NoRRA: Nonlocal Representation and Response Augmentation for Machine Learning Interatomic Models

Anonymous authors
Paper under double-blind review

Abstract

Machine-learning interatomic potentials (MLIPs) achieve efficient atomistic simulation by constructing each atomic representation from neighbors within a finite cutoff, but this locality can miss physical responses that depend on structure beyond the cutoff. Two complementary strategies address this limitation: *nonlocal representation*, which supplies local MLIPs with global structural context, and *nonlocal energy response*, which constructs an explicit nonlocal interaction energy from latent variables such as charges or dipoles. We introduce NoRRA (Nonlocal Representation and Response Augmentation), a plug-in framework that couples both pathways in a local MLIP. Reciprocal-space features inject global periodic context before a replaceable backbone, while a separate nonlocal-response readout predicts constrained latent charges and evaluates an explicit interaction energy through a learnable sum-of-Gaussians kernel. On the same Point Edge Transformer (PET) backbone, NoRRA adds only 0.387% parameters and reduces SPICE relative-energy MAE from the baseline value of 19.798 to 14.297 kJ/mol. This framework also transfers to MatterSim. Across molecular, ionic, interfacial, and electrochemical systems, NoRRA improves energy and force accuracy and downstream physical properties, including molten-salt thermal expansion, the liquid structure of NaCl, and the long-range force tails of ion pairs. Component ablations show that the two pathways are complementary, with the largest gains in ionic regimes where Coulomb interactions dominate.

1 Introduction

Machine-learning interatomic potentials (MLIPs) learn potential-energy surfaces from quantum-mechanical reference data, enabling atomistic simulation at system sizes and time scales that are difficult to reach with direct electronic-structure calculations (Behler & Parrinello, 2007; Bartók et al., 2010; Smith et al., 2017; Deringer et al., 2019; Mishin, 2021). Their efficiency stems largely from a locality assumption—each atomic representation is constructed only from neighbors within a finite cutoff. Directional message passing, equivariant representations, and systematic many-body expansions have greatly increased the expressive power of these local models (Schütt et al., 2017; Gasteiger et al., 2020; 2021; Batzner et al., 2022; Musaelian et al., 2023; Batatia et al., 2022), but the spatial information available to the model remains determined by its interaction neighborhood. A larger cutoff extends that neighborhood at additional cost, but the representation is still a local neighborhood rather than an explicit description of global periodic structure. Local MLIPs may therefore miss interactions whose response depends on the broader structure, including electrostatics, polarization, charge redistribution, and collective periodic effects, with consequences for liquid structure and thermal expansion.

Existing work addresses this locality limitation through two complementary routes: *nonlocal response* and *nonlocal representation*. Nonlocal-response methods introduce charges, dipoles, or charge-equilibration degrees of freedom and use them to construct an explicit nonlocal energy, such as an electrostatic, polarization, or dispersion contribution (Gastegger et al., 2017; Ko et al., 2021; Shaidu et al., 2024; Cheng, 2025; King et al., 2025; Zhong et al., 2025; Kim et al., 2025; Ji et al., 2025; 2026). The advantage of this family is that the long-range contribution is an explicit energy term, so charge constraints and the corresponding forces and stresses can be formulated consistently.

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

Its difficulty is that the latent state is itself nonlocal: charges and polarization depend on the complete atomic arrangement and electronic state, and therefore cannot, in general, be determined from a single cutoff-local environment.

Nonlocal-representation methods instead enrich the structural information available to the model (Grisafi & Ceriotti, 2019; Kosmala et al., 2023; Guo et al., 2026). These approaches can provide global context to a local backbone without replacing its local energy decomposition. However, enriched features alone do not prescribe how two distant atoms interact. If the final energy remains a sum of atom-centered local predictions, the model may learn nonlocal correlations statistically, but the correct Coulomb asymptotics and periodic electrostatic effects are not guaranteed by the representation itself. Thus, nonlocal representation determines what global structural information is available to the model, whereas nonlocal response determines how atoms beyond the local graph interact.

Here we introduce NoRRA—**No**nlocal **R**epresentation and **R**esponse **A**ugmentation—a lightweight plug-in framework that couples nonlocal representation and nonlocal response in local MLIPs. At the representation level, a finite set of reciprocal modes injects global periodic context into the initial atomic and edge embeddings, which are fused with local features before the backbone. After the backbone, a fixed channel split routes one output slice to the base readout and another to the nonlocal response readout. That readout predicts constrained latent charges from its slice and uses a learnable sum-of-Gaussians (SOG) kernel to parameterize the nonlocal interaction, yielding an explicit nonlocal energy (Ji et al., 2025; 2026). This design keeps the local backbone replaceable while adding both global structural context and an explicit nonlocal interaction pathway.

We evaluate NoRRA with PET as the local backbone for the main experiments, and attach the same framework to MatterSim to test portability (Pozdnyakov & Ceriotti, 2023; Mazitov et al., 2025; Yang et al., 2024). The evaluation covers molecules, ionic systems, charged clusters and electrolytes, oxide and aqueous interfaces, bulk inorganic materials, and molten salts, among other long-range-sensitive regimes, and draws on datasets whose sizes range from hundreds of structures to an $\mathcal{O}(10^8)$-record public archive. Beyond supervised errors against DFT labels, we examine downstream physical properties, including thermal-expansion coefficients, radial distribution functions, adsorption and binding-energy diagnostics, and long-range force tails, together with component ablations and inference cost. In PET, NoRRA adds only 0.387% parameters while reducing SPICE relative-energy MAE from 19.798 to 14.297 kJ/mol (Eastman et al., 2026), lowering OC22 ID/OOD energy MAE by 43.6% and 33.8% (Tran et al., 2023), and improving energy or force errors on dedicated nonlocal-sensitive systems by roughly 10–79%. The same framework also transfers to MatterSim and yields substantial improvements in downstream physical properties.

Taken together, our contributions are threefold. First, NoRRA combines nonlocal representation and nonlocal response in one lightweight plug-in framework. Second, the benefit of long-range modeling increases as collective electrostatic effects become more pronounced, with the largest improvements in systems where those effects dominate. Third, NoRRA improves energy and force accuracy, and this gain extends to downstream physical tasks.

2 Related Work

Local MLIP architectures. Modern MLIPs are built on locality: each atom is represented from neighbors inside a finite cutoff. This atom-centered construction was established by high-dimensional neural-network potentials and Gaussian approximation potentials (Behler & Parrinello, 2007; Bartók et al., 2010), then developed through learned graph representations (SchNet, DimeNet, GemNet) (Schütt et al., 2017; Gasteiger et al., 2020; 2021) and equivariant models (NequIP, Allegro, MACE) (Batzner et al., 2022; Musaelian et al., 2023; Batatia et al., 2022). Analyses of many-body design spaces make the same locality–cost trade-off explicit (Glielmo et al., 2018; Batatia et al., 2025). Despite these architectural differences, predictions remain confined to a finite neighborhood. We treat such a local model as the backbone and augment the information available to it, rather than introducing a new message-passing architecture.

Nonlocal energy response. Nonlocal energy-response methods introduce latent physical variables and construct an explicit interaction energy from them. Fourth-generation HDNNPs predict electronegativities and solve charge equilibration for charge transfer (Ko et al., 2021); neural-dipole and

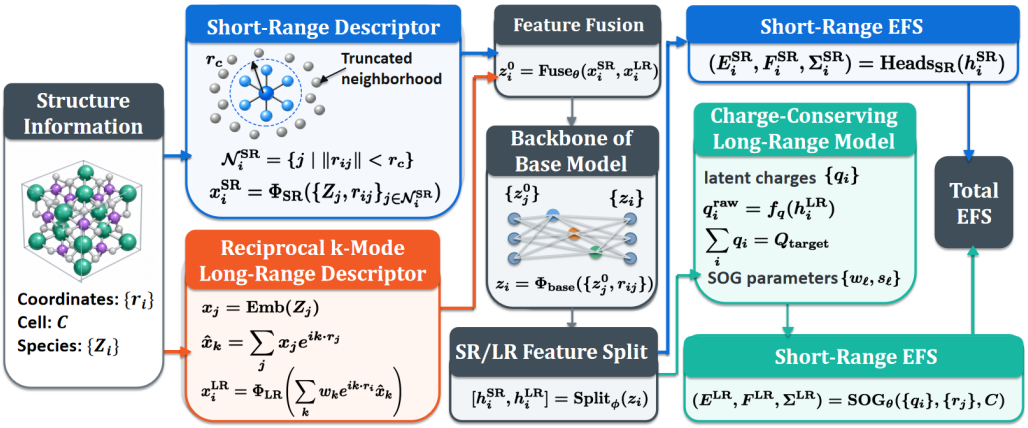

Figure 1: Overview of NoRRA. A cutoff-local descriptor and a finite set of reciprocal k-modes are computed from the same periodic structure. Their atomic and edge features are fused before a replaceable MLIP backbone, so the backbone receives a nonlocal representation while its message passing is left unchanged. After the backbone, one feature slice feeds the base energy, force, and stress heads. A separate slice predicts constrained latent charges and evaluates an explicit nonlocal energy response through a learnable SOG kernel. The two contributions are added to give the final energy, force, and stress.

variational QEq models play the analogous role for polarization and ionic interactions (Gastegger et al., 2017; Shaidu et al., 2024). Latent Ewald Summation couples such variables to reciprocal-space electrostatics, with later work on charge interpretation and electrical response (Cheng, 2025; King et al., 2025; Zhong et al., 2025). Modular wrappers attach similar readouts to different local backbones (Kim et al., 2025). SOG-Net instead parameterizes the interaction kernel as a learnable sum of Gaussians (Ji et al., 2025; 2026). These methods evaluate a nonlocal contribution once suitable latents or local features exist. In our module the SOG operator is that energy response, while a separate reciprocal component supplies the nonlocal representation to the backbone.

Nonlocal representation. Nonlocal representation instead enriches the structural information available to the model. Long-distance equivariant descriptors remap a far-field electrostatic potential onto local environments (Grisafi & Ceriotti, 2019); Ewald message passing inserts Fourier-space updates inside the message-passing layers themselves (Kosmala et al., 2023); and learned reciprocal-space networks encode global periodic context as features (Guo et al., 2026). In this work that periodic information is supplied as the initial node and edge channels of a replaceable backbone, and the backbone's message passing is left unchanged.

3 Method

Overview. Consider a periodic structure with atomic positions $\{\mathbf{r}_i\}_{i=1}^N$, species $\{Z_i\}$, cell matrix $\mathbf{B}$, volume V, and target quantities $\{E, F, S\}$. The method augments a finite-cutoff MLIP with a reciprocal-space interface. The design rationale is complementary: the reciprocal component enriches the representation with global periodic context, whereas the interaction component models an explicit long-range interaction response. Its computation has four stages: local and reciprocal features are constructed from the same structure, the features are fused before a replaceable backbone, the backbone features are used by both the base prediction heads and an interaction-based long-range readout, and the two output contributions are combined. The local neighborhood is

$$\mathcal{N}_i^{\mathrm{SR}} = \{j : \|\mathbf{r}_{ij}\| < r_c\}, \qquad \mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i, \tag{1}$$

where r_c is the local cutoff. The central design goal is to expose information from the complete periodic structure without replacing the local neighbor-processing interface of the base MLIP.

Reciprocal-space representation component. Let $x_j = \mathrm{Emb}(Z_j)$ be a learned species embedding. For an integer reciprocal mode $\mathbf{n}$, the corresponding wave vector is

$$\mathbf{k}_n = 2\pi \mathbf{B}^{-\mathsf{T}} \mathbf{n}. \tag{2}$$

We retain a finite nonzero hemisphere $\mathcal{K}_+$ of integer modes with $\|\mathbf{n}\|_2 \le n_{\mathrm{cut}}$ (Appendix B) and recover the conjugate half implicitly. The species-weighted structure factor is

$$S_n^x = \sum_{j=1}^{N} x_j \exp(\mathrm{i}\mathbf{k}_n \cdot \mathbf{r}_j)\,. \tag{3}$$

The mode response is filtered by a reciprocal-feature function $a_\phi(\mathbf{k}_n, \mathbf{B})$, which can combine an Ewald-inspired envelope with a learned mode filter. The atom-level reciprocal feature is then constructed as

$$z_i^{\mathrm{rec}} = \Pi_{\mathrm{rec}}\left[\frac{2}{V} \sum_{\mathbf{k}_n \in \mathcal{K}_+} a_\phi(\mathbf{k}_n, \mathbf{B})\, S_n^x \exp(-\mathrm{i}\mathbf{k}_n \cdot \mathbf{r}_i)\right], \tag{4}$$

where the factor of two is the hemisphere conjugate and Π_{rec} is a LayerNorm and MLP on that real-valued sum. Equation 4 exposes global periodic context through a finite set of modes: every atom contributes to each structure factor, including atoms outside the local cutoff. The reciprocal feature filter a_ϕ is distinct from the SOG interaction kernel introduced below.

For an edge (i, j) in the local graph, the reciprocal edge feature is

$$z_{ij}^{\mathrm{rec}} = \Phi_{\mathrm{edge}}\big(z_i^{\mathrm{rec}} \,\|\, z_j^{\mathrm{rec}} \,\|\, \mathbf{r}_{ij} \,\|\, \|\mathbf{r}_{ij}\|\big)\,, \tag{5}$$

where Φ_{edge} is a learned MLP and $\|$ denotes concatenation. Thus, the reciprocal component supplies global information to both atomic and edge channels while the local graph itself remains finite-cutoff.

Fusion with a replaceable MLIP backbone. Let e_i^{loc} and e_{ij}^{loc} be the finite-cutoff local node and edge features constructed on $\mathcal{N}^{\mathrm{SR}}$. They are concatenated with the reciprocal channels at the input interface:

$$u_i^0 = e_i^{\mathrm{loc}} \,\|\, z_i^{\mathrm{rec}}, \tag{6}$$

$$u_{ij}^0 = e_{ij}^{\mathrm{loc}} \,\|\, z_{ij}^{\mathrm{rec}}. \tag{7}$$

The fused features are processed by a base MLIP,

$$(h_i^{\mathrm{base}}, h_{ij}^{\mathrm{base}}) = \mathcal{B}_\omega\big(\{u_i^0\}, \{u_{ij}^0\}, \mathcal{N}^{\mathrm{SR}}\big)\,, \tag{8}$$

where $\mathcal{B}_\omega$ is replaceable. PET is the carrier used for the controlled experiments in this paper (Pozdnyakov & Ceriotti, 2023; Mazitov et al., 2025); the interface is specified at the feature and readout level rather than by PET-specific internal states. A channel split of the backbone output produces an SR slice h^{SR} for the base heads and an LR slice for the explicit readout,

$$(h_i^q, h_{ij}^q) = \Pi_q(h_i^{\mathrm{base}}, h_{ij}^{\mathrm{base}}), \tag{9}$$

where Π_q is a fixed width split (about one quarter of the node and edge channels in the reported runs), not a learned projection.

Interaction-based long-range readout. The long-range readout applies an edge MLP to $\{h_{ij}^q\}$, sums the cutoff-weighted results, and concatenates them with h_i^q:

$$q_i^{\mathrm{raw}} = f_q\Big(h_i^q, \mathrm{Agg}_{j \in \mathcal{N}_i^{\mathrm{SR}}}\, h_{ij}^q\Big)\,. \tag{10}$$

The implemented charge q_i is obtained from this residual by the optional residual bound, element prior, second bound, and per-structure projection in Eq. equation 23; it is not a direct map of q_i^{raw}. The resulting charges define a reciprocal structure factor

$$S_n^q = \sum_{i=1}^{N} q_i \exp(\mathrm{i}\mathbf{k}_n \cdot \mathbf{r}_i)\,. \tag{11}$$

The interaction response is represented by a learnable sum of Gaussian basis functions,

$$K_\psi(k) = \sum_{\ell=1}^{M} w_\ell \, \kappa_\ell(k; s_\ell), \qquad \psi = \{w_\ell, s_\ell\}_{\ell=1}^{M}, \tag{12}$$

where the exact positive-width parameterization of κ_ℓ is given in Appendix B. The coefficients w_ℓ and scale parameters s_ℓ are learnable. The SOG interaction energy has the reciprocal-space form

$$E_{\mathrm{LR}} = \frac{1}{2V} \sum_{\mathbf{k}_n \neq \mathbf{0}} W(\mathbf{k}_n, \mathbf{B}) K_\psi(\|\mathbf{k}_n\|) \, |S_n^q|^2 + E_{\mathrm{bc}}, \tag{13}$$

where W contains the reciprocal-grid and boundary-condition factors, and E_{bc} denotes an optional boundary correction (Appendix B). This separates the learned reciprocal feature filter a_ϕ from the learnable SOG interaction kernel K_ψ.

Energy, force, and stress composition. The base heads read the SR slice and the interaction readout reads q:

$$(E_{\mathrm{base}}, F_{\mathrm{base}}, S_{\mathrm{base}}) = H_{\mathrm{base}}(h^{\mathrm{SR}}), \tag{14}$$

$$(E_{\mathrm{LR}}, F_{\mathrm{LR}}, S_{\mathrm{LR}}) = H_{\mathrm{LR}}(q, K_\psi, \{\mathbf{r}_i\}, \mathbf{B}). \tag{15}$$

The reported outputs are the sum of the two contributions:

$$E = E_{\mathrm{base}} + E_{\mathrm{LR}}, \qquad F = F_{\mathrm{base}} + F_{\mathrm{LR}}, \qquad S = S_{\mathrm{base}} + S_{\mathrm{LR}}. \tag{16}$$

Training losses, force/stress target handling, and dataset-specific weights are given in Appendix C.

4 Experiments

The main experiments use PET as the local backbone. Within each comparison, the finite-cutoff PET baseline and PET-NoRRA use the same data split, preprocessing, backbone capacity, training budget, and checkpoint-selection rule; the only architectural difference is whether NoRRA is enabled. MatterSim tests portability of the same framework in Section 4.4, with further diagnostics in Appendix F. The experiments ask when nonlocal modeling improves energy and force accuracy, whether these gains transfer to finite-temperature physical observables, and how the representation and response pathways contribute in different systems. Section 4.5 isolates the two pathways; the remaining main-text comparisons evaluate the full NoRRA framework. Training hyperparameters and evaluation protocols are given in Appendices D and F.

4.1 NoRRA improves molecular energy landscapes and ion-pair tails

SPICE Test Set 1 probes nonlocal electrostatics across dimers, ligands, peptides, and charged species (Eastman et al., 2023; 2026). With the configuration given in Appendix D.1, we attach two nonlocal frameworks to the same PET backbone for this comparison: PET-LES and a direct-space QEq readout with a fixed $1/r$ kernel. In Table 1, bold marks the numerically best value among all listed models in a column, and underline marks the best configuration among PET-backbone models.

Results of the SPICE datasets follow. Panel A reports the within-system pairwise relative-energy error on this test set (Appendix D.1). The relative error removes the effect of an arbitrary energy-zero offset and therefore better reflects the conformational energy differences within each molecule or cluster. We observe that, among all models trained on the same data, PET-NoRRA achieves the best overall result. Within the PET-backbone models, the overall error falls from the baseline value of 19.798 to 14.297 kJ/mol, below both the PET-LES rows (16.606 and 16.960 kJ/mol) and the PET-QEq row (18.270 kJ/mol), and PET-NoRRA is the best PET-backbone configuration in every reported subset. We argue that the reason is that the reciprocal representation supplies conformational context beyond the cutoff, and the response pathway adds an explicit nonlocal energy, so those within-system energy differences change when conformers differ in their long-range electrostatic arrangement.

Panel B focuses on ion pairs because their nonlocal interaction is the strongest in this setting. It uses the 28 diatomic ion-pair groups in SPICE-1.1, each with 51 separations (Appendix D.1). This experiment does not consider the magnitude of the force and instead asks whether the force along the internuclear axis is attractive or repulsive. The three accuracies are whether that direction matches the

Table 1: Results of SPICE dataset. Panel A: Test Set 1 pairwise relative energy (kJ/mol; lower is better). Panel B: ion-pair force-sign accuracy (higher is better). Bold is the best value in a column; underline is the best PET-backbone model. Details are in Appendix D.1.

Panel A: Test Set 1 pairwise relative energy (kJ/mol; lower is better)

Model	Overall	Small	Large	Peptides	Dimers	Neutral	Charged
Public SPICE v1.1 benchmarks							
Egret-1	15.142	**4.256**	**7.046**	**6.184**	43.030	13.252	23.189
FeNNix-Bio1-S	16.759	7.302	9.420	6.919	43.394	15.646	21.463
MACE-OFF24-M	17.726	5.601	8.079	8.122	49.040	15.439	27.460
MACE-OFF23-L	19.778	4.431	7.304	7.461	59.840	17.471	29.598
AceFF-1.1	20.580	8.677	13.926	18.123	41.594	18.207	30.615
MACE-OFF23-S	21.483	7.216	32.396	8.743	37.504	21.610	**20.939**
MACELES-OFF	21.937	4.860	10.490	6.587	65.725	20.976	26.029
Local PET configurations							
PET-NoRRA	**14.297**	9.766	12.829	18.171	**16.422**	**11.256**	24.550
PET-LES, $0.01{:}1$	16.606	12.430	15.095	19.074	19.823	12.698	29.781
PET-LES, $0.05{:}1$	16.960	11.609	14.026	21.336	20.868	12.483	32.053
PET-QEq	18.270	11.118	13.791	28.582	19.589	11.563	40.882
PET baseline	19.798	13.389	16.454	24.228	25.122	14.081	39.074

Panel B: SPICE-1.1 ion-pair tail diagnostics (sign accuracy; higher is better)

Model	Formal tail-force sign	Force sign vs. DFT	Opposite-sign formal tail
PET baseline	0.3330	0.4440	0.3330
PET-NoRRA	**0.8210**	**0.8580**	**1.000**
PET-LES, $0.01{:}1$	0.6670	0.6260	0.5830
PET-LES, $0.05{:}1$	0.5710	0.5740	0.6670
PET-QEq	0.524	0.548	0.250

Coulomb sign implied by the formal-charge product at the largest separations, whether it matches the DFT force over all separations, and whether the same tail test holds for pairs whose formal charges have opposite signs. We use these signs rather than an energy error because, without an isolated-ion reference energy and with an arbitrary energy zero, the sign can still determine whether the long-range tail is attractive or repulsive. We observe that PET-NoRRA surpasses the PET baseline, both PET-LES rows, and the PET-QEq row, raising each sign accuracy from roughly $1/3$ correct to nearly all correct. We argue that the reason is that, once the ion pair is separated beyond the local cutoff, the partner atom leaves the neighborhood and a finite-cutoff model has no basis for deciding whether the tail force should attract or repel, whereas NoRRA obtains a definite nonlocal interaction from the nonlocal representation and the energy response and can therefore determine attraction or repulsion.

4.2 Nonlocal gains depend on the interaction regime

OMat24 (Barroso-Luque et al., 2026) is a public dataset of inorganic crystals spanning a wide range of chemistries. From the full set we take a strict-ionic subset, and from that subset the alkali halides, which are dominated by collective Coulomb interactions. Panel A of Table 2 reports, for each of these regimes, the error on its training set and on its own validation set. Within each regime, the PET baseline and PET-NoRRA use the same records, training budget, and frozen checkpoint. The comparison is the relative reduction within each regime, since absolute errors are not comparable across regimes. We observe that the relative energy and force gains increase from the full set through the strict-ionic subset to alkali halides. This trend appears on both the training set and the validation set. It is especially clear in the relative energy improvement: 3.2%, 5.9%, and 17.2% on the training set, and 2.1%, 3.3%, and 9.3% on the validation set. We argue that, in the full set, valence electrons are mostly shared by neighboring atoms, so the energy and forces come mainly from bonding inside the cutoff and the Coulomb contribution outside it is small. As ionicity increases, the electrons lie more fully on the anions and the atoms behave more like oppositely charged ions. The Coulomb term accumulates as $q_i q_j / r$, and the part beyond the cutoff grows. This is already true for the strict-

Table 2: OMat24 training and downstream physical properties. Panel A: training-set and validation-set errors for each regime; negative Δ is a reduction. Panel B: molten-salt α_V, RDF, and auxiliary LO–TO; closer to the reference is better. Configurations are in Appendix E.

Panel A: domain-concentration train and held-out validation (lower / negative Δ is improvement)

		Energy (meV/atom)			Force (meV/Å)			Stress (GPa)		
Subset	Split	PET	NoRRA E	Δ (%)	PET	NoRRA F	Δ (%)	PET	NoRRA	Δ (%)
Full set	Train	26.58	25.72	-3.200	115.5	113.1	-2.100	0.5764	0.5681	-1.400
	Val	24.17	23.67	-2.100	110.1	108.6	-1.400	0.5560	0.5491	-1.300
Strict ionic	Train	21.72	20.45	-5.900	121.2	117.2	-3.300	0.4915	0.4808	-2.200
	Val	22.76	22.00	-3.300	116.9	113.4	-3.000	0.4970	0.4878	-1.900
Alkali-halide	Train	5.628	4.658	-17.20	37.77	30.38	-19.60	0.08529	0.07822	-8.300
	Val	12.76	11.58	-9.300	46.97	44.15	-6.000	0.09593	0.09100	-5.100

Panel B: molten-salt downstream validation (closer to reference is better)

Observable	Reference	PET baseline	NoRRA	Error reduction (%)
NaCl α_V (10^{-4} K^{-1}) / DFT	2.700	3.837	**2.641**	94.80
LiCl α_V (10^{-4} K^{-1}) / FPMD	2.870	3.475	**2.769**	83.30
LiCl–KCl α_V (10^{-4} K^{-1}) / FPMD	2.730	3.572	**3.030**	64.40
NaCl RDF RMSE / DFT traj.	–	0.04113	**0.03360**	18.30
NaCl LO–TO (THz) / experiment	2.660	2.950	**2.657**	98.80

ionic subset and is strongest for its alkali halides, whose labels are closest to a crystal-wide Coulomb sum, so the relative energy reduction is largest there.

4.3 NoRRA improves downstream physical properties

The intended use of an MLIP is finite-temperature simulation. Panel B of Table 2 asks whether the OMat24 energy and force gains appear in molten-salt observables. The thermal-expansion and RDF results are obtained from molecular-dynamics simulations with the PET baseline and PET-NoRRA models trained on the alkali-halide set in Section 4.2. α_V is obtained from pressure–volume equation-of-state fits to fixed-volume simulations and measures the collective expansion of the melt with temperature. The RDF records ion–ion distances along a molecular-dynamics trajectory and measures the liquid structure produced by the sampled forces. The LO–TO entry is an auxiliary phonon diagnostic: NoRRA is used to test the splitting between the longitudinal and transverse optical branches of NaCl. The configurations of these three experiments are given in Appendix E.

We observe that PET-NoRRA brings the α_V of molten NaCl, LiCl, and LiCl–KCl closer to the DFT/FPMD references (Tovey et al., 2020; Bengtson et al., 2014), reducing the deviations by 94.8%, 83.3%, and 64.4%. Moreover, the molten-NaCl RDF RMSE falls from 0.04113 to 0.03360 (Tovey et al., 2020). In the auxiliary phonon diagnostic, NoRRA moves the NaCl LO–TO splitting from 2.950 to 2.657 THz, close to the experimental 2.660 THz (Raunio et al., 1969).

We argue that the thermal-expansion coefficient of a molten salt is sensitive to nonlocal Coulomb cohesion. As the melt expands, inter-ion distances grow, and ions that lay near the cutoff move outside it. The true Coulomb attraction does not vanish there. It continues to fall off as $1/r$, still pulls the ions inward, and contributes to the pressure at that volume. A finite-cutoff model drops this attraction at the boundary. The larger the expansion, the more cohesion is lost, so the zero-pressure volume moves outward faster with temperature and the baseline α_V lies above the DFT/FPMD references. The nonlocal interaction keeps that energy at every separation, the pressure retains the volume-dependent attraction, and PET-NoRRA moves α_V toward the reference. The smaller RDF error and the LO–TO splitting are consistent with the same long-range electrostatics entering the sampled forces and the vibrational response.

4.4 Generalization and backbone compatibility

This section extends the PET experiments to OC22 (Tran et al., 2023), an oxide-electrocatalyst dataset, and to several dedicated systems that contain long-range interactions. Panel A of Table 3 compares the PET baseline and PET-NoRRA on the ID and OOD splits of OC22. Panel B trains the six dedicated systems with NoRRA. Detailed descriptions of the systems and protocols are given in

Table 3: Generalization across chemistries and a backbone change. Panel A: PET baseline and PET-NoRRA on the OC22 ID and OOD splits. Panel B: six dedicated systems with PET-NoRRA. Panel C: MatterSim-NoRRA, with the backbone and data pipeline unchanged. Negative Δ is a reduction. Lower E/F MAE is better; higher EFwT/cosine is better. Details are in Appendix D.

Panel A: OC22 electrochemical validation

	Energy MAE (meV/system)		Force MAE (meV/Å)		EFwT	Force
Model	ID	OOD	ID	OOD	Avg (%)	Cosine Avg
PET baseline	970.6	1,244	34.63	40.93	0.0004	0.5827
PET-NoRRA	**547.6**	**824.1**	**32.77**	**40.12**	**0.0033**	**0.6153**
Relative change	-43.60%	-33.80%	-5.400%	-2.000%	$\times 8.000$	$+5.600\%$

Panel B: PET dedicated-system training

System	Base E (meV/atom)	NoRRA E (meV/atom)	ΔE (%)	Base F (meV/Å)	NoRRA F (meV/Å)	ΔF (%)
Au/MgO	0.441	0.160	-63.72	40.89	7.935	-80.59
$Na_{8/9}Cl_8^+$	0.363	0.184	-49.14	11.06	2.299	-79.21
Water/Pt interface	6.551	4.827	-26.31	42.03	37.98	-9.630
Deep-water1900	0.533	0.425	-20.17	31.23	22.96	-26.49
LES-BEC water	1.673	1.301	-22.26	58.07	36.07	-37.89
Electrolyte-B	8.213	7.380	-10.15	61.38	35.81	-41.66

Panel C: MatterSim-NoRRA portability

Electrolyte-B	0.419	0.239	-42.90	31.69	17.13	-45.90
$Na_{8/9}Cl_8^+$	1.385	0.492	-64.50	37.90	12.52	-67.00
Deep-water1900	1.053	0.536	-49.10	58.43	44.80	-23.30
Water/Pt interface	0.559	0.400	-28.40	36.42	28.36	-22.10
Au/MgO	2.196	0.089	-95.90	46.83	13.56	-71.10

Appendix D. We observe that, on OC22, both energy and force improve, with an energy reduction of about 40%. On the dedicated systems, both the energy and force errors fall by at least 10%, most strongly for Au/MgO and the $Na_{8/9}Cl_8^+$ clusters. These results indicate that the improvement is broad across systems in which nonlocal effects are present. The systems differ in chemistry, covering an oxide surface, charged ionic clusters, water interfaces, and an electrolyte, but their interactions are not confined to a local neighborhood. NoRRA lowers both energy and force across them, so the gain is not an artifact of one dataset.

Panel C attaches NoRRA to MatterSim. MatterSim-NoRRA reduces error on the five systems in the table. The comparison protocol is given in Appendix F. Moreover, the adsorption-state energy difference on Au/MgO also improves and distinguishes the two adsorption configurations (Table 9).

4.5 Component ablation and inference analysis

This section asks whether the two pathways do the same work. Panel A of Table 4 is a 2×2 on three dedicated systems: NoLR, Repr.-only, Resp.-only, and NoRRA. Repr.-only keeps only the reciprocal representation, feeds it into one total-energy head, and turns off the explicit interaction. Resp.-only keeps only the SOG response and removes the reciprocal representation. We observe that the two pathways are not interchangeable, and NoRRA is lowest on energy and force for all three systems. On the $Na_{8/9}Cl_8^+$ clusters Resp.-only accounts for most of the energy drop, while on Au/MgO Repr.-only accounts for most of the force drop, and on Water/Pt each pathway improves on NoLR. These gains come from different nonlocal effects. The NaCl clusters are missing the Coulomb energy between ions beyond the cutoff, so Resp.-only carries the energy drop. On Au/MgO the force depends on the configuration on the other side of the interface, outside the local neighborhood, so Repr.-only carries the force drop.

The 2×2 leaves open whether a larger local cutoff would recover the same force drop on Au/MgO. Figure 2 uses the same two-layer PET backbone, training budget, and checkpoint rule as Panel A. It reports the PET baseline at local cutoffs from 4 to 12.5 Å and PET-NoRRA at 4 Å; the remaining cutoffs are in Table 7. We observe that the baseline force MAE falls from about 66 meV/Å at 4–5 Å to about 37 meV/Å at 7.5 Å and does not improve at 10–12.5 Å, while PET-NoRRA at 4 Å already reaches about 10 meV/Å, below the saturated local baseline. We argue that the missing information is not a larger neighborhood: the force on Au/MgO still depends on the configuration on the other side of the interface, which NoRRA supplies.

Table 4: Ablation and inference time. Panel A entries are energy/force MAE (meV/atom, meV/Å) at each variant's lowest validation-energy checkpoint. Panel B is the time to compute the energy and forces of one NaCl structure; columns after para. are the number of atoms. Δ is the relative increase (%). Lower energy and force errors are better. The timing protocol is in Appendix G.

Panel A: Ablation

	$Na_{8/9}Cl_8^+$	Water/Pt	Au/MgO
NoLR	0.360/11.09	2.036/84.67	0.441/40.89
Repr.-only	0.468/9.576	1.703/64.50	0.247/11.84
Resp.-only	0.285/6.504	1.788/65.06	0.367/39.04
NoRRA	**0.184/2.299**	**1.511/58.52**	**0.160/7.935**

Panel B: NaCl inference time (ms)

		Atom number N							
	para.	64	128	256	512	1024	2048	4096	8192
Baseline	4.488M	6.9	7.7	8.7	10.6	15.1	31.8	61.0	113.9
NoRRA	4.505M	11.8	12.6	13.7	15.5	20.0	36.8	68.5	131.8
Δ	0.387	71.0	64.6	56.8	46.6	32.6	16.1	14.4	12.0

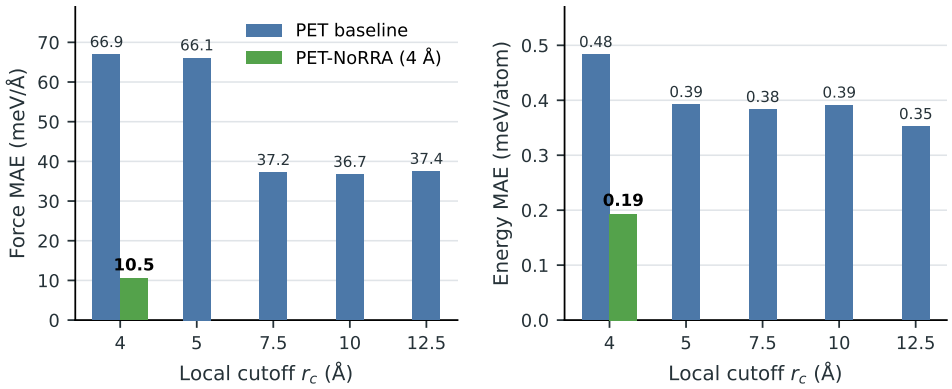

Figure 2: On Au/MgO, comparison between PET baseline with different cutoff and NoRRA at 4 Å.

Panel B of Table 4 compares the time to compute the energy and forces of one structure at a time for the PET baseline and PET-NoRRA trained on the alkali-halide set in Section 4.2. The cells are rock-salt NaCl supercells, from 64 to 8192 atoms. Each reported time is the median over five structures of the per-structure median latency. The protocol is in Appendix G. We observe that the parameter count rises by only 0.387%, while the latency overhead falls from 71.0% at 64 atoms to 12.0% at 8192 atoms. We argue that the extra time is one reciprocal-space calculation over the whole cell. As the number of atoms grows, the local backbone has to process more atoms and neighbors, while this calculation does not grow in the same way, so its share of the total time decreases. In large systems, NoRRA may bring more accurate results at a smaller relative cost in time.

5 Conclusion

We introduce NoRRA, a lightweight plug-in framework for a local machine-learning interatomic potential. It adds a nonlocal representation to the learned features and an explicit nonlocal energy response that describes the nonlocal interaction. The experiments show that NoRRA improves the accuracy of the learned model across multiple systems and also improves downstream physical applications. We attach the framework to two base models, PET and MatterSim. The tests indicate that these improvements depend little on the base model and come mainly from the nonlocal effects introduced by NoRRA. Moreover, the training gain becomes larger as the ionicity of the system increases and Coulomb interactions become more pronounced, which is consistent with the reason for introducing these two nonlocal effects.

Future work will test a wider set of downstream physical properties and, using the lightness of the plug-in, run longer and larger molecular-dynamics simulations. We will also attach the framework to more base models. For systems with weaker ionicity but clear long-range effects, such as dispersion-dominated systems, we will examine whether additional methods can improve the training gain.

AI-USE STATEMENT

In this work, we used generative AI tools to implement parts of the method and to interpret existing experimental results while preparing the manuscript. We also used them for feedback on which comparisons to report; we did not use them to design the experimental protocol, generate synthetic datasets, develop theoretical models, propose hypotheses, or clean or reformat datasets. Additionally, we used generative AI tools to draft and edit the manuscript, reorganize code and LaTeX, format tables and references, and brainstorm wording. We reviewed all AI-assisted text, code, and numerical reporting. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The supplementary material regenerates the manuscript figures and LaTeX tables from frozen numerical exports. It contains those exports, the scripts that build the figures and tables, reference copies of the manuscript-facing outputs, and a manifest linking each item to its inputs. Running the package checks that the reported numbers and figure labels match these frozen inputs.

REFERENCES

Luis Barroso-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M. Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, Matt Uyttendaele, C. Lawrence Zitnick, and Zachary W. Ulissi. The open materials 2024 (OMat24) inorganic materials dataset and models. *Nature Computational Science*, 6:642–652, 2026. doi: 10.1038/s43588-026-00996-w. URL https://www.nature.com/articles/s43588-026-00996-w.

Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical Review Letters*, 104: 136403, 2010. doi: 10.1103/PhysRevLett.104.136403. URL https://doi.org/10.1103/PhysRevLett.104.136403.

Ilyes Batatia, Dávid P. Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems 35*, pp. 11423–11436, 2022.

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

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

Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98:146401, 2007. doi: 10.1103/PhysRevLett.98.146401. URL https://doi.org/10.1103/PhysRevLett.98.146401.

Amelia Bengtson, Hyo On Nam, Saumitra Saha, Ridwan Sakidja, and Dane Morgan. First-principles molecular dynamics modeling of the LiCl–KCl molten salt system. *Computational Materials Science*, 83:362–370, 2014. doi: 10.1016/j.commatsci.2013.10.043. URL https://doi.org/10.1016/j.commatsci.2013.10.043.

Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), 2007. doi: 10.1063/1.2408420. URL https://doi.org/10.1063/1.2408420.

Bingqing Cheng. Latent ewald summation for machine learning of long-range interactions. *npj Computational Materials*, 11(1):80, 2025. doi: 10.1038/s41524-025-01577-7. URL https://doi.org/10.1038/s41524-025-01577-7.

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

Volker L. Deringer, Miguel A. Caro, and Gábor Csányi. Machine learning interatomic potentials as emerging tools for materials science. *Advanced Materials*, 31:1902765, 2019. doi: 10.1002/adma.201902765. URL https://doi.org/10.1002/adma.201902765.

Peter Eastman, Pavan Kumar Behara, David L. Dotson, Raimondas Galvelis, John E. Herr, Josh T. Horton, Yuezhi Mao, John D. Chodera, Benjamin P. Pritchard, Yuanqing Wang, Gianni De Fabritiis, and Thomas E. Markland. SPICE, a dataset of drug-like molecules and peptides for training machine learning potentials. *Scientific Data*, 10:11, 2023. doi: 10.1038/s41597-022-01882-6. URL https://doi.org/10.1038/s41597-022-01882-6.

Peter Eastman, Evan Pretti, and Thomas E. Markland. Accuracy and efficiency benchmarks of pretrained machine learning potentials for molecular simulations. *Journal of Chemical Theory and Computation*, 22(12):6108–6117, 2026. doi: 10.1021/acs.jctc.6c00130. URL https://doi.org/10.1021/acs.jctc.6c00130.

Paul Peter Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921. doi: 10.1002/andp.19213690304. URL https://doi.org/10.1002/andp.19213690304.

Michael Gastegger, Jörg Behler, and Philipp Marquetand. Machine learning molecular dynamics for the simulation of infrared spectra. *Chemical Science*, 8:6924–6935, 2017. doi: 10.1039/c7sc02267k. URL https://doi.org/10.1039/c7sc02267k.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=B1eWbxStPH.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. GemNet: Universal directional graph neural networks for molecules. In *Advances in Neural Information Processing Systems 34*, pp. 6790–6802, 2021.

Aldo Glielmo, Claudio Zeni, and Alessandro De Vita. Efficient nonparametric n-body force fields from machine learning. *Physical Review B*, 97:184307, 2018. doi: 10.1103/PhysRevB.97.184307. URL https://doi.org/10.1103/PhysRevB.97.184307.

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

Ruijie Guo, Hongyu Yu, Liangliang Hong, Shiyou Chen, Xingao Gong, and Hongjun Xiang. Capturing long-range interactions with a reciprocal-space neural network. *Physical Review B*, 113(17):174101, 2026. doi: 10.1103/7yn2-22h6. URL https://doi.org/10.1103/7yn2-22h6.

Yajie Ji, Jiuyang Liang, and Zhenli Xu. Machine-learning interatomic potentials for long-range systems. *Physical Review Letters*, 135(17):178001, 2025. doi: 10.1103/ssp9-7s81. URL https://link.aps.org/doi/10.1103/ssp9-7s81.

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

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

Daniel S. King, Dongjin Kim, Peichen Zhong, and Bingqing Cheng. Machine learning of charges and long-range interactions from energies and forces. *Nature Communications*, 16: 8763, 2025. doi: 10.1038/s41467-025-63852-x. URL https://doi.org/10.1038/s41467-025-63852-x.

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

Dmitry Korogod, Alexander V. Shapeev, and Ivan S. Novikov. Long-range machine-learning potentials with environment-dependent charges enable predicting LO-TO splitting and dielectric constants. *Physical Review B*, 114:024104, 2026. doi: 10.1103/vw8q-1jx8. URL https://doi.org/10.1103/vw8q-1jx8.

Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. Ewald-based long-range message passing for molecular graphs. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17544–17563, 2023. URL https://proceedings.mlr.press/v202/kosmala23a.html.

Yuanshen Lin, Zhenli Xu, Yusu Zhang, and Qi Zhou. Weighted balanced truncation method for approximating kernel functions by exponentials. *Physical Review E*, 112(1):015302, 2025. doi: 10.1103/xsgv-zbvp. URL https://doi.org/10.1103/xsgv-zbvp.

R. P. Lowndes and D. H. Martin. Dielectric dispersion and the structures of ionic lattices. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 308(1495): 473–496, 1969. doi: 10.1098/rspa.1969.0021. URL https://doi.org/10.1098/rspa.1969.0021.

Arslan Mazitov, Filippo Bigi, Matthias Kellner, Paolo Pegolo, Davide Tisi, Guillaume Fraux, Sergey Pozdnyakov, Philip Loche, and Michele Ceriotti. PET-MAD as a lightweight universal interatomic potential for advanced materials modeling. *Nature Communications*, 16: 10653, 2025. doi: 10.1038/s41467-025-65662-7. URL https://doi.org/10.1038/s41467-025-65662-7.

I. S. Messaoudi, A. Zaoui, and M. Ferhat. Band-gap and phonon distribution in alkali halides. *physica status solidi (b)*, 252(3):490–495, 2015. doi: 10.1002/pssb.201451406.

Yuri Mishin. Machine-learning interatomic potentials for materials science. *Acta Materialia*, 214: 116980, 2021. doi: 10.1016/j.actamat.2021.116980. URL https://doi.org/10.1016/j.actamat.2021.116980.

F. D. Murnaghan. The compressibility of media under extreme pressures. *Proceedings of the National Academy of Sciences*, 30(9):244–247, 1944. doi: 10.1073/pnas.30.9.244. URL https://doi.org/10.1073/pnas.30.9.244.

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

Stephen Scott Parker, A. Long, C. Lhermitte, S. Vogel, M. Monreal, and J. M. Jackson. Thermophysical properties of liquid chlorides from 600 to 1600 K: melt point, enthalpy of fusion, and volumetric expansion. *Journal of Molecular Liquids*, 346:118147, 2022. doi: 10.1016/j.molliq.2021.118147. URL https://doi.org/10.1016/j.molliq.2021.118147.

Sergey Pozdnyakov and Michele Ceriotti. Smooth, exact rotational symmetrization for deep learning on point clouds. In *Advances in Neural Information Processing Systems 36*, 2023. doi: 10.52202/075280-3478. URL https://proceedings.neurips.cc/paper/2023/hash/fb4a7e3522363907b26a86cc5be627ac-Abstract.html.

G. Raunio, L. Almqvist, and R. Stedman. Phonon dispersion relations in NaCl. *Physical Review*, 178 (3):1496–1501, 1969. doi: 10.1103/physrev.178.1496. URL https://doi.org/10.1103/physrev.178.1496.

Bernhard Schmiedmayer and Georg Kresse. Derivative learning of tensorial quantities – predicting finite temperature infrared spectra from first principles. *The Journal of Chemical Physics*, 161: 084703, 2024. doi: 10.1063/5.0217243. URL https://doi.org/10.1063/5.0217243.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems 30*, 2017.

Yusuf Shaidu, Franco Pellegrini, Emine Küçükbenli, Ruggero Lot, and Stefano de Gironcoli. Incorporating long-range electrostatics in neural network potentials via variational charge equilibration from shortsighted ingredients. *npj Computational Materials*, 10:47, 2024. doi: 10.1038/s41524-024-01225-6. URL https://doi.org/10.1038/s41524-024-01225-6.

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

Atsushi Togo, Laurent Chaput, Terumasa Tadano, and Isao Tanaka. Implementation strategies in Phonopy and phono3py. *Journal of Physics: Condensed Matter*, 35(35):353001, 2023. doi: 10.1088/1361-648x/acd831. URL https://doi.org/10.1088/1361-648x/acd831.

Samuel Tovey, Anand Narayanan Krishnamoorthy, Ganesh Sivaraman, Jicheng Guo, Chris Benmore, Andreas Heuer, and Christian Holm. DFT accurate interatomic potential for molten NaCl from machine learning. *The Journal of Physical Chemistry C*, 124(47):25760–25768, 2020. doi: 10.1021/acs.jpcc.0c08870. URL https://doi.org/10.1021/acs.jpcc.0c08870.

Richard Tran, Janice Lan, Muhammed Shuaibi, Brandon M. Wood, Siddharth Goyal, Abhishek Das, Javier Heras-Domingo, Adeesh Kolluru, Ammar Rizvi, Nima Shoghi, Anuroop Sriram, Félix Therrien, Jehad Abed, Oleksandr Voznyy, Edward H. Sargent, Zachary Ulissi, and C. Lawrence Zitnick. The open catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, 13(5):3066–3084, 2023. doi: 10.1021/acscatal.2c05426. URL https://pubs.acs.org/doi/10.1021/acscatal.2c05426.

Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, Matthew Horton, Robert Pinsler, Andrew Fowler, Daniel Zugner, Tian Xie, Jake Smith, Lixin Sun, Qian Wang, Lingyu Kong, Chang Liu, Hongxia Hao, and Ziheng Lu. MatterSim: A deep learning atomistic model across elements, temperatures and pressures. *arXiv preprint arXiv:2405.04967*, 2024. URL https://arxiv.org/abs/2405.04967.

Peichen Zhong, Dongjin Kim, Daniel S. King, and Bingqing Cheng. Machine learning interatomic potential can infer electrical response. *npj Computational Materials*, 11:384, 2025. doi: 10.1038/s41524-025-01911-z. URL https://doi.org/10.1038/s41524-025-01911-z.

A Appendix overview and evidence map

This appendix follows the main-text evidence path: implementation and training choices precede dataset protocols; primary physical and portability evidence follows the protocols; secondary robustness checks and omitted configurations are placed after that evidence. The map below distinguishes the protocol needed to interpret a table from the extended evidence that supports or qualifies it.

These notes define the scope of the evidence. Public SPICE rows in Table 1 are literature references on the same metric. OC22 uses the local official-compatible ID/OOD splits at a common epoch-10 checkpoint. OMat24 domains are local materializations. MatterSim runs are listed with their epoch and training settings in Appendix F. The LO–TO protocol is in Appendix E.

Main question	Main evidence	Protocol	Extended evidence
Molecular relative energy	Table 1	Appendix D.1	—
Ion-pair force tails	Table 1	Appendix D.1	—
Interaction-regime trend	Table 2	Appendix D.2	—
Electrochemical generalization (OC22)	Table 3	Appendix D.3	—
Dedicated-system accuracy	Table 3	Appendix D.4	Appendix H
Component roles	Table 4, Figure 2	Appendix D.4	Table 7
Downstream physics	Table 2	Appendix E.1	Appendix E
Backbone portability	Table 3	Appendix F	Appendix H
Efficiency	Table 4	Appendix G	—

Table 5: Appendix evidence map. Protocol sections describe data, metrics, splits, and checkpoint rules; extended-evidence sections report secondary tables, robustness checks, and auxiliary diagnostics.

B NoRRA implementation details

B.1 Reciprocal-space nonlocal representation

For each periodic cell, the implementation enumerates integer modes $\mathbf{n}$ satisfying $\|\mathbf{n}\|_2 \le n_{\mathrm{cut}}$, removes the zero mode, and retains one hemisphere. The second half of the reciprocal sum is recovered by conjugate symmetry, which gives the factor of two in Eq. 4. Fractional coordinates are computed as $\mathbf{r}_i\mathbf{B}^{-1}$ and the phase is evaluated as $2\pi\mathbf{n}^{\mathsf{T}}\mathbf{r}_i\mathbf{B}^{-1}$. Reciprocal modes are processed in chunks to bound temporary memory. The reciprocal feature output is normalized by the cell volume and projected to real-valued node channels.

The reciprocal-feature filter takes $(\|\mathbf{k}\|, \log\|\mathbf{k}\|, \|\mathbf{k}\|^2)$ as inputs. In the Ewald-inspired form, its base envelope is

$$a_0(\mathbf{k}) = \frac{4\pi}{\|\mathbf{k}\|^2}\exp\left[-\frac{\|\mathbf{k}\|^2}{4\alpha^2}\right]. \tag{17}$$

The hybrid filter is $a_0(\mathbf{k})\,(1+\gamma\tanh\delta(\mathbf{k}))$, where δ is a learned scalar and γ is a frozen residual scale; the fully learned filter uses a positive softplus of δ. The implementation also supports in-plane-only and damped-z mode masks for slab-like cells. These options are configuration choices for boundary conditions, not additional model branches.

B.2 Nonlocal response readout and SOG kernel

Following the SOG formulation of long-range interactions (Ji et al., 2025), consider latent channel charges q_{ic} and a translationally invariant pair response $\phi_\psi(r)$. For a periodic structure with lattice vectors $\mathbf{Bn}$, the pair energy is

$$E_{\mathrm{pair}} = \frac{1}{2}\sum_c \sum_{i,j,\mathbf{n}}{}' q_{ic}q_{jc}\,\phi_\psi(\|\mathbf{r}_i - \mathbf{r}_j + \mathbf{Bn}\|)\,, \tag{18}$$

where the prime denotes the exact-self convention selected by the run; the reciprocal expression below represents the configured diagonal correction through E_{self}. The implementation parameterizes ϕ_ψ as a sum of isotropic Gaussians,

$$\phi_\psi(r) = \sum_{\ell=1}^{M} A_\ell \exp\left(-\frac{r^2}{\sigma_\ell^2}\right), \qquad \sigma_\ell = 2\exp(s_\ell). \tag{19}$$

With the Fourier convention $\hat{f}(\mathbf{k}) = \int_{\mathbb{R}^3} f(\mathbf{r})e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\,d\mathbf{r}$, Poisson summation converts this real-space lattice sum to the reciprocal convolution

$$E_{\mathrm{pair}} = \frac{1}{2V}\sum_c \sum_{\mathbf{k}_n\neq\mathbf{0}} K_\psi(\|\mathbf{k}_n\|)\,|\hat{\rho}_c(\mathbf{k}_n)|^2 + E_{\mathrm{self}}, \qquad \hat{\rho}_c(\mathbf{k}) = \sum_i q_{ic}e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}_i}. \tag{20}$$

The structure factor $\hat{\rho}_c$ is S_n^q in the main text, up to the channel index c; conjugating the Fourier kernel leaves $|\hat{\rho}_c|^2$ unchanged. The Gaussian transform gives the exact multiplier used by the code,

$$\kappa_\ell(k; s_\ell) = \exp\left[-\exp(2s_\ell)k^2\right], \qquad K_\psi(k) = \sum_{\ell=1}^{M} w_\ell \kappa_\ell(k; s_\ell). \tag{21}$$

with $w_\ell = \pi^{3/2} A_\ell \sigma_\ell^3$. The implementation either optimizes w_ℓ directly or uses a multiplicative log-residual parameterization; s_ℓ is optimized unless explicitly frozen. The E_{self} and reciprocal-grid factors in Eq. equation 20 are absorbed by W and E_{bc} in the main text. The derivation also clarifies the separation of modules: a_ϕ filters reciprocal features before the backbone, whereas K_ψ is the Fourier multiplier of an explicit latent-charge interaction.

Electrostatic Coulomb interactions provide the limiting case. For $v_{\mathrm{C}}(r) = 1/r$ (up to the electrostatic prefactor and unit convention used by the model), Poisson's equation gives

$$\widehat{v}_{\mathrm{C}}(k) = \frac{4\pi}{k^2}, \qquad \widehat{v}_{\mathrm{Ewald}}(k) = \frac{4\pi}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right), \tag{22}$$

where the second multiplier is the standard Ewald reciprocal form (Ewald, 1921). Thus Ewald splitting corresponds to one fixed, analytically screened Coulomb multiplier. SOG instead uses a learnable Gaussian mixture. Exponential and Gaussian sums are standard accurate bases for inverse-power decaying kernels. In particular, weighted balanced truncation constructs exponential sums that approximate Coulomb-type kernels uniformly over broad intervals (Lin et al., 2025), supporting the use of short-basis SOG expansions in Coulomb-dominated long-range interactions. With positivity constraints, a sufficiently rich mixture can therefore approximate the resolved Coulomb response, but the implemented w_ℓ is not constrained to reproduce $4\pi/k^2$; it represents a flexible long-range response whose singular and short-range contributions remain delegated to the base model. Latent charges are learned coupling variables, not automatically physical partial charges.

B.3 Charge constraints and charge processing

For a graph g with N_g atoms, the charge predictor combines the selected long-range node feature with a cutoff-weighted aggregation of long-range edge features. Let $\Delta q_{ic}^{\mathrm{raw}}$ denote its raw residual. The implementation processes charges in the following order:

$$\begin{aligned}
\Delta q_{ic} &= \begin{cases} a \tanh(\Delta q_{ic}^{\mathrm{raw}}/a), & a > 0, \\ \Delta q_{ic}^{\mathrm{raw}}, & a \text{ disabled}, \end{cases} \\
u_{ic} &= q_{ic}^0(Z_i) + \Delta q_{ic}, \\
\tau_{ic} &= \begin{cases} b \tanh(u_{ic}/b), & b > 0, \\ u_{ic}, & b \text{ disabled}, \end{cases} \\
q_{ic} &= \Pi_Q(\tau_{ic}, Q_g).
\end{aligned} \tag{23}$$

Here a is the optional residual scale, b is the optional global charge bound, and both are frozen per run rather than intrinsic constants of the method. Write $\mathcal{C}_Q$ for the composition of the optional second bound and Π_Q; this is the operator used in the prior loss of Appendix C. A formal prior uses the specified valence table, whereas the weak prior uses a signed initialization of magnitude 0.5.

For the zero-sum and supplied-total-charge modes, the explicit projection is

$$q_{ic} = \tau_{ic} - \frac{1}{N_g}\left(\sum_{j \in g} \tau_{jc} - Q_{gc}\right), \tag{24}$$

with $Q_{gc} = 0$ for zero-sum charges. Thus the common path is bounded nonlinear activation followed by a per-structure mean/total-charge projection. The code additionally supports per-structure charge layer normalization; the soft-neutrality option does not hard-project charges but adds its configured penalty to the interaction energy. The corresponding optional prior loss is defined in Appendix C. If no charge prior is enabled, $q^0 = 0$.

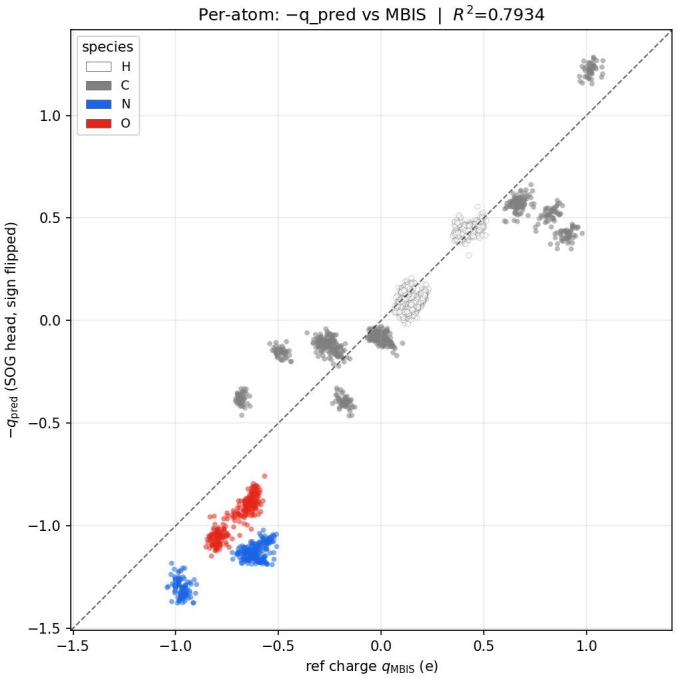

Figure 3: Sign-corrected SOG latent charge versus MBIS partial charge on the SPICE polar-dipeptide subset. The correlation is emergent rather than supervised.

B.4 Latent-charge interpretability

Without direct charge supervision, the sign-corrected latent charge on the SPICE polar-dipeptide subset predicts MBIS charges with $R^2 = 0.793$; the corresponding dipole and primitive-quadrupole component correlations are $R^2 = 0.536$ and $R^2 = 0.458$ (Figure 3).

B.5 Interaction backends

The direct backend evaluates the reciprocal sum on explicit integer modes $\mathbf{n}$. It maps each mode to the physical wave vector $\mathbf{k} = 2\pi\mathbf{B}^{-\mathsf{T}}\mathbf{n}$, retains one hemisphere, and doubles the nonzero-mode contribution by conjugate symmetry. This is the most transparent backend, but its cost grows with the number of retained modes.

The PME/FFT backend instead evaluates the same type of reciprocal response on a uniform fractional-coordinate mesh (Darden et al., 1993). Let $s_i = \mathbf{B}^{-1}\mathbf{r}_i - \lfloor \mathbf{B}^{-1}\mathbf{r}_i \rfloor$ be the wrapped fractional coordinate and let $\mathbf{N}^m = (N_x, N_y, N_z)$ be the mesh size. Charges are assigned to the eight nearest mesh points with trilinear weights $w_{\mathbf{m}}^{\mathrm{mesh}}(s_i)$,

$$\rho_{\mathbf{m}} = \sum_i w_{\mathbf{m}}^{\mathrm{mesh}}(s_i)\, q_i. \tag{25}$$

A three-dimensional FFT gives $\hat{\rho}_{\mathbf{p}}$ on integer FFT modes $\mathbf{p}$. The code maps those modes to physical $\mathbf{k}_{\mathbf{p}} = 2\pi\mathbf{B}^{-\mathsf{T}}\mathbf{p}$, evaluates the SOG filter $K_\psi(\|\mathbf{k}_{\mathbf{p}}\|)$, removes the zero mode, applies the configured slab factor, and divides by the trilinear assignment factor

$$D_{\mathbf{p}} = \prod_{\alpha\in\{x,y,z\}} \mathrm{sinc}^2\left(\frac{p_\alpha}{N_\alpha}\right). \tag{26}$$

After an inverse FFT, the mesh potential is interpolated back to atomic values ϕ_i. For latent charges q_i (possibly vector-valued), the interaction energy is computed as

$$E_{\mathrm{SOG}}^{\mathrm{PME}} = \frac{1}{2V}\sum_i q_i^{\mathsf{T}}\phi_i, \tag{27}$$

with the configured self-interaction subtraction applied when enabled. The mesh is derived from $2\lfloor \|\mathbf{b}_\alpha\| / N_{\mathrm{dl}} \rfloor + 1$, where $\mathbf{b}_\alpha$ is a cell vector. A batch can use one batch-maximum mesh, a structure-specific mesh, or bucket-rounded meshes; the latter group compatible structures to avoid padding every structure to the largest mesh in the batch. Forces and stresses are obtained either by differentiating the energy or with the fixed-charge analytic PME path, according to the run configuration.

The NUFFT backend replaces trilinear grid assignment with a nonuniform FFT. It uses the fractional-coordinate points $2\pi s_i$, applies a type-1 transform to obtain charge structure factors on the FFT mode grid, multiplies by the SOG mode response, and applies a type-2 transform to evaluate the resulting potential at the atomic positions. The energy is again formed from q_i and the nonuniformly sampled potential; the current PET integration uses the NUFFT tolerance default of 10^{-4}. Unlike the PME backend, the current implementation evaluates the SOG response on its dimensionless FFT-mode grid and processes one structure per NUFFT call; it is therefore primarily an efficiency-oriented backend, while the direct and PME paths retain the physical-$\mathbf{k}$ filter convention. Optional CUDA streams overlap multiple per-structure calls.

B.6 Boundary-condition options

For slab-like systems, in-plane mode selection, damped out-of-plane modes, and an optional dipole boundary correction are controlled independently for the feature and interaction components. The self-interaction treatment, mesh policy, N_{dl}, bandwidth count M, backend selection, and boundary flags are recorded with each run configuration.

C Training setup and configurations

C.1 Losses and target normalization

For a structure-level batch, the PET trainer uses target-specific Huber errors. Let $H_\delta(\hat{y}, y)$ denote the elementwise Huber function with transition scale δ, averaged over the normalized tensor elements in the batch. The enabled objective is

$$\mathcal{L} = \lambda_E \mathcal{L}_E + \lambda_F \mathcal{L}_F + \lambda_S \mathcal{L}_S + \lambda_Q \mathcal{L}_Q, \tag{28}$$

where $\mathcal{L}_S$ is included only for stress-supervised tasks and $\mathcal{L}_Q$ is included only when a charge-prior weight is explicitly enabled.

For an energy/force/stress task,

$$\mathcal{L}_E = H_{\delta_E}(\hat{E}_g/N_g, E_g/N_g), \qquad \mathcal{L}_F = H_{\delta_F}(\hat{F}_{i\alpha}, F_{i\alpha}), \qquad \mathcal{L}_S = H_{\delta_S}(\hat{S}_{g,\alpha\beta}, S_{g,\alpha\beta}), \tag{29}$$

where N_g is the number of atoms in structure g. Thus the energy term is optimized on normalized per-atom energies, whereas force and stress retain their tensor shapes. In the conservative mode, the batch stores $-\mathbf{F}$ as the position gradient of the reference energy and $V\mathbf{S}$ as its strain gradient; the corresponding model gradients are compared after target scaling. In the non-conservative mode, force and stress heads are supervised directly. The choice is fixed by the run configuration and is not silently switched during training.

The final SPICE public-style comparison uses the matched qsum configuration $\lambda_E : \lambda_F = 0.25{:}1$, with $\delta_E = 0.015$ and $\delta_F = 0.01$ in normalized target units; stress is absent. The earlier $0.05{:}1$ and $0.1{:}1$ settings are retained only as archived weight-sweep diagnostics and are not the configuration used for the final main-table comparison. The PET baseline and PET-LES controls retain their own prescribed energy weights, as specified in the corresponding run configurations. Representative OMat24 PME configurations use $\lambda_E = 0.01$, $\lambda_F = \lambda_S = 1$, $\delta_E = 0.015$, $\delta_F = 0.01$, and $\delta_S = 0.005$. Other dedicated-system runs retain the same PET loss form but use their frozen configuration weights.

The optional charge-prior term compares the constrained predicted charges with the constrained element prior,

$$\mathcal{L}_Q = \lambda_Q \frac{1}{N C_q} \sum_{i,c} \left[\mathcal{C}_Q(q^0_{ic} + \Delta q_{ic}, Q) - \mathcal{C}_Q(q^0_{ic}, Q) \right]^2, \tag{30}$$

Table 6: Matched training and long-range interface configuration. Panel A lists PET backbone and interface settings; Panel B lists MatterSim settings. "M" in OC22 denotes a larger PET carrier. Training uses Adam with warm-up (10% of epochs) and gradient clipping at norm 1.0; dataset-specific epoch budgets are in the released configs. Matched PET comparisons hold backbone capacity, splits, units, preprocessing, and budget fixed.

Panel A: PET experiments	
Common long-range interface	
Reciprocal cutoff n_{cut} / Ewald α	4 / 1.0
Reciprocal filter	hybrid (Ewald $\times$ learned)
SOG bandwidth M / BSA σ_{init}	12 / 1.6
Charge readout	residual + optional tanh + projection
Backbone and dataset-specific settings	
	r_c / d_{node} / GNN+attn / N_{dl} / backend / charge / batch
SPICE qsum	7.5 / 512 / 2+1 / 2 / direct / q-sum / 1024
SPICE LES	7.5 / 512 / 2+1 / 2 / PME / zero-sum / 1024
OC22 (PET-M)	8.5 / 1536 / 3+2 / 2 / PME / zero-sum / 64
OMat24 broad & strict	7.5 / 512 / 2+1 / 10 / PME / zero-sum / 2048
OMat24 alkali-halide	7.5 / 512 / 2+1 / 10 / PME / zero-sum / 2048
OMat24 downstream	7.5 / 512 / 2+1 / 2 / PME / q-prior / 1024
Dedicated: Au/MgO	7.5 / 512 / 2+1 / 2 / PME / zero-sum / varied
Dedicated: $\mathrm{Na}_{8/9}\mathrm{Cl}_8^+$	7.5 / 512 / 2+1 / 2 / PME / zero-sum / varied
Dedicated: water systems	7.5 / 512 / 2+1 / 2 / PME / zero-sum / varied
Dedicated: Electrolyte-B	7.5 / 512 / 2+1 / 2 / PME / zero-sum / varied
Panel B: MatterSim experiments	
Backbone	MatterSim (1–4 blocks, 64–256 units)
Cutoff (two-body / three-body)	5.0 Å / 4.0 Å
Q predictor	3 layers (64, 32), linear skip
SOG N_{dl} / bandwidth	2 / 12 Gaussians
Charge constraint	zero-sum or λ regularization
$\mathrm{Na}_{8/9}\mathrm{Cl}_8^+$	default MatterSim, NoRRA wrapper
Electrolyte-B	1-block (98K) vs 3-block (228K)
Au/MgO	2 blocks, 64 units, staged refinement
5M-like overhead	4 blocks, 256 units (+3.91% params)
Dedicated: $\mathrm{Na}_{8/9}\mathrm{Cl}_8^+$	default MatterSim, NoRRA wrapper
Dedicated: Electrolyte-B	1-block (98K) vs 3-block (228K)
Dedicated: water-1900	default MatterSim, NoRRA
Dedicated: Water/Pt	1-block, 5000-frame subset
Dedicated: LES-BEC water	1-block, epoch 999

where c indexes latent-charge channels, C_q is the number of channels, and both charge vectors are passed through the same total-charge or neutrality constraint. It is zero when the prior mode or its weight is disabled; for example, the SPICE qsum runs use no charge prior.

Target normalization is applied before the Huber terms. Energy first subtracts a species-dependent atomic baseline and then divides by a fixed energy scale. Non-conservative forces use species-dependent fixed scales, stress uses a fixed stress scale, and conservative force/stress gradients use the energy scale. Reported values undo these normalizations before physical-unit MAE is computed. This distinction is especially important for OC22, where the local training log uses per-atom energy diagnostics but the official-compatible evaluation reports total energy MAE per system.

Force and stress are treated as explicit supervised output targets whenever the selected task requires them. The implementation supports both conservative and non-conservative output configurations; the paper reports the corresponding task metrics without assuming that one output parameterization applies to every experiment.

C.2 PET CONFIGURATIONS

Table 6 summarizes the matched PET configuration. The common block lists the local cutoff, backbone dimensions, reciprocal-mode settings, and SOG kernel parameters shared by the PET experiments. The dataset-specific block reports N_{dl}, the SOG backend, the charge constraint, and the batch size for each experiment family.

The matched comparison holds splits, units, backbone capacity, and budget fixed. The model baseline removes both the reciprocal embedding and explicit SOG readout, whereas NoRRA enables both components. On SPICE, PET-NoRRA uses $\lambda_E : \lambda_F = 0.25{:}1$; the PET baseline and PET-LES

keep their prescribed energy weights (0.01 or 0.05). MatterSim backbone sizes and epochs are listed in Appendix F. Dataset-specific details, including the OC22 ID/OOD manifests, OMat24 domain filters, and dedicated-system train/validation manifests, are kept in the accompanying configuration files and evaluation summaries.

C.3 MATTERSIM CONFIGURATIONS

MatterSim uses its original backbone and data pipeline. The augmented wrapper exposes per-atom representations, predicts latent charges with a lightweight `QPredictor`, and adds the NoRRA nonlocal energy response without modifying the internal MatterSim graph. Run-specific backbone sizes, epochs, and subsets are given in Appendix F.

C.4 CHECKPOINT-SELECTION CONVENTIONS

The main tables use task-specific, predeclared checkpoint rules. SPICE local rows use each run's lowest validation-energy checkpoint after the common 207-epoch budget; OC22 uses the common epoch-10 checkpoint; OMat24 comparisons use the same frozen checkpoint within each domain; dedicated-system retraining uses the energy-calibrated rule described with the data; and the component ablation independently uses each variant's lowest validation-energy checkpoint. Neighboring checkpoints are retained only as stability checks and are not used for test-set model selection.

D DATASETS, SPLITS, AND EVALUATION PROTOCOLS

D.1 SPICE MOLECULAR BENCHMARK

SPICE pairwise relative energy. We use the independent SPICE Test Set 1 HDF5 file and compute the official-style pairwise relative-energy metric (Eastman et al., 2023; 2026). For system s with conformers i, j,

$$\mathrm{MAE}_s^{\mathrm{pair}} = \frac{1}{|P_s|} \sum_{(i,j)\in P_s} \left| (\hat{E}_{s,i} - \hat{E}_{s,j}) - (E_{s,i} - E_{s,j}) \right|, \tag{31}$$

where P_s contains all unordered conformer pairs. The final metric is the macro-average over systems and is reported in kJ/mol. The audited file contains 800 systems and 7,989 available conformations; it is not described as exactly 8,000 conformations. Local PET rows in Panel A of Table 1 use each run's lowest validation-energy checkpoint after matched 207-epoch training. The PET baseline and PET-LES controls keep their prescribed energy weights (0.01 or 0.05) rather than the SOG 0.25 curriculum. Neighboring best-force and last-epoch values are retained in the evaluation archive as a stability check and are not used for model selection.

SPICE-1.1 ion-pair diagnostics. Panel B of Table 1 reports ion-pair tail-sign diagnostics at the same energy-selected PET baseline, PET-NoRRA, and PET-LES 0.01:1 and 0.05:1 checkpoints used in Panel A. The rows use the 28 diatomic ion-pair groups in SPICE-1.1 (51 distances each). Relative pair energies are zeroed by the mean of the 12 largest-distance points in each group. The radial force, with positive values repulsive, is

$$F_R = 0.5(\mathbf{F}_2 - \mathbf{F}_1) \cdot \hat{\mathbf{R}}. \tag{32}$$

Formal tail-force sign accuracy is the fraction of those tail points whose model force sign matches the sign implied by the product of the two formal charges. Force sign vs. DFT is the fraction of all distances whose model radial-force sign matches the DFT radial force. Opposite-sign formal tail restricts the formal-tail metric to the 12 groups with opposite formal charges. SPICE-1.1 does not contain isolated-ion DFT energies, so these curves are not absolute interaction energies.

Direct-space QEq fine-tune. The PET-QEq row in Table 1 is a direct-space, non-periodic charge-equilibration readout with a hard 12 Å cutoff and a fixed $1/r$ kernel. It predicts per-atom electronegativities and hardnesses from the local node features and solves a per-molecule charge-equilibration

problem. It is not a periodic Ewald QEq and not a fourth-generation HDNNP. The reported checkpoint continues the epoch-150 weights for 30 further epochs with a warmup–cosine schedule, learning rate 10^{-3}, batch size 1024, and energy-to-force weight $0.25{:}1$. Selection uses the lowest validation energy. Panel A reports the same subset columns as the other rows.

D.2 OMat24 domain materialization

OMat24 domain filters. We materialize the full OMat24 set, a strict-ionic subset, and an alkali-halide subset (Barroso-Luque et al., 2026). The strict-ionic filter uses electronegativity-separated metal–anion and alkali/alkaline–strong-anion rules; the alkali-halide subset retains structures composed of Li, Na, K, Rb, Cs, F, Cl, Br, and I with at least one alkali and one halogen. Panel A of Table 2 reports same-epoch training MAE together with evaluation of the same frozen checkpoints on the corresponding held-out validation subsets. Absolute MAE values are not comparable across domains because the materializations and epoch budgets differ. Downstream thermal-expansion and RDF simulations use the q-prior PET-NoRRA checkpoint.

D.3 OC22 ID/OOD validation

OC22 metrics. We report local official-compatible $\mathrm{val_id}$ and $\mathrm{val_ood}$ results separately, together with their unweighted average (Tran et al., 2023). The matched table uses the common epoch-10 checkpoint for both the PET baseline and PET-NoRRA under the same split manifests; the checkpoint selection rule and all neighboring-epoch results are retained in the evaluation archive. Energy MAE is the mean absolute error of the predicted total energy per system, in eV/system. Force MAE is the pooled mean absolute error over all Cartesian force components, in eV/Å. EFwT (energy–forces within threshold) is the fraction of systems whose energy and force predictions simultaneously satisfy the OC22 accuracy thresholds; we report it as a percentage. Force cosine is the per-atom cosine similarity between predicted and reference force vectors, averaged over atoms and then over the ID and OOD splits. The split average in Panel A of Table 3 is the unweighted mean of $\mathrm{val_id}$ and $\mathrm{val_ood}$, not a system- or atom-weighted pooled value.

D.4 Dedicated long-range datasets

The dedicated tests include Deep-water1900, Electrolyte-B, $\mathrm{Na}_{8/9}\mathrm{Cl}_8^+$, LES-BEC water, Au/MgO, and Water/Pt(111). The Au/MgO and $\mathrm{Na}_{8/9}\mathrm{Cl}_8^+$ cases follow the long-range charge-transfer benchmark family introduced for 4G-HDNNPs (Ko et al., 2021). LES-BEC water follows the released LES-BEC data and examples (Zhong et al., 2025); its water subset also records the original derivative-learning data source (Schmiedmayer & Kresse, 2024). Water/Pt(111), Deep-water1900, and Electrolyte-B are inherited from the SOG-Net study and its accompanying examples (Ji et al., 2025; 2026). Carbon-chain and Ag-cluster SOG-only runs, and a batch-32 Electrolyte-B component control, are retained in the evaluation archive and are not tabulated here. Dataset manifests, split files, target units, preprocessing, and per-system configuration records accompany the released evaluation archive.

Figure 2 holds the Au/MgO data split, two-layer PET backbone, batch size, learning rate, epoch budget, and neighbor-construction settings fixed at the dedicated Au/MgO protocol, and varies only the local cutoff r_c. The main-text figure reports the PET baseline at 4–12.5 Å and PET-NoRRA only at 4 Å. Table 7 lists the finished energy-best checkpoints for both variants. Neighbor lists follow PET's adaptive-cutoff default in those configs (target of eight neighbors within the search radius), so the scan changes the cutoff upper bound rather than a dense full-neighbor ball. PET-NoRRA runs at $r_c \geq 12.5$ Å and both variants at 20 Å exhausted device memory under the matched batch size.

D.5 Metric and checkpoint conventions

Energy MAE can denote per-atom energy, total energy per system, or a within-system pairwise relative energy depending on the benchmark; the unit is stated in each table caption. Force MAE is reported per Cartesian component. OC22 additionally reports energy–force within-threshold rate and force cosine. Relative changes are computed with respect to the corresponding local baseline. Auxiliary controls and neighboring checkpoints are not used for headline model selection.

Table 7: Au/MgO local-cutoff scan. Entries are validation energy/force MAE (meV/atom, meV/Å) at each run's lowest validation-energy checkpoint, matching Panel A of Table 4. Figure 2 shows the PET baseline at every listed cutoff and PET-NoRRA only at 4 Å.

r_c (Å)	PET baseline	PET-NoRRA
4	0.483/66.89	0.193/10.48
5	0.392/66.06	0.135/7.941
7.5	0.383/37.16	0.217/9.219
10	0.391/36.68	0.218/8.122
12.5	0.352/37.45	—

E Downstream physical properties

E.1 Checkpoint provenance and simulation setup

Panel A reports same-epoch training MAE and domain-matched held-out validation for the same frozen checkpoints: the full set at epoch 1, strict ionic at epoch 6, and alkali-halide at epoch 300 (logged as epoch 299). For the held-out rows, energy and stress MAE are means of per-structure MAE and force MAE is weighted by the number of atoms. The alkali-halide validation subset has N=1394; the full-set and strict-ionic counts are N=769338 and N=238012. Absolute errors remain incomparable across subsets because the materializations and epoch budgets differ.

For Panel B, the thermal-expansion and RDF rows use the q-prior PET-NoRRA 1/4 checkpoints from the molten-salt validation runs. The LO–TO row instead uses an auxiliary response-only checkpoint from the SOG-width sweep. All LO–TO calculations use the same experimental non-analytical correction; no LO–TO result in the table uses model-predicted Born charges.

E.2 Molten-salt thermal expansion from equation-of-state fits

Fixed-volume NVT simulations were used to construct $P(V)$ curves rather than directly sampling volume in an NPT ensemble. Molten NaCl used the 512-atom cubic setup and the Tovey et al. DFT reference (Tovey et al., 2020); LiCl and LiCl–KCl used 216-atom cells and FPMD-based reference data (Bengtson et al., 2014). Literature thermal-expansion comparisons were taken from Parker et al. and Shaidu et al. (Parker et al., 2022; Shaidu et al., 2024). All production runs used a 0.5 fs time step, a Bussi thermostat (Bussi et al., 2007) with 100 fs coupling, 20 ps equilibration, and 50 ps production. NaCl and LiCl–KCl were sampled at 1100, 1150, and 1200 K; LiCl used the 1000, 1096, and 1200 K temperatures aligned to the FPMD reference. Pressures were converted from bar to GPa, and each isotherm was fit with the Murnaghan equation of state (Murnaghan, 1944)

$$P(V) = \frac{B_0}{B_0'}\left[\left(\frac{V_0}{V}\right)^{B_0'} - 1\right], \tag{33}$$

where $V_0(T)$ is the fitted zero-pressure volume, B_0 is the bulk modulus, and B_0' is its pressure derivative. The volumetric expansion coefficient was then computed from a linear fit of $V_0(T)$:

$$\alpha_V = \frac{1}{V_0}\frac{\partial V_0}{\partial T}, \qquad \alpha_L = \frac{\alpha_V}{3}. \tag{34}$$

The error-reduction entry in Table 2 is

$$1 - \frac{|\alpha_V^{\mathrm{LR}} - \alpha_V^{\mathrm{ref}}|}{|\alpha_V^{\mathrm{baseline}} - \alpha_V^{\mathrm{ref}}|}. \tag{35}$$

The fitted zero-pressure volumes of all models remain systematically offset from their references, so Table 2 and Figure 4 compare expansion slopes rather than claiming fully calibrated absolute volumes. Concretely, the fitted α_V values move from 3.837 to 2.641 for NaCl (reference 2.700), from 3.475 to 2.769 for LiCl (reference 2.870), and from 3.572 to 3.030 for LiCl–KCl (reference 2.730), all in units of $10^{-4}\,\mathrm{K}^{-1}$. Table 8 lists the fitted thermal-expansion coefficients and error reductions for all three molten salts; the corresponding NaCl fixed-volume pressure samples are retained in the evaluation archive.

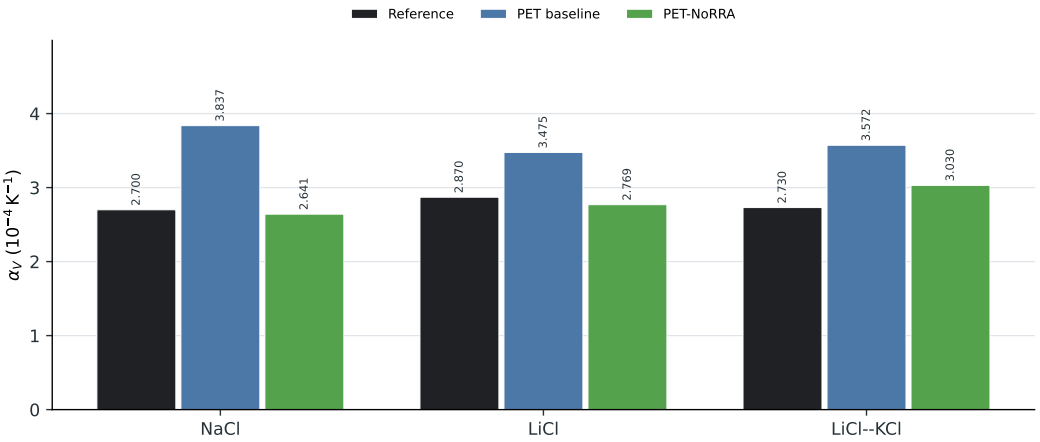

Figure 4: Volumetric thermal-expansion coefficients for molten NaCl, LiCl, and LiCl–KCl. References are DFT for NaCl and FPMD for LiCl and LiCl–KCl.

Table 8: Molten-salt volumetric thermal-expansion coefficients obtained from fixed-volume NVT simulations and Murnaghan equation-of-state fits. $V_0(T)$ is the fitted zero-pressure volume in Å^3/atom; α_V is in $10^{-4}\,\mathrm{K}^{-1}$. $T_1/T_2/T_3$ are 1100/1150/1200 K for NaCl and LiCl–KCl, and 1000/1096/1200 K for LiCl. Model V_0 values remain systematically offset from references; the comparison focuses on the expansion slope. Lower α_V error relative to the reference is better.

System	Model	$V_0^{T_1}$	$V_0^{T_2}$	$V_0^{T_3}$	α_V	Red.
NaCl	DFT ref.	–	–	–	2.70	–
	PET baseline	36.99	37.67	38.44	3.84	–
	PET-NoRRA	39.66	40.19	40.73	**2.64**	94.8%
LiCl	FPMD ref.	–	–	–	2.87	–
	PET baseline	29.95	30.88	32.10	3.47	–
	PET-NoRRA	33.25	34.28	35.15	**2.77**	83.3%
LiCl–KCl	FPMD ref.	–	–	–	2.73	–
	PET baseline	38.99	39.84	40.45	3.57	–
	PET-NoRRA	41.70	42.35	43.39	**3.03**	64.4%

E.3 Molten-NaCl radial distribution functions

The RDF comparison uses four independent 500 ps NVT trajectories at 1100 K, initialized from different frames of the Tovey et al. DFT reference trajectory (Tovey et al., 2020). Each simulation used a 0.5 fs time step and non-conservative model forces. For atom types α and β, the partial radial distribution function is

$$g_{\alpha\beta}(r) = \frac{V}{N_\alpha N_\beta} \left\langle \sum_{i\in\alpha} \sum_{j\in\beta, j\neq i} \delta(r - \|\mathbf{r}_i - \mathbf{r}_j\|) \right\rangle \Big/ 4\pi r^2 \,\Delta r, \tag{36}$$

where N_α and N_β are the counts of type α and β, V is the cell volume, and Δr is the histogram bin width; the angular brackets denote an average over saved trajectory frames including the minimum-image convention. Na–Na, Na–Cl, and Cl–Cl RDFs were evaluated on the common grid of the released reference analysis. The shaded regions in Figure 5 are standard deviations over the four starts; each pair uses its own vertical scale so that the complete curve remains visible. The total RDF error reported in the main text is the mean, over starts and pairs, of

$$\mathrm{RMSE}_{\alpha\beta} = \sqrt{\left\langle \left(g_{\alpha\beta}^{\mathrm{model}}(r) - g_{\alpha\beta}^{\mathrm{ref}}(r) \right)^2 \right\rangle_r}. \tag{37}$$

PET-NoRRA reduces the mean RDF RMSE of all three pair correlations relative to the PET baseline. Across the four starts, the mean total RDF error is 0.04113 ± 0.01033 for the PET baseline and

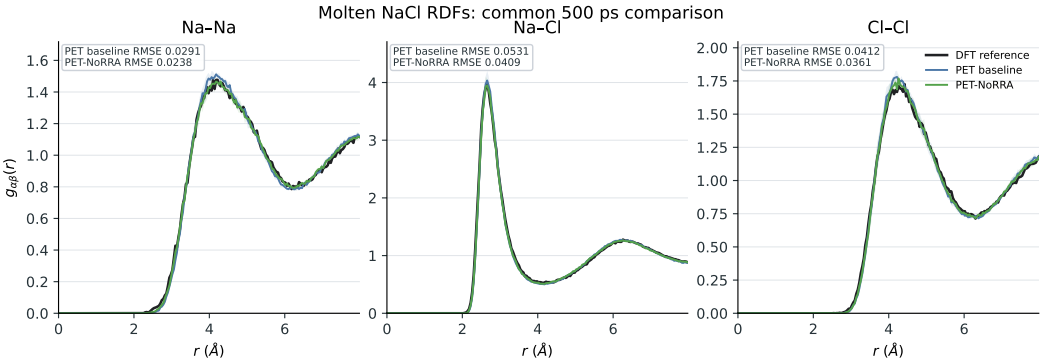

Figure 5: Molten-NaCl partial RDFs from four independent 500 ps simulations. Black, blue, and green curves show the DFT reference, PET baseline, and PET-NoRRA, respectively; shaded bands denote standard deviations across starts. In each panel, the annotated RMSE is the corresponding per-start pair RMSE averaged over the four starts.

0.03360 ± 0.01314 for PET-NoRRA. The corresponding Na–Na, Na–Cl, and Cl–Cl means decrease from 0.02912, 0.05313, and 0.04115 to 0.02378, 0.04087, and 0.03615, respectively.

E.4 Auxiliary NaCl LO–TO splitting diagnostic

As an additional phonon diagnostic, we constructed finite-displacement force constants for a $3\times3\times3$ conventional NaCl supercell (216 atoms) and evaluated Γ-point frequencies with Phonopy (Togo et al., 2023). The splitting arises from the long-range Coulomb field produced by the ionic displacements and tests whether long-range electrostatics enter the vibrational response. The calculation uses the experimental non-analytical correction rather than model-predicted Born charges. The experimental reference splitting is 2.660 THz, from the neutron-scattering TO/LO frequencies of Raunio et al. (Raunio et al., 1969). Following the LO–TO protocol used for environment-dependent-charge potentials (Korogod et al., 2026), a direction-dependent non-analytical correction is added to the dynamical matrix. For a wave vector $\mathbf{q} \to \mathbf{0}$ along the longitudinal direction, the dipole–dipole force-constant term is

$$\Phi^{\mathrm{NAC}}_{\kappa\kappa'\alpha\beta}(\mathbf{q}\to\mathbf{0}) = \frac{4\pi}{\Omega}\frac{(\mathbf{q}\cdot\mathbf{Z}^*_\kappa)_\alpha\,(\mathbf{q}\cdot\mathbf{Z}^*_{\kappa'})_\beta}{\sqrt{M_\kappa M_{\kappa'}}\;\mathbf{q}\cdot\boldsymbol{\epsilon}_\infty\cdot\mathbf{q}}, \tag{38}$$

where Ω is the unit-cell volume, M_κ are atomic masses, and the correction is evaluated with experimental isotropic parameters $Z^*_{\mathrm{Na}} = +1.12$, $Z^*_{\mathrm{Cl}} = -1.12$, and $\epsilon_\infty = 2.33$ (Messaoudi et al., 2015; Lowndes & Martin, 1969). The corrected Γ-point dynamical matrix is diagonalized along a longitudinal direction; TO and LO frequencies are extracted from the threefold-degenerate transverse and split longitudinal optical modes, respectively. This diagnostic uses the auxiliary response-only checkpoint from the SOG-width sweep, not the q-prior NoRRA checkpoint used for thermal expansion and RDF. All PET rows use the same experimental non-analytical correction. The auxiliary checkpoint gives 5.760 and 8.417 THz for the TO and LO branches, so the splitting is 2.657 THz.

F MatterSim backbone portability

F.1 Portability results and run-specific protocols

Panel C of Table 3 lists the systems in which both energy and force improve. The paragraphs below give the corresponding absolute values, epochs, and subsets.

Electrolyte-B. On the 904-frame subset at epoch 1999, the 1-block MatterSim baseline gives 4.187×10^{-4} eV/atom and 3.169×10^{-2} eV/Å, while 1-block MatterSim-NoRRA gives 2.392×10^{-4} eV/atom and 1.713×10^{-2} eV/Å. On the full 1809-frame set at the same epoch, the energy MAE changes from 3.893×10^{-4} to 4.253×10^{-4} eV/atom ($+9.2\%$), while the force MAE improves from 2.686×10^{-2} to 1.416×10^{-2} eV/Å (-47.3%).

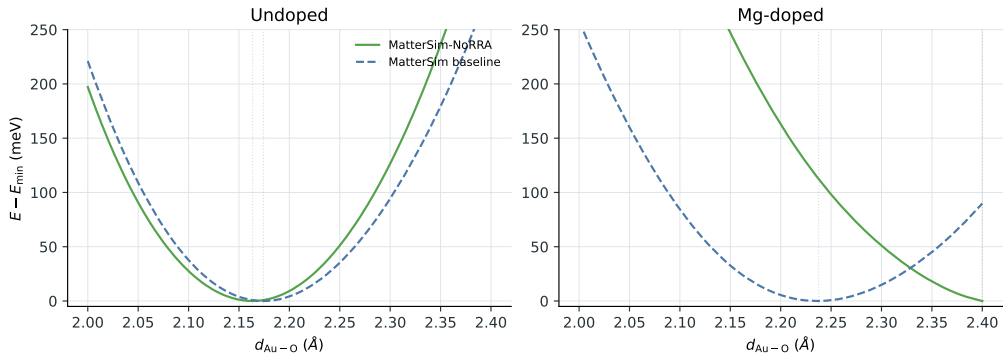

Figure 6: Au–O interaction scans on the Au/MgO support. Energies are relative to each curve minimum. Vertical dotted lines mark the curve minima. In the Mg-doped panel, MATTERSIM-NORRA shifts the equilibrium Au–O distance outward by 0.162 Å relative to the MatterSim baseline, indicating a qualitatively different nonlocal electrostatic response.

Table 9: Au/MgO physical-diagnostic summary. Curve minima are from the Au–O scans in Figure 6; adsorption-state differences are $\Delta E = E(\text{3-state}) - E(\text{1-state})$ with errors relative to DFT.

Quantity	Model	Undoped	Mg-doped	Unit
Curve minimum $d_{\mathrm{Au-O}}$	MatterSim-NoRRA	2.164	2.400	Å
	MatterSim baseline	2.174	2.238	Å
Adsorption ΔE	DFT	+934.8	−66.9	meV
	MatterSim-NoRRA	+935.0	−71.5	meV
	MatterSim baseline	+565.5	+505.2	meV
ΔE error	MatterSim-NoRRA	+0.2	−4.6	meV
	MatterSim baseline	−369.3	+572.1	meV

$Na_{8/9}Cl_8^+$ ionic clusters. The evaluation covers all 5000 frames. Total-energy MAE decreases from 2.318×10^{-2} to 8.226×10^{-3} eV and force MAE from 3.790×10^{-2} to 1.252×10^{-2} eV/Å.

Deep-water1900. Both models use 1900 configurations of 300 atoms each. The MatterSim baseline reaches 1.053×10^{-3} eV/atom, 5.843×10^{-2} eV/Å, and 3.487×10^{-1} stress MAE at epoch 648; MATTERSIM-NORRA reaches 5.361×10^{-4} eV/atom, 4.480×10^{-2} eV/Å, and 2.294×10^{-1} at epoch 424.

Water/Pt interface. The MatterSim experiment uses a 5000-frame subset of PT111. At epochs 999 (MatterSim baseline) and 920 (MATTERSIM-NORRA), energy MAE decreases from 5.587×10^{-4} to 4.001×10^{-4} eV/atom and force MAE from 3.642×10^{-2} to 2.836×10^{-2} eV/Å.

F.2 AU/MGO STAGED CASE STUDY

The original total energies are near 10^6 eV, causing float32 energy differences to be quantized around 0.125 eV. After energy-shift preprocessing, a MatterSim cosine baseline (2 blocks, 64 units, 4000 epochs) gives held-out E/F MAE of 2.196×10^{-3} eV/atom and 4.683×10^{-2} eV/Å. The MATTERSIM-NORRA model uses staged continuation (ordinary training, then force ratio 1.0, then force ratio 0.01, resetting optimizer and scheduler at each stage) and reaches 8.913×10^{-5} eV/atom and 1.356×10^{-2} eV/Å. For the adsorption-state energy difference $\Delta E = E(\text{3-state}) - E(\text{1-state})$, MATTERSIM-NORRA gives −71.50 meV (doped) and +934.99 meV (undoped), within 4.6 and 0.2 meV of DFT, whereas the MatterSim baseline errs by hundreds of meV (Table 9).

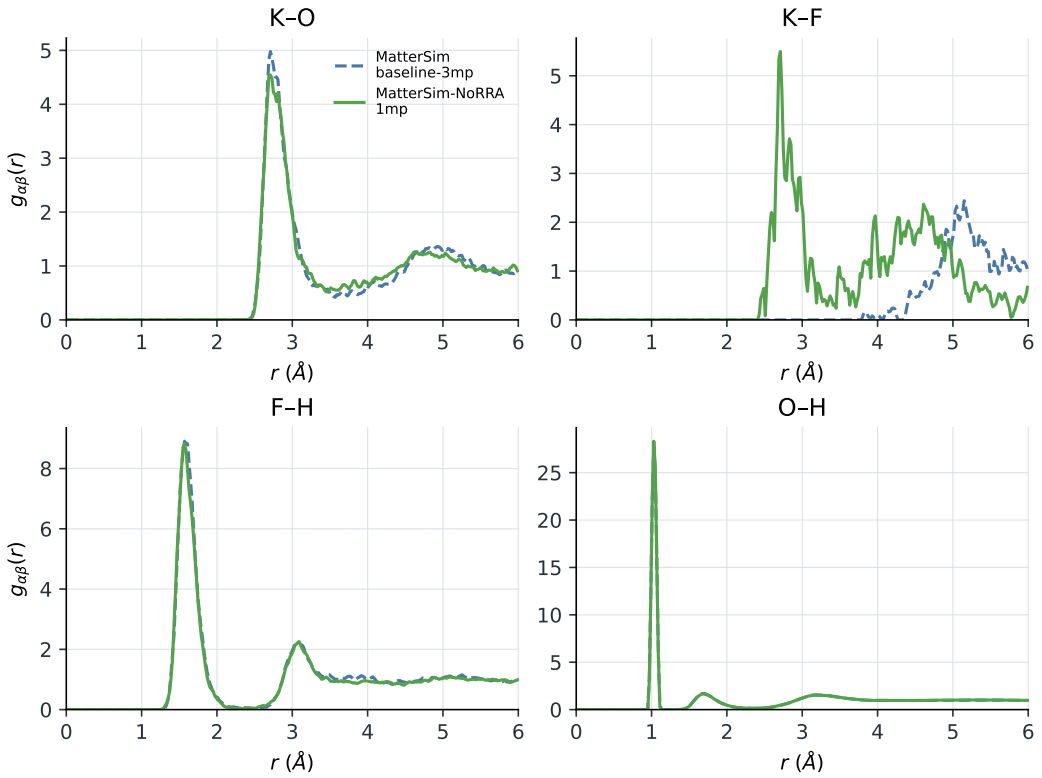

Figure 7: Partial RDFs from the 100 ps KF-electrolyte MD production run. Blue and green curves show the 3-block MatterSim baseline and 1-block MatterSim-NoRRA, respectively. Curves are lightly smoothed for visualization; the archived RDF values are unchanged. The 1-block NoRRA model introduces K–F pairing while preserving the dominant K–O solvation shell and intramolecular O–H peak.

F.3 KF ELECTROLYTE MOLECULAR DYNAMICS

The electrolyte MD used a 196-atom cubic cell (64 H_2O, 2 K, 2 F; 12.42 Å per side), initialized from frame 0 of the Electrolyte-B training structure. All runs used NVT dynamics at approximately 300 K with a 1 fs time step, a 100 ps production segment, and 500 saved frames at 0.2 ps intervals. Here "1mp" and "3mp" denote the number of MatterSim message-passing blocks: the 3-block MatterSim baseline has 227,549 parameters, whereas the 1-block MATTERSIM-NORRA model has 97,863. The 1-block MatterSim baseline collapsed during MD, so the reported control is the stable 3-block baseline.

The MATTERSIM-NORRA model lowers the lowest-k charge structure factor from 9.94×10^{-4} (3-block MatterSim baseline) to 6.60×10^{-4} (-33.6%), consistent with improved charge-structure fidelity (Figure 8). The corresponding K–F coordination of 0.194 is closer to the expected ion-pairing physics than the baseline value of 0. Figure 7 shows the K–O, K–F, F–H, and O–H partial RDFs for the two retained stable models: MATTERSIM-NORRA introduces nonzero K–F pairing while preserving the dominant K–O solvation shell and the intramolecular O–H peak.

F.4 ADDITIONAL MATTERSIM RUNS

LES-BEC water. Both runs complete epoch 999. Energy MAE changes from 6.240×10^{-4} to 6.889×10^{-4} eV/atom ($+10.4\%$), while force MAE improves from 4.652×10^{-2} to 3.889×10^{-2} eV/Å (-16.4%).

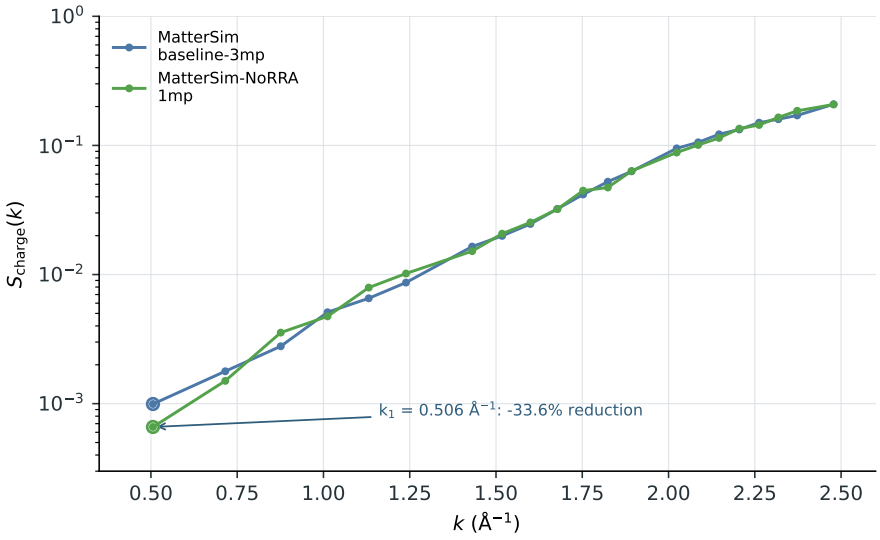

Figure 8: Charge structure factor $S_{\mathrm{charge}}(k)$ in the KF-electrolyte MD run, plotted without k^2 normalization on a logarithmic vertical axis. Open circles mark the first nonzero wave vector k_1, where MatterSim-NoRRA reduces the long-wavelength charge fluctuation by 33.6% relative to the 3-block MatterSim baseline.

G Component controls and inference efficiency

G.1 Pure-short-range NaCl control

To check that the interaction component does not forcibly assign a large contribution to data without long-range physics, we evaluated the ratio $|E_{\mathrm{LR}}|/|E_{\mathrm{total}}|$ on 151 pure-short-range NaCl test configurations. The mean and median ratios are 0.366% and 0.364%, indicating that the interaction component remains small when it is not needed.

G.2 Periodic NaCl latency benchmark

Panel B of Table 4 compares the PET baseline with PET-NoRRA on cubic rock-salt NaCl supercells with lattice constant 5.64 Å. System sizes are 64 to 8192 atoms. Each size uses one ideal structure and four rattled replicas, with rattle standard deviations 0.005, 0.010, 0.020, and 0.030 Å. For each structure, ten warm-up calls are discarded and fifty calls are timed. The two models alternate within every repeat. Timing synchronizes before and after each complete energy and force call, and internal stage profiling is off. The parameter counts are $4{,}487{,}962$ and $4{,}505{,}348$.

H Per-configuration robustness of dedicated-system improvements

Table 3 reports aggregate error changes. This section asks whether those gains are broad across validation configurations rather than driven by a small subset. Panels A and B summarize the aggregate PET and MatterSim relative changes; Panels C and D show per-configuration relative energy and force improvements for three retained systems with frozen per-configuration archives. Relative improvements can be heavy-tailed when the baseline error is close to zero, so the figure reports medians and interquartile ranges rather than means.

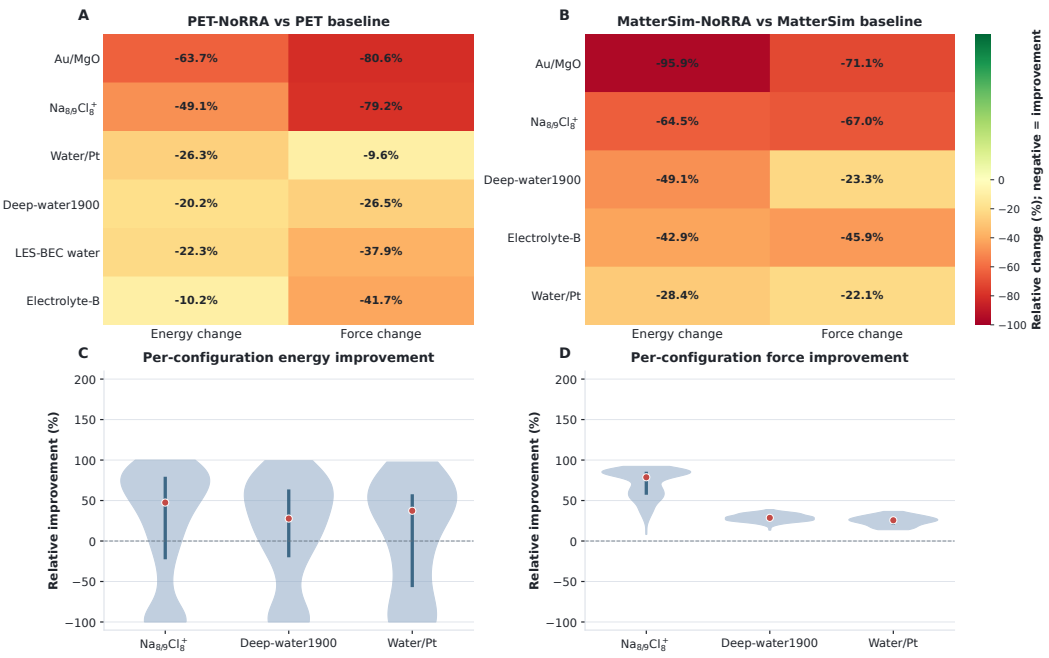

Figure 9: Supplementary dedicated-system relative improvements corresponding to Table 3. (A) PET-NoRRA relative energy and force changes under the energy-calibrated checkpoint rule in Panel B. (B) MatterSim-NoRRA relative changes for the jointly improved systems in Panel C. (C,D) Per-configuration relative energy and force improvement versus the PET baseline on three retained validation sets with archived per-configuration predictions: NaCl clusters (n=500), Deep-water1900 (n=190), and Water/Pt (n=50). Positive values in Panels C and D indicate improvement. Violin tails are clipped to $[-100, 200]\%$ for readability because relative errors are heavy-tailed when the baseline error is near zero; red markers show medians and bars show interquartile ranges.