

Machine learning interatomic potential can infer electrical response

Peichen Zhong,^{1,*} Dongjin Kim,^{2,*} Daniel S. King,^{1,*} and Bingqing Cheng^{1, 2, 3, †}

¹*Bakar Institute of Digital Materials for the Planet, UC Berkeley, California 94720, United States*

²*Department of Chemistry, UC Berkeley, California 94720, United States*

³*The Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria*

(Dated: April 8, 2025)

Modeling the response of material and chemical systems to electric fields remains a longstanding challenge. Machine learning interatomic potentials (MLIPs) offer an efficient and scalable alternative to quantum mechanical methods but do not by themselves incorporate electrical response. Here, we show that polarization and Born effective charge (BEC) tensors can be directly extracted from long-range MLIPs within the Latent Ewald Summation (LES) framework, solely by learning from energy and force data. Using this approach, we predict the infrared spectra of bulk water under zero or finite external electric fields, ionic conductivities of high-pressure superionic ice, and the phase transition and hysteresis in ferroelectric PbTiO₃ perovskite. This work thus extends the capability of MLIPs to predict electrical response—without training on charges or polarization or BECs—and enables accurate modeling of electric-field-driven processes in diverse systems at scale.

I. INTRODUCTION

The polarization \mathbf{P} of a system underlies many electrical response properties including capacitance, dielectric constant, ferroelectricity, piezoelectricity, ionic conductivity, and infrared (IR) spectra. The Born effective charge (BEC) tensor Z^* quantifies the variation in \mathbf{P} due to an atomic displacement at position \mathbf{r}_i of atom i [1]:

$$Z_{i\alpha\beta}^* = \frac{\partial P_\alpha}{\partial r_{i\beta}} = \frac{\partial \mathcal{F}_{i\alpha}}{\partial \mathcal{E}_\beta^0}, \quad (1)$$

where α and β label Cartesian directions. The second part of Eqn. (1) links the electrostatic force \mathcal{F}_i on atom i resulting from an external electric field \mathcal{E}^0 to the system.

Modeling electrical response properties has long been a challenge. The Berry phase definition of the polarization [2, 3] of periodic insulators can be obtained from density functional theory (DFT) calculations. Density-functional perturbation theory (DFPT) [4] or the finite field method [5] can be used to compute BECs and other derivatives of the polarization. However, the computational costs associated with such ab initio methods limit their applications to large systems or long timescales. On the other hand, fixed-charge or polarizable empirical force fields are cheap but may lack quantitative accuracy or transferability [6, 7].

Standard machine learning interatomic potentials (MLIPs) [8, 9], which learn surrogate potential energy surfaces from quantum mechanical reference calculations, are typically short-ranged and do not explicitly consider electrostatics. Several approaches have been developed to incorporate long-range interactions, such as learning DFT-derived partial charges [10–13] or maximally localized Wannier centers [14, 15], or employing long-range descriptors [16, 17] or long-range message-passing [18]. Latent Ewald Summation (LES) [19, 20] is a recent method

that learns a long-range energy contribution E^{lr} by fitting to the total potential energy E and atomic forces F of configurations:

$$E^{\text{lr}} = \frac{1}{2\varepsilon_0 V} \sum_{0 < k < k_c} \frac{1}{k^2} e^{-\sigma^2 k^2/2} |S(\mathbf{k})|^2, \quad (2)$$

where \mathbf{k} is the reciprocal wave vector and V is the cell volume. The structure factor $S(\mathbf{k})$ given by

$$S(\mathbf{k}) = \sum_{i=1}^N q_i^{\text{les}} e^{i\mathbf{k}\cdot\mathbf{r}_i}, \quad (3)$$

and the LES charges q_i^{les} are predicted using a neural network based on local invariant features B_i of atom i . LES can be combined with any short-ranged MLIP architecture, such as descriptor-based [21–23] or message-passing neural networks [24, 25].

However, a natural inclusion of electric response directly within the MLIP frameworks is missing. Currently, \mathbf{P} and BEC have to be learned separately as tensorial properties [26, 27], e.g. directly predicting BEC [28] or local contribution to the dipole moment of a molecule [26] based on atomic local environments, or as the derivatives of the electric enthalpy [29]. Conceptually, this is in contrast with the electronic structure picture of matter: electron density and nuclear positions fully determine how the system will interact with electric field.

Here, we show that polarization and BEC tensors can be naturally derived from long-range MLIPs within the LES framework. This enables accurate predictions of electrical response properties, such as IR spectra and conductivity, solely by learning from energies and forces. Importantly, it is straightforward to add an external electric field to MLIP-driven molecular dynamics (MD) simulations, enabling the exploration of electric-field-driven phenomena in various materials and molecules. We demonstrate this method on a range of complex bulk systems, including molecular liquids, ionic liquids, superionic crystals, and ferroelectric materials.

* These authors contributed equally.

† bingqingcheng@berkeley.edu

II. THEORY

Building on the pioneering Molecular Dynamics in Electronic Continuum (MDEC) model [6, 7, 30, 31], the Coulomb interactions between the “free charges” of atoms are explicitly considered, while the rapidly responding background electrons are treated as a dielectric medium. The electrostatic field produced by the free atomic charge q_i of atom i can then be expressed as

$$\mathcal{E}_i(\mathbf{r}) = \frac{q_i}{4\pi\varepsilon_0\varepsilon_\infty|\mathbf{r} - \mathbf{r}_i|^3} (\mathbf{r} - \mathbf{r}_i), \quad (4)$$

and the resultant electric force between two atoms is given by:

$$\mathcal{F}_{ij} = \mathcal{E}_i(\mathbf{r}_{ij})q_j, \quad (5)$$

where ε_0 represents the vacuum permittivity, and ε_∞ is the high-frequency relative permittivity (also known as the static or electronic dielectric constant), which can be determined experimentally (e.g., from the square of the optical refractive index) or calculated using DFPT with frozen nuclei [32].

We interpret the LES charges q_i^{les} as scaled charges, $q_i^{\text{les}} = q_i/\sqrt{\varepsilon_\infty}$, acting both as the sources and the receivers of the electric field:

$$\mathcal{F}_{ij} = \frac{q_i^{\text{les}} q_j^{\text{les}}}{4\pi\varepsilon_0 r_{ij}^3} \mathbf{r}_{ij}. \quad (6)$$

In the LES algorithm, q_i^{les} are optimized to effectively describe the long-range component of the total potential energy and atomic forces. For modeling the energetics and dynamics of a system under no external field, these LES charges provide a self-contained description of electrostatic interactions, with no further adjustment needed. To model electrical response, however, the LES charges can be unscaled to recover the atomic charges q_i .

The polarization of a finite system such as a gas-phase molecule is

$$\mathbf{P} = \sum_{i=1}^N q_i \mathbf{r}_i, \quad (7)$$

and the BEC Z^* can be obtained by taking its derivative with respect to \mathbf{r} :

$$Z_{i\alpha\beta}^* = \frac{\partial P_\alpha}{\partial r_{i\beta}} = q_i \delta_{\alpha,\beta} + \sum_{j=1}^N r_{j\alpha} \frac{\partial q_j}{\partial r_{i\beta}}. \quad (8)$$

As shown in the second part of Eqn. (8), the BEC tensor Z_i^* comes from two contributions: the charge q_i , and the dependence of the charges on atomic positions.

For modeling either crystalline or disordered bulk systems, it is almost mandatory to apply periodic boundary (PBC) conditions. Importantly, the value of \mathbf{P} cannot be uniquely defined for systems with PBC, according to the modern theory of polarization [2, 3]. To circumvent

such ambiguity, we propose a generalized formulation of polarization $P(k)$ under PBC:

$$P_\alpha(k) = \sum_{i=1}^N \frac{q_i}{ik} \exp(ikr_{i\alpha}), \quad (9)$$

where $P_\alpha(k)$ is the polarization along the α direction, $k = 2\pi/L_\alpha$, and L_α is the length of a periodic cubic cell, while the extension to triclinic cells is straightforward. At the limit of $k \rightarrow 0$, $P(k)$ becomes the finite-system expression in Eqn. (7). The BEC tensor Z^* of atom i can then be evaluated as

$$Z_{i\alpha\beta}^* = \Re \left[\exp(-ikr_{i\alpha}) \frac{\partial P_\alpha(k)}{\partial r_{i\beta}} \right]. \quad (10)$$

The prediction of BEC enables the calculation of several electrical response properties. For instance, the current of the polarization of the system can be obtained as $\mathbf{J}(t) = \sum_{i=1}^N Z_i^*(t) \cdot \mathbf{v}_i(t)$. The current-current autocorrelation function encodes the ionic electrical conductivity σ via the Green-Kubo formula,

$$\sigma = \frac{1}{3Vk_B T} \int_0^\infty dt \langle \mathbf{J}(0)\mathbf{J}(t) \rangle, \quad (11)$$

and the IR spectra via Fourier transform,

$$I(\omega) \propto \int_0^T dt \langle \mathbf{J}(0)\mathbf{J}(t) \rangle e^{-i\omega t}. \quad (12)$$

Moreover, once BEC are computed using Eqn. (10), one can apply a real-valued constant electric field \mathcal{E}^0 to the system by adding the electrostatic force \mathcal{F}_i on each atom using the second part of Eqn. (1) for the linear response regime, thus enabling constant-electric-field simulations under PBC.

III. EXAMPLES

A. Water

The theoretical prediction for the IR spectrum of water is a classic problem but still not fully resolved, and more so with the presence of external electric fields [33]. We used the Cartesian Atomic Cluster Expansion (CACE) [34] potential as the short-ranged MLIP and LES [19, 20] as the long-range part, and thereafter refer to this combination as CACE-LR. The RPBE-D3 bulk water dataset from Ref. [28] contains energies and forces of 654 configurations (90% train/ 10% test split) each of 64 water molecules. Even though we used a compact CACE model with a cutoff of 4.5 Å and no message passing, the test root mean square errors (RMSEs) of energy and forces (0.25 meV/atom and 21 meV/Å) are a fraction compared to the errors in Ref. [28] (0.8 meV/atom and 60 meV/Å).

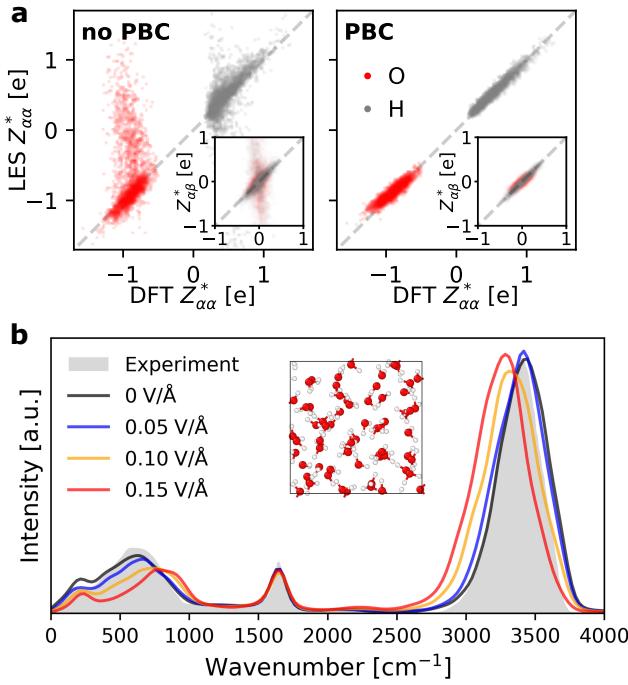


Figure 1. Electrical response of the RPBE-D3 bulk water. **a** compares the Born effective charge tensors (Z^*) computed from DFT and predicted using the LES method. The CACE-LR was trained on the energies and forces of the RPBE-D3 bulk water dataset [28]. The main panels compare the diagonal elements of BEC ($Z_{\alpha\alpha}^*$), and the insets show the off-diagonal elements ($Z_{\alpha\beta}^*$ with $\alpha \neq \beta$). The left panel (no PBC) corresponds to the LES BECs calculated assuming no periodic boundary condition using Eqn. (8), and the right panel (PBC) shows PBC obtained using the generalized polarization form in Eqns. (9) and (10). **b** shows the infrared (IR) absorption spectra of bulk liquid water in the absence of an external field (black line) and under varying external field intensities (colored lines) as indicated in the legend. The experimental IR spectrum in the absence of an external field [35] (gray shading) is included for reference.

Fig. 1a compares the BEC predicted by LES and calculated by the reference DFT for 100 water configurations at experimental density and room temperature [28]. If neglecting PBC and naively using Eqn. (8), the predicted BEC exhibited significant discrepancies from the DFT reference values, particularly for atoms near the edge of the simulation cell. In contrast, by properly accounting for PBC using Eqn. (9) and Eqn. (10), the LES BECs agree well with DFT for both diagonal and off-diagonal components. This shows that it is necessary to resolve the ambiguity of \mathbf{P} for periodic systems using a scheme like Eqn. (9).

We performed MLIP-driven NVT simulations of bulk water (0.25 fs timestep and 200,000 MD steps) under the Nosé-Hoover thermostat, at 300 K and experimental density, and extracted the IR spectra using Eqn. (12) based on the LES BECs. As shown in Fig. 1b (black curve),

not only are the predicted shapes and positions of the intramolecular vibrational modes such as OH stretching band ($\approx 3400 \text{ cm}^{-1}$) and bending ($\approx 1640 \text{ cm}^{-1}$) mode band in excellent agreement with experiment [35], but also the intermolecular low-frequency libration mode band position [36] ($\approx 650 \text{ cm}^{-1}$) and the hydrogen-bond translational stretching mode [37] ($\approx 200 \text{ cm}^{-1}$) are well captured. Notably, it is necessary to use the time-dependent BEC tensors to compute the IR: if using instead the fixed nominal charges of $q^H = +1$ for hydrogen and $q^O = -2$ for oxygen (see Fig. 6 in Methods), the shape of the predicted IR is much worse and the hydrogen-bond translational stretching band is completely absent.

We then performed NVT MD simulations for bulk water under static constant external fields \mathcal{E}^0 (0.05 V/Å, 0.1 V/Å, or 0.15 V/Å) along the z -direction, by adding electrostatic forces on all atoms according to the second part of Eqn. (1). As shown in Fig. 1b (colored curves), at higher electric field intensities, the intermolecular librational band ($\approx 650 \text{ cm}^{-1}$) blue shifts and the intramolecular OH stretching band ($\approx 3400 \text{ cm}^{-1}$) red shifts. These trends are consistent with previous studies using DFT molecular dynamics [33, 38]. The red-shift of the OH stretching band is generally associated with stronger hydrogen bonding [39] and with more ice-like structures [40]. The blue shift of the low-frequency libration mode can be attributed to the enhanced restrictions on the rotational motion of water molecules imposed by the H-bonds [33].

B. Superionic water

This example aims to illustrate the capability of our BEC inference method for ionic and superionic systems, and its generalizability across a wide range of conditions. Water at megabar pressures and thousands of Kelvins exhibits diverse structural and dynamical behaviors: ionic water with partially dissociated hydrogen atoms (Fig. 2a), the face-centred cubic (FCC) superionic phase (ice XVIII) with liquid-like hydrogen atoms while oxygen atoms remain on the crystalline lattice [41, 42] (Fig. 2b), and ice X with a body-centered cubic (BCC) lattice of oxygen atoms [43] (Fig. 2c).

We trained the CACE-LR model using 5,000 configurations (90% train/ 10% test split), randomly selected out of the 17,516 structures in the original MLIP training set that was compiled for predicting the phase diagram of superionic water spanning a wide range of thermodynamic conditions up to thousands of kelvin and megabar pressures [42]. Despite the smaller training set and a relatively light-weight architecture choice of the CACE-LR model, the test errors on energies and forces are halved compared to the original study: 7 meV/atom in energy RMSE and 327 meV/Å in force RMSE, compared to 14 meV/atom and 740 meV/Å of the original MLIP [42], respectively.

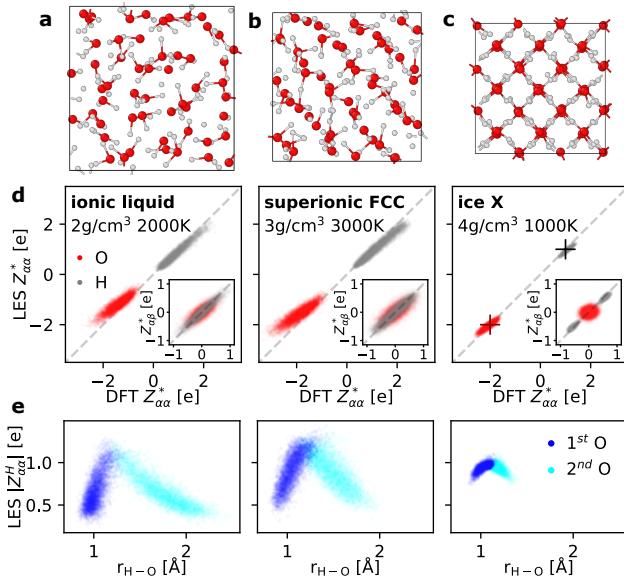


Figure 2. Analysis of the Born effective charges (BECs) in different phases of high-pressure water. **a** corresponds to partially ionic liquid water, **b** shows face-centred cubic (FCC) superionic phase (ice XVIII), and **c** is ice X. The oxygen-hydrogen bonds are drawn with a cutoff of 1.1 Å. **d** compares the BEC tensors (Z^*) computed from DFT and predicted using the LES method, for 100 configurations of each phase at the specified condition. The CACE-LR was trained based on the energies and forces from the superionic water dataset [42]. The main panels compare the diagonal elements of BEC ($Z_{\alpha\alpha}^*$), and the insets show the off-diagonal elements ($Z_{\alpha\beta}^*$ with $\alpha \neq \beta$). **e** illustrates the correlation between the mean diagonal values of Z^* of all hydrogen atoms, and the distances to their nearest two oxygen atoms.

Fig. 2b shows that the predicted LES BECs agree well with the ground-truth DFT values for three distinct phases (illustrated in Fig. 2a-c) under different thermodynamic conditions. At 2 g/cm³ and 2000 K, the liquid water is partially molecular and partially dissociated with frequent proton jumps. The diagonal BEC values for hydrogen atoms ($Z_{\alpha\alpha}^H$) can sometimes exceed the nominal charge of +1. In Fig. 2e, we correlate the distances between all H atoms and their nearest two oxygen atoms with the mean BEC diagonal values ($3|Z_{\alpha\alpha}^H|^2 = (Z_{xx}^H)^2 + (Z_{yy}^H)^2 + (Z_{zz}^H)^2$). This shows that the anomalously large BEC occurs when a hydrogen atom breaks the bond with its nearest oxygen and forms a bond with the second nearest oxygen, analogous to the Grotthuss mechanism. At 3 g/cm³ and 3000 K, the FCC superionic ice exhibits even larger fluctuations of BECs for both hydrogen and oxygen. The anomalous BECs of hydrogen are again correlated to the O-H bond breaking and formation, and such events are more frequent. At 4 g/cm³ and 1000 K, the stable phase is ice X with a BCC lattice of oxygen atoms, and hydrogen atoms are evenly positioned between two neighboring oxygen atoms with straight O-H-O bonds. The diagonal values of BEC

have very small fluctuations, and are centered around the oxidation numbers of +1 for hydrogen and -2 for oxygen ions (shown as crosses in the right panel of Fig. 2d). Intriguingly, the off-diagonal elements of BEC for H show two separate clusters at positive and negative values.

The ionic electrical conductivity σ is crucial for characterizing ionic and superionic systems. To compute σ at 2 g/cm³ and 2000 K, we employed a CACE-LR model that was finetuned using energy, forces, and BEC values of 100 configurations at the same condition. We performed an equilibrium MD simulation of 120 ps duration with 54 water formula units, and such system size was selected to be directly comparable to the previous DFT MD simulation from Ref. [44]. we calculated the current-current correlation functions $C(t) = \langle J(0)J(t) \rangle / 3e$ either using the time-dependent BEC tensor $Z^*(t)$ or the fixed nominal charges of $q^H = +1$ and $q^O = -2$. These correlation functions are shown in Fig. 3a, and at short times they are in perfect agreement with a tour-de-force DFT and BEC calculation of 15 ps from Ref. [44]. Fig. 3b shows the σ values computed by integrating the corresponding $C(t)$ functions using the Helfand-Einstein formula [45], a reduced-variance form of the Green-Kubo relation in Eqn. (11). Interestingly, the estimated σ with time-dependent BEC tensors (32 ± 2 S/cm) is similar to the estimate of $\sigma = 37 \pm 2$ S/cm with the constant $q^H = +1$ and $q^O = -2$. Such observation was also made in Ref. [44], and was rationalized in Ref. [46] from a topological quantization argument albeit only for atomic liquids with all species adiabatically staying in the same motifs without changing oxidation states.

One can also compute σ from non-equilibrium MD simulations under external electric fields. Fig. 3c shows the total displacement of charges in the system over time under different values of the external field \mathcal{E}^0 along the z direction, $D(t) = \sum_i q_i(z_i(t) - z_i(0))$, computed using the constant charges $q^H = +1$ and $q^O = -2$. σ can be estimated from the slope of $D(t)$ as $\sigma = \langle dD(t)/dt \rangle / V\mathcal{E}^0$. Such σ values from these finite-electric-field simulations, as displayed in the legend of Fig. 3c, are consistent with the equilibrium results, but have better statistical convergence and also avoid the problems associated with the Green-Kubo integration to the infinite time limit (Eqn. (11)).

C. Ferroelectric perovskite

Ferroelectric materials are unique in that they exhibit spontaneous and permanent electric polarization, and this polarization can be reversed by applying an electric field [47]. Anomalously large BECs that exceed the nominal charges of ions are often considered hallmarks of ferroelectric materials [48–50]. Here, we will demonstrate that our method can predict the anomalous BECs and model the characteristic P - \mathcal{E} hysteresis loop in the prototypical PbTiO₃ ferroelectric perovskite. At $T = 300$ K, PbTiO₃ exhibits a tetragonal phase characterized by

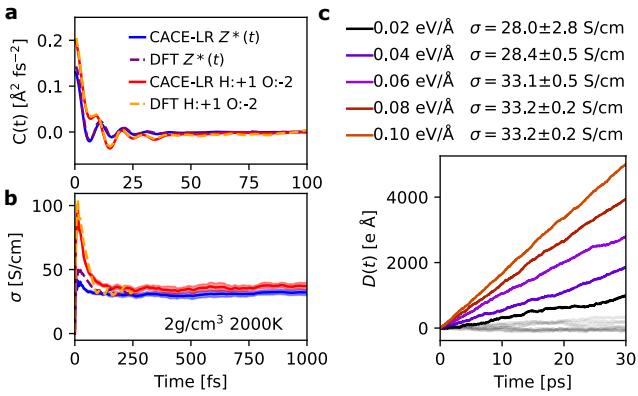


Figure 3. Ionic transport properties of the partially ionic liquid water at 2 g/cm³ and 2000 K. **a** shows the current-current correlation functions $C(t)$ computed using either time-dependent Born effective charge tensors $Z^*(t)$ or fixed nominal charges. **b** plots the corresponding time integrals for estimating the ionic electrical conductivity σ . In **a** and **b**, the DFT molecular dynamics results are from Ref. [44]. **c** illustrates the time-dependent charge displacement from CACE-LR molecular dynamics simulations under constant external electric fields with the specified intensities. The colored lines show the displacements along the direction of the applied field, and gray lines show the displacements along the other orthogonal directions.

a short axis a and a long axis c , with the ratio c/a correlated with the polarization magnitude (see Fig. 4a).

We trained the CACE-LR model using the energies and forces of the PbTiO₃ dataset from Ref. [51], computed using SCAN DFT. The potential achieved test RMSEs of 0.4 meV/atom in energy and 79.8 meV/Å in forces, much reduced from the original model with 1 meV/atom in energy RMSE and 350 meV/Å in force RMSE from Ref. [51]. Without explicitly learning the BEC, Fig. 4b compares the BEC tensors for 45 atomic configurations (including both cubic and tetragonal phase) predicted using LES and computed using PBE DFT. Not only do the LES predictions agree well with the DFT reference, the anomalously large diagonal values of BECs ($Z_{\alpha\alpha}^*$) relative to the nominal ionic charges (indicated by the plus signs in Fig. 4b) for $q^{Pb} = +2$, $q^{Ti} = +4$, and $q^O = -2$ are captured. The anomalous BECs in ferroelectrics are typically considered to come from a complex interplay of global charge transfer, mixed ionic-covalent bonding, and the hybridization between oxygen and transition metal orbitals [49]. The successful prediction of the BECs here thus not only showcases the expressiveness of the LES method, but also indicates that CACE-LR is able to capture the intricate long-range electrostatic interactions in ferroelectric materials [52]. Unlike previous approaches [27, 51, 53] that employ separate models for potential energy surfaces and polarization, CACE-LR properly embeds the latter into the former and do not explicitly train on polarization.

To characterize the spontaneous polarization of

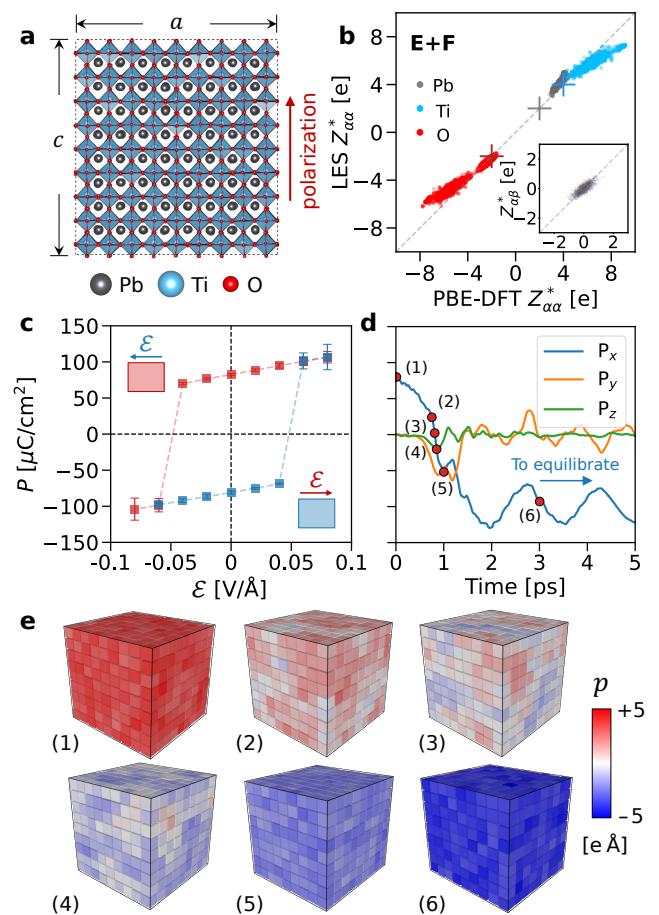


Figure 4. Polarization and Born effective charge tensors in the PbTiO₃ perovskite. **a** A snapshot of equilibrated PbTiO₃ at $T = 300$ K. **b** The BEC tensors (Z^*) computed from LES versus from PBE-DFT calculations for 45 randomly selected configurations. The + symbol indicates the nominal charge of Pb/Ti/O in black/blue/red color. **c** P - \mathcal{E} hysteresis loop computed from CACE-LR MD simulations under different external electric fields. **d** Time evolution of the total polarization during the non-equilibrium MD with $\mathcal{E} = -0.06$ V/Å applied. **e** Local dipole spatial distributions through the polarization reversal event. Each voxel represents a Ti-centered unit cell.

PbTiO₃ in the absence of an external electric field, we performed equilibrium NPT simulations at $T = 300$ K (see Methods). Because SCAN-DFT overestimates $c/a = 1.14$ for the ground state tetragonal PbTiO₃ structure compared to $c/a = 1.06$ from experiments, we applied an isotropic external pressure of 2.8 GPa in the NPT simulations as used in Ref. [51], yielding $c/a = 1.07$ for the equilibrated structures in MD. We then calculated the total polarization of the material as [54]:

$$P_\alpha = \frac{e}{V} \sum_{i\beta} Z_{i\alpha\beta}^* \Delta u_{i\beta}, \quad (13)$$

where $\Delta u_{i\beta}$ indicates the atomic displacement from the non-polar centrosymmetric reference state. The polar-

ization predicted based on the LES BEC is $P = 82 \pm 2 \mu\text{C}/\text{cm}^2$, consistent with the experimentally reported polarization $P(25^\circ\text{C}) = 81 \mu\text{C}/\text{cm}^2$ [55].

We then simulated the P - \mathcal{E} hysteresis loop of PbTiO_3 at $T = 300$ K, by applying an electric field \mathcal{E} along and against the polarization direction of the equilibrated structure with a magnitude ranging from 0.02 to 0.08 V/ \AA . Fig. 4c presents the time-averaged polarization at various external field strengths after equilibration, demonstrating polarization reversal to a negative value at $\mathcal{E} = -0.06$ V/ \AA . Starting from the MD structure at $\mathcal{E} = -0.08$ V/ \AA , we simulated the reverse polarization process by varying the electric field from -0.06 to 0.08 V/ \AA (blue line in Fig. 4c), which exhibits the expected reverse transition behavior and completes the ferroelectric hysteresis loop. Note that the width of the hysteresis loop will be dependent on the system size and simulation time, as phase transition is an activated event, and the hysteresis loop in Fig. 4c aims to demonstrate the existence of hysteresis in the ferroelectric PbTiO_3 .

Fig. 4d illustrates the time evolution of the total polarization during the non-equilibrium MD trajectory after applying $\mathcal{E} = -0.06$ V/ \AA to the original equilibrated structure at $T = 300$ K. The polarization drops rapidly from its spontaneous value to negative values, followed by fluctuations before equilibrating to the new field-induced equilibrium state. To visualize the atomic-level process for the polarization reversal, we calculated the local dipole moment p for each Ti-centered unit cell using $p_\alpha = e \sum_j w_j Z_{j\alpha\beta}^* \Delta u_{j\beta}$, where the sum runs over all atoms j in the unit cell, and w_j represents the weighting factor (1 for Ti, 1/8 for Pb, 1/2 for O). Fig. 4e shows the spatial distributions of p from representative MD snapshots, with the color of each voxel corresponding to the magnitude of p along the c axis. Snapshots (2)–(4) reveal the initial stage of the reversal: domains of opposite polarization nucleate, grow, coalesce, and ultimately form a uniformly negatively polarized state.

IV. DISCUSSION AND CONCLUSIONS

The central thesis of the present paper is that MLIPs can infer the response of a material or chemical system to an external electric field, by fitting latent charges (q^{les}) just from the energies and forces of atomic configurations. Although previous works provide some hints on such a connection, e.g. by showing that machine-learned charges are correlated with DFT charges [20, 56], this paper provides a clear physical picture and provides convincing demonstrations on diverse systems.

Our approach involves several conceptual advances. First, we recognize that LES charges are physical charges that can be determined by fitting to the energies and forces, and that it is not necessary nor advantageous to explicitly fit to the semiclassical partial charges from DFT [20]. Indeed, the DFT partial charges are not physical observables and they depend on the specific partition-

ing strategy used [57–59] and are less indicative of atomic charge states in oxides with charge transfer effects [60]. Our approach of not fitting to the DFT partial charges is in contrast with several other existing long-range MLIP methods [11–14].

Second, we incorporate the background fast-responding charges into the high frequency relative permittivity ϵ_∞ , while explicitly accounting for the Coulomb interactions between the atomic charges. This is similar to the concepts of screened Coulomb interactions and the scaled ion charge in the MDEC model [30, 31]. MDEC is foundational for designing and justifying non-polarizable force fields using scaled ion charges [6, 7], including the popular SPC-type water models [61, 62]. The accuracy of the LES BEC based on $q_i = \sqrt{\epsilon_\infty} q_i^{\text{les}}$ provides a “smoking-gun” validation of the MDEC theory. In addition, our approach is a cleaner way to use the charge scaling framework, as the flexible charges are learned from ab initio data. In comparison, the empirical force fields have the scaled charges and other parameters fitted at the same time to experimental data, which means the errors in describing the electrostatic interactions can be partially canceled by tuning other non-bonded parameters and vice versa, so the resulting charges are less reflective of the true underlying electrostatics. Moreover, the LES framework assigns flexible charges based on local atomic environments. Such environment-dependent charges are more expressive than the fixed charges in empirical force fields. For example, the same LES model is capable of predicting BECs for dramatically different phases of water including isolating ice, superionic ice, and ionic water (Fig. 2). It is difficult to imagine a fixed-charge model to match this level of expressiveness.

Third, we derive the Born effective charge tensor Z_i^* of each atom from the predicted unscaled charges q , by taking the derivative of the total polarization \mathbf{P} with respect to atomic positions. For periodic-boundary-condition systems, where \mathbf{P} is not well-defined, we develop a generalized formulation (Eqn. (9)). Unlike the DFT partial charge, which suffers from the lack of a unified definition, the BEC tensors are physical observables. The fact that the LES BECs align well with the DFT BECs proves that the LES charges are physical charges, despite not being trained on DFT partial charges. Moreover, the link between q and Z_i^* gives the option to train or finetune the MLIP using DFT BECs, e.g. as we did for the water, the superionic water and the PbTiO_3 system in the Methods section. Successful prediction of BECs is also practically useful, as it can be used to predict a number of electrical response properties such as IR and conductivity. It also provides the linear response of forces to an applied electric field (Eqn. (1)), enabling straightforward incorporation of external fields in MLIP molecular dynamics simulations.

We demonstrated the framework on a diverse set of complex bulk systems, including liquid water, ionic high-pressure water, superionic water ice, insulating ice X, and

ferroelectric PbTiO_3 . The LES predicted Z^* are largely in good agreement with DFT, even when trained only on energies and forces, with further improvement possible by explicitly incorporating DFT BECs during the training. Notably, the model generalizes well across different phases and thermodynamic conditions, as seen in high-pressure water (Fig.2). The framework also enables quantitative predictions of key electrical response properties, such as the IR spectra of water, ionic conductivity in high-pressure water, and the P - E hysteresis loop in PbTiO_3 .

To conclude, this paper resolves a critical limitation of state-of-the-art machine learning interatomic potentials: their inability to intrinsically predict electrical response properties. By rigorously linking the latent charges to the experimentally measurable BECs, we bridge the gap between MLIPs and the electrostatics in quantum mechanical systems. Our framework provides a systematic approach to develop and refine MLIPs for modeling polarizable systems under external fields, unlocking numerous applications such as electrolyte design, modeling electrochemical interfaces, piezoelectrics, and pyroelectrics.

V. METHODS

A. Water

The RPBE-D3 bulk water dataset contains energies and forces of 654 configurations (64 water molecules in each snapshot), which were generated via an on-the-fly learning scheme from MD trajectories at different temperatures [28]. The original MLIP trained on this RPBE-D3 data from Ref. [28] has a RMSE of 0.8 meV/atom and 60 meV/ \AA for energies and forces, respectively. Ref. [28] also provided an additional RPBE-D3 dataset of 100 configurations that were separately collected from NVT simulations at experimental density and 298.2 K. This set includes Born effective charges in addition to energies and forces. We will refer to this set as RPBE-D3 + BEC.

We trained four versions of CACE-LR: (1) trained solely on energies and forces with a cutoff radius $r_{\text{cut}} = 5.5 \text{ \AA}$ using RPBE-D3 data, (2) trained solely on energies and forces with $r_{\text{cut}} = 4.5 \text{ \AA}$ using RPBE-D3 data, (3) trained solely on energies and forces with a smaller dataset and $r_{\text{cut}} = 4.5 \text{ \AA}$ using RPBE-D3 + BEC data, and (4) trained on energies, forces, and Born effective charges with a smaller dataset and $r_{\text{cut}} = 4.5 \text{ \AA}$ using RPBE-D3 + BEC data. For the CACE representation, we used 6 trainable Bessel radial functions, $c = 12$, $l_{\text{max}} = 3$, $\nu_{\text{max}} = 3$, $N_{\text{embedding}} = 2$, no message passing, 1-dimensional hidden variable, $\sigma = 1 \text{ \AA}$, and $k_c = \pi$ ($dl = 2 \text{ \AA}$).

Table I summarizes the number of configurations in each data set used for training, along with the applied cutoff settings and the corresponding RMSE values for energies and forces.

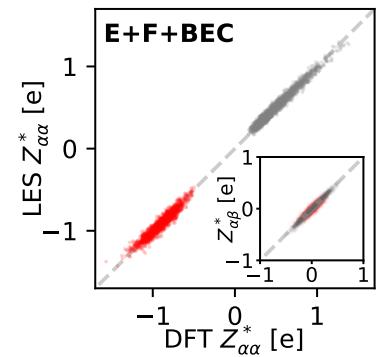


Figure 5. Comparison of DFT BEC and LES BEC using the version (4) potential that is trained 100 structures with energy, forces, and BEC.

DFPT calculations with the RPBE-D3 functional using VASP predict the high-frequency permittivity (ϵ_∞) of water at experimental density and room temperature to be 1.83. We use this value when converting LES charges q^{les} to the physical charges q . Such value is very close to the experimental value of 1.78 for water (with the refractive index of water of 1.333 being the square root of this value).

Fig. 5 compares the reference DFT and LES BEC predicted by the model trained on BEC (version 4) for the same 100 water configurations used in Fig. 1. As expected, training directly on BEC data improves agreement between LES and reference DFT results, although the improvement is modest. This indicates that while prediction accuracy can be further enhanced by training with BEC data, the potential trained exclusively on energies and forces (version 2, Fig. 1) already exhibits sufficiently high accuracy. In the main text, all the reported results are from the version (2).

Version	1 E+F	2 E+F	3 E+F	4 E+F+BEC
N_{config}	654	654	100	100
$r_{\text{cut}}(\text{\AA})$	5.5	4.5	4.5	4.5
E	0.22	0.25	0.26	0.19
F	18.88	21.01	23.84	25.34

Table I. Performance of four versions of the CACE-LR potentials on each test set. Errors are reported via RMSE in meV/atom for energy and in meV/ \AA for forces.

We performed equilibrium NVT simulations in ASE at a density of 0.997 g/cm^3 and 300 K for a system of 64 water molecules, employing the Nosé-Hoover thermostat. The finite-field MD simulations followed the same setups. As the sum of Z^* is not exactly zero due to the small residual prediction errors, in the finite-field MD simulations the total mean forces on all atoms were removed every step to eliminate the non-zero center-of-mass velocities arising under the electric field. Although such mean forces do not affect any physical observables, they

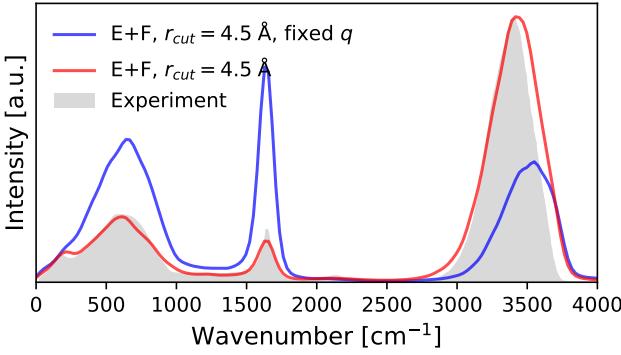


Figure 6. IR spectrum computed using fixed nominal charges ($q^H = +1$ for hydrogen and $q^O = -2$ for oxygen) based on the MD trajectory generated with model version (2). The experimental IR spectrum [35] is included as gray shading for reference.

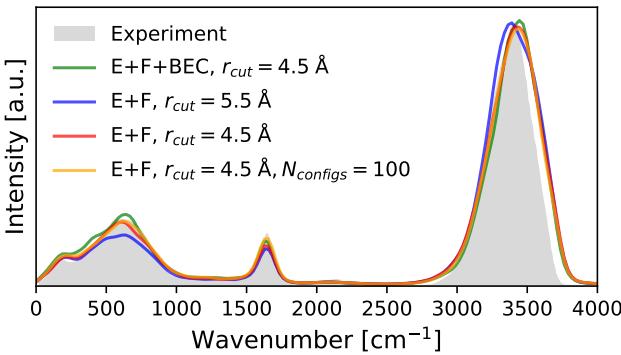


Figure 7. Water IR spectrum with four differently trained MLIP. Comparison of computational IR spectra of liquid water obtained from four differently trained MLIPs based on RPBE-D3 data. The experimental IR spectrum [35] is included as gray shading for reference.

can interfere with the thermostat and the visualization of the trajectories. In all cases, MD simulations were conducted for 50 ps with a time-step of 0.25 fs. IR spectra were calculated from the MD trajectories employing Eqn. (12), which involves computing the polarization current of the system, $\mathbf{J}(t) = \sum_{i=1}^N Z_i^*(t) \cdot \mathbf{v}_i(t)$. A Gaussian filter was applied following the Fourier transform, and each IR spectrum was normalized by its integrated area.

Fig. 6 shows that the IR spectrum calculated using fixed nominal charges does not reproduce peak intensities and shapes well. Moreover, it completely fails to capture the hydrogen-bond stretching mode at approximately 200 cm⁻¹. These results highlight the importance of accurately predicting Born effective charges to obtain detailed and reliable IR spectra.

Fig. 7 presents computational IR spectra of liquid water obtained using four differently trained potentials (Table. I). All four versions of the CACE-LR potentials exhibit consistent peak positions and intensity trends,

closely matching the experimental IR spectrum. Notably, these potentials differ in training dataset size, cutoff radius, and inclusion of Born effective charges, demonstrating robustness and reliability across varying training conditions.

B. Superionic water

The original training set of superionic water has 17,516 configurations (90% train/10% test split), spanning a wide range of thermodynamic conditions (300 K–15000 K, 1 g/cm³–7 g/cm³), and it was trained using N2P2 [63] with a cutoff of 12 Bohr, yielding test RMSE errors of 14 meV/atom in energy, and 740 meV/A in forces [42].

We randomly selected 5,000 configurations (90% train/10% test split) from the original dataset. For the CACE-SR part, we used $r_{\text{cut}} = 3.5$ Å, 6 Bessel radial functions, $c = 12$, $l_{\max} = 3$, $\nu_{\max} = 3$, $N_{\text{embedding}} = 3$, and 1 message passing layer. The LES model uses a one-dimensional hidden variable, $\sigma = 1$ Å, and $k_c = \pi$ ($dl = 2$ Å). The test RMSEs are 7 meV/atom in energy, and 327 meV/A in forces.

For comparing BECs, we randomly selected 100 configurations of 54 water molecules at 3 g/cm³ and 3000 K from DFT MD trajectories from Ref. [44]. We also selected 100 uncorrelated configurations of 54 water molecules from MLP MD simulations at 2 g/cm³ and 2000 K, and at 4 g/cm³ and 1000 K. We employed VASP to calculate the Born effective charge tensor for these configurations using DFPT, with a plane-wave cutoff of 400 eV at the Baldereschi point, consistent with Ref. [44]. From DFPT, we also computed the high-frequency relative permittivity at the three conditions: $\epsilon_\infty = 3.1$ at 2 g/cm³ and 2000 K, $\epsilon_\infty = 4.2$ at 3 g/cm³ and 3000 K, and $\epsilon_\infty = 3.7$ at 4 g/cm³ and 1000 K. These values were used to infer the BECs.

We further finetuned the CACE-LR model using the energy, forces, and BEC values of the 100 configurations (90% train/10% test split) at 2 g/cm³ and 2000 K. Before the finetuning, the RMSE errors in energy, forces, and BEC are 4.5 meV/atom, 103 meV/A, and 0.136 e, respectively. After the finetuning, the test RMSE errors reduced to 0.67 meV/atom and 101 meV/A, and 0.09 e, respectively.

For computing conductivities, we used this finetuned model to perform equilibrium NVT simulations in ASE at 2 g/cm³ and 2000 K for a system of water molecules, employing the Nosé-Hoover thermostat. The timestep was set to 0.3 fs, and the simulation length is 120 ps (400,000 steps) following 3 ps of equilibration. The finite-field MD simulations follow a similar setup, except that a shorter simulation time of 30 ps (100,000 steps) was used.

C. PbTiO₃ perovskite

To fit the CACE-LR potential, we used the original training (4432 configurations) and test datasets (600 configurations) from Ref. [51] and randomly allocated 10% of the original training data as the validation set. Both the training and test dataset contain the SCAN-DFT calculated energy and forces of PbTiO₃ atomic configurations from DP-GEN MD simulations at 300/600/900 K, covering the cubic (ferroelectric) and tetragonal (paraelectric) phases in $3 \times 3 \times 3$ PbTiO₃ unit cells (135 atoms). For the CACE-SR part, we used $r_{\text{cut}} = 6.0$ Å, 6 Bessel radial functions, $c = 12$, $l_{\max} = 3$, $\nu_{\max} = 3$, $N_{\text{embedding}} = 3$, and 1 message passing layer. The LES model uses a one-dimensional hidden variable, $\sigma = 1$ Å, and $k_c = \pi$ ($dl = 2$ Å). The CACE-LR potential trained only on SCAN DFT energy and forces denotes $E + F$ model.

To evaluate the LES BECs against DFT BECs, we randomly selected 443 atomic configurations from the training set for DFT calculations. These configurations were further split into train/val/test sets with a ratio of 8:1:1 for fine-tuning purposes. The test set comprised 45 atomic configurations and was used for comparison in Fig. 4. For computing LES BECs, we replicated the supercells of these configurations $3 \times 3 \times 3$, to eliminate the finite-size effects due to finite k . As a proof of concept, the DFT-BEC was calculated with DFPT at the generalized gradient approximation (GGA) level of accuracy, as the SCAN functional is not currently supported for DFPT. According to our test, the LEC-BEC derived from SCAN-DFT demonstrates good agreement with the PBE-DFT BEC. The $E + F$ model was further fine-tuned with the PBE-DFT energy, forces, and BEC. Fig. 8 displays the comparison after fine-tuning, which demonstrates a marginal improvement from an RMSE of 0.586 e to 0.384 e for diagonal components and from 0.257 e to 0.254 e for off-diagonal components of the BEC tensors when comparing the $E + F$ model to the fine-tuned model. The DFPT calculations were performed using VASP with the PBE54 functional [64], a Gamma-centered k -point, and a plane-wave energy cutoff of 680 eV. The DFT calculations were performed on $3 \times 3 \times 3$ the PbTiO₃ unit cell, with convergence of 10^{-6} eV in total energy.

For computing the ferroelectric properties of PbTiO₃, we employed the $E + F$ model to perform equilibrium NPT simulations using ASE at $P = P_0 + P_a$ with the Nosé-Hoover thermostat. Here, P_0 represents the ambient pressure (1 bar) and $P_a = 2.8$ GPa is an applied correction to compensate for the DFT overestimation of the c/a ratio as suggested in Ref. [65]. The simulation structure was initialized with a $9 \times 9 \times 9$ supercell of the cubic PbTiO₃ unit cell (space group $Pm3\bar{m}$). The MD simulations were conducted with a timestep of 2 fs. For simulations without external electric fields, we performed 100 ps production runs following 10 ps of equilibration. The finite-field MD simulations followed a similar protocol, except that a shorter simulation time of 50 ps was

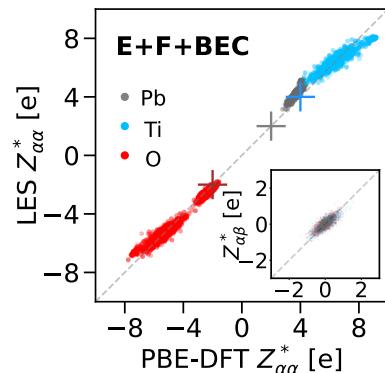


Figure 8. Comparison of DFT BEC and LES BEC using the potential that is trained 354 structures with energy, forces, and BEC.

used.

To estimate the high-frequency dielectric constant ε_∞ , we additionally performed MD simulations with a $3 \times 3 \times 3$ supercell at $T = 300$ K to sample equilibrated structures from the MD trajectories. These atomic configurations were subsequently analyzed using DFPT to obtain the microscopic dielectric constant. We calculated the averaged diagonal value $\varepsilon_\infty = 1/N_{\alpha} \sum_{i\alpha} \varepsilon_{\alpha\alpha} = 7.533$ of the dielectric constant tensor, which was then used as the scaling factor of $\sqrt{\varepsilon_\infty}/9.48933$ to compute the polarization of PbTiO₃ at large scales from MD simulations. For the plots in Fig. 4b-e, a parity transformation of -1 was applied to align the BEC and polarization direction with conventional notation.

D. Notes on implementation

The LES method was implemented in the CACE code, <https://github.com/BingqingCheng/cace>. In practice, we use an **Atomwise** module to predict an internal hidden charge $q_i^{\text{raw}} = Q_\phi(B_i)$ based on a set of local invariant representations B_i . The long-range energy is then computed using an **Ewald** module as

$$E^{\text{lr}} = \frac{2\pi}{V} \sum_{0 < k < k_c} \frac{1}{k^2} e^{-\sigma^2 k^2/2} \left| \sum_{i=1}^N q_i^{\text{raw}} e^{ik \cdot r_i} \right|^2. \quad (14)$$

To obtain the LES charges q^{les} in the unit of [e], the internal hidden charges q^{raw} should be scaled by a factor of $1/9.48933$, due to the internal normalization factor used ($1/2\varepsilon_0 = 1$).

We then use a **Polarization** module to compute the polarization of the system based on q_i^{raw} . If the system is finite, the non-periodic expression (Eqn. (7)) is used, and if the system is periodic, the generalized polarization in Eqn. (9) is used. One can add a normalization factor in this module. The default setting is to remove the mean average charge before computing the polarization. If the

factor of $\sqrt{\epsilon_\infty}/9.48933$ is used, the correct magnitude of the polarization will be recovered.

We then use the `Grad` module to take the derivative of polarization with respect to atomic positions using auto-grad (see Eqn. (8)). For finite systems, this step already provides the BECs. For periodic systems, however, we need to use the `Dephase` module to remove the complex phase factor in Eqn. (10), in order to get the real-valued BECs.

Data availability The training sets, training scripts, and trained CACE potentials are available at <https://github.com/BingqingCheng/LES-BEC>.

Code availability The CACE package is publicly available at <https://github.com/BingqingCheng/>

cace.

Acknowledgments The authors thank for valuable discussions with Pinchen Xie, David Limmer, Jeff Neaton and Greg Voth. The authors thank Sébastien Hamel for providing the DFT MD trajectories for superionic water, and help clarifying questions related to the pseudopotentials. The authors thank Federico Grasselli and Stefano Baroni for providing data and notebooks for computing the conductivity of a molten salt. D.S.K. and P.Z. acknowledge funding from BIDMaP Postdoctoral Fellowship.

Competing Interests B.C. has an equity stake in AIMATX Inc. University of California, Berkeley has filed a provisional patent for the Latent Ewald Summation algorithm.

-
- [1] C-Yu Wang, S Sharma, EKU Gross, and JK Dewhurst, “Dynamical born effective charges,” *Physical Review B* **106**, L180303 (2022).
- [2] Raffaele Resta, “Macroscopic polarization in crystalline dielectrics: the geometric phase approach,” *Reviews of modern physics* **66**, 899 (1994).
- [3] Raffaele Resta and David Vanderbilt, “Theory of polarization: a modern approach,” in *Physics of ferroelectrics: a modern perspective* (Springer, 2007) pp. 31–68.
- [4] Xavier Gonze and Changyol Lee, “Dynamical matrices, born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory,” *Physical Review B* **55**, 10355 (1997).
- [5] Chao Zhang, Thomas Sayer, Jürg Hutter, and Michiel Sprik, “Modelling electrochemical systems with finite field molecular dynamics,” *Journal of Physics: Energy* **2**, 032005 (2020).
- [6] Brian J Kirby and Pavel Jungwirth, “Charge scaling manifesto: A way of reconciling the inherently macroscopic and microscopic natures of molecular simulations,” *The journal of physical chemistry letters* **10**, 7531–7536 (2019).
- [7] Miguel Jorge, “Theoretically grounded approaches to account for polarization effects in fixed-charge force fields,” *The Journal of Chemical Physics* **161** (2024).
- [8] John A Keith, Valentin Vassilev-Galindo, Bingqing Cheng, Stefan Chmiela, Michael Gastegger, Klaus-Robert Müller, and Alexandre Tkatchenko, “Combining machine learning and computational chemistry for predictive insights into chemical systems,” *Chemical reviews* **121**, 9816–9872 (2021).
- [9] Oliver T Unke, Stefan Chmiela, Huziel E Saucedo, Michael Gastegger, Igor Poltavsky, Kristof T Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller, “Machine learning force fields,” *Chemical Reviews* **121**, 10142–10186 (2021).
- [10] Oliver T Unke and Markus Meuwly, “Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges,” *Journal of chemical theory and computation* **15**, 3678–3693 (2019).
- [11] Tsz Wai Ko, Jonas A Finkler, Stefan Goedecker, and Jörg Behler, “A fourth-generation high-dimensional neural network potential with accurate electrostatics includ-
- ing non-local charge transfer,” *Nature communications* **12**, 398 (2021).
- [12] Sheng Gong, Yumin Zhang, Zhenliang Mu, Zhichen Pu, Hongyi Wang, Zhiao Yu, Mengyi Chen, Tianze Zheng, Zhi Wang, Lifei Chen, *et al.*, “Bamboo: a predictive and transferable machine learning force field framework for liquid electrolyte development,” arXiv preprint arXiv:2404.07181 (2024).
- [13] 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).
- [14] Linfeng Zhang, Han Wang, Maria Carolina Muniz, Athanassios Z Panagiotopoulos, Roberto Car, *et al.*, “A deep potential model with long-range electrostatic interactions,” *The Journal of Chemical Physics* **156** (2022).
- [15] Ang Gao and Richard C Remsing, “Self-consistent determination of long-range electrostatics in neural network potentials,” *Nature communications* **13**, 1572 (2022).
- [16] Andrea Grisafi and Michele Ceriotti, “Incorporating long-range physics in atomic-scale machine learning,” *The Journal of chemical physics* **151** (2019).
- [17] Carolin Faller, Merzuk Kaltak, and Georg Kresse, “Density-based long-range electrostatic descriptors for machine learning force fields,” arXiv preprint arXiv:2406.17595 (2024).
- [18] Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann, “Ewald-based long-range message passing for molecular graphs,” in *International Conference on Machine Learning* (PMLR, 2023) pp. 17544–17563.
- [19] Bingqing Cheng, “Latent ewald summation for machine learning of long-range interactions,” arXiv preprint arXiv:2408.15165 (2024).
- [20] Dongjin Kim, Daniel S King, Peichen Zhong, and Bingqing Cheng, “Learning charges and long-range interactions from energies and forces,” arXiv preprint arXiv:2412.15455 (2024).
- [21] Jörg Behler and Michele Parrinello, “Generalized neural-network representation of high-dimensional potential-energy surfaces,” *Physical review letters* **98**, 146401 (2007).

- [22] Alexander V Shapeev, “Moment tensor potentials: A class of systematically improvable interatomic potentials,” *Multiscale Modeling & Simulation* **14**, 1153–1173 (2016).
- [23] Ralf Drautz, “Atomic cluster expansion for accurate and transferable interatomic potentials,” *Physical Review B* **99**, 014104 (2019).
- [24] 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).
- [25] Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi, “Mace: Higher order equivariant message passing neural networks for fast and accurate force fields,” *Advances in Neural Information Processing Systems* **35**, 11423–11436 (2022).
- [26] Andrea Grisafi, David M Wilkins, Gábor Csányi, and Michele Ceriotti, “Symmetry-adapted machine learning for tensorial properties of atomistic systems,” *Physical review letters* **120**, 036002 (2018).
- [27] Elia Stocco, Christian Carbogno, and Mariana Rossi, “Electric-field driven nuclear dynamics of liquids and solids from a multi-valued machine-learned dipolar model,” (2025).
- [28] Bernhard Schmiedmayer and Georg Kresse, “Derivative learning of tensorial quantities—predicting finite temperature infrared spectra from first principles,” *The Journal of Chemical Physics* **161** (2024), 10.1063/5.0217243.
- [29] Stefano Falletta, Andrea Cepellotti, Anders Johansson, Chuin Wei Tan, Albert Musaelian, Cameron J Owen, and Boris Kozinsky, “Unified differentiable learning of electric response,” *arXiv preprint arXiv:2403.17207* (2024).
- [30] IV Leontyev, MV Vener, IV Rostov, MV Basilevsky, and Marshall D Newton, “Continuum level treatment of electronic polarization in the framework of molecular simulations of solvation effects,” *The Journal of chemical physics* **119**, 8024–8037 (2003).
- [31] Igor Leontyev and Alexei Stuchebrukhov, “Accounting for electronic polarization in non-polarizable force fields,” *Physical Chemistry Chemical Physics* **13**, 2613–2626 (2011).
- [32] Ardavan Farahvash, Igor Leontyev, and Alexei Stuchebrukhov, “Dynamic and electronic polarization corrections to the dielectric constant of water,” *The Journal of Physical Chemistry A* **122**, 9243–9250 (2018).
- [33] Giuseppe Cassone, Jiri Sponer, Sebastiano Trusso, and Franz Saija, “Ab initio spectroscopy of water under electric fields,” *Physical Chemistry Chemical Physics* **21**, 21205–21212 (2019).
- [34] Bingqing Cheng, “Cartesian atomic cluster expansion for machine learning interatomic potentials,” *npj Computational Materials* **10**, 157 (2024).
- [35] John E. Bertie and Zhida Lan, “Infrared intensities of liquids xx: The intensity of the oh stretching band of liquid water revisited, and the best current values of the optical constants of h₂o(l) at 25°C between 15,000 and 1 cm⁻¹,” *Applied Spectroscopy* **50**, 1047–1057 (1996).
- [36] Pier Luigi Silvestrelli, M. Bernasconi, and Michele Parrinello, “Ab initio infrared spectrum of liquid water,” *Chemical Physics Letters* **277**, 478–482 (1997).
- [37] Matthias Heyden, Jian Sun, Stefan Funkner, Gerald Mathias, Harald Forbert, Martina Havenith, and Dominik Marx, “Dissecting the thz spectrum of liquid water from first principles via correlations in time and space,” *Proceedings of the National Academy of Sciences* **107**, 12068–12073 (2010).
- [38] Zdenek Futera and Niall J. English, “Communication: Influence of external static and alternating electric fields on water from long-time non-equilibrium ab initio molecular dynamics,” *The Journal of Chemical Physics* **147**, 031102 (2017).
- [39] Deepak Ojha, Kristof Karhan, and Thomas D. Kühne, “On the hydrogen bond strength and vibrational spectroscopy of liquid water,” *Scientific Reports* **8**, 16888 (2018).
- [40] Zhaohui Wang, Andrei Pakoulev, Yoonsoo Pang, and Dana D. Dlott, “Vibrational substructure in the oh stretching transition of water and hod,” *The Journal of Physical Chemistry A* **108**, 9054–9063 (2004).
- [41] M. Millot, F. Coppapi, J. R. Rygg, A. Correa Barrios, S. Hamel, D. C. Swift, and Jon H. Eggert, “Nanosecond X-ray diffraction of shock-compressed superionic water ice,” *Nature* **569**, 251–255 (2019).
- [42] Bingqing Cheng, Mandy Bethkenhagen, Chris J Pickard, and Sébastien Hamel, “Phase behaviours of superionic water at planetary conditions,” *Nature physics* **17**, 1228–1232 (2021).
- [43] Aleks Reinhardt, Mandy Bethkenhagen, Federica Coppapi, Marius Millot, Sébastien Hamel, and Bingqing Cheng, “Thermodynamics of high-pressure ice phases explored with atomistic simulations,” *Nature Communications* **13**, 4707 (2022).
- [44] Martin French, Sébastien Hamel, and Ronald Redmer, “Dynamical screening and ionic conductivity in water from ab initio simulations,” *Physical review letters* **107**, 185901 (2011).
- [45] Federico Grasselli and Stefano Baroni, “Invariance principles in the theory and computation of transport coefficients,” *The European Physical Journal B* **94**, 1–14 (2021).
- [46] Federico Grasselli and Stefano Baroni, “Topological quantization and gauge invariance of charge transport in liquid insulators,” *Nature Physics* **15**, 967–972 (2019).
- [47] Karin M Rabe, Matthew Dawber, Céline Lichtensteiger, Charles H Ahn, and Jean-Marc Triscone, “Modern physics of ferroelectrics: Essential background,” in *Physics of Ferroelectrics: A Modern Perspective* (Springer, 2007) pp. 1–30.
- [48] R Resta and S Sorella, “Many-body effects on polarization and dynamical charges in a partly covalent polar insulator,” *Physical review letters* **74**, 4738 (1995).
- [49] Ph Ghosez, J-P Michenaud, and Xavier Gonze, “Dynamical atomic charges: The case of ab o 3 compounds,” *Physical Review B* **58**, 6224 (1998).
- [50] Alessio Filippetti and Nicola A Spaldin, “Strong-correlation effects in born effective charges,” *Physical Review B* **68**, 045111 (2003).
- [51] Pinchen Xie, Yixiao Chen, Weinan E, and Roberto Car, “Thermal disorder and phonon softening in the ferroelectric phase transition of lead titanate,” *Phys. Rev. B* **111**, 094113 (2025).
- [52] W. Zhong, R. D. King-Smith, and David Vanderbilt, “Giant lo-to splittings in perovskite ferroelectrics,” *Physical Review Letters* **72**, 3618–3621 (1994).
- [53] Lorenzo Gigli, Max Veit, Michele Kotiuga, Giovanni

- Pizzi, Nicola Marzari, and Michele Ceriotti, "Thermodynamics and dielectric response of batio₃ by data-driven modeling," npj Computational Materials **8**, 209 (2022).
- [54] Karin M Rabe and Philippe Ghosez, "First-principles studies of ferroelectric oxides," in *Physics of Ferroelectrics: A Modern Perspective* (Springer, 2007) pp. 117–174.
- [55] Gerald Burns and Bruce A. Scott, "Lattice Modes in Ferroelectric Perovskites: PbTiO₃," Phys. Rev. B **7**, 3088–3101 (1973).
- [56] Zichen Song, Jian Han, Graeme Henkelman, and Lei Li, "Charge-optimized electrostatic interaction atom-centered neural network algorithm," Journal of Chemical Theory and Computation **20**, 2088–2097 (2024).
- [57] F Weinhold, CR Landis, and ED Glendening, "What is nbo analysis and how is it useful?" International reviews in physical chemistry **35**, 399–440 (2016).
- [58] Toon Verstraelen, Steven Vandenbrande, Farnaz Heidar-Zadeh, Louis Vanduyfhuys, Veronique Van Speybroeck, Michel Waroquier, and Paul W Ayers, "Minimal basis iterative stockholder: atoms in molecules for force-field development," Journal of Chemical Theory and Computation **12**, 3894–3912 (2016).
- [59] Aleksandr V Marenich, Steven V Jerome, Christopher J Cramer, and Donald G Truhlar, "Charge model 5: An extension of hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases," Journal of chemical theory and computation **8**, 527–541 (2012).
- [60] C. Wolverton and Alex Zunger, "First-principles prediction of vacancy order-disorder and intercalation battery voltages in Li_xCoO₂," Physical Review Letters **81**, 606–609 (1998).
- [61] IV Leontyev and AA Stuchebrukhov, "Electronic polarizability and the effective pair potentials of water," Journal of chemical theory and computation **6**, 3153–3161 (2010).
- [62] Sergei Izvekov, Michele Parrinello, Christian J Burnham, and Gregory A Voth, "Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching," The Journal of chemical physics **120**, 10896–10913 (2004).
- [63] Andreas Singraber, Tobias Morawietz, Jörg Behler, and Christoph Dellago, "Parallel multistream training of high-dimensional neural network potentials," Journal of chemical theory and computation **15**, 3075–3092 (2019).
- [64] John P Perdew, Kieron Burke, and Matthias Ernzerhof, "Generalized gradient approximation made simple," Physical review letters **77**, 3865 (1996).
- [65] W. Zhong, David Vanderbilt, and K. M. Rabe, "Phase Transitions in BaTiO₃ from First Principles," Phys. Rev. Lett. **73**, 1861–1864 (1994).