



OPEN

DATA DESCRIPTOR

An automatically curated first-principles database of ferroelectrics

Tess E. Smidt^{1,2}, Stephanie A. Mack^{1,2,3}, Sebastian E. Reyes-Lillo ^{1,2,4}, Anubhav Jain ³ & Jeffrey B. Neaton^{1,2,5} 

Ferroelectric materials have technological applications in information storage and electronic devices. The ferroelectric polar phase can be controlled with external fields, chemical substitution and size-effects in bulk and ultrathin film form, providing a platform for future technologies and for exploratory research. In this work, we integrate spin-polarized density functional theory (DFT) calculations, crystal structure databases, symmetry tools, workflow software, and a custom analysis toolkit to build a library of known, previously-proposed, and newly-proposed ferroelectric materials. With our automated workflow, we screen over 67,000 candidate materials from the Materials Project database to generate a dataset of 255 ferroelectric candidates, and propose 126 new ferroelectric materials. We benchmark our results against experimental data and previous first-principles results. The data provided includes atomic structures, output files, and DFT values of band gaps, energies, and the spontaneous polarization for each ferroelectric candidate. We contribute our workflow and analysis code to the open-source python packages *atomate* and *pymatgen* so others can conduct analogous symmetry driven searches for ferroelectrics and related phenomena.

Background & Summary

High-throughput screening of material databases integrated with first-principles calculations has been increasingly successful in the discovery of new functional materials^{1–4}. While many of the individual components for performing high-throughput searches exist, the infrastructure needed to connect and automate all the necessary components is still under development. The identification of ferroelectrics through symmetry arguments has been an active area of research^{5–16}. Moreover, lists of known ferroelectrics have been previously curated^{17–23}. However, the identification of new ferroelectrics has yet to be automated in a manner readily applicable to emerging materials databases^{24–30}. Automated high-throughput searches for ferroelectric candidates would provide a valuable guide for in-depth computational studies and experimental efforts.

Ferroelectrics have important technological applications, such as in tunable capacitors, non-volatile random access memory devices, and electro-optical data storage. In addition, ferroelectrics are capable of displaying couplings between their electronic degrees of freedom with magnetic or lattice degrees of freedom in multiferroic materials. Ferroelectricity often arises from a structural phase transition between a high-symmetry nonpolar structural phase to a low-symmetry polar structural phase with decreasing temperature, resulting in the emergence of a spontaneous polarization^{31–33}. In this scenario, the atomic geometry of the nonpolar structure can continuously distort such that the new polar structure has a subset of the symmetries of the original structure, satisfying the requirements of a second-order phase transition; in these cases, the polar space group must be an isotropy subgroup of the nonpolar space group, which is a stronger requirement than they only share a simple group-subgroup relation^{34–36}.

Thus, certain ferroelectrics can be systematically screened by searching for pairs of nonpolar and polar structures related by a small symmetry-breaking distortion. In the late 1980s, Abrahams performed some of the earliest searches for ferroelectrics in crystallographic databases using symmetry criteria^{5,6}. More recently, automated searches for new ferroelectric candidates have used symmetry arguments to identify nonpolar reference structures for existing polar materials^{7–9}. Other studies have used a combination of group theoretic and first-principles calculations to propose ferroelectric candidates^{10–12}. Bennett and co-workers proposed using high-throughput

¹Department of Physics, University of California, Berkeley, California, 94720, United States. ²Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, United States. ³Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, United States. ⁴Departamento de Ciencias Físicas, Universidad Andres Bello, Santiago, 837-0136, Chile. ⁵Kavli Energy Nanosciences Institute at Berkeley, Berkeley, California, 94720, United States. ✉e-mail: jbneaton@lbl.gov

calculations to perform chemical substitution into structures of known classes of ferroelectrics^{13–15}. Recent work used high-throughput phonon calculations to identify ferroelectrics through polar soft phonon modes of non-polar phases¹⁶.

Previous studies have focused on a limited number of compounds or families of compounds using a relatively narrow set of symmetry conditions. Current curated lists of ferroelectrics only include known ferroelectrics that have been experimentally verified. With shrinking computing costs, high-throughput material searches using first-principles methods provide an efficient strategy to discover and catalog materials. Ferroelectric databases and systematic screening of properties such as band gaps, polarizations, volume expansion, critical temperatures, and coupling to magnetic and/or topological degrees of freedom may lead to new functional materials and potentially new physical phenomena.

In this work, we integrate density functional theory (DFT), crystal structure databases, symmetry tools, workflow software, and a custom analysis toolkit to build a workflow capable of generating libraries of known, previously-proposed and newly-proposed ferroelectrics. This workflow is general and can be used with any crystal structure dataset. We present the results from performing this workflow on the Materials Project database of inorganic crystal structures²⁴. We screen over 67,000 material structures using symmetry relations between nonpolar and polar structure pairs and calculating the polarization from first-principles calculations. We identify 255 ferroelectric candidates, 200 being classified as high-quality candidates by a stringent verification process. Within these high-quality candidates, 74 are known or previously proposed, and 126 are new ferroelectrics. With the workflow developed here, we construct the first automatically-curated first-principles dataset of diverse, multi-class known and new ferroelectrics calculated with a standardized method that permits straightforward comparison. This dataset can be used to develop new tools and criteria for studying ferroelectricity across diverse materials systems. In addition our code for conducting this search has been contributed to the open-source python packages `atomate` and `pymatgen` so others can conduct searches of their own and build directly on this work^{37,38}.

Our automated workflow has three stages: symmetry analysis, first-principles calculations, and post-processing. Accordingly, the rest of the work is organized as follows: the description of the workflow, based in the concept of ferroelectric nonpolar-polar symmetry pair, is described in the Methods section. Technical aspects of our workflow and database are included in the section Data Records. Finally, we validate our workflow method against experimental databases of ferroelectrics and verify our workflow is comparable to previous first-principles results in the section Technical Validation.

Methods

Identifying ferroelectricity from first principles. Ferroelectrics are characterized by a polarization versus electric-field hysteresis loop. Experimentally, the spontaneous polarization can be determined as half of the change in polarization at zero external field³¹. The spontaneous polarization is not a direct observable; one measures the change in spontaneous polarization between two stable configurations of a material³⁹.

In three dimensions, the only space groups compatible with a polarization, meaning they leave a vector invariant under its symmetry operations, are those with polar point groups. Out of the 32 crystallographic point groups, 10 are polar; these polar point groups can keep points along specific lines (point groups 2, 3, 4, 6, *mm2*, *4mm*, *3m*, *6mm*), planes (point group *m*), or all points in three-dimensional space (point group *1*) invariant^{40,41}. All other point groups are nonpolar. We define polar structures as crystal structures with a polar space group, which is composed of a polar point group plus translations, likewise for nonpolar structures.

Following the modern theory of polarization^{39,42–47}, the polarization, \mathbf{P} , of a crystal is defined as,

$$\mathbf{P} = \mathbf{P}_0 + \sum_{i \in \{a,b,c\}} n_i \frac{e\mathbf{R}_i}{\Omega}, \quad (1)$$

where e is the charge of the electron, n_i is an integer, \mathbf{R}_i is a primitive lattice vector, and Ω is the unit cell volume (bold letters denote vectors). \mathbf{P}_0 includes electronic and ionic contributions. The second term on the right hand side of Eq. 1 is the quantum of polarization, $e\mathbf{R}_i/\Omega$, a consequence of translation symmetry. In the general case, \mathbf{P}_0 is defined up to any integer multiple of $e\mathbf{R}_i/\Omega$; for nonmagnetic crystals containing elements from even columns of the Periodic Table, \mathbf{P}_0 is defined up to even integer multiples. Since we are screening many systems here, we use the more general definition in this work.

Only differences of the computed polarization on the “same branch” are physically meaningful, where different branches are related by integer multiples of the quantum of polarization. Equivalently, the evolution of the polarization along an adiabatic path between two states must be smooth. Nonpolar space groups can only host formal polarizations that are zero or one-half modulo the polarization quantum. We use a nonpolar reference structure to calculate the change in polarization due to a polar distortion. In general, paths between two opposite polar configurations are sufficient to compute the change in polarization; however, we choose to use a nonpolar reference structure in this work as is standard for calculating the polarization from first principles.

To recover a smooth polarization path, we ensure the nonpolar structure must be continuously deformable into the polar structure along a path that preserves the symmetry of the polar structure and for which the system remains insulating. We then perform calculations of multiple structures along the distortion path to compute the spontaneous polarization, which can be compared to experiment. Using this approach, the spontaneous polarization can be directly predicted using first-principles methods with good accuracy⁴⁸.

Hence, for the purposes of this work, we only consider ferroelectrics for which a high-symmetry nonpolar reference structure can be readily identified in the database for a lower-symmetry polar structure that can support a polarization⁴². We automate a search for compounds supporting two such phases and then compute the polarization difference along the structural path connecting the two structures.

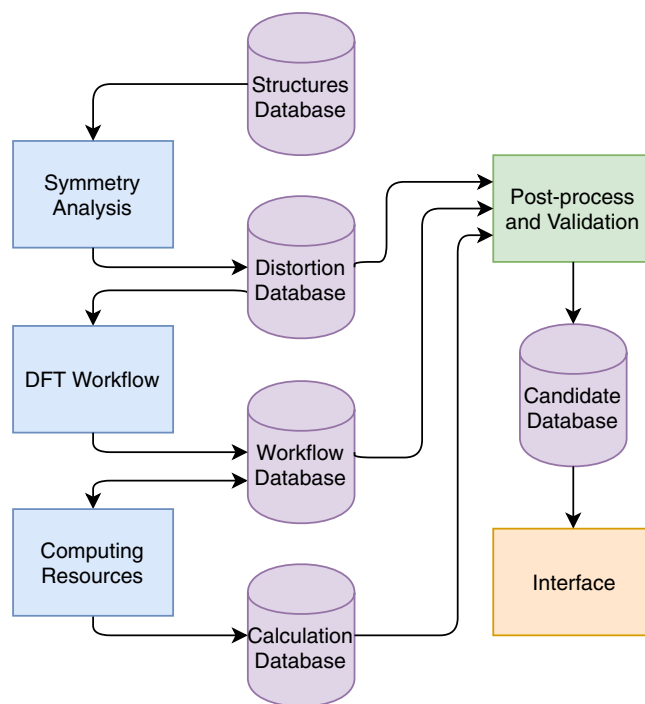


Fig. 1 Diagram of the automated ferroelectric search workflow developed here. Databases are shown as purple cylinders. Processes are shown as rectangles: blue designates processes used to identify and perform first-principles calculations, green designates post-processing and verification, and orange designates the web-interface. Arrow directions indicate the flow of information. For example, the Workflow Database provides information to the Computing Resources about which calculations to compute and the Workflow Database is updated as calculations complete or as errors occur on the Computing Resources.

If a polar ferroelectric structure corresponds to a metastable state, and is higher in energy than a nearby non-polar ground state by a small amount, the system can be considered an antiferroelectric^{49,50}. Antiferroelectrics exhibit double hysteresis loops in polarization versus electric-field measurements; the field-induced first-order phase transition originates with an energy barrier between the nonpolar ground state and the polar metastable phase. In an antiferroelectric, the nonpolar ground state phase is related by a nonpolar distortion to a distinct nonpolar reference structure; the ground state structure is generally characterized as “anti-polar” to distinguish it from the nonpolar reference structure. Symmetry conditions for antiferroelectrics are described in ref.⁵¹. For completeness, we note that to identify antiferroelectrics using the workflow presented here, in addition to finding a reference nonpolar phase, the polar metastable ferroelectric phase and an antipolar ground state, the material would need to display a small energy difference between the polar and antipolar structures on the order of 1–10 meV ⁵².

Workflow overview. We first describe the general workflow diagram comprising symmetry analysis, first-principles calculations, and post-processing. As shown in Fig. 1, the complete workflow involves the passing of data between many separate calculations. In developing our workflow, we automate the following tasks:

1. Identifying candidate materials possessing nonpolar-polar structure pairs related by a continuous distortion (an isotropy subgroup symmetry relation).
2. Performing spin-polarized DFT calculations of changes in total energy, band gap, and polarization for multiple structures along the nonpolar-polar distortion.
3. Post-processing calculation data to compute the spontaneous polarization of the polar ferroelectric phase.
4. Validating the calculation quality for each ferroelectric candidate.
5. Creating an interface for viewing the results for all candidates (see the section Graphical Interface).

We start by choosing a crystal structure database on which to perform the search (see Structure Selection Symmetry Analysis). We emphasize that any crystallographic database (e.g. any of the databases described in^{24–30}) can be used to perform our workflow, as long as the atomic coordinates and lattice parameters of the structures are provided.

Within the chosen database, we perform a symmetry analysis to find candidate materials possessing nonpolar-polar structure pairs related by a continuous symmetry deformation. Any such pairs found to satisfy the symmetry deformation criteria are stored in the Distortion Database as being deformable by symmetry. This criteria includes the following conditions: (1) The polar structure belongs to a space group that is a subgroup of the space group of the nonpolar structure; and (2) There exists a transformation matrix between the high-symmetry setting of the nonpolar structure to the low-symmetry setting of the polar structure. The latter imposes that the

Symmetry	Structures [†]	~67,000
	Polar structures [†]	~15,000
	Distinct polar formula [†]	~10,000
	Nonpolar-polar structure pairs	~17,000
	Structure pairs with distinct chemical formulae	~1,600
	Pairs with continuous transformation	413
DFT	Pairs with metallic endpoints	80
	Pairs with metallic interpolations	59
	Pairs with calculation errors	19
	Pairs that completed successfully	255
Valid	High-quality ferroelectric candidates	200
	Known ferroelectrics in high-quality candidates	74
	New ferroelectrics in high-quality candidates	126

Table 1. Results obtained by applying our workflow to the Materials Project database. The symbol [†] indicates “in Materials Project database at time of search”. Boxes relate numbers by symmetry conditions (Symmetry), first-principles calculations (DFT), and validation processes (Valid.). “Nonpolar-polar structure pairs” satisfy simple group-subgroup relations while “pairs with continuous transformation” satisfy group-isotropy subgroup relations.

distortion of the lattice parameters and atomic coordinates between the nonpolar and polar structures is continuous, meaning the polar structure belongs to an isotropy subgroup of the nonpolar structure.

We then carry out DFT calculations on the candidate pairs to extract the changes in the band gaps, total energies, and polarization along the nonpolar-polar distortion (see Computational Methods). These results are stored in a Workflow Database and then accessed by our Computing Resources to perform the calculations. Next, the information stored in the Distortion, Workflow, and Calculation Databases is used together to post-process quantities such as the computed spontaneous polarization and to validate ferroelectric candidates using experimental and previous first-principles results (see Post-processing and Spontaneous Polarization Values and Verification of Computational Methodology). The information needed to assess the quality and properties of the candidates is then added to the Candidate Database where it can be accessed by our web Interface for viewing the candidate materials in aggregate (see Graphical Interface). Finally, candidates are screened to ensure the polarization and energy profile across the nonpolar-polar distortion are smooth and continuous, i.e. all calculations ended correctly and provide reliable results.

In the sections below, we describe in detail the methods used for creating an automatically curated dataset of ferroelectrics from the Materials Project database²⁴. The Materials Project database is largely based on structures from the Inorganic Crystal Structure Database (ICSD)^{27,28} and includes hypothetical structures created through stoichiometric substitution. We use the Materials Project database to test our workflow. Our results for the Materials Project are not intended as the most general curated list of ferroelectrics; however, as the first automatically obtained list of ferroelectrics, they uncover new candidates and provide a blueprint for further studies. More elaborately curated lists may be constructed by applying our workflow to additional databases in future studies. The results from applying our workflow are described below and summarized in Table 1. We note that our workflow is modular and open-source, so it can be readily adapted and applied by others to expand the search for ferroelectrics and related materials such as antiferroelectrics and multiferroics.

Structure selection. As motivated earlier, the input to our workflow is a collection of candidate materials possessing nonpolar-polar structure pairs. There are several methods that can be used to create candidate nonpolar-polar structure pairs. For example, one can apply a polar distortion to an existing nonpolar structure or create a hypothetical nonpolar reference structure for an existing polar structure. In this work, to classify a material as a ferroelectric candidate, we require both nonpolar and polar structures to be present in the database. As shown below, even this direct approach provides new candidate materials, previously overlooked as ferroelectrics. Future studies may choose to relax this requirement to identify a greater number of promising materials.

To search for compatible nonpolar-polar structure pairs in the Materials Project dataset, we first select compounds possessing nonpolar and polar structures with space groups which are related by a group-subgroup relation. Note that in principle the same compound may display more than one ferroelectric structural transition, and therefore have more than one nonpolar-polar symmetry pair. For each of these pairs, we require that the number of sites in the nonpolar structure is less than or equal to the number of sites in the polar structure. We perform this initial query using pymatgen, spglib, and the Materials Project API^{37,53–55}. We provide the number of nonpolar-polar structures pairs resulting from this query in the top box of Table 1.

At the time of this search, the Materials Project database had approximately 67,000 structures, approximately 15,000 of which are polar. We find approximately 17,000 nonpolar-polar structure pairs related via group-subgroup space group relations. This number is large, in part, because the same polar structure may be paired with multiple possible nonpolar reference structures, and vice versa. This number is also large because the requirement that the polar structure is in a subgroup of the nonpolar space group is a much weaker requirement than the polar structure belongs to an isotropy subgroup of the nonpolar structure—which we check later in the workflow. These roughly 17,000 pairs contain approximately 1,600 of the approximately 10,000 distinct polar compositions in the Materials

Project database. The remaining polar compositions in the Materials Project do not have symmetry compatible nonpolar structures within the same composition present in the database. We note that it is possible to propose hypothetical nonpolar reference structures for polar candidates using group theoretic methods or by relaxing the symmetry tolerance between nonpolar-polar distortions^{7–9,16}; this is left for future work.

Naming conventions. We adopt the pymatgen `alphabetical_formula` method for the Composition class (with spaces and 1s removed) to output consistent formulas for our candidates. We note that this method orders elements in such a way that does not match conventions in the literature. For example, we use O_3PbTi where the standard in the literature is $PbTiO_3$. Compositions printed by pymatgen also differ from those used in mineralogy, such as for boracite, lawsonite and many other minerals in our dataset. In our datafiles, we also provide formula name output using the pymatgen `reduced_formula` method for the Composition class, which sorts elements by electronegativity.

Symmetry analysis. The automated nature of our ferroelectric search relies on strict symmetry criteria. As described in the Structure Selection section, we pre-screen our candidate nonpolar-polar structure pairs using the symmetry tools in pymatgen and spglib to ensure that these pairs satisfy preliminary group-subgroup relationships. We then use the Structure Relations symmetry tool provided by the Bilbao Crystallographic Server (BCS)^{56–58} to impose the symmetry criteria described in the Workflow Overview, namely, to obtain a transformation matrix connecting the lattice parameters and atomic coordinates of the structure pair^{59,60}. The BCS has a freely available web interface for accessing a wide variety of symmetry tools. We create python scripts to automate interaction with and scrape returned data from the BCS to perform our symmetry checks using the python package `mechanize`⁶¹. The Structure Relations tool checks the following:

- 1.1 The group-subgroup index relations are compatible. The index of a group-subgroup relation indicates the number of ferroelectric domains (distinct polar variants) that arise from the symmetry breaking of the high-symmetry structure.
- 1.2 There exists a path of maximal subgroups between the high-symmetry structure and low-symmetry structure.
- 1.3 The Wyckoff position splitting of the high-symmetry structure is compatible with the Wyckoff positions of the low-symmetry structure.
- 1.4 The lattice of the high-symmetry structure in the low-symmetry setting must be within a defined tolerance of the lattice of the low-symmetry structure (see Symmetry Precision section).
- 1.5 Each atom in the high-symmetry structure in the low-symmetry setting can be paired to an atom in the low-symmetry structure such that atom pairs are separated by a distance no greater than a given tolerance (see Symmetry Precision section).

Structure Relations takes Crystallographic Information Files (CIFs) of high-symmetry and low-symmetry structures and tolerance thresholds as arguments. We use a lattice tolerance of 3 Å and 10° for lattice parameters and angles, respectively. These tolerances are generous for materials with average sized unit cells (i.e. with lattice parameters less than 20 Å) and permit a wide variety of distortions. For the present work, high-quality candidates are reported for a maximum pairing distance of 1.5 Å. As shown in Table 1, out of the 17,000 structure pairs that we test with Structure Relations, 413 are found to be deformable by symmetry with a maximum distortion less than or equal to 1.5 Å.

Symmetry precision. Symmetry precision is a tolerance factor used to assess whether an atom is equivalent to another after a symmetry operation up to a maximum distance. A symmetry precision between 10^{-1} and 10^{-5} Å is typically used. In the Materials Project database, a symmetry tolerance of 10^{-1} Å is used for the reported space group stored in the database. We use the same tolerance to generate CIFs sent to the BCS Structure Relations.

We evaluated how varying the symmetry tolerance changes the resolved space group for all the structures in the Materials Project. We were able to determine this efficiently by using a binary search on a \log_{10} scale for a maximum and minimum symmetry tolerance of 10^{-1} Å and 10^{-5} Å, respectively. Out of the 67,000 structures we checked, 50,000 (75%) structures were resolved into one distinct space group for the entire symmetry precision range. For additional discussion about the sensitivity of symmetry precision on resulting space groups in the search for ferroelectrics, see refs.^{16,62}.

DFT calculation details. In our workflow, we perform spin-polarized DFT calculations using the Vienna Ab initio Software Package (VASP) version 5.3.5^{63–65}. We use the generalized gradient approximation (GGA) functional of Perdew, Burke, and Ernzerhof (PBE)⁶⁶. Our calculations use PAW pseudopotentials and an energy cutoff of 520 eV for the plane-wave basis^{67,68}; this is 1.3 times the highest cutoff recommended for the pseudopotentials used⁶⁹. Structures are initialized with ferromagnetic ordering in all cases in this work. Since the default is for parallel alignment of the spins, we expect the workflow to be reliable for nonmagnetic or ferromagnetic materials. Materials with more complicated magnetic ground states, such as antiferromagnets, would require consideration of different antiferromagnetic orderings. Therefore, some common multiferroics possessing antiferromagnetic or near antiferromagnetic ordering may not be identified by this workflow. Extensions to include antiferromagnetic spin arrangements are relegated to future work. These settings correspond to default values used to create the Materials Project database and therefore allow a direct comparison of our results with the Materials Project database.

We use the Berry phase approach from refs.^{39,43,44,70,71}, as implemented in VASP to calculate the electronic part of the macroscopic spontaneous polarization. We calculate the ionic part of the polarization using the point charge and position for each atom in the unit cell, see the `calc_ionic` function in `pymatgen.analysis.ferroelectricity.polarization` for details.

We use the default parameters for VASP inputs as defined in `pymatgen` (and used by Materials Project) and `atomate`^{69,72,73}. For details on these parameters, see the documentation for `pymatgen.io.vasp.sets.MPStaticSet`. We use a Hubbard U correction to correct the DFT-PBE description of d states of select oxides and fluorides following the approach in ref.⁷⁴. To see the guidelines for which compounds we apply a U , see refs.^{24,69,72,73}. We use a reciprocal k -point density of 50 k -points per $(1/\text{\AA})^3$ for structural relaxations and 100 k -points per $(1/\text{\AA})^3$ for static and polarization calculations. We use total energy convergence criteria of 5×10^{-5} eV per atom for the electronic self-consistent loop and 5×10^{-4} eV/ \AA per atom for the ionic relaxation loop for structural optimizations. These convergence parameters were tested against higher-accuracy convergence parameters on a set of 182 chemically diverse compounds in ref.⁷³, yielding total energies within 15 meV/atom and lattice volumes within 7.5% of the higher-accuracy calculations for nearly 96% of the compounds.

We note that while the local density approximation (LDA) is commonly used to describe certain ferroelectric oxides, and therefore to compute their polarizations, use of a generalized gradient approximation (GGA), such as the PBE functional, tends to be standard for wider classes of materials nowadays. PBE is also the default functional used by the Materials Project for structural relaxations and calculating material properties. Thus, we use PBE for this effort. Our results are in line with the typical overestimation of PBE for the lattice parameters, and therefore, we expect a similar overestimation of the polarization. In addition, while DFT-PBE tends to underestimate electronic band gaps, the latter plays a minimal role in the determination of standard ferroelectric materials, and as shown below, will only limit the computation of the spontaneous polarization for a small number of them.

Scientific workflow packages. We construct the scientific workflows to perform the structural relaxations and spin-polarized DFT calculations of energy, band gap, and polarization using the FireWorks and atomate python packages^{38,75}. FireWorks is built for managing computational scientific workflows. atomate is built for constructing workflows for multiple computational material science codes, such as VASP. atomate uses FireWorks classes to develop modules for performing common DFT calculations with VASP, such as structural relaxations and self-consistent calculations of total energy. atomate also provides a framework for building custom modules, which we use to construct our structural interpolations and polarization calculations modules.

DFT workflow. We use the DFT workflow, shown in Fig. 2, to compute the physical properties needed to identify ferroelectric candidates. We perform spin-polarized calculations and for systems with spin-polarized ground states, consider only ferromagnetic ordering. We execute the DFT workflow shown in Fig. 2 for the 413 pairs with continuous nonpolar-polar transformations, with maximum distortions that do not exceed 1.5 \AA .

For each structure pair, we begin with the nonpolar structure in the low-symmetry setting (obtained from BCS Structure Relations in the Symmetry Analysis step) and the polar structure. We use the nonpolar structure transformed into the low-symmetry setting so we can perform structure interpolations and polarization calculations across similar lattices. We perform relaxations of the unit cell and atomic positions of both of these structures twice, using a level of convergence similar to what Materials Project uses for its database entries. As mentioned, all calculations are spin-polarized and use PBE, with a Hubbard U correction for systems with non- d^0 open-shell cations; our workflow assumes ferromagnetic ordering for all systems. Extensions of our workflows to consider antiferromagnetic and other orderings will be the subject of future work. We then fix the relaxed nonpolar and polar structures, and perform a self-consistent DFT calculation to compute the total energy and band gap. If either the nonpolar structure or polar structure is found to be metallic at the DFT-PBE level in our spin-polarized calculations initialized with ferromagnetic orderings and a standard U —here we define metallic as having a DFT-PBE band gap of less than 10 meV—we stop the workflow for that structure pair.

If the polar and nonpolar structures are both insulating, we compute the polarization along the distortion path. As shown in Table 1, 80 of the 413 structure pairs were computed to have metallic endpoints in our spin-polarized calculations: 30 were found to have a metallic nonpolar structure but insulating polar structure, 2 were found to have a metallic polar structure but insulating nonpolar structure, and 24 were found to have both metallic polar and nonpolar structures. 24 additional structures have at least a metallic nonpolar structure, but these workflows were halted before the polar structures had their band gaps computed.

We compute the DFT total energy, band gap, and polarization of eight evenly-distributed linearly-interpolated structure, or interpolations, of the nonpolar to polar structures. We found eight interpolations to be sufficient for reconstructing a smooth polarization trend for at least 75% of our candidates. Similar to the previous step where a metallic calculation causes the workflow to stop, metallic interpolations similarly halt the workflow since we would be unable to calculate the polarization of that structure. 59 candidates were found to have metallic interpolations. The DFT workflow is labeled as complete when all polarization calculations along the path have completed. As shown in Table 1, 255 structure pairs successfully completed the workflow, and satisfy our requirements of a ferroelectric phase transition.

Post-processing spontaneous polarization values. As discussed earlier, only polar space groups are compatible with a polarization vector that is not integer or half integer multiples of the polarization quantum⁴⁶. If a reference nonpolar structure that is continuously deformable into the polar structure can be identified, we can then calculate the polarization of several interpolated structures between the reference nonpolar and the target ferroelectric polar structure. The nonpolar structure is used as a means to calculate the spontaneous polarization; however, we note in general it is not necessary for the nonpolar structure to be experimentally observable for the polar material to be ferroelectric.

We start by calculating the formal polarization of the nonpolar structure, which is either zero or a half quantum of polarization (modulo the quantum of polarization) along the three lattice directions. Then we perform the same calculation for the first interpolated structure along the distortion, and then the next, until we arrive at the final polar structure. For a sufficient number of interpolations between the nonpolar and polar structures, we can trace

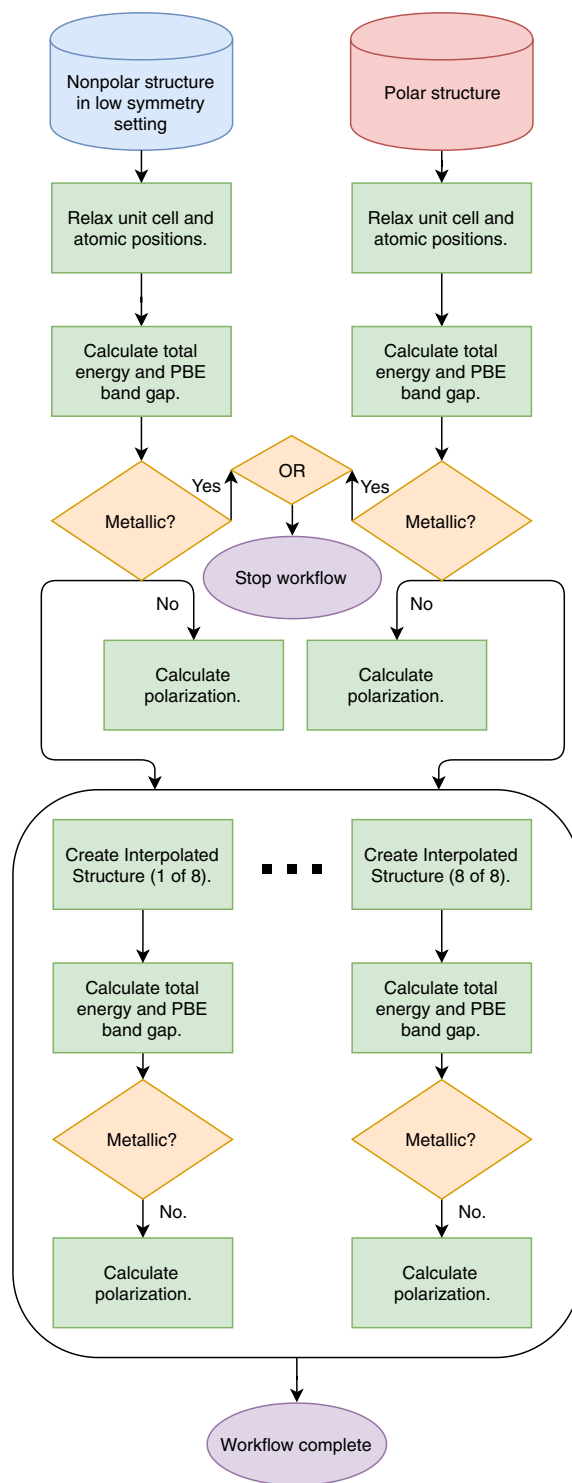


Fig. 2 Diagram of DFT workflow written with atomate and Fire Works. Blue and red boxes denote initial nonpolar and polar structures, respectively, green boxes denote DFT calculations, orange rhombuses denote decision steps, and purple ellipses denote exit steps. Black arrows represent passing of data between different software codes. The metallic check for the interpolated structures leads to a similar condition as earlier, where if any of the interpolations are metallic the workflow stops (not illustrated for clarity in figure).

out smooth, continuous polarization “paths” along the distortion; there will be infinitely many paths due to the periodicity of the polarization lattice. Subtracting the polarization values at the nonpolar and the polar endpoints of the same path or “branch” will give us the spontaneous polarization vector of the polar ferroelectric phase.

We perform the following steps to recover the same branch “proper” polarization, which is independent of choice of branch. The first step, which is crucial, is to readjust the polarization for each structure along the

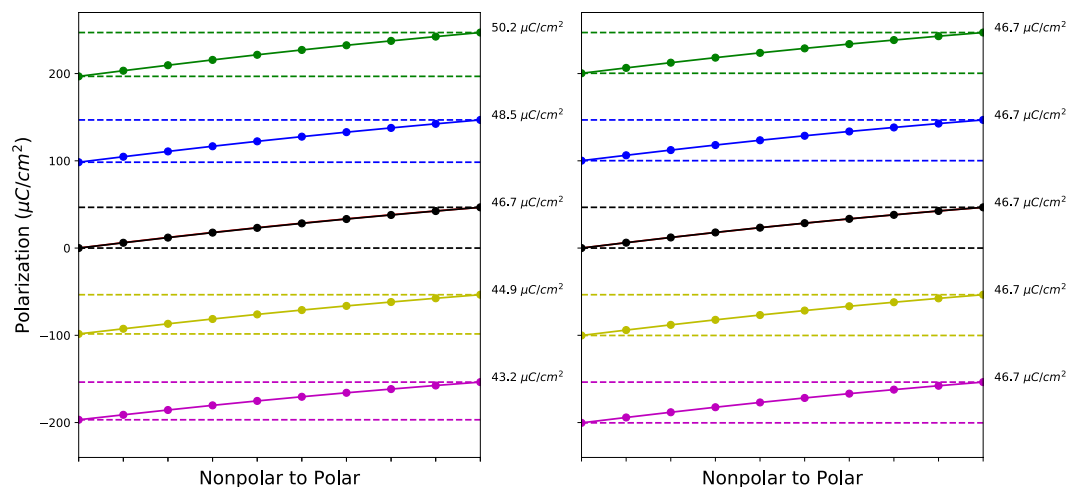


Fig. 3 Examples of improper (left) and proper (right) polarization curves for BaTiO₃ along the \hat{c} direction versus distortion from the nonpolar to polar structure, showing the importance of calculating the proper polarization. Due to the change in lattice parameters and volume across the distortion, the quantum of polarization defined for each structure along the \hat{c} direction is different. Using these different quanta causes the improper spontaneous polarization predicted by different branches to differ, as can be seen in the polarization values given in the right of the image. In contrast for the proper polarization (right), we re-scale the polarization of each intermediate structure to be in the polar structure's polarization lattice and use the quantum of polarization as defined by the polar structure. This results in predictions that are branch independent, which is what we use to assess candidates. Note that while, in this specific case, the calculated polarization values for all interpolations were on the same branch, this need not generally be the case.

distortion to be in the polar polarization lattice. To do this, we modify the polarization of the intermediate structures by the ratio of the quantum of polarizations of the two lattices (the lattice parameters divided by the volume multiplied by the electron charge), i.e.,

$$P_{\eta,i}^{(\text{polar lattice})} = P_{\eta,i} \left(\frac{R_{\text{polar},i} / \Omega_{\text{polar}}}{R_{\eta,i} / \Omega_{\eta}} \right), \quad (2)$$

where $R_{\eta,i}$ and Ω_{η} are the lattice parameters and volume of the η^{th} structure along the distortion, R_{polar} and Ω_{polar} are the lattice parameters and volume of the polar structure, and i is a lattice direction \hat{a} , \hat{b} or \hat{c} . If we do not perform this adjustment, we are calculating what is called the “improper polarization” which will depend on the choice of branch and is therefore unphysical⁷⁶. See Fig. 3 for an example of the differences between proper and improper polarization for BaTiO₃ (here we use conventional notation for the name).

After adjusting our initial polarizations, we then construct a periodic lattice with lengths and angles corresponding to the quantum of polarization along each polar lattice direction; this corresponds to the second term of Eq. (1). For the polarization lattice, the lengths of the lattice vectors are the cell lattice vectors divided by the volume of the unit cell and multiplied by conversion factors for electron charge and length scale.

Our algorithm for adjusting the polarizations to be on the same branch is depicted in Fig. 4. First, we take the nonpolar polarization (adjusted to be in the polar polarization lattice), and choose the “image” (or periodic value) of the polarization value in the nonpolar polarization lattice that is closest to the Cartesian origin (0, 0, 0). The value of the nonpolar polarization along \hat{a} , \hat{b} , and \hat{c} can either be zero or a half-quantum. Then, we find the image of the first interpolation polarization value (again adjusted to be in the polar polarization lattice) that is closest to the Cartesian coordinates of the adjusted nonpolar polarization value. We continue this process until we get to the polarization of the polar structure.

This algorithm will find the polarization path with the smallest difference between polarizations of subsequent interpolations. An issue is that this algorithm can incorrectly find the same branch polarization in cases where the change in polarization between interpolations is larger than the quantum of polarization between branches. One example of this type of failure is the polarization resolved for CrO₃ with search ID 187, see Fig. 5 (the search ID being a simplified unique identifier, defined for the purposes of our work, for pairs of structures used in our workflow search, see Online-only Table 3). In this example, the algorithm chooses a discontinuous path that has a smaller spontaneous polarization of 78.2 $\mu\text{C}/\text{cm}^2$ in red. However, the correct path uses the last three interpolations in the branch shown with a dashed red line and gives a polarization of 122.9 $\mu\text{C}/\text{cm}^2$. To correctly reconstruct this polarization with our existing algorithm, more structures would be needed to better interpolate between the nonpolar and polar structures. Alternatively, the curvature of the spline connecting nonpolar and polar structures could be used to identify points on the same branch; we leave this refinement for future work.

Graphical interface. To view the DFT ferroelectric candidate data in aggregate, we create an interactive web site for viewing polarization and total energy plots, animations of the distortion, and other data. The interface

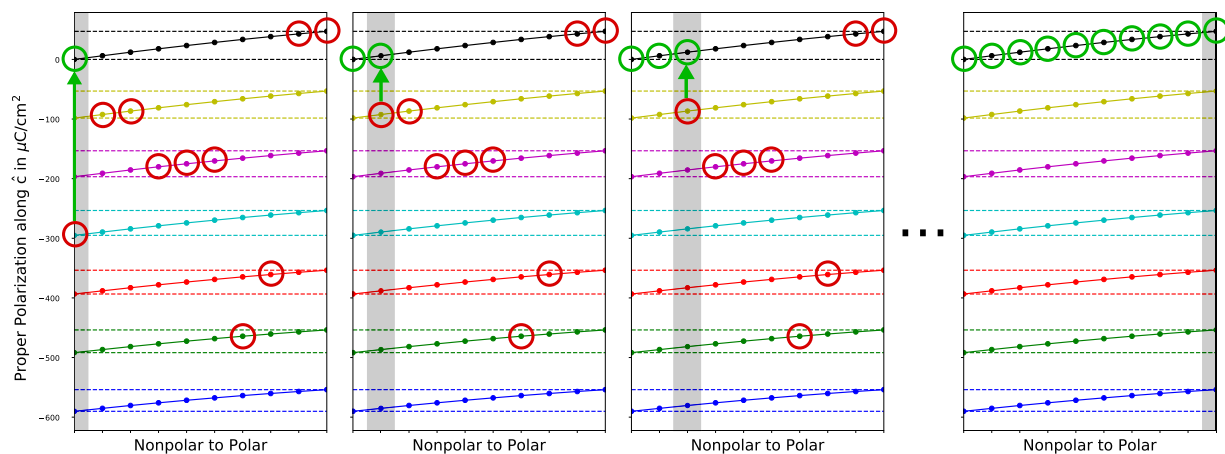


Fig. 4 A visual demonstration of the same branch polarization algorithm demonstrated in one-dimension (rather than three-dimensions) using BaTiO_3 . The values for the polarization for each interpolation are those circled in red. In the first panel, we adjust the nonpolar polarization to be on the branch closest to zero. In the second panel, we move the first interpolated polarization to be on the branch closest to the adjusted nonpolar polarization. In the third panel, we adjust the second interpolated polarization to be on the branch closest to the first interpolation polarization. If the algorithm finishes successfully, all the adjusted polarizations will be on the same branch.

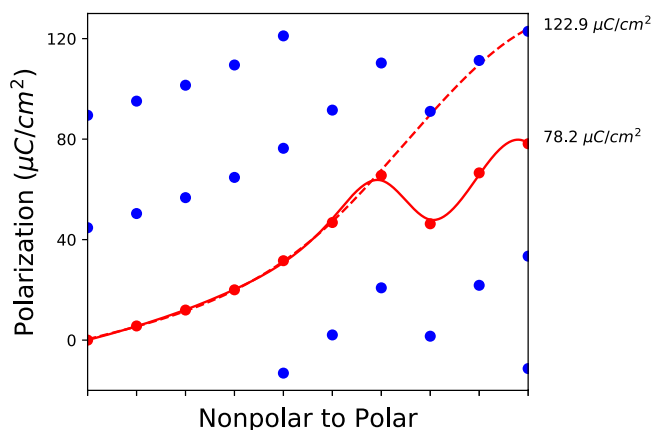


Fig. 5 The polarization reconstruction for CrO_3 with search ID 187. The polarization along the b lattice parameter is incorrectly reconstructed because the different polarization branches are closer than the change in polarization between structure interpolations.

consists of two main pages: (1) a page containing a sortable table of ferroelectric candidates organized by category (whether the candidate had a value of polarization successfully calculated and if so with what level of confidence) and (2) individual candidate pages that show energy and polarization plots, distortion animations, and other data specific to that candidate. This interface is available at https://blondegeek.github.io/ferroelectric_search_site/.

Data Records

This dataset is available as two JSON files deposited in figshare⁷⁷ and our GitHub repository (http://github.com/blondegeek/ferroelectric_search_site)⁷⁸. The JSON files provide details of the symmetry analysis performed for each candidate and data generated by DFT calculations and post-processing from the workflow. Zipped folders of the input and output VASP files for each candidate deposited in figshare⁷⁷. The title of the zipped folder includes the workflow ID to correlate the VASP files to information in the JSON files provided. We also provide an interface for viewing the dataset at http://blondegeek.github.io/ferroelectric_search_site with the code for the interface located at http://github.com/blondegeek/ferroelectric_search_site.

File format. We contribute the following data:

1. JSON file with information on workflow status of each calculated candidate and calculation details extracted from VASP inputs and outputs. This includes total energy, band gap, polarization, post-processed information, and validation criteria for candidates with completed calculation. See Tables 2 and 3 for details.

Key	Type	Description
_id	bson.objectid.ObjectId	Automatically created unique identifier.
wfid	unicode	The workflow id.
cid	unicode	The “connection” or distortion id. The alphanumeric portion of the string after cid_ corresponds to the bson.objectid.ObjectId used in the distortion database.
search_id	unicode	Simplified unique identifier for pairs of structures used in the search.
workflow_status	unicode	Status of workflow denoted by FireWorks.
alphabetical_formula	unicode	Composition with elements sorted alphabetically.
pretty_formula	unicode	Composition with elements sorted by electronegativity.
polar_id	unicode	Materials Project Id.
nonpolar_id	unicode	Materials Project Id.
polar_icsd_ids	floats list	ICSD id numbers, if available.
nonpolar_icsd_ids	floats list	ICSD id numbers, if available.
polar_spacegroup	float	Polar space group, integer between 1 and 230.
nonpolar_spacegroup	float	Nonpolar space group, integer between 1 and 230.
orig_polar_structure	pymatgen.Structure dict	Polar structure as referenced in distortion JSON file.
orig_nonpolar_structure	pymatgen.Structure dict	Nonpolar structure as referenced in distortion JSON file.
structures	pymatgen.Structure dicts list	Static calculation structures. Fully complete workflows have 10.
relaxation_len	float	Number of relaxation calculations performed. Fully complete workflows have 2.
relaxation_task_labels	strs list	The task labels of the relaxation calculations performed.
static_len	float	Number of static calculations performed. Fully complete workflows have 10.
static_task_labels	strs list	The task labels of the static calculations performed.
polarization_len	float	Number of polarization calculations performed. Fully complete workflows have 10.
polarization_task_labels	strs list	The task labels of the polarization calculations performed.
polarization_change_norm	float	The Cartesian norm of the recovered spontaneous polarization.
polarization_change	floats list	The vector along a, b, and c of the recovered spontaneous polarization vector.
raw_electron_polarization	floats lists list	Raw electron polarization per structure from VASP along Cartesian directions.
raw_ionic_polarization_vasp	floats lists list	Raw ionic polarization per structure from VASP along Cartesian directions.
raw_ionic_polarization	floats lists list	Raw ionic polarization per structure from calc_ionic along lattice directions.
polarization_quanta	floats lists list	Structure dependent polarization quanta along a, b, and c lattice vectors.
same_branch_polarization	floats lists list	Same branch polarization along a, b, and c for each polarization calculation structure.
polarization_max_spline_jumps	floats lists list	Max jump between spline and data for polarization along a, b, and c.
polarization_smoothness	floats list	Average jump between spline and data for polarization along a, b, and c.

Table 2. Key, value data type, and value description for `workflow_data.json` entries.

2. JSON file with information describing all 413 nonpolar-polar structure pairs with group-subgroup relations compatible with a second-order phase transition in the Materials Project determined with BCS Structure Relations and used in this search. See Tables 4 and 5 for details.
3. Zipped folders with the VASP 5.3.5 INCAR, KPOINTS, OUTCAR, and POSCAR files.

Reported properties. For each candidate we provide an initial nonpolar-polar pair of structures, including the nonpolar structure in both the nonpolar, high-symmetry setting and polar, low-symmetry setting. We also provide the displacements of each atom and other metrics provided by BCS Structure Relations.

For each successful calculation, we provide the structure used for calculation, the ionic and electronic polarization computed by VASP, the ionic polarization computed via the method of point charges, the total energy and energy per atom of the structure, and other commonly computed quantities such as total magnetization, magnetization per atom, forces, and stresses. We also give details as to which calculations (out of the 22 computed) for a given candidate pair were completed.

For each set of completed calculations we also provide, the recovered spontaneous polarization using the workflow described in the Methods section, as well as spline data characterizing the smoothness of the polarization and energy trends across the nonpolar-polar distortion.

Graphical representation of results. In the top row of Fig. 6, we partition the high-quality candidates found in the Materials Project into known and newly-proposed ferroelectrics and further partition those

Key	Type	Description
energies	floats list	Energy in eV for each static calculation structure.
energies_per_atom	floats list	Energy per atom in eV for each static calculation structure.
energies_per_atom_max_spline_jumps	float	Max jump between spline and data for energy per atom.
energies_per_atom_smoothness	float	Average jump between spline and data for energy per atom.
calculated_max_distance	float	Calculated max distortion distance. Compare to dmax in distortion.json entries.
zval_dict	dict	dict with keys of species and values of ZVAL in number of electrons.
hubbards	dict	dict with keys of species and values of Hubbard U correction in eV pairs.
cbms	floats list	Conduction band minimum per static calculation computed structures.
vbms	floats list	Valence band maximum per static calculation computed structures.
stresses	floats lists list	Stress tensor per static calculation computed structures.
charges	floats dicts lists list	Charges projected onto spd orbitals per atom per static calculation computed structures.
magnetization	floats lists list	Magnetization in Bohr magnetons per atom per static calculation computed structures.
total_magnetization	floats list	Total magnetization in Bohr magnetons per static calculation computed structures.
forces	floats lists list	Cartesian forces per atom per static calculation computed structures.
bandgaps	floats list	list of band gaps in eV for static calculation computed structures.

Table 3. Key, value data type, and value description for `workflow_data.json` entries continued.

ferroelectrics into subclasses. In Fig. 6 we see that known and new ferroelectric candidates are well mixed along the metrics of nonpolar-polar structure energy difference, distortion maximum between nonpolar and polar structures, PBE band gap of polar structure, and energy above hull (a measure of the thermodynamic driving force for decomposition⁷⁹).

In the middle row of Fig. 6, the candidates with large polarizations denoted with red triangles belong to the perovskite family. We recover many known perovskite ferroelectrics, such as: LiNbO_3 , AlBiO_3 , BiInO_3 , KNbO_3 , BaO_3Ti , CdO_3Ti and O_3PbTi . In addition, we recover well-known double perovskite ferroelectrics, such as $\text{Bi}_2\text{Nb}_2\text{O}_9\text{Pb}$ and $\text{Bi}_2\text{O}_9\text{SrTa}_2$. We also recover well-known antiferroelectrics possessing ferroelectric metastable phases, such as NaNbO_3 , HfO_3Pb and O_3PbZr . We again note that we do not use conventional notation for these systems but rather use alphabetical order of elements to provide a consistent ordering with our workflow output.

Other classes in the middle row of Fig. 6 are the organic $(\text{NH}_4)_2\text{SO}_4$ family in blue and structures already proposed by theory to be ferroelectric in purple. These data show that there are many known and proposed ferroelectrics in the literature with polarizations of $10 \mu\text{C}/\text{cm}^2$ or less.

In the bottom row of Fig. 6, we categorize new ferroelectric candidates into different trending compositions or structure types. There are several candidates containing fluorides, carbon-oxygen compounds, and hydroxyl groups. We highlight these candidates because they are very different in composition from oxide ferroelectrics most common in the literature. We also point out some hypothetical non-magnetic hexagonal manganite-like structures found in the Materials Project database that have polarizations of approximately $10 \mu\text{C}/\text{cm}^2$ and half-quantum nonpolar polarizations.

In Fig. 7, we show trends in the number of pairs with continuous deformation with given nonpolar-polar point group transitions. There are 32 crystallographic point groups; nonpolar and polar point groups are shown in blue and red, respectively. The thickness and color of the line connecting nonpolar and polar point groups indicate the number of structures in the dataset with a continuous deformation between those point groups. We find that point group transitions that correspond to orthorhombic structures such as $mmm \rightarrow mm2$, monoclinic structures such as $2/m \rightarrow 2$ or $2/m \rightarrow m$, and hexagonal $6/mmm \rightarrow 6mm$ are the most prevalent.

In Fig. 8, we show the computed polarization of the ferroelectric candidates plotted with respect to their polar point group, similar to the plot of piezoelectric tensor magnitudes in ref. 4. The majority of candidates have polarization less than $5 \mu\text{C}/\text{cm}^2$, shown on the right. The candidates in point group $4mm$ with large polarizations are perovskites with a reference structure in point group $m\bar{3}m$. The degree of shading of a radial cell is proportional to the number of candidates in that region of the plot. For example, there are many candidates with polar point groups 2 and $mm2$ that have polarizations within $25 \mu\text{C}/\text{cm}^2$.

Technical Validation

Verification of computational methodology. Several checks are needed to ensure our automated calculations have completed satisfactorily and the information automatically extracted from them is reliable. We describe these tests below.

Testing smoothness of energy and polarization trends with distortion. We flag any ferroelectric candidates whose calculations cannot be used to reliably assess the quality of the candidate. For example, if the trend in total energy is not continuous, we cannot be confident that we can extract a meaningful polarization trend. Similarly, if the same branch polarization is not continuous, we cannot be confident that an accurate spontaneous polarization has been determined.

Key	Type	Description
_id	bson.objectid.ObjectId	These ids are used to generate cid in workflow_data JSON file.
pretty_formula	unicode	Composition with elements sorted by electronegativity.
polar_id	unicode	Materials Project Id.
nonpolar_id	unicode	Materials Project Id.
polar_icsd	float	ICSD id number, if available.
nonpolar_icsd	float	ICSD id number, if available.
polar_spacegroup	float	Polar space group, integer between 1 and 230.
nonpolar_spacegroup	float	Nonpolar space group, integer between 1 and 230.
bilbao_polar_spacegroup	float	Polar space group from Bilbao Crystallographic Server, integer between 1 and 230.
bilbao_nonpolar_spacegroup	float	Nonpolar space group from Bilbao Crystallographic Server, integer between 1 and 230.
distortion	dict	Details pertaining to distortion between nonpolar and polar structure.
polar_band_gap	float	Materials Project computed band gap.
nonpolar_band_gap	float	Materials Project computed band gap.

Table 4. Key, value data type, and value description for `distortion.json` entries.

Key	Type	Description
high_symm	dict of pymatgen.Structure	Nonpolar structure in high-symmetry setting.
high_low_setting	dict of pymatgen.Structure	Nonpolar structure in low-symmetry setting.
low_symm	dict of pymatgen.Structure	Polar structure in low-symmetry setting.
high_pre	unicode	Structure information as directly output by Bilbao Crystallographic Server website.
high_low_pre	unicode	Structure information as directly output by Bilbao Crystallographic Server website.
low_pre	unicode	Structure information as directly output by Bilbao Crystallographic Server website.
distortion	list	Table of low-symmetry setting of Wyckoff position, the string “(x, y, z)”, species, distortion in x, y, and z, and the magnitude of distortion.
pairings	list	Wyckoff splitting pairing between high symmetry and low symmetry structures.
dmax	unicode	Maximum distortion distance between nonpolar and polar structure.
s	unicode	Degree of lattice distortion (S) is the spontaneous strain (sum of the squared eigenvalues of the strain tensor divided by 3).
dav	unicode	Maximum distortion distance between nonpolar and polar structure.
delta	unicode	The measure of compatibility (Δ) (Bergerhoff <i>et al.</i> 1998).

Table 5. JSON keys, value data type, and value description for distortion dictionary of `distortion.json` entries.

To assess the smoothness of trends in polarization and energy across a distortion, we use `UnivariateSplines` from `scipy.interpolate`⁸⁰. We use cubic splines for fitting polarizations and quartic splines for fitting total energies. We use the default smoothness parameter of 1.0. These splines are generated using the `Polarization` class in `pymatgen.analysis.ferroelectricity.polarization`.

We find that 31 out of the 55 materials that do not have smooth interpolations contain atoms with nonzero magnetic moments, mostly containing 3d elements (V, Cr, Mn, Fe, Co, Ni) and one containing the 5d element W. We found 26 materials to have several discontinuities in total energy (even when these calculations resulted in smooth polarizations). These materials were transition metal oxides, fluorides, carbonates, orthosilicates, and phosphates with alkali or alkaline earth metals (Li, Na, and Ba for these specific examples), many being Li-ion battery cathode candidates. The transition metals in these materials (V, Cr, Mn, Fe, Co, Ni) can take multiple oxidation states. Because these discontinuities in energy were coincident with discontinuities in the total magnetization, we believe these jumps were caused by the transition metal species changing oxidation state through the distortion, suggesting the need to assess their ground state magnetic ordering more carefully.

Metallic endpoints and metallic interpolations. Polarization calculations require a nonzero band gap along the distortion path. Therefore, workflows that have either polar or nonpolar structures that are calculated to be metallic are halted. In the `workflow_data.json`, these workflows are designated by a `workflow_status` of “DEFUSED”. Of the 413 pairs considered, 80 possess metallic endpoints and therefore interpolations were not performed. Occasionally, interpolated structures between two nonmetallic structure endpoints are metallic and there were 59 that had metallic interpolation structures. If any structure along the path from nonpolar to polar structure is metallic, the quality of automated analysis cannot be guaranteed. We include these candidates in our dataset, but they are noted as having a `polarization_len` (see Tables 2 and 3) of less than 10 or do not have a `polarization_change_norm` (in cases where some of the interpolated polarizations are None). If a

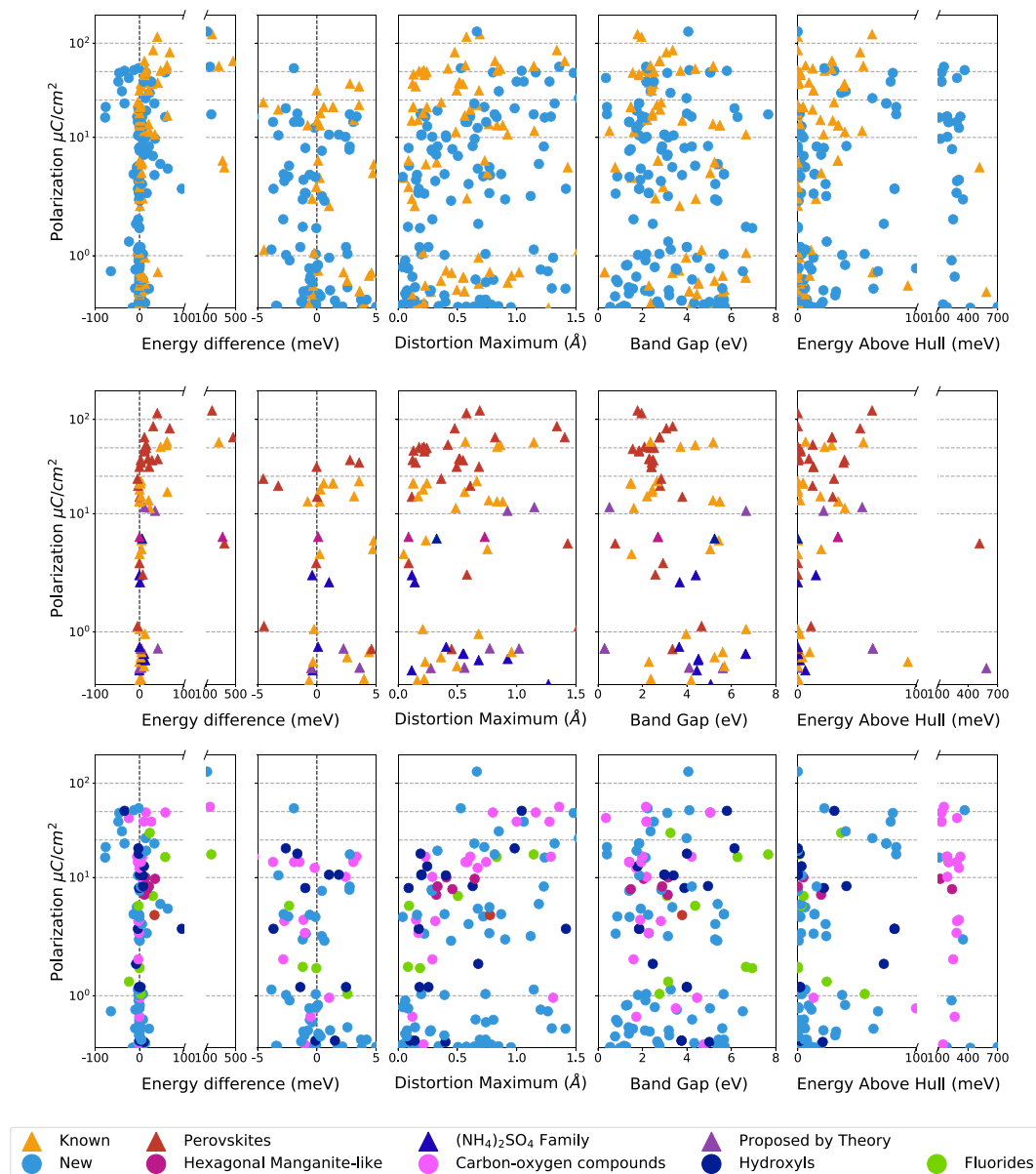


Fig. 6 Validated ferroelectric candidates from our automated search in the Materials Project. Computed spontaneous polarization plotted against nonpolar-polar energy difference, maximum atomic distortion, PBE + U band gap, and energy above hull of the polar structure. All results are generated for spin-polarized DFT-PBE + U calculations. Note, that for spin-polarized systems, we only initialized the calculations with a ferromagnetic ordering. The energy above hull is extracted from the Materials Project. The legend labels different subcategories considered in this work and described in the text.

candidate has a `polarization_len` equal to 10 and have a `polarization_change_norm`, all interpolated polarization calculations completed successfully.

We also note that the method used to determine whether a material is insulating differs for pymatgen and VASP. In our workflow, pymatgen determines the band gap by comparing the energies of the band edges. When VASP determines whether to proceed with calculating the polarization, it checks whether the material is insulating by checking the occupations of the band edges. Since the occupations can be sensitive to the choice of smearing and k-grid density, there are instances where using our default settings in the workflow leads to partial occupancies, while the band edge energies suggest the material has a finite gap; therefore VASP will not proceed with calculating the polarization whereas our workflow deems the material to be insulating. This tends to occur for materials containing 3d elements which, as discussed earlier in the context of magnetism, would require more careful calculations to reliably capture their properties, and therefore not necessarily expected to successfully complete the workflow.

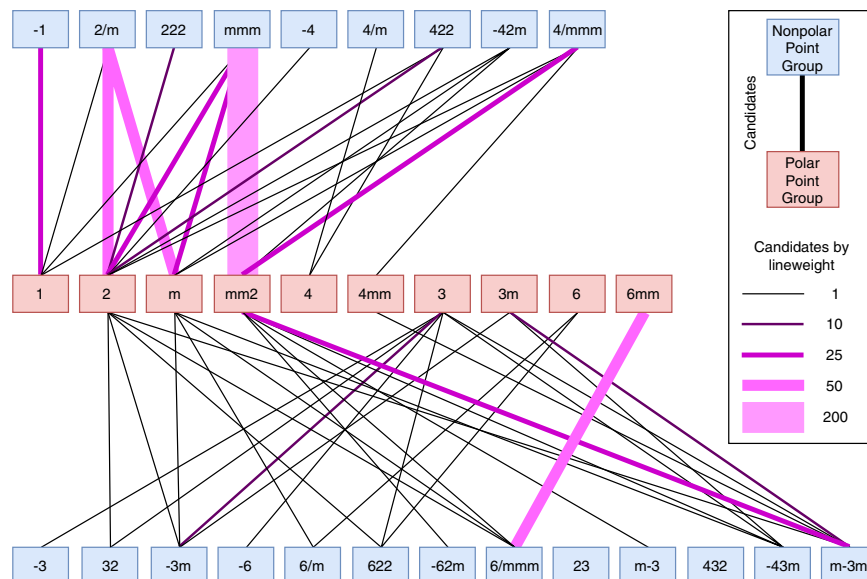


Fig. 7 Schematic of the number of point group transitions between the 32 crystallographic point groups. The thickness and color of the line connecting nonpolar and polar point groups indicates the number of structures in the dataset with a continuous deformation between those point groups. The legend includes a schematic for nonpolar-polar phase transitions and describes the significance of line weights and colors connecting nonpolar and polar point groups.

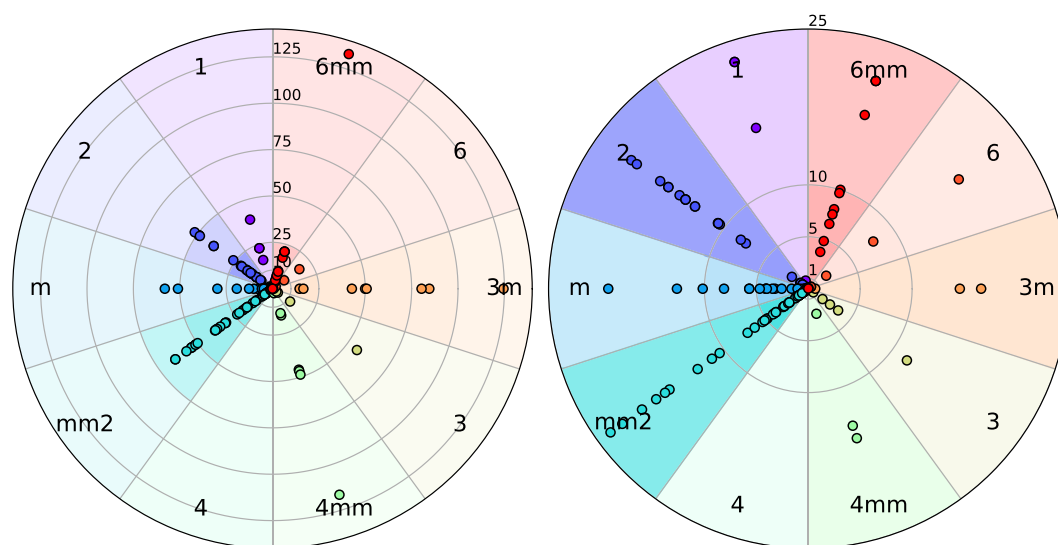


Fig. 8 Diagrams of polarization magnitude vs. point group. The left diagram shows the full range of polarizations from 0 to 140 $\mu\text{C}/\text{cm}^2$ and the right diagram zooms in on polarizations in the range 0 to 25 $\mu\text{C}/\text{cm}^2$. Point groups are grouped according to symmetry. The darkness of a radial cell is proportional to the number of candidates in that region of the plot.

Comparison of materials project to relaxed structures. Two structural relaxation calculations are already performed on all structures in the Materials Project. We perform additional relaxations of the unit cell and atomic positions to ensure total energy convergence. We found only less than 5% (10%) of our relaxed structures have lattice parameter differences of more than 3% (1%) from the original Materials Project structures. Because we perform relaxations of nonpolar structures transformed to the low-symmetry polar setting, we compare the relaxed nonpolar structure to the low-symmetry transformed structure output by BCS of the nonpolar structure from Materials Project.

Identifying high-quality candidates. In any high-throughput search, there are calculations that complete without errors and some that require further scrutiny to interpret its results. We deem as high-quality candidates those calculations where the polarization and total energy trends are smooth and continuous; we define this criteria in

the following way: the maximum difference between the data and spline fitted to the same branch polarization must be less than $10^{-1} \mu\text{C}/\text{cm}^2$, and the maximum difference between the data and the spline fitted to the energy trend must be less than 10^{-2} eV .

As shown in Table 1, out of the 255 candidates, 200 pass through our stringent verification criteria and ensures the polarization and energy trends across the ferroelectric distortion are smooth and continuous. The remaining candidates are still valid candidates; we recommend checking the polarization and energy trends by hand as the algorithms used for analysis may not have reliably recovered the spontaneous polarization in these cases, as in Fig. 5. These high-quality candidates are further described in the section Determining Known and New Ferroelectrics. There are candidates included in this list with polarization values of zero, as computed within the accuracy of our calculations; we include these since they pass the above criteria and could be tuned to host nonzero polarizations such as by chemical substitution. As well, the polarization values may be very small but this is also true of some of the known ferroelectric candidates computed with our workflow, and so for the sake of completeness we also list these candidates. We also call the reader's attention to the materials with magnetic elements. Since all materials are initialized with ferromagnetic ordering, we do not expect to reliably capture materials with other magnetic orderings (e.g. antiferromagnets); as such, materials with nonzero magnetic moments may not necessarily have been calculated using their true magnetic ground state and those results should be interpreted with caution.

Comparison to DFT calculated polarizations from literature. We verify that our workflow reproduces the first-principles calculated polarizations for a variety of ferroelectrics in the literature. DFT calculated values for the ferroelectric polarization depends heavily on the structures and functional used in the calculation, and for magnetic systems, it will also be sensitive to magnetic ordering. For example, while in our work the endpoint structures are fully relaxed, other works constrain the relaxed polar unit cell to have the same volume as the experimental structure⁸¹; these calculated polarizations will be systematically smaller than ours due to PBE optimized structures having larger lattice parameters than experimental values.

We compare to literature where a fully optimized (unit cell, volume, and atomic positions) relaxation procedure is used. For clarity of comparison, we use the chemical formula used by these references rather than the `alphabetical_formula` that is convention for the rest of this work.

The ferroelectric first-principles literature is largely dominated by studies of perovskites. We compare to calculations for the perovskites BaTiO_3 , PbTiO_3 , LiNbO_3 , BiAlO_3 , CdTiO_3 , BiFeO_3 , and we include in our comparison the double perovskite $\text{SrBi}_2\text{Ta}_2\text{O}_9$. These comparisons are summarized in Table 6.

We note again here that because our workflow assumes ferromagnetic ordering, it will fail for many multiferroics since they tend to have other magnetic orderings, such as antiferromagnetic or noncollinear orderings. For example, two important multiferroics are BiFeO_3 and YMnO_3 ; the former is only captured to some extent by our workflow and the latter is not reported at all by our database. The standard polar phase of BiFeO_3 features G-type antiferromagnetic ordering, and that for YMnO_3 features A-type antiferromagnetic ordering. As the present version of our workflow initializes all calculations with ferromagnetic ground states for all magnetic systems, we do not expect to capture most multiferroics using the workflow as implemented for this search with the Materials Project database. Additionally, we note that the standard nonpolar reference structure for BiFeO_3 is in space group 167 ($R\bar{3}c$); however, the $R\bar{3}c$ structure of BiFeO_3 is not in the Materials Project database. That the $R3m$ and $Pm\bar{3}m$ structures of BiFeO_3 calculated in this workflow, with ferromagnetic ordering and the default U value, are found to be insulating is fortuitous and allows it to be reported as a high quality candidate. However, the BiFeO_3 experimentally verified ground state structure, 161 ($R3c$), did not complete our workflow due to metallic interpolations. Conversely, YMnO_3 is DEFUSED and does not proceed to the interpolations because its nonpolar structure is found to be metallic in our calculations. Extending the workflows to account for antiferromagnetic and other spin-ordered ground states, as well as to relax the constraint that a nonpolar reference structure exist in the database, will enable casting a wider net for known and new multiferroics, and will be the subject of future work.

Determining known and new ferroelectrics. We distinguish known from new ferroelectrics in our workflow depending on whether or not a material has been reported in the literature as ferroelectric. Thus, we perform a literature review by hand for every considered candidate, and leave automating such literature searches to future work. We find that out of 200 high quality candidates, 74 are known or previously proposed ferroelectrics and 126 are, to our knowledge, new ferroelectric candidates.

In Fig. 6, we plot calculated spontaneous polarization versus nonpolar-polar total energy difference, maximum distortion distance, PBE band gap, and energy difference between the polar structure and convex hull reported in the Materials Project for known, proposed and new ferroelectric candidates. We also provide tables of the known and new candidates grouped by chemical formula and polar space group in Online-only Tables 1 and 2, respectively. Details connecting these candidate to specific workflow calculations are in Online-only Table 3.

As seen in Fig. 6, known and newly-proposed ferroelectrics display similar dispersion and overlap in the range considered. The middle row of Fig. 6 demonstrates the variety of known ferroelectric candidates that we are able to recover, from perovskites to candidates in the organic $(\text{NH}_4)_2\text{SO}_4$ family to candidates proposed by theory. The bottom row of Fig. 6 shows categories of new ferroelectric candidates we find in the Materials Project, some from previously known ferroelectric classes such as hexagonal manganite-like structures and less-studied categories such as fluorides, carbon-oxygen compounds, and crystals containing hydroxyl groups.

Comparison to hand-curated list of ferroelectrics in the Pauling Files database. The Pauling File is a materials database accessible through SpringerMaterials¹⁸. In this database, there are materials tagged as ferroelectric and antiferroelectric. We use these tagged entries to validate whether the workflow is able to successfully identify diverse ferroelectrics by examining which tagged (anti)ferroelectrics complete the workflow. In the

Formula and Space Group		<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	<i>P_s</i> (μC/cm ²)
BaTiO ₃ (99)	ref. ⁸⁷	4.005	—	4.210	43.5
	ref. ⁸⁸	4.000	—	4.216	47.0
	This work (1)	4.000	—	4.224	46.7
	This work (2)	4.001	—	4.215	45.9
PbTiO ₃ (99)	ref. ⁸⁸	3.844	—	4.767	125.5
	This work (3)	3.871	—	4.594	116.8
LiNbO ₃ (161)	ref. ⁸⁸	5.203	—	14.110	84.7
	This work (4)	5.216	—	14.116	84.5
SrBi ₂ Ta ₂ O ₉ (36)	ref. ⁸⁹	5.550	5.550	25.100	34.1
	This work (5)	5.602	5.614	25.520	36.9
CdTiO ₃ (26)	ref. ^{90†}	5.250	5.387	7.570	21.0
	This work (6)	5.402	5.525	7.694	37.2
CdTiO ₃ (33)	ref. ^{90†}	5.239	5.378	7.619	29.0
	This work (7)	5.360	5.494	7.812	34.8
BiAlO ₃ (161)	ref. ^{91†}	3.840 (cubic)	—	—	75.6
	This work (8)	3.844 (cubic)	—	—	80.3
KH ₂ PO ₄ (43)	ref. ⁸⁸	10.800	10.710	7.110	5.5
	This work (9)	10.730	10.652	7.105	5.2

Table 6. Comparison of this work to other first-principles studies of ferroelectrics, primarily perovskites. For this table, we use the chemical formula conventions used in the works we compare to. We compare to calculations using PBE + U unless otherwise specified. The polar space group is given in parentheses under the chemical formula. The symbol † indicates the reference being compared to used the Local Density Approximation (LDA) functional in their calculations. LDA polarization values tend to be smaller than polarization values calculated with PBE (which we use in this work) due to smaller predicted lattice parameters by LDA than PBE. The search ids for entries in the table are: (1) 69 (2) 70 (3) 389 (4) 331 (5) 80 (6) 146 (7) 147 (8) 23 (9) 257.

Pauling Files, there are 955 distinct compositions tagged as (anti)ferroelectric, 306 of which are pure (not doped) compositions. Out of 306 pure compositions, 95 of those compositions are included in the Materials Project as polar materials. This does not necessarily mean that the Materials Project database contains the same ferroelectric polar structure as referenced in the Pauling Files; rather, it simply means that there exists a polar structure in the Materials Project with the same composition as a tagged ferroelectric or antiferroelectric in the Pauling Files. 57 of these compositions have nonpolar-polar structure pairs in the Materials Project, 40 of which are found to have a continuously distortion by BCS Structure Relations. 32 of the 40 are successfully identified by the workflow as high-quality candidates, meaning the energy and polarization trends are smooth. Out of the 8 candidates that did not successfully make it through the workflow as high quality candidates, 4 of them (CrO₃Y, Eu₂GeSe₄, Cl₃CoTi, and MnO₃Y) had metallic endpoints, 2 candidates (Cl₃CrRb and Br₃MnRb) had metallic interpolation structures, and 2 candidates (Cl₄CoRb₂ and B₇ClCr₃O₁₃) did not have smooth energy trends due to fluctuating band gaps and magnetic moments in the interpolations. We note that most of these 8 candidates are non-*d*⁰ systems, and therefore expected to exhibit magnetic order. This suggests that the primary impact to the robustness of our workflow is the level of DFT used and the magnetic orderings considered; since PBE tends to underestimate band gaps, and since ferromagnetically-ordered systems tend to be overwhelmingly metallic, several of these candidates are experimentally insulating but are metallic in our workflow.

Comparison to experimental measurements in Landolt-Börnstein series. To validate that our workflow calculates polarizations that can be used to guide experimental efforts, we compare to tabulated experimentally measured polarizations of known ferroelectrics in Landolt-Börnstein - Group III Condensed Matter - Ferroelectrics and Related Substances^{19–23}. This series classifies hundreds of ferroelectrics into a 72 class numbering scheme. We note that polarization values for ferroelectrics in the Landolt-Börnstein series volumes 36A, 36B and 36C may be superseded by more recent experimental measurements. Experimentally measured polarizations depend greatly on the quality of the sample and the method used. For many ferroelectrics, polarization measurements made across decades vary greatly depending on these factors. Any comparison between theory must be taken in this context.

Nonetheless, plots of polarizations values calculated by our workflow vs. experimental spontaneous polarizations reported in the Landolt-Börnstein series volumes 36A, 36B and 36C are shown in Fig. 9 and tabulated in Table 7. We only compare search candidates to Landolt-Börnstein entries that match in composition and polar space group. For example, NaNbO₃ is LB Number 1A-1 and has a polarization of 12 μC/cm² for the 161 space group structure. However, all the NaNbO₃ found in our search are orthorhombic (space groups 26 and 29), so we do not make the comparison.

For experimental polarizations greater than 10 μC/cm², the majority of experimental polarizations are between 25% and –50% of those that we calculate, well within an order of magnitude. The exceptions are both polar structures of Bi₂O₉SrTa₂ (space groups 36 and 41), which are calculated to have polarizations much greater

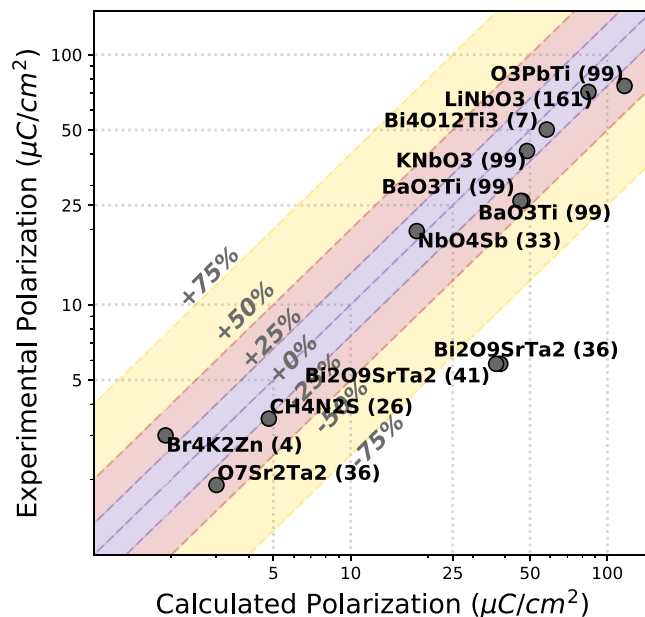


Fig. 9 Log-log plot of experimental vs. calculated polarization for ferroelectric materials in the Landolt-Börnstein series with polarization value larger than $1 \mu\text{C}/\text{cm}^2$. 6 materials with polarization value equal to or smaller than $1 \mu\text{C}/\text{cm}^2$ are not shown in plot but are included in the Table 7. Colored regions show experimental values within $\pm 25\%$, $\pm 50\%$, and $\pm 75\%$ of calculated values.

than their experimental values. Multiple entries of a given formula indicate multiple calculations for different structure pairs in our dataset for the same compound. Compounds with polarizations of less than or equal to $1 \mu\text{C}/\text{cm}^2$ are not shown on the plot given the log-log scale. For polarizations less than $5 \mu\text{C}/\text{cm}^2$, we see our calculations capture the general trends of the experimental polarizations.

We find that the PBE functional we use for our DFT calculations overestimates polarizations. This is partially due to unit cells relaxed with PBE having larger than experimental volumes and thus larger distortions.

Usage Notes

In this work, we present 413 nonpolar-polar structure pairs in the Materials Project database that are compatible with a second-order phase transition as ferroelectric candidates and perform DFT calculations of total energy, band gap, and polarization for these structures pairs.

This dataset offers the first opportunity to compare a large number of known, previously proposed, and new ferroelectrics side by side with the same methodology. We believe by setting strict criteria for ferroelectricity and casting a wide-net using high-throughput searches, we will find candidates that challenge and advance our understanding of ferroelectric phenomena. As seen in our candidates, there may be ferroelectrics waiting to be discovered that defy our expectations. This dataset will be useful for creating new tools and criteria for analyzing diverse ferroelectrics.

The infrastructure provided by the Bilbao Crystallographic Server, FireWorks, pymatgen, and atomate is crucial to being able to perform these types of searches efficiently. Thus, we also provide our code and data for these searches with the hope they will provide access for others to perform and develop similar searches.

Our code for performing structural interpolations and polarization calculations has been incorporated into the pymatgen and atomate packages. We also provide the code that we use to create the interface that we used to view our candidates in aggregate.

The workflow we have presented can be extended to any crystal structure database, experimental and hypothetical. Several modifications can be made to this workflow to extend the scope of these searches. Notably, extending our workflow to treat different magnetic orderings will enable it to treat multiferroics. Additionally, the same DFT workflow can be used to screen any experimentally measured polar structure - even one without an existing nonpolar reference - by generating nonpolar reference structures with BCS Pseudosymmetry⁷. Our workflow can also be adapted to perform species substitutions and find symmetry relations between structure types, classes of structures that share space groups, Wyckoff positions, and other lattice similarities.

Code availability

VASP version 5.3.5 used to perform DFT calculations is a proprietary code. The Bilbao Crystallographic Server (BCS) is freely available on-line at <http://www.cryst.ehu.es>. Fireworks, atomate, and pymatgen are python packages accessible on GitHub. Fireworks and atomate are released under a modified Berkeley Software Distribution (BSD) License. pymatgen is released under a Massachusetts Institute of Technology (MIT) License. Both MIT and BSD licenses are open-source and permit both commercial and non-commercial use. Our workflow code is included since atomate version 0.6.7 and our analysis code is available in pymatgen since

Search ID	Formula	Polarization ($\mu\text{C}/\text{cm}^2$)	LB number	Exp.	Exp. Ref.
51	B ₇ ClMg ₅ O ₁₃ (29)	0.3	18A-1	0.08	92
68	BaO ₃ Ti (38)	49.9	1A-10	Not available	93
69, 70	BaO ₃ Ti (99)	46.7, 45.9	1A-10	26	93
71	BaO ₃ Ti (160)	49.9	1A-10	Not available	93
79	Bi ₂ Nb ₂ O ₉ Pb (36)	38.1	9A-10	Not available	94
81	Bi ₂ O ₉ SrTa ₂ (41)	38.3	9A-12	5.8	95
80	Bi ₂ O ₉ SrTa ₂ (36)	36.9	9A-12	5.8	95
82	Bi ₄ O ₁₂ Ti ₃ (7)	58.1	9A-15	50.2	96
100	Br ₄ K ₂ Zn (4)	1.9	39A-16	3.0	97
111	CH ₄ N ₂ S (26)	4.8	50A-1	3.5	98
146	CdO ₃ Ti (26)	37.2	1A-9	Debated	99
147	CdO ₃ Ti (33)	34.8	1A-9	Debated	99
164, 165	Cl ₄ K ₂ Zn (33)	2.2, 0.5	39A-9	0.13	100
166	Cl ₄ Rb ₂ Zn (33)	0.5	39A-10	0.16	101
292	H ₈ N ₂ O ₄ S (33)	0.0	39A-1	0.6 max at -51.5°C , reversal by 0.05 approaching -260°C	102
302	HfO ₃ Pb (32)	19.8	1A-16	Antiferroelectric (62).	103
309	K ₂ O ₃ Se (33)	0.7	39A-2	0.15	104
313	KNbO ₃ (38)	51.3	1A-2	Not available	105
314	KNbO ₃ (99)	48.7	1A-2	41.2	105
315	KNbO ₃ (160)	50.7	1A-2	Not available	105
331	LiNbO ₃ (161)	84.5	2A-1	71	106,107
365	NbO ₃ Sb (33)	18.1	5A-2	19.7	108
389	O ₃ PbTi (99)	116.8	1A-11	75	109,110
390	O ₃ PbZr (32)	23.5	1A-15	16 on cooling. Antiferroelectric (55).	111
404	O ₃ PTiTi (33)	22.1	35A-13	Not available	112
407	O ₇ Sr ₂ Ta ₂ (36)	3.0	8A-6	1.9	113

Table 7. Table of calculated versus experimentally measured polarizations as given in the Landolt-Börnstein - Group III: Condensed Matter - Ferroelectrics and Related Substances^{19–21}. “Not available” in the Experimental Reference column (Exp. Ref.) indicates a polarization value was not provided in these references. We make comparisons only between search candidates that have a polar space group compatible with the Landolt-Börnstein reference. The search ID being a simplified unique identifier, defined for the purposes of our work, for pairs of structures used in our workflow search.

v2019.2.4. We also use the following python packages in our analysis and Graphical Representation of Results: numpy, scipy, matplotlib, ipython, and jupyter^{80,82–86}. These packages are freely available through the Python Package Index (<https://pypi.org/>).

Our code for recovering the same branch polarization from polarization calculations has been contributed to pymatgen under the `pymatgen.analysis.ferroelectricity` module. Our code for the DFT and polarization analysis workflows for performing polarization calculations has been contributed to atomate under the `atomate.vasp.workflows.base.ferroelectric` module. We also provide code for the interface that we used to view our candidates in aggregate. The web interface for the current work is hosted at http://blondegeek.github.io/ferroelectric_search_site/. The code for the interface can be found at http://github.com/blondegeek/ferroelectric_search_site.

Received: 10 September 2018; Accepted: 21 January 2020;

Published online: 03 March 2020

References

- Jain, A., Shin, Y. & Persson, K. A. Computational predictions of energy materials using density functional theory. *Nature Reviews Materials* **1**, 15004 (2016).
- Yan, Q. *et al.* Solar fuels photoanode materials discovery by integrating high-throughput theory and experiment. *Proceedings of the National Academy of Sciences* **114**, 3040–3043 (2017).
- Chen, R., Po, H. C., Neaton, J. B. & Vishwanath, A. Topological materials discovery using electron filling constraints. *Nature Physics* **14**, 55–61 (2017).
- de Jong, M., Chen, W., Geerlings, H., Asta, M. & Persson, K. A. A database to enable discovery and design of piezoelectric materials. *Scientific Data* **2**, 150053 (2015).
- Abrahams, S. C. Structurally Based Prediction of Ferroelectricity in Inorganic Materials with Point Group 6mm. *Acta Crystallographica Section B* **B44**, 585–595 (1988).
- Abrahams, S. C. Structurally based predictions of ferroelectricity in seven inorganic materials with space group Pba2 and two experimental confirmations. *Acta Crystallographica Section B* **45**, 228–232 (1989).
- Capillas, C. *et al.* A new computer tool at the Bilbao Crystallographic Server to detect and characterize pseudosymmetry. *Zeitschrift für Kristallographie* **226**, 186–196 (2011).

8. Igartua, J. M., Aroyo, M. I., Kroumova, E. & Perez-Mato, J. M. Search for Pnma materials with high-temperature structural phase transitions. *Acta Crystallographica Section B Structural Science* **55**, 177–185 (1999).
9. Kroumova, E., Aroyo, M. I. & Perez-Mato, J. M. Prediction of new displacive ferroelectrics through systematic pseudosymmetry search. Results for materials with Pba2 and Pmc21 symmetry. *Acta Crystallographica Section B Structural Science* **58**, 921–933 (2002).
10. Stokes, H. T., Sadate, C., Hatch, D. M., Boyer, L. L. & Mehl, M. J. Analysis of the ferroelectric phase transition in BaAl₂O₄ by group theoretical methods and first-principles calculations. *Physical Review B* **65**, 064105 (2002).
11. Perez-Mato, J. M., Withers, R. L., Larsson, A. K., Orobengoa, D. & Liu, Y. Distortion modes and related ferroic properties of the stuffed tridymite-type compounds SrAl₂O₄ and BaAl₂O₄. *Physical Review B - Condensed Matter and Materials Physics* **79**, 1–12 (2009).
12. Smith, E. H., Benedek, N. A. & Fennie, C. J. Interplay of Octahedral Rotations and Lone Pair Ferroelectricity in CsPbF₃. *Inorganic Chemistry* **54**, 8536–8543 (2015).
13. Bennett, J. W. Discovery and Design of Functional Materials: Integration of Database Searching and First Principles Calculations. *Physics Procedia* **34**, 14–23 (2012).
14. Bennett, J. W., Garrity, K. F., Rabe, K. M. & Vanderbilt, D. Hexagonal ABC semiconductors as ferroelectrics. *Physical Review Letters* **109**, 167602 (2012).
15. Bennett, J. W. & Rabe, K. M. Integration of first-principles methods and crystallographic database searches for new ferroelectrics: Strategies and explorations. *J. Solid State Chem.* **195**, 21–31 (2012).
16. Garrity, K. F. High-throughput first-principles search for new ferroelectrics. *Phys. Rev. B* **97**, 024115 (2018).
17. Lines, M. E. & Glass, A. M. *Principles and Applications of Ferroelectrics and Related Materials* (Oxford University Press, 2001).
18. Villars, P. (Chief Editor). PAULING FILE in: *Inorganic Solid Phases*, SpringerMaterials (online database), Springer, Heidelberg (ed.) (2012).
19. Shiozaki, Y., Nakamura, E. & Mitsui, T. (eds.) *Landolt-Börnstein - Group III Condensed Matter _Volume 36A1: "Ferroelectrics and Related Substances: Oxides Part 1: Perovskite-type Oxides and LiNbO₃ Family"* (Springer-Verlag, 2001).
20. Shiozaki, Y., Nakamura, E. & Mitsui, T. (eds.) *Landolt-Börnstein - Group III Condensed Matter _Volume 36A2: "Ferroelectrics and Related Substances: Oxides Part 2: Oxides other than Perovskite-type and LiNbO₃ Family"* (Springer-Verlag, 2002).
21. Shiozaki, Y., Nakamura, E. & Mitsui, T. (eds.) *Landolt-Börnstein - Group III Condensed Matter _Volume 36B1: "Ferroelectrics and Related Substances: Inorganic Substances other than Oxides Part 1: SbSI family... TAAP"* (Springer-Verlag, 2004).
22. Shiozaki, Y., Nakamura, E. & Mitsui, T. (eds.) *Landolt-Börnstein - Group III Condensed Matter _Volume 36B2: "Ferroelectrics and Related Substances: Inorganic Substances other than Oxides Part 2: (NH₄)SO₄ family... K3BiCl₆ _2KCl_KH3F₄"* (Springer-Verlag, 2006).
23. Shiozaki, Y., Nakamura, E. & Mitsui, T. (eds.) *Landolt-Börnstein - Group III Condensed Matter _Volume 36C: "Ferroelectrics and Related Substances: Organic crystals, liquid crystals and polymers"* (Springer-Verlag, 2006).
24. Jain, A. *et al.* The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002 (2013).
25. Curtarolo, S. *et al.* AFLOW: An automatic framework for high-throughput materials discovery. *Computational Materials Science* **58**, 218–226 (2012).
26. Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Computational Materials Science* **111**, 218–230 (2016).
27. Bergerhoff, G. & Brown, I. *Inorganic crystal structure database* (International Union of Crystallography, 1987).
28. Hellenbrandt, M. The inorganic crystal structure database (ICSD)|present and future. *Crystallography Reviews* **10**, 17–22 (2004).
29. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).
30. Kirklin, S. *et al.* The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **1**, 15010 (2015).
31. Lines, M. E. & Glass, A. M. chap. Basic Concepts (Oxford University Press, Oxford, 2001).
32. Rabe, K. M., Dawber, M., Lichtensteiger, C., Ahn, C. H. & Triscone, J.-M. Modern physics of ferroelectrics: Essential background. In *Physics of Ferroelectrics: A Modern Perspective*, 1–30 (Springer Berlin Heidelberg, 2007).
33. Strukov, B. A. & Levanyuk, A. P. General characteristics of structural phase transitions in crystals. In *Ferroelectric Phenomena in Crystals*, 1–29 (Springer Berlin Heidelberg, 1998).
34. Chandra, P. & Littlewood, P. B. A Landau primer for ferroelectrics. In *Physics of Ferroelectrics: A Modern Perspective*, 69–116 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007).
35. Stokes, H. T. & Hatch, D. M. *Isotropy Subgroups of the 230 Crystallographic Space Groups* (WORLD SCIENTIFIC, 1989).
36. Müller, U. Symmetry relations at phase transitions. In *Symmetry Relationships between Crystal Structures*, 196–215 (Oxford University Press, 2013).
37. Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319 (2013).
38. Mathew, K. *et al.* Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Computational Materials Science* **139**, 140–152 (2017).
39. Resta, R. Theory of the electric polarization in crystals. *Ferroelectrics* **136**, 51–55 (1992).
40. Aroyo, M. I. (ed.) *International Tables for Crystallography Volume A: Space-group symmetry* (International Union of Crystallography, 2016).
41. Müller, U. Subgroups and supergroups of point and space groups. In *Symmetry Relationships between Crystal Structures*, 86–99 (Oxford University Press, 2013).
42. King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. *Phys. Rev. B* **47**, 1651–1654 (1993).
43. Vanderbilt, D. & King-Smith, R. D. Electric polarization as a bulk quantity and its relation to surface charge. *Physical Review B* **48**, 4442–4455 (1993).
44. Resta, R. Macroscopic polarization in crystalline dielectrics: the geometric phase approach. *Reviews of Modern Physics* **66**, 899–915 (1994).
45. Rabe, K. M. & Waghmare, U. V. First-principles model hamiltonians for ferroelectric phase transitions. *Ferroelectrics* **136**, 147–156 (1992).
46. Resta, R. & Vanderbilt, D. Theory of polarization: A modern approach. In *Physics of Ferroelectrics: A Modern Perspective*, 31–68 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007).
47. Spaldin, N. A. A beginner's guide to the modern theory of polarization. *Journal of Solid State Chemistry* **195**, 2–10 (2012).
48. Rabe, K. M. & Ghosez, P. First-principles studies of ferroelectric oxides. In *Topics in Applied Physics*, 117–174 (Springer Berlin Heidelberg, 2007).
49. Kittel, C. Theory of antiferroelectric crystals. *Phys. Rev.* **82**, 729–732 (1951).
50. Rabe, K. M. Antiferroelectricity in Oxides: A Reexamination. In *Functional Metal Oxides*, 221–244 (Wiley-VCH Verlag GmbH & Co. KGaA, 2013).
51. Tolédano, P. & Guennou, M. Theory of antiferroelectric phase transitions. *Phys. Rev. B* **94**, 014107 (2016).

52. Reyes-Lillo, S. E. & Rabe, K. M. Antiferroelectricity and ferroelectricity in epitaxially strained PbZrO_3 from first principles. *Physical Review B* **88**, 180102R (2013).
53. Toyo, A. spglib, <https://atztogo.github.io/spglib/> (2018).
54. Togo, A. & Tanaka, I. Spglib: a software library for crystal symmetry search, arxiv:1808.01590 (2018).
55. Ong, S. P. *et al.* The materials application programming interface (API): A simple, flexible and efficient API for materials data based on REpresentational state transfer (REST) principles. *Computational Materials Science* **97**, 209–215 (2015).
56. Aroyo, M. I. *et al.* Crystallography online: Bilbao Crystallographic Server. *Bulg. Chem. Commun.* **43**, 183–197 (2011).
57. Aroyo, M. I., Kirov, A., Capillas, C., Perez-Mato, J. M. & Wondratschek, H. Bilbao crystallographic server. II. representations of crystallographic point groups and space groups. *Acta Crystallographica Section A Foundations of Crystallography* **62**, 115–128 (2006).
58. Aroyo, M. I. *et al.* Bilbao Crystallographic Server: I. Databases and crystallographic computing programs. *Zeitschrift für Kristallographie - Crystalline Materials* **221**, 15–27 (2006).
59. Bilbao Crystallographic Server. Structure relations, <http://www.cryst.ehu.es/cryst/rel.html> (2019).
60. Stokes, H. T., Hatch, D. M. & Campbell, B. J. ISOTROPY Software Suite, iso.byu.edu (2019).
61. Aas, G., Lee, J., & Lester, A. Mechanize - stateful programmatic web browsing, version 0.4.4, <https://pypi.python.org/pypi/mechanize/> (2019).
62. Scott, J. F. Searching for new ferroelectrics and multiferroics: A users point of view. *npj Computational Materials* **1**, 15006 (2015).
63. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **6**, 15–50 (1996).
64. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **47**, 558–561 (1993).
65. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **54**, 11169 (1996).
66. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical review letters* **77**, 3865–3868 (1996).
67. Blöchl, P. E. Projector augmented-wave method. *Physical Review B* **50**, 17953 (1994).
68. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **59**, 1758–1775 (1999).
69. Materials Project. Materials project calculation guide, <https://materialsproject.org/docs/calculations> (2019).
70. King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. *Physical Review B* **47**, 1651–1654 (1993).
71. Resta, R. Berry phase in electronic wavefunctions. *Troisième Cycle de la Physique en Suisse Romande* 899–915 (Anne Académique 1995–1996).
72. Jain, A. *et al.* Formation enthalpies by mixing GGA and GGA + U calculations. *Physical Review B - Condensed Matter and Materials Physics* **84**, 1–10 (2011).
73. Jain, A. *et al.* A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science* **50**, 2295–2310 (2011).
74. Dudarev, S. L., Savrasov, S. Y., Humphreys, C. J. & Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Physical Review B* **57**, 1505–1509 (1998).
75. Jain, A. *et al.* FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurrency Computation Practice and Experience* **27**, 5037–5059 (2015).
76. Vanderbilt, D. Berry-phase theory of proper piezoelectric response. *Journal of Physics and Chemistry of Solids* **61**, 147–151 (2000).
77. Smidt, T. E., Mack, S. A., Reyes-Lillo, S. E., Jain, A. & Neaton, J. B. Ferroelectric search distortion and workflow data and VASP files. <https://doi.org/10.6084/m9.figshare.6025634> (2019).
78. Smidt, T. E., Mack, S. A., Reyes-Lillo, S. E., Jain, A. & Neaton, J. B. blondegeek/ferroelectric search site: Ferroelectric Search Site. *Zenodo*. <https://doi.org/10.5281/zenodo.3542303> (2019).
79. Ong, S. P., Wang, L., Kang, B. & Ceder, G. Li-fe-p-o2 phase diagram from first principles calculations. *Chemistry of Materials* **20**, 1798–1807 (2008).
80. Virtanen, P. *et al.* SciPy 1.0—Fundamental Algorithms for Scientific Computing in Python, <https://arxiv.org/abs/1907.10121> (2019).
81. Armiento, R., Kozinsky, B., Fornari, M. & Ceder, G. Screening for high-performance piezoelectrics using high-throughput density functional theory. *Phys. Rev. B* **84**, 014103 (2011).
82. Oliphant, T. NumPy: A guide to NumPy. USA: Trelgol Publishing, <http://www.numpy.org/> (2006–).
83. van der Walt, S., Colbert, S. C. & Varoquaux, G. The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering* **13**, 22–30 (2011).
84. Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* **9**, 90–95 (2007).
85. Prez, F. & Granger, B. E. IPython: A system for interactive scientific computing. *Computing in Science & Engineering* **9**, 21–29 (2007).
86. Kluyver, T. *et al.* Jupyter notebooks – a publishing format for reproducible computational workflows. In Loizides, F. & Schmidt, B. (eds.) *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 87–90 (IOS Press, 2016).
87. Wahl, R., Vogtenhuber, D. & Kresse, G. SrTiO_3 and BaTiO_3 revisited using the projector augmented wave method: Performance of hybrid and semilocal functionals. *Phys. Rev. B* **78**, 104116 (2008).
88. Zhang, Y., Sun, J., Perdew, J. P. & Wu, X. Comparative first-principles studies of prototypical ferroelectric materials by LDA, GGA, and SCAN meta-GGA. *Phys. Rev. B* **96**, 035143 (2017).
89. Yang, Q., Cao, J. X., Ma, Y. & Zhou, Y. C. First principles study of polarization-strain coupling in $\text{SrBi}_2\text{Ta}_2\text{O}_9$. *AIP Advances* **3**, 052134 (2013).
90. Lebedev, A. I. Ferroelectric phase transition in orthorhombic CdTiO_3 : First-principles studies. *Physics of the Solid State* **51**, 802–809 (2009).
91. Baettig, P., Schelle, C. F., LeSar, R., Waghmare, U. V. & Spaldin, N. A. Theoretical prediction of new high-performance lead-free piezoelectrics. *Chemistry of Materials* **17**, 1376–1380 (2005).
92. Adachi, M. *et al.* $\text{Mg}_3\text{B}_7\text{O}_{13}\text{Cl}$ [F], 18A-1: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A2: “Oxides” in SpringerMaterials (2002).
93. Adachi, M. *et al.* BaTiO_3 (cubic) [F] Survey, 1A-10: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A1: “Oxides” in SpringerMaterials (2001).
94. Adachi, M. *et al.* $\text{PbBi}_2\text{Nb}_2\text{O}_9$ [(F)], 9A-10: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A2: “Oxides” in SpringerMaterials (2002).
95. Adachi, M. *et al.* $\text{SrBi}_2\text{Ta}_2\text{O}_9$ [F], 9A-12: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A2: “Oxides” in SpringerMaterials (2002).
96. Adachi, M. *et al.* $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ [F], 9A-15: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A2: “Oxides” in SpringerMaterials (2002).
97. Deguchi, K., Hasebe, K., Gesi, K. & Asahi, T. 39A-16 K_2ZnBr_4 [F]: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36B2: “Inorganic Substances other than Oxides” in SpringerMaterials (2005).
98. Shiozaki, Y. & Asahi, T. 50A-1: 1-6 $\text{SC}(\text{NH}_2)_2$ [F]: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36C: “Organic crystals, liquid crystals and polymers” in SpringerMaterials (2006).

99. Adachi, M. *et al.* Oxides - CdTiO₃, 1A-9: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A1: "Oxides" in SpringerMaterials (2001).
100. Deguchi, K., Hasebe, K., Gesi, K. & Asahi, T. 39A-9 K₂ZnCu₄[F]: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36B2: "Inorganic Substances other than Oxides" in SpringerMaterials (2005).
101. Deguchi, K., Hasebe, K., Gesi, K. & Asahi, T. 39A-10 Rb₂ZnCu₄[F]: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36B2: "Inorganic Substances other than Oxides" in SpringerMaterials (2005).
102. Deguchi, K., Hasebe, K., Gesi, K. & Asahi, T. 39A-1 (NH₄)₂SO₄[F]: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36B2: "Inorganic Substances other than Oxides" in SpringerMaterials (2005).
103. Adachi, M. *et al.* PbHfO₃ [(A)], 1A-16: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A1: "Oxides" in SpringerMaterials (2001).
104. Deguchi, K., Hasebe, K., Gesi, K. & Asahi, T. 39A-2 K₂SeO₄[F]: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36B2: "Inorganic Substances other than Oxides" in SpringerMaterials (2005).
105. Adachi, M. *et al.* Oxides - KNbO₃ [F], 1A-2: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A1: "Oxides" in SpringerMaterials (2001).
106. Wemple, S., DiDomenico, M. Jr. & Camlibel, I. Relationship between linear and quadratic electro-optic coefficients in linbo₃, litao₃, and other oxygen-octahedra ferroelectrics based on direct measurement of spontaneous polarization. *Applied Physics Letters* **12**, 209–211 (1968).
107. Adachi, M. *et al.* LiNbO₃ [F] Survey, 2A-1: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A1: "Oxides" in SpringerMaterials (2001).
108. Adachi, M. *et al.* SbNbO₄ [F, (A)], 5A-2: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A2: "Oxides" in SpringerMaterials (2002).
109. Adachi, M. *et al.* PbTiO₃ [F], 1A-11: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A1: "Oxides" in SpringerMaterials (2001).
110. Terauchi, H. *et al.* EXAFS Studies on Amorphous PbTiO₃. *Journal of the Physical Society of Japan* **50**, 3977–3979 (1981).
111. Adachi, M. *et al.* PbZrO₃ [A], 1A-15: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A1: "Oxides" in SpringerMaterials (2001).
112. Hikita, T. 35A-13 TlTiOPO₄[(F)]: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36B1: "Inorganic Substances other than Oxides. Part 1: SbSI family.. TAAP" in SpringerMaterials (2004).
113. Adachi, M. *et al.* Sr₂Ta₂O₇ [F], 8A-6: Datasheet from Landolt-Börnstein - Group III Condensed Matter · Volume 36A2: "Oxides" in SpringerMaterials (2002).
114. Abrahams, S. C. Systematic prediction of new ferroelectrics in space group R3. I. *Acta crystallographica. Section B, Structural science* **62**(Pt 1), 26–41 (2006).
115. Huang, S. Y., Von Der Mühl, R., Ravez, J. & Hagenmüller, P. Structural, ferroelectric and pyroelectric properties of nonstoichiometric ceramics based on BaAl₂O₄. *Journal of Physics and Chemistry of Solids* **55**, 119–124 (1994).
116. Machida, M. Rotational Dynamics of Trimethylammonium Chloride Studied by MD Simulation and NMR. *Ferroelectrics* **338**, 205–209 (2006).
117. Smith, E. H., Benedek, N. A. & Fennie, C. J. Interplay of octahedral rotations and lone pair ferroelectricity in CsPbF₃. *Inorganic Chemistry* **54**, 8536–8543 (2015).
118. Yeliseyev, A. P. *et al.* Structures and optical properties of two phases of srmg₄. *Phys. Chem. Chem. Phys.* **17**, 500–508 (2015).

Acknowledgements

This work supported by the Materials Project funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05-CH11231: Materials Project program KC23MP. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Additional calculations were performed on the High Performance Computing clusters at Lawrence Berkeley Lab. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy, and Laboratory Directed Research and Development Program at the Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231. T.E.S. acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400. S.E.R. acknowledges partial support from FONDECYT Iniciación en Investigación under Grant No. 11180590. A.J. was funded by the U.S. Department of Energy, Office of Basic Energy Sciences, Early Career Research Program.

Author contributions

All authors contributed to the outline of the workflow. T.E.S. developed the code for the workflow including executing polarization calculations, post-processing calculations to recover the same branch polarization and assess the quality of candidates, and visualizing the results through the graphic web interface. T.E.S. and S.A.M. performed the density functional theory calculations and validated the dataset through comparison to the Pauling Files and Landolt-Börnstein series. T.E.S. and S.E.R. performed the literature search to identify known ferroelectrics in the dataset. S.E.R. and A.J. contributed to verification of the workflow. S.E.R., A.J. and J.B.N. were involved in supervising and planning the work. All authors contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.B.N.

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

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020