What can one learn about material structure given a single first-principles calculation?

Nicholas Rajen[1] and Sinisa Coh[1,2]

[1]*Materials Science and Engineering, University of California Riverside, Riverside, California 92521, USA*
[2]*Mechanical Engineering, University of California Riverside, Riverside, California 92521, USA*

We extract a variable X from electron orbitals $\Psi_{n\mathbf{k}}$ and energies $E_{n\mathbf{k}}$ in the parent high-symmetry structure of a wide range of complex oxides: perovskites, rutiles, pyrochlores, and cristobalites. Even though calculation was done only in the parent structure, with no distortions, we show that X dictates material's true ground-state structure. We propose using Wannier functions to extract concealed variables such as X both for material structure prediction and for high-throughput approaches.

I. INTRODUCTION AND MOTIVATION

There is a growing need for machines to learn about matter from results of a large set of first-principles computer calculations. However, a single such calculation produces $\sim 10^6$–10^9 bits of information and it is unclear how machines can make use of them [1,2]. Therefore, so far, machines set aside most of these bits—such as those describing electron orbitals $\Psi_{n\mathbf{k}}$—and try to learn from the simplest calculated quantities such as the total energy E^{tot} or similar.

One such example where machines are learning from E^{tot} are methods used for predicting positions of atoms in a yet unsynthesized material. These methods look for a set of atom positions ξ that minimize the total energy $E^{\text{tot}}(\xi)$ of the material. Total energy of a periodic solid can be computed from first principles using various approximations to the density functional theory [3,4]. However, minimization of $E^{\text{tot}}(\xi)$ still remains an unsolved problem as ξ is a vector in a very highly dimensional space (it has $\approx 3N$ dimensions and N is a number of atoms in the material). Nevertheless, this problem can be addressed heuristically using machine learning techniques such as evolutionary algorithms [5] or particle swarm optimization [6]. Broadly speaking these methods first use E^{tot} calculated for a wide range of different candidate structures $\xi_1, \xi_2, \ldots, \xi_m$, but the same chemical composition, to learn about the underlying interactions in that material. Next, given this knowledge, the machine makes an informed guess for the next structure ξ_{m+1} and the process repeats until an optimal structure is found. Clearly, if one used in this process information contained in electron orbitals $\Psi_{n\mathbf{k}}$, and not only E^{tot}, one could make a more informed guess of structure ξ_{m+1}.

The need for a machine to learn in the context of materials science is also relevant for the so-called high-throughput approaches such as the materials project [7], the aflow [8], the oqmd [9], aiida [10], or the nomad [11] materials databases. While in the structure prediction problem one considers a single chemical composition at the time, in the high-throughput approach one wishes to learn about materials with a wide range of chemical compositions. Since these databases contain $\sim 10^5$–10^6 materials their total information content, if one were to store electron orbitals, is about 10^{14}–10^{15} bits.

In this paper we do not focus on what or how machines can learn from total energy E^{tot}. Instead, the goal of this paper is to construct a descriptor of electron orbitals $\Psi_{n\mathbf{k}}$ and eigenenergies that can be used for learning in structure prediction and high-throughput approaches. As a proof of principle, we construct here a descriptor—denoted as X—that is strongly correlated with the preferred crystal structure of a material, as described later. Since in our proof of principle work X turns out to be a single number, one can establish a correlation between X and structure just by inspection, without using machine learning. However, for materials with lower symmetry or in the cases of other properties of interest (i.e., not crystal structure) similarly constructed descriptors of electron orbitals will correspond to more than one number and one would therefore have to use machine learning.

Crystal structures of materials can be divided somewhat loosely into families of structures [12] based on polyhedral units present in the structure and their connectivity. Each structure family is derived from a simple high-symmetry structure also called parent, aristotype, or prototype structure. Remaining structures in the family are then derived from the parent by either displacing or substituting atoms [13,14]. These derivative structures are also called hettotypes. When structures are related to each other by distortion that does not preserve polyhedral units or their connectivity, we refer to them only as polymorphs.

This paper is structured as follows. In Sec. II we describe our approach and in Sec. III we present and discuss our results. A comparison of our results with atomic descriptors (Pettifor maps) is done in Sec. III.

II. APPROACH

In what follows we present a general approach to extract a variable X—given electron orbitals $\Psi_{n\mathbf{k}}$ and eigenenergies of a parent structure alone—that dictates which structure is a true ground-state structure of that material. Since such a variable knows about material's structure (i) it has the potential to be used in machine learning approaches described earlier. To that end, a useful variable should also satisfy two additional criteria: (ii) X should be represented with as few numbers as possible,
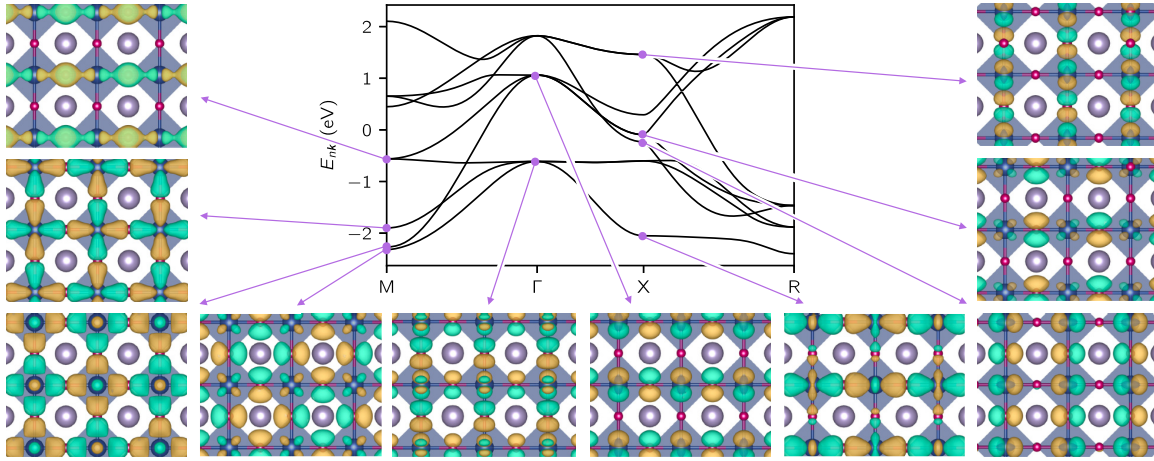
FIG. 1. Selection of several electron orbitals Ψ_{nk} in the high-symmetry cubic phase of SrTiO$_3$ perovskite. Yellow/green: positive/negative isosurfaces of electron orbitals. Blue: Ti-O octahedra. Red: oxygen atoms.

so that learning is as efficient as possible; and (iii) X should be calculable automatically for a given material without any input from the user, so that it can be used by a machine.

To satisfy the criterion (i) we should ideally extract X in some way from the fundamental variable of the problem. In the case of quantum-mechanical descriptions of matter, the fundamental variable is the electron many-body wave function $\widetilde{\Psi}$. We note that the many-body wave function $\widetilde{\Psi}$ is formally not accessible in the commonly used Kohn-Sham [4] scheme for the density functional theory. Therefore we instead use here the so-called Kohn-Sham orbitals Ψ_{nk} (and corresponding eigenenergies E_{nk}) as they are typically reasonably good approximations [15] of the quasiparticle wave function and energy (here n is a band index while k is a crystal momentum). Figure 1 shows some of the Kohn-Sham orbitals for one of the materials we studied in this work (SrTiO$_3$ cubic perovskite).

Orbitals Ψ_{nk} themselves are not a good choice of X as they do not satisfy criterion (ii). Namely, even in the simplest calculations (say, cubic silicon), a machine needs $\sim10^6$ bits to describe Ψ_{nk} reasonably accurately. As discussed later, we circumvent this difficulty by defining X in terms of the Wannier functions [16], as they provide a compact and faithful representation of Ψ_{nk} (see Fig. 2). Wannier functions can be computed without user input [17] so our approach satisfies condition (iii) as well.

To demonstrate the generality of our approach we consider here a diversified set of four structure families: ABO_3 perovskites, AO_2 rutiles, $A_2B_2O_7$ pyrochlores, and AO_2 cristobalites for a range of A and B anions. We chose these four families because they display different bonding environments containing both octahedral and tetrahedral oxygen units, as well as different connectivity of these units (see Fig. 3). The main result of this paper is that we find that our general approach yields an energy scale X that is very well correlated with preferred ground-state structure. We studied 64 representative compounds: 17 perovskites, 16 rutiles, 18 pyrochlores, and 13 cristobalites.

The variable X is constructed in five steps. *First*, we relax the structure of each compound while preserving the symmetry of the parent structure (aristotype) [18].

Second, we identify a set of electron bands of interest given the band structure from the first step. In our case we simply take as bands of interest a complex of bands with a dominant oxygen p-like character, as they are the most dispersive and thus contain the most information about the interatomic interactions in the material. Oxygen bands are also fully occupied and isolated in these materials, which simplifies our Wannier function based analysis. Figure 2 shows bands of interest in the representative case of each of the four families we studied.

Third, we convert electron orbitals Ψ_{nk} of bands selected in step two into a basis of maximally localized Wannier functions W_n [16,19,20]. This procedure significantly simplifies our analysis as Wannier functions are typically strongly localized on a single atom [20,21] and are therefore more convenient descriptors of the chemical environment of a material. Given a set of extended periodic orbitals Ψ_{nk} Wannier functions W_n are defined as

$$W_n(\boldsymbol{r}) = \frac{1}{N_k} \sum_k U_{nmk} \Psi_{mk}(\boldsymbol{r}), \qquad (1)$$

where N_k is a total number of k points. While unitary matrices U_{nmk} are in principle arbitrary, in the case of maximally localized Wannier functions they are chosen so that the resulting function $W_n(\boldsymbol{r})$ is as localized in real space as possible [20]. We stress that basis change in Eq. (1) is exact in the sense that the vector space spanned by W_n (and its periodic images) is the same as the space spanned by Ψ_{mk}. However, W_n is a more convenient object to study than Ψ_{nk} since a single function W_n contains the same information as an entire band of Bloch functions Ψ_{nk} (there is one function for each k point). Therefore any information contained in bands is also contained in the corresponding Wannier functions. One of the Wannier functions for one of the parent structures (perovskites) is denoted in Fig. 2(a) as $W_\parallel$. This Wannier function corresponds to the oxygen p-like function oriented along the B-O-B line (here B is the anion in the ABO_3 perovskite that is in the center of the oxygen octahedron). The remaining two functions ($W_\perp$) centered on the same oxygen atom are perpendicular to the B-O-B line [see also Fig. 2(a)]. In the remaining three families (rutile, pyrochlore, cristobalite) we also find that Wannier
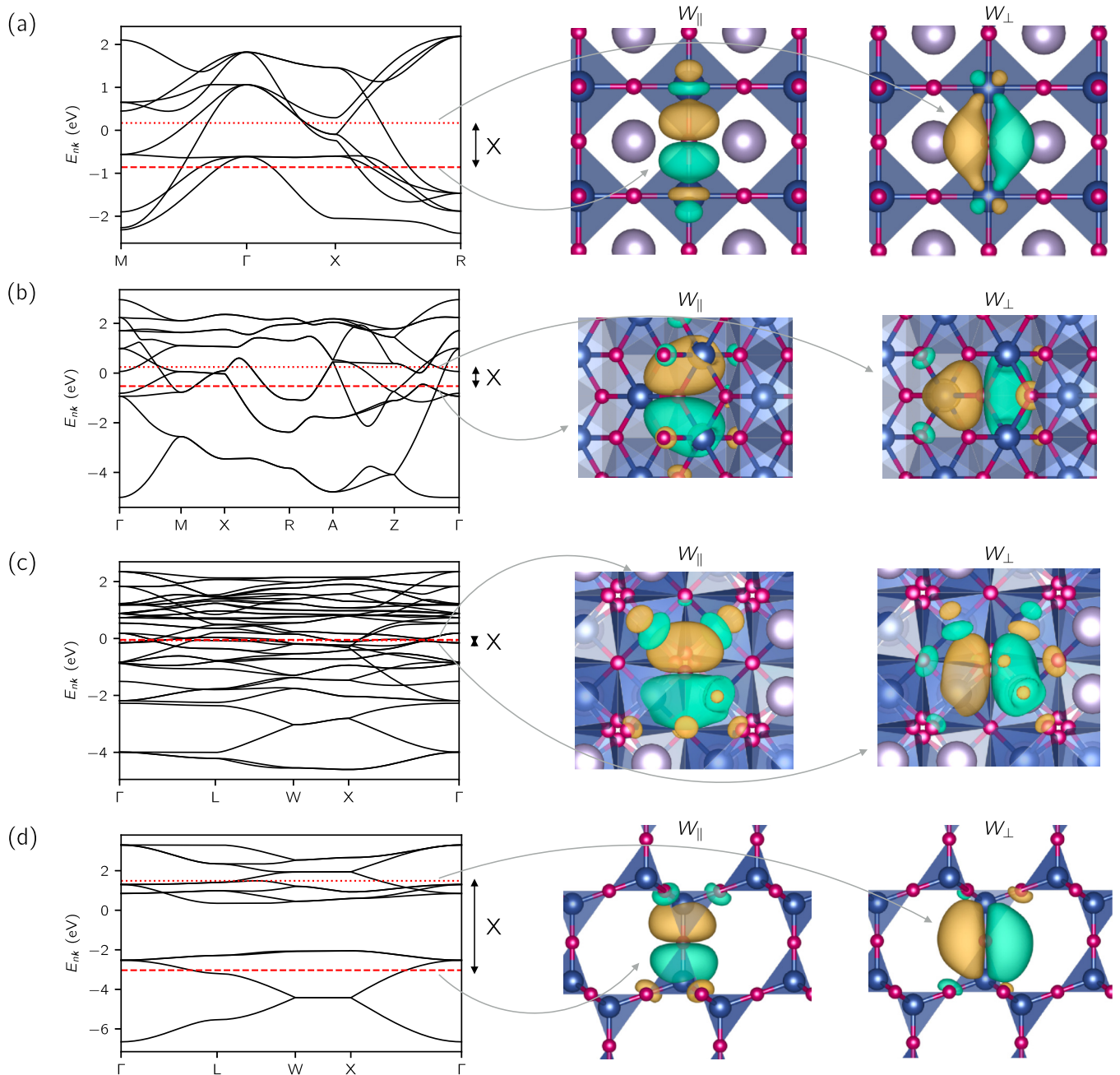
FIG. 2. Localized Wannier functions (right) and band structure (left) for representatives of all four classes of compounds we studied [SrTiO$_3$ perovskite (a), SnO$_2$ rutile (b), Y$_2$Sn$_2$O$_7$ pyrochlore (c), and SiO$_2$ cristobalite (d)]. Yellow/green: positive/negative isosurfaces of Wannier functions $W_\perp$ and $W_\parallel$. Dotted/dashed red line: $\langle W_\perp|H|W_\perp \rangle$ and $\langle W_\parallel|H|W_\parallel \rangle$. The difference is indicated with a black arrow (X).

function on each oxygen atom follows the same logic: there is a singlet orbital (denoted again as $W_\parallel$) and a degenerate pair of orbitals ($W_\perp$). These Wannier functions are shown in Figs. 2(b)–2(d).

Fourth, we construct the representation of the Hamiltonian operator H in the Wannier basis from the previous step,

$$H_{nm\boldsymbol{R}} = \langle W_n|H|W_{m\boldsymbol{R}}\rangle. \quad (2)$$

Here $W_{m\boldsymbol{R}}$ is defined as W_m translated by a lattice vector $\boldsymbol{R}$, $W_{m\boldsymbol{R}}(\boldsymbol{r}) = W_m(\boldsymbol{r} - \boldsymbol{R})$. It is straightforward to show that for structural properties we need not consider $H_{nm\boldsymbol{R}}$ when

either $n \neq m$ or $\boldsymbol{R} \neq \boldsymbol{0}$. These terms are often referred to as hopping integrals. Hopping integrals do not contribute to the integrated band energy $\int_{\mathrm{BZ}} E_{n\boldsymbol{k}}d\boldsymbol{k}$ of a fully occupied band and therefore they do not have the learning potential for determining ground-state structure [22]. Therefore, we are left with $H_{nm\boldsymbol{R}}$ when both $n = m$ and $\boldsymbol{R} = \boldsymbol{0}$. The matrix element $H_{nn\boldsymbol{0}}$ is usually referred to as the on-site energy of the nth Wannier function. The absolute value of on-site energy is ill-defined for a periodic solid as one can change its value by adding an arbitrary constant C to the Hamiltonian of the periodic solid: $H \rightarrow H + C$. However, changing H to $H + C$
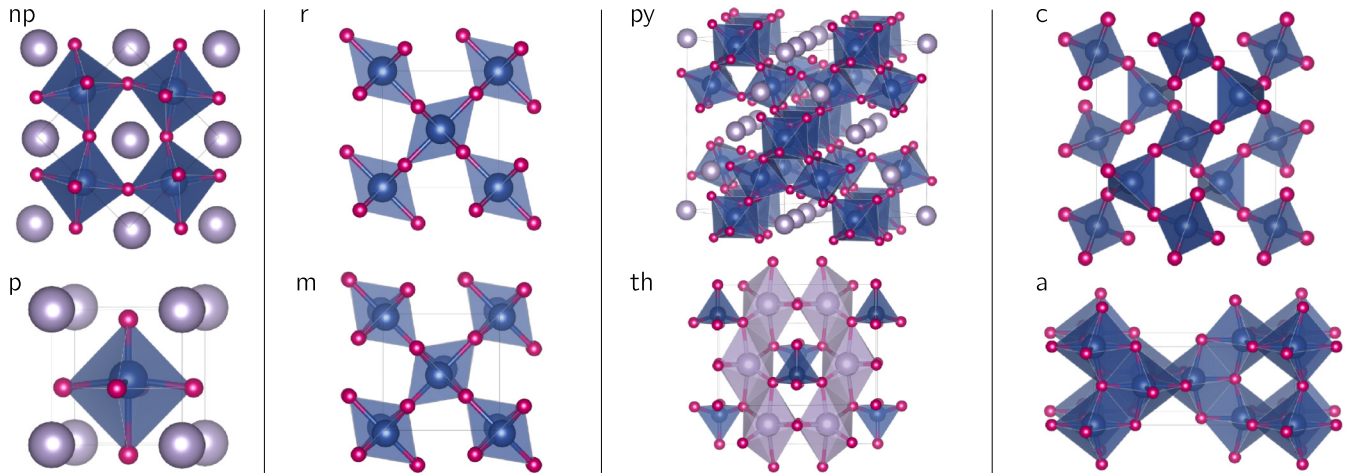
FIG. 3. Crystal structures of the four family classes we considered: perovskites (np and p), rutiles (r and m), pyrochlores (py and th), and cristobalites (c and a). Meaning of symbols is the same as in Table I.

changes on-site energies of all Wannier functions by the same amount. Therefore, even though absolute values are ill-defined, the differences of on-site energies,

$$H_{nn\mathbf{0}} - H_{mm\mathbf{0}} \qquad (3)$$

are well defined.

Fifth, we use symmetry to find all distinct values of $H_{nn\mathbf{0}} - H_{mm\mathbf{0}}$ from the previous step. Since the symmetry of all four parent structures is high, we find that in all four families there remains only one distinct numerical value in which n corresponds to $W_{\perp}$ and m to $W_{\parallel}$. This difference we denote as X,

$$X = \langle W_{\perp}|H|W_{\perp}\rangle - \langle W_{\parallel}|H|W_{\parallel}\rangle. \qquad (4)$$

As an example, Fig. 2(a) shows a complex of oxygen p-like electron bands (black) in SrTiO$_3$ along with calculated $\langle W_{\perp}|H|W_{\perp}\rangle$ and $\langle W_{\parallel}|H|W_{\parallel}\rangle$ (dotted/dashed red line). The black arrow indicates their difference (X). The numerical value of X for all materials we studied is provided in Table I.

We stress here that the energy scale X cannot be inferred from the electron band energies $E_{n\mathbf{k}}$ alone. Instead, one also needs to use wave functions $\Psi_{n\mathbf{k}}$ in the construction. In addition, X is gauge dependent as a different choice of relative phases of wave functions will lead to a different Wannier function and thus different X. For example, one can show that a simple unitary rotation that rotates $W_{\parallel}$ and $W_{\perp}$ will produce two Wannier functions with the difference in on-site

TABLE I. Variable X calculated in a high-symmetry parent state correlates well with the preferred ground-state structure (third subcolumn). The meanings of the crystal structure abbreviations are as follows: np, nonpolar hettotype; p, polar hettotype; r, rutile hettotype; m, manganite hettotype; n, NbO$_2$ hettotype; py, pyrochlore; th, thortveitite; c, cristobalite; a, anatase.

| Perovskite			Rutile			Pyrochlore			Cristobalite		
ABO_3	X (eV)	Struc.	AO_2	X (eV)	Struc.	$A_2B_2O_7$	X (eV)	Struc.	AO_2 ABO_4	X (eV)	Struc.
CaZrO$_3$	0.47	np	ReO$_2$	-0.27	m	Mg$_2$P$_2$O$_7$	-1.47	th	InSbO$_4$	3.62	a
SrZrO$_3$	0.57	np	WO$_2$	-0.26	m	Mg$_2$As$_2$O$_7$	-0.96	th	SnO$_2$	3.68	a
CaHfO$_3$	0.66	np	MoO$_2$	-0.18	m	Y$_2$Si$_2$O$_7$	-0.73	th	PbO$_2$	3.80	a
PbZrO$_3$	0.70	np	TcO$_2$	-0.09	m	In$_2$Si$_2$O$_7$	-0.60	th	AlSbO$_4$	3.95	a
BaZrO$_3$	0.76	np	NbO$_2$	-0.06	n	Cd$_2$V$_2$O$_7$	-0.59	th	GaSbO$_4$	3.98	a
SrHfO$_3$	0.76	np	VO$_2$	0.11	m	Sc$_2$Si$_2$O$_7$	-0.54	th	InPO$_4$	4.00	c
PbHfO$_3$	0.90	np	TiO$_2$	0.17	r	In$_2$Ge$_2$O$_7$	-0.31	th	InAsO$_4$	4.15	c
CdTiO$_3$	0.90	p	CrO$_2$	0.22	r	La$_2$Sn$_2$O$_7$	-0.27	py	GaPO$_4$	4.43	c
CaTiO$_3$	0.91	np	MnO$_2$	0.27	r	Sc$_2$Ge$_2$O$_7$	-0.22	th	AlPO$_4$	4.47	c
BaHfO$_3$	0.96	np	RuO$_2$	0.27	r	Y$_2$V$_2$O$_7$	-0.18	py	SiO$_2$	4.52	c
SrTiO$_3$	1.03	p	OsO$_2$	0.39	r	Bi$_2$Ti$_2$O$_7$	-0.14	py	GaAsO$_4$	4.54	c
PbTiO$_3$	1.20	p	IrO$_2$	0.45	r	Y$_2$Mo$_2$O$_7$	-0.10	py	AlAsO$_4$	4.56	c
BaTiO$_3$	1.25	p	SnO$_2$	0.77	r	Y$_2$Ti$_2$O$_7$	-0.04	py	GeO$_2$	4.67	c
NaNbO$_3$	1.31	p	PbO$_2$	0.82	r	La$_2$Hf$_2$O$_7$	-0.04	py			
KNbO$_3$	1.45	p	SiO$_2$	1.00	r	La$_2$Pb$_2$O$_7$	-0.03	py			
LaAlO$_3$	1.54	np	GeO$_2$	1.06	r	Y$_2$Sn$_2$O$_7$	-0.02	py			
KTaO$_3$	1.66	p				La$_2$Zr$_2$O$_7$	0.01	py			
						Bi$_2$Hf$_2$O$_7$	0.02	py			
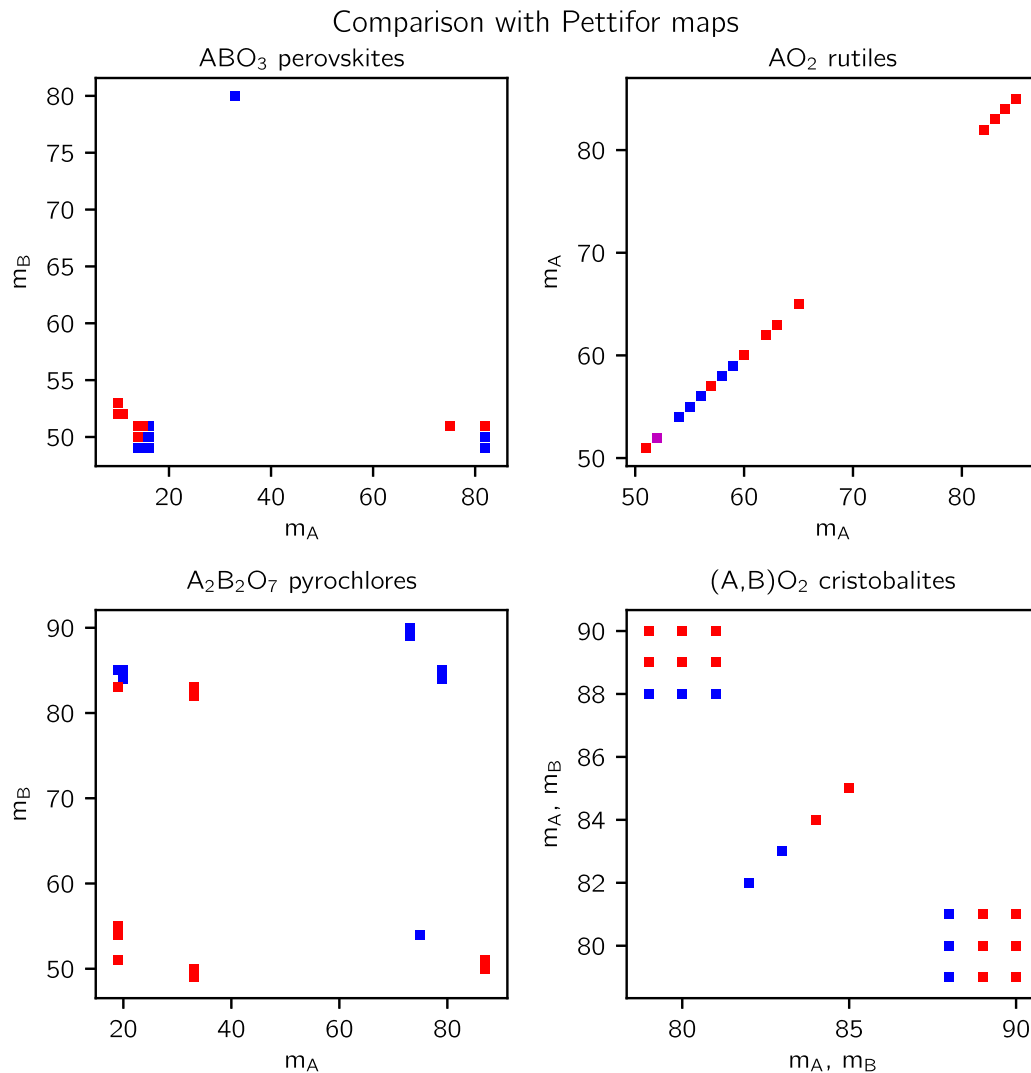
FIG. 4. Comparison with Pettifor maps. Red squares correspond to structures p, r, c, and py while blue crosses correspond to np, m, a, and th. Conventions for structures are the same as in Table I.

energy having any value between $+X$ and $-X$, including zero. However, this procedure will increase the spread of the Wannier functions and will reduce their symmetry. Therefore, in this work we constructed Wannier functions for each family structure in a consistent way that respects the symmetry of the parent phase.

III. RESULTS

Table I contains calculated values of X for the 64 compounds we studied. As can be seen from the table, X is well correlated with the preferred ground-state structure. We discuss now all four structure families in more detail.

Perovskites are one of the most studied complex oxides. The numerical values of X reported in Table I for perovskites was calculated in the parent cubic phase with space group $Pm\bar{3}m$. As can be seen from the table, X is positive for all perovskites and it ranges from 0.47 to 1.66 eV. There is a large number of structures (hettotypes) that derive from the cubic perovskite parent structure. Some of these structures are polar (denoted as p in Table I and Fig. 3) while others are nonpolar and have

rotated oxygen octahedra (np) [23–26]. Our analysis shows that perovskites with X less than $\sim$1 eV tend to condense into a nonpolar state (np) while those with X larger than $\sim$1 eV condense into a polar state (p). The exception is $CdTiO_3$ which is polar but has $X = 0.90$ eV and $LaAlO_3$ which is nonpolar but has $X = 1.54$ eV. While the value of X for $CdTiO_3$ is near the polar–nonpolar boundary, $LaAlO_3$ is not. The anomalous value of X for $LaAlO_3$ likely occurs because its perovskite phase is degenerate with another structure [27].

We note here that the importance of local interactions in perovskites has been discussed earlier in Refs. [25,28–31]. However, we are unaware of any other work in which ground-state density functional calculation of a high-symmetry perovskite alone can be used to infer its low-symmetry ground-state structure.

Rutiles are another common structure of complex oxides [32]. Unlike perovskites, rutiles have a somewhat more complicated structure as their octahedral units are both corner and edge shared. The numerical values of X reported in Table I for rutiles was calculated in the parent rutile structure (r) with tetragonal space group $P4_2/mnm$. As can be seen from

the table, compounds with X larger than 0.17 eV remain in the rutile phase, while those with smaller X distort into a manganite (m) or NbO_2-type (n) structure [32]. Both of these structures contain off-centered A anions and deformed oxygen octahedra. The more common, manganite structure, is monoclinic with space group $P2_1/c$ while the structure unique to NbO_2 is tetragonal $I4_1/a$. In rutiles the calculated value of X ranges between -0.27 and 1.06 eV.

Pyrochlores have the most complex structure among the materials we studied. Variable X was computed in the cubic pyrochlore (py) state with space group $Fd\bar{3}m$. We find that compounds with X above -0.2 eV remain in the pyrochlore state at their ground state while others form into thortveitite (th) [33] polymorph [34] with monoclinic space group $C2/m$. The only outlier we found is $La_2Sn_2O_7$ as it has $X = -0.27$ eV but, as far as we are aware, it remains in a pyrochlore structure at a ground state.

Cristobalite is a well-known structure in which anions are surrounded with oxygen tetrahedra, unlike the other three families in which oxygen atoms form octahedral units. The numerical values of X reported in Table I for cristobalites were calculated in the parent cubic idealized cristobalite structure (space group $F\bar{4}3m$). We again find a strong correlation between X and the preferred ground-state structure. Compounds with X larger than ~ 4 eV remain cristobalite (with small deviations known as α or β cristobalite, space groups $P4_12_12$ and $I\bar{4}2d$). Those with X less than ~ 4 eV are unstable in the cristobalite structure and distort instead into anatase phase (a) with tetragonal space group $I4_1/amd$ [34]. In the cristobalites we studied the calculated values of X range from 3.62 to 4.67 eV.

Comparison with Pettifor maps

While descriptor X in this work was extracted from a first-principles calculation of a periodic solid, there are several descriptors in use that can be extracted just based on the chemical composition of the solid. One such example is Pettifor maps [35] which assign a phenomenological chemical scale χ to each chemical element [36] in the periodic table. Figure 4 shows such maps for all of the materials studied in this work. Horizontal and vertical axes in these plots correspond to the ordering of χ for metal atoms appearing in each compound (this ordering is also called Mendeleev number m).

As can be seen from the Fig. 4, Pettifor maps for these materials correlate well with the ground-state structure only in the case of cristobalites. Namely, compounds with smaller m (χ) tend to deform into anatase phase (blue cross symbols) while those with larger m tend to remain in the cristobalite phase (red square symbols). In the other three classes of materials correlations with Pettifor maps are not as good. For example, for rutiles we find that those with very large m tend to remain in a rutile phase, but those with lower m can be either rutile or manganite. In fact, the one with the lowest m (TiO_2) is in the rutile phase. Similarly, no obvious correlation is found for perovskites and pyrochlores.

IV. CONCLUSION

Our work shows that it is possible to extract information from a first-principles calculation on a parent phase and correlate it with the preference for a ground-state structure. While few of these correlations can be found using empirical parameters such as Pettifor maps, or ionic radii, our work shows that one can also extract relevant parameters from the fundamental variable in the problem (Ψ_{nk}).

We expect that one could use this approach not only to characterize Hamiltonian operator H but also any other quantum-mechanical operator, such as electron-light interaction [37], electron-phonon interaction [38], or other interactions.

[1] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, Nat. Mater. **12**, 191 (2013).

[2] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Phys. Rev. Lett. **114**, 105503 (2015).

[3] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[4] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[5] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, Comput. Phys. Commun. **184**, 1172 (2013).

[6] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Comput. Phys. Commun. **183**, 2063 (2012).

[7] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, Comput. Mater. Sci. **50**, 2295 (2011).

[8] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, Comput. Mater. Sci. **58**, 227 (2012).

[9] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, JOM **65**, 1501 (2013).

[10] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, Comput. Mater. Sci. **111**, 218 (2016).

[11] See https://www.nomad-coe.eu.

[12] U. Muller, *International Tables for Crystallography* (International Union of Crystallography, 2011), Vol. A1, Chap. 1.6, pp. 44–56.

[13] H. D. Megaw, *Crystal Structures—A Working Approach* (Saunders, Philadelphia, 1973).

[14] H. Barnighausen, Commun. Math. Chem. **9**, 139 (1980).

[15] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[16] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012).

[17] J. I. Mustafa, S. Coh, M. L. Cohen, and S. G. Louie, Phys. Rev. B **92**, 165134 (2015).

[18] We use generalized gradient approximation [39] to the density functional theory within the QUANTUM-ESPRESSO package [40–42]. Wannier functions were constructed using the WANNIER90 package [43] .

[19] G. H. Wannier, Phys. Rev. **52**, 191 (1937).

[20] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[21] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, Phys. Rev. Lett. **98**, 046402 (2007).

[22] Total energy E^{tot} in the Kohn-Sham framework does not depend only on $\int_{BZ} E_{nk} d\mathbf{k}$ but also on additional terms that are explicit functions of electron density $n(\mathbf{r})$. However, density can be

computed from the Wannier function as $n(\boldsymbol{r}) = \sum_{n\boldsymbol{R}} |W_n(\boldsymbol{r})|^2$. Therefore, in some cases it is possible that one would have to include a descriptor for the shape of the Wannier function in order to learn about the material structure. That is not the case here as we consider materials in their parent structure, so oxygen Wannier functions have nearly constant shape regardless of anion A or B.

[23] A. M. Glazer, Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. **28**, 3384 (1972).

[24] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **49**, 5828 (1994).

[25] P. M. Woodward, Acta Crystallogr., Sect B: Struct. Sci., Cryst. Eng. Mater. **53**, 44 (1997).

[26] N. A. Benedek and C. J. Fennie, J. Phys. Chem. C **117**, 13339 (2013).

[27] K. Persson, Materials Project Database, doi:10.17188/1203574.

[28] R. E. Cohen, Nature (London) **358**, 136 (1992).

[29] I. B. Bersuker, Ferroelectrics **164**, 75 (1995).

[30] M. H. Palmer, J. Mol. Struct. **405**, 193 (1997).

[31] A. Koleyski and K. Tkacz-miech, Ferroelectrics **314**, 123 (2005).

[32] W. H. Baur, Crystallogr. Rev. **13**, 65 (2007).

[33] F. Brisse, D. J. Stewart, V. Seidl, and O. Knop, Can. J. Chem. **50**, 3648 (1972).

[34] In the strict definition of structure family from Refs. [13,14], thortveitite (or anatase) are not hettotypes but only polymorphs of pyrochlore (or cristobalite) as distortion that relates them does not preserve polyhedral units.

[35] D. G. Pettifor, J. Phys. C **19**, 285 (1986).

[36] D. Pettifor, Solid State Commun. **51**, 31 (1984).

[37] C. S. Ong, S. Coh, M. L. Cohen, and S. G. Louie, Phys. Rev. Mater. **1**, 075408 (2017).

[38] F. Giustino, J. R. Yates, I. Souza, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **98**, 047005 (2007).

[39] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[40] P. Giannozzi *et al.*, J. Phys.: Condens. Matter **21**, 395502 (2009).

[41] D. R. Hamann, Phys. Rev. B **88**, 085117 (2013).

[42] K. F. Garrity, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, Comput. Mater. Sci. **81**, 446 (2014).

[43] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008).