





High-refractive-index materials screening from machine learning and *ab initio* methods

Pedro J. M. A. Carriço, Márcio Ferreira , Tiago F. T. Cerqueira, Fernando Nogueira , and Pedro Borlido *

CFisUC, Department of Physics, University of Coimbra, Rua Larga, 3004-516 Coimbra, Portugal

 (Received 5 June 2023; accepted 8 November 2023; published 3 January 2024)

In this study we analyze the dielectric properties of a recently published dataset to identify high-refractive-index and high-band-gap materials that are crucial for modern optoelectronic applications. We employ advanced crystal graph convolutional neural networks and density functional perturbation theory calculations to accelerate the discovery of such materials. Our analysis confirms the traditional inverse relationship between band gap and dielectric constant, which persists even in this large dataset. However, our study reveals several promising materials that possess competitive properties compared to current industry standards. Our findings provide valuable insights into the field of dielectric materials and demonstrate the potential of advanced machine learning and computational techniques for accelerating materials discovery.

DOI: [10.1103/PhysRevMaterials.8.015201](https://doi.org/10.1103/PhysRevMaterials.8.015201)

I. INTRODUCTION

Modern developments in machine learning models coupled to the increasing size of databases enable the screening of huge chemical composition spaces in a relatively short amount of time. Although we cannot yet train models for every property, either due to the lack of training data or the complicated nature of the properties, several success stories have appeared in recent years [1,2]. An extensive exploration of such works is well beyond the present scope, but we can nonetheless reference the extensive exploration of the convex hull [3,4], the identification of new super-hard materials [5,6], accelerating discovery of superconductors [7,8], and predicting band gaps of solids [9,10], among many others.

In this work we follow this approach and use high-throughput techniques for the identification of high-dielectric-constant materials. The dielectric constant is not only the most basic measure of light-matter interaction (thanks to its connection to the refractive index), but it is also a fundamental quantity in electronics (in part due to the importance of capacitive components). Because of this, high-throughput studies of the dielectric properties of materials can have a high technological impact. In particular, the industry interest resides with materials having large dielectric constants (as we will see in the following), but the interplay of different mechanisms makes the meaning of “high” different in different applications.

At the operating frequencies of optical devices, the most relevant contribution to the dielectric function comes from the response of the electrons to the applied field, a response which is translated into the electronic contribution to the dielectric tensor. This quantity bridges the electronic and optical properties, thanks to its relation to the refractive index, n [11]. Of particular interest to optical devices are materials with high refractive index (referred here as high- n materials), which see

intensive use in the design of waveguides, optical interference filters, mirrors, and antireflective coatings [12,13]. Cr_2O_3 , LiNbO_3 , TiO_2 , and SiO_2 are prototypical materials in such applications. High- n materials can also be good candidates for nonlinear optics. This comes from the empirical observation known as Miller’s rule, which states that the high-order response functions are correlated to the linear response function (and therefore, to some extent, to n) [14].

At the lower frequencies that characterize conventional electronics applications, the response due to the atomic lattice becomes particularly relevant, as often this ionic contribution to the dielectric response is much larger than the electronic one. Electronic applications, as optics, demand materials with large dielectric constants (frequently called high- κ dielectrics). SiO_2 is the textbook example of such materials, being used in the passivation of semiconductor surfaces and gate dielectrics in conventional field-effect transistors. However, its limitations in view of the continuous miniaturization of transistors lead to a rise in popularity of other materials such as HfO_2 , ZrO_2 , Al_2O_3 , and silicon oxynitrides [15–17]. Ceramics such as the SrTiO_3 and BaTiO_3 perovskite titanates are also high- k dielectrics commonly used in ceramic capacitors [18].

For the aforementioned applications, high-dielectric-constants are only one part of the equation. Most optical applications require materials that are transparent in the visible range, while for electronic applications there is the need to suppress both undesirable leakage currents, allow larger operating voltages, and reduce spurious charge generation from thermal and photoexcitation processes. On a fundamental level, such requirements mean that, in addition to the high-dielectric-constant, materials of interest for these applications must also have large band gaps. Unfortunately, the laws of physics present a significant obstacle to achieving our goal of maximizing both the band gap and the dielectric constant simultaneously. Empirically, it has long been observed that these two quantities follow a loose inverse proportionality relationship, a fact that can be understood using, for example,

*pedro.borlido@uc.pt

arguments based on perturbation theory [19]. Nonetheless, the technological importance of discovering new suitable high-dielectric materials is a strong motivator for the community, who plays a constant game of “minimax” in this search. Thanks to the ever-growing size of material databases there is still ample ground to investigate. That is precisely the focus of the present work.

Here we search for materials with large band gaps and high-dielectric-constant, focusing on a recently developed dataset obtained by an exhaustive exploration of ternary compounds by Schmidt *et al.* [3]. This database contains circa 3 million entries (circa 330 thousand within 100 meV/atom of the convex hull), a majority of which was previously unreported in the literature.

To allow for such analysis, we rely on efficient machine learning models to screen the dielectric constants, followed by *ab initio* electronic structure calculations to validate and study the best candidates we found.

II. METHODS

A. DFT *ab initio* calculations

Ab initio calculations were performed using density functional theory [20,21] and density functional perturbation theory [22,23] as implemented in the Vienna Ab Initio Simulation Package (VASP, version 6.3) [24,25]. This code is integrated with the ATOMATE [26] package, which allowed us to automate the calculation of most properties within so-called *workflows* [27,28]. All calculations were performed using the Perdew-Burke-Ernzerhof (PBE) [29,30] approximation to the exchange-correlation functional. This functional is known to underestimate the band gaps of materials, and although this effect is not systematic, we can, on average, disregard the associated spread in the error, especially when focusing on larger gaps (PBE has a mean absolute percentage error of 46% and a mean percentage error of -41% for the band gaps [31,32]). Furthermore, in our analysis we compare with well-established materials which function as standard candles, since both their PBE and experimental gaps are well established. In the future this could be corrected by using available machine learning models [33] or resorting to more expensive hybrid functionals and GW calculations.

For the characterization of the dielectric response, we take into consideration both the ionic and the electronic contributions (ϵ^0 and ϵ^∞ , respectively) [34], such that in the static regime we can write

$$\epsilon_{ij} = \epsilon_{ij}^0 + \epsilon_{ij}^\infty, \quad (1)$$

where i, j are indices defining a Cartesian direction. The high-throughput calculations of optimized geometries, band structures, and static dielectric constants were performed using the defaults of ATOMATE, with the sole exception of increasing the kinetic energy cutoff to 600 eV [35]. For some selected materials we repeated these calculations with slightly tighter parameters: geometry optimizations were performed with a k -point density of 6000 k points per reciprocal atom (kppra) [36] until forces were smaller than 1 meV/Å, followed by density functional perturbation theory (DFPT) calculations with a density of 8000 kppra. The kinetic energy cutoff was kept the same as before.

The directional dependency of tensors makes them unpractical quantities for large-scale comparison and analysis. Instead, it is common to put the discussion in terms of polycrystalline averages, which are bounded by [37]

$$\frac{3}{1/\lambda_1 + 1/\lambda_2 + 1/\lambda_3} < \epsilon_{\text{poly}} < \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}, \quad (2)$$

where λ_i are the eigenvalues of ϵ_{ij} . Usually this translates into a sufficiently small range so that the equality

$$\epsilon_{\text{poly}} \equiv \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (3)$$

suffices. The static refractive index is obtained from the electronic contribution to the static dielectric tensor

$$n_{ii} = \sqrt{\epsilon_{ii}^\infty}, \quad (4)$$

and a static polycrystalline refractive index is defined as $n_{\text{poly}} = \sqrt{\epsilon_{\text{poly}}^\infty}$. For the sake of simplicity, in the following we drop the “poly” subscript whenever unambiguous.

For a small set of materials, we further computed the electronic contribution to the frequency-dependent dielectric function, within the independent particle approximation and without considering local field effects (see, e.g., Ref. [38]),

$$\begin{aligned} \epsilon_{ii}(\omega) = & 1 - \frac{e^2 \hbar^2}{\epsilon_0 m_e^2 \Omega_0} \sum_{\nu\nu'\mathbf{k}} w_{\mathbf{k}} \frac{|\langle \nu\mathbf{k} | p_i | \nu'\mathbf{k} \rangle|^2}{(\epsilon_{\nu\mathbf{k}} - \epsilon_{\nu'\mathbf{k}})^2} \\ & \times \frac{f(\epsilon_{\nu'\mathbf{k}}) - f(\epsilon_{\nu\mathbf{k}})}{\epsilon_{\nu'\mathbf{k}} - \epsilon_{\nu\mathbf{k}} + \hbar(\omega + i\gamma)}, \end{aligned} \quad (5)$$

where $\epsilon_{\nu\mathbf{k}}$ are the eigenvalues of eigenstate $|\nu\mathbf{k}\rangle$, $f(\epsilon_{\nu\mathbf{k}})$ are the corresponding occupation numbers, p_i are the Cartesian components of the momentum operator, $w_{\mathbf{k}}$ is the weight of k -point \mathbf{k} , Ω_0 is the unit-cell volume, and γ is the value for the Lorentzian broadening. For these calculations we used a constant value of 8000 k points per reciprocal atom, an energy cutoff of 600 eV, $\gamma = 0.10$ eV, and the number of bands was chosen as ten times the number of electrons in the unit cell (although for some larger systems this was reduced to 400 bands). This way we avoid expensive convergence calculations while attaining good convergence for the transitions near the band gap.

B. ML prediction

In order to screen the dataset, we use crystal graph convolutional neural networks (CGCNN) [39], which have been shown to perform well across multiple datasets [40] and have the benefit of only using the crystal structure as an input. The dataset for training was generated from the entries of Materials Project [41]. From all the entries containing dielectric properties, we selected only materials with band gaps larger than 0.3 eV, fewer than 20 atoms in the unit cell, and refractive index larger than 1 (as to minimize possible spurious entries). This culminated in a set of 5035 structures and corresponding dielectric tensors out of 7178 entries in Materials Project (at the date of writing). This set was randomly divided into training, test, and validation sets, with 60% /20% /20% size ratios. The training process was allowed to run through 400 epochs but was stopped when the loss function reached a minimum value, L_{min} , and remained in the interval $[L_{\text{min}} -$

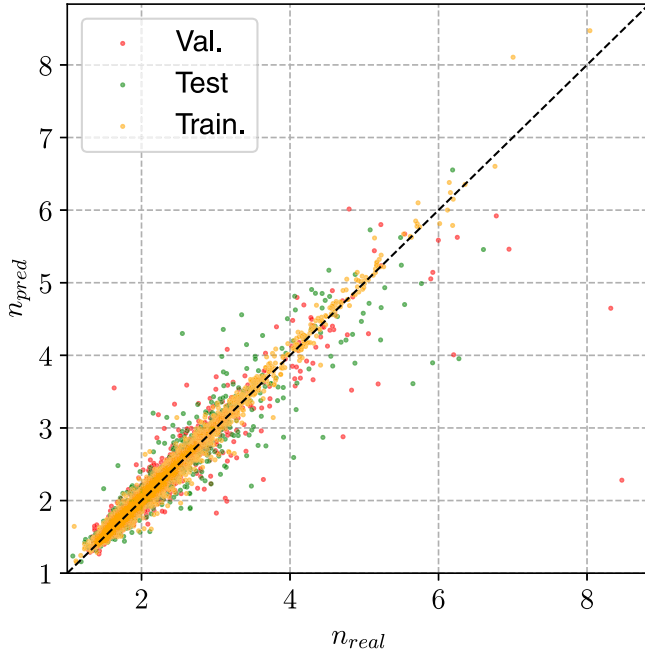


FIG. 1. Scatter plot of the predicted (n_{pred}) vs the calculated (n_{real}) refractive indices, trained on n .

0.001, $L_{\text{min}} + 0.001$] after 20 epochs. We used the stochastic gradient descent (SGD) optimizer [42] with an initial learning rate of 0.02, which was decreased by a factor of 10 every 100 epochs.

C. Target dataset

Our target dataset comes from Ref. [3] and was generated using crystal-graph attention networks to comb through several thousand crystal prototypes in order to study a wide composition space. The dataset is primarily comprised of binaries and ternaries, along with a smaller number of quaternary compounds, and provides an homogeneous sampling of the periodic table [3].

III. RESULTS AND DISCUSSION

A. Refractive index

The first step of the work is the training of the CGCNN. For this we followed two approaches, one by training the model directly on n and another by training on $\log_{10}(n)$, which should reduce problems due to the dispersion of larger values. As it turns out, both training procedures delivered very similar metrics, with the model trained on $\log(n)$ somewhat outperforming the other one. However, given the residual gains, we opted for the simplest option and trained the model directly on n .

The dispersion plot of the results is presented in Fig. 1 for the values in the training, test, and validation sets. A more detailed analysis of the model's performance (including mean errors, Kendall's rank correlation [43,44], quantile analysis, etc.) is given in Sec. S2 of the Supplemental Material (SM) [45].

As can be seen, a good agreement between the calculated and predicted values was achieved. The mean absolute error (mean absolute percentage error) in both validation and test sets is 0.15 (around 6%), which is reasonable given the range of values n takes. The entries with largest absolute errors are MnIn_2O_4 ($n = 8.5$, $\Delta = -6.2$), Sr_2SnHg ($n = 8.3$, $\Delta = -3.7$), Ge ($n = 6.3$, $\Delta = -2.3$), Ca_2AsAu ($n = 6.2$, $\Delta = -2.2$), ThNiSn ($n = 5.7$, $\Delta = -2.0$), and Na_2TlSb ($n = 5.9$, $\Delta = -2.0$). These outliers tend to be materials with very small gap/very high refractive index, and given the distribution of errors, constitute a small percentage of the entire dataset.

Using the trained model, we predicted the static refractive index for the dataset of Ref. [3]. Due to the sheer size of the set and given the goals of this work, we restricted the entries of interest to those that have nonmagnetic ground states, contain no noble gases, have gaps above 0.4 eV, and distances to the hull below 100 meV/atom. This amounts to approximately 55 000 entries.

Before going into a deeper analysis of the results, we should note that two sources of error are present in these plots. First, the values of the band gap in the screened dataset were obtained from self-consistent calculations on relatively loose k -point meshes, with 1000 k kppra. The indicated values are thus upper bounds for the real ones (this being the main reason for the chosen band-gap cutoff), and some metals might be wrongly included in the set, in spite of the precautions taken in their preparation. The second source of error comes from the uncertainty associated with the model used for the prediction, although from the previous analysis this should not bias the distribution too much. It will be seen that these factors do not compromise the end results.

As a point of comparison, we also plot the distribution of values from Naccarato *et al.* [46], which was the subject of a very thorough analysis by its authors and is representative of the general distribution. Although this dataset contains fewer entries than the one used to train the CGCNN model, it does contain information on the direct band gaps of the materials. This information is not available for every entry on the Materials Project, and it will be relevant for the selection of materials.

The resulting distribution of predicted refractive indices is presented in Fig. 2(a). As expected, no material with simultaneously large refractive index and direct gap is observed, and instead the Pareto front for (n, E_g^d) follows a loose relation of inverse proportionality. Comparing with the dataset of Ref. [46], the most notable difference is the high-density region seen below 2 eV and up to almost $n = 6$. A thin stripe of materials above the known distribution is also visible. In order to perform a more detailed analysis of these results, we selected a subset of materials close to the Pareto front. To do so we resorted to the relationship between n and the direct band gap, E_g^d , described in Ref. [46]:

$$n_s^2 = 1 + \left(\frac{\omega_{\text{eff}}}{E_g^d + \alpha + \beta/E_g^d} \right)^3, \quad (6)$$

which allows us to parametrize different regions of the diagram as a function of the parameter ω_{eff} . Using this equation, we selected the value $\omega_{\text{eff}} = 16$ eV for this purpose (see

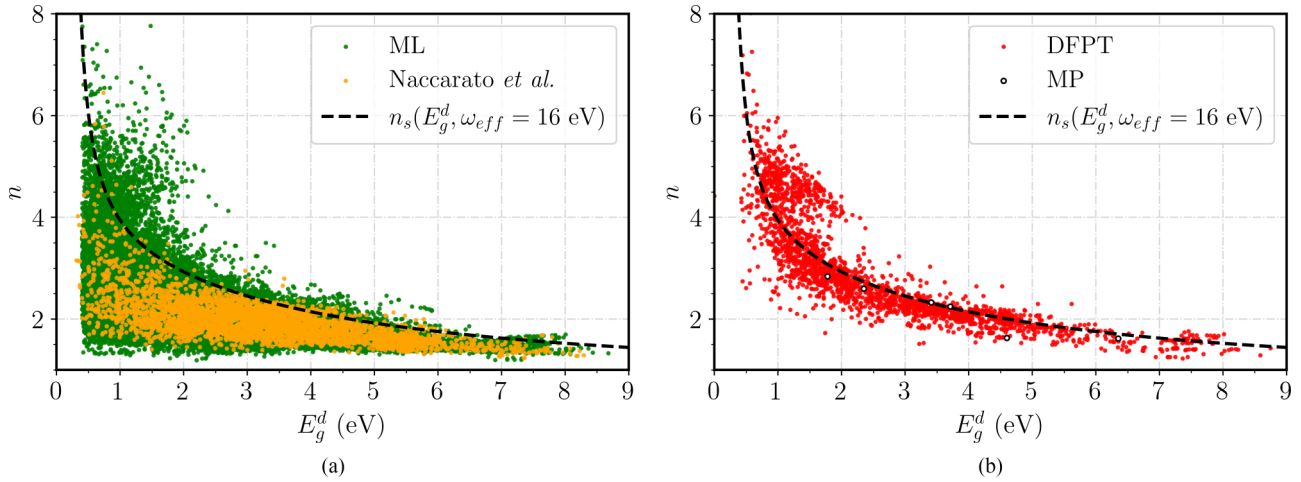


FIG. 2. Left: Dispersion plots of predicted (green) and calculated (from Ref. [46], orange) refractive indices as a function of direct band gap. Right: Dispersion plots of calculated n and E_g^d , selected from the predicted dataset (DFPT). Also indicated are some values for reference materials with corresponding data obtained from Materials Project (MP). In both plots the dashed curve corresponds to Eq. (6) with $\omega_{\text{eff}} = 16$ eV, i.e., the guideline used to separate materials of interest.

Fig. 2), but it should be noted this was done on the basis of visual inspection and considering available computational resources. To avoid redundancy we further focused primarily in materials for which no corresponding entry was found in other databases [41,47,48].

Using the aforementioned criteria, we calculated the ionic and electronic contributions to the dielectric constant using DFPT, as described in the Methods section. As anticipated, several metals had erroneously been identified as semiconductors due to the original k -point sampling, which lead to spurious values of n . In addition, we found that some materials containing Lu lead to strange results and were therefore removed from the analysis. Further neglecting runs that failed to converge, we were left with 1991 entries in the set.

Compared with the original dataset used to train this model, this amounts to an increase of circa 40% in the size. Therefore we decided to do a second round of training including these additional entries. We see a small improvement in the error description (e.g., the MAE in validation and test was reduced to 0.13 and 0.14, respectively), but no large qualitative difference is seen in the distribution. After this retraining, we repeated the previous selection procedure, resulting in a total of 2431 new entries, which we now subject to a more detailed analysis. The resulting distribution is presented in Fig. 2(b).

The error associated with the prediction of the machine learning model, as well as the inaccuracy on the screened band gaps, means that the actual distribution is not exactly as the predicted one. Specifically, a spreading of the data points in the distribution that brings some of the materials below the $\omega_{\text{eff}} = 16$ eV curve is observed. Nonetheless, the calculated distribution correlates well with the predicted one, and the two most prominent features (the hyperbole and the cluster above $n = 4$) remain.

To try and gain more insight on this distribution, we plotted the n vs E_g^d distribution for the subsets of materials containing at least one lanthanide, at least one actinide, at least one alkali, etc., in each material (see Fig. S3 of the SM). Most of the classes present similar looking distributions, which in turn

reflect the global one. The exception are the materials with an alkali and a halogen, for which the cluster is less represented. Although this is not completely unexpected, as materials with these elements tend to have larger band gaps, it is hard to extract detailed information from this observation.

Taking a look at the elemental distribution in Fig. 3 (and Table SV of the SM), some trends become more apparent. As can be seen, oxygen is significantly over-represented (although only present in 664 entries), followed by Te (347), Li (341), S (294), and Se (225). The frequency distribution falls relatively fast, with only 24 elements having more than 100 entries.

Besides O, Te, and Li, either every element is more or less distributed without significative clustering or is not present in a statistically representative number. As such, we will simply focus on the top three most common elements in the dataset. This is not to say that we assign special meaning to these elements as the source of the corresponding combination of n and E_g^d (which likely arises due to a combination of many effects) but rather as a simple categorization tool.

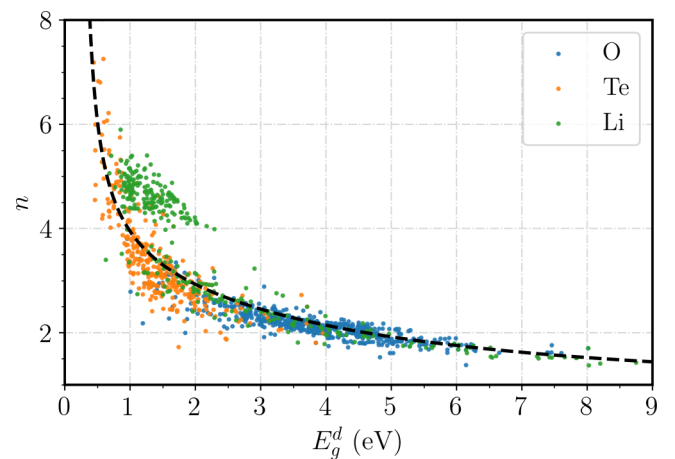


FIG. 3. Scatter plot of n vs E_g^d for materials containing the three most frequent elements of the dataset: O, Li, and Te.

Oxygen-containing materials are primarily located in the neighborhood of the ω_s curve in the direct gap range of 3–5 eV. As shown in Table SVI of the SM, the elements most associated with O in this set are rare earths (Tb, Pr, Dy, Nd, Er, etc.), meaning that this cluster is mostly comprised of rare-earth oxides. Due to their f -shell nature, these tend to be numerically complicated to deal with, but their relation between n and E_g^d is in line with previous observations [46]. Although the electronic structure of lanthanide oxides can present a large variety of behaviors [49], in general these tend to be wide gap systems. In addition, due to the localized nature of oxygen's p orbitals, these compounds tend to have flat valence bands (i.e., high hole effective masses) [50,51] and the associated high density of states contributes to a large number of transitions for the refractive index. The combination of these two factors mostly justifies the location of this group of materials, with the large chemical variety of the (mostly ternary; see Fig. S6 of the SM) systems generating a large spread.

Materials containing the second most represented element, tellurium, are also localized around the ω_s curve but at a lower range of energies, specifically, for E_g^d from 1 to 2 eV. By contrast with oxygen, the elements most associated with tellurium are post-transition elements (Bi, Tl, Pb, As, Sn), as well as Na and Ba.

The third most common element in the set is lithium, associated with which we find mostly group-IV elements (Si, Ge, Sn) but also some transition metals (Ni, Pt, Pd, Rh, etc.). Materials containing this element can be found both amid the O and Te regions and located above the ω_s curve. The predominant distribution is in fact the latter, located in the range of E_g^d from 1 to 2 eV and n above 4 (177 out of the 274 entries with Li lie above the $n = 4$ line). As it turns out, this region is populated almost in its entirety by members of the quaternary Heusler prototype, with chemical formula LiXYZ . This peculiar set of materials has already been extensively studied in Ref. [52], where it was observed that under some conditions the inclusion of lithium in the half-filled sublattice of the corresponding metallic half-Heusler leads to a filled valence band (thus inducing a band gap in the material) and large ϵ^∞ , making them stand out in the (E_g^d, n) distribution.

B. Specific entries

In the following we will be focusing on some of the entries that lie in the neighborhood of known high-dielectric materials of interest. A nonexhaustive list of such materials includes LiNbO_3 (mp-3731), rutile TiO_2 (mp-2657), anatase TiO_2 (mp-390), and HfO_2 (mp-352), which we take as reference in Fig. 2(b). As expected, they lie close to the boundary previously defined using Eq. (6). From this we selected a small set of materials (presented in Table I) for which the frequency-dependent dielectric tensor (within the independent particle approximation) was computed. For each entry we searched the literature to avoid going into details of compounds already studied but not in the databases.

I. CaMgTe_2

CaMgTe_2 belongs to the trigonal $R\bar{3}m$ space group, crystallizing in the same Caswell silverite of NaCrS_2 . In essence,

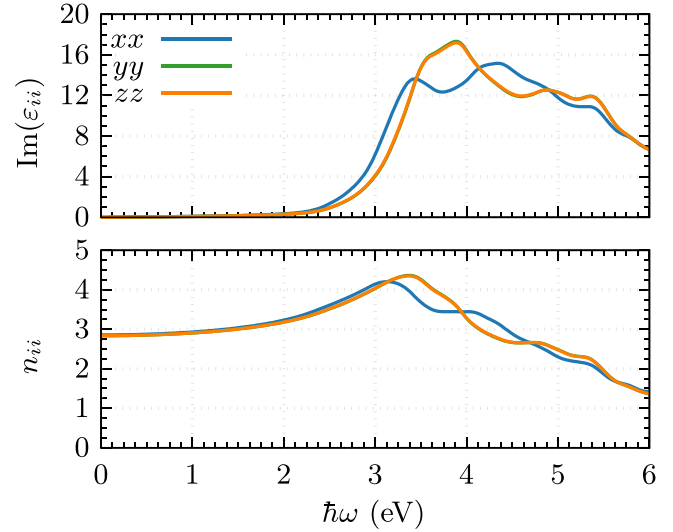


FIG. 4. Frequency-dependent imaginary part of the dielectric function (top) and refractive index (bottom) of CaMgTe_2 .

the structure can be represented as a simple cubic lattice which, when viewed from the (001) direction, has alternating rows (Te-Ca-Te-Mg) along the square diagonals. Due to the different sizes of Ca and Mg, the bonds with Te are distorted accordingly, creating a pattern of pincushion/barrel distortions throughout the solid.

On the electronic side, this material's valence band is characterized by an almost parabolic peak at Γ and made up of p -orbitals of Te. The conduction band is more ridged, with several minima of comparable energy appearing and predominant contributions of Ca d and Mg s orbitals. Given this electronic landscape, the fundamental gap is indirect (due to a $\Gamma - F$ transition), with a PBE value of 1.63 eV. On the other hand, the direct $\Gamma - \Gamma$ gap is substantially larger, at 2.25 eV. Taking the optical gap at face value, $n = 2.7$ and $E_g^d = 2.25$ eV makes CaMgTe_2 a direct competitor with anatase TiO_2 (which for comparison has slightly lower n at 2.6 and slightly larger E_g^d at 2.35 eV). On the other hand, for cases where the fundamental gap is more relevant, the direct competition would be with the rutile TiO_2 phase, which, being a direct gap material with $n = 2.8$ and $E_g^d = 2.35$ eV, is the superior option. The imaginary part of the dielectric function of CaMgTe_2 (see Fig. 4) is dominated by the peak at 4–5 eV, proceeding from the transitions from the valence (at around 24 eV another peak appears due to transitions from deep p states of Ca). This behavior makes the refractive index cover the range of 3 to around 4 within the frequency range of the visible spectra.

As a final comment, the MgTe chalcogenide has already been pointed to as a material with large refractive index [53], but nonetheless, it is more often found as an end component on studies of ternary alloys. One such use case is $\text{Cd}_x\text{Mg}_{1-x}\text{Te}$ in monolithically stacked CdTe -based solar cells [54,55], while $\text{Zn}_x\text{Mg}_{1-x}\text{Te}$ can be used in the production of waveguides [56,57]. Although we could not find reports on the use of $\text{Ca}_x\text{Mg}_{1-x}\text{Te}$, its optical properties suggest it could find use in similar areas as these two examples.

TABLE I. Chemical formula, space group number (Spg.), polycrystalline-averaged static ionic and electronic contributions to the dielectric constant (ϵ^∞ and ϵ^0), polycrystalline-averaged static refractive index (n), fundamental and direct band gap (E_g and E_g^d , in eV) and their difference (Δ), and distance to the convex hull (E_{hull} , in meV/atom) for some materials calculated with the tighter convergence criteria. HfZrO₄ is marked with a † symbol, because in the two different optimizations performed it converged to slightly distorted structures with significantly different ϵ^∞ , as discussed in the main text.

Formula	Spg.	ϵ^0	ϵ^∞	n	E_g	E_g^d	Δ	E_{hull}
KTe ₂ As	166	40.3	15.7	4.0	1.1	1.4	0.3	52
LiTiCoGe	216	15.7	22.2	4.7	1.2	1.6	0.4	0
Ca(AlAs) ₂	166	5.7	11.4	3.4	0.9	1.7	0.8	27
TiGePt	216	5.0	19.8	4.5	0.9	1.7	0.8	0
LiTaSiOs	216	15.7	20.8	4.6	0.5	1.7	1.2	0
LiBeAs	186	5.3	10.8	3.3	1.4	1.7	0.4	25
ZrHN	216	47.1	12.1	3.5	1.6	1.7	0.1	48
LiHfInPt	216	7.8	17.7	4.2	1.0	1.8	0.8	0
NdWN ₃	160	215.7	13.0	3.6	0.6	1.8	1.2	54
AsIrSe	29	4.4	17.4	4.2	1.2	1.8	0.6	0
HfSnPt	216	4.0	16.4	4.1	0.9	1.8	0.9	0
LiZrGeIr	216	11.4	17.8	4.2	1.5	2.0	0.5	0
BiSF	62	65.4	8.1	2.8	2.0	2.1	0.1	20
Ba ₂ NbTiO ₆	225	62.7	5.7	2.4	2.1	2.2	0.1	6
CaMgTe ₂	166	10.0	7.5	2.7	1.6	2.3	0.6	48
HfHN	216	28.3	10.0	3.2	2.0	2.3	0.2	0
CaTe ₂ Pb	166	20.7	10.5	3.2	2.2	2.3	0.1	22
LiHfSiIr	216	11.2	15.9	4.0	1.7	2.3	0.6	16
PfIrS	29	3.1	14.4	3.8	1.7	2.4	0.7	0
Na ₄ HfTe ₄	121	2.5	4.0	2.0	2.2	2.4	0.2	59
ZrZnN ₂	122	6.6	7.4	2.7	2.1	2.4	0.3	65
ScTaO ₄	14	73.0	5.3	2.3	2.6	2.6	0.0	21
NaTaO ₃	161	124.4	5.2	2.3	2.7	2.7	0.0	1
N _a 3TeH	221	8.6	4.8	2.2	2.6	2.8	0.2	0
LiAlC	216	5.6	7.4	2.7	0.8	3.3	2.5	26
TaBiO ₄	30	31.7	5.9	2.4	3.3	3.3	0.0	43
BeZnS ₂	33	2.9	5.6	2.4	3.4	3.4	0.0	16
KTaO ₃	227	63.9	4.4	2.1	3.3	3.5	0.3	23
Li ₃ TeH	221	11.6	6.3	2.5	2.9	3.7	0.8	12
HfZrO ₄ †	50	146.6	5.3	2.3	3.7	3.8	0.1	67
NaBO ₂	122	5.9	2.8	1.7	6.1	6.3	0.2	96
Be ₂ P2O ₇	15	3.5	2.7	1.7	6.3	6.3	0.0	32

2. ZrZnN₂

ZrZnN₂ is a ternary nitride that presents several polymorphs [58–60], although at the present is very underrepresented in the Materials Project. Most of these compounds are metastable, but due to an interplay of effects such as disorder entropy, some can be synthesized by thin-film sputtering [60]. Given that some of these present large piezoelectric constants [59], there is presently some incentive towards researching them.

The present polymorph is a chalcopyrite (space group $I\bar{4}2d$), a family of compounds perhaps best known due to CIGS [61], and at 65 meV/atom from the hull it is also metastable. Although it presents a fundamental indirect gap of 2.09 eV, the valence band is rather flat, making the direct transition only slightly larger at 2.42 eV. While the valency shows major contributions from N's p orbitals (and to some extent Zn's d), the lower conduction bands show predominantly contributions from Zr's d orbitals.

The imaginary part of the dielectric function (see Fig. 5) rises slowly owing to the low density of states of the lower valency and is dominated by a peak at 4–5 eV. This feature is associated with a peak in the refractive index, which brings n from its static value of 2.73 to above 4 in this region (depending on the polarization direction). Overall, these values put ZrZnN₂ in the neighborhood of interest of anatase TiO₂.

3. BeZnS₂

BeZnS₂ is an ordered alloy of wurtzite-BeS and wurtzite-ZnS. ZnS is probably best known for its cubic phase structure, a wide gap material used in optical windows, among other proposed applications. The high-temperature wurtzite phase is rarer, but it can be synthesized and is both predicted [62] and observed [63,64] to have a slightly larger band gap than its cubic counterpart. It also has a rather large excitonic binding energy of around 40 meV [64]. Within the PBE approximation, wurtzite-ZnS (mp-560588) has a band gap of 2.08 eV and a static refractive index of 2.4 [41]. As for BeS, we could

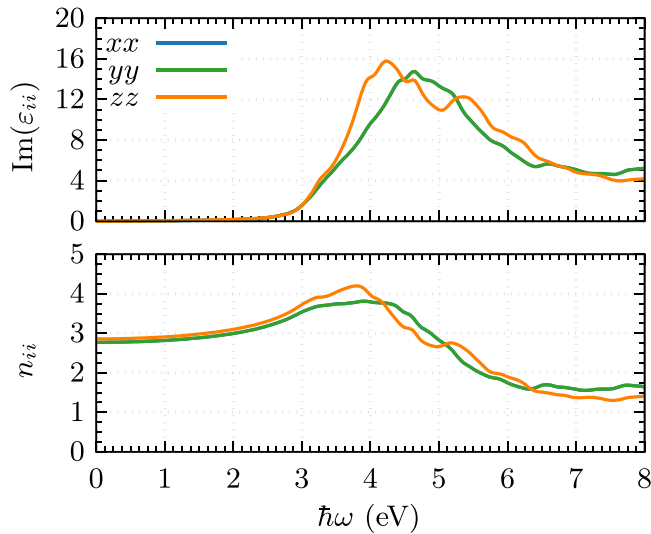


FIG. 5. Frequency-dependent imaginary part of the dielectric function (top) and refractive index (bottom) of ZrZnN_2 .

find little information on the material itself, much less on practical uses, apparently being outshined by its chalcogen cousin BeO . Wurtzite BeS is predicted to have a PBE band gap of 3.70 eV and static refractive index of 2.3 [3]. The BeZnS_2 entry in our dataset presents a PBE band gap of 3.38 eV (higher than would be expected from Vegard's law) and a static refractive index of 2.4 (essentially the same value as its end components), very close to those of LiNbO_3 . BeZnS_2 starts absorbing very close to the band gap, and we do not see a strong effect due to the anisotropy in the bonding directions (see Fig. 6). These results place wurtzite $\text{Zn}_x\text{Be}_{1-x}\text{S}$ as a material of interest, suggesting that one could tune the band gap of the alloy while keeping the refractive index in a small

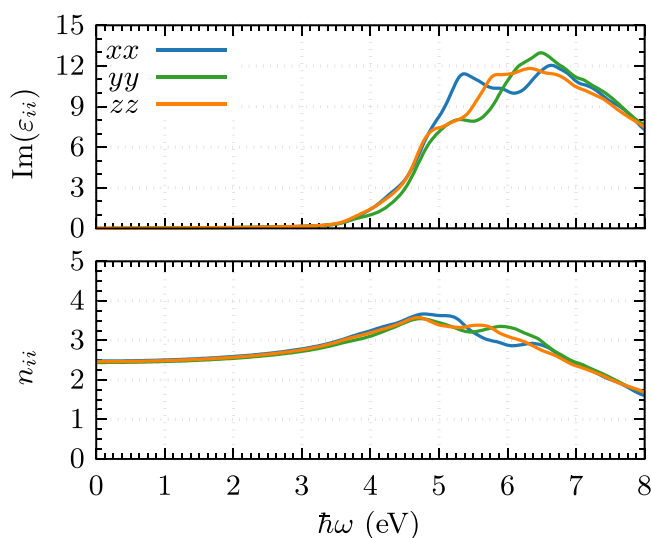


FIG. 6. Frequency-dependent imaginary part of the dielectric function (top) and refractive index (bottom) of BeZnS_2 . As a point of reference, for the unit cell employed here, the x , y , and z directions correspond to the out-of-plane, armchair, and zigzag directions as viewed from the (0001) direction in the wurtzite cell.

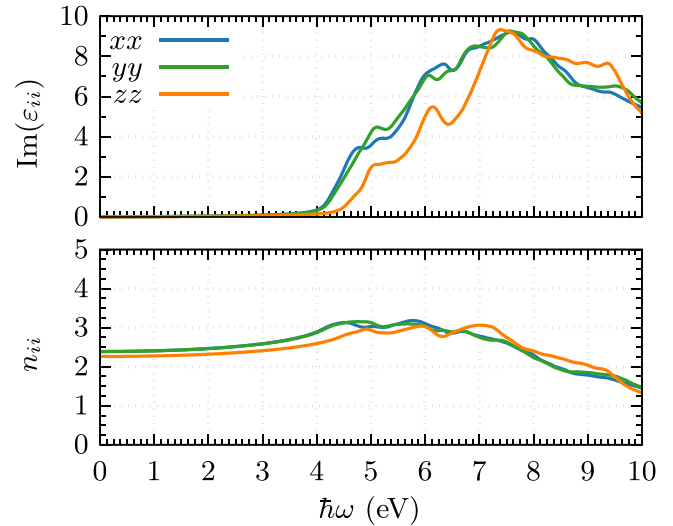


FIG. 7. Frequency-dependent imaginary part of the dielectric function (top) and refractive index (bottom) of HfZrO_4 .

range of values. However, we need to face this with some skepticism. A quick look at the B-Zn-S ternary phase diagram suggests that the alloying of these two materials is not energetically favorable, and therefore no ordered phase of wurtzite $\text{Zn}_x\text{Be}_{1-x}\text{S}$ is expected to exist in reality. In short, although wurtzite $\text{Zn}_x\text{Be}_{1-x}$ raises interest, any practical consideration on this material must be done from an alloy perspective, including the effect of disorder.

4. HfZrO_4

HfO_2 and ZrO_2 are by now well-known materials, not only due to their large dielectric constant and refractive index, but also due to the discovery of ferroelectricity in HfO_2 thin films (which is now a intense point of research. see, e.g., Refs. [65–69]). Both materials present a complicated landscape of phases [69,70], some of which with quite different dielectric constants [71].

With this in mind, finding a HfZrO_4 composition is not surprising—it can rather be seen as validation of the workflow. The present phase is derived from HfO_2 by replacing alternating rows of Hf with Zr, crystallizing in the $P63m$ space group. As expected from oxides, the valence is quite flat and dominated by oxygen's p orbitals, while the lowest conduction band is made up from a bundle of Zr and Hf d states. Because of the dispersive bands, the fundamental gap of 3.7 eV is almost equal to the direct one of 3.8 eV. With a static refractive index of $n = 2.3$ (see Fig. 7), this material lies in the neighborhood of LiTaO_3 and compares favorably to both HfO_2 and ZrO_2 .

5. $\text{Be}_2\text{P}_2\text{O}_7$

We could not find much information on the Be-P-O ternary. The only material of this family that seems to have been studied is the BeO_2P_6 beryllium polyphosphate, and even then, mostly for its structural properties (e.g., [72]). As such it is a bit hard to place $\text{Be}_2\text{P}_2\text{O}_7$ in a general landscape.

$\text{Be}_2\text{P}_2\text{O}_7$ belongs to the $C2/c$ space group, and its Be and P atoms are arranged in tetrahedra with oxygen atoms at the

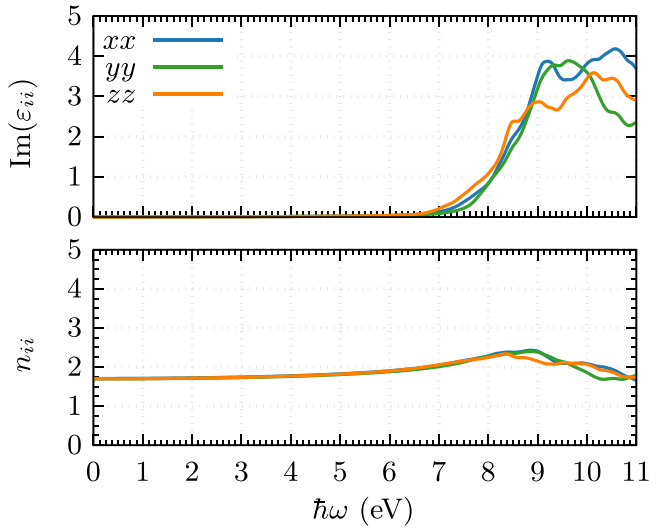


FIG. 8. Frequency-dependent imaginary part of the dielectric function (top) and refractive index (bottom) of $\text{Be}_2\text{P}_2\text{O}_7$.

vertices. Apart from this, the overall distribution of atoms is hard to convey, as the tetrahedra are arranged in different orientations, creating moderately sized gaps in the material.

This is the material with the largest band gap from our selection, at 6.3 eV. The valence band is very flat due to p -orbital contributions from O, but the valence has an almost parabolic minimum, comprised of s and p oxygen orbitals as well as p sulfur orbitals. The imaginary part of $\epsilon(\omega)$ grows slowly (see Fig. 8), although different polarization directions seem to contribute differently for this onset, and it does not attain a well-defined peak. Coupled to its large band gap, this translates into a rather flat variation of the refractive index, which has a static value of 1.65.

IV. IONIC CONTRIBUTION

As mentioned in the Introduction, the ionic part of the dielectric constant is crucial for electronics but is more complex to study than its electronic counterpart. Due to the more complex nature of this quantity, extremely dense k -point grids are necessary to converge ϵ^∞ [73], and doing a high-throughput analysis with such accuracy would be beyond our present capacity. We opted to focus on materials within the previous set that were shown to have large ϵ^0 and to be converged within the chosen calculation parameters. From this set, some interesting compounds appear.

The NaTaO_3 ceramic presents an extensive literature, most of it relating to its use as a photocatalytic for water splitting [74–77]. The present trigonal ($R3c$) phase is so far unreported in the databases, but it contains similar structural elements to those of other perovskitelike phases: Ta atoms form octahedra with O atoms at the vertices, and each Na is bonded to nine O atoms. With $\epsilon^\infty = 124$, it fits somewhere between its cubic ($\epsilon^\infty = 47$, mp-4170) and orthorhombic polymorphs ($\epsilon^\infty = 341$, mp-3858), having a comparable direct band gap of 2.7 eV. Lying a mere 1.3 meV/atom from the convex hull, this polymorph could be relevant for higher temperature dynamics of NaTaO_3 , something that will require further study.

With a similar composition, we find KTaO_3 , a stoichiometry well known in the community thanks to the properties of the stable perovskite phase [78–81]. The present $Fd\bar{3}m$ polymorph is considerably different, composed of two types of tetrahedron, with either Ta or K atoms at the vertices and O atoms in the middle of the edges. These structures are tilted in a way that O atoms are either connected to two Ta and one K or vice versa. Energetically, it is metastable, although it lies a mere 23 meV/atom from the convex hull. With a band gap of 3.3 eV and a total dielectric constant of 68, this phase is more suitable for electronic applications than its more stable polymorph ($E_g = 2.1$ eV, $E_g^d = 6.0$ eV, $\epsilon^0 = 3.3$, $\epsilon^\infty = 5.5$ [41]).

NdWN_3 crystallizes in the $R3c$ space group and is a distorted perovskite nitride, where the Nd atoms lie in the center of the cube, while the N and W lie at the vertices and edges, respectively. This family of compounds has been studied in the scope of hard magnets. Although the present phase has a total dielectric constant close to 228, it also has a rather small indirect band gap of 0.6 eV (even though $E_g^d = 1.9$ eV). Production of nitride perovskites is generally considered challenging, but recently other members of this family have been synthesized [82].

Before concluding we briefly mention HfZrO_4 . From the composition point of view, this is a mixture of the well-established HfO_2 and ZrO_2 , as previously mentioned. These materials have several polymorphs with only slight structural differences, for which ϵ^∞ can significantly vary. This was observed in our study, as the geometry optimization process with different input criteria lead to two structures with different space groups.

Conclusions

Using the available data for the static refractive index of semiconductors, we trained a crystal graph convolutional neural network for the screening of a large-scale dataset of new semiconductors. This model, which requires simply the crystal structure as an input for its prediction, was shown to attain sufficiently good accuracy for the efficient categorization of the static refractive index in a wide variety of compositions and structures. With this tool we were able to screen a recent large-scale dataset in the search for outstanding materials, reducing the target space to a size manageable for high-throughput density functional perturbation theory calculations. Specifically, we used this model to screen 55 000 previously unstudied candidate semiconductors obtained from the database of Schmidt *et al.* [3]. Subsequently, we conducted DFPT calculations on the most promising subset, resulting in the identification of 2431 materials. In the end we found several metastable materials that could be considered promising when compared to current large-refractive-index compounds. In our analysis the typical inverse proportionality relation between gap and refractive index is observed, but the behavior of several quaternary compounds hints at possible performance gains at the cost of stoichiometric complexity. Lastly, as mentioned in the Introduction, studying the ionic contribution to the dielectric function, which typically becomes more important than ϵ^∞ in electronics and circuitry, is also of interest.

From here, more detailed studies of the materials of interest are to be expected. Using more advanced methods such as GW or hybrid functionals will allow for a better description of the materials' electronic structure; analysis of alloying will conclude whether the presented phases are suitable for synthesis; studying the topology of defects will allow a better understanding of the practical behavior of the material; etc. It must be noted that the chemical similarity of some of the elements in these materials may in practice lead to disordered alloys. Whether this disorder is avoidable [83] or if it significantly affects the performance of the material requires analysis with dedicated tools [84–86]. In addition, a more focused approach

to the study of the more complex ionic contribution on the present dataset is reserved for future work.

ACKNOWLEDGMENTS

The authors acknowledge financial support from Fundação para a Ciência e Tecnologia, Portugal (Projects No. UIDB/04564/2020 and No. 2022.09975.PTDC, Contract No. 2020.04225.CEECIND) and the Laboratory for Advanced Computing at University of Coimbra for providing HPC resources that have contributed to the research results reported within this paper.

- [1] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, *npj Comput. Mater.* **5**, 83 (2019).
- [2] H. J. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. L. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, A. Smolyanyuk, S. Curtarolo, A. Tkatchenko, A. P. Bartók, S. Manzhos, M. Ihara, T. Carrington, J. Behler, O. Isayev, M. Veit *et al.*, *Electron. Struct.* **4**, 023004 (2022).
- [3] J. Schmidt, N. Hoffmann, H.-C. Wang, P. Borlido, P. J. M. A. Carriço, T. F. T. Cerqueira, S. Botti, and M. A. L. Marques, *Adv. Mater.* **35**, 2210788 (2023).
- [4] J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, and M. A. L. Marques, *Chem. Mater.* **29**, 5090 (2017).
- [5] A. M. Tehrani, A. O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T. D. Sparks, and J. Brgoch, *J. Am. Chem. Soc.* **140**, 9844 (2018).
- [6] Z. Zhang, A. M. Tehrani, A. O. Oliynyk, B. Day, and J. Brgoch, *Adv. Mater.* **33**, 2005112 (2021).
- [7] K. Choudhary and K. Garrity, *npj Comput. Mater.* **8**, 244 (2022).
- [8] J. Zhang, Z. Zhu, X.-D. Xiang, K. Zhang, S. Huang, C. Zhong, H.-J. Qiu, K. Hu, and X. Lin, *J. Phys. Chem. C* **126**, 8922 (2022).
- [9] Y. Zhuo, A. M. Tehrani, and J. Brgoch, *J. Phys. Chem. Lett.* **9**, 1668 (2018).
- [10] A. C. Rajan, A. Mishra, S. Satsangi, R. Vaish, H. Mizuseki, K.-R. Lee, and A. K. Singh, *Chem. Mater.* **30**, 4031 (2018).
- [11] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999).
- [12] J. M. Senior and M. Y. Jamro, *Optical Fiber Communications: Principles and Practice*, 3rd ed. (Financial Times/Prentice Hall, Harlow, England, New York, 2009).
- [13] T. Higashihara and M. Ueda, *Macromolecules* **48**, 1915 (2015).
- [14] R. C. Miller, *Appl. Phys. Lett.* **5**, 17 (1964).
- [15] A. I. Kingon, J.-P. Maria, and S. K. Streiffer, *Nature (London)* **406**, 1032 (2000).
- [16] J. Robertson, *Rep. Prog. Phys.* **69**, 327 (2006).
- [17] B. Wang, W. Huang, L. Chi, M. Al-Hashimi, T. J. Marks, and A. Facchetti, *Chem. Rev.* **118**, 5690 (2018).
- [18] K. Hong, T. H. Lee, J. M. Suh, S.-H. Yoon, and H. W. Jang, *J. Mater. Chem. C* **7**, 9782 (2019).
- [19] S. Tripathy, *Opt. Mater. (Amsterdam)* **46**, 240 (2015).
- [20] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [21] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [22] S. Baroni, P. Giannozzi, and A. Testa, *Phys. Rev. Lett.* **58**, 1861 (1987).
- [23] X. Gonze, *Phys. Rev. A* **52**, 1086 (1995).
- [24] G. Kresse and J. Furthmüller, *Phys. Rev. B* **54**, 11169 (1996).
- [25] G. Kresse and D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).
- [26] K. Mathew, J. H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I.-h. Chu, T. Smidt, B. Bocklund, M. Horton, J. Dagdelen, B. Wood, Z.-K. Liu, J. Neaton, S. P. Ong, K. Persson, and A. Jain, *Comput. Mater. Sci.* **139**, 140 (2017).
- [27] A. Jain, S. P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, and K. A. Persson, *Concurrency Comput.* **27**, 5037 (2015).
- [28] I. Petousis, D. Mrdjenovich, E. Ballouz, M. Liu, D. Winston, W. Chen, T. Graf, T. D. Schladt, K. A. Persson, and F. B. Prinz, *Sci. Data* **4**, 160134 (2017).
- [29] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [30] R. Elmer, M. Berg, L. Carlen, B. Jakobsson, B. Noren, A. Oskarsson, G. Ericsson, J. Julien, T.-F. Thorsteinsen, M. Guttormsen *et al.*, *Phys. Rev. Lett.* **78**, 1396(E) (1997).
- [31] P. Borlido, T. Aull, A. W. Huran, F. Tran, M. A. L. Marques, and S. Botti, *J. Chem. Theory Comput.* **15**, 5069 (2019).
- [32] P. Borlido, J. Schmidt, A. W. Huran, F. Tran, M. A. L. Marques, and S. Botti, *npj Comput. Mater.* **6**, 96 (2020).
- [33] Á. Morales-García, R. Valero, and F. Illas, *J. Phys. Chem. C* **121**, 18862 (2017).
- [34] M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller, and F. Bechstedt, *Phys. Rev. B* **73**, 045112 (2006).
- [35] I. Petousis, W. Chen, G. Hautier, T. Graf, T. D. Schladt, K. A. Persson, and F. B. Prinz, *Phys. Rev. B* **93**, 115151 (2016).
- [36] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, *Comput. Mater. Sci.* **68**, 314 (2013).
- [37] Z. Hashin and S. Shtrikman, *Phys. Rev.* **130**, 129 (1963).
- [38] F. Bechstedt, *Many-Body Approach to Electronic Excitations: Concepts and Applications*, 1st ed., Springer Series in Solid-State Sciences No. 181 (Springer Berlin Heidelberg, Imprint: Springer, Berlin, Heidelberg, 2015).
- [39] T. Xie and J. C. Grossman, *Phys. Rev. Lett.* **120**, 145301 (2018).
- [40] A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, *npj Comput. Mater.* **6**, 138 (2020).
- [41] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, *APL Mater.* **1**, 011002 (2013).
- [42] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook* (Springer, Cham, Switzerland, 2018).
- [43] M. G. Kendall, *Biometrika* **30**, 81 (1938).

- [44] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, *Nat. Methods* **17**, 261 (2020).
- [45] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevMaterials.8.015201> for an analysis of the general distribution of refractive indices and gaps, as well as densities of states and band structures of the materials discussed in more detail.
- [46] F. Naccarato, F. Ricci, J. Suntivich, G. Hautier, L. Wirtz, and G.-M. Rignanese, *Phys. Rev. Mater.* **3**, 044602 (2019).
- [47] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, *Comput. Mater. Sci.* **58**, 227 (2012).
- [48] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, *npj Comput. Mater.* **1**, 15010 (2015).
- [49] L. Petit, A. Svane, Z. Szotek, and W. M. Temmerman, *Phys. Rev. B* **72**, 205118 (2005).
- [50] H. Kawazoe, H. Yanagi, K. Ueda, and H. Hosono, *MRS Bull.* **25**, 28 (2000).
- [51] A. Banerjee and K. Chattopadhyay, *Prog. Cryst. Growth Charact. Mater.* **50**, 52 (2005).
- [52] J. He, S. S. Naghavi, V. I. Hegde, M. Amsler, and C. Wolverton, *Chem. Mater.* **30**, 4978 (2018).
- [53] A. Kuhn, A. Chevy, and M.-J. Naud, *J. Cryst. Growth* **9**, 263 (1971).
- [54] A. M. Bothwell, J. A. Drayton, and J. R. Sites, *Sol. Energy Mater. Sol. Cells* **237**, 111549 (2022).
- [55] X. Mathew, J. Drayton, V. Parikh, N. R. Mathews, X. Liu, and A. D. Compaan, *Semicond. Sci. Technol.* **24**, 015012 (2009).
- [56] F. Kazami, W.-C. Sun, K. Taguri, T. Nakasu, T. Aiba, S. Yamashita, S. Hattori, T. Kizu, M. Kobayashi, and T. Asahi, *Phys. Status Solidi B* **253**, 635 (2016).
- [57] W. Sun, T. Nakasu, K. Taguri, T. Aiba, S. Yamashita, M. Kobayashi, H. Togo, and T. Asahi, *Phys. Status Solidi C* **11**, 1252 (2014).
- [58] Y. Hinuma, T. Hatakeyama, Y. Kumagai, L. A. Burton, H. Sato, Y. Muraba, S. Iimura, H. Hiramatsu, I. Tanaka, H. Hosono, and F. Oba, *Nat. Commun.* **7**, 11962 (2016).
- [59] C. Tholander, C. B. A. Andersson, R. Armiento, F. Tasnádi, and B. Alling, *J. Appl. Phys.* **120**, 225102 (2016).
- [60] R. Woods-Robinson, V. Stevanović, S. Lany, K. N. Heinselman, M. K. Horton, K. A. Persson, and A. Zakutayev, *Phys. Rev. Mater.* **6**, 043804 (2022).
- [61] S. Körbel, D. Kammerlander, R. Sarmiento-Pérez, C. Attacalite, M. A. L. Marques, and S. Botti, *Phys. Rev. B* **91**, 075134 (2015).
- [62] C.-Y. Yeh, S.-H. Wei, and A. Zunger, *Phys. Rev. B* **50**, 2715 (1994).
- [63] A. Ebina, E. Fukunaga, and T. Takahashi, *Phys. Rev. B* **12**, 687 (1975).
- [64] H. C. Ong and R. P. H. Chang, *Appl. Phys. Lett.* **79**, 3612 (2001).
- [65] K. D. Kim, Y. H. Lee, T. Gwon, Y. J. Kim, H. J. Kim, T. Moon, S. D. Hyun, H. W. Park, M. H. Park, and C. S. Hwang, *Nano Energy* **39**, 390 (2017).
- [66] R. Materlik, C. Künneth, and A. Kersch, *J. Appl. Phys.* **117**, 134109 (2015).
- [67] S. Migita, H. Ota, S. Asanuma, Y. Morita, and A. Toriumi, *Appl. Phys. Express* **14**, 051006 (2021).
- [68] J. Müller, T. S. Böske, U. Schröder, S. Mueller, D. Bräuhäus, U. Böttger, L. Frey, and T. Mikolajick, *Nano Lett.* **12**, 4318 (2012).
- [69] U. Schroeder, M. H. Park, T. Mikolajick, and C. S. Hwang, *Nat. Rev. Mater.* **7**, 653 (2022).
- [70] B. Johnson and J. L. Jones, in *Ferroelectricity in Doped Hafnium Oxide: Materials, Properties and Devices* (Elsevier, New York, 2019), pp. 25–45.
- [71] X. Zhao and D. Vanderbilt, *Phys. Rev. B* **65**, 233106 (2002).
- [72] E. Schultz and F. Liebau, *Z. Kristallogr. Cryst. Mater.* **154**, 115 (1981).
- [73] J. Qu, D. Zagaceta, W. Zhang, and Q. Zhu, *Sci. Data* **7**, 81 (2020).
- [74] P. Kanhere, P. Shenai, S. Chakraborty, R. Ahuja, J. Zheng, and Z. Chen, *Phys. Chem. Chem. Phys.* **16**, 16085 (2014).
- [75] Y.-L. Liu, C.-L. Yang, M.-S. Wang, X.-G. Ma, and Y.-G. Yi, *Mater. Res. Bull.* **107**, 125 (2018).
- [76] G. R. Portugal, S. F. Santos, and J. T. Arantes, *Appl. Surf. Sci.* **502**, 144206 (2020).
- [77] G. A. S. Alves, H. A. Centurion, J. R. Sambrano, M. M. Ferrer, and R. V. Gonçalves, *ACS Appl. Energy Mater.* **4**, 671 (2021).
- [78] K. Zou, S. Ismail-Beigi, K. Kisslinger, X. Shen, D. Su, F. J. Walker, and C. H. Ahn, *APL Mater.* **3**, 036104 (2015).
- [79] S. Mallik, G. C. Ménard, G. Saïz, H. Witt, J. Lesueur, A. Gloter, L. Benfatto, M. Bibes, and N. Bergeal, *Nat. Commun.* **13**, 4625 (2022).
- [80] K. Ueno, S. Nakamura, H. Shimotani, H. T. Yuan, N. Kimura, T. Nojima, H. Aoki, Y. Iwasa, and M. Kawasaki, *Nat. Nanotechnol.* **6**, 408 (2011).
- [81] C. Liu, X. Yan, D. Jin, Y. Ma, H.-W. Hsiao, Y. Lin, T. M. Bretz-Sullivan, X. Zhou, J. Pearson, B. Fisher, J. S. Jiang, W. Han, J.-M. Zuo, J. Wen, D. D. Fong, J. Sun, H. Zhou, and A. Bhattacharya, *Science* **371**, 716 (2021).
- [82] K. R. Talley, C. L. Perkins, D. R. Diercks, G. L. Brennecke, and A. Zakutayev, *Science* **374**, 1488 (2021).
- [83] A. Mascarenhas, *Spontaneous Ordering in Semiconductor Alloys* (Kluwer Academic/Plenum Publishers, 2002).
- [84] A.-B. Chen and A. Sher, *Semiconductor Alloys: Physics and Materials Engineering*, Microdevices (Plenum Press, New York, 1995).
- [85] D. B. Laks, L. G. Ferreira, S. Froyen, and A. Zunger, *Phys. Rev. B* **46**, 12587 (1992).
- [86] A. Zunger, S.-H. Wei, L. G. Ferreira, and J. E. Bernard, *Phys. Rev. Lett.* **65**, 353 (1990).