

Physics-Based Factorized Machine Learning for Predicting Ionic Dielectric Tensors

Atsushi Takigawa^{1,2,*} Shin Kiyohara^{1,*†} and Yu Kumagai^{1,3,‡}

¹*Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan*

²*Graduate School of Engineering, Tohoku University,
6-6 Aramaki Aoba, Aoba-ku, Sendai 980-8579, Japan*

³*Organization for Advanced Studies, Tohoku University,
2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan*



(Received 9 October 2025; accepted 25 February 2026; published 7 April 2026)

Considerable effort continues to be devoted to the exploration of next-generation high- κ materials that combine a high dielectric constant with a wide band gap. However, machine learning (ML)-based virtual screening has remained challenging, primarily due to the low accuracy in predicting the ionic contribution to the dielectric tensor, which dominates the dielectric performance of high- κ materials. We propose a joint ML model that predicts Born effective charges using an equivariant graph neural network, and phonon properties using a highly accurate pretrained ML potential. The ionic dielectric tensor is then computed analytically from these quantities. This approach significantly improves the accuracy of ionic contribution. Using the proposed model, we successfully identified 31 novel high- κ oxides from a screening pool of over 8000 candidates.

DOI: [10.1103/28wr-w896](https://doi.org/10.1103/28wr-w896)

Subject Areas: Computational Physics, Electronics,
Materials Science

I. INTRODUCTION

High- κ materials play an essential role in a wide range of advanced electronic and energy storage devices, e.g., transistors and capacitors [1–3]. In these applications, a high dielectric constant is essential for maximizing capacitance, while a wide band gap (E_g) remains necessary to ensure insulation and minimize leakage [3–6]. For example, in transistor gate insulators, SiO_2 has traditionally been the dominant material. Although its wide band gap ($E_g \approx 8.1$ eV) [7] is beneficial, the low dielectric constant ($\epsilon \approx 3.8$) [7] requires thinning of the capacitor to realize high capacitance, which in turn leads to an increase in leakage current due to quantum tunneling [3]. In capacitor applications, perovskite oxides, such as BaTiO_3 and SrTiO_3 , have been widely employed [8]. Yet, achieving both high permittivity and low leakage current in ultrathin perovskite oxide films remains a major challenge during device miniaturization [3,9]. In addition, they require

high-temperature processing and are therefore difficult to integrate with CMOS technology [5,10].

In both applications, attention is now shifting to HfO_2 as it offers a relatively high dielectric constant ($\epsilon \approx 20$ –30), a wide band gap ($E_g \approx 5.7$ eV), and excellent thermal and chemical stability [2,4,5,11–13]. However, HfO_2 exhibits a smaller band gap than SiO_2 and a lower conduction band offset with silicon, which can lead to increased leakage current [7,13]. Furthermore, its dielectric constant is lower than those of the other high- κ materials such as BaTiO_3 , limiting the achievable capacitance. Thus, considerable effort has also been devoted to exploring next-generation high- κ materials that combine a high dielectric constant with a wide band gap [14,15], and this effort is still ongoing.

Discovering novel materials solely through experiments significantly constrains the candidate space. In recent years, high-throughput first-principles calculations have played an important role in guiding such explorations. Still, their high computational cost limits the scope of searches, and machine learning (ML)-based virtual screening has emerged as a promising alternative [16–18]. Although the band gap can now be predicted accurately by ML [19–21], the accuracy for dielectric constants remains relatively low, which lowers the success rate of virtual screening.

The static dielectric tensor (ϵ) is described as the sum of the electronic (ϵ^{ele}) and the ionic (ϵ^{ion}) contributions, in addition to the vacuum permittivity: $\epsilon = \mathbf{1} + \epsilon^{\text{ele}} + \epsilon^{\text{ion}}$. Here, dielectric constants are described using the relative

*These authors contributed equally to this work.

†Contact author: sin@tohoku.ac.jp

‡Contact author: yukumagai@tohoku.ac.jp

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

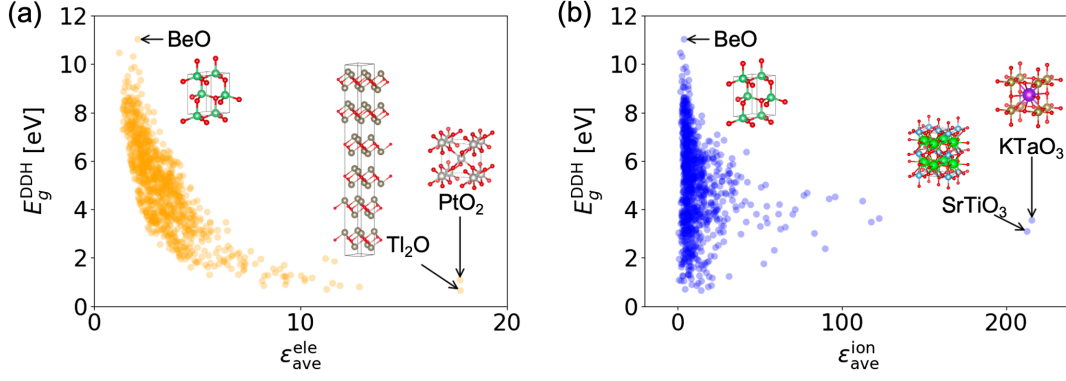


FIG. 1. Relationship between dielectric constants and band gaps. Spherically averaged (a) electronic ($\epsilon_{\text{ave}}^{\text{ele}}$) and (b) ionic ($\epsilon_{\text{ave}}^{\text{ion}}$) dielectric constants are plotted against band gaps (E_g^{DDH}). The dielectric constants were computed using PBEsol, while the band gaps were obtained using the DDH (see text for details). All data are derived from first-principles calculations performed on the 928 oxides.

permittivity. The electronic part $\epsilon_{\alpha\beta}^{\text{ele}}$ is derived via the Kramers-Kronig relation:

$$\epsilon_{\alpha\beta}^{\text{ele}} = \frac{2}{\pi} \mathcal{P} \int_0^\infty \frac{\epsilon_{\alpha\beta}^{\text{img}}(\omega')}{\omega'} d\omega', \quad (1)$$

where α and β denote directions in Cartesian coordinates, $\mathcal{P}$ denotes the principal value, and $\epsilon_{\alpha\beta}^{\text{img}}$ is the imaginary part of the dielectric function [22]. As seen in the integrand, the denominator ω' means that lower transition energy (i.e., smaller band gap) contributes more to $\epsilon_{\alpha\beta}^{\text{ele}}$ because $\epsilon_{\alpha\beta}^{\text{img}}$ vanishes below the band gap [23]. Therefore, ϵ^{ele} exhibits a trade-off relationship with the band gap, as shown in Fig. 1(a).

In contrast, as illustrated in Fig. 1(b), the ϵ^{ion} does not necessarily correlate with a band gap because it is determined by the Born effective charges (BECs) and phonon properties, as will be shown later. In addition, ϵ^{ion} can reach significantly higher values, such as several hundred, meaning that accurately predicting ϵ^{ion} is pivotal for screening materials that exhibit a high dielectric constant.

Several studies have reported ML-based predictions of ϵ^{ele} and ϵ^{ion} . Despite the diversity in target materials (e.g., oxides [23–25] and other materials [18,26,27]), dielectric properties (scalars [23–28] or tensors [18]), and ML models (including random forest [23,25], support vector regression [24,27], gradient-boosting-based regression [26], and graph neural networks (GNN) [18,27–29]), the prediction accuracy for ϵ^{ion} consistently lags behind that of ϵ^{ele} . In particular, the predictions tend to be less accurate for materials with higher ϵ^{ion} as they are likely to lie in the extrapolation regime. Thus, the previous prediction models are hardly applicable to the screening of high- κ materials.

In this work, we develop a high-accuracy ML model for predicting ϵ^{ion} by employing separate ML models for BEC tensors and phonon properties and integrating these predicted quantities to compute ϵ^{ion} analytically. We focus

on oxides since they are usually easy to synthesize under ambient conditions and have larger band gaps. The commercially used and intensively investigated dielectrics are indeed mainly oxides such as SiO_2 , BaTiO_3 , and HfO_2 . Leveraging our framework, we conducted a virtual screening of 8717 oxides and identified 31 high- κ oxides.

II. RESULTS

A. ML predictions of ionic dielectric tensors

The ionic dielectric tensor $\epsilon_{\alpha\beta}^{\text{ion}}$ is given by the following expression [30,31]:

$$\epsilon_{\alpha\beta}^{\text{ion}} = \frac{4\pi}{V_0} \sum_{j\alpha', j'\beta'} Z_{j,\alpha\alpha'}^* (\Phi^{-1})_{j\alpha', j'\beta'} Z_{j',\beta\beta}^*, \quad (2)$$

where V_0 is the volume of the unit cell, α, α', β and β' are directions in Cartesian coordinates, j and j' are the atom indices within the unit cell. $Z_{j,\alpha\alpha}'$ denotes the BEC tensor for ion j , and $(\Phi^{-1})_{j\alpha', j'\beta'}$ is the inverse of the interatomic force constant matrix between ion j and j' , respectively (see Sec. IV). The BEC tensor is defined as

$$Z_{j,\alpha\beta}^* = \frac{V_0}{e} \frac{\partial P_\alpha}{\partial u_{j\beta}}, \quad (3)$$

where e is the elementary charge, P_α is the polarization along direction α , and $u_{j\beta}$ is the displacement of atom j along β . It should be noted that the BEC tensors, being 3×3 second-rank tensors, are in general *nonsymmetric*, as they are defined as second derivatives of the total energy with respect to atomic displacements and applied electric fields [30]. In first-principles calculations, they are calculated based on the density functional perturbation theory (DFPT). The limited accuracy of previous ML models for the ϵ^{ion} can be attributed to the inherent difficulty of simultaneously capturing both BECs and phonon properties within a single ML model. To overcome this, we predict

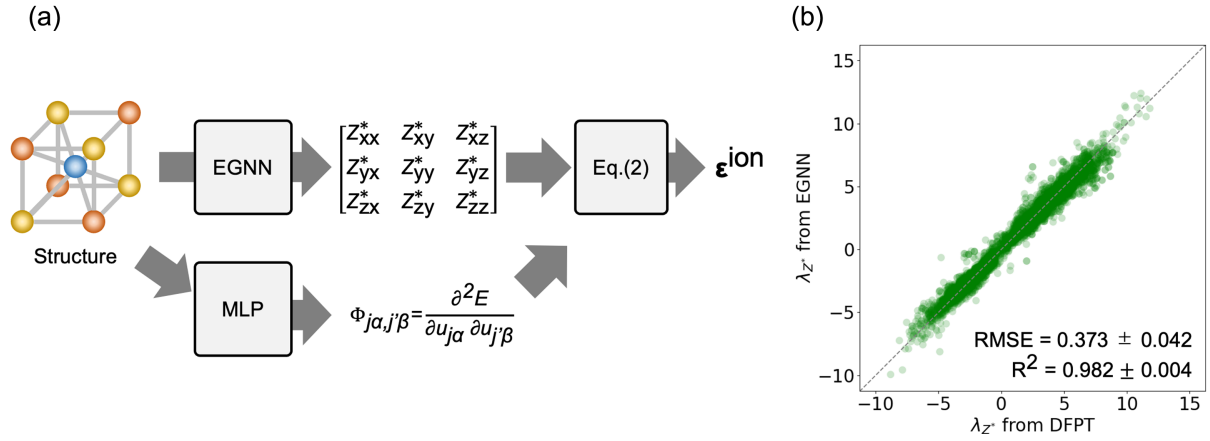


FIG. 2. (a) Schematic workflow for predicting the ϵ^{ion} in our joint model. From a given crystal structure, an EGNN is used to predict the BEC tensors ($\mathbf{Z}^*$) at all the atomic sites, while an MLP provides the force constant matrix (Φ). These quantities are then substituted into Eq. (2) to compute the ϵ^{ion} . (b) Parity plot of the eigenvalues of $\mathbf{Z}^*$ (λ_{Z^*}) between the reference DFPT values and ML predicted values (14 202 inequivalent data points at 4734 sites in the 928 oxides). Results from all test sets in fivefold cross validation are shown.

each separately and analytically combine their outputs using Eq. (2) as shown in Fig. 2(a). We refer to this approach as the *joint* model.

High- κ materials often exhibit anomalously large BECs. For instance, in cubic SrTiO₃, the BECs of Ti and O along the Ti-O bonding direction are +7.26 and −5.73, respectively, which significantly deviate from their nominal ionic charges of +4 and −2 [32]. Such large deviation is known to be related to the dynamic charges [30,33]. In Supplemental Material Fig. S1 [34], we show distributions of the BECs for each element in our dataset. In an extreme case, the BEC for an oxygen atom in NbVO₅ reaches as high as −8.80. Thus, they must be accurately inferred via ML from the surrounding atomic environments.

We developed an ML model for the BECs that is applicable to a broad spectrum of oxides based on an E(3)-Equivariant GNN (EGNN) [43–45] (see Sec. IV and Appendix A). The model was trained on first-principles-calculated data. The dataset, containing 928 oxides composed of 45 elements, was retrieved from our previous studies [46], and covers a wide range of crystal structures (see Supplemental Material Figs. S2–S3 [34]). The EGNN performance was evaluated using fivefold cross validation. The prediction accuracy was assessed by comparing the eigenvalues of the BEC tensors. Figure 2(b) presents the performance of our EGNN model, showing an excellent agreement with an average coefficient of determination (R^2) of 0.982 and root mean squared error (RMSE) of 0.373. While previous approaches for predicting the BECs are typically restricted to a single material [47–49] or a limited class of materials such as perovskites [50], our work is the first to demonstrate that BECs can be predicted with consistently high accuracy across a broad range of chemical compositions and crystal structures (see Appendix B). Note that the acoustic sum rule [30,51] is not explicitly enforced in the EGNN for the BECs; nevertheless, the predicted

BECs satisfy this rule to high accuracy owing to the overall predictive fidelity of the model (see Appendix C).

Another requirement to predict ϵ^{ion} via Eq. (2) is an accurate prediction of phonon properties. Recent advances in machine learning potentials (MLPs) have substantially enhanced the accuracy of their predictions in a wide range of material systems [52–57]. However, the evaluation of ϵ^{ion} is particularly sensitive to the accuracy of low-frequency phonon modes. When the frequencies of optical phonon modes become marginally unstable, the ϵ^{ion} diverges, as seen in Eq. (7) in Sec. IV. To assess the performance of MLPs in predicting phonon properties, we evaluated four pretrained MLPs with a particular focus on their ability to predict the ϵ^{ion} . To isolate the error originating from the MLPs, we calculated the ϵ^{ion} using the force constants predicted by the MLPs together with the BECs obtained from first-principles calculations.

The MLPs include eSEN-uma-s-1-omat (eSEN) [52], ORB-v3 (ORB) [53], SevenNet-MF-ompa (SevenNet) [55,57], and MACE-MPA-0 (MACE) [54]. Based on the Matbench discovery [58], the symmetric relative mean errors in predicted phonon mode contributions to thermal conductivity of eSEN, ORB, SevenNet, and MACE are 0.170, 0.210, 0.317, and 0.412, respectively.

Initially, we discuss the dynamical stability. In this study, a structure is considered dynamically unstable when the fourth lowest phonon frequency, $f_{4\text{th}}$, at the Γ point is smaller than 0.2 THz. Note that the training datasets of these MLPs were generated using the PBE functional [59], while we adopted the PBEsol functional [60]. Partly due to these, some structures were predicted to be unstable at the Γ point although they are stable in our PBEsol-based DFPT calculations. The numbers of such dynamically unstable oxides are 87, 62, 88, and 89 when using eSEN, ORB, SevenNet, and MACE, respectively. In principle, their ϵ^{ion} values are not calculated because imaginary modes at the Γ point give

TABLE I. R^2 and RMSE metrics for the contribution of the predicted force constant matrices to ϵ^{ion} . The evaluation was performed on 786 oxides in the training dataset, which were predicted by all MLPs to be dynamically stable at the Γ point. ϵ^{ion} were estimated based on the force constant matrices predicted by MLPs, and the BECs obtained from DFPT. The predicted ϵ^{ion} were converted to eigenvalues, followed by a base-10 logarithmic conversion, and subsequently compared with the corresponding reference DFPT values. The MLP achieving the highest accuracy is highlighted in bold.

MLP model	R^2	RMSE
eSEN-uma-s-1-omat	0.965	0.062
ORB-v3	0.960	0.066
MACE-MPA-0	0.952	0.073
SevenNet-MF-ompa	0.968	0.059

unphysical negative contribution to the ϵ^{ion} as shown in Eq. (7). Thus, we evaluated the predictive accuracy using 786 oxides, which are predicted to be dynamically stable by all the MLPs. Consequently, we realized that SevenNet achieves the highest accuracy for predicting ϵ^{ion} , although the differences among the MLPs are marginal as shown in Table I (see Supplemental Material Fig. S4 [34] for more details). Accordingly, SevenNet was selected for subsequent calculations involving phonon properties.

Finally, we predicted the ϵ^{ion} via Eq. (2) using the EGNN for the BECs and the SevenNet for phonon properties (the joint model). Its predictive performance, evaluated using fivefold cross validation, was compared with that of a model that directly predicts the ϵ^{ion} (referred to as the direct

model), as shown in Fig. 3 (see Supplemental Material Note 1 [34] for more details). The accuracy metrics were evaluated using 839 oxides that were predicted to be dynamically stable at the Γ point by SevenNet. The direct model exhibits limited predictive capability, yielding an average R^2 of 0.411 and an average RMSE of 0.254. In contrast, the joint model achieves substantially improved accuracy, achieving a higher R^2 of 0.911 and a lower RMSE of 0.099. Although the direct comparison with previous studies [18,23–28] is not strictly appropriate due to differences in datasets and the target material systems, our joint model demonstrates the highest performance in predicting the ϵ^{ion} , highlighting the effectiveness of our joint model (see Appendix D).

To closely examine the sources of the prediction errors, we decomposed them into two contributions: (i) combinations of ML-predicted BECs with DFPT-calculated phonons, and (2) combinations of DFPT-calculated BECs with ML-predicted phonons. This decomposition allows us to isolate how inaccuracies in each component propagate into the final prediction of ϵ^{ion} . The results were further classified into low- and high- ϵ^{ion} regimes using $\lambda_{\epsilon^{\text{ion}}} = 10$ as the dividing threshold, where $\lambda_{\epsilon^{\text{ion}}}$ denotes an eigenvalue of the ionic dielectric tensor. The resulting BEC- and phonon-related contributions to the prediction errors are summarized in Table II. We emphasize that the BEC- and phonon-related errors do not necessarily sum to the total prediction error, as partial cancellation between the two contributions can occur.

While the overall errors arise from both BECs and phonon contributions, in the high- ϵ^{ion} regime, the

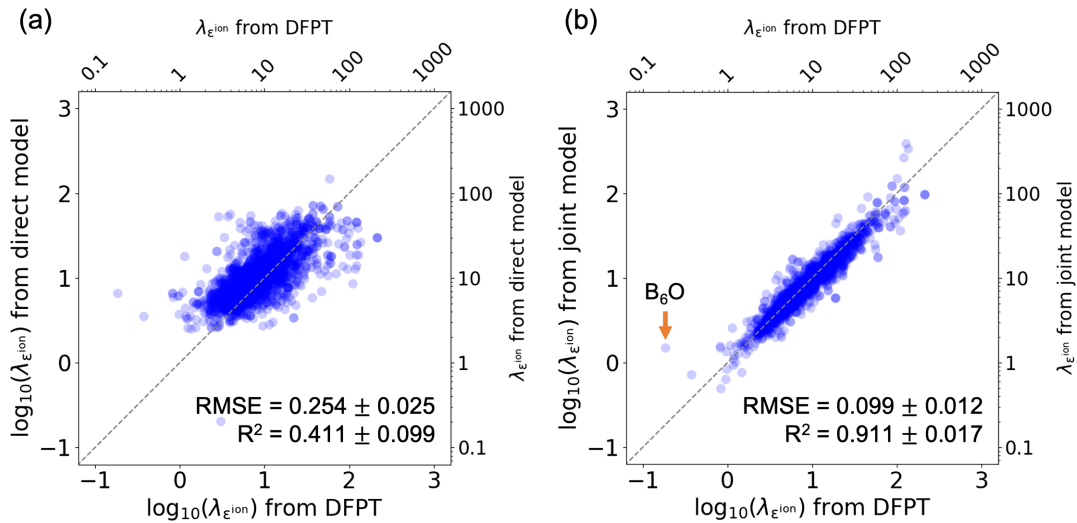


FIG. 3. Performance comparison of the conventional direct model and the proposed joint model. Parity plots of the eigenvalues of the ionic dielectric tensors are shown, comparing the reference DFPT values with the ML-predicted values obtained using (a) the direct model and (b) the joint model. The prediction accuracy was evaluated for the 839 oxides, which were predicted to be dynamically stable by SevenNet. For comparison, the ionic dielectric tensors were converted to eigenvalues ($\lambda_{\epsilon^{\text{ion}}}$) and transformed using a base-10 logarithm to mitigate scale differences. One predicted value of $\lambda_{\epsilon^{\text{ion}}}$ for $\text{Ba}_3\text{Nb}_2\text{ZnO}_9$, specifically -5.135 , is omitted from panel (a) due to its physically ill-defined nature.

TABLE II. Decomposition of ϵ^{ion} prediction errors by Seven-Net into contributions from the BECs and phonon properties. RMSEs were computed for the base-10 logarithm of the eigenvalues of ϵ^{ion} ($\lambda_{\epsilon^{\text{ion}}}$). Results are reported for a dataset comprising 839 dynamically stable oxides. The corresponding mean absolute error (MAE) and R^2 are provided in Supplemental Material Note 2 [34]. RMSEs in the high- and low- ϵ^{ion} regions, separated by the threshold $\lambda_{\epsilon^{\text{ion}}} = 10$, are also tabulated. The “Total” column reports the prediction errors of the joint model, whereas the “Phonon” and “BEC” columns represent the phonon- and BEC-related errors, respectively (see text for details).

Dataset	Total	Phonon	BEC
All	0.100	0.064	0.071
High- ϵ^{ion}	0.120	0.092	0.075
Low- ϵ^{ion}	0.086	0.041	0.069

prediction errors are predominantly governed by phonon-related inaccuracies. This is consistent with the inverse relationship between the ionic dielectric constants and phonon frequencies, as shown in Eq. (7). A single outlier appears in the low- ϵ^{ion} regime: cubic B_2O_3 , which exhibits an exceptionally small $\lambda_{\epsilon^{\text{ion}}}$ of 0.3. Accurate prediction of such small ϵ^{ion} values require highly precise predictions of the BECs. Indeed, the ML-predicted BECs show relatively large deviations from the DFPT values, which arises not only from the intrinsically small magnitude of BECs but also from the lack of chemically similar

environments in the training dataset. Further details of the error decomposition are provided in Supplemental Material Note 2 [34].

B. Virtual screening of high- κ oxides

Using our joint model, we performed a virtual screening of promising high- κ oxides from a candidate pool extracted from the materials project database (MPD) [61]. A stepwise screening procedure was employed, as illustrated in Fig. 4(a). Specifically, we considered oxides for which dielectric constants had not yet been reported but for which band gaps were available. The candidate pool was further restricted to compositions containing only the 45 elements covered by our EGNN for prediction of BECs, and the 928 oxides used in the training dataset were excluded to avoid the data leakage. Application of these criteria yielded 19 569 oxides from the MPD on May 20, 2025.

We then removed metallic oxides, reducing the number of candidates from 19 569 to 13 261. Magnetic oxides were further excluded, as dielectric tensors computed for spin-polarized systems can be highly sensitive to the choice of effective U parameters and to specific spin configurations, thereby limiting the predictive reliability for such magnetic materials. As a result, the final candidate set comprised 8717 oxides. The distributions of elements in the training dataset and the screening candidates exhibit substantial overlap, supporting the applicability of our joint model to virtual screening (see Supplemental Material Figs. S2–S3 [34]).

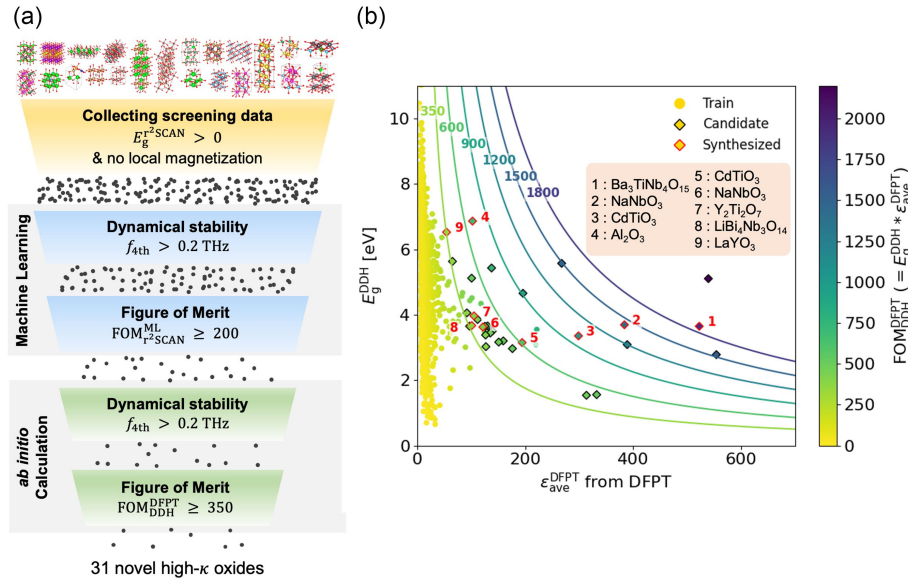


FIG. 4. (a) Schematic of the screening workflow to identify novel high- κ oxides. Details are provided in the main text and the *Methods* section. (b) Plot of the spherically averaged dielectric constant ($\epsilon_{\text{ave}}^{\text{DFPT}}$) versus the band gap (E_g^{DDH}) for the 928 oxides in the ML training dataset and the 31 high- κ oxides identified by virtual screening ($\text{FOM}_{\text{DDH}}^{\text{DFPT}} \geq 350$). The dielectric constants were calculated using PBEsol-based DFPT, and the band gaps were computed using the DDH. The color scale represents the $\text{FOM}_{\text{DDH}}^{\text{DFPT}}$. Circles denote oxides in the training set, while diamonds indicate the identified high- κ oxides in this study. Among the diamonds, those enclosed in red lines correspond to synthesized oxides, whose formula are also described in the inset.

We first predicted the phonon properties at the Γ point using the SevenNet and excluded dynamically unstable oxides. Consequently, 4654 oxides survived.

In the second stage, we performed screening based on the r^2 SCAN band gaps reported in the MPD and the ML-predicted dielectric constants. Although the r^2 SCAN band gaps underestimate experimental values by approximately 50% (see Supplemental Material Fig. S5 [34]), their strong correlation with experimental values supports their use for screening purposes. To evaluate the static dielectric tensor (ϵ), both ϵ^{ele} and ϵ^{ion} are required. Accordingly, we additionally constructed an EGNN model to predict the ϵ^{ele} . The prediction accuracy was assessed using fivefold cross validation for the 839 oxides predicted to be dynamically stable by SevenNet. For $\log_{10}(\lambda_{\epsilon^{\text{ele}}})$ and $\log_{10}(\lambda_{\epsilon})$, the models achieved an average R^2 values of 0.714 and 0.905, respectively (see Sec. IV and Supplemental Material Note 1 [34]). Finally, the EGNNs for both the ϵ^{ele} and BECs were retrained using full dataset of 928 oxides to further improve the predictive accuracy.

Since superior dielectric materials exhibit both a high dielectric constant and a wide band gap, we introduced a figure of merit (FOM) that captures these two key properties simultaneously. Here, FOM was defined as $\epsilon_{\text{ave}} \cdot E_g$, where ϵ_{ave} denotes the spherically averaged static dielectric constant, and the band gap E_g is in eV, following previous studies [14,15,17,18]. Using these ML-predicted dielectric constants ($\epsilon_{\text{ave}}^{\text{ML}}$) and r^2 SCAN band gaps ($E_g^{\text{r}^2\text{SCAN}}$), we calculated the FOM, which is referred to as $\text{FOM}_{\text{r}^2\text{SCAN}}^{\text{ML}}$, for each oxide and applied a screening threshold of $\text{FOM}_{\text{r}^2\text{SCAN}}^{\text{ML}} \geq 200$. Consequently, 142 oxides remained.

In the third stage, we conducted DFPT calculations using PBEsol to validate both the predicted $\epsilon_{\text{ave}}^{\text{ML}}$ and the dynamical stability of the screened candidates. Candidates for which the first-principles calculations failed to converge were also excluded. This step reduced the pool to 121 oxides. We then calculated the band gaps using dielectric-dependent hybrid functional (DDH) calculations [62], which predict the band gaps of oxides accurately (see Supplemental Material Fig. S5 [34]). Applying a more stringent $\text{FOM}_{\text{DDH}}^{\text{DFPT}} \geq 350$ criterion subsequently reduced the number of candidates to 31. Identified novel high- κ oxides are plotted together with the training data in Fig. 4(b).

Among the training data, 9 out of 928 oxides exhibited $\text{FOM}_{\text{DDH}}^{\text{DFPT}} \geq 350$ (see Table III). In contrast, among 142 high- κ oxides with $\text{FOM}_{\text{r}^2\text{SCAN}}^{\text{ML}} \geq 200$, 31 oxides satisfy $\text{FOM}_{\text{DDH}}^{\text{DFPT}} \geq 350$. The properties of these 31 candidates are summarized in Table III, which lists the top ten oxides ranked by $\text{FOM}_{\text{DDH}}^{\text{DFPT}}$, together with those that are either experimentally synthesized or exhibit an energy above the convex hull, E_{hull} (in eV/atom), of zero. A complete list of

the 31 identified high- κ oxides, along with the nine oxides from the training dataset, is provided in Supplemental Material Note 3 [34]. An important benchmark is KTaO_3 , which exhibits the highest $\text{FOM}_{\text{DDH}}^{\text{DFPT}}$ among the 928 training oxides ($E_g^{\text{DDH}} = 3.5$ eV, $\epsilon_{\text{ave}}^{\text{DFPT}} = 221$, and $\text{FOM}_{\text{DDH}}^{\text{DFPT}} = 785$). Among the 31 identified oxides, eight exhibit higher $\text{FOM}_{\text{DDH}}^{\text{DFPT}}$ values than KTaO_3 , demonstrating a marked improvement over the best-performing oxide in the training dataset.

For both the nine oxides in the training dataset and the 31 identified high- κ oxides with $\text{FOM}_{\text{DDH}}^{\text{DFPT}} \geq 350$, the ratios of ionic contributions to the total dielectric constants exceed 90%, reinforcing the notion that accurate prediction of ϵ^{ion} is the key to identifying novel high- κ oxides.

Of the 31 identified oxides, experimental synthesis has already been reported for 9 oxides. For four of these oxides, experimental values of dielectric constant ($\epsilon_{\text{ave}}^{\text{Exp}}$) have been reported, and all four exhibit high $\epsilon_{\text{ave}}^{\text{Exp}}$ values, consistent with our predictions. For the remaining five oxides, no $\epsilon_{\text{ave}}^{\text{Exp}}$ have been reported. Although CdTiO_3 in the $Pmc2_1$ phase has been experimentally synthesized [74], previous studies suggest that alternative phases, $Pbnm$ at room temperature and $Pna2_1$ below 80 K, are more stable [76–78]. NaNbO_3 in the $Pna2_1$ phase was investigated by Parker *et al.*, who identified the $Pna2_1$ space group as the most suitable structure for its polar phase [69–72]. Al_2O_3 in the $Cmcm$ phase adopts the CaIrO_3 -type post-perovskite structure, which typically forms only under high-pressure conditions [75]. In contrast, $\text{LiBi}_4\text{Nb}_3\text{O}_{14}$ and LaYO_3 have been successfully synthesized under ambient conditions [88,89]. $\text{LiBi}_4\text{Nb}_3\text{O}_{14}$ has been studied as a photocatalyst, though its dielectric properties have not yet been measured. LaYO_3 possesses a perovskite structure and has been predicted as a promising high- κ oxide by Coh *et al.* [136], but no experimental ϵ_{ave} has been reported to date. Thus, $\text{LiBi}_4\text{Nb}_3\text{O}_{14}$ and LaYO_3 are experimentally accessible as-yet-unexplored high- κ oxides that require future experimental validation.

Analysis of the 31 identified oxides revealed that 29 include transition metals with empty d -orbitals, namely Y^{3+} , Ti^{4+} , Zr^{4+} , Hf^{4+} , Nb^{5+} , Ta^{5+} , and W^{6+} . This makes sense as such d^0 -ness condition is commonly associated with large dynamic charges and, consequently, an enhanced dielectric response [137]. Given that such d^0 oxides exhibit anomalous BECs that deviate significantly from the nominal ionic charges, our EGNN predictions are indispensable for reliably identifying these oxides. Exceptions are only CaIrO_3 -type Al_2O_3 , and Sr_3CaO_4 . Unfortunately, their syntheses under ambient conditions are expected to be difficult, as the former case is a high-pressure phase [75], while the latter is a theoretically predicted structure [138] and is expected to show strong cation disordering.

TABLE III. Properties of identified novel high- κ oxides as well as those in the training dataset, including the calculated band gaps (E_g^{DDH} , in eV), spherically averaged electronic and ionic dielectric constants ($\epsilon_{\text{ave}}^{\text{ele,DFPT}}$, $\epsilon_{\text{ave}}^{\text{ion,DFPT}}$), and figures of merit ($\text{FOM}_{\text{DDH}}^{\text{DFPT}}$). Band gaps were computed using the DDH, whereas dielectric constants were calculated using PBEsol. The “Group” column classifies each oxide as follows: “Top” denotes the top 10 screened oxides ranked by $\text{FOM}_{\text{DDH}}^{\text{DFPT}}$; “Expt” corresponds to oxides that have been experimentally reported; “Stable” exhibits the oxides with an energy above the convex hull (E_{hull}) of zero; “Train” indicates oxides from the training dataset. The “ E_{hull} ” column corresponds to the values reported in the MPD, and “...” means that E_{hull} is zero. The column “Synthesized” denotes whether the oxide has been experimentally synthesized. When available, experimental dielectric constants ($\epsilon_{\text{ave}}^{\text{Exp}}$) and characteristic crystal structure names (e.g., perovskite, pyrochlore) are noted. References for previously reported syntheses and experimental dielectric constants are also provided in Supplemental Material Note 3 [34].

Group	Formula	E_g^{DDH}	$\epsilon_{\text{ave}}^{\text{ele,DFPT}}$	$\epsilon_{\text{ave}}^{\text{ion,DFPT}}$	$\text{FOM}_{\text{DDH}}^{\text{DFPT}}$	E_{hull}	Synthesized	$\epsilon_{\text{ave}}^{\text{Exp}}$	Structure
Top	SrZrO ₃	5.1	3.7	535	2752	0.02			Perovskite
	Ba ₃ TiNb ₄ O ₁₅	3.7	5.0	516	1912	0.02	✓ [63]	900 [64–68]	ReO ₃
	Pb ₂ TiZrO ₆	2.8	5.8	547	1540	0.04			Perovskite
	Sr ₃ Hf ₂ O ₇	5.6	3.2	263	1490	0.02			Layered perovskite
	NaNbO ₃	3.7	4.8	377	1421	0.00	✓ [69–72]	200 [73]	Perovskite
	AgTaO ₃	3.1	5.7	382	1200	0.05			Perovskite
	CdTiO ₃	3.4	6.7	290	1001	0.04	✓ [74]		Perovskite
	Ta ₂ Zn ₄ O ₉	4.7	4.2	190	911	0.10			
	Sr ₃ CaO ₄	5.4	2.9	134	747	0.03			Caswellsilverite-like
	Al ₂ O ₃	6.9	2.5	98	699	0.28	✓ [75]		CaIrO ₃
Expt	CdTiO ₃	3.2	6.7	186	611	0.04	✓ [76–78]	2100 [79]	Perovskite
	NaNbO ₃	3.6	4.9	115	439	0.00	✓ [69]		Perovskite
	Y ₂ Ti ₂ O ₇	4.0	5.2	99	416	0.01	✓ [80–86]	54 [87]	Pyrochlore
	LiBi ₄ Nb ₃ O ₁₄	3.7	5.8	94	371	...	✓ [88]		Aurivillius derived
	LaYO ₃	6.5	3.6	49	351	0.04	✓ [89]		Perovskite
Stable	Ta ₄ Cd ₃ HgO ₁₄	3.0	5.0	170	522	...			
	SrTa ₄ Cd ₃ O ₁₄	3.9	4.5	106	430	...			
	Ta ₂ CdPbO ₇	3.4	5.2	120	427	...			
	Ta ₂ HgPbO ₇	3.0	5.5	121	384	...			
	CaTa ₄ Cd ₃ O ₁₄	4.1	4.5	86	372	...			
Train	KTaO ₃	3.5	4.5	216	785	...	✓ [90,91]	5000 [92]	Perovskite
	SrTiO ₃	3.1	5.7	213	678	...	✓ [93–105]	4500 [106]	Perovskite
	NaTaO ₃	4.3	4.3	113	512	...	✓ [107–110]	147 [111]	Perovskite
	Cd ₂ Ta ₂ O ₇	3.6	4.8	123	466	...	✓ [112,113]	1200 [114]	Perovskite
	Ba ₃ Nb ₂ CdO ₉	4.5	4.2	96	457	...			Perovskite derived
	CaTiO ₃	3.5	5.6	118	433	...	✓ [77,115–131]	180 [132]	Perovskite
	Ba ₃ CaNb ₂ O ₉	4.4	4.0	92	431	...	✓ [133]	40 [134]	Perovskite derived
	KNbWO ₆	4.7	4.0	79	392	...	✓ [135]		Pyrochlore derived
	TaTiWO ₆	4.6	4.4	77	376	...			

Almost all final candidates are composed of oxygen octahedra network centered around transition metal cations. Such octahedral frameworks are commonly observed in crystal structures that exhibit high dielectric constants, including perovskite structures, such as BaTiO₃ and SrTiO₃ [8]. The perovskite-related frameworks identified in this study encompass distorted perovskites, layered perovskites, post-perovskites, pyrochlores, and ReO₃ structures. LiBi₄Nb₃O₁₄ is specifically classified as a variant of the Aurivillius structure, which is a layered perovskite. The predicted high dielectric constants are primarily attributed to their structural flexibility, particularly the adaptability of their octahedral networks. One exceptional structure is fluorite TiO₂ that does not exhibit oxygen octahedra

network but are known to be stable only under high-pressure conditions. Although identified oxides mostly possess similar oxygen octahedral networks, some fall outside the conventional chemical design space. This highlights the importance of virtual screening conducted without expert knowledge, which can otherwise bias the search space.

III. CONCLUSIONS AND OUTLOOK

In this study, we propose an approach that enables accurate prediction of the ionic contributions to the dielectric constants (ϵ^{ion}), the dominant contributor to the dielectric response in high- κ materials. The key idea

is to factorize ϵ^{ion} into BECs and phonon properties, predicting the former using an EGN and the latter with a highly accurate pretrained MLP. We achieved the accurate predictions of BECs using a training dataset comprising 928 oxides. Compared to conventional single-model approaches, our decomposition and integration strategy leads to a significant improvement in prediction of ϵ^{ion} . Using this framework, we successfully identified 31 high- κ oxides from a pool of 8717 candidates.

We emphasize that, as indicated by Eq. (7) in Sec. IV, larger ϵ^{ion} are associated with marginally stable optical phonon frequencies. Consequently, the prediction of high dielectric constants becomes highly sensitive to even small errors in phonon frequencies. Moreover, if the phonon frequencies are erroneously predicted to be imaginary, the corresponding dielectric constants cannot be evaluated. Therefore, accurate prediction of phonon properties is a critical aspect of this study. This implies that improvements in both the first-principles calculations and the MLP trained based on their outputs directly contribute to enhancing the accuracy of predicting the ϵ^{ion} .

As future directions following this study, two main avenues can be considered: extending the material systems to nonoxides and incorporating finite-temperature effects. The former can be realized simply by expanding the dataset to include nonoxides. On the other hand, the latter can be addressed by performing anharmonic phonons [139] or molecular dynamics calculations [140] using MLP. These extensions will improve the screening of superior high- κ dielectrics. Furthermore, the screened candidates can be theoretically investigated in depth in terms of band alignment with contact metals and defects that affect conductivity and dielectric constants.

Finally, we emphasize that our physics-based factorization approach is applicable not only to dielectric tensors, but also to a broad range of material properties that can be derived from distinct fundamental quantities. We anticipate

that this study will offer new perspectives for predicting physical properties.

IV. METHODS

A. Definition and computational approach of force constant matrix

Force constant matrix $\Phi_{ja,j'\beta}$ represents the rate of change of the restoring force that arises in the α direction on atom j when atom j' is slightly displaced in the β direction from the equilibrium positions. For atoms j and j' , and coordinate components α and β , the elements of the force constant matrix are defined as

$$\Phi_{ja,j'\beta} = \frac{\partial^2 E}{\partial u_{j\alpha} \partial u_{j'\beta}} = -\frac{\partial F_{j\alpha}}{\partial u_{j'\beta}}, \quad (4)$$

where E denotes potential energy of the system, and $u_{j\alpha}$ denotes displacement of atom j in the α direction.

In this calculation, the force constant matrix is obtained using the finite-difference method, whereby an atom is displaced by 0.01 Å, and the resulting forces on the other atoms are subsequently evaluated using MLPs [141] as follows:

$$\Phi_{ja,j'\beta} \approx -\frac{\Delta F_{j\alpha}}{\Delta u_{j'\beta}}. \quad (5)$$

The training dataset of the MLPs adopted in this study were generated using the PBE functional. Nevertheless, as summarized in Table I, all the models attain reasonable accuracy of predicting the ϵ^{ion} .

B. Derivation of the ionic dielectric constant in the static limit

The ionic dielectric tensor, which denotes the ionic contribution to the dielectric constants, is expressed by the following expression [30,31]:

$$\epsilon_{\alpha\beta}^{\text{ion}}(\omega) = \frac{4\pi}{V_0} \sum_m \frac{(\sum_{j\alpha'} Z_{j,\alpha\alpha'}^* U_{j\alpha',mq=0})(\sum_{j'\beta'} Z_{j',\beta\beta'}^* U_{j'\beta',mq=0})}{\omega_m^2 - \omega^2}, \quad (6)$$

where V_0 is the volume of the unit cell, and $U_{j\alpha',mq=0} = ([e_{m,j\alpha'}]/[\sqrt{M_j}])$ denotes the eigendisplacement of phonon mode m at the Γ point ($q = 0$), mass-normalized by the atomic mass M_j of atom j . ω_m is the frequency of phonon mode m , and ω is the frequency of the applied electric field. Taking the static limit ($\omega \rightarrow 0$), Eq. (6) simplifies to

$$\epsilon_{\alpha\beta}^{\text{ion}}(0) = \frac{4\pi}{V_0} \sum_m \frac{(\sum_{j\alpha'} Z_{j,\alpha\alpha'}^* U_{j\alpha',mq=0})(\sum_{j'\beta'} Z_{j',\beta\beta'}^* U_{j'\beta',mq=0})}{\omega_m^2}. \quad (7)$$

To proceed, we employ the spectral decomposition of the dynamical matrix $\mathbf{D}$, whose inverse is given by

$$(\mathbf{D}^{-1})_{j\alpha',j'\beta'} = \sum_m \frac{e_{m,j\alpha'} e_{m,j'\beta'}}{\omega_m^2}. \quad (8)$$

Using the relation between the normalized displacements and the dynamical matrix:

$$\frac{1}{\sqrt{M_j M_{j'}}} (\mathbf{D}^{-1})_{j\alpha', j'\beta'} = \sum_m \frac{U_{j\alpha', m\mathbf{q}=0} U_{j'\beta', m\mathbf{q}=0}}{\omega_m^2}, \quad (9)$$

we can rewrite the ionic dielectric tensor as

$$\epsilon_{\alpha\beta}^{\text{ion}}(0) = \frac{4\pi}{V_0} \sum_{j\alpha', j'\beta'} Z_{j\alpha'}^* \left(\frac{1}{\sqrt{M_j M_{j'}}} (\mathbf{D}^{-1})_{j\alpha', j'\beta'} \right) Z_{j'\beta'}. \quad (10)$$

Finally, by introducing the force constant matrix Φ , Eq. (2) is obtained (see Appendix E for details). Note that this expression is valid only when all optical phonon frequencies are real ($\omega_m^2 > 0$). Thus, the presence of imaginary phonon modes, which signal dynamical instability of the lattice, invalidates the above formulation and prevents a meaningful evaluation of the ionic dielectric tensor. In this study, $(\Phi^{-1})_{j\alpha', j'\beta'}$ were evaluated using the PHONOPY code [141].

C. EGNN model

The overall architecture of our ML model is illustrated in Appendix A. The model architecture is inspired by EGNNs such as Nequip, which also leverage symmetry-aware message passing mechanisms [43–45, 142]. The model is implemented primarily using the e3nn library [143], which enables equivariant operations under Euclidian symmetry. In this model, node and edge features are represented using irreducible representations of the three-dimensional rotation and inversion group $O(3)$. Each representation is labeled by two indices, namely, the angular momentum degree ($l = 0, 1, 2, \dots$), and the parity, where “e” and “o” stand for even and odd parity, respectively. The nodes are initialized using atomic embedding vectors in the same manner as in CGCNN [144], while the edges are encoded based on interatomic distances via radial basis functions. The initialized node features and edge features are represented as 0e. Additionally, the normalized relative vectors between atoms are encoded into the 0e, 1o, and 2e components using spherical harmonics. Furthermore, by taking tensor products between 0e edge features and corresponding spherical harmonic representations, the edge features are expanded into the 0e, 1o, and 2e components. Through convolution layers, tensor products are computed between node features and edge features, and the node features ultimately have up to 0e, 1o, 1e, and 2e components. Finally, the model outputs the 0e and 2e components for the dielectric tensor, and the 0e, 1e, and 2e components for the BEC. More details are described in Appendix A.

The EGNN model was trained using specific hyperparameters. The feature dimensions were set to 64 for 0e,

16 for 1e and 1o, and 32 for 2e. The edge dimension was fixed at 32, and three convolutional layers with SiLU activations [145] were used. The batch size was set to 16. Except for the initial learning rate, all hyperparameters were kept fixed through all experiments. The initial learning rate was optimized experimentally, starting from 6.25×10^{-4} and doubled incrementally up to a maximum value of 1.60×10^{-1} . The optimized learning rates determined for each task are summarized in Supplemental Material, Table I [34]. For model training, the mean squared error (MSE) was used as the loss function, and optimization was performed using the Adam optimizer [146]. The selected learning rate decayed by a factor of 0.9 every 30 epochs. To ensure reproducibility, a fixed random seed was used throughout all experiments. All training runs were performed on four NVIDIA A100 PCIe GPU with 80 GB of memory.

D. Training and test datasets

Model evaluation was carried out using fivefold cross validation. In each fold, 20% of the data was reserved for testing, while the remaining 80% was split into 80% for training and 20% for validation. Early stopping was applied based on validation loss, with a maximum of 300 epochs. Training was terminated if no improvement occurred over 100 consecutive epochs.

After cross validation, the model was retrained on the entire dataset of 928 oxides using the optimized initial learning rate to produce the final predictive model. This step ensured the use of all available data and maximized prediction accuracy. The total number of training epochs in this retraining phase was calculated as the average across the five folds. This final predictive model was then used for virtual screening.

E. Details of first-principles calculations

First-principles calculations were performed using the projector augmented-wave (PAW) [147] method, as implemented in the Vienna *ab initio* simulation package (VASP) [148]. Structure relaxation and dielectric tensor calculations were performed using PBEsol + U , and band gap calculations were carried out with the DDH. The PAW dataset, the effective U parameters, cutoff energies, densities of the k -point mesh, structure optimization criteria were the same as those used in our previous study [46].

ACKNOWLEDGMENTS

This work has been supported by JST FOREST Program (JPMJFR235S), and KAKENHI (25K01486) and the Asahi Glass Foundation.

DATA AVAILABILITY

The data that support the findings of this article are openly available [149], embargo periods may apply.

APPENDIX A: DETAILS OF EGNN ARCHITECTURE IN THIS STUDY

In this study, EGNN architectures were constructed and trained to predict three target properties: BECs, ϵ^{ele} , and ϵ^{ion} . An overview of the architecture is provided in Fig. 5. The model is implemented using the e3nn [143] library. The SiLU function is adopted as the activation function.

The architecture operates as follows. Nodes and edges are first vectorized, with edges constructed using a cutoff distance of 5 Å. Convolutional operations are then performed over three layers, processing features corresponding to 0e, 1o, 1e, and 2e representations. For the dielectric constants, only the 0e and 2e representations are used in the final output, whereas for the prediction of BECs, the 0e, 1e, and 2e representations are employed, considering the symmetries of the matrices.

The EGNN components are described below.

Node embedding (Embedding). Node features are constructed from atomic numbers using a dictionary-based approach, employing the same vector representation as used in CGCNN [144]. These vectors are 92-dimensional one-hot encodings incorporating descriptors such as atomic number and electronegativity.

Radial basis function (RBF). The edge vectorization is carried out using the `soft_one_hot_linspace` method from the `math` module of e3nn, enabling the construction of smooth radial features based on interatomic distance r_{ij} through Gaussian basis functions.

Fully connected net layer (fcnet). The function $\text{fcnet}(x)$ is implemented using the `FullyConnectedNet` class from the `nn` module of e3nn, which is defined as

$$\text{fcnet}(x) = \frac{xW}{\sqrt{h^{\text{in}}\sigma^2}}. \quad (\text{A1})$$

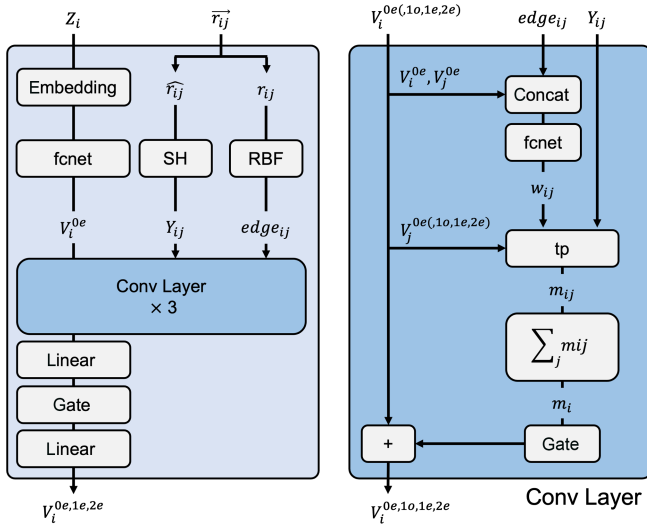


FIG. 5. The EGNN architecture adopted in this study.

Here, h^{in} and h^{out} denote the input and output dimensions, $W \in \mathbb{R}^{h^{\text{in}} \times h^{\text{out}}}$ is a learnable weight matrix, and σ represents the standard deviation of the input x .

Linear layer (Linear). This layer combines a fully connected transformation with a nonlinear activation function, while maintaining $O(3)$ equivariance. The transformation is formulated as

$$\text{Linear}(x) = \left(\bigoplus_l (W^l \otimes I_{2l+1}) x^l \right) + b^0. \quad (\text{A2})$$

Here, I_{2l+1} is the identity matrix in the representation space of degree l , $W^l \in \mathbb{R}^{n_l^{\text{out}} \times n_l^{\text{in}}}$ are learnable weights and b^0 is the bias term added only to scalar features. This operation is implemented using the `Linear` class from the `o3` module of e3nn.

Equivariant nonlinear layer (Gate). This layer applies nonlinear transformations to both scalar ($l = 0$) and higher-order ($l > 0$) features. The operation is defined as

$$\text{Gate}(x, g, y) = \left(\bigoplus_i \phi(x_i) \right) \oplus \left(\bigoplus_j \phi(g_j) y_j \right). \quad (\text{A3})$$

Here, ϕ denotes the activation function, x_i and g_j are scalar features, and y_j represents higher-order features. The scalar inputs x_i and g_j are typically obtained by splitting the input scalar features into two parts. This component is implemented via the `Gate` class from the `nn` module of e3nn.

Tensor product (tp). In our model, tensor product operations are employed to lift scalar features into higher-order representations such as vectors and second-rank tensors, while preserving spatial equivariance. This enables the network to encode structural information in a symmetry-aware manner. Specifically, the tensor product is computed between node-assigned features and spherical harmonics, which serve as irreducible representations of the $O(3)$ group and capture angular dependencies. For instance, taking the tensor product of a scalar feature and a spherical harmonic allows the model to construct features of higher angular momentum (i.e., $l > 0$). The resulting features are decomposed into components corresponding to their total angular momentum l . When combining two features of angular momenta l_1 and l_2 , the output contains components with $|l_1 - l_2| \leq l \leq l_1 + l_2$.

In this work, the model is designed to handle components up to $l = 2$. Through this mechanism, the model effectively incorporates geometric information such as interatomic distances and directions while maintaining compliance with spatial symmetries. As a result, it naturally constructs representations that are invariant or equivariant under transformations like rotation and reflection.

The tensor product operation is formulated in e3nn as:

$$\text{tp}(x, y, w) = \sum_{u,v} w_{uvw} x_u \otimes y_v. \quad (\text{A4})$$

Here, x_u represents the features assigned to node j (denoted V_j), y_v corresponds to the spherical harmonic Y_{ij} , and w_{uvw} are weights derived from the edge features w_{ij} , described below. This operation is implemented efficiently via the FullyConnectedTensorProduct class within o3 module of the e3nn library.

Convolution layer (Conv Layer). Our model primarily updates atomic features using operations based on tensor products. The tensor product weights w_{ij} are generated from the vectorized edge features (denoted edge_{ij}) and scalar features of atom i and j , denoted $V_i^{l=0}$ and $V_j^{l=0}$:

$$w_{ij} = \text{fcnet}(\text{concatenate}(V_i^{l=0}, V_j^{l=0}, \text{edge}_{ij})). \quad (\text{A5})$$

Using the generated weight w_{ij} , the feature vector V_i associated with atom i is updated. Inspired by the ResNet architecture [45], a residual connection is employed to facilitate efficient gradient propagation:

$$V_i^{(t+1)} = V_i^{(t)} + \sum_j m_{ij}, \quad (\text{A6})$$

where m_{ij} is a message-passing feature, defined as follows:

$$m_{ij} = \text{tp}(V_j, Y_{ij}, w_{ij}), \quad (\text{A7})$$

where $V_i^{(t)}$ is the feature vector at the t -th convolution.

APPENDIX B: PREVIOUS STUDIES ON MACHINE-LEARNING PREDICTION OF BORN EFFECTIVE CHARGES

Here, we summarize previous studies on the ML prediction of the BECs.

- (i) Falletta *et al.* [47] constructed MLP-based models to simultaneously predict energies, forces, polarization, polarizability, and BECs from differential relationships between a generalized potential function and applied external fields. Models were trained for α -SiO₂ and BaTiO₃ using *ab initio* molecular dynamics (AIMD) snapshots independently. The resulting MAEs for BECs are 0.005 for α -SiO₂ and 0.011 for BaTiO₃.
- (ii) Shimizu *et al.* [49] developed a neural-network model based on vector atomic fingerprints for predicting BECs from atomic structures to study ion migration under electric fields. The model was trained on BECs from 17 900 Li₃PO₄ structures,

including both pristine and defect-containing configurations. The model achieved an RMSE of 0.038.

- (iii) Kutana *et al.* [50] predicted the BECs in perovskite oxides (ABO_3 in the $Pnma$ space group; $A = \text{Ca, Sr, Ba, Pb}$ and $B = \text{Ti, Zr, Hf}$) as well as those in Li₃PO₄ and ZrO₂ using EGNNs. The reported MAEs for BECs are 0.020–0.032 when trained jointly and 0.009–0.033 when trained separately for each material class.
- (iv) Schmiedmayer and Kresse [48] developed a tensorial equivariant derivative learning that predict BECs using λ -SOAP descriptors combined with linear regression for predicting finite-temperature infrared spectra. The resulting RMSEs were 0.032 for liquid water and 0.040 for MAPbI₃.

Although the numerical errors reported in previous studies [46–49] are approximately 1 order of magnitude smaller than those obtained with the present model (MAE = 0.231 ± 0.018 , RMSE = 0.373 ± 0.042), these approaches are generally applied to specific materials or limited classes of materials.

APPENDIX C: ACOUSTIC SUM RULE FOR PREDICTED BORN EFFECTIVE CHARGES

The BEC tensors satisfy the acoustic sum rule

$$\sum_j Z_{j,\alpha\beta}^* = 0, \quad (\text{C1})$$

because rigid translation of the crystal does not modify the internal charge distribution, so the macroscopic polarization must remain unchanged.

$$\nabla \mathcal{P}_\alpha = \sum_j \frac{\partial P_\alpha}{\partial u_{j\beta}} u_{j\beta} = \frac{1}{V_0} \left(\sum_j Z_{j,\alpha\beta}^* \right) u_\beta = 0. \quad (\text{C2})$$

In the present work, this constraint is not explicitly imposed during the training or inference of the EGNN for predicting the BECs. To assess its impact, we quantitatively evaluated the degree to which the acoustic sum rule is satisfied. Specifically, we introduce direction-resolved neutrality metrics that characterize the degree of electrical neutrality along each Cartesian direction $\alpha = x, y, z$, defined as

$$\Delta = \frac{1}{3} \sum_\alpha \overline{Z_{\alpha\alpha}^*}, \quad (\text{C3})$$

where

$$\overline{Z_{\alpha\alpha}^*} = \frac{1}{N} \left| \sum_{j=1}^N Z_{j,\alpha\alpha}^* \right|. \quad (\text{C4})$$

These metrics measure the residual net effective charge along each direction, and thus provide a measure of the

accuracy with which electrical neutrality is satisfied. Across all oxides in all test sets of the fivefold cross validation, we obtain $\Delta = 0.061 \pm 0.060$, indicating that the relative residual violation is typically small compared with the absolute value of the nominal oxygen charge of -2 .

In addition, we explicitly enforced the acoustic sum rule on the predicted ϵ^{ion} by constantly shifting the BECs of all atoms so that the ASR is satisfied as

$$Z_{j,\alpha\beta}^{\text{corr},*} = Z_{j,\alpha\beta}^* - \frac{1}{N} \sum_{j'} Z_{j',\alpha\beta}^*, \quad (\text{C5})$$

where N donates the number of atoms in the cell.

With the original Z_j^* , RMSE is 0.099 ± 0.012 and R^2 is 0.911 ± 0.017 , while with the $Z_j^{\text{corr},*}$, RMSE is 0.097 ± 0.012 and R^2 is 0.914 ± 0.017 . Here, the mean values and standard deviations are evaluated over the five test sets of the fivefold cross validation used for BEC training, indicating that explicit enforcing of the acoustic sum rule has no measurable impact on the prediction accuracy of the ϵ^{ion} .

APPENDIX D: PREVIOUS STUDIES ON MACHINE-LEARNING PREDICTION OF DIELECTRIC CONSTANTS

Several studies have reported ML-based predictions of the electronic, ionic, and total dielectric constants. The dielectric constant is a second-rank tensor, and its spherical average is obtained by averaging the diagonal components. In evaluating the prediction accuracies of tensors, the eigenvalues are utilized for assessment. Here, previous studies are summarized, and the accuracy metrics are tabulated in Table IV.

- (i) Takahashi *et al.* [23] employed a random forest (RF) model with compositional and structural descriptors.

They predicted spherically averaged electronic and ionic dielectric constants ($\epsilon_{\text{ave}}^{\text{ele}}$ and $\epsilon_{\text{ave}}^{\text{ion}}$) across 1266 oxides.

- (ii) Morita *et al.* [27] reported ML models for predicting only $\epsilon_{\text{ave}}^{\text{ele}}$ in 1364 crystalline materials. They benchmarked a descriptor-based support vector regression (SVR) model against a crystal-graph MEGNet [150] model with transfer learning.
- (iii) Kim *et al.* [26] constructed a gradient boosting regression model for predicting dielectric properties using first-principles data of 7113 inorganic crystals curated in the MPD [61].
- (iv) Shimano *et al.* [25] employed RF and graph convolutional neural network (SchNet [29]) models to predict the dielectric constants of 2150 oxides derived from elemental substitutions in rutile, perovskite, Ruddlesden-Popper, and double-perovskite prototypes.
- (v) Hu *et al.* [24] predicted the averaged electronic and ionic dielectric constants of 722 binary and ternary oxides using SVR models incorporating detailed structural features. Although their prediction accuracy improves upon that of Takahashi *et al.* [23], this is likely due to the restriction of the compositional space to binary and ternary oxides.
- (vi) Mao *et al.* [18] applied an EGNN combined with transfer learning from an MLP, targeting 6648 inorganic crystals, whose dielectric tensor data were collected from the MPD. They limited the target crystals to those for which all components of the dielectric tensors are below 100. Only the prediction accuracies for the summed dielectric constants were reported.
- (vii) Lou *et al.* [28] proposed an EGNN (AnisoNet) to directly predict the ϵ^{ele} from crystal structures. The model was trained and evaluated on 6706 inorganic crystals registered in the MPD, with the dataset

TABLE IV. R^2 of the prediction of dielectric constants in previous studies and this study. There are cases where only spherically averaged quantities are predicted, as well as cases where eigenvalues are predicted. “...” indicate the absence of the data. “Scale” refers to whether the evaluation metrics are computed on a linear or base-10 logarithmic scale.

Study	Dataset	Model	Scale	$\epsilon_{\text{ave}}^{\text{ele}}$	$\epsilon_{\text{ave}}^{\text{ion}}$	ϵ_{ave}	$\lambda_{\epsilon^{\text{ele}}}$	$\lambda_{\epsilon^{\text{ion}}}$	λ_{ϵ}
Takahashi <i>et al.</i> [23]	1266 oxides	RF	log ₁₀	0.89	0.73	...	...	...	...
Morita <i>et al.</i> [27]	1364 inorganic crystals	SVR	Linear	0.86	...	...	...	...	...
		MEGNet	Linear	0.84	...	...	...	...	...
Kim <i>et al.</i> [26]	7113 inorganic crystals	GBR	log ₁₀	0.83	0.67	...	...	...	...
Shimano <i>et al.</i> [25]	2150 oxides in limited structural framework	RF	log ₁₀	0.92	0.89	...	...	...	...
		SchNet	log ₁₀	...	0.85	...	...	...	...
Hu <i>et al.</i> [24]	722 binary and ternary oxides	SVR	log ₁₀	0.94	0.79	0.89	...	...	...
Mao <i>et al.</i> [18]	6648 inorganic crystals	EGNN	log ₁₀	...	...	0.89	...	...	0.78
This study	928 oxides for ϵ^{ele} 839 oxides for ϵ^{ion} and ϵ	EGNN, Joint model	log ₁₀	0.74	0.91	0.91	0.71	0.91	0.91

restricted to entries for which all diagonal components (ε_{ii}) satisfy $\varepsilon_{ii} \geq 1$. In addition, entries with $\varepsilon_{ave}^{ele} \geq 15$ were excluded. They achieved MAE = 0.311 for ε_{ave}^{ele} in a base-10 logarithmic scale. As the R^2 was not reported, it is not included in Table IV.

APPENDIX E: FORMULATION AND PROPERTIES OF DYNAMICAL MATRIX

The harmonic contribution E to the potential energy of a system can be expressed using the force constant matrix $\Phi_{j\alpha,j'\beta}$ as follows:

$$E = \frac{1}{2} \sum_{j\alpha,j'\beta} u_{j\alpha} \Phi_{j\alpha,j'\beta} u_{j'\beta}, \quad (E1)$$

where $u_{j\alpha}$ is the displacement of atom j in the α direction. Introducing the mass-normalized displacement $\tilde{u}_{j\alpha} = \sqrt{M_j} u_{j\alpha}$, where M_j is the mass of atom j , Eq. (E1) can be reformulated as

$$E = \frac{1}{2} \sum_{j\alpha,j'\beta} \tilde{u}_{j\alpha} \left(\frac{1}{\sqrt{M_j}} \Phi_{j\alpha,j'\beta} \frac{1}{\sqrt{M_{j'}}} \right) \tilde{u}_{j'\beta}. \quad (E2)$$

Defining the reduced force constant matrix as $\tilde{\Phi}_{j\alpha,j'\beta} = ([1]/[\sqrt{M_j}]) \Phi_{j\alpha,j'\beta} ([1]/[\sqrt{M_{j'}}])$, the symmetric matrix $\tilde{\Phi}$ can be diagonalized according to

$$\tilde{\Phi} = \mathbf{Q} \mathbf{\Omega}^2 \mathbf{Q}^T, \quad (E3)$$

where $\mathbf{\Omega}^2 = \text{diag}(\dots, \omega_m^2, \dots)$, $\mathbf{Q}$ is an orthogonal matrix whose columns are the eigenvectors $\mathbf{e}_m$ corresponding to the phonon modes indexed by m .

The associated eigenvalue problem of Eq. (E3) is given explicitly by

$$\sum_{j'\beta} \tilde{\Phi}_{j\alpha,j'\beta} e_{m,j'\beta} = \omega_m^2 e_{m,j\alpha}. \quad (E4)$$

By the Bloch theorem, the eigenvector $e_{m,i\alpha}$ of a phonon mode with wave vector $\mathbf{q}$ take the form [35,151]

$$e_{m,j\alpha}(\mathbf{q}) = W_{m,j\alpha} \exp(i\mathbf{q} \cdot \mathbf{R}_j), \quad (E5)$$

where $W_{m,i\alpha}$ is a periodic function of the crystal, and $\mathbf{R}_j$ denotes the position of atom j . Substituting Eq. (E5) into Eq. (E4), multiplying both sides by $\exp(-i\mathbf{q} \cdot \mathbf{R}_j)$, yields

$$\sum_{j'\beta} D_{j\alpha,j'\beta}(\mathbf{q}) W_{m,j'\beta} = \omega_m^2 W_{m,j\alpha}, \quad (E6)$$

where the dynamical matrix $D_{j\alpha,j'\beta}(\mathbf{q})$ is defined as

$$\begin{aligned} D_{j\alpha,j'\beta}(\mathbf{q}) &= \exp(-i\mathbf{q} \cdot \mathbf{R}_j) \tilde{\Phi}_{j\alpha,j'\beta} \exp(i\mathbf{q} \cdot \mathbf{R}_{j'}) \\ &= \frac{1}{\sqrt{M_j M_{j'}}} \Phi_{j\alpha,j'\beta} \exp(i\mathbf{q} \cdot (\mathbf{R}_{j'} - \mathbf{R}_j)). \end{aligned} \quad (E7)$$

At the Γ point, i.e., when $\mathbf{q} = \mathbf{0}$, Eq. (E5) simplifies to

$$e_{m,j\alpha} = W_{m,j\alpha} \quad (E8)$$

And the dynamical matrix elements become

$$D_{j\alpha,j'\beta}(0) = \frac{1}{\sqrt{M_j M_{j'}}} \Phi_{j\alpha,j'\beta} = \tilde{\Phi}_{j\alpha,j'\beta}. \quad (E9)$$

Based on this formulation, it is possible to construct the force constant matrix from the potential energy E and atomic forces predicted by MLPs. Subsequently, the dynamical matrix can be computed, and through its diagonalization, phonon frequencies and eigendisplacement vectors are obtained.

-
- [1] R. P. Ortiz, A. Facchetti, and T. J. Marks, *High- κ Organic, inorganic, and hybrid dielectrics for low-voltage organic field-effect transistors*, *Chem. Rev.* **110**, 205 (2010).
 - [2] B. Wang, W. Huang, L. Chi, M. Al-Hashimi, T. J. Marks, and A. Facchetti, *High- κ gate dielectrics for emerging flexible and stretchable electronics*, *Chem. Rev.* **118**, 5690 (2018).
 - [3] A. I. Kingon, J.-P. Maria, and S. K. Streiffer, *Alternative dielectrics to silicon dioxide for memory and logic devices*, *Nature (London)* **406**, 1032 (2000).
 - [4] G. D. Wilk, R. M. Wallace, and J. M. Anthony, *High- κ gate dielectrics: Current status and materials properties considerations*, *J. Appl. Phys.* **89**, 5243 (2001).
 - [5] J. Robertson, *High dielectric constant oxides*, *Eur. Phys. J. Appl. Phys.* **28**, 265 (2004).
 - [6] C. H. Ahn, K. M. Rabe, and J.-M. Triscone, *Ferroelectricity at the nanoscale: Local polarization in oxide thin films and heterostructures*, *Science* **303**, 488 (2004).
 - [7] M. Esro, O. Kolosov, P. J. Jones, W. I. Milne, and G. Adamopoulos, *Structural and electrical characterization of SiO₂ gate dielectrics deposited from solutions at moderate temperatures in air*, *ACS Appl. Mater. Interfaces* **9**, 529 (2017).
 - [8] K. Eisenbeiser *et al.*, *Field effect transistors with SrTiO₃ gate dielectric on Si*, *Appl. Phys. Lett.* **76**, 1324 (2000).
 - [9] O. Trithaveesak, J. Schubert, and Ch. Buchal, *Ferroelectric properties of epitaxial BaTiO₃ thin films and heterostructures on different substrates*, *J. Appl. Phys.* **98** (2005).
 - [10] D. G. Schlom, L. Chen, X. Pan, A. Schmehl, and M. A. Zurbuchen, *A thin film approach to engineering functionality into oxides*, *J. Am. Ceram. Soc.* **91**, 2429 (2008).
 - [11] S. M. George, *Atomic layer deposition: An overview*, *Chem. Rev.* **110**, 111 (2010).
 - [12] J. Robertson, *High dielectric constant gate oxides for metal oxide Si transistors*, *Rep. Prog. Phys.* **69**, 327 (2006).

- [13] M. Balog, M. Schieber, M. Michman, and S. Patai, *Chemical vapor deposition and characterization of HfO₂ films from organo-hafnium compounds*, *Thin Solid Films* **41**, 247 (1977).
- [14] K. Yim, Y. Yong, J. Lee, K. Lee, H.-H. Nahm, J. Yoo, C. Lee, C. Seong Hwang, and S. Han, *Novel high- κ dielectrics for next-generation electronic devices screened by automated ab initio calculations*, *NPG Asia Mater.* **7**, e190 (2015).
- [15] I. Petousis, D. Mrdjenovich, E. Ballouz, M. Liu, D. Winston, W. Chen, T. Graf, T.D. Schladt, K. A. Persson, and F.B. Prinz, *High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials*, *Sci. Data* **4**, 160134 (2017).
- [16] A. Gopakumar, K. Pal, and C. Wolverton, *Identification of high-dielectric constant compounds from statistical design*, *npj Comput. Mater.* **8**, 146 (2022).
- [17] J. Riebesell, T. W. Surta, R. E. A. Goodall, M. W. Gaultois, and A. A. Lee, *Discovery of high-performance dielectric materials with machine-learning-guided search*, *Cell Rep. Phys. Sci.* **5**, 102241 (2024).
- [18] Z. Mao, W. Li, and J. Tan, *Dielectric tensor prediction for inorganic materials using latent information from preferred potential*, *npj Comput. Mater.* **10**, 265 (2024).
- [19] R. Ruff, P. Reiser, J. Stühmer, and P. Friederich, *Connectivity optimized nested line graph networks for crystal structures*, *Digital Discovery* **3**, 594 (2024).
- [20] S. S. Omee, S.-Y. Louis, N. Fu, L. Wei, S. Dey, R. Dong, Q. Li, and J. Hu, *Scalable deeper graph neural networks for high-performance materials property prediction*, *Patterns* **3**, 100491 (2022).
- [21] K. Choudhary and B. DeCost, *Atomistic line graph neural network for improved materials property predictions*, *npj Comput. Mater.* **7**, 185 (2021).
- [22] M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller, and F. Bechstedt, *Linear optical properties in the projector-augmented wave methodology*, *Phys. Rev. B* **73**, 045112 (2006).
- [23] A. Takahashi, Y. Kumagai, J. Miyamoto, Y. Mochizuki, and F. Oba, *Machine learning models for predicting the dielectric constants of oxides based on high-throughput first-principles calculations*, *Phys. Rev. Mater.* **4**, 103801 (2020).
- [24] Y. Hu *et al.*, *Accurate prediction of dielectric properties and bandgaps in materials with a machine learning approach*, *Appl. Phys. Lett.* **125** (2024).
- [25] Y. Shimano, A. Kutana, and R. Asahi, *Machine learning and atomistic origin of high dielectric permittivity in oxides*, *Sci. Rep.* **13** (2023).
- [26] E. Kim, J. Kim, and K. Min, *Prediction of dielectric constants of ABO₃-type perovskites using machine learning and first-principles calculations*, *Phys. Chem. Chem. Phys.* **24**, 7050 (2022).
- [27] K. Morita, D.W. Davies, K. T. Butler, and A. Walsh, *Modeling the dielectric constants of crystals using machine learning*, *J. Chem. Phys.* **153**, 024503 (2020).
- [28] Y. Lou and A. M. Ganose, *Discovery of highly anisotropic dielectric crystals with equivariant graph neural networks*, *Faraday Discuss.* **256**, 255 (2024).
- [29] K. T. Schütt, H.E. Sauceda, P.J. Kindermans, A. Tkatchenko, and K.R. Müller, *SchNet—A deep learning architecture for molecules and materials*, *J. Chem. Phys.* **148** (2018).
- [30] X. Gonze and C. Lee, *Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory*, *Phys. Rev. B* **55**, 10355 (1997).
- [31] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation*, 2nd ed. (Academic Press, New York, 1971), Vol. 3.
- [32] Ph. Ghosez, J.-P. Michenaud, and X. Gonze, *Dynamical atomic charges: The case of ABO₃ compounds*, *Phys. Rev. B* **58**, 6224 (1998).
- [33] R. Resta, *Macroscopic polarization in crystalline dielectrics: the geometric phase approach*, *Rev. Mod. Phys.* **66**, 899 (1994).
- [34] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/28wr-w896> for the distributions of the eigenvalues of the BEC tensors, and a systematic comparison of r²SCAN and dielectric-dependent hybrid functionals with experimental band-gap data, the prediction accuracies for dielectric constants, detailed information on the 31 identified high- κ oxides and the nine oxides included in the training dataset, which includes Refs. [35–42].
- [35] J. M. Ziman, *Electrons and Phonons* (Oxford University Press, Oxford, U.K., 2001).
- [36] S. H. Wemple, *Polarization fluctuations and the optical-absorption edge in BaTiO₃*, *Phys. Rev. B* **2**, 2679 (1970).
- [37] M. Balog, M. Schieber, M. Michman, and S. Patai, *The chemical vapour deposition and characterization of ZrO₂ films from organometallic compounds*, *Thin Solid Films* **47**, 109 (1977).
- [38] S. Maj, *Energy gap and density in SiO₂ polymorphs*, *Phys. Chem. Miner.* **15**, 271 (1988).
- [39] M. Batzill and U. Diebold, *The surface and materials science of tin oxide*, *Prog. Surf. Sci.* **79**, 47 (2005).
- [40] Y. Hinuma, Y. Kumagai, I. Tanaka, and F. Oba, *Effects of composition, crystal structure, and surface orientation on band alignment of divalent metal oxides: A first-principles study*, *Phys. Rev. Mater.* **2**, 124603 (2018).
- [41] Y. Hinuma, T. Gake, and F. Oba, *Band alignment at surfaces and heterointerfaces of Al₂O₃, Ga₂O₃, In₂O₃, and related group-III oxide polymorphs: A first-principles study*, *Phys. Rev. Mater.* **3**, 084605 (2019).
- [42] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, *Scaling deep learning for materials discovery*, *Nature (London)* **624**, 80 (2023).
- [43] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, *E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials*, *Nat. Commun.* **13**, 2453 (2022).
- [44] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, *Tensor field networks: Rotation and translation-equivariant neural networks for 3D point clouds*, *arXiv:1802.08219*.
- [45] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in 2016 IEEE Conference

- on *Computer Vision and Pattern Recognition (CVPR)* (IEEE, Amsterdam, 2016), pp. 770–778.
- [46] Y. Kumagai, N. Tsunoda, A. Takahashi, and F. Oba, *Insights into oxygen vacancies from high-throughput first-principles calculations*, *Phys. Rev. Mater.* **5**, 123803 (2021).
- [47] S. Falletta, A. Cepellotti, A. Johansson, C. W. Tan, M. L. Descoteaux, A. Musaelian, C. J. Owen, and B. Kozinsky, *Unified differentiable learning of electric response*, *Nat. Commun.* **16**, 4031 (2025).
- [48] B. Schmiedmayer and G. Kresse, *Derivative learning of tensorial quantities—predicting finite temperature infrared spectra from first principles*, *J. Chem. Phys.* **161** (2024).
- [49] K. Shimizu, R. Otsuka, M. Hara, E. Minamitani, and S. Watanabe, *Prediction of Born effective charges using neural network to study ion migration under electric fields: applications to crystalline and amorphous Li_3PO_4* , *Sci. Technol. Adv. Mater.* **3**, 2253135 (2023).
- [50] A. Kutana, K. Shimizu, S. Watanabe, and R. Asahi, *Representing Born effective charges with equivariant graph convolutional neural networks*, *Sci. Rep.* **15**, 16719 (2025).
- [51] Max. Born and Kun. Huang, *Dynamical Theory of Crystal Lattices* (Clarendon Press, Oxford, 1998).
- [52] X. Fu, B. M. Wood, L. Barroso-Luque, D. S. Levine, M. Gao, M. Dzamba, and C. L. Zitnick, *Learning smooth and expressive interatomic potentials for physical property prediction*, [arXiv:2502.12147](https://arxiv.org/abs/2502.12147).
- [53] B. Rhodes, S. Vandenhaute, V. Šimkus, J. Gin, J. Godwin, T. Duignan, and M. Neumann, *Orb-v3: atomistic simulation at scale*, [arXiv:2504.06231](https://arxiv.org/abs/2504.06231).
- [54] I. Batatia et al., *A foundation model for atomistic materials chemistry*, [arXiv:2401.00096](https://arxiv.org/abs/2401.00096).
- [55] J. Kim, J. Kim, J. Kim, J. Lee, Y. Park, Y. Kang, and S. Han, *Data-efficient multifidelity training for high-fidelity machine learning interatomic potentials*, *J. Am. Chem. Soc.* **147**, 1042 (2025).
- [56] A. Loew, D. Sun, H.-C. Wang, S. Botti, and M. A. L. Marques, *Universal machine learning interatomic potentials are ready for phonons*, *npj Comput. Mater.* **11**, 178 (2025).
- [57] Y. Park, J. Kim, S. Hwang, and S. Han, *Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations*, *J. Chem. Theory Comput.* **20**, 4857 (2024).
- [58] J. Riebesell, R. E. A. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A. A. Lee, A. Jain, and K. A. Persson, *A framework to evaluate machine learning crystal stability predictions*, *Nat. Mach. Intell.* **7**, 836 (2025).
- [59] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized gradient approximation made simple*, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [60] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, *Restoring the density-gradient expansion for exchange in solids and surfaces*, *Phys. Rev. Lett.* **100**, 136406 (2008).
- [61] A. Jain et al., *Commentary: The materials project: A materials genome approach to accelerating materials innovation*, *APL Mater.* **1**, 011002 (2013).
- [62] J. H. Skone, M. Govoni, and G. Galli, *Self-consistent hybrid functional for condensed systems*, *Phys. Rev. B* **89**, 195112 (2014).
- [63] P. B. Jamieson and S. C. Abrahams, *Re-examination of the crystal structure of ferroelectric tetragonal bronze-type $\text{Ba}_6\text{Ti}_2\text{Nb}_8\text{O}_{30}$* , *Acta Crystallogr. Sect. B* **24**, 984 (1968).
- [64] R. Someya, Y. J. Shan, and K. Tezuka, *Crystal structure and dielectric properties of $\text{Ba}_{6-x}\text{Ca}_x\text{Ti}_2\text{Nb}_8\text{O}_{30}$ tungsten bronze ceramics*, *J. Ceram. Soc. Jpn.* **131**, 236 (2023).
- [65] Y. Yuan, X. M. Chen, and Y. J. Wu, *Diffused ferroelectrics of $\text{Ba}_6\text{Ti}_2\text{Nb}_8\text{O}_{30}$ and $\text{Sr}_6\text{Ti}_2\text{Nb}_8\text{O}_{30}$ with filled tungsten-bronze structure*, *J. Appl. Phys.* **98** (2005).
- [66] G. L. Roberts, R. J. Cava, W. F. Peck, and J. J. Krajewski, *Dielectric properties of barium titanium niobates*, *J. Mater. Res.* **12**, 526 (1997).
- [67] Y. Itoh, S. Miyazawa, T. Yamada, and H. Iwasaki, *Dielectric property of $\text{Ba}_6\text{Ti}_2\text{Nb}_8\text{O}_{30}$ single crystal*, *Jpn. J. Appl. Phys.* **9**, 157 (1970).
- [68] K. Praveena and K. B. R. Varma, *Ferroelectric and optical properties of $\text{Ba}_5\text{Li}_2\text{Ti}_2\text{Nb}_8\text{O}_{30}$ ceramics potential for memory applications*, *J. Mater. Sci. Mater. Electron.* **25**, 3103 (2014).
- [69] K. E. Johnston, C. C. Tang, J. E. Parker, K. S. Knight, P. Lightfoot, and S. E. Ashbrook, *The polar phase of NaNbO_3 : A combined study by powder diffraction, solid-state nmr, and first-principles calculations*, *J. Am. Chem. Soc.* **132**, 8732 (2010).
- [70] H. Xu, Y. Su, M. Lou Balmer, and A. Navrotsky, *A new series of oxygen-deficient perovskites in the $\text{NaTi}_x\text{Nb}_{1-x}\text{O}_{3-0.5x}$ system: Synthesis, crystal chemistry, and energetics*, *Chem. Mater.* **15**, 1872 (2003).
- [71] A. C. Sakowski-Cowley, K. Łukaszewicz, and H. D. Megaw, *The structure of sodium niobate at room temperature, and the problem of reliability in pseudosymmetric structures*, *Acta Crystallogr. Sect. B* **25**, 851 (1969).
- [72] S. K. Mishra, N. Choudhury, S. L. Chaplot, P. S. R. Krishna, and R. Mittal, *Competing antiferroelectric and ferroelectric interactions in NaNbO_3 : Neutron diffraction and theoretical studies*, *Phys. Rev. B* **76**, 024110 (2007).
- [73] D. O. Mishchuk, O. I. V'yunov, O. V. Ovchar, and A. G. Belous, *Structural and dielectric properties of solid solutions of sodium niobate in lanthanum and neodymium niobates*, *Inorg. Mater. (USSR)* **40**, 1324 (2004).
- [74] Y. J. Shan, H. Mori, R. Wang, W. Luan, H. Imoto, Mitsuru Itoh, and T. Nakamura, *Powder X-ray diffraction study of ferroelectric phase transition in perovskite oxide CdTiO_3* , *Ferroelectrics* **259**, 85 (2001).
- [75] R. Caracas and R. E. Cohen, *Prediction of a new phase transition in Al_2O_3 at high pressures*, *Geophys. Res. Lett.* **32**, 1 (2005).
- [76] B. J. Kennedy, Q. Zhou, and M. Avdeev, *The ferroelectric phase of CdTiO_3 : A powder neutron diffraction study*, *J. Solid State Chem.* **184**, 2987 (2011).
- [77] S. Sasaki, C. T. Prewitt, J. D. Bass, and W. A. Schulze, *Orthorhombic perovskite CaTiO_3 and CdTiO_3 : structure and space group*, *Acta Crystallogr. Sect. C* **43**, 1668 (1987).
- [78] H. F. Kay and J. L. Miles, *The structure of cadmium titanate and sodium tantalate*, *Acta Crystallogr.* **10**, 213 (1957).

- [79] T. Sugai and M. Wada, *Single crystal growth and some properties of $\text{Cd}_2\text{Ti}_2\text{O}_3\text{F}_2$ and CdTiO_3* , *Jpn. J. Appl. Phys.* **18**, 1709 (1979).
- [80] A. Garbout, N. Kallel-Kchaou, and M. Férid, *Relationship between the structural characteristics and photoluminescent properties of $\text{LnEuTi}_2\text{O}_7$ ($\text{Ln} = \text{Gd}$ and Y) pyrochlores*, *J. Lumin.* **169**, 359 (2016).
- [81] L. Kong, Z. Zhang, M. de los Reyes, I. Karatchevtseva, G. R. Lumpkin, G. Triani, and R. D. Aughterson, *Soft chemical synthesis and structural characterization of $\text{Y}_2\text{Hf}_x\text{Ti}_{2-x}\text{O}_7$* , *Ceram. Int.* **41**, 5309 (2015).
- [82] F. Matteucci, G. Cruciani, M. Dondi, G. Baldi, and A. Barzanti, *Crystal structural and optical properties of Cr-doped $\text{Y}_2\text{Ti}_2\text{O}_7$ and $\text{Y}_2\text{Sn}_2\text{O}_7$ pyrochlores*, *Acta Mater.* **55**, 2229 (2007).
- [83] O. Knop, F. Brisse, and L. Castelliz, *Pyrochlores. V. Thermoanalytic, x-ray, neutron, infrared, and dielectric studies of $\text{A}_2\text{Ti}_2\text{O}_7$ titanates*, *Can. J. Chem.* **47**, 971 (1969).
- [84] B. J. Sun, Q. L. Liu, J. K. Liang, J. B. Li, L. N. Ji, J. Y. Zhang, Y. H. Liu, and G. H. Rao, *Subsolidus phase relations in the ternary system $\text{SnO}_2 - \text{TiO}_2 - \text{Y}_2\text{O}_3$* , *J. Alloys Compd.* **455**, 265 (2008).
- [85] M. Glerup, O. F. Nielsen, and F. W. Poulsen, *The structural transformation from the pyrochlore structure, $\text{A}_2\text{B}_2\text{O}_7$, to the fluorite structure, AO_2 , studied by raman spectroscopy and defect chemistry modeling*, *J. Solid State Chem.* **160**, 25 (2001).
- [86] J. M. Farmer, L. A. Boatner, B. C. Chakoumakos, M.-H. Du, M. J. Lance, C. J. Rawn, and J. C. Bryan, *Structural and crystal chemical properties of rare-earth titanate pyrochlores*, *J. Alloys Compd.* **605**, 63 (2014).
- [87] T. Ting-ting, W. Li-xi, and Z. Qi-tu, *Study on the composite and properties of $\text{Y}_2\text{O}_3 - \text{TiO}_2$ microwave dielectric ceramics*, *J. Alloys Compd.* **486**, 606 (2009).
- [88] B. Muktha, M. H. Priya, G. Madras, and T. N. Guru Row, *Synthesis, structure, and photocatalysis in a new structural variant of the Aurivillius phase: $\text{LiBi}_4\text{M}_3\text{O}_{14}$ ($\text{M} = \text{Nb}$, Ta)*, *J. Phys. Chem. B* **109**, 11442 (2005).
- [89] M. Bharathy, A. H. Fox, S. J. Mugavero, and H.-C. zur Loye, *Crystal growth of inter-lanthanide $\text{LaLn}'\text{O}_3$ ($\text{Ln}' = \text{Y}$, Ho-Lu) perovskites from hydroxide fluxes*, *Solid State Sci.* **11**, 651 (2009).
- [90] E. A. Zhurova, Y. Ivanov, V. Zavodnik, and V. Tsirelson, *Electron density and atomic displacements in KTaO_3* , *Acta Crystallogr. Sect. B* **56**, 594 (2000).
- [91] A. Tkach, P. M. Vilarinho, and A. Almeida, *Role of initial potassium excess on the properties of potassium tantalate ceramics*, *J. Eur. Ceram. Soc.* **31**, 2303 (2011).
- [92] O. Aktas, S. Crossley, M. A. Carpenter, and E. K. H. Salje, *Polar correlations and defect-induced ferroelectricity in cryogenic KTaO_3* , *Phys. Rev. B* **90**, 165309 (2014).
- [93] Y. Wu, *Synthesis of SrTiO_3 nanoparticles for photocatalytic applications*, *Highlights Sci. Eng. Technol.* **116**, 9 (2024).
- [94] W. Jauch and A. Palmer, *Anomalous zero-point motion in SrTiO_3 : Results from γ -ray diffraction*, *Phys. Rev. B* **60**, 2961 (1999).
- [95] S. Keshri, L. Joshi, and S. S. Rajput, *Studies on $\text{La}_{0.67}\text{Ca}_{0.33}\text{MnO}_3 - \text{SrTiO}_3$ composites using two-phase model*, *J. Alloys Compd.* **509**, 5796 (2011).
- [96] S. A. Howard, J. K. Yau, and H. U. Anderson, *Structural characteristics of $\text{Sr}_{1-x}\text{La}_x\text{Ti}_{3+\delta}$ as a function of oxygen partial pressure at 1400 °C*, *J. Appl. Phys.* **65**, 1492 (1989).
- [97] Yu. A. Abramov, V. G. Tsirelson, V. E. Zavodnik, S. A. Ivanov, and I. D. Brown, *The chemical bond, and atomic displacements in SrTiO_3 from x-ray diffraction analysis*, *Acta Crystallogr. Sect. B* **51**, 942 (1995).
- [98] J. M. Kiat and T. Roisnel, *Rietveld analysis of strontium titanate in the Müller state*, *J. Phys. Condens. Matter* **8**, 3471 (1996).
- [99] L. Fang, W. Dong, F. Zheng, and M. Shen, *Effects of Gd substitution on microstructures and low temperature dielectric relaxation behaviors of SrTiO_3 ceramics*, *J. Appl. Phys.* **112** (2012).
- [100] J. Brous, I. Fankuchen, and E. Banks, *Rare earth titanates with a perovskite structure*, *Acta Crystallogr.* **6**, 67 (1953).
- [101] M. Schmidbauer, A. Kwasniewski, and J. Schwarzkopf, *High-precision absolute lattice parameter determination of SrTiO_3 , DyScO_3 and NdGaO_3 single crystals*, *Acta Crystallogr. Sect. B* **68**, 8 (2012).
- [102] S. Parida, S. K. Rout, V. Subramanian, P. K. Barhai, N. Gupta, and V. R. Gupta, *Structural, microwave dielectric properties and dielectric resonator antenna studies of $\text{Sr}(\text{Zr}_x\text{Ti}_{1-x})\text{O}_3$ ceramics*, *J. Alloys Compd.* **528**, 126 (2012).
- [103] R. H. Mitchell, A. R. Chakhmouradian, and P. M. Woodward, *Crystal chemistry of perovskite-type compounds in the taunonite-loparite series, $(\text{Sr}_{1-2x}\text{Na}_x\text{La}_x)\text{TiO}_3$* , *Phys. Chem. Miner.* **27**, 583 (2000).
- [104] J. Hutton and R. J. Nelmes, *High-resolution studies of cubic perovskites by elastic neutron diffraction. II. SrTiO_3 , KMnF_3 , RbCaF_3 and CsPbCl_3* , *J. Phys. C* **14**, 1713 (1981).
- [105] K. Tsuda and M. Tanaka, *Refinement of crystal structure parameters using convergent-beam electron diffraction: the low-temperature phase of SrTiO_3* , *Acta Crystallogr. Sect. A* **51**, 7 (1995).
- [106] Z. Yang, D. Lee, J. Yue, J. Gabel, T.-L. Lee, R. D. James, S. A. Chambers, and B. Jalan, *Epitaxial SrTiO_3 films with dielectric constants exceeding 25,000*, *Proc. Natl. Acad. Sci. U.S.A.* **119** (2022).
- [107] M. Ahtee and L. Unonius, *The structure of NaTaO_3 by x-ray powder diffraction*, *Acta Crystallogr. A* **33**, 150 (1977).
- [108] M. Ahtee and C. N. W. Darlington, *Structures of NaTaO_3 by neutron powder diffraction*, *Acta Crystallogr. Sect. B* **36**, 1007 (1980).
- [109] S. W. Arulnesan, P. Kayser, B. J. Kennedy, and K. S. Knight, *The impact of room temperature polymorphism in K doped NaTaO_3 on structural phase transition behaviour*, *J. Solid State Chem.* **238**, 109 (2016).
- [110] B. J. Kennedy, A. K. Prodjosantoso, and C. J. Howard, *Powder neutron diffraction study of the high temperature phase transitions in NaTaO_3* , *J. Phys. Condens. Matter* **11**, 6319 (1999).
- [111] V. Shanker, S. L. Samal, G. K. Pradhan, C. Narayana, and A. K. Ganguli, *Nanocrystalline NaNbO_3 and NaTaO_3 : Rietveld studies, Raman spectroscopy and dielectric properties*, *Solid State Sci.* **11**, 562 (2009).

- [112] F. Brisse, D. J. Stewart, V. Seidl, and O. Knop, *Pyrochlores. VIII. Studies of some 2–5 pyrochlores and related compounds and minerals*, *Can. J. Chem.* **50**, 3648 (1972).
- [113] K. Łukaszewicz, A. Pietraszko, J. Stepień-Damm, and N. N. Kolpakova, *Temperature dependence of the crystal structure and dynamic disorder of cadmium in cadmium pyroniobates [Cd₂Nb₂O₇ and Cd₂Ta₂O₇]*, *Mater. Res. Bull.* **29**, 987 (1994).
- [114] S. N. Al-Refaie and S. A. Alboon, *On the analysis of Cd₂Nb₂O₇ dielectric dispersion*, *Phys. Procedia* **25**, 15 (2012).
- [115] N. Takani and H. Yamane, *Structure analysis of CaTi_{1-x}Sn_xO₃ (x = 0.0–1.0) solid solutions*, *Powder Diff.* **29**, 254 (2014).
- [116] X. Liu and Robert C. Liebermann, *X-ray powder diffraction study of CaTiO₃ perovskite at high temperatures*, *Phys. Chem. Miner.* **20** (1993).
- [117] M. Yashima and R. Ali, *Structural phase transition and octahedral tilting in the calcium titanate perovskite CaTiO₃*, *Solid State Ionics* **180**, 120 (2009).
- [118] A. Chandra and D. Pandey, *Evolution of crystallographic phases in the system (Pb_{1-x}Ca_x)TiO₃: A Rietveld study*, *J. Mater. Res.* **18**, 407 (2003).
- [119] M. Tsubota, F. Iga, K. Uchihira, T. Nakano, S. Kura, T. Takabatake, S. Kodama, H. Nakao, and Y. Murakami, *Coupling between orbital and lattice degrees of freedom in Y_{1-x}Ca_xTiO₃ (0 < x ≤ 0.75): A resonant X-ray scattering study*, *J. Phys. Soc. Jpn.* **74**, 3259 (2005).
- [120] K. S. Knight, *Structural and thermoelastic properties of CaTiO₃ perovskite between 7K and 400K*, *J. Alloys Compd.* **509**, 6337 (2011).
- [121] R. C. S. Costa, A. D. S. Bruno Costa, F. N. A. Freire, M. R. P. Santos, J. S. Almeida, R. S. T. M. Sohn, J. M. Sasaki, and A. S. B. Sombra, *Structural properties of CaTi_{1-x}(Nb_{2/3}Li_{1/3})_xO_{3-δ} (CNLTO) and CaTi_{1-x}(Nb_{1/2}Ln_{1/2})_xO₃ (Ln = Fe (CNFTO), Bi (CNBTO)), modified dielectric ceramics for microwave applications*, *Physica (Amsterdam)* **404B**, 1409 (2009).
- [122] S. Yoon *et al.*, *Improved photoluminescence and afterglow of CaTiO₃: Pr³⁺ by ammonia treatment*, *Opt. Mater. Express* **3**, 248 (2013).
- [123] V. Vashook, L. Vasylechko, M. Knapp, H. Ullmann, and U. Guth, *Lanthanum doped calcium titanates: synthesis, crystal structure, thermal expansion and transport properties*, *J. Alloys Compd.* **354**, 13 (2003).
- [124] R. Ali and M. Yashima, *Space group and crystal structure of the perovskite CaTiO₃ from 296 to 1720 K*, *J. Solid State Chem.* **178**, 2867 (2005).
- [125] K. S. Knight, *Parameterization of the crystal structures of centrosymmetric zone-boundary-tilted perovskites: An analysis in terms of symmetry-adapted basis-vectors of the cubic aristotype phase*, *Can. Mineral.* **47**, 381 (2009).
- [126] A. R. Chakhmouradian and R. H. Mitchell, *A structural study of the perovskite series CaTi_{1-2x}Fe_xNb_xO₃*, *J. Solid State Chem.* **138**, 272 (1998).
- [127] R. H. Buttner and E. N. Maslen, *Electron difference density and structural parameters in CaTiO₃*, *Acta Crystallogr. Sect. B* **48**, 644 (1992).
- [128] V. Vashook, D. Nitsche, L. Vasylechko, J. Rebello, J. Zosel, and U. Guth, *Solid state synthesis, structure and transport properties of compositions in the CaRu_{1-x}Ti_xO_{3-δ} system*, *J. Alloys Compd.* **485**, 73 (2009).
- [129] L. H. Oliveira, A. P. de Moura, T. M. Mazzo, M. A. Ramírez, L. S. Cavalcante, S. G. Antonio, W. Avansi, V. R. Mastelaro, E. Longo, and J. A. Varela, *Structural refinement and photoluminescence properties of irregular cube-like (Ca_{1-x}Cu_x)TiO₃ microcrystals synthesized by the microwave–hydrothermal method*, *Mater. Chem. Phys.* **136**, 130 (2012).
- [130] H. Yang, Y. Ohishi, K. Kurosaki, H. Muta, and S. Yamanaka, *Thermomechanical properties of calcium series perovskite-type oxides*, *J. Alloys Compd.* **504**, 201 (2010).
- [131] H. F. Kay and P. C. Bailey, *Structure and properties of CaTiO₃*, *Acta Crystallogr.* **10**, 219 (1957).
- [132] H. Y. Zhou, X. Q. Liu, X. L. Zhu, and X. M. Chen, *CaTiO₃ linear dielectric ceramics with greatly enhanced dielectric strength and energy storage density*, *J. Am. Ceram. Soc.* **101**, 1999 (2018).
- [133] J. E. F. S. Rodrigues, P. J. Castro, P. S. Pizani, W. R. Correr, and A. C. Hernandez, *Structural ordering and dielectric properties of Ba₃CaNb₂O₉-based microwave ceramics*, *Ceram. Int.* **42**, 18087 (2016).
- [134] J. E. Rodrigues, D. M. Bezerra, and A. C. Hernandez, *Ordering effect on the electrical properties of stoichiometric Ba₃CaNb₂O₉-based perovskite ceramics*, *Ceram. Int.* **43**, 14015 (2017).
- [135] A. P. Gorshkov, N. S. Volkova, D. G. Fukina, S. B. Levichev, and L. A. Istomin, *Impurity defect absorption and photochromic effect in KNbWO₆*, *J. Solid State Chem.* **298**, 122099 (2021).
- [136] S. Coh *et al.*, *Si-compatible candidates for high-κ dielectrics with the Pbnm perovskite structure*, *Phys. Rev. B* **82**, 064101 (2010).
- [137] N. A. Spaldin, *A beginner's guide to the modern theory of polarization*, *J. Solid State Chem.* **195**, 2 (2012).
- [138] M. A. Ghebouli, B. Ghebouli, A. Bouhemadou, M. Fatmi, and K. Bouamama, *Structural, electronic, optical and thermodynamic properties of Sr_xCa_{1-x}O, Ba_xSr_{1-x}O and Ba_xCa_{1-x}O alloys*, *J. Alloys Compd.* **509**, 1440 (2011).
- [139] I. Errea, M. Calandra, and F. Mauri, *First-principles theory of anharmonicity and the inverse isotope effect in superconducting palladium-hydride compounds*, *Phys. Rev. Lett.* **111**, 177002 (2013).
- [140] M. Sharma, R. Resta, and R. Car, *Intermolecular dynamical charge fluctuations in water: A signature of the H-bond network*, *Phys. Rev. Lett.* **95**, 187401 (2005).
- [141] A. Togo, *First-principles phonon calculations with PHONOPY and PHONO3PY*, *J. Phys. Soc. Jpn.* **92**, 012001 (2023).
- [142] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, *Quantum-chemical insights from deep tensor neural networks*, *Nat. Commun.* **8**, 13890 (2017).
- [143] M. Geiger and T. Smidt, *e3nn: Euclidean neural networks*, [arXiv:2207.09453](https://arxiv.org/abs/2207.09453).
- [144] T. Xie and J. C. Grossman, *Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties*, *Phys. Rev. Lett.* **120**, 145301 (2018).
- [145] S. Elfving, E. Uchibe, and K. Doya, *Sigmoid-weighted linear units for neural network function approximation in reinforcement learning*, *Neural Netw.* **107**, 3 (2018).

- [146] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [147] P. E. Blöchl, *Projector augmented-wave method*, [Phys. Rev. B **50**, 17953 \(1994\)](#).
- [148] G. Kresse and J. Furthmüller, *Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set*, [Phys. Rev. B **54**, 11169 \(1996\)](#).
- [149] https://github.com/kumagai-group/high-k_oxides_screening.
- [150] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, *Graph networks as a universal machine learning framework for molecules and crystals*, [Chem. Mater. **31**, 3564 \(2019\)](#).
- [151] D. C. Wallace, *Thermodynamics of Crystals*, Dover ed. (Dover Publications, New York, 1998).