

---

# Pushing the Pareto front of band gap and permittivity: ML-guided search for dielectric materials

---

Janosh Riebesell<sup>1</sup>

University of  
Cambridge

Rhys Goodall<sup>1</sup>

University of  
Cambridge

T. Wesley Surta<sup>2</sup>

University of Liverpool

Michael Gaultois<sup>2</sup>

University of Liverpool

Alpha Lee<sup>1</sup>

University of  
Cambridge

2024-01-11

## ABSTRACT

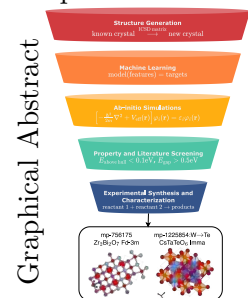
Materials with high-dielectric constant easily polarize under external electric fields, allowing them to perform essential functions in many modern electronic devices. Their practical utility is determined by two conflicting properties: high dielectric constants tend to occur in materials with narrow band gaps, limiting the operating voltage before dielectric breakdown. We present a high-throughput workflow that combines element substitution, ML pre-screening, ab initio simulation and human-in-the-loop filtering to efficiently explore the vast space of unknown materials for potential dielectrics, leading to the synthesis and characterization of two novel dielectric materials, CsTaTeO<sub>6</sub> and Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub>. This work constitutes a successful demonstration of ML-guided multi-objective materials screening reporting the first high-purity synthesis and dielectric characterization of Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub> with a band gap of 2.27 eV and a permittivity of 20.5, meeting all target metrics of our multi-objective search. The second synthesized candidate, CsTaTeO<sub>6</sub>, was a structure generated via element substitution not present in our reference data sources, thus exemplifying successful de-novo materials design of a thermodynamically stable material. However, unfortunately, the low observed bandgap of 1.05 eV renders CsTaTeO<sub>6</sub> unusable as a dielectric material. We believe our successful synthesis results highlight both the potential and current limitations of applying ML in materials discovery. We emphasize the wider need for experimental validation of ML predictions to better understand and address limitations in high-throughput databases powering ML models for computational material science.

## Keywords

Pareto front · data-driven · density functional theory · dielectric materials  
· high-throughput · machine learning · materials discovery

## 1 Introduction

Dielectric materials are indispensable in numerous modern electronic devices including central processing units (CPUs), random access memory (RAM), solid-state disks (SSDs), high-frequency (5G) antennas, photovoltaics, and light-emitting diodes (LEDs) [1], [2]. Their utility hinges on the intricate balance between dielectric constant and band gap, two anti-correlated properties that rarely co-occur in a single material. High band gaps are crucial for reducing leakage current and preventing dielectric breakdown when subjected to high voltage. Conversely, a large dielectric constant is desirable for minimizing the energy required for polarization, which is especially



important in applications like transistor gates. As transistors continue to shrink, the need for materials that can serve as ultra-thin gate dielectrics while withstanding operating voltages grows.

Historically, the discovery of dielectric materials has often relied on trial and error. Recent advancements, particularly in automated workflows for computational screening using density functional perturbation theory (DFPT) have shown promise in systematically searching for high-performance dielectrics, e.g. mapping the bandgap-dielectric Pareto front of binary and ternary oxides [3]. Improvements in compute power and workflow robustness have since enabled the scaling to several thousand diverse materials [4], [5], [6].

However, the sheer size of the space of  $\sim 10^5$  known, let alone the  $\sim 10^{10}$  hypothesized materials (up to quaternary order) [7], prohibits sampling without inductive bias and presents a daunting challenge for existing computational methods. Consequently, the dielectric properties of the vast majority of the  $\sim 10^7$  simulated inorganic crystals remain unknown, making it likely a more comprehensive exploration of the space should yield novel high-performance materials. To screen even a small subset of the full space requires orders of magnitude cheaper methods. Worse, to go beyond the  $\sim 10^5$  known materials introduces another layer of computational complexity in the form of thermodynamic stability prediction on top of estimating band gap and dielectric constant.

To address this, we propose a new dielectric discovery workflow that judiciously integrates machine learning (ML) as the first filter in a multi-step funnel. ML, while less reliable than traditional methods like DFPT, is orders of magnitude faster and quickly improving in accuracy. Our ML-guided approach uses surrogate models for band gaps, dielectric constants, and formation energies. Instead of exact Cartesian coordinates, we employ Wyckoff positions for a coordinate-free, coarse-grained crystal structure representation [8]. This enables rapid generation and stability prediction of novel structures through elemental substitutions. Following DFPT validation of the most promising candidates, the last selection step is an expert committee to incorporate human intuition when weighing the risks, precursor availability and ease of experimental synthesis of all high-expected-reward materials. Finally, we validate the whole workflow by deploying it from start to finish which culminated in making and characterizing two new metastable materials in the process:  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$  which partially and fully satisfy our target metrics, respectively.

Finding exceptional materials that extremize a single property necessarily requires extrapolation from the training data, for example maximizing hardness [9], [10], [11]. This is fundamentally at odds with the statistical nature of ML, leading to increased error and less reliable predictions. Our approach diverges from previous efforts by choosing a target class of materials where the path to application relevance requires balancing multiple conflicting properties. This was driven by the hypothesis that doing so may allow the ML models to operate within regions of good training support while still predicting materials with exceptional figures of merit. This type of tradeoff is ubiquitous in material science and is seen in other materials classes such as thermoelectrics (need high low thermal but high electrical conductivity) [12], [13], catalysts (need high activity for fast reactions which tends to lower selectivity, increasing unwanted side reactions) [14], [15], high-strength and shape-memory alloys (need high strength and high ductility) [16], [17], and many more.

Despite a nascent but growing body of work on automated and high-throughput synthesis [18], [19], [20], [21], [22], [23], experimental validation remains a key bottleneck in the design of materials. The process of manually developing experimental synthesis recipes for theoretical materials is very time-consuming, often taking months to a year per material. The central claim of ML-guided screening and related efforts in rational materials design is that we can reduce the downside risk of attempting novel synthesis procedures by increasing the hit rate of successful materials. To test

the performance of our ML-guided approach we developed synthesis procedures for two materials predicted to be high-performing -  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$ , with the structure of  $\text{CsTaTeO}_6$  created by our generative workflow. Both materials displayed dielectric character with measured permittivities in the 43rd and 81st percentile, respectively, of 136 experimental reference results for dielectric materials reported in [4], [24], validating the benefits of our ML-guided workflow.

In summary, our work showcases an advancement in ML-guided materials discovery, demonstrating its potential in efficiently navigating the vast landscape of dielectric materials and balancing multiple material properties for optimal device performance.

## 2 Results

We first report the computational output of our workflow and then present experimental validation of two novel dielectric materials,  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$ .

## 2.1 A scalable machine learning workflow for dielectric discovery

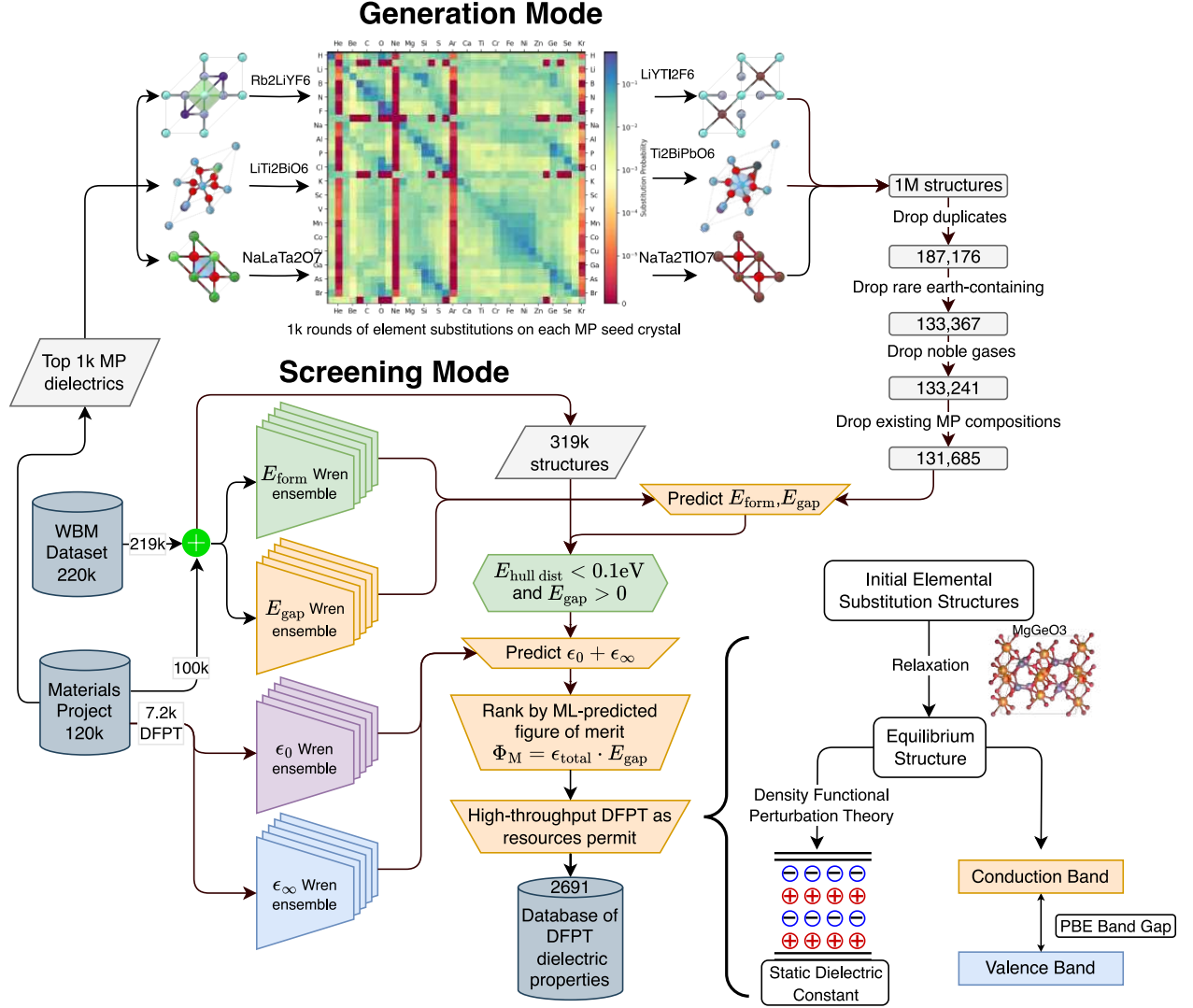


Figure 1: Diagram of our dielectric material discovery workflow, integrating ML pre-screening and elemental substitution for generating novel crystals with high-throughput DFPT validation. The discovery pipeline can operate in two modes: screening and generation. Screening mode searches for large permittivity among known materials. In generation mode, we feed the top 1k MP structures by figure of merit  $\Phi_M$  into an element substitution process.

This section describes the components and design decisions of our dielectric discovery workflow visualized in Figure 1.

The large search space and high cost of experimental validation demand a funnel approach to dielectric materials discovery. To maximize the size of the initial candidate pool and still retain tractable computational cost, less auspicious materials must be discarded by a hierarchy of successively more expensive but higher-fidelity computational filters. Such an approach maximizes return on invested effort by allotting more resources to candidates which accumulated evidence of expected utility in earlier filters. Our proposed implementation for such a funnel workflow depicted in Figure 1 precedes high-throughput DFPT with 5-6 orders of magnitude cheaper ML pre-screening to reduce a large list of 131685 candidate materials down to 2532 with computed dielectric properties.

We pre-screen based on 4 quantities - thermodynamic stability derived from predicted formation energy  $E_{\text{form}}$ , band gap  $E_{\text{gap}}$ , ionic permittivity  $\epsilon_0$  and electronic permittivity  $\epsilon_\infty$  - each of which

is predicted by a separate ensemble of 10 Wren models [8] independently trained from random initializations. This allows us to both massively expand the search pool of initial candidates and waste fewer resources on unpromising compounds. The formation energy and band gap training sets each consist of 319601 data points, the combination of 98850 Materials Project (MP) [25] calculations and 220751 from the WBM dataset [26] (named WBM from the author’s last name initials) which was generated with MP-compatible VASP settings. The  $\epsilon_0$  and  $\epsilon_\infty$  ensembles are trained on the much smaller dataset of 7172 DFPT calculations in MP (database version 2020-09-08) due to the lack of additional MP-compatible dielectric datasets.

While simply screening materials within large ab-initio databases for which properties of interest have yet to be calculated is a viable strategy, it is also important to demonstrate the generative capabilities of ML-based workflows. To this end, we identify the top 1k MP structures by figure of merit  $\Phi_M = \epsilon_{\text{tot}} \cdot E_{\text{gap PBE}}$  and use them as seed crystals for element substitution. The expectation is that this generates novel structures with increased likelihood of high  $\Phi_M$ . The substitution process involves replacing all sites of one element in the structure with a chemically similar element (e.g. Na  $\rightarrow$  K), as determined by a similarity matrix mined from the ICSD [27]. After filtering out duplicates (compositions that already exist in MP or WBM, i.e. we do not consider structural degrees of freedom) as well as compounds containing noble gases, lanthanides or actinides, we are left with 131685 potential new dielectric materials.

Using the trained Wren ensembles, we predict  $E_{\text{form}}$ ,  $E_{\text{gap}}$ ,  $\epsilon_{\text{ionic}}$  and  $\epsilon_{\text{elec}}$  for all candidates, both those sourced from high-throughput databases and those produced using our generative methodology. We estimate the convex hull distance for each crystal from these predicted energies and discard those more than 0.1 eV/atom above the hull. This is motivated by the observation that 90% of crystals in ICSD are predicted to be less than 67 meV/atom above the convex hull [28]. This tolerance towards instability accounts for errors in DFT energies and the fact that some thermodynamically unstable materials are kinetically or entropically meta-stable and hence synthesizable.

The remaining candidates are ranked by their ML-predicted figure of merit  $\Phi_M^{\text{Wren}}$  and subjected to a high-throughput DFPT workflow as our computational budget permits, resulting in a database of 2532 dielectric properties.

## 2.2 Computational discovery of dielectrics beyond the Pareto front

The violin plot in Figure 2 shows Gaussian kernel density estimates (KDE) of all 2532 DFPT-computed electronic and ionic dielectric constants split by crystal system. Unlike the electronic contribution which is lower-bounded by the vacuum permittivity, the ionic dielectric constant can be zero in all crystal systems. We observe a general trend of higher dielectric constant the higher the crystal symmetry, especially for the ionic contribution. Only cubic crystals reach significant electronic permittivity with a median of 10.

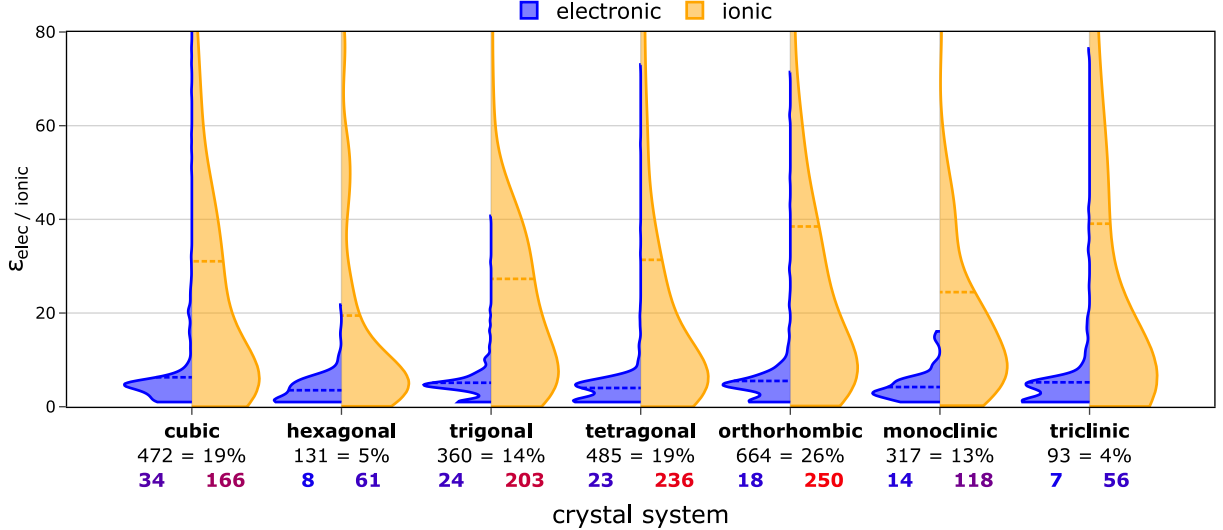


Figure 2: Violin plot showing Gaussian KDEs of DFPT-computed electronic (blue left halves) and ionic (orange right halves) contributions to the dielectric constant split by crystal system. The dashed horizontal lines in each violin show the median. Below each crystal system is the number of materials we have for it as well as its share of the total DFPT dataset in percent. The colored bold numbers (blue = low, red = high) show the mean of the top 30 electronic/ionic dielectric constants for each crystal system.

Figure 3 compares the results from our methodology against those published in Petousis *et al.* [4] and Qu *et al.* [29] by plotting the PBE band gap on the  $y$ -axis against the total dielectric constant on the  $x$ -axis on a log-log scale. The blue circles show the 2532 DFPT results we computed. The 441 orange diamonds show data generated by [29] while the 139 green squares are from [4]. The dark blue dashed isolines indicate constant figure of merit at values  $\Phi_M = E_{\text{gap}} \cdot \epsilon_{\text{tot}} = c \in \{30, 60, 120, 240\}$  for band gap  $E_{\text{gap}}$  and total dielectric constant  $\epsilon_{\text{tot}}$ . Our results achieve a larger number of materials beyond the highest  $\Phi_M$  isoline of 240 than both previous works combined. We also achieve a higher hit rate per DFPT calculation of such high-merit materials as shown in Table 1. For J. Qu, D. Zagaceta, W. Zhang, and Q. Zhu [29] 15/441 = 3.4% of materials achieve  $\Phi_M > 240$ , while I. Petousis *et al.* [4] reach 7/139 = 5.0% and our data has 155/2680 = 5.8% materials with  $\Phi_M > 240$ . Note that our hit rate increases even further when post-hoc excluding metals, i.e. filtering the hit rate analysis for materials with a band gap of at least 0.1 eV. While the other works started from DFT structures with known band gaps and hence were able to filter out metals from the outset, the same is not possible when generating novel crystals with unknown electronic structures. Our workflow instead relies on ML band gap prediction to filter out metals. This step unfortunately suffers from a high false positive rate (metals misclassified as semiconductors/insulators). Thus by upgrading to a better band gap model, a future realization of our workflow could achieve a high-merit hit rate in excess of 154/1911 = 7.5%.

Study	Number of Hits / Total	Hit Rate (%)
Petousis <i>et al.</i> [4]	7/139	5.0
Qu <i>et al.</i> [29]	15/441	3.4
This work	155/2532	5.8
This work (with $E_{\text{gap}} > 0.1$ eV)	154/1911	7.5

Table 1: Hit rate comparison for materials with  $\Phi_M > 240$ . Excluding metals misclassified as insulators by our band gap models (which did not enter the other works in the first place), we achieve a  $\Phi_M > 240$  hit rate of 7.5%. This validates our approach of creating candidate structures from known dielectrics and pre-screening with ML.

Figure 3 compares our DFPT data to the results of I. Petousis *et al.* [4] and J. Qu, D. Zagaceta, W. Zhang, and Q. Zhu [29]. Our workflow generates more high- $\Phi_M$  materials than both previous works combined and at a higher hit rate per expensive DFPT calculation than either I. Petousis *et al.* [4] or J. Qu, D. Zagaceta, W. Zhang, and Q. Zhu [29]. We believe this hit rate increase is attributable to ML pre-screening and substituting elements into known dielectric materials.

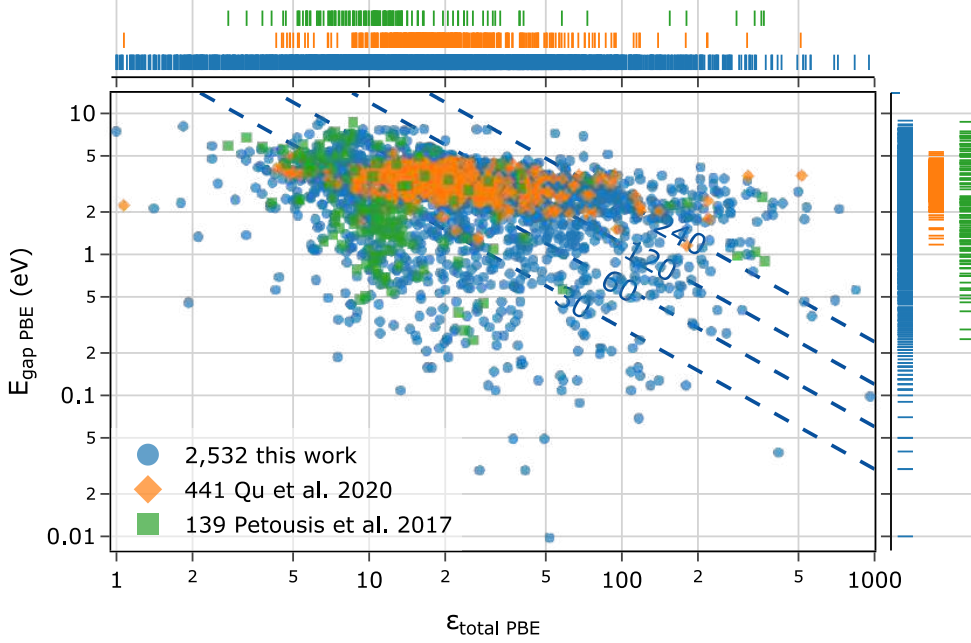


Figure 3: Log-log plot of PBE band gap  $E_{\text{gap}}$  vs. total dielectric constant  $\epsilon_{\text{tot}}$  visualizing the hit rates for high- $\Phi_M$  materials from different studies. Many of our DFPT data points (blue circles) reach into regions far beyond the 240 eV isoline. The orange diamonds and green squares show results from I. Petousis *et al.* [4] and J. Qu, D. Zagaceta, W. Zhang, and Q. Zhu [29] which produce fewer  $\Phi_M > 240$  materials, both in absolute numbers and as a fraction of dataset size (see Table 1). The dark blue lines indicate constant figure of merit  $\Phi_M = E_{\text{gap}} \cdot \epsilon_{\text{tot}}$ . The stacked marginal rugs along the top and right show the distribution of band gaps and dielectric constants in each dataset.

### 2.3 Prospective Experimental Validation

To validate our workflow’s ability to produce viable dielectric materials in practice, we selected  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$  for experimental synthesis and characterization. Our selection criteria incorporated DFPT results, prior literature or related materials appearing in the ICSD [30], as well as precursor availability and expected ease of synthesis. The selection process was facilitated by a custom web interface to visualize DFPT results on the Pareto front hooked up to a shared database for note-taking and collecting prior literature appearances on individual candidate materials detailed in Section 3.7. Even so, making  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$  required several trial-and-error iterations to optimize the synthesis conditions which we detail in this section.



### 2.3.1 Optimization and Purity

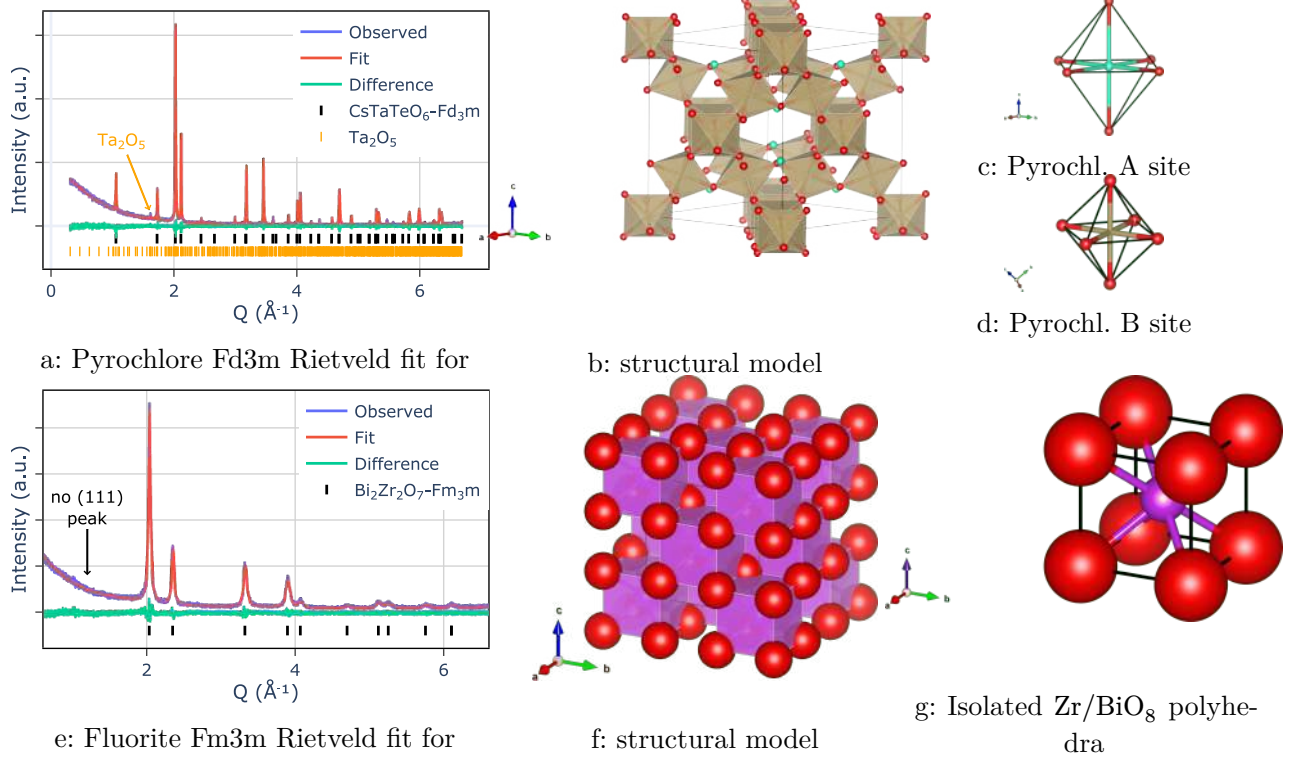


Figure 4: Structural determination of  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$  ([mp-756175](#)) using XRD and Rietveld refinement.  $Q = 2\pi \cdot d^{-1} [\text{\AA}^{-1}]$  is the scattering vector. [Figure 4b](#) Crystal structure of the best Rietveld fit for  $\text{CsTaTeO}_6$  with [Figure 4c](#) and [Figure 4d](#) showing the defect-pyrochlore A and B site octahedra. [Figure 4f](#) Crystal structure of the best Rietveld fit for  $\text{Bi}_2\text{Zr}_2\text{O}_7$  with [Figure 4g](#) showing the isolated  $\text{Zr}/\text{BiO}_8$  polyhedra. Notable impurities were detected in the XRD scan [Figure 4a](#).  $\text{CsTaTeO}_6$  has many  $hkl$  reflections, most of which are not distinguishable from the background noise. The most prominent observable peak at  $Q = 1.7$  as marked by the orange arrow. The absence of a (111) peak in the Rietveld fit [Figure 4e](#) suggests a defect-fluorite structure with one oxygen vacancy, in contrast to the literature-proposed pyrochlore model.



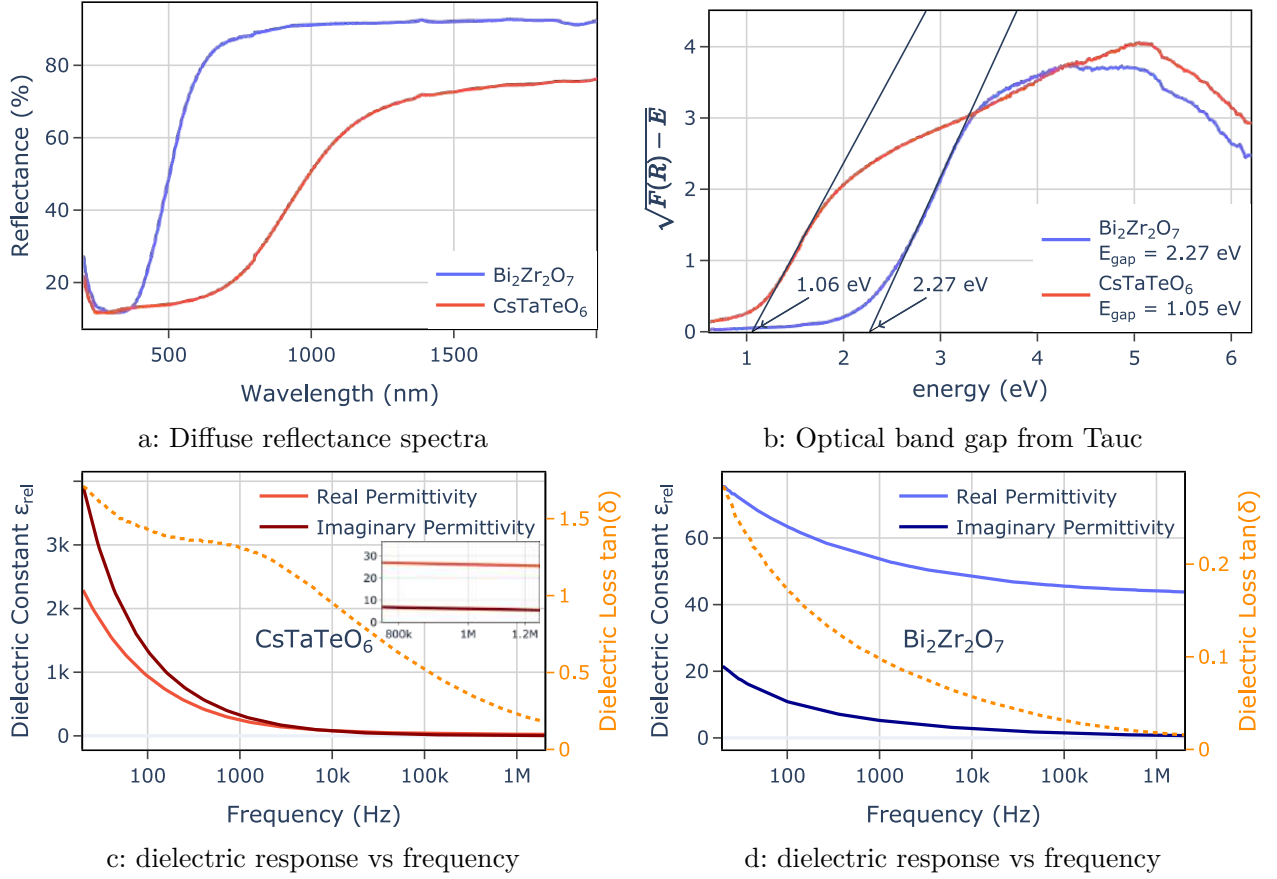


Figure 5: Dielectric measurements of  $\text{Bi}_2\text{Zr}_2\text{O}_7$  and  $\text{CsTaTeO}_6$ . **Figure 5a** Diffuse reflectance spectra for both compounds exhibit distinctive absorption edges, indicating ordered crystalline structures. **Figure 5b** Tauc plot measuring absorption coefficient  $\alpha(E_{\text{ph}})$  vs photon energy  $E_{\text{ph}} = h\nu$  for both compounds. The extracted optical band gaps are  $E_{\text{gap}} = 2.27$  eV for  $\text{Bi}_2\text{Zr}_2\text{O}_7$  and  $1.05$  eV for  $\text{CsTaTeO}_6$ . **Figure 5c** Dielectric response of  $\text{CsTaTeO}_6$  as a function of frequency. We measure  $\epsilon_{\text{tot}} = 26$  at 1 MHz electric field (compared to 67 from DFPT) Its unwelcome high dielectric loss of  $\tan(\delta) = 0.23$  at 1 MHz confirms the semiconducting nature observed in the Tauc plot's spectroscopic data. **Figure 5d** Dielectric response of  $\text{Bi}_2\text{Zr}_2\text{O}_7$  as a function of frequency yields  $\epsilon_{\text{tot}} = 20.5$  at 1 MHz (compared to 206 from DFPT) We highlight  $\text{Bi}_2\text{Zr}_2\text{O}_7$ 's dielectric loss of less than 0.1 above 1 kHz, a sufficiently low value for many practical applications.

We use X-ray diffraction (XRD) data and Rietveld fits to test our structural models for  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$ .

The measured XRD pattern for our sample at 80% of the theoretical weight density readily fits a defect-pyrochlore model. The atomic displacement parameters were small but positive within error. Even though multiple disordered models were explored, the defect-pyrochlore provided the best fit. We detected impurities constituting  $4.20 \pm 0.12$  % of total weight that are highlighted in **Figure 4a**.

For  $\text{Bi}_2\text{Zr}_2\text{O}_7$ , we explored optimal synthesis temperatures between  $550^\circ\text{C} - 750^\circ\text{C}$ . An extensive 8-hour XRD scan of after 48 h of heating at  $650^\circ\text{C}$  confirmed the absence of and impurities in the sample, which significantly surpasses existing literature in terms of purity [31]. After sintering, we obtained a ceramic sample with 92+% of the theoretical density of a single crystal. Contrary to literature reports that typically describe an impure pyrochlore with a noticeable (111) reflection [31], our samples exhibit no such peaks. Prolonged heating did result in a broad (111) peak but was accompanied by undesired and impurities. Avoiding prolonged heat, the Rietveld analysis in **Figure 4e** shows the (111) peak to be absent, favoring a defect-fluorite model for , in contrast

to the literature-proposed pyrochlore models. The compound exhibited large atomic displacement parameters ( $B_{\text{iso}}$ ) which may arise from two superimposed crystallographic positions or due to off-stoichiometry (occupancy). Both commonly result in models with large atomic displacement parameters that simulate the distribution of electron density from these sites. However, attempts to reduce atomic displacement using site splitting and occupancy refinement for disordered materials did not yield better fits. Higher-quality diffraction data, e.g. from neutron scattering, would likely be required for more accurate modeling.

Further details on synthesis development, equipment used and XRD fitting for both  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$  are provided in methods [Section 3.8](#) and [Section 6](#) and [Section 7](#).

### 2.3.2 Dielectric Characterization

Having selected, synthesized at high purity, and confirmed the structures of  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$ , we investigate their physical properties. The band gaps of both materials were identified using UV-vis impedance spectroscopy on powders using diffuse reflectance and an integrating sphere. These data can be seen in [Figure 5a](#) and they were modified and fit using the Kubelka Munk [\[32\]](#) equation to extract the bandgap, seen in [Figure 5b](#). [Figure 5a](#) shows diffuse reflectance measurements for  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$  exhibiting distinctive absorption edges. The extracted band gaps are 2.27 eV for  $\text{Bi}_2\text{Zr}_2\text{O}_7$  and 1.05 eV for  $\text{CsTaTeO}_6$ . It is worth noting that the measurements for both  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$  turned out much lower than the DFT-calculated values of 2.09 eV and 2.96 eV respectively. This is surprising given PBE’s tendency to underestimate experimental band gaps. For  $\text{CsTaTeO}_6$ , this may be due to complex defect effects not captured by DFT arising from Cs or Te volatility [\[33\]](#). A more accurate ML band gap model that provides a more specific filter for metals and semiconductors would save future implementations of our workflow from spending unwarranted compute and lab time on semiconducting compounds like  $\text{CsTaTeO}_6$ . However, given the limitations of PBE observed for these materials it would be advisable to train the model on reference data obtained from higher levels of theory.

The low value of  $E_{\text{gap}} = 1.05$  eV for  $\text{CsTaTeO}_6$  is consistent with its observed black color and unfortunately renders it unusable as a dielectric material. The dielectric measurements in [Figure 5c](#) confirm a band gap-related high dielectric loss<sup>1</sup>. It is worth noting that despite its low band gap,  $\text{CsTaTeO}_6$  exhibits high polarizability of  $\epsilon_{\text{real}} = 26$  at 1 MHz up to its low breakdown voltage. However, its high dielectric loss of  $\tan(\delta) = 0.23$  at 1 MHz confirms the semiconducting behavior observed in the spectroscopic data. We also caveat the measured dielectric constant with the fact that 23% loss makes the extracted  $\epsilon_{\text{real}}$  value less reliable.

The compound  $\text{Bi}_2\text{Zr}_2\text{O}_7$  has an observed band gap of 2.27 eV, making it a useful dielectric. Importantly, the observed band gap is 0.27 eV (12.5%) higher than the previously reported mixed phase [\[31\]](#) who report  $E_{\text{gap}} = 2$  eV. This suggests reduced defect states and further substantiates the high purity and distinct phase of our synthesized materials. Dense ceramics were only accessible using spark plasma sintering, due to the metastable nature of the compound. Room temperature dielectric properties as a function of frequency can be seen in [Figure 5d](#). Dielectric properties arise from a variety of different mechanisms: space charge, dipolar, ionic, and electronic polarization. Measuring as a function of frequency allows mechanisms with slower response times, such as space charge polarization arising from ionic conductivity, to be isolated from more meaningful mechanisms. At high frequency (1 MHz) the dielectric response shows a dielectric permittivity

---

<sup>1</sup>The dielectric loss measures dissipation of electromagnetic energy propagating inside a dielectric material to heat. It is defined as the phasor in the complex plane between the real resistive (lossy) and imaginary reactive (lossless) components of the relative permittivity  $\epsilon_{\text{rel}} = \epsilon_{\text{real}} + i\epsilon_{\text{imag}}$  and is commonly given as the tangent of that angle,  $\tan(\delta) = \frac{\epsilon_{\text{imag}}}{\epsilon_{\text{real}}}$

( $\epsilon_{\text{real}}$ ) of 44 and a dielectric loss of  $\tan(\delta) = 0.018$ . The low dielectric loss ( $<0.1$ ) indicates that the value of  $\epsilon_{\text{real}}$  is free from conductive contributions. The permittivity of 44 is similar to doped  $\text{Bi}_2\text{O}_3$  with fluorite-related structures, such as a 10%-doped  $\text{Ta}^{5+}$  with a  $\epsilon_{\text{real}}$  of 42 [34]. However,  $\text{Bi}_2\text{Zr}_2\text{O}_7$  has a higher  $\epsilon_{\text{real}}$  than  $\text{HfO}_2$  or  $\text{ZrO}_2$  ( $\epsilon_{\text{real}}$  between 22-25) fluorites which are used as high-k dielectrics industrially [35], making it a worthwhile material to consider for real-world application. Furthermore, the aqueous-based synthesis with low calcination temperature of  $\text{Bi}_2\text{Zr}_2\text{O}_7$  presents promising opportunities for solution processing of dielectrics which are compatible with existing industrial MOSFET processing technologies.

## 3 Methods

### 3.1 Derivation of $\Phi_{\text{M}}$

Since dielectric constant and band gap are both crucial factors when considering electronic device applications, we measure materials by a figure of merit defined as

$$\Phi_{\text{M}} = E_{\text{gap}} \cdot \epsilon_{\text{tot}} \quad \text{where} \quad \epsilon_{\text{tot}} = \epsilon_{\text{ionic}} + \epsilon_{\text{elec}}. \quad (1)$$

A product ensures materials exhibit at least intermediate levels of band gap *and* permittivity. This follows Y.-C. Yeo, T.-J. King, and C. Hu [36] who define this semi-empirical expression for the leakage current through a MOSFET gate dielectric:

$$J_G \propto \exp \left\{ -\frac{4\pi\sqrt{2q}}{h} \cdot (m_{\text{eff}} \Phi_b)^{1/2} \epsilon_{\text{tot}} \cdot t_{\text{ox,eq}} \right\} \quad (2)$$

with charge  $q$ , effective tunneling mass  $m_{\text{eff}}$  of the electron or hole, injection barrier of the gate dielectric  $\Phi_b$ , and the  $\text{SiO}_2$ -equivalent-capacitance oxide thickness  $t_{\text{ox,eq}} = (\epsilon_{\text{SiO}_2}/\epsilon_{\text{tot}}) \cdot t_{\text{phys}}$ . Increasing  $(m_{\text{eff}} \Phi_b)^{1/2} \epsilon_{\text{tot}}$  exponentially suppresses the tunneling current. Thus MOSFET device miniaturization requires materials that maximize this quantity. The effective tunneling mass  $m_{\text{eff}}$  and the carrier injection barrier  $\Phi_b$  are expensive to compute from first principles and out of reach for high throughput workflows. C. L. Hinkle, C. Fulton, R. J. Nemanich, and G. Lucovsky [37] therefore approximate their product as proportional to the band gap,  $E_{\text{gap}} \propto (m_{\text{eff}} \Phi_b)^{1/2}$ . Increasing  $\Phi_{\text{M}} = E_{\text{gap}} \cdot \epsilon_{\text{tot}}$  should therefore result in exponentially suppressed tunneling current.

### 3.2 Initial Candidate Generation

As shown in Figure 1, we begin our discovery campaign by generating a large set of initial candidates. The Materials Project currently holds 7172 materials with DFPT-calculated permittivity. Starting with the 1000 highest FOM MP dielectric materials, we perform 1000 rounds of elemental substitution on each source structure. Substitutions are guided by a chemical similarity matrix [26] mined from the Inorganic Crystal Structure Database (ICSD) [27], resulting in 1 million potential new structures.

The chemical similarity matrix offers a likelihood score for elemental substitution, based on their co-occurrence in the same space group in ICSD. This approach is inspired by previous works [8], [38]. During substitution, we swap out one element for another across the entire structure and limit ourselves to the 89 elements present in the Materials Project. This process yields 187,176 potential candidates, which we then filter as follows:

1. Remove duplicates:  $10^6 \rightarrow 187.2 \text{ k}$
2. Exclude structures containing rare earths (lanthanides and actinides):  $187.2 \text{ k} \rightarrow 133.4 \text{ k}$
3. Exclude structures containing noble gases:  $133.4 \text{ k} \rightarrow 133.2 \text{ k}$

#### 4. Remove existing Materials Project compositions: 133.2 k $\rightarrow$ 131.7 k

We remove rare earths because DFT is well-known to struggle with the 4f electrons [39], making any DFPT on such compounds less reliable. We filter noble gases because they are chemically inert and hence unlikely to occur in stable compounds. We remove structures with matching compositions in the Materials Project since many MP structures are sourced from the ICSD and hence the experimentally observed ground state. As such, any structures we generate with polymorphs in MP have an increased risk of being metastable at best.

### 3.3 Training Data

We trained the Wren ensembles for formation energy and band gap on the combination of two large datasets:

- The **Materials Project (MP) database** [25] is a well-curated database of high-throughput DFT calculations. We used MP database version 2020-09-08 powered by `pymatgen` version 2022.0.8 containing 146,323 crystal structures [40].
- Wang et al. [26] calculated energies and properties for a large number of crystal structures generated from MP source structures via elemental substitution with chemically similar elements as pioneered in [38]. After substitution, the structures were relaxed using MP-compatible workflows. Using the author’s initials, we refer to this as the **WBM data set**. After de-duplication and cleaning, WBM contains 220k structures.

Together, MP and WBM provide 319601 formation energies, 319601 band gaps. The Materials Project also contains dielectric properties for 7172 materials which we used to train Wren ensembles that predict ionic and electronic permittivity.

### 3.4 Machine Learning

To predict formation energy, band gap and permittivity in the ML pre-filtering step for each of the 131,685 generated candidate materials (Section 3.2), we utilize Wren ensembles [8] which use a coarse-grained Wyckoff position-based material representation that discards exact atomic coordinates in favor of discrete, enumerable symmetry labels identifying groups of sites that map onto each other under the crystal’s symmetry operations. Each Wyckoff position is embedded into a vector space and concatenated with the crystal site’s Matscholar element embeddings [41] before being placed in a fully connected graph with all other Wyckoff sites. Each node in the graph is then allowed to contextualize to its neighbors via multiple message-passing layers and finally mean-pooled to get a permutation- and relaxation-invariant, fixed-length, symmetry-aware crystal descriptor which is much cheaper to obtain than relaxed atom positions. A simple feed-forward net with skip connections [42] and ReLU [43] activations then maps the Wren crystal embedding onto one or multiple target variables. This featurization becomes more informative with higher symmetry in the structure. For our use case of filtering out unrelaxed structures immediately after elemental substitution, its distinct advantage is invariance under structure relaxations as long as the relaxation does not affect the structure’s symmetry (many DFT relaxations enforce keeping the initial symmetry throughout the relaxation, e.g. by setting `ISYM > 0` for VASP).

For each of the four material properties of interest – formation energy, ionic and electronic dielectric constants, and band gap – we train Deep Ensembles [44] of 10 independent Wren models. Trained on identical data but with different initializations, these ensembles offer two advantages:

- The ensemble average yields more reliable point estimates compared to single models.

- Ensemble variance allows us to assess epistemic model uncertainty, which we incorporate into a risk-aware figure of merit via error propagation. This reduces false positives at the cost of increased false negatives.

The ensemble-risk-aware figure of merit  $\Phi_M^{\text{std-adj}}$  including uncertainty propagation reads:

$$\Phi_M^{\text{std-adj}} = \sqrt{\left(\varepsilon_{\text{tot}} \cdot \sigma_{E_{\text{gap}}}\right)^2 + \left(E_{\text{gap}} \cdot \sigma_{\varepsilon_{\text{tot}}}\right)^2}, \quad (3)$$

where  $\varepsilon_{\text{tot}}^{\text{Wren}}$  and  $\sigma_{\varepsilon_{\text{tot}}}$  are the Wren ensemble mean and standard deviation for the predicted total dielectric constant. Likewise  $E_{\text{gap}}^{\text{Wren}}$  and  $\sigma_{E_{\text{gap}}}$  are the ensemble mean/std. dev. for the Wren-predicted band gap. We use the  $\Phi_M^{\text{std-adj}}$  (rather than the standard  $\Phi_M$ ) to rank element substitution structures for priority when allocating compute budget for DFPT calculations.

Moreover, for the formation energy ensemble, we also estimate aleatoric uncertainty, i.e. uncertainty that is inherent to the data, by using a “robust” loss function. This loss requires changing the final output layer of each model to predict two numbers per sample. The loss function interprets the first number as the predicted mean and the second as predicted aleatoric uncertainty. This uncertainty enters the loss function as an attenuation term on the  $L^p$  norm. This allows the model to deweight the loss on predictions it attributes higher uncertainty to at the cost of incurring a higher regularization penalty.

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\sigma(\mathbf{x}_i)^2} \|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|^2 + \frac{1}{2} \log \sigma(\mathbf{x}_i)^2 \quad (4)$$

where  $\mathbf{x}_i$  are the model inputs,  $\mathbf{y}_i$  the corresponding target value,  $\mathbf{f}(\mathbf{x}_i)$  the model predictions and  $\sigma(\mathbf{x}_i)$  is the observation noise parameter, the second predicted by the model. The observation noise is learned by the model as a function of the input, making the loss heteroscedastic (i.e. sample-dependent). Thus the model can learn to deweight the standard  $L^2$  loss in the first term by increasing the predicted observation noise.

We do not set a robust loss on the other 3 ensembles due to an increase in validation error which we did not observe for the formation energy ensemble.

For all 4 Wren ensembles (formation energy, band gap, ionic + electronic dielectric constant), we adopt the same hyperparameters as [8] to which we refer for details on the model architecture. In summary, each ensemble member consists of 3 message passing layers, each with a single attention head. Both parts of the soft-attention mechanism use single-hidden layers with 256 hidden units and LeakyReLU activation functions. The output network following the message-passing layers is a simple feed-forward net with skip connections and ReLU activation functions. Its 4 hidden layers have sizes 64, 256, 256, and 1, respectively.

### 3.5 DFT Structure Relaxation

We used the Vienna ab-initio Simulation Package (VASP) [45], [46] in projector augmented wave (PAW) mode [47] to relax artificial crystal structures generated via elemental substitution of known structure prototypes. The exchange-correlation energy was computed in the generalized gradient approximation (GGA) [48] using the Perdew-Burke-Ernzerhof (PBE) functional [49]. Input files were auto-generated by pymatgen [50]. To perform high-throughput DFT, we used the Materials Project workflow library atomate [51], the job launcher, queue manager and progress monitor Fireworks [52], and the automatic error handler Custodian [50]. Structures were relaxed until all interatomic forces fell below  $10^{-2}$  eV/Å and the total energy change between self-consistent field (SCF) cycles fell below  $10^{-7}$  eV.

### 3.6 Dielectric Properties from DFPT

Candidates that pass our ML filters are fed into high-throughput density functional perturbation theory (DFPT). This stage offers more accurate property estimates at 3-4 orders of magnitude increase in computational cost.

We computed the electronic permittivity using linear response theory at the generalized gradient approximation (GGA) level of density functional perturbation theory as implemented in VASP 6.2.0. High-throughput calculations were orchestrated on the Cambridge CSD3 cluster using the `wf_dielectric_constant` workflow in `atomate` v1.1.0 [51]. This yields Born effective charges and phonon modes at the  $\Gamma$  point [53].

We depart from standard MP dielectric settings in several respects. First, by using the (at the time) most recent `PBE_54` release of VASP POTPAW pseudopotentials (MP uses `PBE`). We used the default structure-dependent  $k$ -point grid as implemented in `pymatgen` which constructs Gamma-centered meshes for hexagonal and face-centered cells, and Monkhorst-Pack grids otherwise. However, we increased the grid density to 3000  $k$ -points per atom despite the significant cost increase for a high-throughput workflow to accommodate the sensitivity of linear-response calculations to  $k$ -point sampling. We also set tight convergence criteria of `EDIFF` =  $10^{-7}$  eV (default =  $10^{-5}$  eV) and a high kinetic energy cutoff for the plane wave basis set of `ENCUT` = 700 eV (default = 520 eV). We expect these changes to increase the fidelity of our results, or at worst, increase compute cost at no benefit.

The total dielectric tensor splits into ionic ( $\varepsilon^0$ ) and electronic ( $\varepsilon^\infty$ ) contributions:

$$\varepsilon_{ij}^{\text{total}} = \varepsilon_{ij}^0 + \varepsilon_{ij}^\infty \quad (5)$$

with  $i, j \in x, y, z$  the 3 spatial dimensions and 0 and  $\infty$  representing the electric field frequency. The ionic contribution  $\varepsilon^0$  is computed from the Born effective charges  $Z^*$  and the phonon modes  $\omega$  [54],

$$\varepsilon_{ij}^0 = \frac{4\pi}{\Omega} \sum_m \frac{Z_{m,i}^* Z_{m,j}^*}{\omega_m^2} \quad (6)$$

with  $\Omega$  the unit cell volume,  $m$  the phonon mode index,  $\omega_m$  the infrared phonon frequency of mode  $m$  and  $Z_{m,i}^*$  the  $i$ th component of the Born effective charge of mode  $m$ .

The scalar dielectric constant that enters our figure of merit equation eq. ([eq:fom] 1) is the mean of the eigenvalues  $\varepsilon_i$  of the total dielectric tensor:

$$\varepsilon_{\text{tot}} = \frac{1}{3} \sum_{i=1}^3 \varepsilon_i^{\text{total}} \quad (7)$$

The ionic contribution  $\varepsilon_{\text{ionic}}$  to the permittivity is known to be sensitive to low-frequency phonon modes which are incorrectly softened by the lattice parameter overestimation typical for GGA [3], resulting in higher mean error than LDA. We chose to run GGA DFPT despite this known GGA shortcoming to retain compatibility with existing MP dielectric data [4], [24].



### 3.7 Web Interface for Collaborative Synthesis Selection

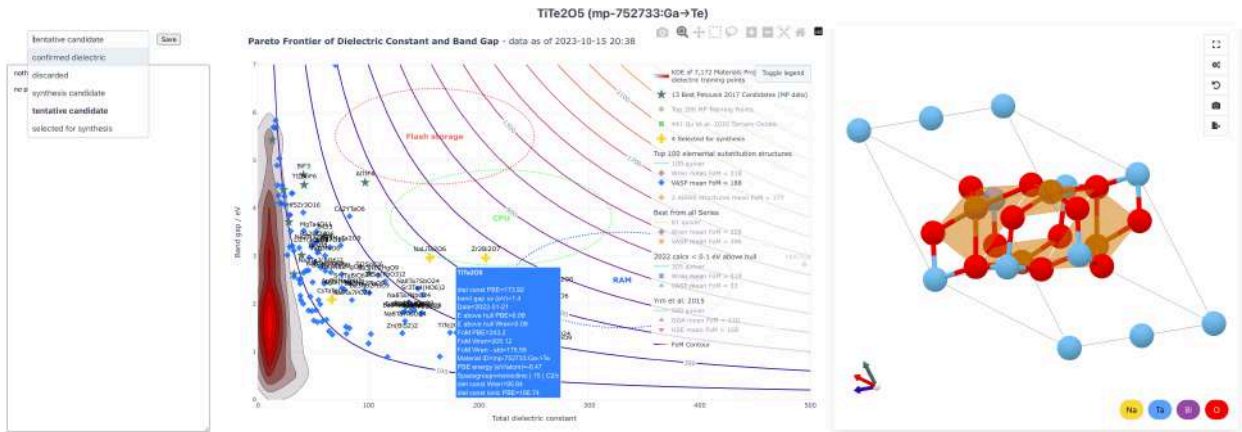


Figure 6: Screenshot of the [web app](#) that aids with synthesis selection. The centerpiece of the app is an interactive scatter plot similar to [Figure 3](#) showing the Pareto front of band gap and total dielectric constant. The legend on the right enables toggling between our own data set and data from prior works discussed in [Section 5](#). For our own data, we have legend groups for different calculation batches and  $\Phi_M$ -based subsets of the data that allow switching between viewing ML predictions and/or their corresponding DFPT results with a third option to show a quiver plot that renders arrows between these points. This makes it easy to visualize how ML predictions differ from DFPT results and to look for trends in ML errors w.r.t. chemistry. Ellipses again indicate regions of particular interest for specific device applications (CPU, RAM, Flash storage). The density contours show lines of constant figure of merit.

The last step in our discovery workflow before experimental synthesis involves a custom-built web interface shown in [Figure 6](#) powered by a MongoDB Atlas M2 instance on the backend. This database is automatically updated when new **atomate** [\[51\]](#) DFPT workflows finish on our compute cluster. To implement the frontend, we leveraged multiple open-source technologies:

- **pymongo** [\[55\]](#) for fetching the latest calculation results from our **atomate** `tasks` collection.
- **plotly** powers the interactive scatter plot used to show computed and/or ML-predicted band gaps and dielectric constants, switch between different calculation series or best-of subslices of the data as well as clicking points to select individual materials for closer inspection.
- **CrystalToolkit** [\[56\]](#) renders the 3d structure of the material selected in the scatter plot with pan and zoom functionality.
- **dash** [\[57\]](#) stitches the above 3 components together with callback functions. The two main ones are updating the structure viewer whenever a new point is selected from the scatter plot and updating the database when new free-form notes or categorizations are recorded in the text area and dropdown menu on the left.

In our case, selecting individual points from the scatter plot and annotating them with free-form text took place during remote live discussions between theorists and experimentalists while screen sharing. Since these meetings took place over months, the web app massively helped with keeping track of reasons for categorizing a given material as discarded or tentative/firm synthesis candidate or recording links to prior art for materials categorized as already confirmed dielectrics. Given this web app proved an enabler of effective remote collaboration between computational and experimental labs, we emphasize the importance of developing more custom tools that improve information flow and data visualization. Moreover, we found our tool significantly facilitates the

process of keeping provenance. Ideally, this process should be automated entirely in the future as this area is extremely prone to human error.

The final verdict of these discussions results in a classification as one of

**confirmed dielectric** prior experimental literature exists confirming our candidate material to be a dielectric. No point in synthesizing and re-characterizing, but increases trust in our workflow.

**selected for synthesis** Promising in every way, i.e. high calculated band gap and permittivity, cheap and easily accessible precursors, synthesis procedure matches our experimentalists’ area of expertise and has ideally been demonstrated in earlier experimental works but without dielectric characterization.

**strong candidate** promising in some ways, i.e. high calculated band gap and permittivity but perhaps no existing literature reporting successful prior synthesis or compound looks challenging to make (e.g. might require aerobic environment)

**weak candidate** less promising in terms of simulated properties but potentially easier to make than other materials with superior expected properties

**discarded** failures of our screening method, usually due to existing literature indicating properties are not as we predict such as when a material was previously synthesized but reported as black, indicating a small band gap.

This interactive selection tool proved invaluable for extracting maximum utility and insight from the data we generated and resulted in identifying two candidates for final selection as suitable candidates for experimental synthesis.

We use GitHub Pages to host a figure-only version of this web interface at [janosh.github.io/dielectrics](https://janosh.github.io/dielectrics). It is set up with continuous integration to update automatically as new data is generated. It has no write access to the database and hence cannot be used to annotate or categorize candidate materials but serves as a user-friendly public entry point to the most promising results in our database that requires no setup nor technical knowledge to use.

### 3.8 Synthesis Details

CsTaTeO<sub>6</sub> was synthesized using standard solid-state synthesis techniques. Stoichiometric amounts of Cs<sub>2</sub>CO<sub>3</sub> (Alfa Aesar, 99.95), Ta<sub>2</sub>O<sub>5</sub> (Alfa Aesar, 99.999), and Te(OH)<sub>6</sub> (Aldrich, 99.5) were added to an agate pestle and mortar and ground to homogenize the precursors before calcining at 400 °C for 24 h in an Al<sub>2</sub>O<sub>3</sub> crucible. After calcining, samples were reground and pressed into a 10 mm disk and annealed at 750 °C for 48 h in a covered Al<sub>2</sub>O<sub>3</sub> crucible to form the final product.

Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub> was synthesized using an ethylenediaminetetraacetic acid (EDTA) and nitrate chelation and combustion, similar to a sol-gel process, a reaction modified from [31]. Equal molar quantities of Bi<sub>2</sub>O<sub>3</sub> and ZrO(NO<sub>3</sub>)<sub>2</sub> were added to separate beakers and dissolved in minimal amounts of concentrated nitric acid by stirring with a magnetic stir bar. Once dissolved, these two solutions were mixed with a 4 times molar excess of EDTA to ensure chelation. The solution was then heated at 80 °C until all liquid evaporated, leaving a brownish-white powder that was amorphous to X-rays. The amorphous powder was then calcined as a loose powder in a furnace inside a Al<sub>2</sub>O<sub>3</sub> crucible at temperatures from 550 °C–750 °C in 50 °C increments for 1 h. The samples heated at 650 °C produced the sharpest XRD peaks, without any trace of impurities. Samples heated higher than 650 °C or for longer than 1 h resulted in the decomposition of the sample into Bi<sub>2</sub>O<sub>3</sub> and ZrO<sub>2</sub>, indicating metastability.

Both samples were sintered using spark plasma sintering. Pure phase samples were loaded into 10 mm graphite dies in a Thermal Technology LLC DCS10 furnace. Samples ( $\sim 0.75$  g) were loaded into a 10 mm diameter graphite die lined with a graphite foil and loaded into a sample chamber which was evacuated and backfilled with He three times. The sample was pressed uniaxially at 60 MPa, heated to the desired temperature at a rate of  $200^\circ\text{C}/\text{min}$ , held for 1 min, and cooled at the same rate. The  $\text{CsTaTeO}_6$  sample was heated to a maximum temperature of  $750^\circ\text{C}$  and the  $\text{Bi}_2\text{Zr}_2\text{O}_7$  sample was heated to  $600^\circ\text{C}$  resulting in samples with 92% and 94% of theoretical densities, respectively.

Diffuse reflectance measurements were taken on powdered  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$  using a Cary 5000 UV–Vis–NIR Spectrometer. Dielectric permittivity data was collected on sintered samples that had been thinned to a thickness of 1 mm and sputtered with gold electrodes. Data was collected using an Agilent 4980A instrument with a home-built sample holder and a program created in LABVIEW. X-ray diffraction data was collected using a Paralytical X’pert Pro instrument with  $\text{Co K}\alpha 1$  ( $\lambda = 1.789 \text{ \AA}$ ) radiation. Rietveld analysis was carried out using Topas Academic on these X-ray data. Initial refinements started with parameters identified using the Pawley method. Final refinements included lattice parameters, atomic positions, atomic displacement parameters, profile parameters and the background.

## 4 Discussion

We have demonstrated a high-throughput workflow for dielectric materials discovery that combines data-driven and first-principles methods. We show in [Table 1](#) that this combination achieves improved enrichment of high  $\Phi_{\text{M}}$  materials than ab-initio methods alone.

By deploying this workflow into practice, we identified and synthesized two candidate materials,  $\text{CsTaTeO}_6$  and  $\text{Bi}_2\text{Zr}_2\text{O}_7$ . After careful Rietveld analysis to verify we realized the target structures, we measured their band gaps and dielectric properties.  $\text{Bi}_2\text{Zr}_2\text{O}_7$  shows strong promise for electronic applications given its measured band gap of 2.27 eV, dielectric constant of 20.5 and its relatively available constituent elements.  $\text{CsTaTeO}_6$  is a black semiconductor with a low band gap of 1.05 eV and dielectric constant of 26, making it unsuitable for electronic applications. However, we emphasize this structure was generated via element substitution by our workflow with no prior reports in the ICSD or MP. We thus demonstrated successful de novo synthesis on a challenging metastable phase and established a prior for the dielectric properties of similar materials in this largely unexplored region of chemical space. This outcome shows that ML-driven thermodynamic stability prediction has matured enough in reliability to be effectively incorporated into a complex multi-step workflow. This requires sufficient trust in the method to attempt a risky metastable synthesis in an unknown chemical system.

The biggest failure mode in our funnel search was the weakness of our band gap ML model. It incurred a high false-positive rate, predicting many generated metallic structures as semiconductors or insulators. Although there is significant room for improvement in ML band gap prediction, it was not the main focus of this work. We consider accurate band gap models to be an unsolved problem in materials informatics and encourage more efforts be directed at it. Models that predict a spectrum rather than a single scalar may be an interesting avenue to pursue. Predicting the electronic density of states (eDOS) like Mat2Spec [58] and inferring the band gap from that also opens the door to more nuanced loss functions and increased regularization during training. Sufficiently complex models with good inductive bias may learn more subtle trends from this approach. It should be noted, however, that Mat2Spec refrained from reporting band gaps inferred from their eDOS predictions, potentially indicating more work is required to unlock such benefits. N. Shoghi, A.

Kolluru, J. R. Kitchin, Z. W. Ulissi, C. L. Zitnick, and B. M. Wood [59], [59] is a more recent work demonstrating impressive band gap accuracy on the matbench MP  $E_{\text{gap}}$  task after pre-training on many large but non-cognate materials prediction tasks. This suggests that perhaps current model architectures and training methods can be sufficient. Achieving reliable ML band gap prediction could be a matter of careful data curation and model pre-training.

However, the challenge of predicting band gaps in our workflow is not restricted to ML but carries through to DFT. PBE exhibits an unusual severe overestimation ( $E_{\text{gap}}^{\text{PBE}} > 2.09 \text{ eV}$ ) of the experimental band gap of  $\text{CsTaTeO}_6$  ( $E_{\text{gap}}^{\text{exp}} = 1.05 \text{ eV}$ ). Although defect chemistry may play a role in this effect, there are obvious computing limitations in a high-throughput workflow, making the simulation of defect effects cost-prohibitive. One obvious improvement to narrow the gap between simulation and reality is to employ higher levels of theory such as r2SCAN or even to incorporate a third computational filter to the funnel in the form of hybrid functionals such as HSE, applied sparingly to compounds that have passed ML and PBE filters but before attempting experimental synthesis.

## Code and Data Availability

The MIT-licensed code for this work can be found at <https://github.com/janosh/dielectrics> and as a Zenodo archive at <https://doi.org/10.5281/zenodo.10456384>. Zenodo includes a complete dump of our DFPT dataset. Our live data is also publicly accessible through a MongoDB M2 Atlas instance with the schema of an `atomate tasks` collection. It can be queried free of charge and without registration using the read-only database credentials and example code snippet provided in the GitHub readme. This requires `pymongo` or any other MongoDB language driver. The query syntax will be familiar to users of the (legacy) Materials Project `MPRester` API. We used Materials Project data from the (v2020.09.08) database release and a cleaned version of the WBM dataset [26] available at <https://figshare.com/articles/dataset/22715158>.

## Acknowledgements

J.R. acknowledges support from the German Academic Scholarship Foundation ([Studienstiftung](#)). A.A.L. acknowledges support from the Royal Society.

## Supplementary Information

### 5 Related Work

While previous studies have made significant strides in automating high-throughput DFPT to uncover new dielectrics, our work diverges in 3 important regards. We prefix DFPT with generative and pre-filtering ML which allows us to consider a much larger initial candidate pool as well as venture into uncharted regions of material space in our search for high dielectrics. Using ML-preselection and biasing the structure generation to crystals similar in chemistry to known high dielectric materials in MP allows us to nonetheless maintain a higher hit rate of materials with high  $\Phi_{\text{M}} > 240$  than previous works as shown in [Table 1](#). Third, we built a web UI that enabled effective collaboration with experimentalists to select 2 promising candidates which we successfully synthesized and characterized.

To our knowledge, Yim et al. [3] were the first to develop codes that fully automate ab-initio calculation of band gaps and dielectric permittivities. They calculated 1800 structures of binary

and ternary oxides from the ICSD to generate a dielectric property map which confirmed the inverse correlation between band gap and permittivity for most oxides, with occasional outliers that exhibit both large permittivity despite large band gaps.

I. Petousis *et al.* [24], [24] calculated electronic and ionic dielectric tensors for 88 compounds to test the predictive power of DFPT against experiment for total dielectric constant and refractive index. While they observed a Mean Average Deviation (MARD) of 16.2% when using PBE as compared to LDA, they noted that DFPT is less accurate for compounds with complex structural effects or strong anharmonicity. Their results, however, showed a high Spearman correlation factor of 0.92, demonstrating the utility of DFPT in identifying promising materials by ranking.

The following year, I. Petousis *et al.* [24] extended their previous work by running high-throughput DFPT on 1,056 inorganic compounds. The resulting database of dielectric tensors was integrated into the Materials Project for public access. While this greatly improved explorability of the data and likely may have helped expand the search pool for experimentalists seeking synthesis candidates, the scale of the data remained too limited to cover more than a small fraction of compositional and even less of the configurational space of potential high dielectrics.

While the above works resulted in novel and promising candidate materials, they relied exclusively on expensive DFPT calculations, making truly high-throughput screening of hundreds of thousands of materials cost-prohibitive. Yet they produced a sizeable pool of DFT dielectric properties with which we are now able to train ML models to accelerate and amortize the high cost of DFPT in the search for dielectrics, allowing screening of a much more expansive chemical space.

## 6 Synthesis Development and Structure Fitting

$\text{Bi}_2\text{Zr}_2\text{O}_7$  is known and has seen research interest for its use as a photocatalyst [60], [61], [62]. In these reports, the compound has been said to have either a stoichiometric pyrochlore structure ( $\text{A}_2\text{B}_2\text{O}_7$ ) [31], [61], [62], [63], [64], [65] or the structurally related defect-fluorite structure [60], [66], [67], [68], [69]. Our results show that a pyrochlore could not be isolated without additional  $\text{Bi}_2\text{O}_3$  or  $\text{ZrO}_2$  impurities due to the metastable nature of this compound. Though the pyrochlore and fluorite structures yield similar XRD patterns, with the most intense peaks located in the same positions, the absence of the (111) peak at  $Q = 1.01 \text{ \AA}$  favors assignment of the fluorite structure, Figure 4e.

The XRD data was fit using Rietveld refinement. Attempts were made to fit the data with a pyrochlore structure. When using both a standard pyrochlore model and models with oxygen and  $\text{Bi}^{3+}$ , displacive disorder produces calculated patterns that fail to fit the data properly. Intensity mismatch is observed for low-angle pyrochlore peaks, specifically the (111) reflection. No amount of disorder was sufficient to reduce the intensity of this peak in the model to noise levels in the data, further confirming that this compound does not crystallize as a pyrochlore.

Using a defect-fluorite structure ( $\text{Bi}_{0.5}\text{Zr}_{0.5}\text{O}_{1.75}$ , Figure 4f results in rapid model-to-data convergence with a good visual fit, Figure 4e. The resultant model shows atomic displacement parameters of  $3.28(17) \text{ \AA}^2$  for the cations and  $8.4(4) \text{ \AA}^2$  for the oxygen, which are large. Large atomic displacement parameters are commonly found in disordered compounds, and experimentally observable in the form of broad diffraction peaks, compared to Figure 4a. Attempts to account for the disorder in this compound in our structural model were not successful. Splitting the position of Zr and Bi to account for chemical displacements off their position in the center of the cubic polyhedra resulted in the cations refining back to their undisplaced positions. The same process was used for the oxygen positions but resulted in much larger atomic displacement parameters, leading us to discount this



distortion. The occupancies of sites were also refined, resulting in the cations maintaining a 1:1 ratio, within error. This allowed us to conclude that a simple defect-fluorite structure is the most sensible model. This final model can be seen in Figure 4f and an isolated Zr/BiO<sub>8</sub> polyhedra can be seen in Figure 4g. This model produced sensible metal-oxygen bond lengths of 2.3160(5) Å, which were expectedly longer than ZrO<sub>2</sub> bond lengths of 2.25 Å.

## 7 Synthesis Development and Structure Fitting

The targeted CsTaTeO<sub>6</sub> pyrochlore compound was initially investigated as it both met the figure of merit criterion and had not been reported previously in the ICSD or MP. However, we did find mention of this compound and its crystallographic analysis in [70] after completing synthesis and characterization. Moreover, a related defect-pyrochlore with composition CsNbTeO<sub>6</sub> had been reported in [33], [71] from which we extracted initial synthesis parameters. With minor modifications of the synthetic procedure, the new compound was isolated in high purity, with only a 4.20 wt% Ta<sub>2</sub>O<sub>5</sub> impurity.

Figure 4a shows XRD data of the final CsTaTeO<sub>6</sub> product. This pattern indexes readily to the symmetry and lattice parameters of a cubic defect-pyrochlore (Figure 4b), consistent with both the Nb<sup>5+</sup>-based analog and the computational predictions. Rietveld refinements were initiated using parameters taken from Pawley fitting and readily converged to a defect-pyrochlore structural model. The structural model was taken from the refinement of CsNbTeO<sub>6</sub> which places Cs on the larger site 8b (Figure 4c) site and the Ta<sup>5+</sup> and Te<sup>6+</sup> in equimolar amounts on the 16c site (Figure 4d). This formulation is that of a defect-pyrochlore (AB<sub>2</sub>X<sub>6</sub>), which is distinct from the traditional A<sub>2</sub>B<sub>2</sub>X<sub>6</sub>Y pyrochlore structure. Relative to a traditional pyrochlore, this structure has both cation and anion vacancies, while maintaining the same anion packing and BO<sub>6</sub> connectivity. After refining all parameters simultaneously, a good visual fit to the data is obtained with sensible atomic positions, sensible atomic displacement parameters 0.087(15) – 0.78(2) Å<sup>2</sup>, a lattice parameter (10.29894(5) Å) close to that of the Nb<sup>5+</sup> analog (10.288 Å), and a fit quality parameter (Rwp = 8.095%) approaching that of the minimum set by the Pawley Fit (Rwp = 7.399%) [72].

Due to the defect nature of this pyrochlore formulation, the Cs<sup>+</sup> (A-site) adopts an octahedral polyhedral environment (Figure 4c) with six equal bond lengths of 3.183(6) Å, instead of the cubic (AO8) environment found in stoichiometric A<sub>2</sub>B<sub>2</sub>O<sub>7</sub> pyrochlores. The observed bond lengths are consistent with AO<sub>6</sub> polyhedral environments seen in other Cs<sup>+</sup> pyrochlores such as CsNbTeO<sub>6</sub> or CsMoTeO<sub>6</sub> which range from 3.180 – 3.421 Å [72], [73]. The Ta<sup>5+</sup> and Te<sub>6+</sub> occupy the smaller octahedral environment (Figure 4d) found in traditional and defect-pyrochlores. This environment generates bond lengths of 1.9430(18) Å, again falling within the expected range of related materials such as the Nb<sup>5+</sup> and Mo<sup>5+</sup> analogs previously mentioned, compounds that range from 1.941 – 2.013 Å. This consistency of the structural environments found in CsTaTeO<sub>6</sub> with similar chemistries further validates the quality of our model.

## 8 Tradeoffs in Dielectric Materials for Computing Applications

As indicated by the shaded regions in Figure 7, while ideal dielectric materials all push into the top right of this plot, different applications have different requirements. Materials for flash storage require especially large band gaps to minimize leakage current and maintain polarization over extended periods. CPU gate dielectrics trade off lower band gaps in exchange for increased permittivity which lowers the gate voltage required to achieve polarization and hence decreases



power consumption. For random access memory (RAM) applications, increased leakage current resulting from a lower band gap is acceptable since RAM is memory-refreshed hundreds of times a second (stored data is read and immediately rewritten unmodified to preserve integrity to avoid polarization sapping over time). Instead, optimal RAM performance relies on exceptionally high permittivity so that each repolarization costs minimal energy. Our goal is to discover materials in any of these regions beyond the green isoline ( $\Phi_M = 240$ ). Figure 8 shows that the principal contributions to the permittivity are due to the ionic permittivity of the materials rather than their electronic permittivity.

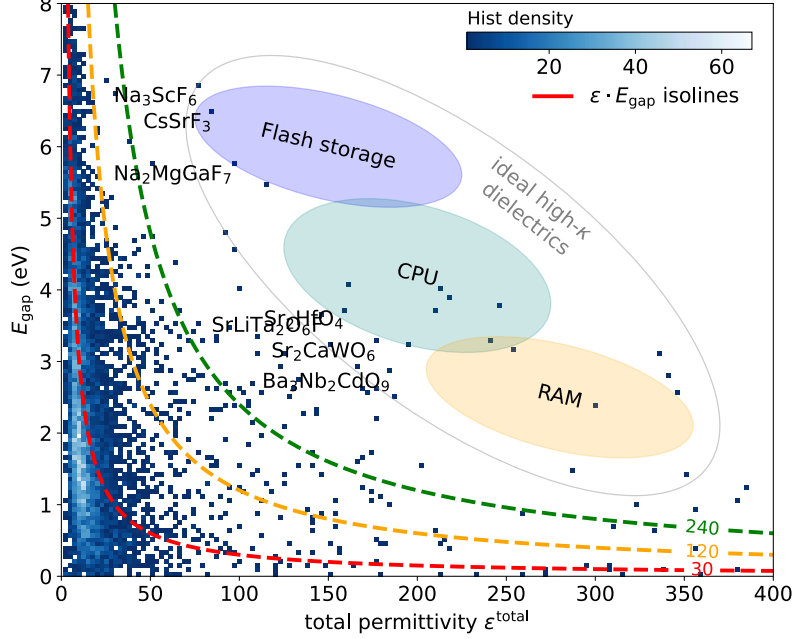


Figure 7: 2d histogram showing the  $1/x$  relationship between band gap and dielectric constant for 7.2k MP materials. The dashed isolines represent levels of constant figure of merit ( $\epsilon_{\text{tot}} \cdot E_{\text{gap}}$ ). The colored ellipses highlight the optimal trade-offs between band gap and permittivity for specific device applications. See Figure 8 for the same plot split by electronic and ionic contributions to the permittivity.

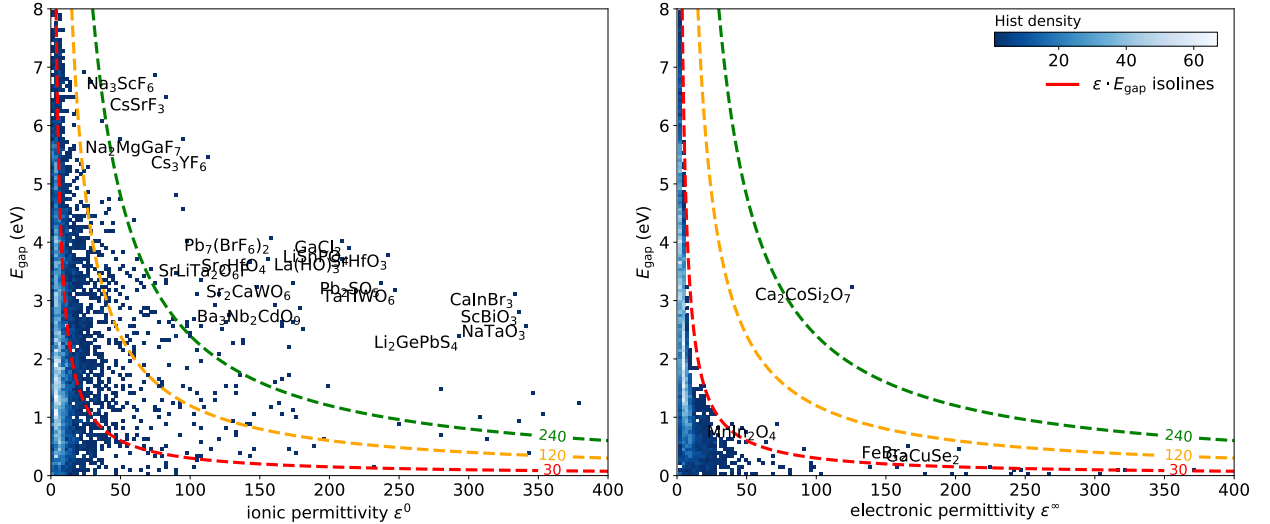


Figure 8: Ionic (left) and electronic (right) parts of the dielectric constant. Ionic permittivity makes a much larger contribution than the total compared to electronic permittivity and is also much more likely to break the  $1/x$  relationship with band gap.

## 9 DFPT Validation

To validate our DFPT results, [Figure 9](#) compares our ab-initio results generated using the `wf_dielectric_constant` workflow in `atomate` [51] against available experimental dielectric constants collected in [4], [24]. While we achieve better agreement with experiment than Petousis as indicated by the lower MAE of 16.5 (vs 20.4) and higher  $R^2$  of 0.41 (vs 0.0), and similar performance to MP (MAE = 14.9,  $R^2 = 0.12$ ), we incur a slightly larger fraction of outliers than either of them at 14% (vs 9% and 10%, respectively). We define outliers as points with absolute relative deviation greater than  $\pm 50\%$  relative to experiment. The reason we nonetheless achieve higher  $R^2$  is due to the lack of extreme outliers; we have more but they are less severe. This is advantageous in high-throughput settings where the goal is to guide experiment. Even rare cases of extreme outliers will show up given sufficient throughput and extreme permittivity overpredictions are more likely to result in wasted experimental effort.

We note that while the data in MP was generated with the same `atomate` workflow as designed and benchmarked by I. Petousis *et al.* [24], our data is expected to deviate from MP/Petousis due to our departure in choice of VASP parameters described in [Section 3.6](#), most notably the use of PBE\_54 POTCARs, increased  $k$ -point density of 3000 points per atom, increased ENCUT = 700 eV plane wave energy cutoff and decreased EDIFF=  $10^{-7}$  eV SCF convergence criterion. All of the above, though most notably the newer pseudopotentials may explain the less extreme outliers with respect to experiment. Overall, the variations with respect to MP/Petousis are within reason for run-to-run variability using slightly modified settings.

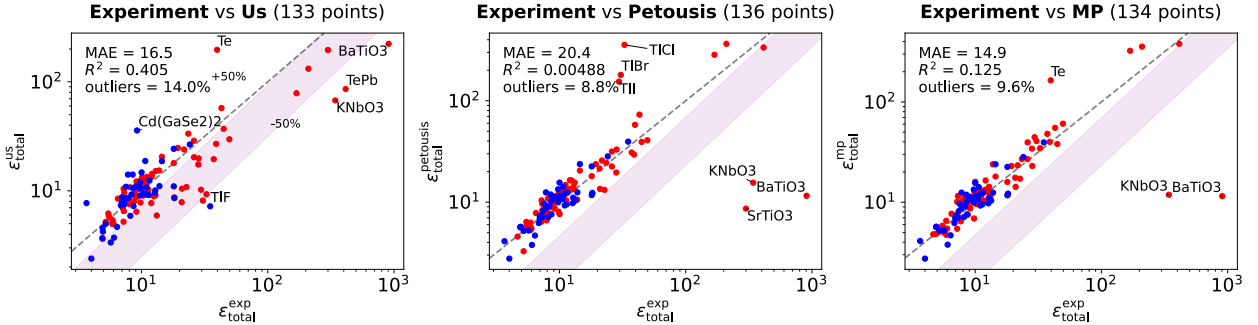
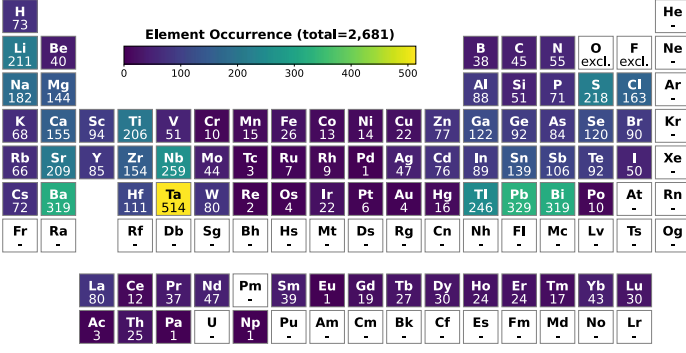
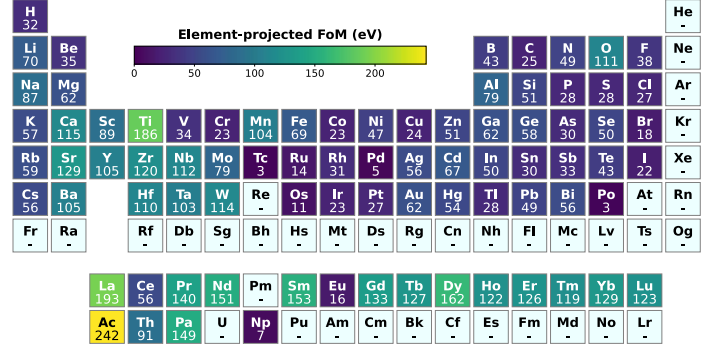


Figure 9: Comparison of experimental and DFPT-computed values for total permittivity  $\epsilon_{\text{tot}}$ . Our data shows lower MAE and higher  $R^2$  but more outliers (defined as points with  $> 50\%$  error) compared to Petousis *et al.* Comparing our DFPT dielectric constants with experimental values, we achieve an MAE of 16.5 and  $R^2$  of 0.4 while MP results attain a slightly lower MAE of 14.9 and  $R^2$  of 0.125. A CSV file with the plotted experimental data is available on [GitHub](#).

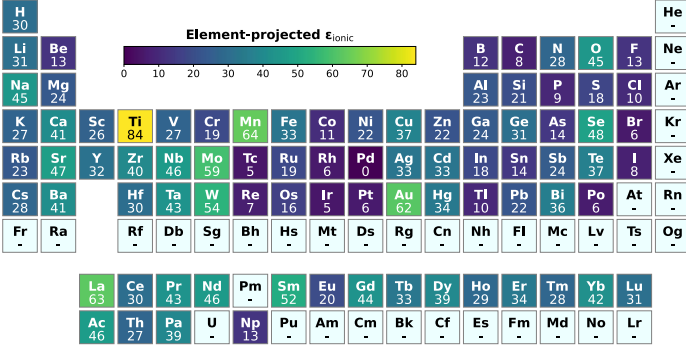
# 10 Exploratory Data Analysis



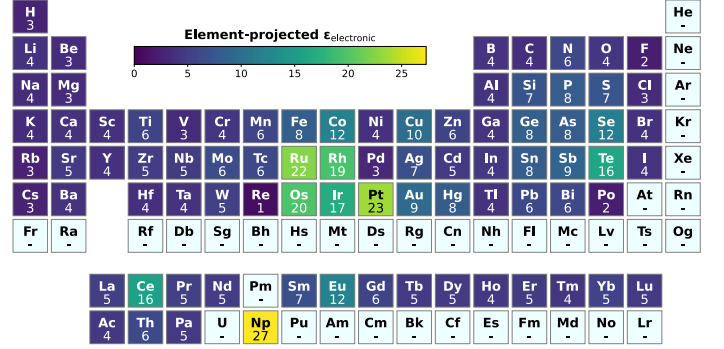
a: Element occurrences



b:  $\Phi_M = \varepsilon_{\text{tot}} \cdot E_{\text{gap}}$



c: Ionic permittivity  $\varepsilon_0$



d: Electronic permittivity  $\varepsilon_\infty$

Figure 10: [Figure 10a](#) Element occurrence counts, i.e. the number of structures among 2532 DFPT results containing a given element. [Figure 10b](#) Figure of merit  $\Phi_M = \varepsilon_{\text{tot}} \cdot E_{\text{gap}}$  projected onto elements by composition and averaged over all 2532 structures. E.g. a  $\text{Fe}_2\text{O}_3$  with a  $\Phi_M$  of 100 would contribute a sample of 40 to the mean heatmap value of Fe and 60 to O. [Figure 10c](#) same as [Figure 10b](#) but for ionic permittivity  $\varepsilon_0$ . [Figure 10d](#) same as [Figure 10b](#) but for electronic permittivity  $\varepsilon_\infty$ . We filtered out 10 untrustworthy calculations with electronic permittivity  $\varepsilon_\infty > 100$ .

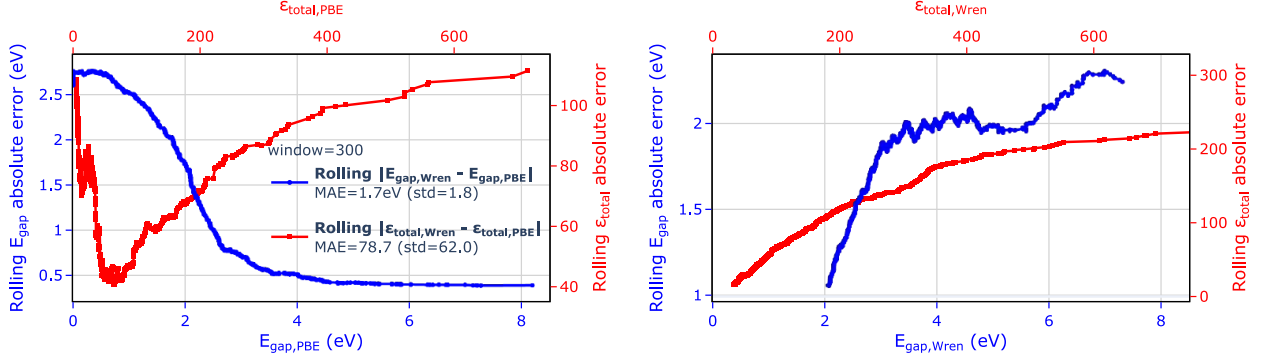
[Figure 10](#) shows the distribution of elements in our DFPT dataset ([Figure 10a](#)) and their element-projected figure of merit  $\Phi_M$  ([Figure 10b](#)) and electronic ([Figure 10d](#)) and ionic ([Figure 10c](#)) permittivities  $\varepsilon_\infty$  and  $\varepsilon_0$ . The most prevalent elements in our dataset are Ta (514), Pb (329), Bi (319), Ba (319), Nb (259) where the number in parentheses is the number of structures containing that element. Our data recovers well-known trends for elements that tend to be present in high dielectrics. In particular, [Figure 10c](#) shows titanium has the highest ionic permittivity when averaged over all Ti-containing structures in our dataset. This matches the prevalence of high dielectric alkaline earth metal titanates such as the perovskites  $\text{BaTiO}_3$  and  $\text{SrTiO}_3$ . [Figure 10d](#) reveals that late transition metals like Ru, Rh, Os, Ir and Pt tend to yield the highest observed electronic permittivities.

[Table 2](#) lists all DFPT results in our dataset with  $\Phi_M > 350$  sorted by  $\Phi_M$ . The highest- $\Phi_M$  materials are almost exclusively oxides with only two fluorides and one selenide in the mix ( $\text{AcF}_3$ ,  $\text{LiY}_2\text{F}_7$  and  $\text{Sm}_2\text{CdSe}_4$ ). Some of the top materials, unfortunately, contain toxic or rare elements (e.g. Cd, Nd, Dy) which are undesirable for environmental, economic and lab-safety/logistic reasons. Others contain lanthanides and actinides, f-block elements which DFT is known to struggle with due to strong electron correlation effects in the atomic-like  $4f$  orbitals near the valence band [39]. Both are strong arguments against attempting experimental synthesis, explaining why we did not simply select the top materials in this list.

	Material ID	Formula	Spacegroup	$\epsilon_{\text{elec}}$	$\epsilon_{\text{ionic}}$	$\epsilon_{\text{total}}$	$E_{\text{gap}}$ (eV)	$\Phi_{\text{M}}$ (eV)	$n_{\text{sites}}$	$n_{\text{elems}}$
1	mp-14550	TiCdO3	62	7.51	709	716	2.12	1,520	20	3
2	mp-997585:La->Y	Y8Al7GaO24	221	4.69	490	495	2.83	1,402	40	4
3	mp-32244:W->Mo	LiNbMoO6	113	5.76	528	534	2.19	1,172	18	4
4	mp-1097026	LaGaO3	221	5.23	305	311	3.42	1,062	5	3
5	mp-754225	YbTiO3	62	6.37	422	428	2.30	986	20	3
6	mp-3335	Sm2Ti2O7	227	6.13	331	337	2.82	951	22	3
7	wbm-1-40021	YbTiO3	62	6.36	387	393	2.33	916	20	3
8	mp-1226157	Cs2Ti(WO4)3	166	5.36	315	321	2.83	908	18	4
9	wbm-3-54931	Sr2LuTaO6	225	4.34	247	252	3.49	878	10	4
10	mvc-3783	MgTiO3	62	6.34	370	376	2.33	877	20	3
11	mp-556003	CaTiO3	74	6.39	384	390	2.16	843	10	3
12	mp-1202153	Na2Nb4O11	9	6.20	301	307	2.74	842	34	3
13	wbm-1-42539	La2MgZrO6	148	4.74	199	204	4.09	832	10	4
14	mp-1217978	SrPrScO4	107	4.64	224	228	3.52	804	7	4
15	mp-556925	TiMnO3	62	7.29	515	523	1.54	803	20	3
16	mp-979932:Sr->Pb	SiPb3O5	140	6.55	387	393	1.93	757	18	3
17	mp-1225952	CsNbWO6	74	5.43	248	254	2.92	741	18	4
18	wbm-1-25816	LiTaO3	167	5.41	223	228	3.21	733	10	3
19	mp-9890	TaAgO3	167	6.71	364	371	1.95	724	10	3
20	mp-754128:Ba->Ca	Ca3Hf2O7	139	4.24	199	203	3.47	704	12	3
21	mp-1219587	RbNbWO6	74	5.26	231	236	2.93	691	18	4
22	mp-1222804	LaY3Ti4O14	166	6.09	233	240	2.81	674	22	4
23	mp-1222933	LaTiNbO6	33	6.06	226	232	2.80	649	36	4
24	mp-755367:Cu->Rb	RbLiTa2O6	155	5.51	265	271	2.36	639	10	4
25	mp-4019	CaTiO3	62	6.39	257	263	2.31	606	20	3
26	wbm-1-38346	CaTiO3	62	6.39	254	261	2.32	605	20	3
27	mp-1218162	SrNd2Ti4O12	123	6.68	333	340	1.73	588	19	4
28	mp-756175	Zr2Bi2O7	227	6.53	200	206	2.80	578	22	3
29	mp-3858	NaTaO3	62	5.26	203	208	2.59	539	20	3
30	mp-755367:Cu->Ag	LiTa2AgO6	155	6.11	260	266	2.02	538	10	4
31	wbm-1-40022	YbTiO3	74	6.34	230	236	2.23	527	10	3
32	mp-1222976	LaTaBi2O7	74	6.39	262	268	1.96	526	22	4
33	mp-39511	LiCaTa2O6F	74	4.49	138	143	3.66	524	22	5
34	wbm-1-38356	CaZrO3	127	4.65	163	168	3.09	519	10	3
35	wbm-1-39241	NaTaO3	62	5.26	193	198	2.59	513	20	3
36	mp-6440	CaTiSiO5	15	4.51	171	175	2.91	509	16	4
37	wbm-3-51397	NdScO3	140	5.14	166	171	2.95	504	10	3
38	mp-1173711	Na3ErTi2Nb2O12	26	6.08	235	241	2.09	502	40	5
39	mp-559482	Ti2Bi2O7	227	8.42	180	189	2.62	495	22	3
40	mp-675778:Na->Li	LiY2F7	12	2.32	68	70	6.99	489	10	3
41	mp-1218781	Sr2La2MgTi3O12	1	5.67	233	239	1.97	471	20	5
42	mp-1220301	NbTiWO6	74	5.86	152	158	2.93	462	18	4
43	mp-755367:Cu->Na	NaLiTa2O6	155	5.32	150	155	2.96	460	10	4
44	mp-754936	DyAlO3	167	4.25	83	87	5.19	453	10	3
45	mp-1199037	NaNbO3	52	5.76	245	251	1.76	441	40	3
46	wbm-3-55347	Ba2ZrTiO6	225	5.68	192	198	2.23	441	10	4
47	wbm-2-35353	Ca2DyTaO6	87	4.46	111	116	3.79	439	10	4
48	mp-1218358	SrCaTi2O6	26	6.40	202	209	2.09	435	20	4
49	mp-38125	Sm2CdSe4	122	10.72	298	309	1.40	433	14	3
50	mp-1227686	Ca3Ta4(O6F)2	166	4.50	113	118	3.57	420	21	4
51	wbm-3-51131	Na2ZrO3	65	3.76	128	131	3.15	414	6	3
52	mp-1223520	KCa2Ta3O10	38	4.09	184	189	2.18	411	16	4
53	mp-5986	BaTiO3	99	5.63	218	223	1.78	397	5	3
54	mp-977360	AcF3	225	3.01	51	54	7.30	395	4	2
55	mp-759812	Li4Ti11O24	12	5.14	166	172	2.30	394	39	3
56	mp-769280	Ca5Ta4O15	164	4.67	127	131	2.98	391	24	3
57	mp-545665	WO3	191	6.14	313	319	1.19	381	12	2
58	mp-1518526	Sr2ZrTiO6	225	5.33	156	161	2.35	378	10	4
59	mp-1227336	BaSrTi2O6	123	6.65	206	212	1.75	371	10	4
60	mp-1227473	Ca2La2MgTi3O12	1	5.66	172	178	2.08	369	20	5
61	mp-1078457	Ba2ZrTiO6	225	5.68	157	162	2.22	359	10	4
62	wbm-1-40008	SrTiO3	99	5.96	190	196	1.80	354	5	3
63	wbm-1-38267	CaHfO3	127	4.39	91	96	3.67	352	10	3
64	mp-756214	YAlO3	167	4.25	66	70	5.00	351	10	3

Table 2: Table of DFPT results with  $\Phi_{\text{M}} > 350$  sorted by  $\Phi_{\text{M}}$ .

# 11 Band Gap Prediction



a: RAE as a function of PBE band gap

b: RAE as a function of Wren band gap

Figure 11: Rolling absolute error (RAE) of Wren band gap and dielectric constant predictions relative to DFPT.

In this screening campaign, we emphasize that substantial challenges remain concerning band gap prediction, due in part to a metal-heavy dataset imbalance and in part to band gaps being an inherently non-local property of the electronic rather than ionic structure. This makes the prediction problem poorly, if not ill-defined for a coarse-grained structural model with no concept of electronic degrees of freedom. We describe some attempts to mitigate this issue that achieved limited success but ultimately consider ML band gap prediction a high-impact but unsolved problem (see discussion Section 4).

Our ensemble of 10 Wren band gap models trained with L1 loss achieved a deceivingly low  $\text{MAE} = 0.151 \text{ eV}$  and high coefficient of determination  $R^2 = 0.969$ . This is largely due to the aforementioned dataset imbalance.  $243095 / 319601 = 76.1\%$  of the combined MP + WBM dataset are PBE metals. Not wanting to discard 3/4 of our training data, we attempted naive equal loss weighting across all samples as well as increased loss weighting of non-metals. Finally, we tried prepending a metal-nonmetal classifier to our band gap regressor to only predict the band gap for materials classified as non-metals. While the latter slightly decreased the false-positive rate, neither managed to significantly improve the overall performance of our band gap model nor fix this main failure mode in our discovery pipeline of metals classified as insulators/semiconductors. Many of the generated elemental substitution structures we predicted to have sizable band gaps turned out to be PBE metals. More recent efforts in training foundation models on giant datasets and then fine-tuning on smaller cognate datasets [59] have achieved impressive sub-100 meV band gap MAEs and may be able to overcome this issue.

Figure 11 plots the rolling absolute error of our Wren band gap and dielectric constant ensembles with respect to DFPT using a variable window size of 300 samples. In Figure 11a, the bottom x-axis spans the range of PBE-computed band gaps for which we also have Wren predictions. Similarly, the top x-axis spans the range of DFPT-computed dielectric constants for which we also have Wren predictions. In Figure 11b, we swap the x-axis values to be PBE instead. That is Wren band gap predictions on the bottom x-axis and Wren dielectric constants on the top x-axis. The y-axis is identical in both subplots: the rolling Wren-vs-DFPT absolute error for band gaps on the left and dielectric constants on the right.

Figure 11a reveals that the error in dielectric constant shows a pronounced dip at intermediate ranges from about 40 to 120. This supports our initial argument for choosing dielectrics as the target material class for this discovery campaign. We hypothesized that by screening for materials

near the pareto front when considering the trade-off between two opposing material properties, we can operate both the dielectric and band gap models in regions of good training support where ML models are most reliable and still discover materials with high  $\Phi_M$ . In the case of the band gap model, this argument is less supported by the data. While the error in band gap prediction indeed drops significantly in our target region of  $E_{\text{gap}} > 2 \text{ eV}$ , the error does not increase again for extreme values but stays low even for outlier points beyond 5 eV. However, small errors on large band gaps do not negatively affect the chances of dielectric materials discovery and so are not in conflict with our objective. The issue with the band gap model is that its error for small band gaps is  $> 2 \text{ eV}$  and therefore large enough to predict metals as insulators, thereby introducing false positives into our discovery pipeline.

Figure 11b reveals that our workflow suffered from a negative feedback loop in that we purposely selected materials with large band gaps according to Wren which drew the bulk of our selection towards the lower end of the blue line. This line ends at a minimum band gap of 2 eV, indicating that no smaller Wren band gaps made it into our DFPT validation set. However, this is precisely the region where model error and its prediction are almost equal, resulting in a large number of false positive insulator predictions that turned out to be PBE metals.

## Bibliography

- [1] B. Wang, W. Huang, L. Chi, M. Al-Hashimi, T. J. Marks, and A. Facchetti, “High- k Gate Dielectrics for Emerging Flexible and Stretchable Electronics,” *Chemical Reviews*, vol. 118, no. 11, pp. 5690–5754, Jun. 2018, doi: [10.1021/acs.chemrev.8b00045](https://doi.org/10.1021/acs.chemrev.8b00045).
- [2] R. Ponce Ortiz, A. Facchetti, and T. J. Marks, “High-k Organic, Inorganic, and Hybrid Dielectrics for Low-Voltage Organic Field-Effect Transistors,” *Chemical Reviews*, vol. 110, no. 1, pp. 205–239, Jan. 2010, doi: [10.1021/cr9001275](https://doi.org/10.1021/cr9001275).
- [3] K. Yim *et al.*, “Novel High- $\kappa$  Dielectrics for next-Generation Electronic Devices Screened by Automated Ab Initio Calculations,” *NPG Asia Materials*, vol. 7, no. 6, p. e190–e190, Jun. 2015, doi: [10.1038/am.2015.57](https://doi.org/10.1038/am.2015.57).
- [4] I. Petousis *et al.*, “High-Throughput Screening of Inorganic Compounds for the Discovery of Novel Dielectric and Optical Materials,” *Scientific Data*, vol. 4, no. 1, p. 160134, Jan. 2017, doi: [10.1038/sdata.2016.134](https://doi.org/10.1038/sdata.2016.134).
- [5] G. Petretto *et al.*, “High-Throughput Density-Functional Perturbation Theory Phonons for Inorganic Materials,” *Scientific Data*, vol. 5, no. 1, p. 180065, May 2018, doi: [10.1038/sdata.2018.65](https://doi.org/10.1038/sdata.2018.65).
- [6] K. Choudhary, K. F. Garrity, V. Sharma, A. J. Biacchi, A. R. Hight Walker, and F. Tavazza, “High-Throughput Density Functional Perturbation Theory and Machine Learning Predictions of Infrared, Piezoelectric, and Dielectric Responses,” *npj Computational Materials*, vol. 6, no. 1, pp. 1–13, May 2020, doi: [10.1038/s41524-020-0337-2](https://doi.org/10.1038/s41524-020-0337-2).
- [7] D. W. Davies *et al.*, “Computational Screening of All Stoichiometric Inorganic Materials,” *Chem*, vol. 1, no. 4, pp. 617–627, Oct. 2016, doi: [10.1016/j.chempr.2016.09.010](https://doi.org/10.1016/j.chempr.2016.09.010).
- [8] R. E. A. Goodall, A. S. Parackal, F. A. Faber, R. Armiento, and A. A. Lee, “Rapid Discovery of Stable Materials by Coordinate-Free Coarse Graining,” *Science Advances*, vol. 8, no. 30, p. eabn4117, Jul. 2022, doi: [10.1126/sciadv.abn4117](https://doi.org/10.1126/sciadv.abn4117).



- [9] Z. Zhang, A. Mansouri Tehrani, A. O. Oliynyk, B. Day, and J. Brgoch, "Finding the Next Superhard Material through Ensemble Learning," *Advanced Materials*, vol. 33, no. 5, p. 2005112, 2021, doi: [10.1002/adma.202005112](https://doi.org/10.1002/adma.202005112).
- [10] Y. Zuo *et al.*, "Accelerating Materials Discovery with Bayesian Optimization and Graph Deep Learning," *Materials Today*, Oct. 2021, doi: [10.1016/j.mattod.2021.08.012](https://doi.org/10.1016/j.mattod.2021.08.012).
- [11] J. Schmidt *et al.*, "Machine-Learning-Assisted Determination of the Global Zero-Temperature Phase Diagram of Materials," *Advanced Materials*, vol. 35, no. 22, p. 2210788, 2023, doi: [10.1002/adma.202210788](https://doi.org/10.1002/adma.202210788).
- [12] M. W. Gaultois, A. O. Oliynyk, A. Mar, T. D. Sparks, G. J. Mulholland, and B. Meredig, "Perspective: Web-based Machine Learning Models for Real-Time Screening of Thermoelectric Materials Properties," *APL Materials*, vol. 4, no. 5, p. 53213, May 2016, doi: [10.1063/1.4952607](https://doi.org/10.1063/1.4952607).
- [13] J. Yan *et al.*, "Material Descriptors for Predicting Thermoelectric Performance," *Energy & Environmental Science*, vol. 8, no. 3, pp. 983–994, Mar. 2015, doi: [10.1039/C4EE03157A](https://doi.org/10.1039/C4EE03157A).
- [14] Q. Guan *et al.*, "Bimetallic Monolayer Catalyst Breaks the Activity–Selectivity Trade-off on Metal Particle Size for Efficient Chemoselective Hydrogenations," *Nature Catalysis*, vol. 4, no. 10, pp. 840–849, Oct. 2021, doi: [10.1038/s41929-021-00679-x](https://doi.org/10.1038/s41929-021-00679-x).
- [15] A. Liutkova, N. Kosinov, and E. J. M. Hensen, "Ca/ZSM-5 Catalysts for the Methanol-to-Hydrocarbons Reaction: Activity – Selectivity Trade-Off?," *Journal of Catalysis*, vol. 428, p. 115169, Dec. 2023, doi: [10.1016/j.jcat.2023.115169](https://doi.org/10.1016/j.jcat.2023.115169).
- [16] W. Chiu, K. Wakabayashi, A. Umise, M. Tahara, T. Inamura, and H. Hosoda, "Investigations of Mechanical Properties and Deformation Behaviors of the Cr Modified Ti–Au Shape Memory Alloys," *Journal of Alloys and Compounds*, vol. 897, p. 163134, Mar. 2022, doi: [10.1016/j.jallcom.2021.163134](https://doi.org/10.1016/j.jallcom.2021.163134).
- [17] W. W. Li, H. Y. Song, J. L. Dai, J. Y. Wang, M. R. An, and Y. L. Li, "Effect of Shape Memory Alloys on the Mechanical Properties of Metallic Glasses: A Molecular Dynamics Study," *Computational Materials Science*, vol. 187, p. 110088, Feb. 2021, doi: [10.1016/j.commatsci.2020.110088](https://doi.org/10.1016/j.commatsci.2020.110088).
- [18] R. D. King, "Rise of the Robo Scientists," *Scientific American*, vol. 304, no. 1, pp. 72–77, 2011, Accessed: Jan. 03, 2024. [Online]. Available: <https://www.jstor.org/stable/26002355>
- [19] A.-C. Bédard *et al.*, "Reconfigurable System for Automated Optimization of Diverse Chemical Reactions," *Science*, vol. 361, no. 6408, pp. 1220–1225, Sep. 2018, doi: [10.1126/science.aat0650](https://doi.org/10.1126/science.aat0650).
- [20] S. Steiner *et al.*, "Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language," *Science*, vol. 363, no. 6423, p. eaav2211, Jan. 2019, doi: [10.1126/science.aav2211](https://doi.org/10.1126/science.aav2211).
- [21] B. Burger *et al.*, "A Mobile Robotic Chemist," *Nature*, vol. 583, no. 7815, pp. 237–241, Jul. 2020, doi: [10.1038/s41586-020-2442-2](https://doi.org/10.1038/s41586-020-2442-2).
- [22] N. J. Szymanski *et al.*, "An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials," *Nature*, vol. 624, no. 7990, pp. 86–91, Dec. 2023, doi: [10.1038/s41586-023-06734-w](https://doi.org/10.1038/s41586-023-06734-w).
- [23] A. Lunt *et al.*, "Modular, Multi-Robot Integration of Laboratories: An Autonomous Workflow for Solid-State Chemistry," *Chemical Science*, 2024, Accessed: Jan. 04, 2024. [Online]. Available: <https://pubs.rsc.org/en/content/articlehtml/2024/sc/d3sc06206f>

- [24] I. Petousis *et al.*, “Benchmarking Density Functional Perturbation Theory to Enable High-Throughput Screening of Materials for Dielectric Constant and Refractive Index,” *Physical Review B*, vol. 93, no. 11, p. 115151, Mar. 2016, doi: [10.1103/PhysRevB.93.115151](https://doi.org/10.1103/PhysRevB.93.115151).
- [25] A. Jain *et al.*, “Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation,” *APL Materials*, vol. 1, no. 1, p. 11002, Jul. 2013, doi: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323).
- [26] H.-C. Wang, S. Botti, and M. A. L. Marques, “Predicting Stable Crystalline Compounds Using Chemical Similarity,” *npj Computational Materials*, vol. 7, no. 1, pp. 1–9, Jan. 2021, doi: [10.1038/s41524-020-00481-6](https://doi.org/10.1038/s41524-020-00481-6).
- [27] G. Bergerhoff, R. Hundt, R. Sievers, and I. D. Brown, “The Inorganic Crystal Structure Data Base,” *Journal of Chemical Information and Computer Sciences*, vol. 23, no. 2, pp. 66–69, May 1983, doi: [10.1021/ci00038a003](https://doi.org/10.1021/ci00038a003).
- [28] W. Sun *et al.*, “The Thermodynamic Scale of Inorganic Crystalline Metastability,” *Science Advances*, vol. 2, no. 11, p. e1600225, 2016, doi: [10.1126/sciadv.1600225](https://doi.org/10.1126/sciadv.1600225).
- [29] J. Qu, D. Zagaceta, W. Zhang, and Q. Zhu, “High Dielectric Ternary Oxides from Crystal Structure Prediction and High-Throughput Screening,” *Scientific Data*, vol. 7, no. 1, p. 81, Mar. 2020, doi: [10.1038/s41597-020-0418-6](https://doi.org/10.1038/s41597-020-0418-6).
- [30] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, and S. Rehme, “Recent Developments in the Inorganic Crystal Structure Database: Theoretical Crystal Structure Data and Related Features,” *Journal of Applied Crystallography*, vol. 52, no. 5, pp. 918–925, Oct. 2019, doi: [10.1107/S160057671900997X](https://doi.org/10.1107/S160057671900997X).
- [31] J. Pandey, V. Shrivastava, and R. Nagarajan, “Metastable Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub> with Pyrochlore-like Structure: Stabilization, Oxygen Ion Conductivity, and Catalytic Properties,” *Inorganic Chemistry*, vol. 57, no. 21, pp. 13667–13678, Nov. 2018, doi: [10.1021/acs.inorgchem.8b02258](https://doi.org/10.1021/acs.inorgchem.8b02258).
- [32] P. Kubelka and F. Munk, “An Article on Optics of Paint Layers,” *Z. Tech. Phys*, no. 12, pp. 593–601, Sep. 1931.
- [33] M. Weiss, B. Wirth, and R. Marschall, “Photoinduced Defect and Surface Chemistry of Niobium Tellurium Oxides ANbTeO<sub>6</sub> (A = K, Rb, Cs) with Defect-Pyrochlore Structure,” *Inorganic Chemistry*, vol. 59, no. 12, pp. 8387–8395, Jun. 2020, doi: [10.1021/acs.inorgchem.0c00811](https://doi.org/10.1021/acs.inorgchem.0c00811).
- [34] M. Valant and D. Suvorov, “Dielectric Characteristics of Bismuth Oxide Solid Solutions with a Fluorite-Like Crystal Structure,” *Journal of the American Ceramic Society*, vol. 87, no. 6, pp. 1056–1061, Jun. 2004, doi: [10.1111/j.1551-2916.2004.01056.x](https://doi.org/10.1111/j.1551-2916.2004.01056.x).
- [35] J. Choi, Y. Mao, and J. Chang, “Development of Hafnium Based High-k Materials—A Review,” *Materials Science and Engineering: R: Reports*, vol. 72, no. 6, pp. 97–136, Jul. 2011, doi: [10.1016/j.mser.2010.12.001](https://doi.org/10.1016/j.mser.2010.12.001).
- [36] Y.-C. Yeo, T.-J. King, and C. Hu, “MOSFET Gate Leakage Modeling and Selection Guide for Alternative Gate Dielectrics Based on Leakage Considerations,” *IEEE Transactions on Electron Devices*, vol. 50, no. 4, pp. 1027–1035, Apr. 2003, doi: [10.1109/TED.2003.812504](https://doi.org/10.1109/TED.2003.812504).
- [37] C. L. Hinkle, C. Fulton, R. J. Nemanich, and G. Lucovsky, “A Novel Approach for Determining the Effective Tunneling Mass of Electrons in HfO<sub>2</sub> and Other High-K Alternative Gate Dielectrics for Advanced CMOS Devices,” *Microelectronic Engineering*, vol. 72, no. 1, pp. 257–262, Apr. 2004, doi: [10.1016/j.mee.2003.12.047](https://doi.org/10.1016/j.mee.2003.12.047).

- [38] H. Glawe, A. Sanna, E. K. U. Gross, and M. A. L. Marques, “The Optimal One Dimensional Periodic Table: A Modified Pettifor Chemical Scale from Data Mining,” *New Journal of Physics*, vol. 18, no. 9, p. 93011, Sep. 2016, doi: [10.1088/1367-2630/18/9/093011](https://doi.org/10.1088/1367-2630/18/9/093011).
- [39] P. Söderlind, P. E. A. Turchi, A. Landa, and V. Lordi, “Ground-State Properties of Rare-Earth Metals: An Evaluation of Density-Functional Theory,” *Journal of Physics: Condensed Matter*, vol. 26, no. 41, p. 416001, Sep. 2014, doi: [10.1088/0953-8984/26/41/416001](https://doi.org/10.1088/0953-8984/26/41/416001).
- [40] K. Persson, “Materials Project :: About.” Accessed: Feb. 21, 2022. [Online]. Available: <https://materialsproject.org/about#db-stats>
- [41] L. Weston *et al.*, “Named Entity Recognition and Normalization Applied to Large-Scale Information Extraction from the Materials Science Literature,” *Journal of Chemical Information and Modeling*, vol. 59, no. 9, pp. 3692–3702, Sep. 2019, doi: [10.1021/acs.jcim.9b00470](https://doi.org/10.1021/acs.jcim.9b00470).
- [42] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 10, 2015. Accessed: Aug. 05, 2020. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [43] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” presented at the ICML, Jan. 2010. Accessed: Feb. 15, 2022. [Online]. Available: <https://openreview.net/forum?id=rkb15iZdZB>
- [44] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles,” Dec. 05, 2016. Accessed: Mar. 03, 2019. [Online]. Available: <http://arxiv.org/abs/1612.01474>
- [45] G. Kresse and J. Furthmüller, “Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set,” *Computational materials science*, vol. 6, no. 1, pp. 15–50, 1996.
- [46] G. Kresse and J. Furthmüller, “Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set,” *Physical Review B*, vol. 54, no. 16, pp. 11169–11186, Oct. 1996, doi: [10.1103/PhysRevB.54.11169](https://doi.org/10.1103/PhysRevB.54.11169).
- [47] P. E. Blöchl, “Projector Augmented-Wave Method,” *Physical Review B*, vol. 50, no. 24, pp. 17953–17979, Dec. 1994, doi: [10.1103/PhysRevB.50.17953](https://doi.org/10.1103/PhysRevB.50.17953).
- [48] D. C. Langreth and M. J. Mehl, “Beyond the Local-Density Approximation in Calculations of Ground-State Electronic Properties,” *Physical Review B*, vol. 28, no. 4, pp. 1809–1834, Aug. 1983, doi: [10.1103/PhysRevB.28.1809](https://doi.org/10.1103/PhysRevB.28.1809).
- [49] J. P. Perdew, M. Ernzerhof, and K. Burke, “Rationale for Mixing Exact Exchange with Density Functional Approximations,” *The Journal of Chemical Physics*, vol. 105, no. 22, pp. 9982–9985, Dec. 1996, doi: [10.1063/1.472933](https://doi.org/10.1063/1.472933).
- [50] S. P. Ong *et al.*, “Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis,” *Computational Materials Science*, vol. 68, pp. 314–319, Feb. 2013, doi: [10.1016/j.commatsci.2012.10.028](https://doi.org/10.1016/j.commatsci.2012.10.028).
- [51] K. Mathew *et al.*, “Atomate: A High-Level Interface to Generate, Execute, and Analyze Computational Materials Science Workflows,” *Computational Materials Science*, vol. 139, pp. 140–152, Nov. 2017, doi: [10.1016/j.commatsci.2017.07.030](https://doi.org/10.1016/j.commatsci.2017.07.030).
- [52] A. Jain *et al.*, “FireWorks: A Dynamic Workflow System Designed for High-Throughput Applications,” *Concurrency and Computation: Practice and Experience*, vol. 27, no. 17, pp. 5037–5059, 2015, doi: [10.1002/cpe.3505](https://doi.org/10.1002/cpe.3505).

- [53] C.-K. Lee, E. Cho, H.-S. Lee, K. S. Seol, and S. Han, “Comparative Study of Electronic Structures and Dielectric Properties of Alumina Polymorphs by First-Principles Methods,” *Physical Review B*, vol. 76, no. 24, p. 245110, Dec. 2007, doi: [10.1103/PhysRevB.76.245110](https://doi.org/10.1103/PhysRevB.76.245110).
- [54] X. Gonze and C. Lee, “Dynamical Matrices, Born Effective Charges, Dielectric Permittivity Tensors, and Interatomic Force Constants from Density-Functional Perturbation Theory,” *Physical Review B*, vol. 55, no. 16, pp. 10355–10368, Apr. 1997, doi: [10.1103/PhysRevB.55.10355](https://doi.org/10.1103/PhysRevB.55.10355).
- [55] A. Fedorova, K. Smith, K. Bostic, A. Gorrod, S. LoVerso, and M. Cahill, “Writes Hurt: Lessons in Cache Design for Optane NVRAM.” Accessed: Oct. 16, 2023. [Online]. Available: <http://arxiv.org/abs/2205.14122>
- [56] M. Horton *et al.*, “Crystal Toolkit: A Web App Framework to Improve Usability and Accessibility of Materials Science Research Algorithms.” Accessed: Feb. 16, 2023. [Online]. Available: <http://arxiv.org/abs/2302.06147>
- [57] S. Hossain, “Visualization of Bioinformatics Data with Dash Bio,” *Proceedings of the 18th Python in Science Conference*, pp. 126–133, 2019, doi: [10.25080/Majora-7ddc1dd1-012](https://doi.org/10.25080/Majora-7ddc1dd1-012).
- [58] S. Kong, F. Ricci, D. Guevarra, J. B. Neaton, C. P. Gomes, and J. M. Gregoire, “Density of States Prediction for Materials Discovery via Contrastive Learning from Probabilistic Embeddings,” Oct. 21, 2021. Accessed: Oct. 28, 2021. [Online]. Available: <http://arxiv.org/abs/2110.11444>
- [59] N. Shoghi, A. Kolluru, J. R. Kitchin, Z. W. Ulissi, C. L. Zitnick, and B. M. Wood, “From Molecules to Materials: Pre-training Large Generalizable Models for Atomic Property Prediction.” Accessed: May 12, 2024. [Online]. Available: <http://arxiv.org/abs/2310.16802>
- [60] D. Wu, T. He, J. Xia, and Y. Tan, “Preparation and Photocatalytic Properties of Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub> Photocatalyst,” *Materials Letters*, vol. 156, pp. 195–197, Oct. 2015, doi: [10.1016/j.matlet.2015.05.107](https://doi.org/10.1016/j.matlet.2015.05.107).
- [61] V. Jayaraman, C. Ayappan, B. Palanivel, and A. Mani, “Bridging and Synergistic Effect of the Pyrochlore like Bi<sub>2</sub> Zr<sub>2</sub> O<sub>7</sub> Structure with Robust CdCuS Solid Solution for Durable Photocatalytic Removal of the Organic Pollutants,” *RSC Advances*, vol. 10, no. 15, pp. 8880–8894, 2020, doi: [10.1039/D0RA00644K](https://doi.org/10.1039/D0RA00644K).
- [62] Y. Luo, L. Cao, J. Huang, L. Feng, and C. Yao, “A New Approach to Preparing Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub> Photocatalysts for Dye Degradation,” *Materials Research Express*, vol. 5, no. 1, p. 15039, Jan. 2018, doi: [10.1088/2053-1591/aaa584](https://doi.org/10.1088/2053-1591/aaa584).
- [63] X. Liu *et al.*, “Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub> Nanoparticles Synthesized by Soft-Templated Sol-Gel Methods for Visible-Light-Driven Catalytic Degradation of Tetracycline,” *Chemosphere*, vol. 210, pp. 424–432, Nov. 2018, doi: [10.1016/j.chemosphere.2018.07.040](https://doi.org/10.1016/j.chemosphere.2018.07.040).
- [64] Y. Luo *et al.*, “Synthesis, Characterization and Photocatalytic Properties of Nanoscale Pyrochlore Type Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub>,” *Materials Science and Engineering: B*, vol. 240, pp. 133–139, Jan. 2019, doi: [10.1016/j.mseb.2019.01.017](https://doi.org/10.1016/j.mseb.2019.01.017).
- [65] P. Kurlla *et al.*, “Green-Engineered Synthesis of Bi<sub>2</sub>Zr<sub>2</sub>O<sub>7</sub> NPs: Excellent Performance on Electrochemical Sensor and Sunlight-Driven Photocatalytic Studies,” *Environmental Science and Pollution Research*, Jan. 2023, doi: [10.1007/s11356-022-24760-5](https://doi.org/10.1007/s11356-022-24760-5).

- [66] S. Sorokina and A. Sleight, “New Phases in the  $\text{ZrO}_2\text{--Bi}_2\text{O}_3$  and  $\text{HfO}_2\text{--Bi}_2\text{O}_3$  Systems,” *Materials Research Bulletin*, vol. 33, no. 7, pp. 1077–1081, Jul. 1998, doi: [10.1016/S0025-5408\(98\)00076-2](https://doi.org/10.1016/S0025-5408(98)00076-2).
- [67] V. M. Sharma, D. Saha, G. Madras, and T. N. G. Row, “Synthesis, Structure, Characterization and Photocatalytic Activity of  $\text{Bi}_2\text{Zr}_2\text{O}_7$  under Solar Radiation,” *RSC Advances*, vol. 3, no. 41, p. 18938, 2013, doi: [10.1039/c3ra43518k](https://doi.org/10.1039/c3ra43518k).
- [68] A. Rajashekharaiyah *et al.*, “NUV Light-Induced Visible Green Emissions of Erbium-doped Hierarchical  $\text{Bi}_2\text{Zr}_2\text{O}_7$  Structures,” *Optical Materials*, vol. 95, p. 109237, Sep. 2019, doi: [10.1016/j.optmat.2019.109237](https://doi.org/10.1016/j.optmat.2019.109237).
- [69] X. Feng *et al.*, “Unraveling the Principles of Lattice Disorder Degree of  $\text{Bi}_2\text{B}_2\text{O}_7$  ( $\text{B} = \text{Sn}, \text{Ti}, \text{Zr}$ ) Compounds on Activating Gas Phase  $\text{O}_2$  for Soot Combustion,” *ACS Catalysis*, vol. 11, no. 19, pp. 12112–12122, Oct. 2021, doi: [10.1021/acscatal.1c03075](https://doi.org/10.1021/acscatal.1c03075).
- [70] C. F. Simon, “The Synthesis and Characterisation of Pyrochlore Frameworks,” 2010. Accessed: Jan. 06, 2024. [Online]. Available: <https://eprints.soton.ac.uk/203757/>
- [71] D. G. Fukina *et al.*, “Structure Analysis and Electronic Properties of  $\text{ATe}_4+0.5\text{Te}_6+1.5\text{-xM}_6+\text{xO}_6$  ( $\text{A}=\text{Rb}, \text{Cs}$ ,  $\text{M}_6+=\text{Mo}, \text{W}$ ) Solid Solutions with Beta-Pyrochlore Structure,” *Journal of Solid State Chemistry*, vol. 293, p. 121787, Jan. 2021, doi: [10.1016/j.jssc.2020.121787](https://doi.org/10.1016/j.jssc.2020.121787).
- [72] R. Galati, C. Simon, P. F. Henry, and M. T. Weller, “Cation Displacements and the Structures of the Superconducting Pyrochlore Osmates  $\text{A Os}_2\text{O}_6$  ( $\text{A} = \text{K}, \text{Rb}, \text{and Cs}$ ),” *Physical Review B*, vol. 77, no. 10, p. 104523, Mar. 2008, doi: [10.1103/PhysRevB.77.104523](https://doi.org/10.1103/PhysRevB.77.104523).
- [73] D. Fukina *et al.*, “Crystal Structure and Thermal Behavior of Pyrochlores  $\text{CsTeMoO}_6$  and  $\text{RbTe}_{1.25}\text{Mo}_{0.75}\text{O}_6$ ,” *Journal of Solid State Chemistry*, vol. 272, pp. 47–54, Apr. 2019, doi: [10.1016/j.jssc.2019.01.026](https://doi.org/10.1016/j.jssc.2019.01.026).