scientific reports

OPEN

Representing Born effective charges with equivariant graph convolutional neural networks

Alex Kutana[1], Koji Shimizu[2], Satoshi Watanabe[3] & Ryoji Asahi[1✉]

Graph convolutional neural networks have been instrumental in machine learning of material properties. When representing tensorial properties, weights and descriptors of a physics-informed network must obey certain transformation rules to ensure the independence of the property on the choice of the reference frame. Here we explicitly encode such properties using an equivariant graph convolutional neural network. The network respects rotational symmetries of the crystal throughout by using equivariant weights and descriptors and provides a tensorial output of the target value. Applications to tensors of atomic Born effective charges in diverse materials including perovskite oxides, Li_3PO_4, and ZrO_2, are demonstrated, and good performance and generalization ability is obtained.

Since their introduction in materials science and chemistry, graph convolutional neural networks (GCNNs) have shown remarkable results[1]. GCNNs are being used to represent various molecular and materials properties[1–3], with typical tasks of predicting scalar values for electronic band gaps, ionization potentials/electron affinities, formation energies, as well as performing classification. In particular, in interatomic potentials and force fields applications, GCNNs presently achieve state of the art performance[4–6]. While most applications have been limited to scalar target values, many of the fundamental properties of materials are represented by tensors. When constructing a neural network with tensorial outputs, it is essential to incorporate relevant physics-informed[7] constraints and inductive biases, thus greatly improving the model performance in comparison with data augmentation approaches[8,9].

Tensors must obey certain transformation rules to ensure the independence of physical properties on the choice of the reference frame[10]. In order for the network features and outputs to follow these rules, it is necessary to encode their equivariance explicitly into the network architecture. Stemming from the original idea of steerable filters for vision and image processing[11], feature equivariance ensures their proper transformations in convolutional neural networks[12,8,13–15]. In physics and materials science, equivariance with respect to translations and rotations in three-dimensional Euclidean space is most essential. Utilization of Euclidean equivariant networks[9] has greatly increased the accuracy of interatomic potentials[5,6], and benefited other tasks, including prediction of optical, phonon, and neutron scattering spectra[16–18], density functional Hamiltonians[19,20], ab initio wavefunctions[21], reaction activation barriers[22], and spin dynamics[23]. At the same time, direct, end-to-end applications of machine learning to tensorial atomic quantities have been scarce. Recent developments include predictions of NMR chemical shift[24] and elasticity[25,26] tensors. Our earlier efforts utilized a symmetry-restricted network to predict part of the tensor of atomic Born effective charges[27]. The network was designed for a special case of the external field directed along one of the Cartesian axes, and accommodating the general case would have been substantially more complex. Recently, tensors of atomic Born charges of liquids, dielectrics, and ferroelectrics have also been represented by derivative learning using deep potentials[28], kernel-based regression[29], as well as differentiable deep NNs[30] and equivariant GCNNs[31]. Here, we report Equivar, a simple equivariant GCNN (EGCNN) for direct prediction of atomic tensorial quantities, which observes the symmetries of the physical three-dimensional Euclidean space, and operates with geometric quantities throughout. It utilizes local message passing based on the geometric input data. As only equivariant operations are performed on the input geometric data, all constraints imposed by the point group symmetry are automatically satisfied by the outputs. We demonstrate the application of Equivar to end-to-end machine learning of the target property of full tensors

[1]Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan. [2]National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan. [3]The University of Tokyo, Hongo, Tokyo, Japan. ✉email: ryoji.asahi@chem.material.nagoya-u.ac.jp
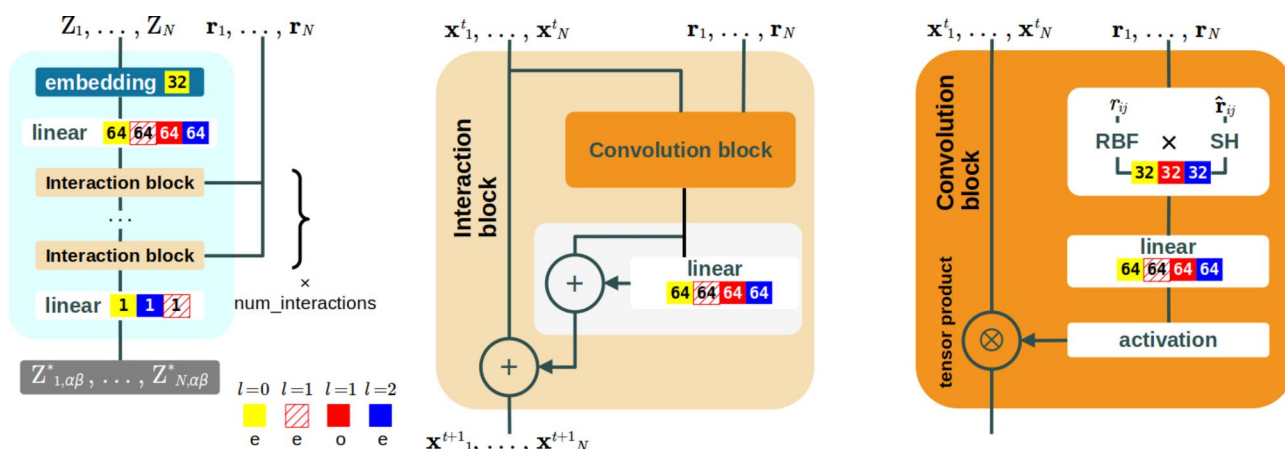
of atomic Born effective charges, achieving good performance. Due to equivariance constraints, high accuracy is achieved, while keeping the number of parameters small.

Network architecture

The atomic structure is represented by a graph[32], with atoms mapped to graph nodes and neighbor connections to edges, each equipped with attributes/features. Node features are updated via message passing from the neighbor nodes in the interaction block. The basic architecture of Equivar is shown in Fig. 1. It has a simple layout first employed in SchNet[33], consisting of the main unit with inputs/embeddings, interaction layers, and outputs. Each of the interaction layers contains the interaction and convolution blocks. The simple architecture helps maximize the transferability and generalizability of the model. The input attributes are scalars (atomic numbers Z_i, interatomic distances r_{ij}), and vectors (bond directions $\hat{r}_{ij}$), while the output features are the node values of tensors of atomic Born effective charges $Z^*_{i,\alpha\beta}$. Node embedding is a lookup dictionary with the mapping of the atomic numbers into an array of learnable scalar weights. The embedding dimension is a hyperparameter. Edge attributes represent the length and direction of the vector connecting neighbors. Edges are indexed by the bond vector $\mathbf{r}_{ab}$ connecting the central atom a with neighbor atom b. One-hot edge encodings $\delta(\mathbf{r}-\mathbf{r}_{qb})$ are projected onto a fixed basis of products of radial and angular functions, yielding the mapping $R^3 \to R \times R^{2l+1}$ for a radial basis function $R_n(r)$ and a set of angular basis functions of real spherical harmonics $Y_{l,m}(\theta,\varphi)$. A similar expansion is employed to obtain the rotational power spectrum and smooth overlap of atomic positions (SOAP) similarity kernel[34]. Uniformly shifted Gaussians $R_n(r) = \exp(-\gamma(r-\mu_n)^2)$[33] were used as radial basis functions[35], and a 3 Å radial cutoff was used for neighbor lists. The set of spherical harmonics $Y_{l,m}$ of degree l, $|m|\leq l$, produces equivariant geometric edge features enabling the "information flow"[13] within the channels of a geometric object upon rotation. The projections are passed through linear and activation blocks, producing a trainable tensor product convolution kernel. To provide a nonlinearity, a shifted softplus activation function[33], $\log(1 + e^x) - \log(2)$, is applied to the scalar edge features.

After the embedding, an equivariant linear layer introduces geometric node features, initialized to invariant (zero) values. The geometric (equivariant) features are irreducible representations of $O(3)$, a three-dimensional orthogonal group. Unlike the invariant (scalar) features, the equivariant features are "space-aware" and transform with rotations according to certain rules. Encoding these rules into the weights of the neural network increases its accuracy significantly[7], as the target quantities share same transformation properties as tensors. The features are stored sequentially in conventional arrays of floats, and typed according to how they transform under $O(3)$[9]. The equivariant linear layer acts by taking linear combinations of irreducible representations with the same rotational order l and parity p, and storing them in each of the matching outputs. The scalar coefficients of the linear combinations are learnable parameters. The equivariant node features are then updated via message passing in a sequence of interaction blocks, with the number of interactions being a hyperparameter. Within the interaction block, the message passed from a node to its neighbor is a tensor product of the node features with the features of the edge connecting the nodes, and the messages are aggregated by target nodes in a ResNet-style update[36]. The tensor product is calculated using Clebsch-Gordan coefficients[15]. For convenience, identical descriptor types and sizes for node and edge are used in the interaction block.

After the interactions, the final linear layer yields a decomposition of a general tensor of the second rank into irreducible representations of the $O(3)$ group, namely $0e + 2e + 1e$, which has 9 components that are converted to Cartesian form, accounting for full anisotropy. The equivariance of outputs ensures that the transformation rules for tensor components are always satisfied exactly upon coordinate system rotations. For equivariant operations with geometric features, the primitives implemented in e3nn[9] are used. Charge neutrality acoustic sum rule constraint is applied to model outputs using Eq. (49) from Ref[37].

Fig. 1. Architecture of the Equivar EGCNN. The inputs are atomic numbers and positions, and outputs are full tensors of atomic Born effective charges. $l=0$, $l=1$, and $l=2$ equivariant features are colored yellow, red, and blue, respectively; "e" and "o" stand for even and odd parity. Numbers indicate feature multiplicities. The values shown correspond to those used in "base model 1".

Datasets

The equivariant model was trained with full tensors of Born effective charges, calculated from first principles using density functional perturbation theory (DFPT)[37], as implemented in the VASP package[38]. PBEsol functional[39] was used for perovskites, and PBE functional[40] for Li_3PO_4 and ZrO_2. Ion-electron interactions were represented by all-electron projector augmented wave potentials[41], and a plane wave basis was used with cutoff energy of 500 eV (Li_3PO_4) and 520 eV (perovskites and ZrO_2). Tensors of atomic Born effective charges are fundamental quantities determining the long-range long-wavelength part of the force constants, LO-TO gamma point optical phonon splitting, and static dielectric response[37,42]. They are defined as the linear part of the change in polarization due to gamma point ion displacement, and given by the mixed second derivatives of the total energy E with respect to atomic positions and electric field[37,43–46]:

$$Z^*_{i,\alpha\beta} = -\frac{1}{|e|}\frac{\partial^2 E}{\partial \mathcal{E}_\alpha \partial u_{i,\beta}}\tag{1}$$

Here, $Z^*_{i,\alpha\beta}$ is the tensor of the Born effective charge of atom i, $\mathbf{u}_i=\mathbf{u}_i(q=0)$ is the gamma point displacement of atom i, e is elementary charge, and $\mathcal{E}$ is the electric field. A definition with swapped Cartesian indices α and β has also been used[47–49]. In the DFPT approach used here, the ground state and first derivatives of the wavefunction are calculated to obtain the Born effective charges. Alternatively, Born charges can be obtained using the finite electric field method. Both methods yield similar values, as seen from the comparison in Fig. S1 for $CaTiO_3$. After training with DFPT results, the equivariant end-to-end ML model quickly predicts the Born charges tensors from the structural input.

We use three datasets representing different systems of interest for training. The datasets were generated in-house by performing substitutions or creating vacancy defects in pristine bulk structures obtained from the Materials Project[50] database. This approach is complementary to other high throughput efforts to building databases of Born effective charges and other response properties based on the structures in the Materials Project database, e.g., JARVIS-DFT[51]. Structure ids of the structures used are given in the Supplementary Information. The first dataset contains substituted perovskite oxides, relevant to electric energy storage applications[52]. Substitutions are a primary method for boosting the dielectric permittivity of transition metal oxides[53–55], leading to greater energy storage capacity. Model validation with this dataset also demonstrates its capability for a wide range of chemical elements. The dataset was generated via cation substitutions in mineral perovskite, Pnma $CaTiO_3$. Ca^{2+}, Sr^{2+}, Ba^{2+}, and Pb^{2+} isovalent substitutions were performed on the alkaline earth metal site, and Ti^{4+}, Zr^{4+}, and Hf^{4+} on the transition metal site[54,55] to generate a dataset of 1,224 materials. Born charges of optimized structures were used for training. The second dataset contains various structures of the Li-ion battery material, Li_3PO_4, one of the most widely used solid electrolytes[56–58]. The dataset of MD snapshots including defects may test sensitivity of local configurations, and provides a more realistic and diverse distribution, compared to, e.g. employing random displacements sampled from the normal distributions centered on equilibrium positions[30]. This dataset was used to elucidate the Li ion conduction behaviors[27,58]. Pristine Li_3PO_4 ($Li_{12}P_4O_{16}$ in the adopted supercell), as well as systems with Li and Li_2O vacancy defects ($Li_{11}P_4O_{16}$ and $Li_{22}P_8O_{31}$, respectively)[27] were used. The pristine subset consists of snapshots from NVT-ensemble ab initio molecular dynamics (AIMD) simulations at 300 and 2,000 K. A time step of 1 fs was used. The Li vacancy structure set contains images from nudged elastic band (NEB) calculations, whereas Li_2O vacancy structures are snapshots of AIMD at 2,000 K. All frames from the ab initio MD calculations were used. The total number of systems is 17,991; additionally, a dataset with 1,870 larger structures ($Li_{46}P_{16}O_{63}$) with a Li_2O vacancy was used for testing. The third dataset has structures of zirconia (ZrO_2), a high permittivity material[59], with applications in microelectronics, energy storage, and as structural ceramics[60]. The dataset with three different crystal structures — cubic, tetragonal, and monoclinic — contains the subtle effects of these long-range orders. It is planned to use this dataset to analyze the mechanism of plastic deformation enhancement under electric field application[61]. Materials Project structure ids of the structures used are given in the Supplementary Information. The dataset consists of 10,103 NVT-ensemble empirical potential simulation[62] snapshots of cubic, tetragonal, and monoclinic zirconia ($Zr_{16}O_{32}$ in the supercell for the pristine models) at 1,300, 1,500, 1,700, and 1,900 K, with isotropic lattice constant changes (-2%, -1%, 0%, +1%, +2%), and with some of the systems hosting an oxygen vacancy. For the oxygen vacancy structures, +2 charge state was considered. The snapshots were taken every 1000 fs.

Results

We benchmarked the model while tuning the hyperparameters using a small dataset of 224 substituted perovskites. The model was trained for 250 epochs using L^1-norm loss function and AdamW optimizer[63] with an initial learning rate of 0.005, and best validation mean absolute error (MAE) on a held-out subset of 45 structures was recorded. The results are given in Table S1. The model with ~131k trainable parameters (6 interaction layers and feature size of 32 in the interaction block) has the MAE of 0.0117 for the small set validation, compared to the cross-validation MAE of 0.0088 for the full set of 1,224 perovskites. The model demonstrates a good generalization with a small dataset of substituted perovskites, and stable performance with respect to hyperparameter variation. The performance can be further improved by increasing the model size and the number of interactions. Here, we use a 510k-parameter model, 'base model 1' (BM1), with 6 interaction layers and feature size of 64 in the interaction block, which provides a good tradeoff among the accuracy, speed, and size, and compares with the smaller/faster 131k-parameter model (BM2).

The two models were trained for 500 epochs with a combined dataset of three systems types (perovskites, Li_3PO_4, and ZrO_2, 29,318 structures), as well as several subsets, using a 0.8/0.2 train/validation split. The training history with validation loss is shown in Fig. 2. BM1 shows consistently better performance, due to larger size of

the descriptor in the interaction block. Further increase in the descriptor size leads to even better performance, but is not pursued due to memory constraints. Another important hyperparameter is the number of interactions. Iterative interactions allow the nodes to exchange messages beyond the nearest neighbor cutoff, which allows capturing long-range interactions. In case of 3 Å cutoff radius and 6 interactions, the effective cutoff radius for the total receptive field of each node is 18 Å, sufficient to represent most long-range interactions. Here, 6 interactions were used, although the accuracy kept improving even at 12 interactions, the largest value tested. While a small charge drift MAE of 0.154 and 0.173 were recorded for the BM1 and BM2 models, respectively, subsequent correction is applied to the model outputs to ensure that the sum rule is satisfied exactly. The definition and further details are provided in the Supplementary Information.

The scalings of the training and inference times with the dataset size are shown in Figs. S2 and S3, respectively. Calculations were performed on a single NVIDIA RTX A6000 GPU. Training and inference times of Equivar are found to scale approximately linearly with both the number of atoms and number of model parameters. The training times are 0.493 ms/atom/epoch for BM1, and 0.156 ms/atom/epoch for BM2, whereas the inference times are 0.340 ms/atom for BM1, and 0.117 ms/atom for BM2. These timings can be compared with the result of 1.05 ms/atom/epoch for SchNet on the Nvidia GTX 1080 GPU[64].

The model performance summary is given in Table 1. The results for the models BM1 and BM2 trained with all three datasets are compared with those for "specialized" models, which were trained with only one
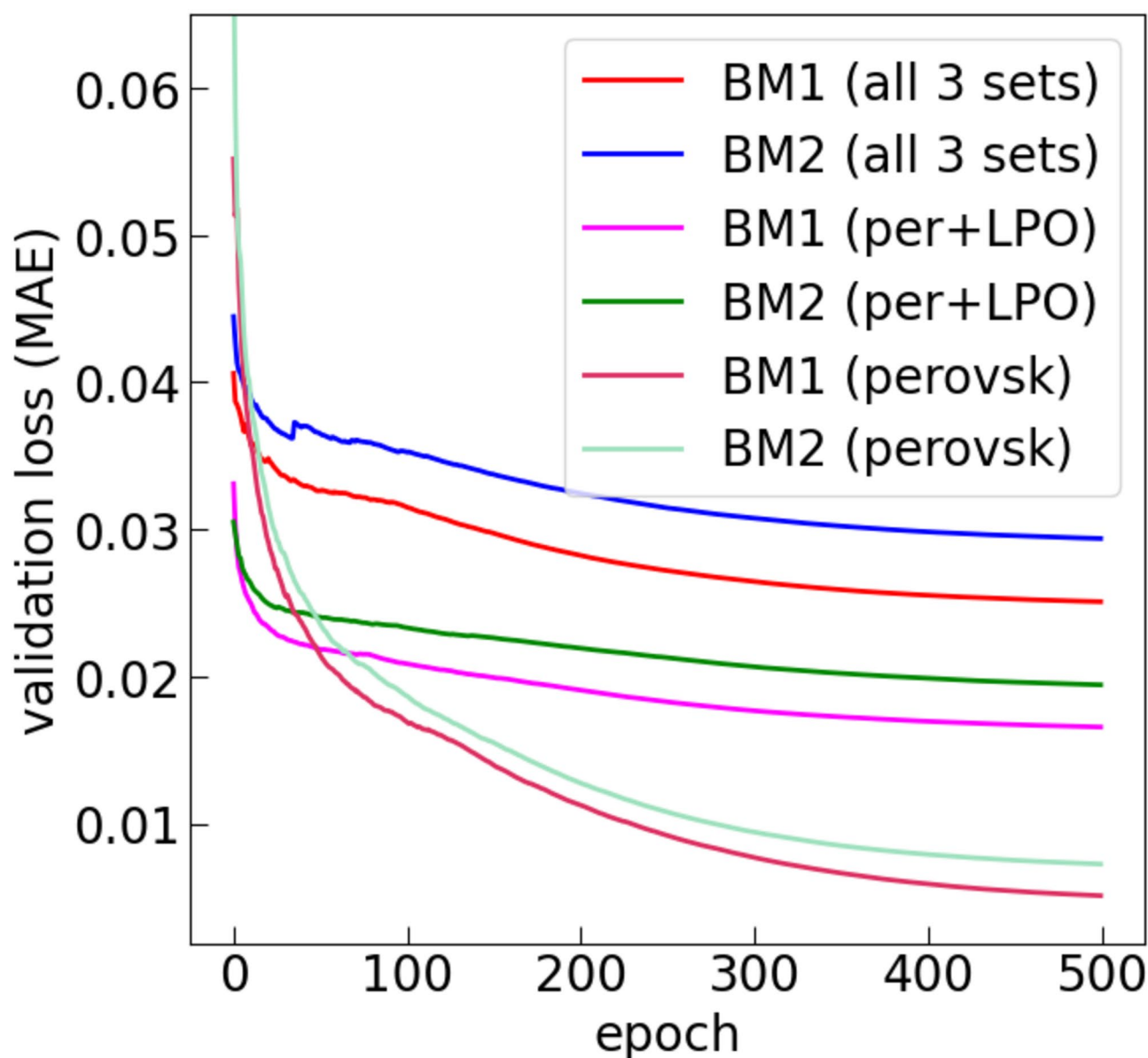
Fig. 2. Training history for EGCNN base models BM1 and BM2 for regression of tensors of atomic Born effective charges. Models were trained for 500 epochs using combinations of dataset of three systems types - perovskites, Li_3PO_4, and ZrO_2. The larger BM1 model shows a consistently better performance.

dataset. It is seen that a specialized model can be trained to achieve a better performance with a particular dataset. The results of five-fold cross-validation for the tensor of Born effective charges of 1,224 perovskites are shown in Fig. 3. For cross-validation, the dataset was randomly split into five approximately equal-sized subsets. Each of the subsets was sequentially used for validation, while the remaining subsets for training the model, and validation results were recorded in each run. Cross-validation MAEs for diagonal and off-diagonal tensor components and individual atomic species are shown in Fig. S4.

The obtained diagonal and off-diagonal values for these elements are typical for the ABO_3 compounds[48,45]. Born effective charges of Ca, Sr, and Ba are rather similar, with average values for diagonal components of 2.5–2.7 and standard deviations of 0.15–0.17, whereas Pb has a larger average Born charge of 3.5 with a standard deviation of 0.26. Ti, Zr, and Hf have the average values of 6.5, 5.7, and 5.6, and standard deviations of 0.56, 0.41, and 0.42 respectively. The diagonal components for O range between − 1.3 and − 6.5, with the average of -2.9 and standard deviation of 0.91. Larger standard deviations in B-type cations and O as compared to A-type cations reflect the bimodal distribution of their Born charges, corresponding to two distinct oxidations states. Of note are the large variations of the off-diagonal components in O, ranging from − 2.2 to 2.1, while in Ca the range is much smaller, from − 0.26 to 0.34. The larger variability of the O values does not result in the larger error for the model prediction.

Validation results for training with a Li_3PO_4 dataset are shown in Fig. 4. In contrast with the perovskites, the O Born charges in Li_3PO_4 show smaller values and variability, with the average being − 1.53 and standard deviation 0.33 for the diagonal components. The average charge of Li is 1.06, corresponding to its nominal valence state, whereas that of P is 2.97. The prediction error is worse than for the perovskites dataset that only contains optimized structures without any defects. Thus, the larger prediction MAE in Li_3PO_4 is most likely due to the presence of anomalous values for a small number of structures in the dataset. The noise in the data could appear when atoms are displaced far from equilibrium in the high temperature MD.

Cross-validation results for the model trained with Born effective charges of ZrO_2 are shown in Fig. 5. In ZrO_2, the average O charge is -2.78, and std is 0.45, Zr average charge is 5.50 and std is 0.40. The error is larger in ZrO_2, which we surmised was due to the presence of a small number of anomalous data. We also trained an Equivar model using a cutoff radius of 5 Å and 3 interactions, to check whether a larger cutoff would better account for the nearest neighbor coordination and improve the performance. As seen from the validation plot in Fig. S5, the model performance in ZrO_2 is similar to that of the original model, showing that the cutoff radius of 3 Å is adequate.

Discussion

The Equivar model demonstrates good performance and generalization ability while retaining simplicity and small size. The model is comparable in size to SchNet[33], which has ~ 121k parameters (taking radial basis size of 32, same as used here, rather than 300 in the original SchNet). The larger size of Equivar is mainly due to its use of equivariant vector/tensor weights and features, which have larger dimensions than SchNet's scalars. Overall, while showing good performance, Equivar implements a rather minimalistic EGCNN for tensor representations, and further improvements are possible by employing more complex architectures. Recently, an equivariant model for regression of atomic Born effective charges based on λ-SOAP descriptors[65] has been reported[29]. The model archives a root mean square percentage error (RMSPE) of 3% for Born effective charges isolated water dimers. Equivar shows a better or similar accuracy for bulk systems, with RMSPE of 0.5%, 3%, and 4% for the diagonal elements of effective charge tensors of perovskite, Li_3PO_4, and ZrO_2 datasets, respectively. Note that here we do not consider another popular mechanism of attention/transformers[66], although it may further improve the performance. Based on the size scaling, rather large systems become readily accessible, e.g., evaluation times of ~ 1 s could be attained for a structure with ~ 3000 atoms. This presents a prospect for studying a variety of local doping/substitution configurations quite efficiently, enabling the design of new materials.

We now reiterate some of the important aspects contributing to the high accuracy of the model, making it an efficient approach for "second-principles"[67] calculations of tensors of effective charge in oxides. The underlying approach is that of a convolution on a graph, whereby the target values can be gradually learned from the node and edge attributes by performing a sequence of interactions/convolutions with neighbor nodes. Message passing at each interaction allows the information to accumulate and propagate beyond nearest neighbors. Continuous filters are assigned to graph edges, enabling smooth modulation of trainable filter weights according to edge type. Unlike hand-designed representations of the local atomic environments used in the SOAP/λ-SOAP approach[34,65], hidden node features of a GCNN are determined automatically through backpropagation. In SOAP/λ-SOAP, the atomic neighbor density function plays a central role and is a starting point for constructing SOAP descriptors. It is computed by taking the sum of atomic positions over all neighbors around the central

|	perovskites	Li_3PO_4	ZrO_2
base model BM1 (510k) MAE	0.0207	0.0196	0.0323
base model BM1 (510k) RMSE	0.0300	0.0319	0.0510
base model BM2 (140k) MAE	0.0260	0.0227	0.0370
base model BM2 (140k) RMSE	0.0377	0.0364	0.0569
specialized model (131k) MAE	0.0088	0.0199	0.0325
specialized model (131k) RMSE	0.0135	0.0325	0.0493

Table 1. Model performance summary.
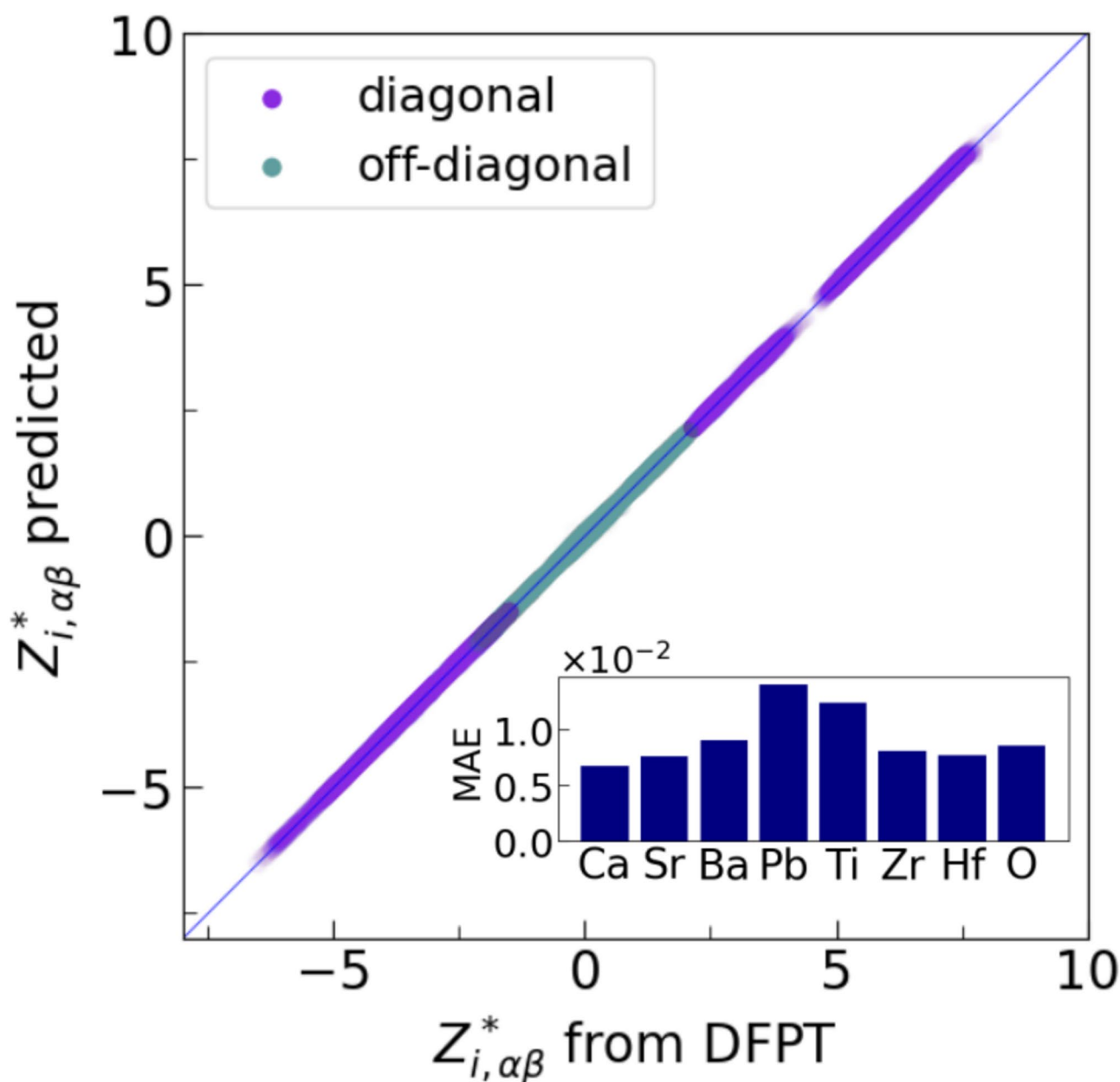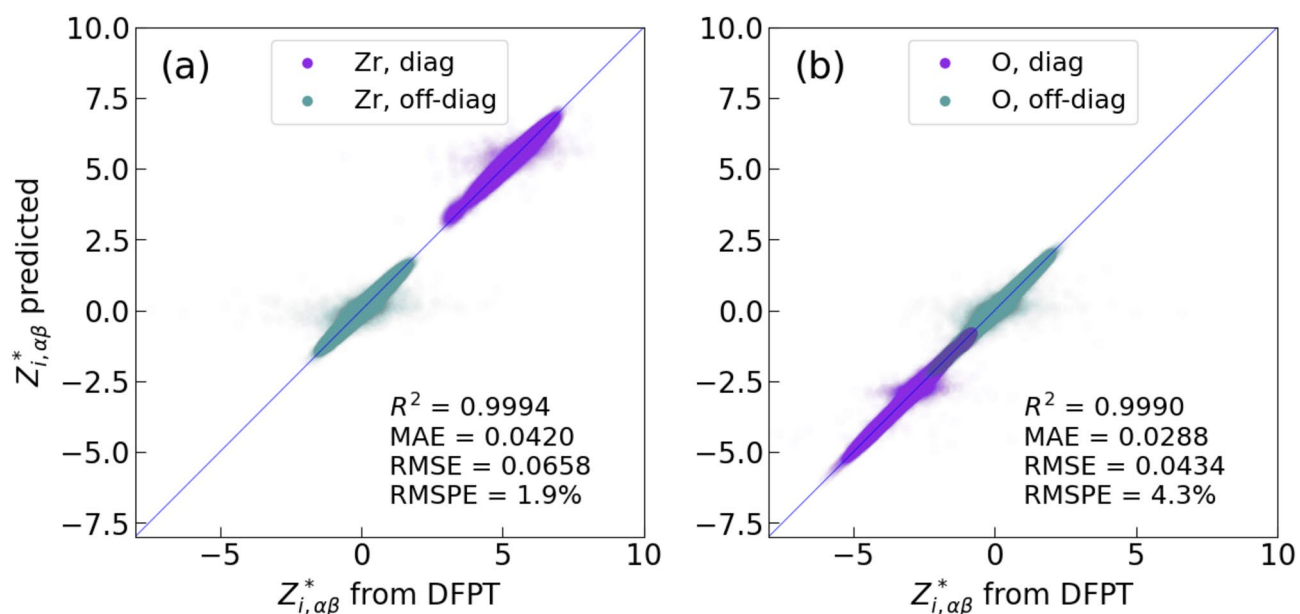
Fig. 3. Five-fold cross-validation results for the prediction of the tensor of Born effective charges of Pnma ABO_3 perovskites (A = Ca, Sr, Ba, Pb; B = Ti, Zr, Hf) by the equivariant GCNN model. Diagonal and off-diagonal tensor components are shown. The inset shows the MAE for all components for individual atomic species.

atom. Power spectrum/bispectrum descriptors and similarity kernel in SOAP/λ-SOAP are based on the atomic neighbor density function having a fixed predetermined form, in contrast with hidden equivariant representations of an EGCNN. Graph edge descriptors are created from bond vectors individually for each pair of neighbors, and sum is taken at the last step after being passed through a neural network. These important distinctions allow greater flexibility for the automatically determined descriptors, endowing GCNN models with greater regressive power. Equivariant edge features and descriptors allow continuous one-hot encoding and modulation of the convolution weights not only by the neighbor-neighbor distance, but also the direction of the interatomic vector, while equivariant node features naturally represent tensorial target values. The empirically established capability for regressing tensors of Born effective charges suggests sufficient expressiveness of the EGCNN architecture.

In the future, the analysis of model limitations and components that are essential for performing efficient target tensor decomposition[68] should be performed. Relevant information is expected to propagate from further neighbors over multiple interaction/message passing steps, but the efficiency of this path has not been tested yet, and more efficient pathways could be possibly designed. While the model shows good accuracy for our datasets, its extrapolation ability to new structures and compositions is still largely untested. Thus, among the
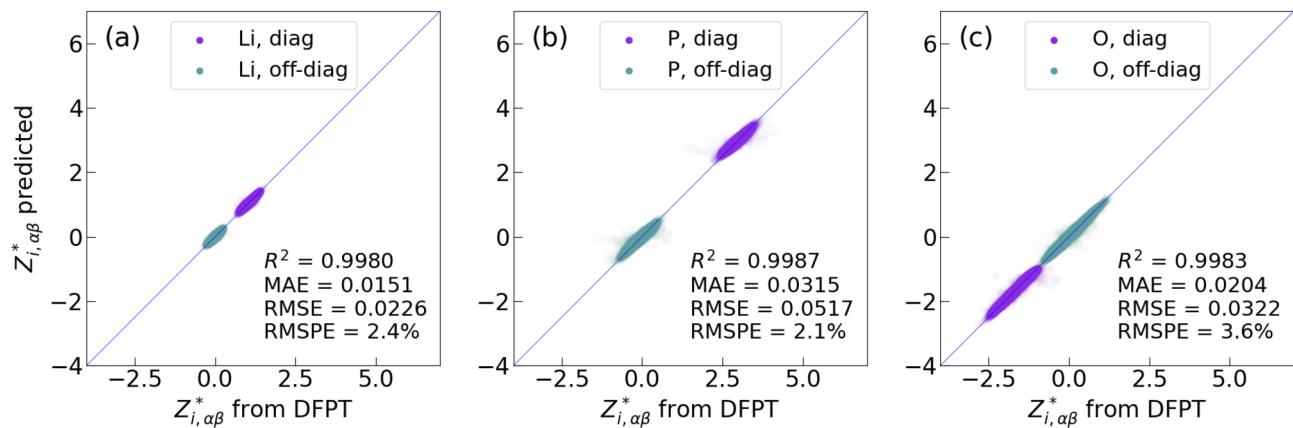
Fig. 4. Validation results for tensors of Born effective charges of Li_3PO_4. Panels show results for different atomic species: (**a**) Li, (**b**) P, and (**c**) O. The structures used for model training are the snapshots of ab initio molecular dynamics of pristine Li_3PO_4 as well as structures with single Li or Li_2O vacancies at 300 K and 2,000 K. RMSPE is given for diagonal elements.

Fig. 5. Five-fold cross-validation results for the tensor of Born effective charges of ZrO_2. The dataset consists of AIMD snapshots of cubic, tetragonal, and monoclinic zirconia, with some of the systems hosting an oxygen vacancy. Panels show results for different atomic species: (**a**) Zr, (**b**) O. RMSPE is given for diagonal elements.

current challenges that need to be addressed in the future are interpretability and strong generalization ability to diverse materials. Establishing causality[69,70] in the training data could improve interpretability and help models make better predictions. Other possible strategies for improving model performance include using more diverse training data and employing regularization.

Fundamentally, Born effective charges define the coupling of atomic vibrations to electric fields, and are responsible for long-range Coulomb interactions, which determine the TO-LO phonon mode splitting in polar crystals[48,71]. An early empirical model[72] yielded reasonable values for Born charges of various ABO_3 compounds from the fitting to experimental phonon mode oscillator strengths. With the advent of ab initio approaches, theoretical values have been calculated for many systems and found in good agreement with experiments based on comparison with infrared and inelastic neutron scattering spectra[73–76]. Recent developments include the extension of Born charges to metals[77,78], where they can be used to probe electron-phonon interactions.

Besides their utility for calculating physical observables, Born effective charges can also provide insights into the nature of microscopic interactions. Qualitatively, one can consider the two limits, one of purely electrostatic (or ionic), and the other of purely electronic interactions. In the former limit, a rigid ion model gives a good approximation of the interaction energy and effective charges. This ionic model of interactions is most applicable

in Li_3PO_4, where the variation of the effective charges is smallest, indicating mostly ionic nature of bonding. This is especially true for Li, where the effective charge corresponds to the nominal value and shows small variability, whereas variability is somewhat larger in P and O.

On the other hand, perovskites and ZrO_2 display a mixture of covalent and ionic bonding. "Semicovalent" bonding is one of the characteristic features of perovskites, giving rise to several interesting phenomena, including the indirect (double) exchange coupling of cations[79]. The mixture of ionic and covalent bonding is responsible for anomalously high Born effective charges, as well as other disparities with the simple rigid ion model, such as the inequivalence of the charges of O anions[44,80]. This enhancement of Born effective charges has been traced to dynamical changes of O $2p$ and metal d orbital hybridization[81] during ion displacement. The resulting flow of electrons augments the polarization change. Anomalous Born charges are also responsible for the large LO-TO gamma point optical phonon splitting[73] and appearance of the soft ferroelectric mode[82], and are central for achieving colossal permittivity materials[54,55]. We note that these effects are present in the studied systems and should be well captured by our EGCNN model.

Conclusion

In conclusion, we apply equivariant graph convolutional neural networks to learning the tensors of atomic Born effective charges. Good performance is achieved with three diverse systems: ABO_3 perovskites, Li_3PO_4, and ZrO_2. The model possibly demonstrates a good generalization ability by effectively learning from a small dataset, although a broader dataset[51] would provide a more rigorous test. Performance is further improved when using a larger dataset. The important hyperparameters affecting the model performance are the number of interactions/ convolutions and the size of the descriptor in the interaction block. The training and inference times scale linearly with the number of atoms, and model architecture allows treating a wide range of combinations of chemical elements, enabling future creation of universal foundation models trained on a vast chemical space. Our model represents an important step towards fast and accurate modeling of microscopic quantum phenomena in response to the electric field.

Data availability

References

1. Reiser, P. et al. Graph neural networks for materials science and chemistry. *Commun. Mater.* **3**, 93 (2022).
2. Wu, Z. et al. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 4–24 (2021).
3. Corso, G., Stark, H., Jegelka, S., Jaakkola, T. & Barzilay, R. Graph neural networks. *Nat. Rev. Methods Primer.* **4**, 17 (2024).
4. Gasteiger, J., Becker, F., Günnemann, S. & GemNet Universal directional graph neural networks for molecules. in Advances in Neural Information Processing Systems (eds Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S. & Vaughan, J. W.) vol. 34 6790–6802 (Curran Associates, Inc., (2021).
5. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
6. Batatia, I. et al. A foundation model for atomistic materials chemistry. Preprint at (2023). https://arxiv.org/abs/2401.00096
7. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
8. Worrall, D. E., Garbin, S. J., Turmukhambetov, D. & Brostow, G. J. Harmonic Networks: Deep Translation and Rotation Equivariance. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017).
9. Geiger, M. & Smidt, T. e3nn: Euclidean Neural Networks. Preprint at (2022). https://arxiv.org/abs/2207.09453
10. Arfken, G. B. *Mathematical Methods for Physicists* (Academic, 1985).
11. Freeman, W. T. & Adelson, E. H. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 891–906 (1991).
12. Cohen, T. & Welling, M. Group Equivariant Convolutional Networks. in Proceedings of The 33rd International Conference on Machine Learning (eds. Balcan, M. F. & Weinberger, K. Q.) vol. 48 2990–2999PMLR, New York, New York, USA, (2016).
13. Weiler, M., Geiger, M., Welling, M., Boomsma, W. & Cohen, T. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data. Preprint at (2018). https://arxiv.org/abs/1807.02547
14. Kondor, R. & Trivedi, S. On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups. in Proceedings of the 35th International Conference on Machine Learning (eds. Dy, J. & Krause, A.) vol. 80 2747–2755 (PMLR, 2018).
15. Thomas, N. et al. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. Preprint at (2018). https://arxiv.org/abs/1802.08219
16. Schütt, K., Unke, O. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. in Proceedings of the 38th International Conference on Machine Learning (eds. Meila, M. & Zhang, T.) vol. 139 9377–9388 (PMLR, 2021).
17. Okabe, R. et al. Virtual Node Graph Neural Network for Full Phonon Prediction. Preprint at (2023). https://arxiv.org/abs/2301.02197
18. Cheng, Y. et al. Direct prediction of inelastic neutron scattering spectra from the crystal structure. *Mach. Learn. Sci. Technol.* **4**, 015010 (2023).
19. Gong, X. et al. General framework for E(3)-equivariant neural network representation of density functional theory hamiltonian. *Nat. Commun.* **14**, 2848 (2023).
20. Remme, R. et al. Deep learning a transferable kinetic energy functional for orbital-free density functional theory. *J. Chem. Phys.* **159**, 144113 (2023).
21. Hermann, J. et al. Ab initio quantum chemistry with neural-network wavefunctions. *Nat. Rev. Chem.* **7**, 692–709 (2023).
22. Vijay, S. et al. CoeffNet: predicting activation barriers through a chemically-interpretable, equivariant and physically constrained graph neural network. *Chem. Sci.* **15**, 2923–2936 (2024).

23. Miyazaki, Y. Equivariant neural networks for spin dynamics simulations of itinerant magnets. *Mach. Learn. Sci. Technol.* **4**, 045006 (2023).
24. Venetos, M. C., Wen, M. & Persson, K. A. Machine learning full NMR chemical shift tensors of silicon oxides with equivariant graph neural networks. *J. Phys. Chem. A.* **127**, 2388–2398 (2023).
25. Pakornchote, T., Ektarawong, A., Chotibut, T. & StrainTensorNet Predicting crystal structure elastic properties using SE(3)-equivariant graph neural networks. *Phys. Rev. Res.* **5**, 043198 (2023).
26. Wen, M., Horton, M. K., Munro, J. M., Huck, P. & Persson, K. A. An equivariant graph neural network for the elasticity tensors of all seven crystal systems. *Digit. Discov.* **3**, 869–882 (2024).
27. Koji Shimizu, E. M., Otsuka, R., Watanabe, S. & Masahiro Hara & Prediction of born effective charges using neural network to study ion migration under electric fields: applications to crystalline and amorphous Li3PO4. *Sci. Technol. Adv. Mater. Methods.* **3**, 2253135 (2023).
28. Zhang, L. et al. A deep potential model with long-range electrostatic interactions. *J. Chem. Phys.* **156**, 124107 (2022).
29. Schmiedmayer, B. & Kresse, G. Derivative learning of tensorial quantities—Predicting finite temperature infrared spectra from first principles. *J. Chem. Phys.* **161**, 084703 (2024).
30. Malenfant-Thuot, O., Ryczko, K., Tamblyn, I. & Côté, M. Efficient determination of Born-effective charges, LO-TO splitting, and Raman tensors of solids with a real-space atom-centered deep learning approach. *J. Phys. Condens. Matter.* **36**, 425901 (2024).
31. Falletta, S. et al. Unified Differentiable Learning of Electric Response. Preprint at (2024). https://arxiv.org/abs/2403.17207
32. Scarselli, F., Gori, M., Tsoi, A. C. & Hagenbuchner, M. Monfardini. The graph neural network model. *IEEE Trans. Neural Netw.* **20**, 61–80 (2009).
33. Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A. & Müller, K. R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
34. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B.* **87**, 184115 (2013).
35. Buhmann, M. D. Radial basis functions. *Acta Numer.* **9**, 1–38 (2000).
36. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).
37. Gonze, X. & Lee, C. Dynamical matrices, born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B.* **55**, 10355–10368 (1997).
38. Gajdos, M., Hummer, K., Kresse, G., Furthmüller, J. & Bechstedt, F. Linear optical properties in the projector-augmented wave methodology. *Phys. Rev. B.* **73**, 045112 (2006).
39. Perdew, J. P. et al. Restoring the Density-Gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **100**, 136406 (2008).
40. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
41. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B.* **50**, 17953–17979 (1994).
42. Pick, R. M., Cohen, M. H. & Martin, R. M. Microscopic theory of force constants in the adiabatic approximation. *Phys. Rev. B.* **1**, 910–920 (1970).
43. Giannozzi, P., de Gironcoli, S., Pavone, P. & Baroni, S. Ab initio calculation of phonon dispersions in semiconductors. *Phys. Rev. B.* **43**, 7231–7242 (1991).
44. Resta, R., Posternak, M. & Baldereschi, A. Towards a quantum theory of polarization in ferroelectrics: the case of KNbO₃. *Phys. Rev. Lett.* **70**, 1010–1013 (1993).
45. Cockayne, E. & Burton, B. P. Phonons and static dielectric constant in CaTiO3 from first principles. *Phys. Rev. B.* **62**, 3735–3743 (2000).
46. Baroni, S., de Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73**, 515–562 (2001).
47. Ghosez, P. First-principles study of the dielectric and dynamical properties of barium titanate. Dr. Thesis Univ. Cathol. Louvain (1997).
48. Ghosez, P., Michenaud, J. P. & Gonze, X. Dynamical atomic charges: the case of ABO₃ compounds. *Phys. Rev. B.* **58**, 6224–6240 (1998).
49. Wu, X., Vanderbilt, D. & Hamann, D. R. Systematic treatment of displacements, strains, and electric fields in density-functional perturbation theory. *Phys. Rev. B.* **72**, 035105 (2005).
50. Jain, A. et al. Commentary: the materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
51. Choudhary, K. et al. High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses. *Npj Comput. Mater.* **6**, 64 (2020).
52. Yang, Z., Du, H., Jin, L. & Poelman, D. High-performance lead-free bulk ceramics for electrical energy storage applications: design strategies and challenges. *J. Mater. Chem. A.* **9**, 18026–18085 (2021).
53. Taniguchi, H., Sato, D., Nakano, A. & Terasaki, I. Permittivity boosting in yellow (Nb + In) co-doped TiO2. *J. Mater. Chem. C.* **8**, 13627–13631 (2020).
54. Kutana, A., Shimano, Y. & Asahi, R. Permittivity boosting by induced strain from local doping in titanates from first principles. *Sci. Rep.* **13**, 3761 (2023).
55. Shimano, Y., Kutana, A. & Asahi, R. Machine learning and atomistic origin of high dielectric permittivity in oxides. *Sci. Rep.* **13**, 22236 (2023).
56. Kuwata, N., Iwagami, N., Tanji, Y., Matsuda, Y. & Kawamura, J. Characterization of Thin-Film Lithium batteries with stable Thin-Film Li3PO4 solid electrolytes fabricated by ArF excimer laser deposition. *J. Electrochem. Soc.* **157**, A521 (2010).
57. Deng, Y. et al. Structural and mechanistic insights into fast Lithium-Ion conduction in Li4SiO4–Li3PO4 solid electrolytes. *J. Am. Chem. Soc.* **137**, 9136–9145 (2015).
58. Li, W., Ando, Y., Minamitani, E. & Watanabe, S. Study of Li atom diffusion in amorphous Li3PO4 with neural network potential. *J. Chem. Phys.* **147**, 214106 (2017).
59. Zhao, X. & Vanderbilt, D. Phonons and lattice dielectric properties of zirconia. *Phys. Rev. B.* **65**, 075105 (2002).
60. Pešić, M., Hoffmann, M., Richter, C., Mikolajick, T. & Schroeder, U. Nonvolatile random access memory and energy storage based on antiferroelectric like hysteresis in ZrO2. *Adv. Funct. Mater.* **26**, 7486–7494 (2016).
61. Motomura, H., Tamao, D., Nambu, K., Masuda, H. & Yoshida, H. Athermal effect of flash event on high-temperature plastic deformation in Y2O3-stabilized tetragonal ZrO2 polycrystal. *J. Eur. Ceram. Soc.* **42**, 5045–5052 (2022).
62. Wang, Y., Zahid, F., Wang, J. & Guo, H. Structure and dielectric properties of amorphous high-割 oxides: HfO_2, ZrO_2, and their alloys. *Phys. Rev. B.* **85**, 224110 (2012).
63. Loshchilov, I. & Hutter, F. Decoupled Weight Decay Regularization. Preprint at (2017). https://arxiv.org/abs/1711.05101
64. Park, C. W. et al. Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *Npj Comput. Mater.* **7**, 73 (2021).
65. Grisafi, A., Wilkins, D. M., Csányi, G. & Ceriotti, M. Symmetry-Adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.* **120**, 036002 (2018).
66. Vaswani, A. et al. Attention is all you need. in Advances in Neural Information Processing Systems (ed Guyon, I.) vol. 30 (Curran Associates, Inc., (2017).

67. Ghosez, P. & Junquera, J. Modeling of ferroelectric oxide perovskites: from first to second principles. *Annual Rev. Condens. Matter Phys.* **13**, 325–364 (2022).
68. Kolda, T. G. & Bader, B. W. Tensor decompositions and applications. *SIAM Rev.* **51**, 455–500 (2009).
69. Ghosh, A., Palanichamy, G., Trujillo, D. P., Shaikh, M. & Ghosh, S. Insights into cation ordering of double perovskite oxides from machine learning and causal relations. *Chem. Mater.* **34**, 7563–7578 (2022).
70. Ghosh, A. Towards physics-informed explainable machine learning and causal models for materials research. *Comput. Mater. Sci.* **233**, 112740 (2024).
71. Born, M. & Göppert-Mayer, M. Dynamische Gittertheorie der Kristalle. in Aufbau Der Zusammenhängenden Materie 623–794Springer Berlin Heidelberg, Berlin, Heidelberg, (1933). https://doi.org/10.1007/978-3-642-91116-3_4
72. Axe, J. D. Apparent ionic charges and vibrational eigenmodes of $BaTiO_3$ and other perovskites. *Phys. Rev.* **157**, 429–435 (1967).
73. Zhong, W., King-Smith, R. D. & Vanderbilt, D. Giant LO-TO splittings in perovskite ferroelectrics. *Phys. Rev. Lett.* **72**, 3618–3621 (1994).
74. Öğüt, S. & Rabe, K. M. Anomalous effective charges and far-IR optical absorption of Al_2Ru from first principles. *Phys. Rev. B.* **54**, R8297–R8300 (1996).
75. Balan, E., Saitta, A. M., Mauri, F. & Calas, G. First-principles modeling of the infrared spectrum of kaolinite. *Am. Mineral.* **86**, 1321–1330 (2001).
76. Schmalzl, K., Strauch, D. & Schober, H. Lattice-dynamical and ground-state properties of CaF_2 studied by inelastic neutron scattering and density-functional methods. *Phys. Rev. B.* **68**, 144301 (2003).
77. Wang, C. Y., Sharma, S., Gross, E. K. U. & Dewhurst, J. K. Dynamical born effective charges. *Phys. Rev. B.* **106**, L180303 (2022).
78. Marchese, G. et al. Born effective charges and vibrational spectra in superconducting and bad conducting metals. *Nat. Phys.* **20**, 88–94 (2024).
79. Goodenough, J. B. Theory of the role of covalence in the Perovskite-Type manganites [La, M(II)]MnO_3. *Phys. Rev.* **100**, 564–573 (1955).
80. Posternak, M., Resta, R. & Baldereschi, A. Role of covalent bonding in the polarization of perovskite oxides: the case of $KNbO_3$. *Phys. Rev. B.* **50**, 8911–8914 (1994).
81. Ghosez, P., Michenaud J. P. A microscopic study of barium titanate. *Ferroelectrics* **164**, 113–121 (1995).
82. Resta, R. & Vanderbilt, D. Theory of polarization: A modern approach. in Physics of Ferroelectrics: A Modern Perspective 31–68 (Springer Berlin Heidelberg, Berlin, Heidelberg, doi:https://doi.org/10.1007/978-3-540-34591-6_2. (2007).

Acknowledgements

Author contributions

K.A.: computations, generating data, discussion, writing. S.W. and K.S.: generating training data, discussion. R.A: supervision, discussion, writing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01250-5.

Correspondence and requests for materials should be addressed to R.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.