Transferable long-range interactions in machine-learned interatomic potentials

Viktor Zaverkin[1,*] **Matheus Ferraz**[2] **Francesco Alesiani**[1] **Henrik Christiansen**[1]
Makoto Takamoto[1] **Federico Errica**[3] **Mathias Niepert**[1,4]
[1]NEC Laboratories Europe [2]NEC OncoImmunity AS
[3]NEC Italy [4]University of Stuttgart

Abstract

Machine-learned interatomic potentials (MLIPs) enable efficient large-scale atomistic simulations with near first-principles accuracy. However, their limited ability to model long-range dispersion and electrostatics often reduces accuracy in many practical applications. Achieving consistent modeling of these interactions is also crucial for ensuring transferability between molecular and bulk systems. In this work, we present a framework that augments local MLIPs with analytic long-range electrostatic and dispersion terms parameterized through latent partial charges learned directly from total energies and forces. By ensuring a consistent treatment of long-range interactions across molecular and bulk systems, the framework enables the improved transferability of MLIPs. We assess the framework's effectiveness on the synthetic point-charge datasets and further demonstrate the improved accuracy and generalization ability of the resulting models on the SPICE-v2 dataset and in molecular dynamics simulations.

1 Introduction

Machine-learned interatomic potentials (MLIPs) have become an essential tool in molecular and materials modeling, offering accuracy close to that of first-principles methods such as density functional theory (DFT) at a fraction of the computational cost. A broad range of MLIPs have been developed in recent years [1–11], with their accuracy and generalization ability determined mainly by the local representations, which depend on configurational (atomic positions) and compositional (atomic types) degrees of freedom. While these approaches have achieved remarkable success across diverse applications [12–14], including recent advances in foundation MLIPs [15–18], their limited ability to capture long-range dispersion and electrostatics often restricts accuracy in systems such as electrolytes, interfacial structures, and biomolecules. A growing body of work has, therefore, focused on augmenting local MLIPs with explicit long-range interactions to improve their accuracy.

Prior work has provided significant advances in physics-based [18–37] and data-driven approaches to modeling long-range interactions [21, 38–43]. However, to the best of our knowledge, no existing framework has been explicitly designed to ensure transferability between molecular and bulk systems. This limitation is becoming increasingly important with the rise of foundation MLIPs, which are trained on large-scale molecular datasets such as SPICE [44, 45] and OMol25 [46]. Existing physics-based methods, while conceptually transferable, often introduce inductive biases that hinder this in practice. For example, some approaches impose specific choices of Ewald-related parameters when training on molecular data [31–33], which are system-dependent and, thus, limit the model's transferability. Others inconsistently modify the short-range component of Coulomb interactions [21] or rely on simple long-range cutoffs [18]. Fully data-driven models offer a conceptually important and promising direction, but typically rely on design choices that do not guarantee consistent transferability [43].

European Conference on Neural Information Processing Systems (EurIPS) Workshop: SimBioChem 2025.

Contributions. In this work, we address the limitations of existing physics-based approaches and demonstrate that, when treated consistently, learned long-range interactions transfer between molecular and bulk systems. In particular, our framework augments local, symmetry-preserving MLIPs with analytic long-range dispersion and electrostatic terms parameterized through latent partial charges learned directly from total energies and forces. In addition, we combine local MLIPs, transferred to bulk systems, with the smooth particle-mesh Ewald (SPME) method [47], an efficient approximation of Ewald summation that can be naturally combined with automatic differentiation through its smooth interpolation scheme and offers tunable interpolation accuracy through the choice of spline order [48]. We demonstrate the framework's transferability using synthetic point-charge datasets and further validate its accuracy and generalization on the SPICE-v2 dataset and in molecular dynamics (MD) simulations.

2 Background

Machine-learned interatomic potentials. We define a structure with N_{at} atoms as $S = \{\mathbf{r}_u, Z_u\}_{u=1}^{N_{\mathrm{at}}}$, where $\mathbf{r}_u \in \mathbb{R}^3$ are Cartesian coordinates and $Z_u \in \mathbb{N}$ is the atomic number of atom u. Provided a structure S, MLIPs approximate the potential energy by learning a mapping $f_{\boldsymbol{\theta}} : S \mapsto E \in \mathbb{R}$ with trainable parameters $\boldsymbol{\theta}$. To ensure linear scaling with system size, the total energy is typically expressed as a sum of site energies for all atoms [1]

$$E\left(S, \boldsymbol{\theta}\right) = \sum\nolimits_{u=1}^{N_{\mathrm{at}}} E_u\left(S_u, \boldsymbol{\theta}\right), \tag{1}$$

where S_u denotes the local environment of atom u within a cutoff radius r_{c}. Atomic forces are computed as negative gradients of the total energy with respect to atomic coordinates, i.e., $\mathbf{F}_u = -\nabla_{\mathbf{r}_u} E$. The trainable parameters $\boldsymbol{\theta}$ are learned from datasets containing total energies and atomic forces obtained by a first-principles method.

Equivariant message-passing models. An MLIP has to satisfy several symmetries. These are the invariances with respect to global rotations, translations, and reflections of a structure, as well as with respect to the exchange of atoms of the same atomic species. The latter requirement is ensured by the site-energy decomposition in Eq. (1), while other symmetries are satisfied by constructing atom-centered representations of atomic environments. The expressive power of these representations depends on capturing symmetries of local environments and many-body correlations [49, 50]. For this reason, equivariant message passing has become a widely used approach [6–8, 10, 11].

In message-passing models, molecular and bulk systems are represented as graphs in a three-dimensional Euclidean space. Atoms are treated as nodes, with an edge $\{u, v\}$ connecting two atoms u and v if their distance is within a cutoff distance r_{c}, i.e., $r_{uv} = \|\mathbf{r}_u - \mathbf{r}_v\|_2 \leq r_{\mathrm{c}}$. Through iterative processing of local information, we learn the representation of atom u, capturing local atomic environments beyond the cutoff radius. Directional information from atomic environments is embedded into higher-dimensional tensor spaces and applies tensor products to compute many-body features [2–5, 7–9, 11].

Long-range dispersion and electrostatics. The energy $E_u\left(S_u, \boldsymbol{\theta}\right)$ in Eq. (1) is determined exclusively by its local environment S_u. As a result, MLIPs typically neglect long-range interactions, which can be crucial in accurate simulations of molecular and materials systems [51, 52]. These interactions are often characterized by a power-law decay $1/r_{uv}^p$, with the most common cases being $p = 1$ for Coulomb electrostatics and $p = 6$ for dispersion interactions. Accurate MLIPs, therefore, need to capture both short-range many-body correlations and long-range effects, whose range can extend beyond the effective cutoff of message-passing architectures.

3 Related work

Physics-based approaches. An established strategy for modeling long-range effects is to augment short-range learned energies with explicit, physics-inspired interaction terms that decay slowly with distance. These terms are often written in pairwise additive form, $E_{\mathrm{lr}} = \sum_{\mathbf{l}}^{*} \sum_{u<v} g\left(\|\mathbf{r}_{uv} + \mathbf{l}\|_2, q_u, q_v\right)$, where g is a physically motivated kernel (e.g., Coulomb or dispersion), q_u are per-atom latent charges predicted from local environments, $\mathbf{l}$ are lattice vectors, and the asterisk denotes omission of self-interaction. Some approaches constrain predicted charges

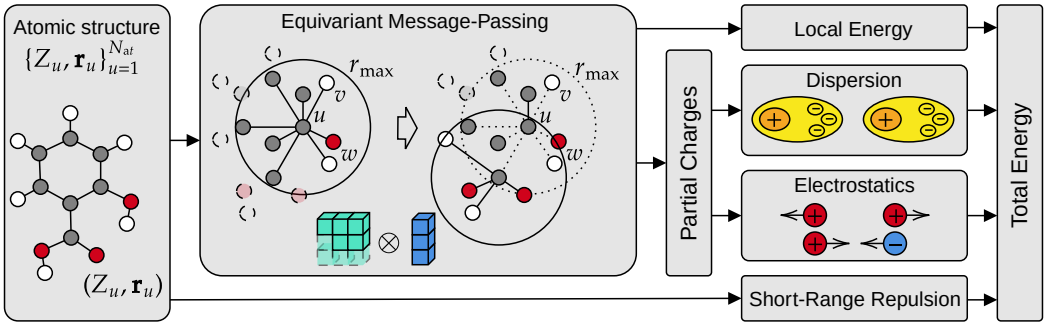

Figure 1: **Schematic illustration of a local MLIP augmented with a charge-readout layer for predicting partial charges.** Learned partial charges are used to compute long-range electrostatic and dispersion contributions to the total energy. A short-range repulsion term is added to ensure the correct asymptotic behavior as $r \to 0$.

using reference partial charges or dipole moments [18, 20, 21, 23, 25, 30], while others relax this constraint [31–33, 36] or instead use kernels directly to build invariant or equivariant long-range features [19, 26, 27, 34]. Charge equilibration schemes have also been proposed to determine q_u, allowing learned charges to capture information beyond local environments [22, 28, 29, 35].

Prior work has not been explicitly designed to transfer between both molecular and bulk systems. For example, a closely related approach employs Ewald summation [31], where the Coulomb kernel is split into real- and reciprocal-space contributions to efficiently evaluate long-range electrostatics. In its original form, however, only the reciprocal component was used, which can introduce inductive biases that hinder transferability. Later work subtracts the real-space contribution when training on molecules [32, 33], but relies on a fixed smearing parameter, which can again restrict transferability. Additionally, no existing approach has systematically extended these ideas beyond the $\mathcal{O}(N_{\text{at}}^{3/2})$ scaling of traditional Ewald summation, despite a growing body of work implementing more efficient methods [53]. For example, particle-mesh methods with $\mathcal{O}(N_{\text{at}} \log N_{\text{at}})$ scaling [47, 54–56], offer a more efficient alternative and can be naturally combined with automatic differentiation if a smooth interpolation scheme is used [48].

Data-driven approaches. Another strategy is to capture long-range effects in an entirely data-driven manner, without introducing explicit physics-inspired corrections or kernels. For example, some approaches compute distance-dependent pairwise interactions between all atoms [38, 39], resulting in an $\mathcal{O}(N_{\text{at}}^2)$ scaling. Others employ self-attention mechanisms to model long-range dependencies, but often neglect explicit distance and directional information [21]. More recent work addresses these limitations but requires numerical symmetrization [43], which may restrict transferability across both molecular and bulk systems. A different class of methods combines local models with spectral filters [41], but at the cost of unfavorable $\mathcal{O}(N_{\text{at}}^3)$ scaling. It is worth noting, however, that approaches inspired by similar ideas but more firmly grounded in physical principles can capture long-range interactions beyond simple power-law decays [40]. Furthermore, recent advances demonstrate that graph models can adaptively learn the required number of message-passing layers for capturing long-range dependencies [42]. Despite being conceptually appealing, fully data-driven strategies continue to face challenges in transferability between non-periodic and periodic systems.

4 Methods

Our approach combines local, symmetry-preserving MLIPs with explicit analytic models of long-range interactions. Local MLIPs, such as MACE [8] and ICTP [11], excel at capturing bonded and short-range non-bonded interactions. However, their finite cutoff can prevent them from describing slowly decaying electrostatic and dispersion contributions to the total energy. To overcome this limitation, we extend these MLIPs by introducing latent partial charges that couple the learned representations of atomic environments to analytic long-range potentials. The resulting model, illustrated schematically in Fig. 1, integrates the strengths of both ML and physics-based approaches, while remaining transferable across molecular and material systems.

4.1 General framework: Equivariant MLIPs with latent partial charges

State-of-the-art MLIPs achieve high accuracy by constructing local atomic representations that respect translation, rotation, reflection, and permutation symmetries. Architectures such as MACE [8] and its Cartesian extension ICTP [11] generate equivariant many-body features of atomic environments by combining learnable radial functions with directional embeddings of pair distance vectors, followed by tensor products of the resulting features. MACE relies on spherical tensors and Clebsch–Gordan products, whereas ICTP uses irreducible Cartesian tensors as an alternative basis. Both approaches yield comparable accuracy and efficiency, and we can use both in a message-passing framework with residual updates. We provide further details on ICTP in Appendix A.

MACE and ICTP yield scalar and tensor-valued features $\mathbf{h}_{uL}^{(t)}$ for each atom u and message-passing step t. Scalar features with $L = 0$ are mapped to site energies $E_u^{(t)}$, and the short-range contribution to the total energy is obtained by summing these energies over all atoms and message-passing layers

$$E_{\text{local}} = \sum\nolimits_{u=1}^{N_{\text{at}}} \sum\nolimits_{t=0}^{T} E_u^{(t)}. \tag{2}$$

We further introduce a charge readout layer, similar to Eq. (A8), that predicts latent partial charges from scalar features

$$\tilde{q}_u = \sum\nolimits_{t=0}^{T} f_q^{(t)}\left(\mathbf{h}_{u(L=0)}^{(t)}\right). \tag{3}$$

Here, $f_q^{(t)}$ denotes a readout function that depends on the message-passing step t. To ensure transferability between the Coulomb potential and Ewald or SPME methods, the ML model is constrained to conserve total charge Q. In particular, we uniformly redistribute any residual net charge across all atoms [21]

$$q_u = \tilde{q}_u + \frac{1}{N_{\text{at}}}\left(Q - \sum\nolimits_{v=1}^{N_{\text{at}}} \tilde{q}_v\right). \tag{4}$$

With both site energies and latent charges available, we can define the total energy as

$$E = E_{\text{local}} + E_{\text{rep}} + E_{\text{es}}(\{q_u\}, \{\mathbf{r}_u\}) + E_{\text{disp}}(\{q_u\}, \{\mathbf{r}_u\}), \tag{5}$$

where E_{rep} denotes the short-range repulsion, E_{local} is defined in Eqs. 1 and 2, and E_{es} and E_{disp} are electrostatic and dispersion energies. Contributions from atoms within the cutoff are excluded from E_{es} and E_{disp}, ensuring that short-range interactions are described solely by the MLIP.

Unlike most prior work [18, 20, 21, 23, 30], the learned charges in our model are not trained against reference values or dipole moments. Instead, we treat them as latent variables trained exclusively on reference energies and forces, allowing the model to recover missing long-range contributions while remaining consistent with short-range energies. This approach is related to earlier work [31], but we avoid using periodic boxes during training on isolated systems. Therefore, we remove inductive biases that hinder the transferability of the resulting models across molecular and bulk systems. Similarly, later work [33] considers electrostatic contributions in isolated systems, but subtracts the real-space contribution of the Ewald method using a fixed smearing parameter, which affects transferability. For more details, we refer to Sections 5 and C.3.

The expression in Eq. (5) offers a modular framework for integrating analytic long-range physics into MLIPs. Since it relies only on scalar features, it is architecture-agnostic. Thus, we can extend any MLIP with a charge readout layer and combine it with analytic long-range terms without modification. The local energy captures bonded and short-range non-bonded interactions, while analytic Coulomb and dispersion terms ensure the correct long-range asymptotics. Latent charges couple the learned local representations to these analytic potentials, ensuring that all contributions are learned jointly during training. Furthermore, excluding interactions within the cutoff from E_{es} and E_{disp} ensures that analytic terms complement rather than compete with the short-range model.

4.2 Long-range dispersion and electrostatics

The following sections detail the specific forms of the dispersion and electrostatic energies. The latent partial charges, introduced in Eq. (4) and learned implicitly during training, are used in the analytic expressions for both terms. This design ensures that long-range interactions are not fixed corrections but remain coupled to the learned representations and optimized jointly with the short-range model.

4.2.1 Dispersion

Van der Waals interactions arise from electron density fluctuations and include short-range repulsion and long-range attraction terms [57, 58]. The latter is referred to as dispersion interactions. A local potential model can learn the short-range repulsion or explicitly include it through an empirical repulsion potential; see Section B. Similarly, a local MLIP with sufficiently large (effective) cutoff can learn two-body dispersion interactions, which decay as r^{-6}. However, including an explicit analytic expression can provide useful inductive biases, such as the correct asymptotics [59].

We explicitly include the analytic two-body term of the D4 dispersion correction [60]

$$E_{\text{disp}} = -\frac{1}{2}\sum_{u=1}^{N_{\text{at}}}\sum_{v\neq u}^{N_{\text{at}}}\sum_{n=6,8} s_n \frac{C_{(n)}^{uv}}{r_{uv}^n} f_{\text{damp}}^{(n)}\left(r_{uv}\right), \tag{6}$$

where s_n are scaling factors and $f_{\text{damp}}^{(n)}$ denotes the Becke–Johnson damping function. Pairwise dispersion coefficients $C_{(n)}^{uv}$ follow earlier work [60] but use learned partial charges. Similar to prior work [21], we treat all density-functional-dependent parameters as learnable and introduce an additional learnable parameter s_q that scales the tabulated reference charges used to determine $C_{(n)}^{uv}$. Unlike that work, partial charges are learned exclusively from reference energies and forces.

Due to the fast decay of two-body dispersion interactions, we truncate them at a cutoff radius and account for this by interpolating the coordination number between its actual value and zero at the cutoff using a switch function [21]. To ensure that the potential and the forces approach zero at the cutoff radius, we modify the two-body dispersion term using the shifted force potential [61]. For bulk systems, we include the periodic images of atoms in the local neighborhood defined by the cutoff.

We further remove the contributions from atoms within the cutoff radius, which are already accounted for by the local energy term. In particular, we subtract

$$-\frac{1}{2}\sum_{u=1}^{N_{\text{at}}}\sum_{v\neq u}^{N_{\text{at}}}\sum_{n=6,8} s_n \frac{C_{(n)}^{uv}}{r_{uv}^n} f_{\text{damp}}^{(n)}\left(r_{uv}\right) f_{\text{c}}\left(r_{uv}\right)$$

from the dispersion energy in Eq. (6).

Many-body effects can result in a slower decay of dispersion interactions that the effective cutoff radius may not capture, and including these effects through known analytic expressions may increase the computational cost of the resulting model. We decided not to account for these effects explicitly, assuming that introducing analytic expressions for electrostatic interactions that decay as r^{-1} and using learnable partial charges expanded into many-body contributions may implicitly capture some portion of these slower-decaying dispersion interactions.

4.2.2 Coulomb potential

We model long-range electrostatic interactions using the Coulomb potential such that the total electrostatic energy is given by

$$E_{\text{es}} = \frac{1}{2}\sum_{u=1}^{N_{\text{at}}}\sum_{v\neq u}^{N_{\text{at}}} \frac{q_u q_v}{r_{uv}}, \tag{7}$$

where q_u and q_v denote learned partial charges and the factor $1/2$ accounts for double-counting pair interactions. For simplicity, we omit the pre-factor $1/\left(4\pi\epsilon_0\right)$, where ϵ_0 is the permittivity of the vacuum. The summation over all atom pairs scales as $\mathcal{O}\left(N_{\text{at}}^2\right)$, making the calculation of the electrostatic energy inefficient for large systems.

Calculating the electrostatic energy in bulk systems is even more challenging

$$E_{\text{es}} = \frac{1}{2}{\sum_{\mathbf{l}}}^{*}\sum_{u,v=1}^{N_{\text{at}}} \frac{q_u q_v}{\left\|\mathbf{r}_{uv}+\mathbf{l}\right\|_2}, \tag{8}$$

where the sum runs over all atom pairs, including $u = v$, and lattice vectors $\mathbf{l}$. The asterisk $*$ indicates that terms where $\left\|\mathbf{r}_{uv}+\mathbf{l}\right\|_2 = 0$ are excluded. Due to the slow decay of the Coulomb potential, this sum is only conditionally convergent. Therefore, the following sections discuss Ewald summation [62, 63], one of the main approaches for addressing this problem, as well as more computationally efficient mesh-based approaches [47, 54–56, 64].

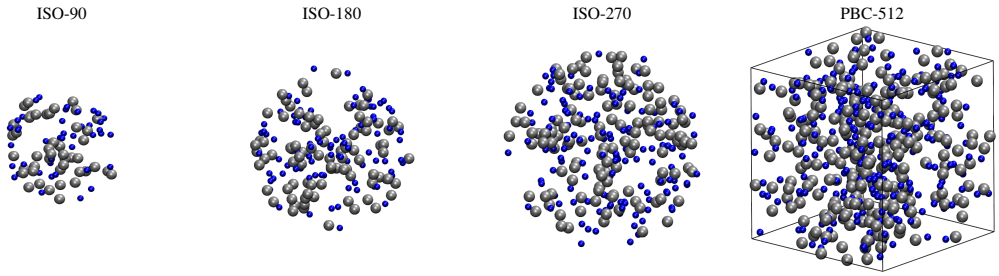

Figure 2: **Representative configurations from the isolated (ISO-90, ISO-180, ISO-270) and periodic (PBC-512) datasets.** These datasets are used to assess models' transferability between isolated and periodic systems.

We ensure the transferability of resulting MLIPs across both molecular clusters and bulk systems. In particular, we include electrostatic interactions for molecular clusters using Eq. (7) and use Ewald sum or mesh-based approaches for Eq. (8) for bulk systems. Furthermore, prior work [65] has demonstrated the transferability of local energies, which we now apply to learned partial charges.

As for the dispersion energy, we exclude the contributions from atoms within the cutoff by subtracting

$$\begin{cases} \frac{1}{2}\sum_{u=1}^{N_{\mathrm{at}}}\sum_{v\neq u}^{N_{\mathrm{at}}}\frac{q_u q_v}{r_{uv}} f_{\mathrm{c}}\left(r_{uv}\right), & \text{for isolated molecules,} \\ \frac{1}{2}\sum_{\mathbf{l}}^{*}\sum_{u,v=1}^{N_{\mathrm{at}}}\frac{q_u q_v}{\|\mathbf{r}_{uv}+\mathbf{l}\|_2} f_{\mathrm{c}}\left(r_{uv}\right), & \text{for bulk systems.} \end{cases}$$

4.2.3 Ewald summation

Ewald summation decomposes the sum in Eq. (8) into two separate terms

$$E_{\mathrm{es}} = \frac{1}{2}\sum_{\mathbf{l}}^{*}\sum_{u,v=1}^{N_{\mathrm{at}}}\frac{q_u q_v \phi\left(r_{uv}\right)}{\|\mathbf{r}_{uv}+\mathbf{l}\|_2} + \frac{1}{2}\sum_{\mathbf{l}}^{*}\sum_{u,v=1}^{N_{\mathrm{at}}}\frac{q_u q_v \left(1-\phi\left(r_{uv}\right)\right)}{\|\mathbf{r}_{uv}+\mathbf{l}\|_2} \tag{9}$$

The convergence function $\phi\left(r\right) = \mathrm{erfc}\left(\alpha r\right)$, with parameter α, is chosen to be smooth for small r and decay rapidly as $r \to \infty$. This choice ensures fast convergence of the first sum in real space and efficient evaluation of the second sum through Fourier transformation.

Splitting the total electrostatic energy in Eq. (9) yields the general expression

$$E_{\mathrm{es}} = E_{\mathrm{es}}^{\mathrm{r}} + E_{\mathrm{es}}^{\mathrm{f}} + E_{\mathrm{es}}^{\mathrm{self}}, \tag{10}$$

$$E_{\mathrm{es}}^{\mathrm{r}} = \frac{1}{2}\sum_{\mathbf{l}}^{*}\sum_{u,v=1}^{N_{\mathrm{at}}}\frac{q_u q_v \phi\left(r_{uv}\right)}{\|\mathbf{r}_{uv}+\mathbf{l}\|_2}, \tag{11}$$

$$E_{\mathrm{es}}^{\mathrm{f}} = \frac{2\pi}{V}\sum_{\mathbf{k}\neq 0}\frac{1}{k^2}\exp\left(-\frac{k^2}{4\alpha^2}\right)\left[S\left(\mathbf{k}\right)S\left(-\mathbf{k}\right)\right], \tag{12}$$

$$E_{\mathrm{es}}^{\mathrm{self}} = -\frac{\alpha}{\sqrt{\pi}}\sum_{u=1}^{N_{\mathrm{at}}} q_u^2. \tag{13}$$

Here, V is the unit cell volume, $k = \|\mathbf{k}\|_2$ is the norm of the reciprocal space vector $\mathbf{k}$, and $S\left(\mathbf{k}\right) = \sum_u q_u \exp\left(i\mathbf{k}\cdot\mathbf{r}_u\right)$ is the structure factor. The self-energy term $E_{\mathrm{es}}^{\mathrm{self}}$ corrects for the interaction of an atom with itself in $E_{\mathrm{es}}^{\mathrm{f}}$. In this work, we ensure that the bulk systems are electrically neutral, $\sum_u q_u = 0$, allowing us to skip the introduction of a homogeneous background charge.

The real- and reciprocal-space sums are convergent but require real-space and reciprocal-space cutoffs, r_{c} and k_{c}, for their efficient evaluation. These cutoffs introduce errors in energy and force calculations. Therefore, the convergence parameter α and both cutoffs must be chosen appropriately to achieve the desired accuracy. In this work, we determine α and k_{c} from the prescribed tolerance δ and real-space cutoff r_{c} following established practices [48, 66]. While naive Ewald summation scales as $\mathcal{O}\left(N_{\mathrm{at}}^2\right)$, an optimal choice of cutoffs and α reduces it to $\mathcal{O}\left(N_{\mathrm{at}}^{3/2}\right)$.

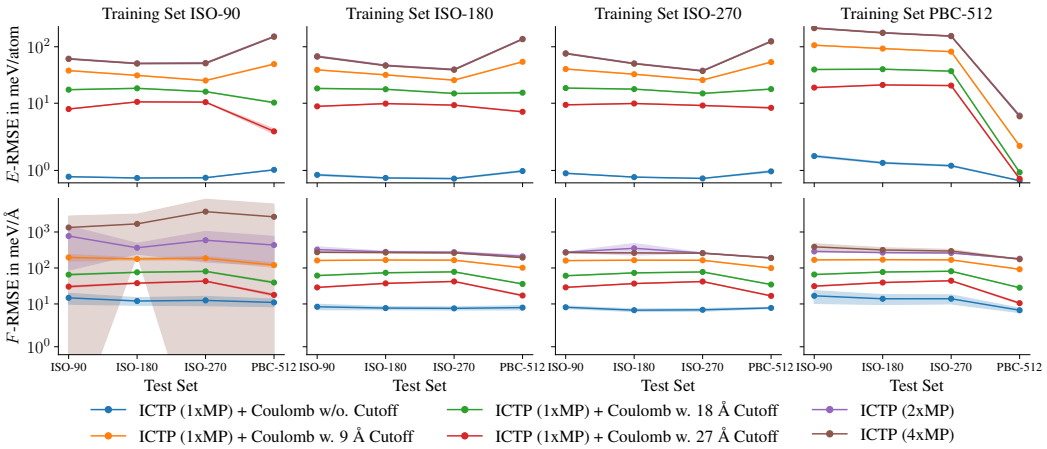

Figure 3: **RMSEs in energies (E) and forces (F) for ICTP models trained with and without explicit long-range electrostatics, and for Coulomb interactions modeled with and without a cutoff.** Models with analytic long-range terms without a cutoff achieve superior transferability across isolated and periodic systems.

4.2.4 Smooth particle mesh Ewald method

For a fixed real-space cutoff, the reciprocal-space sum in Eq. (12) scales as $\mathcal{O}\left(N_{\mathrm{at}}^2\right)$, making the Ewald summation inefficient for large systems. This issue can be addressed by interpolating the charges q_u onto an equally spaced mesh, which enables the use of the Fast Fourier Transform [47, 53–56, 64]. This approach reduces the computational complexity of the reciprocal part of the Ewald sum to $\mathcal{O}\left(N_{\mathrm{at}}\log\left(N_{\mathrm{at}}\right)\right)$. We refer to earlier work [56] for a detailed overview of mesh-based methods. In this work, we employ the smooth particle mesh Ewald (SPME) method [47]. Through a smooth charge interpolation scheme, SPME allows for the analytic differentiation of the energy and can fully exploit the automatic differentiation capabilities of `PyTorch` [48]. The parameters of SPME are determined following established practices [48, 66].

5 Experiments and results

Point-charge dataset. We first evaluate ICTP models with explicit long-range Coulomb interactions on the synthetic point-charge datasets described in Section C.1. We also compare them to ICTP models with truncated Coulomb interactions and to short-range ICTP models with increased message-passing depth. The point-charge datasets enable a systematic assessment of how explicitly modeled long-range electrostatics affects transferability of the potential energy models between isolated and periodic systems. Although these datasets are synthetic and use geometry-independent charges, they are constructed so that electrostatics dominates the energies and forces, allowing us to isolate errors arising from the treatment of Coulomb interactions rather than from the learned partial charges. Representative configurations are shown in Fig. 2. The corresponding energy and force root-mean-square errors (RMSEs) for models trained with and without explicit long-range electrostatics, and for those employing Coulomb interactions with and without a cutoff, are shown in Fig. 3, and the numerical values are provided in Table A2.

Models with an explicit Coulomb term and no cutoff achieve substantially lower errors than short-range ICTP baselines with increased message-passing depth or those using truncated Coulomb interactions, especially when transferred between isolated and periodic systems. Further increasing the number of message-passing layers, or enlarging the cutoff radius in models with truncated Coulomb interactions, yields marginal improvements and does not recover the accuracy of models with unrestricted Coulomb interactions. Even with an excessively large cutoff of 27 Å, the model with truncated Coulomb interactions only partially matches the force accuracy but remains considerably less accurate in energies. Furthermore, such a large cutoff introduces an unfavorable computational trade-off compared to our Ewald- and SPME-based approaches. Adding analytic Coulomb and

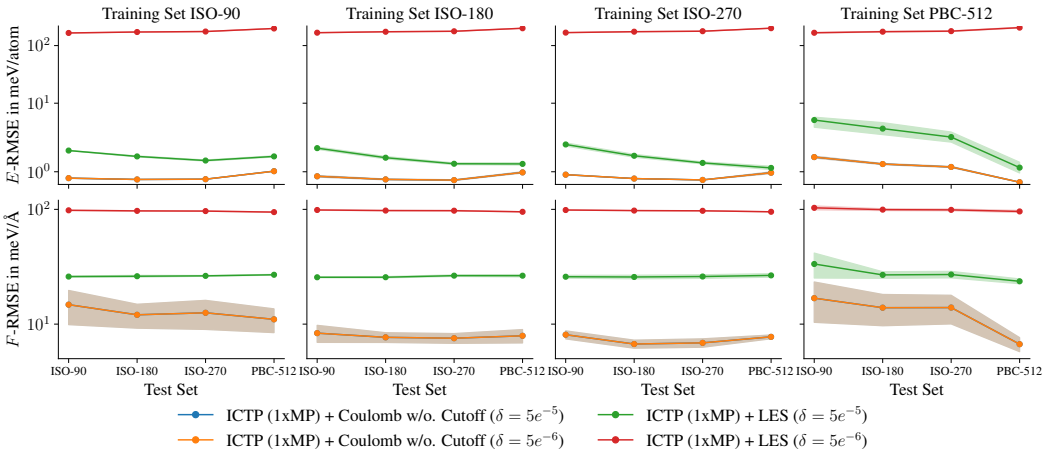

Figure 4: **RMSEs in energies (E) and forces (F) for ICTP models with unrestricted Coulomb interactions and with LES [31–33].** Models with unrestricted Coulomb interactions achieve superior transferability across isolated and periodic systems. Furthermore, once real- and reciprocal-space sums are converged (tight δ), they show no dependence on Ewald/SPME hyperparameters.

dispersion terms also requires only a small charge-readout layer with a negligible number of additional parameters compared to increasing message-passing depth.

Figure 4 further compares ICTP models trained with unrestricted Coulomb interactions to those trained with the latent Ewald summation (LES) method [31–33]. Table A3 provides the corresponding numerical values. LES models long-range electrostatics using

$$E_{\mathrm{es}} = \frac{1}{2} \sum\nolimits_{\mathbf{l}}^{*} \sum\nolimits_{u,v=1}^{N_{\mathrm{at}}} \frac{q_u q_v \left(1 - \phi\left(r_{uv}\right)\right)}{\left\|\mathbf{r}_{uv} + \mathbf{l}\right\|_2},$$

where the asterisk indicates that self-interactions are excluded. The convergence function is defined as $\phi(r) = \mathrm{erfc}(\alpha r)$, with α controlling the splitting between real and reciprocal space. Following the original publications, we model the long-range electrostatics as a real-space summation of the above expression for molecular clusters, and as a reciprocal-space summation for bulk systems, corresponding to the combined contributions of Eqs. 12 and 13. For the short-range energy, we use the ICTP model to ensure that any observed differences arise from long-range interactions.

The choice of α is coupled to the selected real-space cutoff, which we set to 9 Å, consistent with typical values in atomistic simulations. Using a tolerance of 5×10^{-5}, as in all other experiments, this yields $\alpha = 0.337$ Å^{-1}. To assess the robustness of our and the LES framework to small variations in hyperparameters, we also consider a tighter convergence tolerance of 5×10^{-6}, which also results in $\alpha = 0.377$ Å^{-1}. These choices determine the reciprocal-space cutoffs for the Ewald summation, resulting in $k_{\max} = 15$ and $k_{\max} = 18$. For the SPME approach, the corresponding values are $k_{\max} = 32$ and $k_{\max} = 57$. For the computation of all these parameters, we follow established practices [48, 66].

The unrestricted Coulomb model yields the lowest errors, 0.18–2.15 meV/atom for energies and 6.69–16.83 meV/Å for forces, and transfers well between isolated and periodic systems. For LES, when training and testing both use $\delta = 5 \times 10^{-5}$, the corresponding RMSEs are 1.30–5.10 meV/atom and 23.62–33.42 meV/Å. Tightening the tolerance during testing to $\delta = 5 \times 10^{-6}$ is plausible for Ewald/SPME, since δ only controls numerical convergence of the r^{-1} approximation. With $\delta = 5 \times 10^{-6}$ at test time, the results for the model with unrestricted Coulomb interactions remain essentially unchanged, while LES degrades to large, nearly dataset-independent errors of 167.01–204.96 meV/atom and 94.64–103.12 meV/Å.

SPICE-based dataset. Building on previous work [67], we use the earlier short-range ICTP-SR and long-range ICTP-LR models, and introduce an ICTP-LR(D) model that constrains the learned partial charges using reference molecular dipole moments during training. All models are trained on the extended SPICE-v2 dataset [67], which we describe in Section C.2.

Table 1: **RMSEs in the predicted energies/atomic forces for the individual test datasets.** Results are shown for the short-range ICTP-SR model, the long-range ICTP-LR model, and the ICTP-LR(D) model, which constrains learned partial charges using reference dipole moments during training. Energy RMSEs are given in meV/atom and atomic force errors are given in meV/Å.

Dataset	ICTP-LR [67]	ICTP-LR(D)	ICTP-SR [67]
Solvated Amino Acids	1.04/33.11	0.91/33.46	1.50/38.34
Dipeptides	1.29/33.80	1.21/33.41	1.89/35.16
DES370K Monomers	2.04/25.12	1.88/25.80	3.64/25.16
DES370K Dimers	3.17/28.19	2.34/28.20	3.63/29.43
PubChem	3.34/52.34	3.40/52.67	3.56/52.81
Ion Pairs[a]	32.97/79.09	22.50/48.92	143.38/451.87
Solvated PubChem	2.90/66.58	2.85/67.38	3.07/69.12
Water Clusters	0.76/20.53	0.77/19.66	1.13/24.88
Amino Acid Ligand	2.00/32.87	2.11/32.77	2.58/35.52
QMugs	1.15/55.77	1.21/56.23	1.29/55.55
NaCl-Water Clusters	2.11/38.12	1.57/36.48	4.57/53.68
Test-only datasets			
Large Ligands[b]	3.03/104.85	2.91/130.12	4.33/121.96
Small Ligands[b]	7.68/87.68	7.83/83.71	7.83/88.24
Pentapeptides	1.85/41.11	1.78/40.58	4.92/44.31

[a] For Ion Pairs, we report force RMSEs only for the z-component, as the x and y components are zero.
[b] A few outliers affect the reported errors for Large and Small Ligands. Using more robust MAEs reduces the values to 1.47/38.08 and 2.06/37.24 for the ICTP-LR model, respectively. For more details, see Table A4.

Table 1 demonstrates that long-range models consistently outperform the short-range ICTP-SR baseline across all datasets. The largest improvement is observed for Ion Pairs, where ICTP-SR fails to capture the r^{-1} behavior and exhibits errors more than four times larger than those of ICTP-LR. Notable reductions in errors are also observed for Solvated Amino Acids, Water Clusters, and NaCl-Water Clusters, underscoring the importance of explicit long-range interactions in systems with unscreened charges and complex solvation environments. For other molecular systems, such as Dipeptides and Amino Acid–Ligand complexes, the improvements are smaller but systematic, indicating that explicit long-range interactions enhance overall model accuracy across diverse chemical environments. To assess generalization, we tested on SPICE-v2 subsets that were not included in the training set. Both long-range models retained their improved accuracy on molecular systems larger than those seen during training.

The ICTP-LR(D) model, trained with reference dipole moments, achieved slightly lower energy errors but showed no consistent improvement in forces compared to ICTP-LR. This result demonstrates that, even when trained exclusively on total energies and forces, ICTP-LR can recover the missing long-range contributions while remaining consistent with short-range energetics. Molecular-dynamics simulations further revealed differences between the two models. While both performed well for pure water, the ICTP-LR(D) model became unstable after 500–600 ps at low NaCl concentration (0.99 mol/kg) and failed to run at higher concentrations (5.0 mol/kg). In contrast, the ICTP-LR model remained stable in all tested systems [67], suggesting that the specific way dipole information is incorporated in the current training setup may not provide the most suitable inductive bias.

We note that our comparison of ICTP-LR and ICTP-LR(D) models in molecular-dynamics simulations is intended to evaluate how introducing a specific inductive bias, in this case training with reference dipole moments, affects model behavior. A broader analysis of the impact of explicit long-range interactions in bulk simulations is outside the scope of this work. For such analyses, we refer to previous results [67], which also show that the ICTP-SR model with two message-passing layers produces stable molecular-dynamics trajectories and can reproduce properties such as the density and radial distribution functions of NaCl–water mixtures on par with ICTP-LR.

A key advantage of the proposed approach is that the analytic long-range terms introduce negligible computational overhead. Table A5 reports the throughput (in ns/day) and per-atom inference times (μs/atom) for molecular-dynamics simulations with up to 8192 atoms. The cost of ICTP-LR remains within 10 % of the short-range ICTP-SR baseline across all system sizes, and the difference further

diminishes as the system size increases. This demonstrates that explicit electrostatics and dispersion can be included without sacrificing computational efficiency, even for large-scale simulations.

6 Conclusions and limitations

We presented an approach that augments local MLIPs with analytic long-range electrostatic and dispersion terms, enabling consistent modeling of molecular clusters and bulk systems. Systematic evaluations on synthetic point-charge datasets and the extended SPICE-v2 benchmark demonstrate that explicit long-range interactions substantially improve accuracy and transferability while maintaining near-baseline computational efficiency. The method allows models trained only on total energies and atomic forces to recover missing long-range contributions without compromising short-range interaction energies. Moreover, the resulting models yield stable molecular-dynamics simulations and accurately reproduce fundamental properties, such as densities and radial distribution functions [67], highlighting the reliability of the approach in practical simulations.

The proposed framework is modular and architecture-agnostic, offering a practical approach to systematically integrate physically motivated long-range terms into existing MLIPs. The proposed treatment of Coulomb interactions avoids inductive biases introduced by approaches like LES and yields models that remain robust to Ewald hyperparameter choices. Taken together, the work establishes the proposed method as a useful building block for developing more transferable MLIPs and future foundation models applied across molecular and condensed-phase systems.

Limitations. This work employs invariant latent charges, although the framework could, in principle, be extended to incorporate equivariant features. The use of molecular dipole moments introduces further limitations. The current way of incorporating them may not provide an optimal inductive bias, and results may depend on the number of charged systems and the quality of the dipole data in SPICE-v2. Models could also be trained on the atomic dipole moments available in SPICE-v2, which is a different target and may lead to different models' behavior in molecular-dynamics simulations. Moreover, we apply the framework only to ICTP models and leave applications to other architectures for future studies. Finally, a Coulomb pair potential alone cannot capture all long-range effects, requiring more advanced approaches.

Data availability

The training and test datasets will be made available at `https://10.5281/zenodo.17652034` upon publication.

Code availability

The ICTP source code is available on GitHub and can be accessed via this link: `https://github.com/nec-research/ictp`.

Acknowledgements

MN acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 – 390740016 and the Stuttgart Center for Simulation Science (SimTech).

References

[1] J. Behler and M. Parrinello: *Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces*. Phys. Rev. Lett. **98**, 146401 (2007)

[2] A. V. Shapeev: *Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials*. Multiscale Model. Simul. **14**, 1153–1173 (2016)

[3] R. Drautz: *Atomic cluster expansion for accurate and transferable interatomic potentials*. Phys. Rev. B **99**, 014104 (2019)

[4] V. Zaverkin and J. Kästner: *Gaussian Moments as Physically Inspired Molecular Descriptors for Accurate and Scalable Machine Learning Potentials*. J. Chem. Theory Comput. **16**, 5410–5421 (2020)

[5] V. Zaverkin, D. Holzmüller, I. Steinwart, and J. Kästner: *Fast and Sample-Efficient Interatomic Neural Network Potentials for Molecules and Materials Based on Gaussian Moments*. J. Chem. Theory Comput. **17**, 6658–6670 (2021)

[6] K. T. Schütt, O. T. Unke, and M. Gastegger: *Equivariant message passing for the prediction of tensorial properties and molecular spectra*. Int. Conf. Mach. Learn. **139**, 9377–9388 (2021)

[7] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa et al.: *E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials*. Nat. Commun. **13**, 2453 (2022)

[8] I. Batatia, D. P. Kovacs, G. N. C. Simm, C. Ortner, and G. Csanyi: *MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields*. Adv. Neural Inf. Process. Syst. **35**, 11423–11436 (2022)

[9] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen et al.: *Learning local equivariant representations for large-scale atomistic dynamics*. Nat. Commun. **14**, 579 (2023)

[10] G. Simeon and G. D. Fabritiis: *TensorNet: Cartesian Tensor Representations for Efficient Learning of Molecular Potentials*. Adv. Neural Inf. Process. Syst. **36**, 37334–37353 (2023)

[11] V. Zaverkin, F. Alesiani, T. Maruyama, F. Errica, H. Christiansen et al.: *Higher-Rank Irreducible Cartesian Tensors for Equivariant Message Passing*. Adv. Neural Inf. Process. Syst. **37**, 124025–124068 (2024)

[12] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti et al.: *Origins of structural and electronic transitions in disordered silicon*. Nature **589** (7840), 59–64 (2021)

[13] V. Kapil, C. Schran, A. Zen, J. Chen, C. J. Pickard et al.: *The first-principles phase diagram of monolayer nanoconfined water*. Nature **609**, 512–516 (2022)

[14] K. Gubaev, V. Zaverkin, P. Srinivasan, A. I. Duff, J. Kästner et al.: *Performance of two complementary machine-learned potentials in modelling chemically complex systems*. npj Comput. Mater. **9**, 129 (2023)

[15] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács et al.: *A foundation model for atomistic materials chemistry*. `https://arxiv.org/abs/2401.00096` (2023)

[16] H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi et al.: *MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures*. `https://arxiv.org/abs/2405.04967` (2024)

[17] D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton et al.: *MACE-OFF: Short-Range Transferable Machine Learning Force Fields for Organic Molecules*. J. Am. Chem. Soc. **147**, 17598–17611 (2025)

[18] A. Kabylda, J. T. Frank, S. Suárez-Dou, A. Khabibrakhmanov, L. Medrano Sandonas et al.: *Molecular Simulations with a Pretrained Neural Network and Universal Pairwise Force Fields*. J. Am. Chem. Soc. **147**, 33723–33734 (2025)

[19] A. Grisafi and M. Ceriotti: *Incorporating long-range physics in atomic-scale machine learning*. J. Chem. Phys. **151**, 204105 (2019)

[20] O. T. Unke and M. Meuwly: *PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges*. J. Chem. Theory Comput. **15** (6), 3678–3693 (2019)

[21] O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda et al.: *SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects*. Nat. Commun. **12** (1), 7273 (2021)

[22] T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler: *A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer*. Nat. Commun. **12**, 398 (2021)

[23] J. Behler: *Four Generations of High-Dimensional Neural Network Potentials*. Chem. Rev. **121**, 10037–10072 (2021)

[24] A. Gao and R. C. Remsing: *Self-consistent determination of long-range electrostatics in neural network potentials*. Nat. Commun. **13**, 1572 (2022)

[25] T. Plé, L. Lagardère, and J.-P. Piquemal: *Force-field-enhanced neural network interactions: from local equivariant embedding to atom-in-molecule properties and long-range effects*. Chem. Sci. **14**, 12554–12569 (2023)

[26] K. K. Huguenin-Dumittan, P. Loche, N. Haoran, and M. Ceriotti: *Physics-Inspired Equivariant Descriptors of Nonbonded Interactions*. J. Phys. Chem. Lett. **14** (43), 9612–9618 (2023)

[27] A. Kosmala, J. Gasteiger, N. Gao, and S. Günnemann: *Ewald-based Long-Range Message Passing for Molecular Graphs*. Int. Conf. Mach. Learn. **202**, 17544–17563 (2023)

[28] M. Gubler, J. A. Finkler, M. R. Schäfer, J. Behler, and S. Goedecker: *Accelerating Fourth-Generation Machine Learning Potentials Using Quasi-Linear Scaling Particle Mesh Charge Equilibration*. J. Chem. Theory Comput. **20**, 7264–7271 (2024)

[29] Y. Shaidu, F. Pellegrini, E. Küçükbenli, R. Lot, and S. de Gironcoli: *Incorporating long-range electrostatics in neural network potentials via variational charge equilibration from shortsighted ingredients*. npj Comput. Mater. **10**, 47 (2024)

[30] O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel et al.: *Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments*. Sci. Adv. **10** (14), eadn4397 (2024)

[31] B. Cheng: *Latent Ewald summation for machine learning of long-range interactions*. npj Comput. Mater. **11**, 80 (2025)

[32] D. S. King, D. Kim, P. Zhong, and B. Cheng: *Machine learning of charges and long-range interactions from energies and forces*. Nat. Commun. **16**, 8763 (2025)

[33] D. Kim, X. Wang, P. Zhong, D. S. King, T. J. Inizan et al.: *A universal augmentation framework for long-range electrostatics in machine learning interatomic potentials*. 2507.14302 (2025)

[34] E. Rumiantsev, M. F. Langer, T.-E. Sodjargal, M. Ceriotti, and P. Loche: *Learning Long-Range Representations with Equivariant Messages*. 2507.19382 (2025)

[35] E. Kocer, A. Singraber, J. A. Finkler, P. Misof, T. W. Ko et al.: *Iterative charge equilibration for fourth-generation high-dimensional neural network potentials*. J. Chem. Phys. **162**, 124106 (2025)

[36] Y. Ji, J. Liang, and Z. Xu: *Machine-Learning Interatomic Potentials for Long-Range Systems*. Phys. Rev. Lett. **135**, 178001 (2025)

[37] J. Thomas, W. Baldwin, G. Csanyi, and C. Ortner: *Self-consistent Coulomb interactions for machine learning interatomic potentials*. Nonlinearity **38**, 095024 (2025)

[38] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt et al.: *Machine learning of accurate energy-conserving molecular force fields*. Sci. Adv. **3**, e1603015 (2017)

[39] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko: *Towards exact molecular dynamics simulations with machine-learned force fields*. Nat. Commun. **9**, 3887 (2018)

[40] I. Batatia, L. L. Schaaf, H. Chen, G. Csányi, C. Ortner et al.: *Equivariant Matrix Function Neural Networks*. Int. Conf. Learn. Represent. https://arxiv.org/abs/2310.10434 (2024)

[41] S. Geisler, A. Kosmala, D. Herbst, and S. Günnemann: *Spatio-Spectral Graph Neural Networks*. Adv. Neural Inf. Process. Syst. **37**, 49022–49080 (2024)

[42] F. Errica, H. Christiansen, V. Zaverkin, T. Maruyama, M. Niepert et al.: *Adaptive Message Passing: A General Framework to Mitigate Oversmoothing, Oversquashing, and Underreaching*. 2312.16560 (2025)

[43] J. T. Frank, S. Chmiela, K.-R. Müller, and O. T. Unke: *Euclidean Fast Attention – Machine Learning Global Atomic Representations at Linear Cost*. 2412.08541 (2025)

[44] P. Eastman, P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr et al.: *SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials*. Sci. Data **10** (1), 11 (2023)

[45] P. Eastman, B. P. Pritchard, J. D. Chodera, and T. E. Markland: *Nutmeg and SPICE: Models and Data for Biomolecular Machine Learning*. J. Chem. Theory Comput. **20** (19), 8583–8593 (2024)

[46] D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, M. G. Taylor, M. R. Hasyim et al.: *The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models*. https://arxiv.org/abs/2505.08762 (2025)

[47] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee et al.: *A smooth particle mesh Ewald method*. J. Chem. Phys. **103** (19), 8577–8593 (1995)

[48] H. Christiansen, T. Maruyama, F. Errica, V. Zaverkin, M. Takamoto et al.: *Fast, modular, and differentiable framework for machine learning-enhanced molecular simulations*. J. Chem. Phys. **163**, 182501 (2025)

[49] S. N. Pozdnyakov, M. J. Willatt, A. P. Bartók, C. Ortner, G. Csányi et al.: *Incompleteness of Atomic Structure Representations*. Phys. Rev. Lett. **125**, 166001 (2020)

[50] C. K. Joshi, C. Bodnar, S. V. Mathis, T. Cohen, and P. Lio: *On the Expressive Power of Geometric Graph Neural Networks*. Int. Conf. Learn. Represent. https://arxiv.org/abs/2301.09308 (2023)

[51] S. Yue, M. C. Muniz, M. F. Calegari Andrade, L. Zhang, R. Car et al.: *When do short-range atomistic machine-learning models fall short?* J. Chem. Phys. **154**, 034111 (2021)

[52] S. P. Niblett, M. Galib, and D. T. Limmer: *Learning intermolecular forces at liquid–vapor interfaces*. J. Chem. Phys. **155**, 164101 (2021)

[53] P. Loche, K. K. Huguenin-Dumittan, M. Honarmand, Q. Xu, E. Rumiantsev et al.: *Fast and flexible long-range models for atomistic machine learning*. 2412.03281 (2025)

[54] R. W. Hockney and J. W. Eastwood: *Computer Simulation Using Particles*. CRC Press, 1st edition (1988)

[55] T. Darden, D. York, and L. Pedersen: *Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems*. J. Chem. Phys. **98** (12), 10089–10092 (1993)

[56] M. Deserno and C. Holm: *How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines*. J. Chem. Phys. **109** (18), 7678–7693 (1998)

[57] J. Hermann, R. A. J. DiStasio, and A. Tkatchenko: *First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications*. Chem. Rev. **117** (6), 4714–4758 (2017)

[58] D. M. Anstine and O. Isayev: *Machine Learning Interatomic Potentials and Long-Range Physics*. J. Phys. Chem. A **127** (11), 2417–2431 (2023)

[59] M. Esders, T. Schnake, J. Lederer, A. Kabylda, G. Montavon et al.: *Analyzing Atomic Interactions in Molecules as Learned by Neural Networks*. J. Chem. Theory Comput. **21** (2), 714–729 (2025)

[60] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher et al.: *A generally applicable atomic-charge dependent London dispersion correction*. J. Chem. Phys. **150** (15), 154122 (2019)

[61] C. J. Fennell and J. D. Gezelter: *Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics*. J. Chem. Phys. **124** (23), 234104 (2006)

[62] P. P. Ewald: *Die Berechnung optischer und elektrostatischer Gitterpotentiale*. Ann. Phys. **369** (3), 253–287. https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19213690304 (1921)

[63] B. A. Wells and A. L. Chaffee: *Ewald Summation for Molecular Simulations*. J. Chem. Theory Comput. **11** (8), 3684–3695 (2015)

[64] Darden, TA, Toukmaji, A, and Pedersen, LG: *Long-range electrostatic effects in biomolecular simulations*. J. Chim. Phys. **94**, 1346–1364 (1997)

[65] V. Zaverkin, D. Holzmüller, R. Schuldt, and J. Kästner: *Predicting properties of periodic systems from cluster data: A case study of liquid water*. J. Chem. Phys. **156** (11), 114103 (2022)

[66] P. Eastman, R. Galvelis, R. P. Peláez, C. R. A. Abreu, S. E. Farr et al.: *OpenMM 8: Molecular dynamics simulation with machine learning potentials*. J. Phys. Chem. B **128** (1), 109–116 (2023)

[67] V. Zaverkin, M. Ferraz, F. Alesiani, and M. Niepert: *Performance of universal machine-learned potentials with explicit long-range interactions in biomolecular simulations*. 2508.10841 (2025)

[68] D. R. Lehman and W. C. Parke: *Angular reduction in multiparticle matrix elements*. J. Math. Phys. **30**, 2797–2806 (1989)

[69] J. Gasteiger, J. Groß, and S. Günnemann: *Directional Message Passing for Molecular Graphs*. Int. Conf. Learn. Represent. https://arxiv.org/abs/2003.03123 (2020)

[70] K. He, X. Zhang, S. Ren, and J. Sun: *Deep Residual Learning for Image Recognition*. IEEE Conf. Comput. Vis. Pattern Recognit. https://doi.org/10.1109/CVPR.2016.90 (2016)

[71] S. Elfwing, E. Uchibe, and K. Doya: *Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning*. Neural Netw. **107**, 3–11 (2018)

[72] P. Ramachandran, B. Zoph, and Q. V. Le: *Searching for Activation Functions* (2017)

[73] J. F. Ziegler and J. P. Biersack: *The Stopping and Range of Ions in Matter*, pp. 93–129. Springer US, Boston, MA (1985)

[74] P. Ramachandran, B. Zoph, and Q. V. Le: *Searching for activation functions*. Int. Conf. Learn. Represent. https://arxiv.org/abs/1710.05941 (2018)

[75] S. J. Reddi, S. Kale, and S. Kumar: *On the Convergence of Adam and Beyond*. Int. Conf. Learn. Represent. https://arxiv.org/abs/1904.09237 (2018)

[76] C. Isert, K. Atz, J. Jiménez-Luna, and G. Schneider: *QMugs, quantum mechanical properties of drug-like molecules*. Sci. Data **9** (1), 273 (2022)

[77] D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber et al.: *PSI4 1.4: Open-source software for high-throughput quantum chemistry*. J. Chem. Phys. **152** (18), 184108 (2020)

[78] A. Najibi and L. Goerigk: *The Nonlocal Kernel in van der Waals Density Functionals as an Additive Correction: An Extensive Analysis with Special Emphasis on the B97M-V and ωB97M-V Approaches*. J. Chem. Theory Comput. **14** (11), 5725–5738 (2018)

[79] F. Weigend and R. Ahlrichs: *Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy*. Phys. Chem. Chem. Phys. **7** (18), 3297–3305 (2005)

[80] D. Rappoport and F. Furche: *Property-optimized Gaussian basis sets for molecular response calculations*. J. Chem. Phys. **133** (13), 134105 (2010)

[81] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg: *A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu*. J. Chem. Phys. **132** (15), 154104 (2010)

[82] S. Grimme, S. Ehrlich, and L. Goerigk: *Effect of the damping function in dispersion corrected density functional theory*. J. Comput. Chem. **32** (7), 1456–1465 (2011)

[83] C. Schran, F. L. Thiemann, P. Rowe, E. A. Müller, O. Marsalek et al.: *Machine learning potentials for complex aqueous systems made simple*. Proc. Natl. Acad. Sci. **118** (38), e2110077118 (2021)

Appendices

Appendix Contents

A Irreducible Cartesian tensor potential

The ICTP model has been presented as an extension of the state-of-the-art MACE architecture to the Cartesian basis [8, 11]. Thus, in ICTP, invariant many-body features are computed similarly to MACE. However, instead of using spherical tensors to represent atomic features and local environments and defining (higher-body-order) tensor products with (generalized) Clebsch–Gordan coefficients, we employ irreducible Cartesian tensors and their products. ICTP models are a promising alternative to equivariant models based on spherical tensors, achieving comparable prediction accuracy to leading state-of-the-art approaches [11].

A.1 Irreducible Cartesian tensors and tensor product

The two key operations required for constructing equivariant models are embedding unit directional vectors and the tensor product, which combines two geometric features in an equivariant manner to create new ones. The following offers a high-level overview of these operations. We refer to the original publication for more details on irreducible Cartesian tensors and tensor products, including their use in equivariant message-passing architectures [11].

We embed a unit vector $\hat{\mathbf{r}} \in \mathbb{R}^3$ into the tensor space of maximal rank $l_{\max}$ using irreducible Cartesian tensors $\mathbf{T}_l(\hat{\mathbf{r}}) \in \left(\mathbb{R}^3\right)^{\otimes l}$, with l dimensions, each of size 3. The corresponding tensor is defined as [68]

$$\mathbf{T}_l(\hat{\mathbf{r}}) = \sum_{m=0}^{\lfloor l/2 \rfloor} C_{l,m} \big\{ \hat{\mathbf{r}}^{\otimes(l-2m)} \otimes \mathbf{I}^{\otimes m} \big\}, \tag{A1}$$

where $C_{l,m} = C \times (-1)^m (2l-2m-1)!!/(2l-1)!!$ ensures the traceless property, and $C = (2l-1)!!/l!$ normalizes the tensor such that the l-fold contraction of $\mathbf{T}_l$ with $\hat{\mathbf{r}}$ yields unity. The matrix $\mathbf{I}$ is the 3×3 identity matrix, while $\hat{\mathbf{r}}^{\otimes(l-2m)} = \hat{\mathbf{r}} \otimes \cdots \otimes \hat{\mathbf{r}}$ and $\mathbf{I}^{\otimes m} = \mathbf{I} \otimes \cdots \otimes \mathbf{I}$ are the corresponding $(l-2m)$- and m-fold outer products. The curly brackets denote summation over all permutations of the l unsymmetrized indices, i.e., $\{\mathbf{T}_l\}_{i_1 \cdots i_l} = \sum_{\pi \in S_l} (\mathbf{T}_l)_{i_{\pi(1)} \cdots i_{\pi(l)}}$, with S_l being the set of permutations.

We now introduce the product of two irreducible Cartesian tensors $\mathbf{x}_{l_1} \in (\mathbb{R}^3)^{\otimes l_1}$ and $\mathbf{y}_{l_2} \in (\mathbb{R}^3)^{\otimes l_2}$, which yields an irreducible Cartesian tensor of rank l_3, i.e., $\mathbf{z}_{l_3} = (\mathbf{x}_{l_1} \otimes_{\text{Cart}} \mathbf{y}_{l_2})_{l_3} \in (\mathbb{R}^3)^{\otimes l_3} \, \forall \, l_3 \in \{|l_1 - l_2|, \cdots, l_1 + l_2\}$. In this work, we focus exclusively on the even tensor product with $l_1 + l_2 - l_3 = 2n$ defined as [68]

$$(\mathbf{x}_{l_1} \otimes_{\text{Cart}} \mathbf{y}_{l_2})_{l_3} = \sum_{m=0}^{\min(l_1,l_2)-n} C_{l_1 l_2 l_3, m} \big\{ (\mathbf{x}_{l_1} \cdot (n+m) \cdot \mathbf{y}_{l_2}) \otimes \mathbf{I}^{\otimes m} \big\}, \tag{A2}$$

where $(\mathbf{x}_{l_1} \cdot (n+m) \cdot \mathbf{y}_{l_2}) = \sum_{i_1, \cdots, i_{n+m}} (\mathbf{x}_{l_1})_{i_1 \cdots i_{n+m}} (\mathbf{y}_{l_2})_{i_1 \cdots i_{n+m}}$ denotes an $(n+m)$-fold tensor contraction, which results in a tensor of rank $l_1 + l_2 - 2(n+m)$. Here, $C_{l_1 l_2 l_3, m} = C_{l_1 l_2 l_3} \times (-1)^m 2^m (2l_3 - 2m - 1)!!/(2l_3 - 1)!!$ ensures the traceless property, similar to $C_{l,m}$ in Eq. (A1). Furthermore, the constant

$$C_{l_1 l_2 l_3} = \frac{l_1! l_2! (2l_3 - 1)!! ((L_1+1)/2)! ((L_2+1)/2)!}{l_3! L_1!! L_2!! L_3!! (L/2)!},$$

with $L = l_1 + l_2 + l_3$ and $L_i = L - 2l_i - 1$, normalizes the resulting tensor such that the l_3-fold contraction of $\mathbf{z}_{l_3}$ with $\hat{\mathbf{r}}$ yields unity, provided $\mathbf{x}_{l_1}$ and $\mathbf{y}_{l_2}$ are obtained from Eq. (A1).

A.2 Two- and many-body features

The following presents the construction of equivariant two- and many-body features based on irreducible Cartesian tensors and their tensor product. We split vectors $\mathbf{r}_{uv} = \mathbf{r}_u - \mathbf{r}_v \in \mathbb{R}^3$ from atom u to atom v (see Fig. 1 (a)) into their radial and angular components (unit vectors), i.e., $r_{uv} = \|\mathbf{r}_{uv}\|_2 \in \mathbb{R}$ and $\hat{\mathbf{r}}_{uv} = \mathbf{r}_{uv}/r_{uv} \in \mathbb{R}^3$, respectively. In the t-th message-passing layer, edges $\{u, v\}$ are embedded using a fully connected neural network $R^{(t)}_{k l_1 l_2 l_3} : \mathbb{R}^{d_{\text{Bessel}}} \to \mathbb{R}$ with k output feature channels. The radial function $R^{(t)}_{k l_1 l_2 l_3}$ takes as an input radial distances r_{uv},

which are embedded through d_{Bessel} Bessel functions and multiplied by a smooth polynomial cutoff function [69].

We achieve the rotation equivariance by constraining convolution filters to be the products between learnable radial functions and irreducible Cartesian tensors, i.e., $R^{(t)}_{kl_1l_2l_3}(r_{uv})\mathbf{T}_{l_1}(\hat{\mathbf{r}}_{uv})$. The two-body features $\mathbf{A}^{(t)}_{ukl_3} \in \left(\mathbb{R}^3\right)^{\otimes l_3}$ are obtained through the Cartesian tensor product of the filters and the neighbors' equivariant features $\mathbf{h}^{(t)}_{vkl_2} \in \left(\mathbb{R}^3\right)^{\otimes l_2}$, as follows

$$\mathbf{A}^{(t)}_{ukl_3} = \sum_{v\in\mathcal{N}(u)} R^{(t)}_{kl_1l_2l_3}(r_{uv})\mathbf{T}_{l_1}(\hat{\mathbf{r}}_{uv}) \otimes_{\text{Cart}} \frac{1}{\sqrt{d^{(t)}_{\text{feat}}}}\mathbf{W}^{(t)}_{kl_2} \cdot \mathbf{h}^{(t)}_{vl_2}. \tag{A3}$$

Permutational invariance is enforced by pooling over the neighbors $v \in \mathcal{N}(u)$. The learnable weights $\mathbf{W}^{(t)}_{l_2} \in \mathbb{R}^{d^{(t)}_{\text{feat}}\times d^{(t)}_{\text{feat}}}$, with $d^{(t)}_{\text{feat}}$ being the number of feature channels in the t-th message-passing layer, are initialized from a normal distribution with zero mean and unit variance. Here, $\cdot$ denotes the scalar product between the neighbors' equivariant features and the corresponding learnable weights for the k-th output feature channel.

In the first message-passing layer, node embeddings are initialized as learnable weights $\mathbf{W}_{Z_v} \in \mathbb{R}^{d^{(1)}_{\text{feat}}}$ which embed the atom type Z_v. Thus, constructing equivariant two-body features simplifies to

$$\mathbf{A}^{(1)}_{ukl_1} = \sum\nolimits_{v\in\mathcal{N}(u)} R^{(1)}_{kl_1}(r_{uv})\mathbf{T}_{l_1}(\hat{\mathbf{r}}_{uv})W_{kZ_v}. \tag{A4}$$

When training on molecules in different charge states simultaneously, we embed the total charge $Q \in \mathbb{Z}$ in addition to the atom type Z_v. In particular, we obtain new node embeddings $W_{kZ_v} \to W_{kZ_v} + W_{kQ}$, with W_{kQ} computed using an attention-like mechanism described elsewhere [21].

Many-body equivariant features are constructed from $\mathbf{A}^{(t)}_{ukl_\xi}$ through successive applications of the Cartesian tensor product. Specifically, the ν-fold Cartesian tensor product yields $(\nu+1)$-body features, which are represented by an irreducible Cartesian tensor of rank L and are given by

$$\mathbf{B}^{(t)}_{u\eta_\nu kL} = \underbrace{\tilde{\mathbf{A}}^{(t)}_{ukl_1} \otimes_{\text{Cart}} \cdots \otimes_{\text{Cart}} \tilde{\mathbf{A}}^{(t)}_{ukl_\nu}}_{\nu\text{-fold}}. \tag{A5}$$

Here, η_ν counts all possible ν-fold products of $\{l_1, \cdots, l_\nu\}$-rank tensors, resulting in rank-L irreducible Cartesian tensors. Additionally, we define $\tilde{\mathbf{A}}^{(t)}_{ukl_\xi} = \mathbf{W}^{(t)}_{kl_\xi} \cdot \mathbf{A}^{(t)}_{ul_\xi}/\sqrt{d^{(t)}_{\text{feat}}N_{l_\xi}}$, where N_{l_ξ} is the number of rank-l_ξ two-body features obtained from Eq. (A3) or Eq. (A4), and $\mathbf{W}^{(t)}_{l_\xi} \in \mathbb{R}^{d^{(t)}_{\text{feat}}\times\left(d^{(t)}_{\text{feat}}N_{l_\xi}\right)}$ is initialized from a normal distribution with zero mean and unit variance. Finally, since the two-fold tensor products in Eq. (A5) are symmetric to permutations of the involved tensors, the number of distinct ν-fold tensor products, $\text{len}(\eta_\nu)$, is further reduced [11].

A.3 Many-body message-passing and readout layers

The many-body equivariant features in Eq. (A5) are combined using the linear expansion

$$\mathbf{m}^{(t)}_{ukL} = \sum_\nu \sum_{\eta_\nu} W^{(t)}_{Z_u\eta_\nu kL}\mathbf{B}^{(t)}_{u\eta_\nu kL}, \tag{A6}$$

where $W^{(t)}_{Z_u\eta_\nu kL}$ denotes a learnable weight matrix which depends on the chemical element Z_u and rank L. The corresponding entries are initialized from a normal distribution with zero mean and a standard deviation of $1/\text{len}(\eta_\nu)$. The updated node embeddings are further obtained as a linear combination of $\mathbf{m}^{(t)}_{ukL}$ and the residual connection [70]

$$\mathbf{h}^{(t+1)}_{ukL} = \frac{1}{\sqrt{d^{(t)}_{\text{feat}}}}\mathbf{W}^{(t)}_{kL} \cdot \mathbf{m}^{(t)}_{uL} + \frac{1}{\sqrt{d^{(t)}_{\text{feat}}N_Z}}\mathbf{W}^{(t)}_{Z_ukL} \cdot \mathbf{h}^{(t)}_{uL}, \tag{A7}$$

where N_Z denotes the number of atom types and learnable weights $\mathbf{W}^{(t)}_{L} \in \mathbb{R}^{d^{(t)}_{\text{feat}}\times d^{(t)}_{\text{feat}}}$ and $\mathbf{W}^{(t)}_{Z_uL} \in \mathbb{R}^{d^{(t)}_{\text{feat}}\times d^{(t)}_{\text{feat}}}$ are again initialized by picking the respective entries from a normal distribution with zero mean and unit variance.

The site energy for each atom u is expanded into a series of many-body contributions, i.e., $E_u = E_u^{(0)} + E_u^{(1)} + \cdots + E_u^{(T)}$, and is obtained by applying readout layers to scalar many-body features $\mathbf{h}_{u(L=0)} \in \mathbb{R}^{d_{\text{feat}}}$

$$E_u^{(t)} = \begin{cases} \frac{1}{\sqrt{d_{\text{feat}}^{(t)}}} \mathbf{W}_1^{(t)} \mathbf{h}_{u(L=0)}^{(t)} & \text{if } t < T, \\ \frac{1}{\sqrt{d_1^{(t)}}} \mathbf{W}_2^{(t)} \sigma \left(\frac{1}{\sqrt{d_{\text{feat}}^{(t)}}} \mathbf{W}_1^{(t)} \mathbf{h}_{u(L=0)}^{(t)} \right) & \text{if } t = T. \end{cases} \tag{A8}$$

The learnable weights $\mathbf{W}_{l+1}^{(t)} \in \mathbb{R}^{d_{l+1}^{(t)} \times d_l^{(t)}}$ are initialized from a normal distribution with zero mean and unit variance, while $d_0^{(t)} = d_{\text{feat}}^{(t)}$. The Swish/SiLU activation function $\sigma(x) = \alpha x / (1 + \exp(-x))$ with $\alpha \approx 1.6765$ is used as a non-linearity [71, 72]. The corresponding message passing layer is denoted by t with a maximal depth of T.

B Short-range repulsion

By including an empirical repulsion potential, we aid the training process and ensure correct asymptotic behavior as $r \to 0$. Therefore, we introduce the Ziegler-Biersack-Littmark (ZBL) potential [73]

$$E_{\text{rep}} = \frac{1}{2} \sum_u^{N_{\text{at}}} \sum_{v \neq u}^{N_{\text{at}}} \frac{Z_u Z_v}{r_{uv}} f_{\text{c}}(r_{uv}) \times \left[\sum_{k=1}^{4} c_k \exp\left(-d_k \frac{r_{uv}}{a_{uv}} \right) \right], \tag{A9}$$

where $a_{uv} = a / \left(Z_u^b + Z_v^b \right)$. The parameters a, b, $\{c_k\}$, and $\{d_k\}$ are initialized from their tabulated values and treated as learnable during training. A smooth cutoff function f_c is used with cutoff radii defined as the sum of the covalent radii of atoms u and v.

C Experiments and results

C.1 Point-charge dataset

The following sections describe the generation of point-charge datasets and the training of ICTP models using these data. The focus is on the careful design of synthetic reference datasets that enable assessing the transferability of learned long-range electrostatic and dispersion interactions across molecular clusters (isolated, non-periodic atomic structures) and bulk systems (periodic atomic structures). For this purpose, consistent physical models were employed to describe both electrostatic and van der Waals interactions in the isolated and bulk systems.

C.1.1 Dataset construction

Bulk reference data. We generated bulk reference data by simulating a uniform mixture of point charges ($\pm 1e$) comprising 512 particles in a cubic periodic box with an edge length of 40 Å. Simulations were performed with OpenMM 8.2 [66], where the interactions included only long-range electrostatic and truncated van der Waals forces. Long-range electrostatic interactions were treated using Ewald summation, while van der Waals interactions were described by a truncated Lennard–Jones potential. The particle masses were set to 22.99 amu for the positively charged species ($+1e$) and 35.45 amu for the negatively charged species ($-1e$). The Lennard–Jones parameters were defined as $\sigma = 2.3$ Å and $\varepsilon = 0.45$ kJ mol^{-1} for the positive species, and $\sigma = 4.3$ Å and $\varepsilon = 0.42$ kJ mol^{-1} for the negative species. For the Ewald summation, a real-space cutoff of 9 Å was applied. The same cutoff was used for the Lennard–Jones potential. For the latter, a smooth switching function was employed, starting at 6.75 Å.

The molecular dynamics simulation was performed in the canonical (NVT) statistical ensemble using OpenMM 8.2, with a Langevin thermostat at a temperature of 5000 K and a timestep of 4 fs. The simulation was initialized from a random configuration, followed by energy minimization and 1 ns of equilibration. The subsequent 8 ns of dynamics were used to generate reference data. From this trajectory, 2000 configurations were sampled sequentially at 4 ps intervals. The first 1000 frames were assigned to the training dataset, and the remaining 1000 to the test dataset. These datasets are referred to as PBC-512 in the main text.

Cluster reference data. We generated complementary isolated (non-periodic) cluster datasets of controlled size by extracting subsets of particles from the training and test PBC-512 datasets. For each periodic frame, a random central particle was selected, and clusters containing the central particle and its 90, 180, or 270 nearest neighbors were constructed. The resulting datasets are referred to as ISO-90, ISO-180, and ISO-270 in the main text. Energies and forces for the isolated clusters were computed using the Coulomb potential without any cutoff and a truncated Lennard–Jones potential, employing the same parameters as in the bulk reference data.

If a selected cluster carried a net charge, particle charges were randomly flipped until overall charge neutrality was achieved. The physical plausibility of each modified configuration was verified by evaluating atomic forces: the absolute value of the largest force component on any flipped particle was required to remain below 3 eV/Å. The final procedure yielded 1000 configurations for each cluster size.

C.1.2 Training details

All ICTP models were trained using a single NVIDIA RTX 6000 GPU. Models incorporating explicit long-range electrostatics and dispersion terms were trained with a single message-passing layer, whereas models without explicit long-range interactions employed two or four message-passing layers. Irreducible Cartesian tensors up to rank 2 were used to encode directional information, and three-fold tensor products were employed in the product-basis layer. The number of uncoupled feature channels was fixed to 32 for all models.

Radial features were defined using eight Bessel basis functions with a polynomial envelope for the cutoff with $p = 3$ [69]. These features are fed into a fully connected neural network with hidden sizes $[64, 64, 64]$ and SiLU non-linearities [71, 74]. Following the first message-passing layer, the readout functions, which map geometric features to short-range atomic energies and latent atomic charges, are implemented as linear layers. After the second message-passing layer, the readout functions are single-layer fully connected neural networks with 16 hidden neurons each.

A cutoff radius of 5 Å was used for message passing. Long-range dispersion interactions were computed with a cutoff of 9 Å. When long-range electrostatics were modeled via Ewald summation or SPME, the same real-space cutoff of 9 Å was applied. For simulations employing truncated Coulomb interactions, the cutoff radius was varied between 9 Å, 18 Å, and 27 Å, and to ensure that the potential and forces smoothly approached zero at the cutoff, the Coulomb term was modified according to the shifted-force potential [61]. When electrostatic interactions in isolated systems were modeled using the full Coulomb potential without a cutoff, the resulting potentials were fully compatible with Ewald summation and SPME.

All parameters of ICTP models are optimized by minimizing the combined squared loss on training data $\mathcal{D}_{\text{train}} = (\mathcal{X}_{\text{train}}, \mathcal{Y}_{\text{train}})$, where $\mathcal{X}_{\text{train}} = \{S^{(k)}\}_{k=1}^{N_{\text{train}}}$ and $\mathcal{Y}_{\text{train}} = \{E_k^{\text{ref}}, \{\mathbf{F}_{u,k}^{\text{ref}}\}_{u=1}^{N_{\text{at}}^{(k)}}\}_{k=1}^{N_{\text{train}}}$

$$\mathcal{L}\left(\boldsymbol{\theta}, \mathcal{D}_{\text{train}}\right) = \sum_{k=1}^{N_{\text{train}}} \left[C_{\text{e}} \left\| E_k^{\text{ref}} - E(S^{(k)}, \boldsymbol{\theta}) \right\|_2^2 + C_{\text{f}} \sum_{u=1}^{N_{\text{at}}^{(k)}} \left\| \mathbf{F}_{u,k}^{\text{ref}} - \mathbf{F}_u\left(S^{(k)}, \boldsymbol{\theta}\right) \right\|_2^2 \right]. \quad \text{(A10)}$$

When training ICTP models, we set $C_{\text{e}} = 1/N_{\text{at}}^{(k)}$ and $C_{\text{f}} = 0.05$ Å^2 to balance the relative contributions of energies and atomic forces, respectively. Here, $E(S^{(k)}, \boldsymbol{\theta})$ and $\mathbf{F}_u\left(S^{(k)}, \boldsymbol{\theta}\right)$ are energies and atomic forces predicted by ICTP models.

For the ISO-90, ISO-180, and ISO-270 datasets, all ICTP models were trained for 250 epochs, whereas for the PBC-512 dataset we used 500 epochs. Training was performed using the AMSGrad variant of the Adam optimizer [75] with default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. A learning rate of 0.01 and a batch size of 10 were used for all experiments. We also applied gradient clipping with a maximum norm of 100. The learning rate was reduced using an on-plateau scheduler based on the validation loss with a patience of 10 and a decay factor of 0.8. We utilized an exponential moving average with a weight of 0.99 for evaluation on the validation set and for the final model. Additionally, we applied exponential weight decay of 5×10^{-7} on the weights of Supplementary Equations (A5) and (A6). No per-atom energy shift was applied in these experiments. A per-atom scale is determined as the root-mean-square of the components of the forces over the training configurations.

Table A1: **Overview of the SPICE-based training and test datasets used in this work.** Most datasets are taken from SPICE-v2 [44, 45]. For all other datasets, the corresponding references are explicitly provided.

Dataset	$N_{\mathrm{train+valid}}$	N_{test}	N_{at}	Chemical elements
Dipeptides	32157	1693	26–60	H, C, N, O, S
Solvated Amino Acids	1235	65	79–96	H, C, N, O, S
DES370K Dimers	328390	17286	2–34	H, Li, C, N, O, F, Na, Mg, P, S, Cl, K, Ca, Br, I
DES370K Monomers	17765	935	3–22	H, C, N, O, F, P, S, Cl, Br, I
PubChem	1327459	69878	3–50	H, B, C, N, O, F, Si, P, S, Cl, Br, I
Solvated PubChem	13230	697	63–110	H, C, N, O, F, P, S, Cl, Br, I
Amino Acid Ligand Pairs	184316	9702	24–72	H, C, N, O, F, P, S, Cl, Br, I
Ion Pairs	1356	72	2	Li, F, Na, Cl, K, Br, I
Water Clusters[a]	2546	135	3–150	H, O
QMugs[17, 76]	2746	146	50–90	H, C, N, O, F, P, S, Cl, Br, I
NaCl-Water Clusters	1036	56	60–150	H, O, Na, Cl
Test-only datasets				
Small Ligands[45]	–	1996	40–50	H, B, C, N, O, F, P, S, Cl, Br, I
Large Ligands[45]	–	1994	70–80	H, C, N, O, F, P, S, Cl, Br, I
Pentapeptides[45]	–	2000	68–110	H, C, N, O, S

[a] We combine water clusters from SPICE-v2 [45] with those used elsewhere [17].

C.2 SPICE-based dataset

C.2.1 Dataset description

We use a dataset built on the SPICE-v2 dataset [67], which comprises 2,008,628 molecules and molecular clusters with reference energies, atomic forces, and dipole moments obtained from density functional theory (DFT) calculations [44, 45]. The corresponding DFT calculations for SPICE-v2 we performed using the PSI4 software package [77] at the ωB97M-D3(BJ)/def2-TZVPPD level of theory [78–82]. From the 2,008,628 structures, 890 were identified as outliers and removed from SPICE-v2.

Following prior work [17], the SPICE-v2 dataset was further expanded to improve the generalization ability of resulting models by including 2,892 larger molecules (50–90 atoms) from QMugs [76] and 1,681 water clusters containing up to 50 water molecules [83], in addition to those already present in SPICE-v2. Additionally, the SPICE-v2 dataset was augmented with 1092 water clusters containing one or two Na^+Cl^- ion pairs [67]. All additional subsets were computed at the same level of theory as the original SPICE-v2 data. Table A1 provides an overview of the training and test datasets, including an additional SPICE-v2 test set to evaluate the model's generalization ability [45].

All datasets from Table A1, except those used exclusively for testing, were split such that 95 % of atomic structures were used for training and validation, and the remaining 5 % for testing. This split was performed independently for each subset of the SPICE-v2 dataset. Model performance was then evaluated separately on each test subset. The split was performed at the subset level, meaning that conformers of the same molecule could appear in both the training/validation and test datasets. To further test our models on data not seen during training, we used an additional test dataset 45.

C.2.2 Training details

All ICTP models were trained using a single NVIDIA A100 GPU. Each model employed two message-passing layers, irreducible Cartesian tensors up to a maximal rank of two to encode directional information, and two-fold tensor products in the product basis layer. The number of uncoupled feature channels was fixed to 128. Depending on whether we include long-range dispersion and electrostatics, the employed models are ICTP-SR or ICTP-LR, with suffixes (D) indicating whether reference dipole moments were used during training.

Most of the remaining settings are identical to those described for the point-charge dataset. A cutoff radius of 5 Å was employed for message passing. During training, electrostatic interactions were modeled using a Coulomb pair potential without a cutoff. All ICTP parameters were optimized by

Table A2: **RMSEs in the predicted energies and atomic forces for the point-charge test datasets.** Results are shown for the ICTP models trained with and without explicit long-range electrostatics, and for Coulomb interactions modeled with and without a cutoff. Energy RMSEs are given in meV/atom and atomic force errors are given in meV/Å.

Method	ISO-90		ISO-180		ISO-270		PBC-512	
	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces
Training Set ISO-90								
ICTP (1×MP) + Coulomb w/o. Cutoff	0.50±0.02	14.78±4.88	0.40±0.01	12.06±2.84	0.42±0.00	12.55±3.56	1.05pm0.02	11.01±2.58
ICTP (1×MP) + Coulomb w. 9 Å Cutoff	37.86±0.09	196.01±38.64	31.12±0.06	177.68±12.26	25.24±0.14	186.54±22.56	49.33±0.99	120.24±16.88
ICTP (1×MP) + Coulomb w. 18 Å Cutoff	17.38±0.05	65.24±2.24	18.38±0.05	75.83±1.20	16.04±0.06	80.39±1.30	10.26±0.20	39.34±1.87
ICTP (1×MP) + Coulomb w. 27 Å Cutoff	7.89±0.03	30.30±1.41	10.58±0.02	37.86±0.81	10.46±0.04	43.01±0.49	4.11±0.20	17.81±1.19
ICTP (2×MP)	61.66±0.32	770.33±683.27	50.98±0.68	363.71±118.36	51.71±1.09	589.97±444.12	152.64±1.83	433.51±316.17
ICTP (4×MP)	60.71±0.34	1330.58±1430.09	50.27±0.36	1680.18±1445.51	50.82±0.80	3678.66±4417.98	150.17±0.94	2639.73±3324.19
Training Set ISO-180								
ICTP (1×MP) + Coulomb w/o. Cutoff	0.65±0.06	8.35±1.41	0.40±0.03	7.67±0.75	0.35±0.03	7.54±0.74	0.95±0.05	7.91±1.06
ICTP (1×MP) + Coulomb w. 9 Å Cutoff	39.02±0.15	161.60±0.66	31.80±0.09	166.06±0.53	25.63±0.08	164.76±0.21	54.21±0.34	101.44±0.40
ICTP (1×MP) + Coulomb w. 18 Å Cutoff	18.28±0.13	61.41±0.28	17.68±0.11	73.55±0.46	14.89±0.08	78.00±0.41	15.37±0.43	35.94±0.24
ICTP (1×MP) + Coulomb w. 27 Å Cutoff	8.80±0.02	28.80±0.45	9.85±0.03	37.30±0.48	9.26±0.03	42.20±0.43	7.06±0.11	17.15±0.55
ICTP (2×MP)	68.12±1.17	324.13±62.70	47.08±0.54	280.29±11.78	39.75±0.20	275.28±10.69	137.87±0.71	210.64±16.02
ICTP (4×MP)	66.74±0.57	273.48±2.42	46.28±0.39	267.58±3.00	39.05±0.54	261.86±2.15	136.05±2.04	193.99±1.86
Training Set ISO-270								
ICTP (1×MP) + Coulomb w/o. Cutoff	0.77±0.02	8.09±0.65	0.47±0.01	6.71±0.54	0.36±0.00	6.87±0.58	0.93±0.07	7.75±0.30
ICTP (1×MP) + Coulomb w. 9 Å Cutoff	40.35±0.13	160.26±0.29	32.65±0.07	164.42±0.20	25.73±0.04	163.47±0.22	53.50±0.37	99.24±0.33
ICTP (1×MP) + Coulomb w. 18 Å Cutoff	18.55±0.13	60.79±0.22	17.78±0.02	72.99±0.17	14.92±0.03	77.49±0.16	17.81±0.19	34.62±0.18
ICTP (1×MP) + Coulomb w. 27 Å Cutoff	9.37±0.04	28.89±0.06	9.89±0.02	36.87±0.12	9.12±0.02	41.97±0.21	8.29±0.16	16.76±0.24
ICTP (2×MP)	76.84±0.47	273.84±2.25	50.86±0.29	352.34±124.93	37.79±0.15	260.34±1.25	125.73±0.36	191.59±1.04
ICTP (4×MP)	75.43±0.13	268.70±1.17	50.08±0.10	262.57±0.75	37.13±0.12	256.75±0.70	123.75±1.22	188.59±0.83
Training Set PBC-512								
ICTP (1×MP) + Coulomb w/o. Cutoff	2.15±0.07	16.83±6.47	1.60±0.05	13.91±4.23	1.38±0.05	13.94±3.90	0.18±0.01	6.69±0.94
ICTP (1×MP) + Coulomb w. 9 Å Cutoff	106.56±0.15	167.94±0.12	92.93±0.11	170.05±0.07	81.93±0.09	167.81±0.06	2.96±0.00	91.96±0.04
ICTP (1×MP) + Coulomb w. 18 Å Cutoff	39.58±0.08	65.76±0.70	40.08±0.06	77.11±0.22	36.99±0.05	81.03±0.15	0.85±0.00	28.31±0.05
ICTP (1×MP) + Coulomb w. 27 Å Cutoff	18.92±0.14	31.17±0.62	21.09±0.12	39.36±0.40	20.55±0.11	44.18±0.38	0.33±0.00	10.52±0.47
ICTP (2×MP)	215.69±0.61	287.32±8.31	177.93±0.40	266.88±0.74	155.62±0.36	260.88±0.68	6.06±0.03	181.09±0.65
ICTP (4×MP)	213.23±1.81	386.92±82.66	175.84±0.87	316.60±47.24	153.99±0.62	294.95±31.17	5.85±0.04	174.50±1.50

minimizing the combined squared loss on the training data. For ICTP-LR(D), the loss in Eq. (A10) was augmented with an additional dipole-moment term

$$C_d \sum\nolimits_{k=1}^{N_{\mathrm{train}}} \left\| \boldsymbol{\mu}_k^{\mathrm{ref}} - \boldsymbol{\mu}\left(S^{(k)}, \boldsymbol{\theta}\right) \right\|_2^2,$$

with $C_d = 0.005\ (\mathrm{eV})^2 / \left(e\text{Å}\right)^2$, $\boldsymbol{\mu}_k^{\mathrm{ref}}$ denoting reference dipole moments, and $\boldsymbol{\mu}\left(S^{(k)}, \boldsymbol{\theta}\right) = \sum_{u=1}^{N_{\mathrm{at}}} q_u\left(S^{(k)}, \boldsymbol{\theta}\right) \mathbf{r}_u$.

All models were trained for 300 epochs with a batch size of 256. Furthermore, we incorporated a per-atom shift of total energies by solving a linear regression problem [5], taking into account the energies of individual atoms.

C.3 Additional results

Point-charge dataset. Tables A2 and A3 complement the main text by reporting the numerical values of RMSEs for predicted energies and atomic forces across all point-charge datasets and training/test configurations.

SPICE-based dataset. Tables A4 and A5 provide additional results complementing those discussed in the main text. Table A4 reports the mean absolute errors (MAEs) of predicted energies and atomic forces across all test datasets, comparing the short-range ICTP-SR model with the long-range ICTP-LR and ICTP-LR(D) counterparts. Table A5 summarizes the ICTP throughput and per-atom inference times for various system sizes, as obtained from MD simulations with DIMOS [48, 67].

D Broader social impact

This section discusses the broader social impact of the presented work. Our work has important implications for the chemical sciences and engineering, as many problems in these fields require atomistic simulations. Although this work focuses on benchmark data sets, our experiments demonstrate the scalability of our method to larger atomic systems. Beyond constructing machine-learned interatomic potentials, models with explicit long-range interactions can be applied for molecular property prediction, protein structure prediction, protein generation, ribonucleic acid structure ranking, and many more.

Table A3: **RMSEs in the predicted energies and atomic forces for ICTP models with unrestricted Coulomb interactions and with LES [31–33].** All models were trained with a tolerance of $\delta = 5 \times 10^{-5}$ and the associated Ewald/SPME parameters. Results are provided for models evaluated using $\delta = 5 \times 10^{-5}$ and $\delta = 5 \times 10^{-6}$. Energy RMSEs are given in meV/atom and atomic force errors are given in meV/Å.

Method	ISO-90		ISO-180		ISO-270		PBC-512	
	Energy	Forces	Energy	Forces	Energy	Forces	Energy	Forces
Training Set ISO-90								
ICTP (1×MP) + Coulomb w/o. Cutoff ($\delta = 5 \times 10^{-5}$)	0.50±0.02	14.78±4.88	0.40±0.01	12.06±2.84	0.42±0.00	12.55±3.56	1.05±0.02	11.01±2.58
ICTP (1×MP) + Coulomb w/o. Cutoff ($\delta = 5 \times 10^{-6}$)	0.50±0.02	14.78±4.88	0.40±0.01	12.06±2.84	0.42±0.00	12.55±3.56	1.04±0.02	11.01±2.58
ICTP (1×MP) + LES ($\delta = 5 \times 10^{-5}$)	2.66±0.01	25.96±0.32	2.20±0.04	26.14±0.46	1.88±0.04	26.34±0.21	2.20±0.05	26.92±0.22
ICTP (1×MP) + LES ($\delta = 5 \times 10^{-6}$)	165.72±0.58	98.10±0.51	172.08±0.64	97.00±0.38	175.41±0.65	96.58±0.37	198.07±0.75	94.64±0.35
Training Set ISO-180								
ICTP (1×MP) + Coulomb w/o. Cutoff ($\delta = 5 \times 10^{-5}$)	0.65±0.06	8.35±1.41	0.40±0.03	7.67±0.75	0.35±0.03	7.54±0.74	0.95±0.05	7.91±1.06
ICTP (1×MP) + Coulomb w/o. Cutoff ($\delta = 5 \times 10^{-6}$)	0.65±0.06	8.35±1.41	0.40±0.03	7.67±0.75	0.35±0.03	7.54±0.74	0.95±0.05	7.91±1.06
ICTP (1×MP) + LES ($\delta = 5 \times 10^{-5}$)	2.84±0.09	25.60±0.27	2.10±0.09	25.62±0.34	1.63±0.07	26.45±0.43	1.61±0.09	26.43±0.63
ICTP (1×MP) + LES ($\delta = 5 \times 10^{-6}$)	167.44±0.58	98.75±0.57	173.85±0.67	97.61±0.52	177.19±0.68	97.37±0.61	199.92±0.75	95.21±0.30
Training Set ISO-270								
ICTP (1×MP) + Coulomb w/o. Cutoff ($\delta = 5 \times 10^{-5}$)	0.77±0.02	8.09±0.65	0.47±0.01	6.71±0.54	0.36±0.00	6.87±0.58	0.93±0.07	7.75±0.30
ICTP (1×MP) + Coulomb w/o. Cutoff ($\delta = 5 \times 10^{-6}$)	0.77±0.02	8.09±0.65	0.47±0.01	6.71±0.54	0.36±0.00	6.87±0.58	0.92±0.07	7.75±0.30
ICTP (1×MP) + LES ($\delta = 5 \times 10^{-5}$)	3.13±0.12	25.88±0.74	2.25±0.11	25.75±0.93	1.69±0.08	25.99±0.95	1.30±0.13	26.58±0.94
ICTP (1×MP) + LES ($\delta = 5 \times 10^{-6}$)	167.98±0.34	98.67±0.27	174.32±0.41	97.58±0.17	177.66±0.41	97.15±0.16	200.37±0.42	95.26±0.16
Training Set PBC-512								
ICTP (1×MP) + Coulomb w/o. Cutoff ($\delta = 5 \times 10^{-5}$)	2.15±0.07	16.83±6.47	1.60±0.05	13.91±4.23	1.38±0.05	13.94±3.90	0.18±0.01	6.69±0.94
ICTP (1×MP) + Coulomb w/o. Cutoff ($\delta = 5 \times 10^{-6}$)	2.15±0.07	16.83±6.47	1.60±0.05	13.91±4.23	1.38±0.05	13.94±3.90	0.18±0.01	6.69±0.94
ICTP (1×MP) + LES ($\delta = 5 \times 10^{-5}$)	5.10±0.62	33.42±8.21	4.37±0.47	26.85±1.78	3.71±0.40	27.05±1.77	1.32±0.42	23.62±1.15
ICTP (1×MP) + LES ($\delta = 5 \times 10^{-6}$)	167.01±2.19	103.12±4.04	174.25±2.36	99.43±1.84	178.13±2.45	98.99±1.85	204.96±3.19	95.92±2.00

Table A4: **MAEs in the predicted energies/atomic forces for the individual test datasets.** Results are shown for the short-range ICTP-SR model, the long-range ICTP-LR model, and the ICTP-LR(D) model, which constrains learned partial charges using reference dipole moments during training. Energy MAEs are given in meV/atom and atomic force errors are given in meV/Å.

Dataset	ICTP-LR [67]	ICTP-LR(D)	ICTP-SR [67]
Solvated Amino Acids	0.78/23.21	0.71/23.25	1.18/27.10
Dipeptides	0.91/22.91	0.88/22.79	1.32/23.78
DES370K Monomers	1.35/16.31	1.27/16.52	1.87/16.34
DES370K Dimers	1.51/15.24	1.33/28.20	1.70/15.94
PubChem	1.74/30.55	1.75/30.80	1.78/30.84
Ion Pairs	26.39/59.93	18.78/48.92	114.12/355.50
Solvated PubChem	1.58/31.63	1.46/31.41	1.77/34.46
Water Clusters	0.57/15.44	0.60/14.77	0.89/18.74
Amino Acid Ligand	1.02/17.39	0.99/17.28	1.19/18.80
QMugs	0.83/36.24	0.90/36.33	0.96/36.11
NaCl Water Clusters	1.72/27.07	1.22/25.66	3.57/38.03
Test-only datasets			
Large Ligands	1.47/38.08	1.39/38.40	1.48/38.50
Small Ligands	2.06/37.24	2.04/37.06	2.09/37.45
Pentapeptides	1.34/28.42	1.38/28.20	2.08/30.14

[a] For Ion Pairs, we report force MAEs only for the z-component, as the x and y components are zero.

Our work has no obvious negative social impact. As long as it is applied to the chemical sciences and engineering in a way that benefits society, it will have positive effects.

Table A5: **Throughput performance and inference time of ICTP models.** Throughput performance (ns/day) and inference time (μs/atom) are measured for bulk liquid water across increasing system sizes.

N_{at}	ICTP-LR [67]	ICTP-SR [67]
512	1.79/31.46	1.96/28.73
1024	1.12/25.10	1.26/22.40
2048	0.62/22.72	0.66/21.26
4096	0.29/24.47	0.30/23.13
8192	0.14/24.71	0.15/23.33

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction include the presented work's main contributions, assumptions, and results. Each of these claims are backed up through empirical results in the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of the presented work in Section 6.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We describe the entire experimental setting and all details on the performed experiments in Section 5 and Appendix C. We also provide a detailed description of the proposed framework for incorporating long-range dispersion and electrostatics in Section 4. We discuss the details on the ICTP architecture in Section A.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The ICTP source code is publicly available on GitHub at `https://github.com/nec-research/ictp`. The point-charge datasets will be released at `https://10.5281/zenodo.17652034` upon publication.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The details on the experimental setting are provided in Section 5 and Appendix C.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer:[Yes]

   Justification: We run all point-charge experiments using several random seeds and provide the corresponding standard deviations and error bars in all tables and figures. For the SPICE-based dataset, we train a single model for each configuration due to the large size of the dataset.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide details on the computer resources in Appendix C.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The presented work follows the NeurIPS Code of Ethics. We also include Appendix D, which discusses the broader social impact.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We include Appendix D, which discusses the broader social impact.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper does not release any data or models with a high risk for misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We reference all employed data sets and source codes in the main text.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We provide detailed descriptions for generating the point-charge datasets and model training. These datasets will be released at `https://10.5281/zenodo.17652034` upon publication. The ICTP source code is available on GitHub and can be accessed via `https://github.com/nec-research/ictp`.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: We performed neither crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: We performed neither crowdsourcing nor research with human subjects.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.