} = \Big[\boldsymbol{W}^V_\tau \boldsymbol{\Psi}_s \otimes_{\theta_\tau} \bigoplus_{\ell=0}^{p} \frac{\|\boldsymbol{r}_<\|^\ell}{\|\boldsymbol{r}_>\|^{\ell+1}} \boldsymbol{\rho}^{(\ell)}_{\theta,\tau}\big(\boldsymbol{R}(\|\boldsymbol{r}_{st}\|)\big)\, \boldsymbol{Y}^{(\ell)}(\hat{\boldsymbol{r}}_{st})\Big]^{(\ell_M)}, \tag{19}$$

The analytical stages fix the path weights and couple each harmonic only to the transported state. Learned stages may also weight other allowed paths, and intermediate gates depend on intermediate features, so a composition of learned stages is not restricted to the analytical coupling. Suppressing τ and channel indices, each stage applies

$$\boldsymbol{\Psi}^{(\ell_M)}_t = \sum_{s\in\mathcal{N}_t} \alpha^{(\ell_M)}_{\theta,st} \boldsymbol{T}^{(\ell_M)}_{st}, \qquad \alpha^{(\ell_M)}_{\theta,st} = \boldsymbol{\phi}_\theta\big(\mathcal{G}_{st} \oplus \mathcal{M}_{st} \oplus \mathcal{A}_{st}\big)^{(\ell_M)}. \tag{20}$$

The gate $\alpha^{(\ell_M)}_{\theta,st}$ is an $\mathrm{SO}(3)$-invariant scalar: $\mathcal{G}_{st}$ contains distances and other geometric invariants, $\mathcal{M}_{st}$ contains block norms of source and target irreps, and $\mathcal{A}_{st}$ contains invariant contractions of projected source and target tensors. Thus, multiplication by $\alpha^{(\ell_M)}_{\theta,st}$ preserves the irrep type of $\boldsymbol{T}^{(\ell_M)}_{st}$ and cannot break equivariance:

$$\alpha^{(\ell_M)}_{\theta,st} \boldsymbol{T}^{(\ell_M)}_{st} \mapsto \alpha^{(\ell_M)}_{\theta,st} \boldsymbol{D}^{(\ell_M)}(R) \boldsymbol{T}^{(\ell_M)}_{st} = \boldsymbol{D}^{(\ell_M)}(R)\big(\alpha^{(\ell_M)}_{\theta,st} \boldsymbol{T}^{(\ell_M)}_{st}\big). \tag{21}$$

The gate therefore has the same algebraic role as a classical scalar damping, screening, or polarization prefactor: it changes the strength of a multipole translation without changing its tensor transformation law.

A useful physical analogy is radial damping. Classical models regularize singular inverse-power interactions at short separation by replacing r^{-n} with a damped kernel,

$$r^{-n} \quad \longmapsto \quad f_n(br)\, r^{-n}, \qquad f_n(br) = 1 - e^{-br} \sum_{k=0}^{n} \frac{(br)^k}{k!}, \tag{22}$$

as in Tang–Toennies damping (Tang & Toennies, 1984). Suppressing alignment inputs for this comparison, the learned gate has the analogous multiplicative role:

$$f_n(br) \quad \rightsquigarrow \quad \alpha^{(\ell_M)}_{\theta,st} = \boldsymbol{\phi}_\theta\big(\mathcal{G}_{st} \oplus \mathcal{M}_{st}\big)^{(\ell_M)}. \tag{23}$$

Distances enter through $\mathcal{G}_{st}$, and feature magnitudes $\mathcal{M}_{st}$ allow the modulation to depend on the source and target environments. The output degree ℓ_M permits different responses in different tensor channels. This is an analogy in operator structure: learned feature norms are not physical hardness parameters, and the angular degree does not determine the damping order n. Log-compressing feature norms controls their dynamic range; it does not by itself establish extensivity or impose the analytical short-distance limit of f_n.

A second connection is induced polarization. In a polarizable multipole model, a local field induces a moment

$$\boldsymbol{\mu}^{\mathrm{ind}}_t = \boldsymbol{\alpha}_t \boldsymbol{E}_t, \qquad E_{\mathrm{pol},t} = -\frac{1}{2}\boldsymbol{\mu}^{\mathrm{ind}}_t \cdot \boldsymbol{E}_t = -\frac{1}{2}\boldsymbol{E}_t^\top \boldsymbol{\alpha}_t \boldsymbol{E}_t, \tag{24}$$

with analogous tensor contractions for higher multipoles (Ponder et al., 2010; Stone, 2013). These expressions motivate using invariant angular contractions to modulate tensorial messages. NEMORA constructs the alignment features

$$\mathcal{A}_{st} = \bigoplus_{\ell} \left\langle \boldsymbol{q}_t^{(\ell)}, \boldsymbol{k}_s^{(\ell)} \right\rangle, \qquad \boldsymbol{q}_t^{(\ell)} = \boldsymbol{W}_Q \boldsymbol{\Psi}_t^{(\ell)}, \qquad \boldsymbol{k}_s^{(\ell)} = \boldsymbol{W}_K \boldsymbol{\Psi}_s^{(\ell)}. \tag{25}$$

The query and key transform in the same irrep, so their inner product is invariant and records their relative alignment. They are learned features, without an imposed identification with a physical susceptibility or electric field. Passing $\mathcal{A}_{st}$ through the gate gives

$$\alpha_{\theta,st}^{(\ell_M)} = \phi_\theta\big(\mathcal{G}_{st} \oplus \mathcal{M}_{st} \oplus \mathcal{A}_{st}\big)^{(\ell_M)}, \tag{26}$$

so the transmitted message is amplified or suppressed according to the alignment of the source and target features while preserving equivariance.

Classical polarizable force fields may solve self-consistently for induced moments (Ponder et al., 2010). NEMORA instead learns a feed-forward, environment-dependent modulation from its training targets; the gate alone does not enforce those self-consistency equations. Setting $\alpha_{\theta,st}^{(\ell_M)} \equiv 1$ in Equation 20 removes the state-dependent modulation. Recovering analytical FMM additionally requires its fixed translation coefficients, normalizations, and aggregation conventions (Appendix E).