Computational Materials Science 163 (2019) 108–116

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Finding the needle in the haystack: Materials discovery and design through computational *ab initio* high-throughput screening

Geoffroy Hautier

Institute of Condensed Matter and Nanoscience, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

A B S T R A C T

Essential materials properties can now be assessed through *ab initio* methods. When coupled with the exponential rise in computational power, this predictive power provides an opportunity for large-scale computational searches for new materials. We can now screen thousands of materials by their computed properties even before the experiments. This computational paradigm allows experimentalists to focus on the most promising candidates, and enable researchers to efficiently and rapidly explores new chemical spaces. In this paper, I present the challenges and opportunities in materials discovery using high-throughput *ab initio* computing. Focusing on my own research activities in the last five years with examples from the field of transparent conducting oxides, thermoelectrics and electrides, I illustrate how computational screening can be used to accelerate the discovery of new materials as well as to detect new relationships between chemistry, structures, and properties.

1. The quest for new materials

The discovery and development of new materials have enabled major technological breakthrough: airplanes, cell-phones, or transistors all rely on the control and understanding of materials. Finding the material with the adequate properties for a given application is a very challenging task. The specific property can sometimes only be present in very unique chemistries and structures. Moreover, not only one but several properties often need to align perfectly to lead to high performances. For instance, a cathode material would require adequate voltage, stability and ionic diffusion and a solar absorber would need not only the right band gap but also adequate carrier transport and defects. Unwanted and inherent correlations between materials properties (e.g., between voltage and stability in batteries) make the task at hand even more arduous. This is without even mentioning the important economical constraints (e.g., cost of raw materials and process). The development of new materials has been for most of its long history very empirical using past experience, simple models, trial-and-error and a good dose of serendipity and intuition. This is a very long and wasteful process with one of the most extreme examples being the search for the best bulb lamp filament by Thomas Edison. Edison's team had to try more than 3000 filaments and evaluate the bulb lamp performances for each of them.

The advent of computational techniques that are affordable and accurate enough to be predictive has offered a new avenue to accelerate the materials discovery process. In this review, I will present how high-throughput (HT) computational screening combined with data mining has the potential to revolutionize the materials discovery process by accelerating it tremendously. In the spirit of this special "Rising Stars" issue, I will present my views on the opportunities and challenges in the area and illustrate those with examples from my own research activities in the last 5 years. I will not exhaustively review the emerging field of HT computing. By no means this paper implies that these are the only activities in the very dynamic field of HT computing. We refer the interested reader to the more exhaustive reviews written on the topic (e.g., [1] or [2]).

2. Materials modeling and *ab initio* theory

A full understanding of the inherent atomistic behavior of materials was only achievable with the advent of quantum mechanics. The Schrödinger equation was discovered in 1926. The potential to use quantum mechanics to compute the properties for molecules and solids was very quickly identified with the first computation of the H_2 molecule by Heitler and London performed in 1927. Enthusiastic claims of having all the tools at hand to solve most of the problems in molecular and solid chemistry were quickly made but with the important caveat that the task at end would be extremely computationally expensive. It was only with the emergence of generally programmable computers that the computational study of the properties of matter through so-called first principles or *ab initio* methods became possible beyond simple molecules. However, realistic materials were still out of reach and it is a combination of two developments that led to a revolution in atomistic modeling of materials. The first development is the microelectronic industry (itself relying on important materials developments) offering an exponential growth in computational power following Moore's law. However, the first *ab initio* methods scaled so poorly with the number of atoms that even Moore's law was not enough. The poor scaling is inherent to the quantum theory and more specifically to the complex multi-particle wavefunction describing any solid or molecule. A conceptual leap forward was needed. These developments came from fundamental work from Kohn, Hohenberg and Sham who put together a

https://doi.org/10.1016/j.commatsci.2019.02.040
Received 30 October 2018; Received in revised form 22 February 2019; Accepted 25 February 2019

series of theorems and approximations enabling the development of a quantum theory of solids and molecules that could precisely bypass the description by the complex wavefunction and work with the much simpler charge density. These developments led to Density Functional Theory (DFT), which is currently the most, used *ab initio* technique. While based on an exact theoretical framework, any practical DFT computation relies on approximations. The power of DFT lies in its ability to develop approximations schemes that are very powerful, leading to an excellent compromise between accuracy and computational time. DFT has dramatically impacted materials science, condensed matter physics and chemistry as recognized by the Nobel Prize in Chemistry in 1998. While the early works using DFT and other *ab initio* techniques focused mainly on the understanding and rationalization of experimental observations, the last decades have seen more and more examples of true predictions made by DFT that were confirmed afterwards experimentally [3,4]. This growing predictive power has enabled computational materials science to reach a new status in the materials discovery and design process: from a tool dedicated to understanding, its predictive potential is used more and more.

3. High-throughput computational screening

The rising predictive power of *ab initio* techniques combined with the access to unprecedented computational power open the avenue to a very simple yet powerful idea: why do we not simply compute properties for hundreds to thousands of materials using DFT and search among these materials to find the best candidates? This screening process is illustrated in Fig. 1. Starting with a database of known or yet-to-be-made materials (i.e., a composition and a crystal structure), essential materials properties for the targeted application are automatically computed. As materials selection problems are often multi-property optimizations, the screening process relies on a series of steps each of them focusing on one property. The screening is tiered with the first properties to be computed being computationally inexpensive while more and more advanced and expensive properties are computed as we go down the funnel for which less and less materials have to be considered. We note that the same property can be present at different stages with different levels of accuracy. For instance, a DFT band gap could be used in the first steps while the next steps include more accurate but more expensive GW or hybrid functional band gap computations.

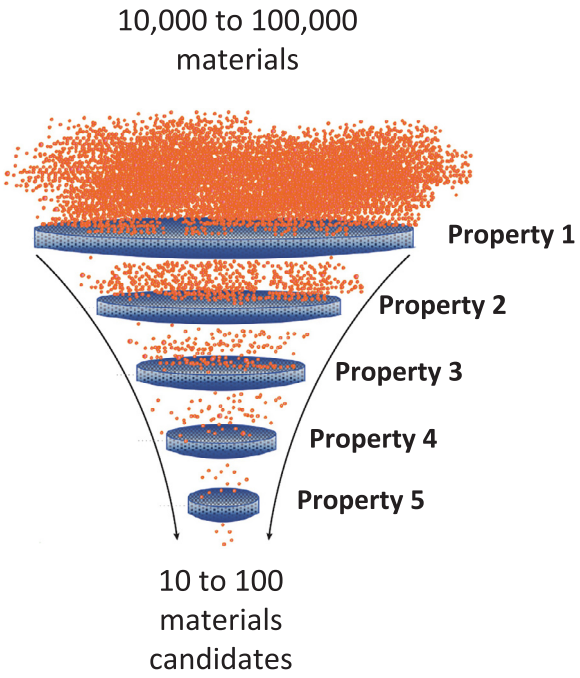

Fig. 1. High-throughput tiered screening. A series of properties are computed in a funnel screening. The easier properties are computed first followed by more computationally expensive ones. Adapted from [5].

The concept of combinatorial screening was originally applied for experiments and a whole field of combinatorial or HT experimental research exists. While there are common challenges and opportunities between experimental and computational HT research, the advantages of computations versus experiments are numerous. Certain experimental measurements are extremely complex and very difficult to automatize. For instance, measuring a cathode voltage requires building an electrochemical cell that can be contacted and tested. This is extremely challenging to perform combinatorially. Computations also offer more control as the measurement of an experimental property can be difficult to deconvolute from a combination of fundamental processes. Finally, much more limits in terms of chemistry and processing conditions are present experimentally. Introducing new elements in a sputtering chamber can be problematic because of contamination. The first computational HT studies were essential proof of concepts. These pioneer projects targeted for instance alloy stability [6], thermoelectrics [7], and electrocatalysis [8]. Starting in 2006, while working in Professor Ceder's group at MIT, I developed with a series of colleagues and collaborators the first HT search for new cathode materials. This work has led to the computational discovery followed by experimental synthesis of several new cathode materials [9–14]. Next to these findings, the very large data set generated with information on more than 30,000 potential cathode candidates led to a new understanding of the limits and opportunities in lithium ion battery cathodes.

4. The challenges and opportunities of high-throughput computing

While the idea of HT computing is quite simple, its practical implementation brings important challenges. I will group these challenges in three general categories: relevance, infrastructure and accuracy. The first issue of relevance relates to the limits of *ab initio* computing. Even with DFT, standard *ab initio* computations are limited to systems of at most a few hundred atoms. Many materials properties depend dramatically on processes driven at higher scales (e.g., plasticity). For these properties, *ab initio* computations are typically difficult to use in a predictive manner although multi-scale approaches or indirect descriptors (e.g., stacking fault energy for plasticity) offer interesting opportunities. In other cases, the property has some dependence on microstructure but is still strongly driven by atomistic level mechanisms. For instance, ionic conduction in a polycrystalline solid-state electrolyte material will depend on the bulk properties but could be also affected by the nature and amount of grain boundaries. Here, the situation is subtler. Bulk properties are of importance but do not tell everything about the final materials behavior. Often the atomistic level will lead to necessary but not sufficient properties. If one looks for a good solid-state electrolyte, atomistic level bulk conductivity is needed but does not guarantee that a sample with poor grain boundaries would not lead to low performances. All these considerations are very field and property dependent and the take home message is that HT computing needs to be used on the adequate problem. Moreover, a good HT study requires knowing what is important to compute, what we can compute with a certain accuracy and computational cost and put everything together in an efficient screening approach. One very illustrative example of the need for a good understanding of what is driving the material property is the search for high temperature superconductor (i.e., not following BCS theory). This would not be possible nowadays with HT screening because the fundamental theory beyond this type of superconductivity is not understood and simply put, no-one understands exactly what to compute.

HT computing also leads to challenges in software infrastructure. Widely distributed, robust and documented *ab initio* codes are available

both commercially and open source. However, the task of automatizing procedures that are typically performed one material at a time is not trivial. As computing one material property typically requires a series of inter-linked computations, software that can handle complex workflows of computations necessary. These software workflow toolsoffer for instance clear procedures for restarting computations with alternative parameters (e.g., in case of non-convergence), for spotting problematic results and to keep track of failed jobs. These needs have motivated the development of open sources packages often python-based (e.g., pymatgen [15], Fireworks [16]) that interact with *ab initio* codes to make them automatic and runnable in a HT fashion. We note that every new material property requires a non-trivial effort in developing the algorithms and the software around it. While simple ionic relaxation giving access to energetics were among the first properties computed, the field is moving to more and more complex workflows with recent efforts towards phonons [17,18], defects [19] or advanced electronic structure such as GW [20].

The final challenge to be mentioned is the one related to the accuracy of the methods, which directly affects the predictive power. As we already mentioned any practical implementation of DFT relies on an approximate treatment of the electronic interactions. There are currently many flavors of DFT with different level of complexity. In addition to these effects of the approximate functional, there are always other assumptions, for instance the neglect of temperature. The properties computed are thus always deviating from the exact experimental ground truth. While HT researchers always need to be aware of these limitations, the inaccuracy of the methods is not in itself preventing their use. Not being entirely accurate does not preclude making predictions as long as there is some knowledge of the uncertainties involved. Unfortunately, it is often difficult to know a priori how well an approximation would work for a given property. Empirical knowledge has been built and experts can have a good intuition about what methods would work for what property. However, the most rational way to assess the predictive power of an *ab initio* technique is to benchmark it versus experiments or extremely accurate computations. We note that the field of molecular *ab initio* theory has been following this path for longer than for solids and benchmark studies of certain functionals or methods on standard sets of molecules are common in this field. I am convinced these types of benchmarks are going to be more and more necessary on solids as HT computing and predictive *ab initio* modeling is growing. Recent efforts towards the benchmarking of effective mass, band gaps and vibrational entropy are illustrated in Fig. 2[17,20,21]. We note that one challenge here is the lack of reliable and easily accessible experimental data. Additionally, the field of solids do not have clear “gold-standard” *ab initio* methods which would be as accurate as experimental techniques. The important work towards the implementation of highly accurate techniques from the molecular field to periodic solids is moving fast but is still in its infancy [22].

5. High-throughput in action: searching for new transparent conductors

After this general overview, we will now delve into a few specific examples of HT computational searches starting with the field of transparent conducting materials (TCMs). High performance TCMs are necessary to many applications including modern electronic devices (e.g., the touch screen on a smart-phone), thin film solar cells and energy efficient windows [25]. Combining transparency and conductivity is difficult as they are antagonistic properties: good conductors (e.g., metals) are usually not transparent and transparent materials (e.g., glass) are not conductive. TCMs have been developed following a very simple idea: using a large band gap (> 3eV for transparency in the visible) material that is sufficiently doped either p- or n-type to bring high conductivity. TCMs have been developed for long now with materials such as n-type In_2O_3 doped with tin (or ITO) being widely produced and commercialized. However, their p-type counterparts are

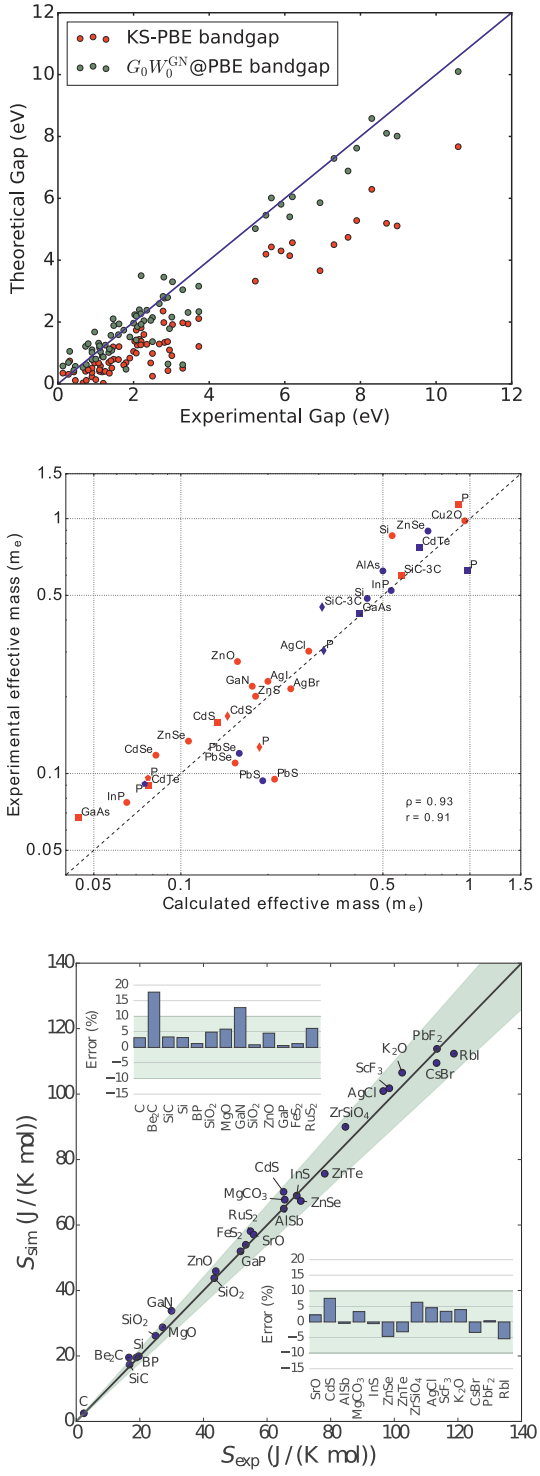

Fig. 2. Theory versus experiments for band gaps (G_0W_0 and DFT-GGA-PBE), DFT effective masses and DFT vibrational entropy at 300 K from phonons. Figures extracted from [20,23,24].

lagging behind and this prevents the development of new applications such as transparent electronics. One of the key metrics illustrating the gap between n- and p-type TCMs is carrier mobility. While the best n-type materials exhibit mobilities around 100 cm^2/Vs, the very best p-type materials do not exceed 5–10 cm^2/Vs. The search for high performance p-type TCMs is a major challenge in materials science and a very interesting field of application for HT computing [26]. We published the first HT search for TCMs focusing at p-type transparent

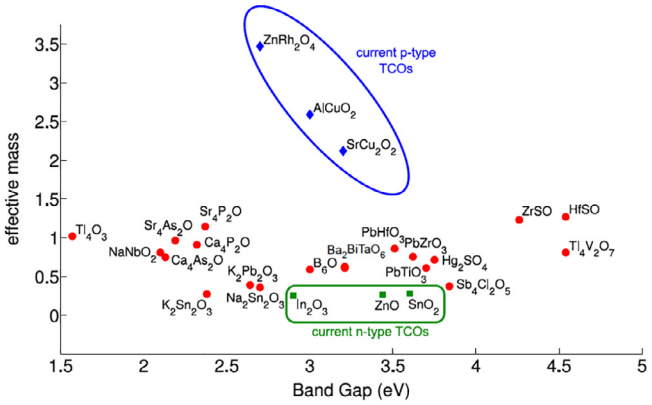

Fig. 3. Effective mass versus band gap for the low hole effective mass candidates from the ICSD (red dots). The current p-type TCOs are in blue and the electron effective mass of current n-type TCOs are shown in green. Figure adapted from [27].

conducting oxides (TCOs) [27]. The tiered screening was built identifying three important properties: mobility directly evaluated by effective mass, band gap and dopability (through defect computations). Good p-type TCOs should exhibit a low hole effective mass, a high band gap (> 3 eV ideally) and defects compatible with p-type doping. The first screening step on effective mass was by far the most restrictive. Among the 6,000 known oxides considered and coming from the Inorganic Crystal Structure Database (ICSD), only 20 showed hole effective masses 2 times lower than previously known p-type TCOs such as $CuAlO_2$. This example illustrates very well the title of this paper as it shows the needle in the haystack problem often faced by materials scientists. Among these 20 compounds, a more advanced method namely GW was used to evaluate more accurate band gaps. The band gap versus hole effective mass is plotted in Fig. 3 for all newly identified low hole effective mass materials (red dots) and for current p-type TCOs (in blue). For comparison, the electron effective mass of a series of high-performance n-type TCOs (SnO_2, ZnO, In_2O_3) are plotted as well.

The first observation is that hole effective masses can reach values close to the electron effective masses of the best n-type TCOs at least for some exceptional materials. There is therefore hope for p-type TCOs to reach mobilities similar to their n-type counterpart. The list of candidates materials can be further screened out by element abundance and toxicity, by band gap (ideally the band gap should be higher than 3 eV) and finally by defect computations. We performed defect computations on the most promising handful materials to assess their dopability (i.e., the ability for a material to be p- or n-type doped). Simple intrinsic point defects (e.g., vacancies) often indicate if a certain doping type will be favored in the material. For instance, hole-killer defects such as the oxygen vacancy will provide electron when the Fermi level is pushed closer to the valence band minimum through extrinsic p-type doping. This is of course a detrimental process for p-type material and materials with favorable hole-killer defects are not good candidates for p-type TCOs. Among the most interesting materials combining high band gap, low effective masses and favorable dopability are $PbMO_3$ (M = Ti, Hf, Zr), B_6O and Ba_2BiTaO_6. Fig. 4 plots the DFT band structure with corrected GW band gap for all these materials. The very dispersive valence band leading to a low hole effective mass is striking in all of them.

$K_2Sn_2O_3$ is likely to be more absorbing because of its smaller band gap but offering the lowest hole effective mass in our study. Experimental indications of the p-type nature of $PbTiO_3$ or boron suboxide (B_6O) were available but with unclear results on the hole mobility. Our further defect computations confirmed the p-type dopability of B_6O and identify C–H on O complex defects as potential shallow acceptors [29].

Ba_2BiTaO_6 on the other hand did not present any experimental optical and electrical data but offered many attractive computed features: high band gap, low effective mass and potential for p-type dopability. This motivated its follow-up experimental study [28]. The experiments confirmed the transparency to the visible and the possibility to p-type dope the material (through potassium) although to low carriers concentrations only (10^{15} cm^{-3}). The measured mobility was impressive with a value of 38 cm^2/Vs to compare with the previously known p-type oxide showing mobilities around 5–10 cm^2/Vs. The exceptional behavior of Ba_2BiTaO_6 could be after the fact rationalized by identifying the ingredients responsible for its high band gap and low hole effective mass. The presence of Bi^{3+} allowed the mixing of 6*s* states with O-2*p* in the valence band while the Ta nature of the conduction band guaranteed a large band gap. This mixing of metallic *s* states with O-2*p* leads naturally to a delocalization of the wavefunction and to a dispersive valence band (i.e., a lower hole effective mass). Despite its attractive features, no one in the TCO field would have identified this material *a priori* as a high band gap, low hole effective mass compound. In fact, previous attempts to make Bi^{3+}-based p-type TCOs (e.g., in BiCuOS[30]) were unsuccessful as the crystal structure did not allow the right mixing of the Bi-6*s* and O-2*p*. The example of Ba_2BiTaO_6 is an interesting demonstration of the predictive power of HT *ab initio* screening and shows very well how the approach can propose new and unexpected avenues. The main limitation of Ba_2BiTaO_6 is still its low carrier concentrations and further work will be needed to increase the carriers in this material offering complex defect chemistry. Nevertheless, the demonstration of a Bi^{3+} based p-type transparent oxide of high hole mobility opens interesting new opportunities in p-type TCOs.

Next to the identification of potential candidates, HT computing provides value in understanding trends and limits for certain chemistries. The data generated by our p-type TCOs study directly points to two strategies to form high performance TCOs: using reduced main group elements (i.e., Sn^{2+}, Pb^{2+}, Bi^{3+},....) or oxyanions (e.g., oxysulfides) [27]. It also quantitatively indicates that the copper-based chemistry, which had been the main focus of the p-type TCO community, would have a very hard time competing with those alternative chemistries. This analysis generates further questions and new avenues of research. The need for certain ions such as reduced main group elements could be directly inferred from our computational database but the presence of these ions do not necessarily lead to low hole effective mass. This points out to an important structural component. We performed a further study to determine what structural features would lead to low hole effective mass in the Sn^{2+} chemistry [31]. Effective masses are known to relate to the alignment of orbitals and this is why reduced main group elements lead to low hole effective masses. Indeed, the *s* states of these ions tend to align relatively well in energy with the O-2*p* orbitals. Additionally, the effective mass is also set by the overlap between orbitals: the larger the overlap the lower the effective mass. This hinted at the importance of the Sn-O-Sn angle. Indeed, the overlap between the Sn-5*s* of Sn^{2+} and the O-2*p* would be the largest when the angle is close to 180° naturally inducing low effective masses. Our large data set confirmed this trend and Fig. 5 shows the reasonable correlation between Sn-O-Sn angle and hole effective mass in all Sn^{2+}-compounds in the ICSD. This figure rationalizes why the structure of $K_2Sn_2O_3$ favoring Sn-O-Sn angles close to 180° outperforms the more traditional p-type TCO SnO in hole effective mass. Recent experimental work measuring mobilities on ternary Sn^{2+} TCOs showed an agreement with our computational analysis [32].

When they have been built, high-throughput infrastructures can be up-scaled relatively easily on a wider range of materials. From our original set of 6000 oxides, we have extended now the data set to non-oxides. Oxides have been the preferred chemistry for developing TCMs as they tend to be easy to process and stable. However, it remains an open question if other chemistries would not be able to offer better p-type TCMs. Large computational data sets can again be useful to answer this question [33]. Fig. 6a shows the distribution of hole effective mass within oxides, sulfides, nitrides and phosphides in our computational

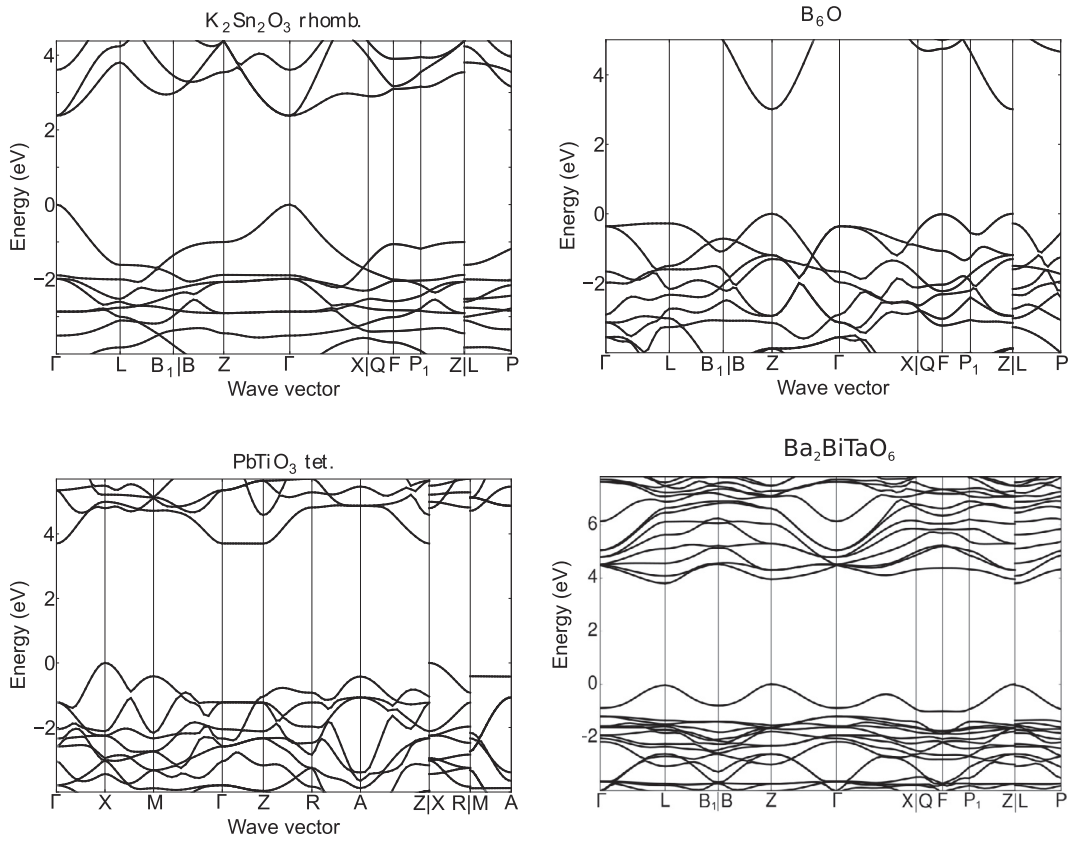

Fig. 4. Band structure of materials candidates ($K_2Sn_2O_3$, B_6O, $PbTiO_3$ and Ba_2BiTaO_6) identified by HT screening. The band structures have been obtained by DFT-GGA-PBE with a scissor shift for the band gap according to G_0W_0 computations. Figures adapted from [27,28].

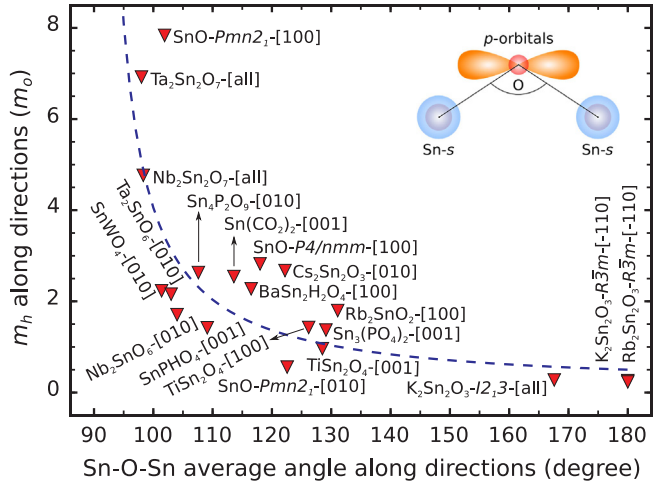

Fig. 5. Dependence of the hole effective mass on the Sn-O-Sn angle in Sn^{2+} oxides from the ICSD. Computations are performed in DFT-GGA-PBE. Figure adapted from [31].

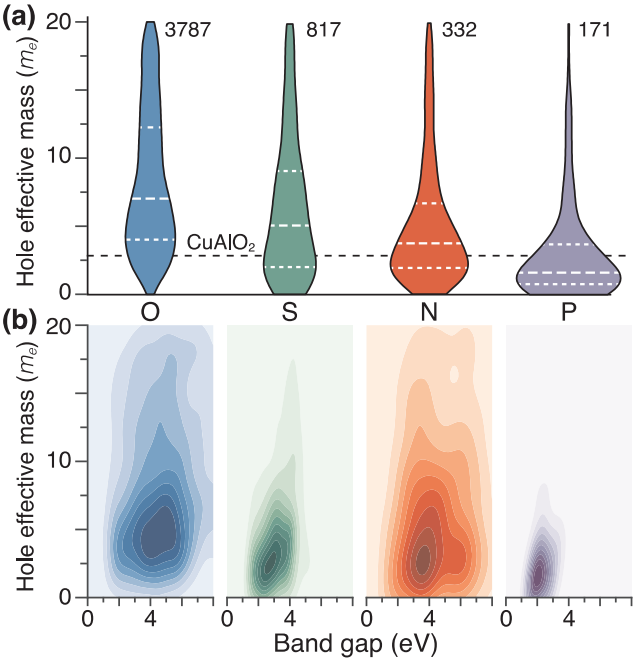

Fig. 6. Distribution of hole effective mass (a) and hole effective mass vs band gap (b) in different chemistries. Hole effective masses are obtained from DFT-GGA-PBE and band gaps are computed from DFT with an empirical shift. Figures from [33].

database.

Oxides present the highest hole effective mass and as we move to other anions the more delocalized and higher energy of the anion *p* orbitals lead to more dispersive valence bands. Among the four chemistries we compared, oxides are by far the most challenging to find low hole effective mass materials. However, when effective mass is plotted against band gap (Fig. 6b), we observe that the lower hole effective mass of non-oxides chemistry strongly correlates with lower band gaps and thus lower transparency. As in many materials science problem, the intrinsic correlations between properties (here band gap and effective mass) makes the material selection problem very difficult to solve. Here, there is fortunately a work-around when keeping in mind that band gaps can be direct or indirect. Indirect band gaps will

absorb much less than direct band gaps and, when used as thin-films with a few 100 nm thickness, materials with small indirect gap materials can remain transparent. Using this simple principle, a slightly more complex screening can be performed focusing on materials with large *direct* band gaps. We used this screening approach on a data set of non-oxides, especially phosphides, and identified boron phosphide (BP) as a p-type TCM of great interest. BP shows all the characteristics of a high efficiency p-type TCM if used in a thin film of a few hundred nanometers [33]. This simple III-V material had been previously studied experimentally in the 60 s and 70 s pointing at experimental data in agreement with our computational predictions. Our work motivates the reinvestigation of BP for p-type TCM applications. HT computing is here a way to reconsider a material that is not actively studied anymore and to pinpoint what specific application it could be good for.

The very large data set generated is still widely used to search for new TCMs. The whole effective mass data set has been published and is available on the Materials Project [21,34]. We also used this data to study n-type TCOs establishing that current n-type TCOs are difficult to outperform in terms of effective masses and identifying a few interesting candidates [35]. We note that, since its publication, our work demonstrated the great potential of the approach and has motivated a series of other HT screening studies focusing on TCMs [36–39].

6. Recycling data sets for new applications: thermoelectrics

HT infrastructure can easily be extended to additional chemistries and materials as we demonstrated in extending our TCO search to non-oxides. Additionally, the properties computed for one application can be of interest to other fields. While our TCM work focused on effective masses, many other "by-products" such as Seebeck coefficient could be extracted from our analysis of electronic transport. We have been collaborating with a series of theory and experimental groups to harness this electronic transport data set to search for thermoelectric materials. Thermoelectrics devices can transform difference in temperature in difference in electrical potential they are therefore considered in waste-energy recovery or cooling applications [40]. The efficiency of the device built on a certain material is linked to the figure of merit $ZT = \frac{S^2\sigma}{\kappa}T$ which directly links different materials properties: S is the Seebeck coefficient, σ the electrical conductivity and K the thermal conductivity. The infrastructure built for the TCM project offers a way to assess the numerator of the ZT computing the power factor $S^2\sigma$. Other properties such as lattice thermal conductivity will need to be assessed through other computations or directly experimentally. One of the difficulties facing the design of high ZT thermoelectrics is the correlation of Seebeck and conductivity. Materials with high Seebeck tend to be not very conductive and vice-versa. Indeed, large density of states and flat bands increase the Seebeck while they limit conductivity. Fig. 7 shows the Seebeck and conductivity extracted from our large database of transport properties. The color of the data points indicates the power factor. Only a small fraction of the materials leads to the high Seebeck combined with a high conductivity required to maximize the power factor.

Without going into details here, we note that all this work is performed in the so-called constant relaxation time approximation where the scattering processes for the electrons are considered independent of temperature and of energy. While this approximation is known to especially underestimate resistivity and to lead to an inaccurate quantitative agreement with experiment, it is remarkable that power factors under the constant relaxation time approximation can already identify the best current thermoelectrics. We have indicated in Fig. 7 the data points related to a few known typical thermoelectrics. They all sit in a region offering a good compromise between Seebeck and conductivity and maximizing the power factor. So, the computed data can be used at the very least to search for materials with potentially high power factors. Using the database, we identified a few new thermoelectrics

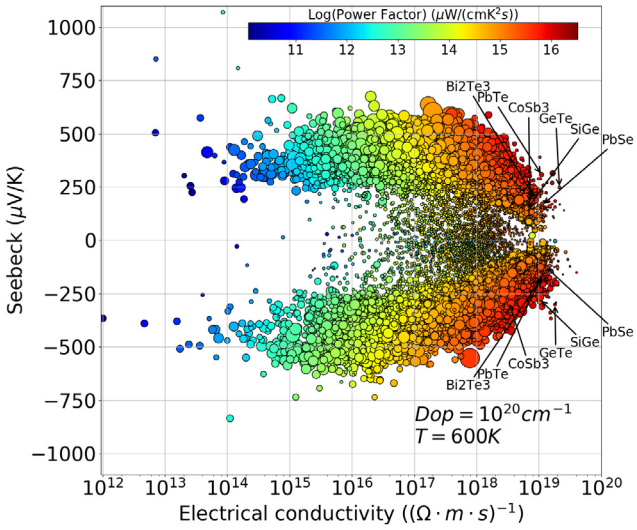

Fig. 7. Seebeck versus conductivity in a database of 40,000 materials. The transport properties are computed on DFT-GGA-PBE band structures using the constant relaxation time approximation and a doping of 10^{20} cm^{-3} and a temperature of 600 K. The color indicates the power factor ($S^2\sigma$) and the size of the dot the band gap. The computed values are indicated for common thermoelectrics. Adapted from [23].

among which $TmAgTe_2$ and $YCuTe_2$ were subsequently synthesized and characterized. Their experimental test showed zT values from 0.4 to 1.0 [41,42]. These are attractive numbers for non-optimized materials, keeping in mind the challenge of finding good thermoelectrics that the very best thermoelectrics show ZTs between 1 and 2 Again, the exciting perspective of truly identifying materials *in silico* that are synthesized afterwards and demonstrate attractive properties is still uncommon and a very exciting development in our field.

The identified materials are tellurides which are very common chemistries for thermoelectrics (e.g., Bi_2Te_3 and PbTe) but new chemical playgrounds can be identified and motivated by HT computing as well. We recently demonstrated through HT computations and experiments that phosphides show attractive electronic properties and could lead to lower thermal conductivity than one would expect [43]. Fundamental limits to certain chemistries can as well be discovered: oxides, for instance, clearly show power factor values much lower than tellurides or other chemistries [44]. This indicates that the band structures of oxides intrinsically lead to lower power factors. This provides a data-driven justification for the poor ZTs of oxides thermoelectrics (among other factors such as poor thermal conductivity).

7. Challenging established rules: electrides

High-throughput computing is especially of great interest in fields where common sense and chemical intuition is difficult to build because of the lack of data. This is for instance the case in the field of electrides. Electrides are fascinating materials presenting electrons in off-nuclei positions. While most materials have their electron density localized around the nuclei, electrides present electrons localized in empty "pockets" in the crystal structure. The first inorganic electrides $Ca_{12}Al_{14}O_{32}$ was discovered by Hosono in 2003 [45]. Since then, only a handful electrides have been reported in the literature. Using the power of high-throughput computing, we performed the first large scale search for electrides in the ICSD [46]. The idea is again simple yet powerful. By looking at the electron localization of the DFT wavefunctions, we can detect if their localization is mainly off-nuclei. Screening among 30,000 metallic compounds, our approach identified more than 60 new electrides (recovering the handful known ones). Our work has uncovered structural as well as chemical factors (i.e., the presence of certain

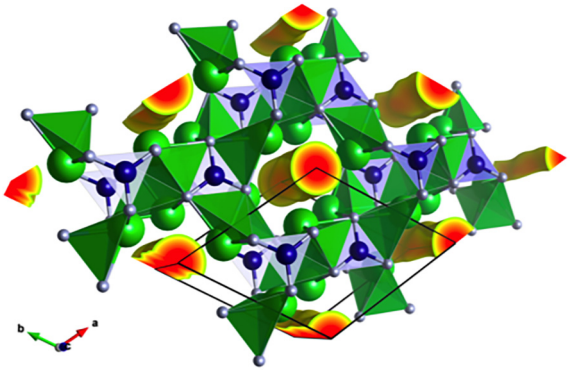

Fig. 8. Electron density near the Fermi level for Ba_3CrN_3 a newly identified electride. Results are from DFT-GGA-PBE computations. Figure from [46].

elements such as alkali-earths) leading to electride formation. It also has led to an unexpected finding as we identified the first electrides containing a redox active transition metal: Ba_3CrN_3 and Sr_3CrN_3. Fig. 8 shows the electron distribution around the Fermi level for these 2 new electrides. The electron is present in a tubular shape cavity showing the signs of a one-dimensional electride. This finding defies chemical intuition as it was long expected that redox active elements here Cr would energetically favor a change in oxidation state rather than the localization of the electron in a structural pocket. This example outlines once more the potential of HT searches to question established rules and find new and unexpected outliers.

8. Towards more complex properties

The first high-throughput studies mentioned in this review paper on TCMs used relatively simple computations such as standard DFT ionic relaxation or band structure computations. All the more complex follow-up computations such as GW band gaps or defect computations were performed in a low-throughput manner with a good amount of human intervention. This approach worked very efficiently as the first screening criteria on effective mass was so strong that not many of those low-throughput computations had to be performed. However, in other fields such as thermoelectrics where more candidates appear to pass criteria based on DFT band structures, the later steps will have to also be automatized. Defects are for instance often limiting thermoelectrics performance and we are currently leading efforts towards performing HT defect computations with first automation steps provided in our pycdt python code [19].

Another example of new property brought to automation is our recent work on HT computations of phonons and vibrational properties. We will use this example to illustrate the different steps necessary to achieve the computations of a property in a HT framework. First of all, one needs to develop a “recipe” to compute the property with high accuracy in an automatized way. This procedure includes the different steps needed to perform the computations, how they connect to each other and if certain error checks are necessary. It also requires setting a series of typical parameters for which the HT computations will be converged to a given accuracy. This procedure is based on expert knowledge but also on systematic benchmarks. We performed such a study for phonons and evaluated the required k-point and q-point grids for high quality phonon band structures [18]. When this is established, all this knowledge needs to be coded within a workflow framework that will be used to run the computations in production [17]. Here, for the phonons we used Fireworks interfaced with pymatgen and abipy using the ABINIT *ab initio* code [15,16,47]. This phonon data set will be of great use to many applications including thermoelectrics (where thermal conductivity is important) but also superconductivity or ferroelectricity.

We note that automation does not benefit only to HT studies. For instance, while G_0W_0 computations are currently not run on a very large scale (because of their computational cost), the development of an automatized infrastructure to perform those complex computations is of great interest for obtaining reliable converged values [20].

9. Sharing large data sets and the materials project

High-throughput studies generate by definition large sets of data. These are often produced for a specific purpose but of interest to many fields. For instance, the transport data we generated for our TCM work is of great interest for thermoelectrics as we already mentioned but also for photovoltaics. Sharing efficiently those data sets is therefore of great importance. Several initiatives have been setting freely accessible databases through web interfaces: AFLOW, Materials Project, OQMD or NOMAD [34,48,49]. This opens the possibility for users to browse among properties for thousands of compounds often not only through web interfaces but also directly querying the database using common programming language such as python. The Materials Project, which is the initiative we have been contributing to, offers a large series of materials properties: energies, elastic constants, band structures, phonons etc… for a wide range of materials [34]. The impact of the Materials Project is major with 60,000 registered users and more than 1,400 sessions/day. Large computational databases such as the Materials Project have been changing the way materials science is performed and driving the field to be more data-driven. While computations have their limits and need to be used with care, it is nevertheless a very exciting moment for materials science as no such amount of data as ever been accessible to researchers.

10. Data mining, trends and machine learning on large computational databases

Using data sets to observe trends and build predictive models has always been an important part of materials science. One could argue that the periodic table is one of the first examples of data mining but structure maps which are used to predict what structure would form in binaries depending on the element attributes (ionic radius, Mendeleev number, etc…) are another very nice example of early data mining. Coupled with the growing data sets easily accessible, it is expected that more and more trends will be extracted from computational data. Materials scientists have been also benefiting from the very powerful methods that are built in the field of machine learning [50–53]. New and unexpected structure-property relationships, which are central to materials science, will likely emerge from these studies. However, this approach requires efficient structural descriptors. Indeed, while crystal structures can be described as unit cell parameters and site positions, this is unlikely to be informative enough to build structure-property relationships. This is even more the case when the same crystal structure can be defined in many different settings.

I strongly believe that, when building new tools, we should not forget the experience of the decades of research in materials science, solid-state physics and chemistry. If one keeps this in mind, it is natural to think that local or coordination environments are very important crystal structure descriptors. Solid-state chemists would often describe their structure by the different local environments of the ions (e.g., Ti^{4+} in rutile TiO_2 is octahedral coordinated). Local environments are often correlated with certain properties and are natural structural descriptors. Despite their importance for the description of crystalline solids, local environment assignment is however most of the time performed “manually” looking at crystal structures in visualization software. This is of course not adequate to HT computing and motivated us to develop an algorithm and a code (ChemEnv) to perform automatic local environment detection. The principle of the algorithm is quite simple. We consider a series of perfect local environments (52 in total) and the algorithm computes a “similarity measure” of a true potentially distorted environment to each of the perfect local environments.

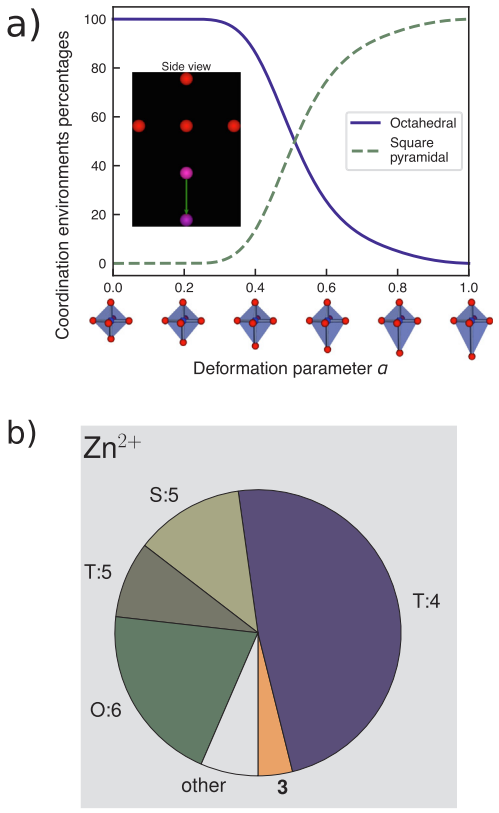

Fig. 9. Local/coordination environment assignment using ChemEnv. (a) Percentage of coordination environment for a distorted octhedra transforming from octahedral to square pyramidal. (b) Pie chart of the distribution of local environment of Zn^{2+} in oxides. T:4 is tetrahedral, S:5 square pyramidal, T:5 triangular bipyramidal, O:6 is octahedral, 3 is three-fold coordinated and other regroups all other environments. Figures from [55].

Comparing these distance results allows to assign the most likely coordination environment. One of the important features of ChemEnv is that the algorithm is robust to distortions that occur naturally in real crystals. Fig. 9a illustrates this feature: as an octahedron is distorted in one direction the metric will smoothly go from a pure octahedron assignment to a mixed octahedral and square pyramidal and finally to a pure square pyramidal environment.

We used ChemEnv to perform the first statistical search for local environments in oxides. Experienced materials scientists can often tell what local environments certain cations would prefer. However, no statistically robust study had been performed to identify the distribution of coordination environments in oxides. We published recently an exhaustive study indicating the local environment distribution for each cation in oxides. For instance, Fig. 9b shows the distribution of local environment for Zn^{2+} in oxides. While literature describes Zn^{2+} as a tetrahedral cation [54], our analysis outlines a more complex crystal chemistry with more than half of the local environments being different than tetrahedral. This is a first use of ChemEnv but further use towards structure-property relationships will come soon.

11. Conclusion

First principles approaches have strongly impacted materials science in the last few decades. Techniques such as DFT have offered an unprecedented way of probing how atomistic mechanisms influence materials properties. The impact of *ab initio* computations is still growing as the methods become more and more accurate and used on scales that had never been experienced through high-throughput computing. This review outlined the challenges but also the opportunities in HT computing highlighting examples from a series of applications and from my own research activities. The strength of HT computing lies especially in identifying materials combining exceptional properties among the vast landscape of structures and chemistries and helping in finding the needle in the haystack. These findings are often unexpected and question the admitted knowledge based on smaller data sets. As HT data sets are growing and made widely accessible the mining of this data through modern machine learning techniques will become more and more important and is expected to lead to new insight into structure-chemistry-property relationships. There is no doubt these developments will make materials science more and more data-driven and strongly impact the traditional way new materials are discovered and developed.

CRediT authorship contribution statement

Geoffroy Hautier: Writing - review & editing.

Acknowledgements

I would like to acknowledge the many students, postdoctoral researchers, collaborators and mentors whom I had a chance to work with in the last few years. I also acknowledge the financial support from funding agency and private partners especially the FNRS-F.R.S., the European Union, and BASF. The Materials Project is 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.

References

[1] S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, Nat. Mater. 12 (2013) 191–201, https://doi.org/10.1038/nmat3568.
[2] K. Alberi, M.B. Nardelli, A. Zakutayev, L. Mitas, S. Curtarolo, A. Jain, M. Fornari, N. Marzari, I. Takeuchi, M.L. Green, M. Kanatzidis, M.F. Toney, S. Butenko, B. Meredig, S. Lany, U. Kattner, A. Davydov, E.S. Toberer, V. Stevanovic, A. Walsh, N.-G. Park, A. Aspuru-Guzik, D.P. Tabor, J. Nelson, J. Murphy, A. Setlur, J. Gregoire, H. Li, R. Xiao, A. Ludwig, L.W. Martin, A.M. Rappe, S.-H. Wei, J. Perkins, The 2019 materials by design roadmap, J. Phys. D. Appl. Phys. 52 (2019) 013001, , https://doi.org/10.1088/1361-6463/aad926.
[3] G. Hautier, A. Jain, S.P. Ong, From the computer to the laboratory: materials discovery and design using first-principles calculations, J. Mater. Sci. 47 (2012) 7317–7340, https://doi.org/10.1007/s10853-012-6424-0.
[4] A. Jain, Y. Shin, K.A. Persson, Computational predictions of energy materials using density functional theory, Nat. Rev. Mater. 1 (2016), https://doi.org/10.1038/natrevmats.2015.4.
[5] A. Aspuru-Guzik, J. Aguilera-Iparraguirre, C. Suh, E.O. Pyzer-Knapp, R. Gómez-Bombarelli, What is high-throughput virtual screening? A perspective from organic materials discovery, Annu. Rev. Mater. Res. 45 (2015) 195–216, https://doi.org/10.1146/annurev-matsci-070214-020823.
[6] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, G. Ceder, Predicting crystal structures with data mining of quantum calculations, Phys. Rev. Lett. 91 (2003) 135503, https://doi.org/10.1103/PhysRevLett. 91.135503.
[7] G.K.H. Madsen, Automated search for new thermoelectric materials: the case of LiZnSb, J. Am. Chem. Soc. 128 (2006) 12140–12146, https://doi.org/10.1021/ja062526a.
[8] J. Greeley, T.F. Jaramillo, J. Bonde, J.K. Nørskov, I.B. Chorkendorff, Computational high-throughput screening of electrocatalytic materials for hydrogen evolution, Nat. Mater. 5 (2006) 909–913, https://doi.org/10.1038/nmat1752.
[9] G. Hautier, A. Jain, H. Chen, C. Moore, S.P. Ong, G. Ceder, Novel mixed polyanions lithium-ion battery cathode materials predicted by high-throughput *ab initio* computations, J. Mater. Chem. 21 (2011) 17147–17153, https://doi.org/10.1039/c1jm12216a.
[10] A. Jain, G. Hautier, C.J. Moore, B. Kang, J. Lee, H. Chen, N. Twu, G. Ceder, A computational investigation of $Li_9M_3(P_2O_7)_3(PO_4)_2$ (M = V, Mo) as cathodes for Li ion batteries, J. Electrochem. Soc. 159 (2012) A622–A633.
[11] X. Ma, G. Hautier, A. Jain, R. Doe, G. Ceder, Improved capacity retention for LiVO2 by Cr substitution, J. Electrochem. Soc. 160 (2013) A279–A284, https://doi.org/10.1149/2.046302jes.
[12] H. Chen, G. Hautier, A. Jain, C. Moore, B. Kang, R. Doe, L. Wu, Y. Zhu, Y. Tang, G. Ceder, Carbonophosphates: a new family of cathode materials for Li-ion batteries identified computationally, Chem. Mater. 24 (2012), https://doi.org/10.1021/cm203243x.
[13] J.C. Kim, C.J. Moore, B. Kang, G. Hautier, A. Jain, G. Ceder, Synthesis and Electrochemical Properties of Monoclinic $LiMnBO_3$ as a Li Intercalation Material, J.

Electrochem. Soc. 158 (2011) A309, https://doi.org/10.1149/1.3536532.
[14] G. Hautier, A. Jain, S.P. Ong, B. Kang, C. Moore, R. Doe, G. Ceder, Phosphates as lithium-ion battery cathodes: an evaluation based on high-throughput *ab Initio* calculations, Chem. Mater. 23 (2011) 3945–4508, https://doi.org/10.1021/cm200949v.
[15] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis, Comput. Mater. Sci. 68 (2013) 314–319, https://doi.org/10.1016/j.commatsci.2012.10.028.
[16] A. Jain, S.P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, K.A. Persson, FireWorks: a dynamic workflow system designed for high-throughput applications, Concurr. Comput. Pract. Exp. 27 (2015) 5037–5059, https://doi.org/10.1002/cpe.3505.
[17] G. Petretto, S. Dwaraknath, H.P.C. Miranda, D. Winston, M. Giantomassi, M.J. van Setten, X. Gonze, K.A. Persson, G. Hautier, G.-M. Rignanese, High-throughput density-functional perturbation theory phonons for inorganic materials, Sci. Data 5 (2018) 180065, https://doi.org/10.1038/sdata.2018.65.
[18] G. Petretto, X. Gonze, G. Hautier, G.-M. Rignanese, Convergence and pitfalls of density functional perturbation theory phonons calculations from a high-throughput perspective, Comput. Mater. Sci. 144 (2018), https://doi.org/10.1016/j.commatsci.2017.12.040.
[19] D. Broberg, B. Medasani, N.E.R. Zimmermann, G. Yu, A. Canning, M. Haranczyk, M. Asta, G. Hautier, PyCDT: a Python toolkit for modeling point defects in semiconductors and insulators, Comput. Phys. Commun. (2018) 1–15, https://doi.org/10.1016/j.cpc.2018.01.004.
[20] M.J. van Setten, M. Giantomassi, X. Gonze, G.-M. Rignanese, G. Hautier, Automation methodologies and large-scale validation for GW: Towards high-throughput GW calculations, Phys. Rev. B. 96 (2017) 155207, https://doi.org/10.1103/PhysRevB.96.155207.
[21] F. Ricci, W. Chen, U. Aydemir, G.J. Snyder, G.-M. Rignanese, G.-M. Rignanese, A. Jain, G. Hautier, An ab initio electronic transport database for inorganic materials, Sci. Data 4 (2017) 170085, https://doi.org/10.1038/sdata.2017.85.
[22] G.H. Booth, A. Grüneis, G. Kresse, A. Alavi, Towards an exact description of electronic wavefunctions in real solids, Nature 493 (2012) 365–370, https://doi.org/10.1038/nature11770.
[23] F. Ricci, W. Chen, U. Aydemir, G.J. Snyder, G.-M. Rignanese, A. Jain, G. Hautier, An *ab initio* electronic transport database for inorganic materials, Sci. Data 4 (2017) 170085, https://doi.org/10.1038/sdata.2017.85.
[24] G. Petretto, X. Gonze, G. Hautier, G.-M. Rignanese, Convergence and pitfalls of density functional perturbation theory phonons calculations from a high-throughput perspective, Comput. Mater. Sci. 144 (2017) 331–337, https://doi.org/10.1016/j.commatsci.2017.12.040.
[25] K. Ellmer, Past achievements and future challenges in the development of optically transparent electrodes, Nat. Photon. 6 (2012) 809–817, https://doi.org/10.1038/nphoton.2012.282.
[26] S. Sheng, G. Fang, C. Li, S. Xu, X. Zhao, P-type transparent conducting oxides, Phys. Status Solidi. 203 (2006) 1891–1900, https://doi.org/10.1002/pssa.200521479.
[27] G. Hautier, A. Miglio, G. Ceder, G.-M. Rignanese, X. Gonze, Identification and design principles of low hole effective mass p-type transparent conducting oxides, Nat. Commun. 4 (2013) 2292, https://doi.org/10.1038/ncomms3292.
[28] A. Bhatia, G. Hautier, T. Nilgianskul, A. Miglio, J. Sun, H.J. Kim, K.H. Kim, S. Chen, G.-M. Rignanese, X. Gonze, J. Suntivich, High-mobility bismuth-based transparent p-type oxide from high-throughput material screening, Chem. Mater. 28 (2016) 30–34, https://doi.org/10.1021/acs.chemmater.5b03794.
[29] J.B. Varley, V. Lordi, A. Miglio, G. Hautier, Electronic structure and defect properties of B_6O from hybrid functional and many-body perturbation theory calculations: a possible ambipolar transparent conductor, Phys. Rev. B 90 (2014) 045205, https://doi.org/10.1103/PhysRevB.90.045205.
[30] H. Hiramatsu, H. Yanagi, T. Kamiya, K. Ueda, M. Hirano, H. Hosono, Crystal structures, optoelectronic properties, and electronic structures of layered oxychalcogenides MCuOCh (M = Bi, La; Ch = S, Se, Te): Effects of electronic configurations of M^{3+} Ions, Chem. Mater. 20 (2008) 326–334, https://doi.org/10.1021/cm702303r.
[31] V.-A. Ha, F. Ricci, G.-M. Rignanese, G. Hautier, Structural design principles for low hole effective mass s-orbital-based p-type oxides, J. Mater. Chem. C 5 (2017) 5772–5779, https://doi.org/10.1039/C7TC00528H.
[32] N. Kikuchi, A. Samizo, S. Ikeda, Y. Aiura, K. Mibu, K. Nishio, Carrier generation in a p-type oxide semiconductor: $Sn_2(Nb_{2-x}Ta_x)O_7$, Phys. Rev. Mater. 1 (2017) 021601, https://doi.org/10.1103/PhysRevMaterials.1.021601.
[33] J.B. Varley, A. Miglio, V.-A. Ha, M.J. van Setten, G.-M. Rignanese, G. Hautier, High-throughput design of non-oxide p-type transparent conducting materials: data mining, search strategy, and identification of boron phosphide, Chem. Mater. 29 (2017) 2568–2573, https://doi.org/10.1021/acs.chemmater.6b04663.
[34] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: the materials project: a materials genome approach to accelerating materials innovation, APL Mater. 1 (2013), https://doi.org/10.1063/1.4812323.
[35] G. Hautier, A. Miglio, D. Waroquiers, G. Rignanese, X. Gonze, Influence electron effective mass in oxides? A high-throughput computational analysis, Chem. Mater. 26 (2014) 5447–5458, https://doi.org/10.1021/cm404079a.
[36] K. Yim, Y. Youn, M. Lee, D. Yoo, J. Lee, S.H. Cho, S. Han, Computational discovery of p-type transparent oxide semiconductors using hydrogen descriptor, Npj Comput. Mater. 4 (2018), https://doi.org/10.1038/s41524-018-0073-z.
[37] R. Kormath Madam Raghupathy, T.D. Kühne, C. Felser, H. Mirhosseini, Rational design of transparent p-type conducting non-oxide materials from high-throughput calculations, J. Mater. Chem. C. 6 (2018) 541–549, https://doi.org/10.1039/c7tc05311h.
[38] J. Shi, T.F.T. Cerqueira, W. Cui, F. Nogueira, S. Botti, M.A.L. Marques, High-throughput search of ternary chalcogenides for p-type transparent electrodes, Sci. Rep. 7 (2017) 43179, https://doi.org/10.1038/srep43179.
[39] N. Sarmadian, R. Saniz, B. Partoens, D. Lamoen, Easily doped p-type, low hole effective mass, transparent oxides, Sci. Rep. 6 (2016) 20446, https://doi.org/10.1038/srep20446.
[40] G.J. Snyder, E.S. Toberer, Complex thermoelectric materials, Nat. Mater. 7 (2008) 105–114, https://doi.org/10.1038/nmat2090.
[41] H. Zhu, G. Hautier, U. Aydemir, Z.M. Gibbs, G. Li, S. Bajaj, J.-H. Pöhls, D. Broberg, W. Chen, A. Jain, M.A. White, M. Asta, G.J. Snyder, K. Persson, G. Ceder, Computational and experimental investigation of $TmAgTe_2$ and XYZ_2 compounds, a new group of thermoelectric materials identified by first-principles high-throughput screening, J. Mater. Chem. C 3 (2015) 10554–10565, https://doi.org/10.1039/C5TC01440A.
[42] U. Aydemir, J.-H. Pöhls, H. Zhu, G. Hautier, S. Bajaj, Z.M. Gibbs, W. Chen, G. Li, S. Ohno, D. Broberg, S.D. Kang, M. Asta, G. Ceder, M.A. White, K. Persson, A. Jain, G.J. Snyder, $YCuTe_2$: a member of a new class of thermoelectric materials with CuTe4-based layered structure, J. Mater. Chem. A 4 (2016) 2461–2472, https://doi.org/10.1039/c5ta10330d.
[43] J.-H. Pöhls, A. Faghaninia, G. Petretto, U. Aydemir, F. Ricci, G. Li, M. Wood, S. Ohno, G. Hautier, G.J. Snyder, G.-M. Rignanese, A. Jain, M.A. White, Metal phosphides as potential thermoelectric materials, J. Mater. Chem. C 5 (2017), https://doi.org/10.1039/c7tc03948d.
[44] W. Chen, J.-H. Pöhls, G. Hautier, D. Broberg, S. Bajaj, U. Aydemir, Z.M. Gibbs, H. Zhu, M. Asta, G.J. Snyder, B. Meredig, M.A. White, K. Persson, A. Jain, Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment, J. Mater. Chem. C 4 (2016) 4414–4426, https://doi.org/10.1039/C5TC04339E.
[45] S. Matsuishi, Y. Toda, M. Miyakawa, K. Hayashi, T. Kamiya, M. Hirano, I. Tanaka, H. Hosono, High-density electron anions in a nanoporous single crystal: $[Ca_{24}Al_{28}O_{64}]^{4+}$(4e-), Science (80-) 301 (2003) 626–629, https://doi.org/10.1126/science.1083842.
[46] L.A. Burton, F. Ricci, W. Chen, G.-M. Rignanese, G. Hautier, High-throughput identification of electrides from all known inorganic materials, Chem. Mater. 30 (2018) 7521–7526, https://doi.org/10.1021/acs.chemmater.8b02526.
[47] X. Gonze, F. Jollet, F. Abreu Araujo, D. Adams, B. Amadon, T. Applencourt, C. Audouze, J.-M. Beuken, J. Bieder, A. Bokhanchuk, E. Bousquet, F. Bruneval, D. Caliste, M. Côté, F. Dahm, F. Da Pieve, M. Delaveau, M. Di Gennaro, B. Dorado, C. Espejo, G. Geneste, L. Genovese, A. Gerossier, M. Giantomassi, Y. Gillet, D.R. Hamann, L. He, G. Jomard, J. Laflamme Janssen, S. Le Roux, A. Levitt, A. Lherbier, F. Liu, I. Lukačević, A. Martin, C. Martins, M.J.T. Oliveira, S. Poncé, Y. Pouillon, T. Rangel, G.-M. Rignanese, A.H. Romero, B. Rousseau, O. Rubel, A.A. Shukri, M. Stankovski, M. Torrent, M.J. Van Setten, B. Van Troeye, M.J. Verstraete, D. Waroquiers, J. Wiktor, B. Xu, A. Zhou, J.W. Zwanziger, Recent developments in the ABINIT software package, Comput. Phys. Commun. 205 (2016) 106–131, https://doi.org/10.1016/j.cpc.2016.04.003.
[48] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L.W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, AFLOWLIB.ORG, A distributed materials properties repository from high-throughput *ab initio* calculations, Comput. Mater. Sci. 58 (2012) 227–235, https://doi.org/10.1016/j.commatsci.2012.02.002.
[49] S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, Npj Comput. Mater. 1 (2015) 15010, https://doi.org/10.1038/npjcompumats.2015.10.
[50] T. Mueller, A.G. Kusne, R. Ramprasad, Machine learning in materials science: Recent progress and emerging applications, in: Rev. Comput. Chem., (2016) 186–273. doi:10.1002/9781119148739.ch4.
[51] N. Wagner, J.M. Rondinelli, Theory-guided machine learning in materials science, Front. Mater. 3 (2016) 28, https://doi.org/10.3389/fmats.2016.00028.
[52] A. Jain, G. Hautier, S.P. Ong, K. Persson, New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships, J. Mater. Res. (2016) 977–994, https://doi.org/10.1557/jmr.2016.80.
[53] L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, Big data of materials science: critical role of the descriptor, Phys. Rev. Lett. 114 (2015) 105503, https://doi.org/10.1103/PhysRevLett. 114.105503.
[54] P.A. Cox, Transition Metal Oxides, (2010).
[55] D. Waroquiers, X. Gonze, G.-M. Rignanese, C. Welker-Nieuwoudt, F. Rosowski, M. Göbel, S. Schenk, P. Degelmann, R. André, R. Glaum, G. Hautier, Statistical analysis of coordination environments in oxides, Chem. Mater. 29 (2017) 8346–8360, https://doi.org/10.1021/acs.chemmater.7b02766.