Journal of Materials Chemistry C

PAPER

View Article Online
View Journal | View Issue

Cite this: J. Mater. Chem. C, 2016, 4, 4414

Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment†

Wei Chen,[ab] Jan-Hendrik Pöhls,[c] Geoffroy Hautier,[d] Danny Broberg,[e] Saurabh Bajaj,[af] Umut Aydemir,[fg] Zachary M. Gibbs,[f] Hong Zhu,[h] Mark Asta,[e] G. Jeffrey Snyder,[fg] Bryce Meredig,[i] Mary Anne White,[c] Kristin Persson[ae] and Anubhav Jain*[a]

Received 21st December 2015,
Accepted 31st March 2016

DOI: 10.1039/c5tc04339e

www.rsc.org/MaterialsC

We present an overview and preliminary analysis of computed thermoelectric properties for more than 48 000 inorganic compounds from the Materials Project (MP). We compare our calculations with available experimental data to evaluate the accuracy of different approximations in predicting thermoelectric properties. We observe fair agreement between experiment and computation for the maximum Seebeck coefficient determined with MP band structures and the BoltzTraP code under a constant relaxation time approximation (R^2 = 0.79). We additionally find that scissoring the band gap to the experimental value improves the agreement. We find that power factors calculated with a constant and universal relaxation time approximation show much poorer agreement with experiment (R^2 = 0.33). We test two minimum thermal conductivity models (Clarke and Cahill–Pohl), finding that both these models reproduce measured values fairly accurately (R^2 = 0.82) using parameters obtained from computation. Additionally, we analyze this data set to gain broad insights into the effects of chemistry, crystal structure, and electronic structure on thermoelectric properties. For example, our computations indicate that oxide band structures tend to produce lower power factors than those of sulfides, selenides, and tellurides, even under the same doping and relaxation time constraints. We also list families of compounds identified to possess high valley degeneracies. Finally, we present a clustering analysis of our results. We expect that these studies should help guide and assess future high-throughput computational screening studies of thermoelectric materials.

I. Introduction

Thermoelectric materials, which convert thermal to electrical energy and *vice versa*, could play an important role towards building an efficient and sustainable energy portfolio.[1–3] When integrated with traditional or renewable energy generation systems, thermoelectric devices provide attractive possibilities to increase energy efficiency. For instance, thermoelectrics can potentially harvest waste heat from engine exhaust to improve vehicle efficiency or can be employed as cooling systems when extended reliability, silent operation or small size is required.

The maximum efficiency of a thermoelectric material is related to its figure of merit (zT), an intrinsic material parameter that depends on the Seebeck coefficient S, electrical conductivity σ, thermal conductivity κ, and the absolute temperature T:

$$zT = \frac{S^2\sigma T}{\kappa}. \quad (1)$$

To achieve high zT values, one must both maximize the thermoelectric power factor, $S^2\sigma$, and minimize κ.

A wealth of literature has demonstrated the power of atomic-scale modeling approaches based on density functional theory (DFT) in understanding thermoelectric properties of diverse materials.[4–8] Using related techniques, we have initiated a

[a] Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. E-mail: AJain@lbl.gov
[b] Department of Mechanical, Materials and Aerospace Engineering, Illinois Institute of Technology, Chicago, IL 60616, USA
[c] Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada
[d] Institute of Condensed Matter and Nanosciences (IMCN), Université catholique de Louvain, Chemin des étoiles 8, bte L7.03.01, Louvain-la-Neuve 1348, Belgium
[e] Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA
[f] Materials Science, California Institute of Technology, Pasadena, CA 91125, USA
[g] Department of Materials Science and Engineering, Northwestern University, 2220 Campus Drive, Evanston, IL 60208, USA
[h] University of Michigan – Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China
[i] Citrine Informatics, Redwood City, CA 94063, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5tc04339e

high-throughput search for new thermoelectric materials in a similar spirit to earlier studies by Madsen,[9] Wang *et al.*,[10] Carrete *et al.*,[11] Joshi *et al.*[12] and Gorai *et al.*[13,14] The specific materials uncovered and investigated during this long-term study are the topics of other papers.[15] Herein, we present an overview and preliminary analysis of computed thermoelectric properties for more than 48 000 inorganic compounds from the Materials Project (MP). Our aims are two-fold. First, we desire to gain broad insights into the effects of chemistry, crystal structure, and electronic structure on thermoelectric properties. Second, we compare our calculations with experimental data to assess the accuracy of different approximations in predicting thermoelectric properties, including the Seebeck coefficient, power factor and thermal conductivity. We expect that these data can help provide an overview of the thermoelectric chemical space and help shape future computational screening studies.

II. Computational methodology

The high-throughput calculations of thermoelectric properties were performed in two steps. First, DFT calculations were performed for relaxed structures from the MP database to generate electronic band structures. Second, thermoelectric transport coefficients were derived using the Boltztrap software package by solving the Boltzmann transport equation based on a Fourier expansion of the band energies from the first step.[16] We developed the thermoelectric workflow by leveraging the existing MP high-throughput infrastructure that performs automated electronic structure calculations. The workflow and relevant data analysis codes are available in the MP software stack: pymatgen,[17] FireWorks,[18] and MPWorks (available at www.github.com/materialsproject).

a. Band structure calculations

The DFT calculations in this work were performed using the Vienna *ab initio* Simulation Package (VASP)[19] with the Perdew–Burke–Ernzerhof (PBE)[20] generalized gradient approximation (GGA) and projector augmented-wave (PAW)[21] pseudopotentials. Starting with relaxed structures from the MP database, we generated standard conventional cells using the conventions of Setyawan and Curtarolo[22] for band structure calculations. Spin-polarized self-consistent static calculations were performed for these standard conventional cells to converge the charge density using a moderate k-point density of 90 k-points per Å^{-3} (reciprocal lattice volume). Hubbard corrections were considered for transition metal oxides using the standard parameters for MP.[23] Non-self-consistent calculations were then performed using the converged charge density to calculate band structures. These band structure calculations employed a much denser uniform k-point grid. For large band gap systems ($\geq$0.5 eV, estimated from self-consistent runs), 1000 k-points per Å^{-3} were used; for small band gap systems ($<$0.5 eV), we used 1500 k-points per Å^{-3}. Spin–orbit coupling was not considered in the current study, as the goal in this first stage of the screening was to rapidly characterize overall trends.

b. Thermoelectric properties calculations

The electronic transport properties were calculated using the Boltztrap code.[16] Boltztrap performs a Fourier expansion of the band energies using symmetry-conserving star functions to build a smooth analytical representation of the bands, which is then used to solve the semi-classical Boltzmann equations to obtain the conductivity tensor and other electronic transport coefficients. A constant relaxation time approximation, with universal $\tau = 10^{-14}$ s independent of the compound, was used for all calculations. The constant relaxation time approximation is based on an assumption that relaxation time that determines electrical conductivity does not vary strongly at the energy scale of $k_{\text{B}}T$.[24] The universal value for this constant relaxation time employed in this study is a hypothesis that qualitative results (suitable for ranking compounds) for thermoelectric quantities can be obtained without specifying differences in relaxation time, *e.g.*, as was previously assumed for a more narrow class of materials by Madsen.[9] We discuss the level of accuracy obtained through this method in the Results section.

In this study, thermoelectric properties were calculated at theoretical doping levels from 10^{15} cm^{-3} to 10^{20} cm^{-3} with a step size of one order of magnitude for both p-type and n-type doping. Temperature was considered from 100 K to 1300 K at intervals of 100 K. We note that temperature was only introduced within the electronic transport integrals, and do not simulate more complex phenomena such as melting. Additional calculations at specific values of temperature and doping were performed to allow comparisons with experiments under matching conditions.

c. Data set

The full data set includes 48 770 compounds from the MP database.[25] The majority of the data set represents compounds first reported in the Inorganic Crystal Structure Database[26,27] (previously reported compounds). The data set also contains compounds from previous high-throughput projects, *e.g.*, a Li-ion battery screening project,[28] as well as structures from other databases, *e.g.*, the Open Quantum Materials Database.[29] We intend to release the full data set as part of the MP database in the future.

III. Results

a. Distribution of properties in the data set

To better understand the distribution of properties exhibited within this large space of compounds, we plot a flow diagram representation of the data set in Fig. 1.

The left side of Fig. 1 indicates the preferred doping of compounds, which represents the type of carriers that would in theory lead to the highest computed power factor (provided that such doping is achievable in practice). Most of the compounds in our study perform better as n-type materials. This trend might stem in part from the fact that our data set is largely composed of oxygen-containing compounds ($\sim$70% of the data set) and that such compounds tend to achieve higher power factors as n-type ($\sim$61% of the oxygen-containing materials in

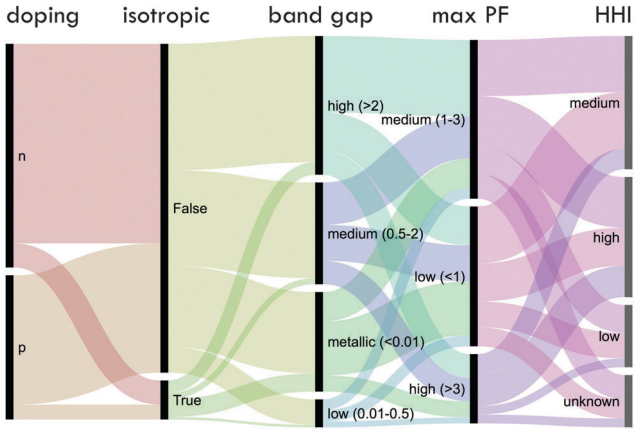

Fig. 1 Flow diagram representation of the data set of 48 770 compounds in this study. The term "doping" refers to the doping type that maximizes the power factor in our calculations. Compounds with power factors within 10% in all directions are considered "isotropic". The unit of band gap is eV and the unit of power factor is mW K^{-2} m^{-1}. HHI represents the Herfindahl–Hirschman Index of elemental reserves for the compound and is a measure of the resource criticality of the compound's component elements, with "high" indicating that resources are geographically confined (undesirable) and "low" indicating geographically dispersed (desirable).

our study perform better as n-type). However, the preponderance of more promising n-type band structures might also be an artifact of our data set. Thus far, we note that the zT value of the best n-type oxide thermoelectrics is only about 0.3 whereas the p-type oxides have attained zT values about four times higher[30,31] (and has motivated a search for new n-type oxide thermoelectrics).

The second column of Fig. 1 represents whether the power factor is isotropic, *i.e.*, whether all eigenvalues of the power factor tensor are within 10% of one another. The vast majority of compounds is not isotropic, and can only achieve their optimum power factor in preferred directions. This is potentially problematic as it indicates that achieving high performance might require the use of single crystals or textured polycrystals.

The third column of Fig. 1 indicates the band gap as calculated by the GGA/GGA+U functional. One caveat when interpreting this column is that these functionals are known to underestimate experimental band gaps. The amount of underestimation is system-dependent and can be difficult to predict. Chan and Ceder[32] reported that the GGA-PBE functional underestimates semiconductor and insulator band gaps by an average of 0.73 eV compared with experiment, and in another study Setyawan *et al.* (focused on large-gap systems) observed an error of 42% (this error could be reduced to 24% using a linear shift).[33] With this consideration in mind, most compounds in our study can be classified as large-gap semiconductors or insulators, although a significant number of systems are metallic. The relatively small fraction of small-gap semiconductors is one shortcoming of this data set.

The fourth column of Fig. 1 indicates the maximum power factor along any direction. As stated in the Methodology, this value is calculated under a constant and universal relaxation time approximation and typically assumes that relatively high doping (up to 10^{20} cm^{-3}) can be achieved. We expect that compounds with power factor greater than 3 mW K^{-2} m^{-1} are of high potential interest to form next-generation thermoelectrics, and values less than 1 mW K^{-2} m^{-1} are of low interest. This criterion is based on a goal of reaching zT higher than unity at temperatures of 1000 K or below and assuming a thermal conductivity of 1 W m^{-1} K^{-1}. The band gap does not appear to have a major effect on the fraction of compounds that achieve a particular power factor, except that metallic compounds are more likely to have lower power factors (as would be expected, because metals typically possess low Seebeck coefficients).

The last column of Fig. 1 represents the Herfindahl–Hirschman Index[34] (HHI) for known reserves of the elements comprising the compounds (in the ESI,† we also present the elemental production data in Fig. S1). The data are derived from the report of Gaultois *et al.*,[35] who introduced this measure to the thermoelectrics community as a way to gauge resource constraints on new thermoelectric materials. Specifically, this measure indicates whether known reserves of the elements comprising the compound are geographically dispersed or confined. Unfortunately, the majority of materials in our study would be considered "highly concentrated" or "moderately concentrated" under definitions drafted by the U.S. Department of Justice and the Federal Trade Commission.[36] In particular, only a small subset of the total number of compounds possesses both "high" power factor and "low" HHI.

Overall, Fig. 1 visually represents many of the tradeoffs that make thermoelectrics design particularly challenging, especially in identifying isotropic systems with low HHI index that also meet other performance criteria.

b. Effects of chemistry and crystal structure on power factor

High-performance thermoelectrics (such as PbTe[37] and Bi_2Te_3[38]) typically include heavy anions such as Te^{2-} or Sb^{3-}. It is well-known that it is much more challenging to develop high-performance thermoelectrics within lighter anion chemistries such as oxides. Compared with oxides, the smaller differences in the element electronegativity of tellurides and selenides tend to increase carrier scattering time and improve electronic mobility.[3,39] High atomic mass is also likely to reduce the atomic vibration frequency and the lattice thermal conductivity. Our high-throughput data set provides an opportunity to explore if further fundamental differences exist between chemistries and study how the thermoelectric power factor varies with the selection of anion.

We focused specifically on anions with a −2 charge state: tellurides, selenides, sulfides and oxides. We only considered materials with a clear oxidation state assignment from a bond valence algorithm[40] and without anion mixing (*e.g.*, we excluded oxysulfides). Fig. 2 depicts violin plots for the distribution of the highest power factor (averaged over the three eigenvalues of the power factor tensor) achievable in the four chemistries. In the violin plots, the probability density of the computed power factor is plotted at both sides of the y axis for each chemistry. This allows one to observe the distribution of data values over large data sets. It is important to note that temperature and doping were varied freely up to 1300 K and 10^{20} cm^{-3} charge

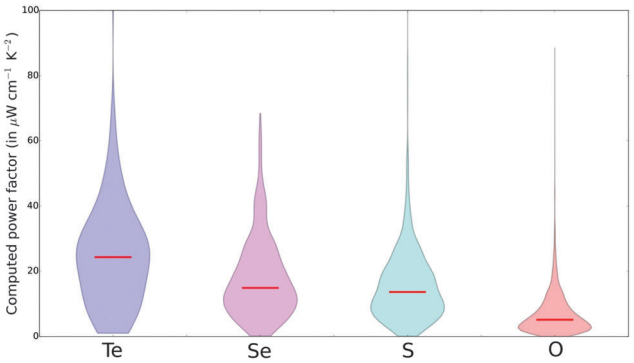

Fig. 2 Violin plots depicting the distribution of computed maximum power factor under a fixed relaxation time approximation across compounds in the MP database, separated by anion type. The red lines indicate the median computed power factor for different anions. The distribution of power factors for tellurides tends to be higher than that for selenides and sulfides, which are in turn higher than those of oxides.

carrier density, respectively, for all chemistries; thus, variations in the power factor reported here do not reflect differences in achievable doping. Similarly, the relaxation time was set equal for all materials (10^{-14} s). In this instance, we consider this to be an advantage of the computational method in that it allows one to more clearly separate the factors that cause differences in attainable power factor. In Fig. 2, the reported differences stem only from systematic differences in the calculated band structure.

Under these constraints, Fig. 2 illustrates that tellurides lead to higher power factors than the other three chemistries. Selenide and sulfide compounds exhibit very similar power factors on average, which are slightly lower than the tellurides. The oxides exhibit by far the lowest power factors with a significantly lower median power factor than the tellurides, sulfides, and selenides. In addition to a lower median value, the variance within the oxides is much smaller, leading to many fewer materials with extremely high power factors. It is important to reiterate that the maximum achievable power factor in this analysis depends only on the band structure characteristics. One explanation for this trend could be that a greater hybridization in tellurides[41] could lead to a larger tendency to form materials with high valley degeneracy that results in both high Seebeck coefficients and high electrical conductivities.

It is also interesting to consider how these results might relate to the attainable zT, which will depend on the thermal conductivity and the dopability differences between materials. The thermal conductivity is generally expected to be higher for lighter anion elements such as oxygen.[42] In terms of dopability, materials with large band gap are typically more difficult to dope.[43] Because band gaps increase from the tellurides to the oxides, the expected dopability should decrease on average. Therefore, moving from tellurides to oxides (*i.e.*, going to lighter anions) could be detrimental on three accounts: thermal conductivity, dopability, and maximum achievable power factor (based on band structure characteristics that are independent of doping). These three factors combined make it very challenging to develop high-performance oxide thermoelectrics. It has been discovered that layered $Na_{0.5}CoO_2$ has good thermoelectric properties.[3,44] However, it remains to be determined whether the diversity of oxide chemistry in combination with other advantages, such as abundance and potentially low cost, can produce more promising candidates.

c. Analysis of valley degeneracy

Another aspect of band structure that can be examined is the thermoelectric quality factor $B = \frac{2k_B^2\hbar}{3\pi}\frac{N_v C_l}{m_I^* E_{def}^2 \kappa_L}$, in which N_v is the valley degeneracy, C_l is the average longitudinal elastic modulus, m_I^* is the inertial mass, E_{def} is the deformation potential, and κ_L is the lattice thermal conductivity. The thermoelectric quality factor scales with the maximum attainable power factor for a given material,[45] and is directly proportional to the valley degeneracy N_v, which is the focus of our analysis. N_v is a fundamental materials property that is related to both the crystal symmetry and the symmetry of the k-points in reciprocal space at which the primary conduction and valence band extrema occur.

In this work, we have calculated N_v for compounds within the MP database in a manner similar to ref. 46, in which band degeneracies were tabulated for 412 compounds. In our analysis, only isotropic compounds of energies within 50 meV per atom of the convex hull were considered for the calculations, resulting in a total of approximately 1400 materials for which we analyzed valley degeneracy. Brillouin zone symmetry operations were obtained using the spglib library[47,48] and pymatgen.[17] The symmetry operations (rotations) were performed on the k-points corresponding to the primary valence band maximum (VBM) and conduction band minimum (CBM). After rotation, N_v was determined by counting the number of unique points that existed within the first Brillouin zone. All other extrema that occurred at the same energy of the VBM or CBM were also considered.

The overall trends of our analysis (Fig. S2, ESI†) indicate that many compounds possess N_v = 3 for the valence band and N_v = 1 for the conduction band. The triply degenerate valence band is likely due to convergence of separate bands at the Γ-point (as is the case for all group VI and III–V semiconductors in our study and in the absence of spin–orbit coupling). Some compounds noted to have high degeneracies are discussed next.

One family of compounds with high valley degeneracies is the half-Heusler compounds with space group 216. Interesting compound examples (and their corresponding identification numbers within the MP database) include VSbRu (mp-31455), NbSbRu (mp-505297), TaSbRu (mp-31454), NbSbFe (mp-9437) and ZrBiCo (mp-31451), which possess degeneracies of 8 and 3 for the VB and CB, respectively. The VB in these compounds is pushed up in energy at the L-point, making it the VBM, as illustrated in Fig. 3 of the electronic band structure of NbSbRu. In most other half-Heusler compounds, such as ZrNiSn (mp-924129), LiZnP (mp-10182) and HfNiSn (mp-924128), the VBM is at the Γ-point, causing the VB degeneracy to be lower at 3. A common trend in the high degeneracy half-Heusler compounds mentioned above is that the tetrahedral zinc blende-type structure is built up of an electropositive element

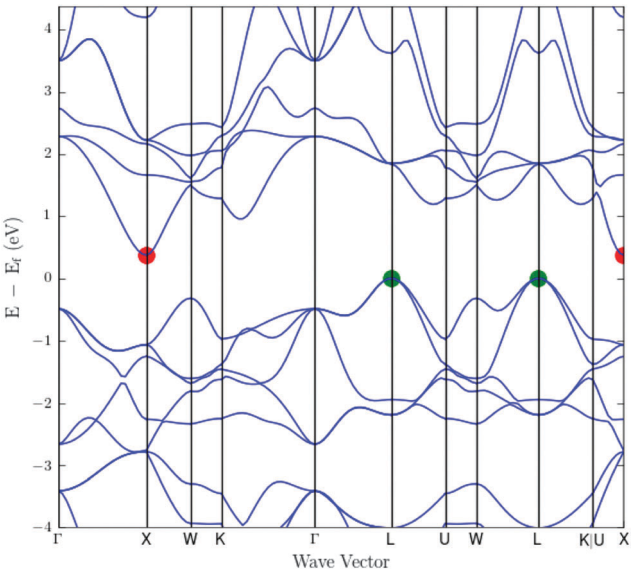

Fig. 3 Electronic band structure of NbSbRu (MP materials id = mp-505297) computed using DFT-GGA.

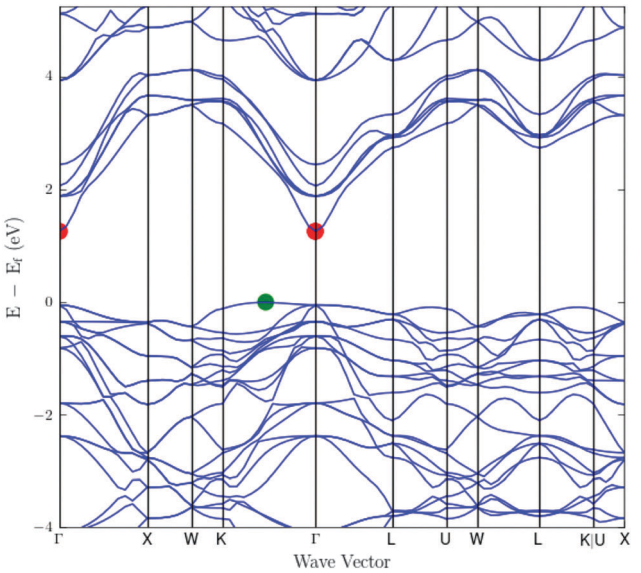

Fig. 4 Electronic band structure of Al_2CdSe_4 (MP materials id = mp-3807) computed using DFT-GGA.

from the s^2d^3 group of the periodic table (V, Nb, Ta), and an element of intermediate electronegativity from the s^2d^6 group of the periodic table (Ru, Fe).[49] None of these elements is present in the lower degeneracy compounds. This class of compounds has recently been observed to have high zT and is highly optimizable for high-temperature thermoelectric modules.[49,50]

A second family of compounds with high valley degeneracy is chalcogenide compounds with space group 227 of the A_2BC_4 family, including: Al_2CdSe_4 (mp-3807), Al_2CdS_4 (mp-9993), Al_2ZnS_4 (mp-4842), In_2ZnS_4 (mp-22052), In_2HgS_4 (mp-22356), In_2CdS_4 (mp-559200), In_2MnS_4 (mp-22168), Cr_2CdS_4 (mp-4338), and Cr_2CdSe_4 (mp-22605) which all possess VB degeneracies of 12. As can be observed in Fig. 4 of the electronic band structure of Al_2CdSe_4, the VBM in these compounds is shifted from the Γ-point and lies along the Γ–K line, causing high degeneracies. However, the effective mass may be prohibitively high in this case. In other compounds within the same family, such as Tm_2CdSe_4 (mp-14620), Lu_2MnS_4 (mp-14305), and In_2MgS_4 (mp-20493), the VBM is at the Γ-point, leading to lower VB degeneracies of 3. The most electropositive element in high-degeneracy compounds belongs to the s^2p^1 group (Al, In) or Cr, in combination with an element of (similar) intermediate electronegativity from the s^2d^{10} group or the s^2d^5 group of the periodic table (Zn, Cd, Hg, Fe). The VB degeneracy is lowered to 3 if either the most electropositive element is a rare-earth (Dy, Ho, Er, Tm, Yb, Lu, and Sc) or Mg, and seems to be associated with greater differences in electronegativity between the two most electropositive elements.

Other chalcogenide compounds of space group 217 include Tl_3VS_4 (mp-5513), with degeneracies of 6 for both the VB and CB, and Tl_3TaSe_4 (mp-10644) and Tl_3TaS_4 (mp-7562), which possess degeneracies of 6 and 3 for the VB and CB, respectively. These compounds share a few commonalities: their VBM are at the N-point, and they all contain Tl together with an element from the s^2d^3 group of the periodic table (V, Ta). However, the compounds K_3SbS_4 (mp-9911) and Na_3SbSe_4 (mp-8703), which also fall within the same family, have their VBM at the H-point, causing the VB degeneracy to decrease to 3.

The well-studied lead chalcogenides of space group 225, PbTe (mp-19717), PbSe (mp-2201), and PbS (mp-21276) have N_v = 4 for VB and CB.[13] The combination of an element from the s^2p^2 group of the periodic table (Pb, Ge), and an element from the s^2p^4 group of the periodic table (S, Se, Te), causes the L-point to shift up in energy, making it the VBM. In many other compounds within the same family, the degeneracy is lowered to 3 at the Γ-point.

d. Assessment of high-throughput electronic property calculations

A major concern in high-throughput assessments of thermoelectrics properties is the level of accuracy obtained. Therefore, in this section, we compare electronic properties from our high-throughput computation procedure against experiments. For this comparison, both the Seebeck and power factor were computed as stated in the methodology, but with two modifications. First, the temperature and carrier concentrations used for solving the Boltzmann transport equation were set to the conditions under which the experiments attained the maximum values for these quantities. Second, we averaged the eigenvalues differently depending on whether the experimental data was taken from a single crystal measurement or from a polycrystalline measurement. When comparisons were made with single crystal results, we used the maximum calculated eigenvalue. When comparisons were made with polycrystalline experimental results, we averaged the three eigenvalues. Through this procedure, we attempt to match the computational and experimental conditions as closely as possible. We calculated the transport properties for the host compounds instead of the doped samples.

First, we discuss results for the Seebeck coefficients (see Table S1, ESI† for the full data set). As depicted in Fig. 5(a), the calculated Seebeck coefficient correlates strongly with the experimental data for most materials, but tends to be underestimated. The smaller computed Seebeck coefficient can be traced to the underestimation of the experimental band gap energy, which is on average approximately 0.15 eV (38%) lower than the experimental band gap energy over our data set for the comparison. In particular, thermoelectric compounds with low GGA band gap energies (<0.2 eV; black dots in Fig. 5(a)) appear to have greater discrepancies with experiment. It is important to note that compounds computed to have zero band gap can exhibit an incorrect sign of the Seebeck coefficient (negative S for p-type and positive S for n-type; see red circles in Fig. 5(a) and Table S1, ESI†). The Pearson correlation coefficient for this data set is 0.79 and the Spearman rank correlation coefficient is 0.62.

To see whether correcting the band gap problem would lead to better agreement, we recalculated the data set after applying a 'scissor operation'[51,52] to compounds for which the discrepancy in band gap was larger than 50%. The scissor operation matches the computational band gap to the experimental value but retains the other features of the original band structure. With the scissored band gap energies, the agreement between experimental and computed Seebeck coefficients significantly improves as indicated in Fig. S3(a) and Table S2 (ESI†). With the application of scissoring, the Pearson correlation coefficient increases to 0.90 and the Spearman's rank correlation coefficient increases to 0.91. We note that the band gap energies of Ag_2Se (mp-568889) and Cu_2Te (mp-1861), for which the sign of Seebeck coefficient is incorrect, could not be scissored due to zero GGA band gap.

We note that Mg_2Sn (mp-2343) exhibits a higher calculated Seebeck coefficient (-213 μV K^{-1}) than the experimental value (-71.5 μV K^{-1}) after the 'scissor' operation.[53] It is possible that this stems from the difficulty in determining an accurate band gap energy (the range of reported experimental band gap energies is E_g = 0.16–0.33 eV[54,55]). An additional source of error in this analysis is that carrier concentrations are taken from experimental Hall measurements, which tend to underestimate the carrier concentration.[56] Despite these issues, we find that computational assessments of the Seebeck coefficients under the approximations employed generally match well with experimental data and are appropriate for high-throughput screening.

In addition to analyzing the Seebeck coefficient, we also compare the computed power factor under a constant and universal relaxation time approximation to experiment (Fig. 5(b), and Table S3, ESI†). Although Fig. 5(b) shows qualitative agreement between computation and experiment, we observe a larger discrepancy than for the Seebeck coefficient. This is also reflected in the lower Pearson (0.33) and Spearman rank (0.48) coefficients. In contrast to the Seebeck coefficient, scissoring the band gap does not appear to improve the result (ESI,† Fig. S3(b) and Table S4). For the scissored data, the Pearson coefficient is 0.30 and the Spearman rank correlation coefficient is 0.33. It is surprising that the scissored power factors exhibit a weaker correlation than the un-scissored values, especially because the Seebeck coefficient improved with scissoring.

One major reason for the difficulty in computing power factors is likely the constant, universal relaxation time approximation ($\tau = 10^{-14}$ s) used in our high-throughput calculations. The relaxation time should instead depend on several factors, including the selection of material, temperature, and level of impurities in the microstructure. Further, it will be energy-dependent;[57,58] in most thermoelectric materials, the relaxation time decreases with temperature due to acoustic phonon–electron interactions.[59,60] Our results indicate that ignoring these factors leads to only very rough qualitative agreement.

As a potential example of how non-uniform relaxation time might have affected our results, in Bi_2S_3 (mp-22856) thin films, the experimentally measured power factor is higher than the calculated value. This could be because the relaxation time in the highly crystalline and oriented crystals[61] used in experiments is likely larger than the value assumed in the calculation.

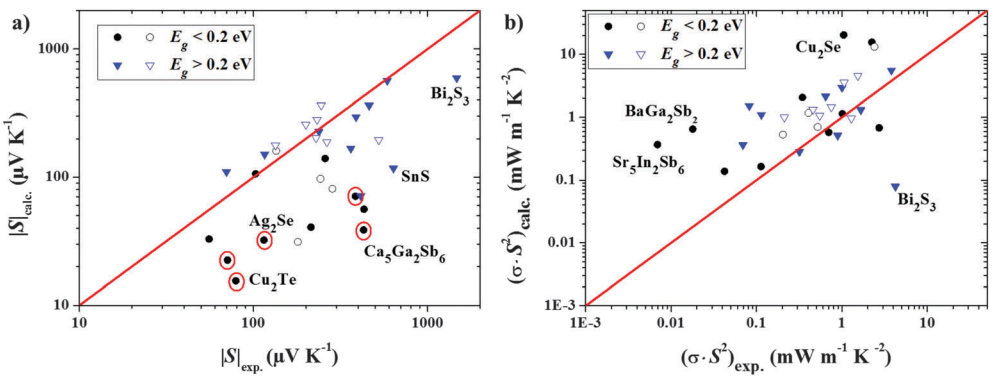

Fig. 5 Comparison of the calculated and experimental (a) Seebeck coefficients and (b) power factors for different thermoelectric materials (filled symbols) and the extrinsic doped compounds (hollow symbols). The red circles indicate that the sign of the Seebeck coefficient was incorrect (see Table S1, ESI†), and the red lines indicate equality of computation and experiment. The calculations are based on the Boltzmann transport equation applied to MP band structures under a constant relaxation time approximation, with temperature and carrier concentration set to conditions where experiments reach the maximum value of the quantity being investigated (Seebeck coefficient in panel (a), power factor in panel (b)). A major source of error appears to be underestimation of band gaps in DFT, as indicated by the poorer agreement for materials with small DFT band gaps (black circles) *versus* intermediate/large gaps (blue triangles).

For $BaGa_2Sb_2$ (mp-29938), which has a higher calculated power factor than the experimental value (see Fig. 5(b)),[59] the reverse situation may be true due to the smaller relaxation time of polycrystalline samples.

We next compare calculated and experimental electrical conductivities and mobilities. In Fig. S4 (ESI†), the ratio of the calculated to experimental mobilities ($\mu = \sigma/ne$), where e is the elementary charge and n is the measured carrier concentration, is plotted as a function of the calculated band-gap energies. We note that because the carrier concentration is set equal between computation and experiment, the mobility comparison is equivalent to a conductivity comparison. Our results indicate that smaller calculated gaps correspond to greater overestimation of the mobility (Fig. S4(a) and Table S5, ESI†). One culprit could be inaccurate band curvatures (and therefore effective masses) for smaller band gaps, especially for near metallic materials. Although DFT often reproduces accurate effective masses for wide band gap materials,[62,63] this not always the case.[64] Calculated and experimental mobilities agree more closely if the band gap energies are set to the experimental band gap energy (Fig. S4(b) and Table S6, ESI†), suggesting that the issues may stem both from band curvature as well as the band gap. Techniques beyond DFT such as the Tran–Blaha[65] or Heyd–Scuseria–Ernzerhoff (HSE)[66] functionals that are known to open the band gap might be better to model accurately effective masses in small gap semiconductors.

Finally, we examined temperature dependence for the example of the half-Heusler compound ZrNiSn (mp-924129).[67] As in the earlier comparisons, the carrier concentration and temperature were set equal for the computations and experiments. The results for n-type ZrNiSn are plotted in Fig. 6. Although the Seebeck coefficient agrees well over the full temperature range of 300 K to 800 K, the power factor agrees well only at low temperatures. At temperatures above 600 K, the calculated power factor is overestimated, most likely due to an overestimation of the relaxation time since this quantity is held fixed in our calculations but generally decreases with temperature due mostly to phonon–electron interaction.

In conclusion, computed Seebeck coefficients demonstrate generally good agreement with experiment; however, estimating the power factor in high-throughput calculations under a universal and constant relaxation time approximation is less predictive and yields only qualitative agreement. Further refinement of the relaxation time, or alternate metrics for high-throughput screening,[46] might be required to enhance the accuracy of the calculations.[68–70] Finally, it should be noted that these results tested the models assuming that the carrier concentration was known. For high-throughput screening purposes, an additional uncertainly generally exists in estimating the achievable carrier concentration.

e. Assessment of high-throughput thermal conductivity models

Low thermal conductivity is required for thermoelectric materials to maintain the temperature gradient across the material. Calculating the thermal conductivity from first-principles is difficult because it consists of both electron and phonon

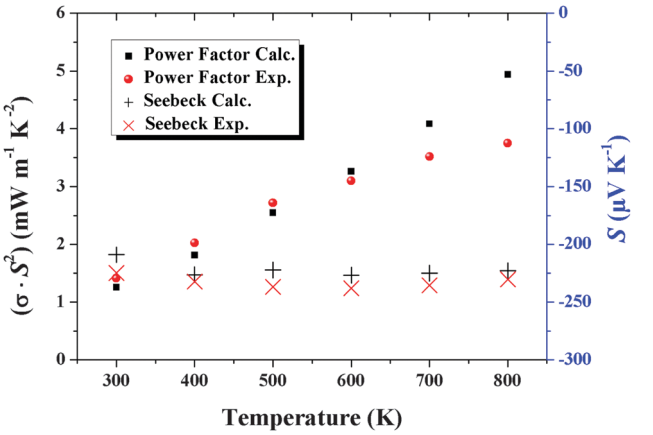

Fig. 6 Comparison of calculated and experimental power factor and Seebeck coefficient for n-type ZrNiSn as a function of temperature.[67] The calculations apply the Boltzmann transport equation to MP band structures under a constant relaxation time approximation; carrier concentrations for the computations were set equal to the experimental value at that temperature. The Seebeck coefficient shows good agreement over the full range of temperatures, whereas the computed power factor deviates from experimental measurements at high temperatures.

contributions. The electronic contribution can be determined from the Lorenz number and the electrical conductivity, whereas the phonon contribution requires an enhanced understanding of the lattice dynamics. Many authors have suggested methods to predict the phonon thermal conductivity; however, the more reliable methods require computationally intensive density functional perturbation theory calculations.[71,72] These methods are currently outside the scope of high-throughput screening methods, mostly due to the extended computational time required to run these simulations. In addition, the thermal conductivity can vary widely based on microstructure,[73] which fundamentally limits the scope of DFT-based approaches, which focus on perfect bulk materials.

In this study, the thermal conductivities were computed using simpler models and compared with experimental thermal conductivities. Because we are targeting low thermal conductivity, we discuss amorphous-limit thermal conductivity values (κ_{min}) using the Clarke model[74] and the Cahill–Pohl model,[75,76] which provide a lower bound for the thermal conductivities in the high temperature limit ($T \gg \theta_{Debye}$). Although the two models have different starting points, they arrive at similar expressions by assuming that the mean free path takes the limiting value of the interatomic spacing.

The minimum thermal conductivity in the Clarke model is defined by

$$\kappa_{min}^{Clarke} = 0.87 k_B \left(\frac{N_A m \rho}{M}\right)^{2/3} \left(\frac{E}{\rho}\right)^{1/2} \quad (2)$$

in which E is the Young's modulus, M is the mass per unit cell, m is the number of atoms per unit cell, N_A is Avogadro's number, and ρ is the mass density.[74] The speed of sound in the Clarke model is calculated by $v_s = \left(\frac{E}{\rho}\right)^{1/2}$.

The minimum thermal conductivity using the Cahill–Pohl model is given by:

$$\kappa_{\min}^{\text{Cahill}} = \frac{1}{2.48} k_B n^{2/3} (v_L + 2v_T) \quad (3)$$

in which n is the number density of atoms, v_L and v_T represent longitudinal and transverse sound velocities, respectively.[75,76] In principle, the sound velocities needed for the Cahill–Pohl model could be obtained from first-principles phonon calculations. However, in our application of the Cahill–Pohl model, we obtain these values from elastic properties of materials for which we recently published a database[77] of first-principles values, utilizing the Voigt–Reuss–Hill average for bulk and shear moduli of a material.[78] To compute these moduli, the automated formalism produced 24 different deformed structures by applying small strains (±1% and ±0.5%) to the ground state structure and then solving for the corresponding stress/strain tensor. The full procedure is described in ref. 77. Thus, in our implementation, the longitudinal and transverse sound velocities were computed from the following formulas appropriate for an elastically isotropic solid:

$$v_L = \left(\frac{K + 4/3G}{\rho}\right)^{1/2} \quad (4)$$

$$v_T = \left(\frac{G}{\rho}\right)^{1/2} \quad (5)$$

where K is the bulk modulus, G is the shear modulus, and ρ is the mass density.

Fig. 7 plots $\kappa_{\min}$ computed using the Clarke and Cahill models, with parameters taken from our computations, compared with the lowest measured thermal conductivity above 300 K (Table S7, ESI†). The experimental minimum thermal conductivities agree qualitatively with both models. The Clarke and the Cahill model both have a Pearson correlation coefficient of 0.82 and a Spearman rank correlation coefficient of 0.78 and 0.77, respectively. However, the calculated minimum thermal conductivity is in general lower than in experiments. This may not be a failure of the model, but rather because the glass-like localized lattice vibrations assumed in the models are not completely realized in experimental samples.[74,76] With precise phonon engineering, the thermal conductivity of these compounds might be further reduced to the calculated $\kappa_{\min}$ limit.[73] We note that although it is possible for the experimental thermal conductivity to be lower than $\kappa_{\min}$ from these models in 'ultralow'[76] materials (*e.g.*, WSe_2[79]), in our study this condition is more likely due to overestimation of the speed of sound from the computation (the elastic tensor calculations have uncertainties of roughly 15%[77]). We note that in a recent study, a modified minimum thermal conductivity model explained the origins of ultralow thermal conductivity in PCBM and C_{60}/C_{70}.[80] For the model, the experimental Debye temperature from heat capacity measurements was used and the heat transfer was limited by the atomic density.

Overall, we conclude that the Clarke and the Cahill–Pohl models are appropriate for high-throughput screening to determine the lowest limit of the thermal conductivity in the materials (*i.e.*, after phonon engineering).

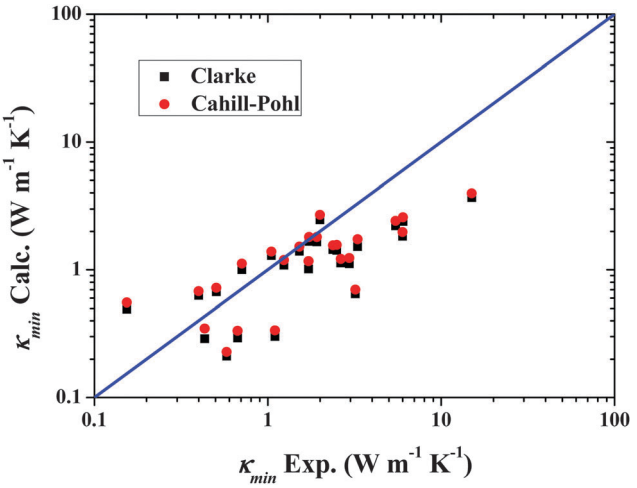

Fig. 7 Calculated thermal conductivities in the amorphous limit and experimental minimum thermal conductivities from two models (Clarke and Cahill Pohl) with parameter values computed from DFT-GGA plotted on a log–log scale. The blue line plots equivalence between computation and experiment.

f. Clustering materials for thermoelectrics

Given that the present work involves many tens of thousands of diverse compounds, we used large-scale data analysis techniques to organize and interpret the results. In particular, here we employ a clustering analysis to determine if it is possible to define useful groups of thermoelectric materials based on intrinsic descriptors. We use 58 descriptor values for each compound. These descriptors include various DFT-derived and empirical characteristics of the materials in the data set; examples include band gap, nearest-neighbor bond length, elemental electronegativities, and atomic radii. Because our objective is to predictively model the key thermoelectric properties of Seebeck coefficient and electrical conductivity, we exclude these target thermoelectric quantities from our clustering process. We excluded materials for which we lack some descriptor values (these are mostly rare earth and noble gas elements) and furthermore exclude materials that are thermodynamically unstable (energy above hull not equal to zero in MP). These reductions limited our consideration to a subsample of 5431 ground-state structures.

We employ the DBSCAN algorithm (density-based spatial clustering of applications with noise)[81] as our clustering method. This algorithm groups data into clusters of high density and assigns points that fall between these clusters as outliers. Importantly, in the DBSCAN approach, the user does not need to pre-define the number of desired clusters (as is required with, for example, *k*-means clustering[82]).

The results of our clustering of 5431 candidate thermoelectrics are presented in Fig. 8. We utilized t-distributed stochastic neighbor embedding (t-SNE)[83] to project the results of our 58-dimensional clusters onto two dimensions for visualization purposes. A list of detailed clustering results can be observed in the ESI.† Our clustering reveals six distinct clusters (colored points in Fig. 8), while materials that do not clearly belong to one of these clusters are classified as noisy (gray points encompassing

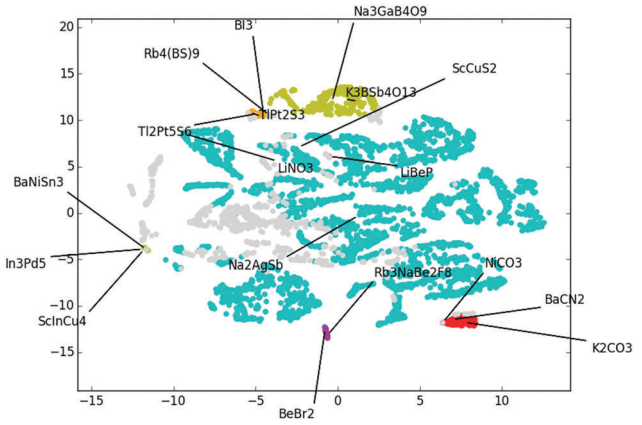

Fig. 8 Visualization of the clustering of materials in the dataset. The axes are t-SNE dimensionality-reduced coordinates, which have no physical meaning, but are used to flatten the very high-dimensional materials property space onto two dimensions for visualization.

830 materials, or 15% of the data set). In Fig. 9, we plot how the average values for some key descriptors we considered (y axis) vary by cluster (x axis), normalized by the averages of these descriptors across all materials. We discuss trends and key descriptors within each cluster below. The list of compounds within each cluster is presented in the ESI.†

The largest cluster (cyan; 4076 materials; colors refer to Fig. 8) contains 75% of all materials we considered, and thus its average descriptor values are similar to those of the entire dataset. Notably, this massive cluster includes many canonical thermoelectrics, including Bi_2Te_3, PbTe, SnTe, and Mg_2Si.

The yellow cluster (297 materials) consists largely of borates and materials with complex anions containing several of the set {B, F, Cl, O, Pt}. This cluster exhibits the third-highest average Seebeck coefficient but low electrical conductivities. This latter limitation is expected because many borates and compounds with electronegative ions such as F, Cl, and O form insulators and highly ionic compounds. Based on Fig. 9, we note that short nearest-neighbor bonds, large negative formation energies, and the presence of elements with high electronegativity distinguish the borate cluster.

The red cluster (104 materials) contains a large number of carbonate and related compounds. This cluster exhibits the second-highest average Seebeck coefficient, but also has low electrical conductivity. Fig. 9 demonstrates that this cluster exhibits large deviations from overall descriptor averages in nearest-neighbor bond length (the shortest among all clusters), number of elements per compound (the largest among all clusters), and average atomic number (very low).

The orange cluster (54 materials) is a group of heavy element chalcogenides (many Pt compounds) and pnictides. In particular, Pt and BS_3 thioborates are abundant in this cluster. This cluster is not an obvious outlier in terms of Seebeck coefficient or electrical conductivity, yet the behavior of these two properties in concert leads to the largest maximum theoretical power factors on average among the clusters. This cluster is relatively high in density, has less-negative formation energies, is unlikely to contain any strongly electronegative elements, and tends to exhibit semiconductor-magnitude band gaps. This cluster could potentially yield unexpected, interesting thermoelectric candidate materials.

The magenta cluster (51 materials) consists mostly of Be-containing, strongly ionic chalcogenides and halides (*e.g.*, BeI_2, $BaBeF_4$, $MgBe_2N_2$). This cluster exhibits the highest average Seebeck coefficient and low electronic conductivities. Based on Fig. 9, we note that this cluster has the largest (*i.e.*, most negative) formation energies among the clusters, is the least dense, is likely to have a wide direct band gap, and is much more likely than average to contain alkaline earth elements (*e.g.*, Be).

The light green cluster (19 materials) contains binary and ternary Sc and In-based intermetallic compounds. As might be expected, this cluster exhibits exceptionally low Seebeck coefficients and very high conductivities, leading overall to low power factors.

IV. Discussion

Within the last decade, many new thermoelectric materials have been discovered, such as SnSe,[84] tetrahedrite,[85] colusite,[86]

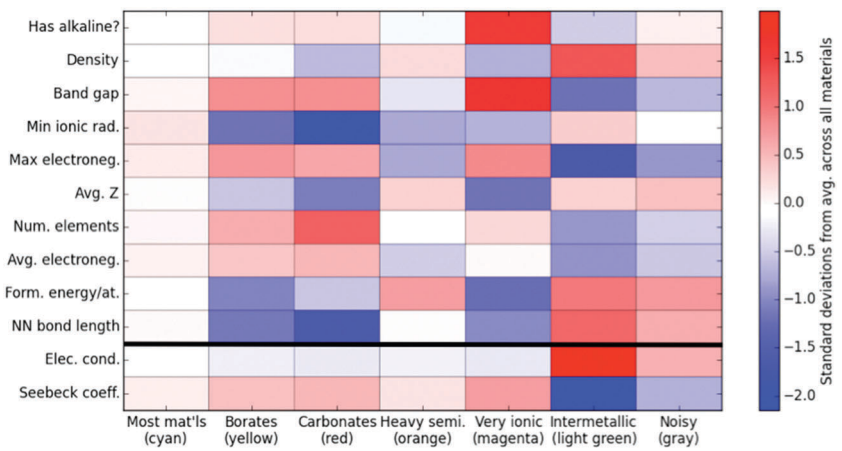

Fig. 9 Heat map of selected descriptors used in clustering of materials. The 7 clusters are listed on the x-axis, and the map is used to illustrate the difference in properties between clusters. The thermoelectric properties below the black lines were not used in clustering materials.

Cu_2S/Se,[87] and others.[88] Some of these new materials exhibit very high performance; for example, the peak zT of SnSe was recently measured to be 2.6 at 923 K, the highest for any material reported thus far.[84] Other new materials are abundant and potentially low cost, such as the tetrahedrite structures[85] that are already being tested commercially by Alphabet Energy in large 25 kW thermoelectric generators. Thus, there may be further interesting opportunities in searching for new bulk thermoelectric materials. Approaches based on high-throughput screening and data mining offer a possibility to efficiently and comprehensively explore these opportunities.

One concern with high-throughput methods is accuracy. We evaluated the accuracy of a high-throughput approach to calculating electronic transport using a combination of standard density functional theory (*i.e.*, GGA-PBE functional) and Boltztrap, using a constant (no dependence on E or k) and universal (no dependence on material composition or structure) relaxation time. Even under these approximations, we observed that calculations of the Seebeck coefficient matched well with experiments, with the few disagreements stemming from the underestimation of the band gap with GGA. These results, performed within a high-throughput study, are consistent with previous individual studies on Mg_2Si,[89] Bi_2Te_3[9] and $CoSb_3$,[9] and half-Heusler compounds.[90]

In contrast to the Seebeck coefficient, we determined the power factor to be more difficult to predict in a high-throughput mode. This is likely due to the larger sensitivity of electronic conductivity to the selection of relaxation time, carrier concentration, and band gap. Our results indicate a systematic bias in the high-throughput method for which compounds with the lowest GGA band gap yielded the largest errors, in particular having calculated mobilities that were too high. Based on these results, one suggestion would be to correct the band gaps of compounds from GGA calculations in an efficient way, *e.g.*, by calculating the energies of the band edge symmetry points with a more advanced functional such as HSE.[91,92] Alternatively, one might use a formalism like one recently developed by Yan *et al.*,[46] which employs the bulk modulus and band effective mass to estimate the mobility along with experimental fitting parameters.

One component of accuracy that we did not test in this study was the effect of spin–orbit coupling, which might play a greater role in compounds with heavy elements and high degeneracies. In addition, compounds with highly localized orbitals (*e.g.*, f-electron systems) remain difficult to model with high-throughput methods. Another potential issue, which we did not investigate here, is faithful interpolation of the band structure at band crossings in the Boltztrap code.[16] Recently, a new code based on Wannier functions, BoltzWann, was developed to provide greater fidelity of interpolated band structures and better treatment of band crossings.[93] However, currently BoltzWann is difficult to employ within a high-throughput framework because one must determine the appropriate parameters to yield a good interpolation.

One particularly difficult problem is calculating thermal conductivity in a high-throughput mode. We evaluated the accuracy of two models, one by Clarke and another by Cahill and Pohl, to predict a minimum thermal conductivity. We observed that these models correlate similarly well with experimentally measured minimum thermal conductivities. In recent work, we demonstrated that it is possible to calculate the necessary moduli on a large scale,[77] making these models capable of mid- to high-throughput. We note that these models are not the only method by which lattice thermal conductivity can be estimated. Toberer *et al.*[73] developed a model to estimate thermal conductivity for single crystals using materials parameters that are accessible from DFT calculations. Subsequently, Yan *et al.*[46] demonstrated that this method could be used for high-throughput computational screening of materials by assuming a Grüneisen parameter equal to one and rebalancing some of the model weights. Another method, proposed by Toher *et al.*,[94] uses the GIBBS method to estimate thermal conductivity, although this method is formally only valid for isotropic materials and is likely better suited to high-symmetry structures.[95] The accuracy of the GIBBS-based model was recently evaluated by Madsen *et al.* for rocksalt and zinc-blende structures, and furthermore compared with another model that calculates the Debye temperature from the full phonon band structure.[96] In this study, a mode-averaged squared version of the Grüneisen parameter was demonstrated to be important to obtaining more accurate results.[96] Another possibility for evaluating thermal conductivity is to explicitly calculate the phonon–phonon interactions. This method has traditionally been limited by high computational expense. However, a recent study by Seko *et al.* demonstrated that third-order phonon interactions can now be computed in high-throughput for several hundred compounds when those compounds are relatively simple.[97] New methods might soon make such approaches tractable for more complex compounds and even for higher-order phonon interactions. For example, one recent method introduced by Zhou *et al.* uses compressed sensing to obtain high-order force constants with very few snapshot calculations.[98]

We also used the MP database to assess differences in the maximum power factor for group VIA elements. We observed that the power factor distributions were in the order tellurides $>$ selenides $\sim$ sulfides $>$ oxides. This trend assumes the same doping limits for each compound, and is thus independent of any differences that could arise from systematic differences in the difficulty of doping. Unfortunately, while the power factor increases moving down the periodic table, resource availability decreases (decreased abundance and increased HHI[35]). However, we emphasize that such ''macro'' analyses do not necessarily hold for individual compounds, and a well-designed oxide thermoelectric might be an outlier from such overall trends. We note that many oxide compounds have been investigated in the past,[99–103] with only a few reaching high zT (*e.g.*, p-type $NaCoO_2$ reaching zT of 1.2 at 800 K[104]).

An aspect of thermoelectrics design that represents both an opportunity as well as a difficulty is the diversity of materials chemistries that encompass potentially interesting thermoelectric properties. We presented a first-order clustering analysis to begin the process of grouping together materials with similar features. This type of classification can be interesting in a qualitative sense, *i.e.*, to see how similar physical properties result in similar

thermoelectric performance metrics, and could also be incorporated in the future into quantitative models such as cluster-rank-model,[105] which builds predictive models for separate materials chemistries. For example, one limitation of our high-throughput methods is that the approximation of an electronic relaxation time and thermal conductivity is held constant even across diverse chemistries. In the future, clustering analyses could also present a way forward to rationally and automatically apply more targeted approximations for different regions of chemical space.

V. Conclusion

Leveraging the Materials Project database, we have developed a comprehensive database of thermoelectric properties for solid-state materials and assessed its accuracy against available experimental data. We observed good agreement between computational and experimental Seebeck coefficients; however, the evaluation of power factor in high-throughput is less accurate. Thermal conductivity models from Clarke and from Cahill and Pohl exhibit similar qualitative agreement with experiment. Our data set indicates that designing oxide thermoelectrics will be very challenging, and that overall thermoelectrics design involves several difficult trade-offs. The data set identifies the correlations between some crystal structures and high electronic band valley degeneracy, which can help guide search for new thermoelectric materials. Combined with empirical models (*e.g.*, lattice thermal conductivity estimation) and advanced data mining techniques, such high-throughput investigations might uncover new opportunities for the design of new thermoelectric materials.

Acknowledgements

This work was intellectually led by the U.S. Department of Energy, Office of Basic Energy Sciences, Early Career Research Program. Additional funding was provided by the Materials Project, which is supported by the Department of Energy Basic Energy Sciences program under Grant No. EDCBEE, DOE Contract DE-AC02-05CH11231. J-HP and MAW are grateful to the NSERC CREATE DREAMS (Dalhousie University Research in Energy, Advanced Materials and Sustainability) for funding. U.A. acknowledges the financial assistance of The Scientific and Technological Research Council of Turkey. G. H. acknowledges the F. R. S.-FNRS and the European Union Marie Curie Career Integration (CIG) grant HTforTCOs PCIG11-GA-2012-321988 for financial support. 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.

References

1 J. R. Sootsman, D. Y. Chung and M. G. Kanatzidis, *Angew. Chem., Int. Ed.*, 2009, **48**, 8616–8639.
2 H. Wang, Y. Pei, A. D. Lalonde and G. J. Snyder, *Thermoelectric Nanomaterials*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, vol. 182.
3 G. J. Snyder and E. S. Toberer, *Nat. Mater.*, 2008, **7**, 105–114.
4 H. Zhu, W. Sun, R. Armiento, P. Lazic and G. Ceder, *Appl. Phys. Lett.*, 2014, **104**, 082107.
5 Y. Zhang, E. Skoug, J. Cain, V. Ozoliņ, D. Morelli and C. Wolverton, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **85**, 045306.
6 K. P. Ong, D. J. Singh and P. Wu, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 115110.
7 S. Bhattacharya and G. K. H. Madsen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **92**, 085205.
8 C. Bera, S. Jacob, I. Opahle, N. S. H. Gunda, R. Chmielowski, G. Dennler and G. K. H. Madsen, *Phys. Chem. Chem. Phys.*, 2014, **16**, 19894–19899.
9 G. K. H. Madsen, *J. Am. Chem. Soc.*, 2006, **128**, 12140–12146.
10 S. Wang, Z. Wang, W. Setyawan, N. Mingo and S. Curtarolo, *Phys. Rev. X*, 2011, **1**, 021012.
11 J. Carrete, W. Li, N. Mingo, S. Wang and S. Curtarolo, *Phys. Rev. X*, 2014, **4**, 011019.
12 G. Joshi, R. He, M. Engber, G. Samsonidze, T. Pantha, E. Dahal, K. Dahal, J. Yang, Y. Lan, B. Kozinsky and Z. Ren, *Energy Environ. Sci.*, 2014, **7**, 4070–4076.
13 P. Gorai, P. Parilla, E. S. Toberer and V. Stevanović, *Chem. Mater.*, 2015, **27**, 6213–6221.
14 P. Gorai, D. Gao, B. Ortiz, S. Miller, S. A. Barnett, T. Mason, Q. Lv, V. Stevanović and E. S. Toberer, *Comput. Mater. Sci.*, 2016, **112**, 368–376.
15 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 and G. Ceder, *J. Mater. Chem. C*, 2015, **3**, 10554–10565.
16 G. K. H. Madsen and D. J. Singh, *Comput. Phys. Commun.*, 2006, **175**, 67–71.
17 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2013, **68**, 314–319.
18 A. Jain, S. P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter and K. A. Persson, *Concurr. Comput. Pract. Exp.*, 2015, **27**, 5037–5059.
19 G. Kresse, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169–11186.
20 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
21 G. Kresse and D. Joubert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 1758–1775.
22 W. Setyawan and S. Curtarolo, *Comput. Mater. Sci.*, 2010, **49**, 299–312.
23 https://www.materialsproject.org/wiki/index.php/GGA+U_calculations.
24 D. J. Singh, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **81**, 1–6.
25 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
26 A. Belsky, M. Hellenbrandt, V. L. Karen and P. Luksch, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 364–369.

27 G. Bergerhoff, R. Hundt, R. Sievers and I. Brown, *J. Chem. Inf. Comput. Sci.*, 1983, **23**, 66–69.
28 A. Jain, G. Hautier, C. J. Moore, S. Ping Ong, C. C. Fischer, T. Mueller, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2011, **50**, 2295–2310.
29 J. Saal, S. Kirklin, M. Aykol, B. Meredig and C. Wolverton, *JOM*, 2013, **65**, 1501–1509.
30 S. Hébert and A. Maignan, *Functional Oxides*, John Wiley & Sons, Ltd, 2010, pp. 203–255.
31 L. Bocher, M. H. Aguirre, D. Logvinovich, A. Shkabko, R. Robert, M. Trottmann and A. Weidenkaff, *Inorg. Chem.*, 2008, **47**, 8077–8085.
32 M. Chan and G. Ceder, *Phys. Rev. Lett.*, 2010, **105**, 196403.
33 W. Setyawan, R. M. Gaume, S. Lam, R. S. Feigelson and S. Curtarolo, *ACS Comb. Sci.*, 2011, **13**, 382–390.
34 A. O. Hirschman, *Am. Econ. Rev.*, 1964, **54**, 761.
35 M. W. Gaultois, T. D. Sparks, C. K. H. Borg, R. Seshadri, W. D. Bonificio and D. R. Clarke, *Chem. Mater.*, 2013, **25**, 2911–2920.
36 U.S. Department of Justice and the Federal Trade Commission, *Horizontal Merger Guidelines*, 2010.
37 J. P. Heremans, V. Jovovic, E. S. Toberer, A. Saramat, K. Kurosaki, A. Charoenphakdee, S. Yamanaka and G. J. Snyder, *Science*, 2008, **321**, 1457–1461.
38 L. D. Hicks and M. S. Dresselhaus, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 727–731.
39 F. J. DiSalvo, *Science*, 1999, **285**, 703–706.
40 M. O. Kee, N. E. Brese and M. Okeeffe, *J. Am. Chem. Soc.*, 1991, **113**, 3226–3229.
41 Y. Pei, X. Shi, A. LaLonde, H. Wang, L. Chen and G. J. Snyder, *Nature*, 2011, **473**, 66–69.
42 Z. Tian, S. Lee and G. Chen, *J. Heat Transfer*, 2013, **135**, 61605.
43 A. Zunger, *Appl. Phys. Lett.*, 2003, **83**, 57–59.
44 W. Koshibae, K. Tsutsui and S. Maekawa, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2000, **62**, 6869–6872.
45 Y. Pei, H. Wang and G. J. Snyder, *Adv. Mater.*, 2012, **24**, 6124.
46 J. Yan, P. Gorai, B. Ortiz, S. Miller, S. A. Barnett, T. Mason, V. Stevanovic, E. S. Toberer, V. Stevanović and E. S. Toberer, *Energy Environ. Sci.*, 2015, **8**, 983–994.
47 R. Grosse-Kunstleve, *Acta Crystallogr.*, 1999, **55**, 383–395.
48 R. W. Grosse-Kunstleve, N. K. Sauter, N. W. Moriarty and P. D. Adams, *J. Appl. Crystallogr.*, 2002, **35**, 126–136.
49 C. Fu, S. Bai, Y. Liu, Y. Tang, L. Chen, X. Zhao and T. Zhu, *Nat. Commun.*, 2015, **6**, 8144.
50 E. Rausch, B. Balke, J. M. Stahlhofen, S. Ouardi, U. Burkhardt and C. Felser, *J. Mater. Chem. C*, 2015, **3**, 10409–10414.
51 X. Gonze and C. Lee, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1997, **55**, 10355–10367.
52 Z. H. Levine and D. C. Allan, *Phys. Rev. Lett.*, 1989, **63**, 1719–1722.
53 S. Kim, B. Wiendlocha, H. Jin, J. Tobola and J. P. Heremans, *J. Appl. Phys.*, 2014, **116**, 153706.
54 F. Vazquez, R. Forman and M. Cardona, *Phys. Rev.*, 1968, **176**, 905–908.
55 R. Blunt, H. Frederikse and W. Hosler, *Phys. Rev.*, 1955, **571**, 663–666.
56 A. May, E. Toberer, A. Saramat and G. Snyder, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2009, **80**, 125205.
57 T.-I. Jeon and D. Grischkowsky, *Phys. Rev. Lett.*, 1997, **78**, 1106–1109.
58 G. Samsonidze, B. Kozinsky, 2015, arXiv, 1511.08115.
59 U. Aydemir, A. Zevalkink, A. Ormeci, Z. M. Gibbs, S. Bux and G. J. Snyder, *Chem. Mater.*, 2015, **27**, 1622–1630.
60 S. I. Johnson, A. Zevalkink and G. J. Snyder, *J. Mater. Chem. A*, 2013, **1**, 4244.
61 S. C. Liufu, L. D. Chen, Q. Yao and C. F. Wang, *Appl. Phys. Lett.*, 2007, **90**, 112106.
62 G. Hautier, A. Miglio, D. Waroquiers, G.-M. Rignanese and X. Gonze, *Chem. Mater.*, 2014, **26**, 5447–5458.
63 G. Hautier, A. Miglio, G. Ceder, G.-M. Rignanese and X. Gonze, *Nat. Commun.*, 2013, **4**, 2292.
64 C. Persson and S. Mirbt, *Braz. J. Phys.*, 2006, **36**, 286–290.
65 F. Tran and P. Blaha, *Phys. Rev. Lett.*, 2009, **102**, 226401.
66 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **18**, 8207.
67 H. Xie, H. Wang, C. Fu, Y. Liu, G. J. Snyder, X. Zhao and T. Zhu, *Sci. Rep.*, 2014, **4**, 6888.
68 A. Bulusu and D. G. Walker, *Superlattices Microstruct.*, 2008, **44**, 1–36.
69 T. J. Scheidemantel, C. Ambrosch-Draxl, T. Thonhauser, J. V. Badding and J. O. Sofo, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2003, **68**, 125210.
70 W. W. Schulz, P. B. Allen and N. Trivedi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1992, **45**, 10886–10890.
71 Z. Tian, J. Garg, K. Esfarjani, T. Shiga, J. Shiomi and G. Chen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **85**, 184303.
72 D. A. Broido, M. Malorny, G. Birner, N. Mingo and D. A. Stewart, *Appl. Phys. Lett.*, 2007, **91**, 2013–2016.
73 E. S. Toberer, A. Zevalkink and G. J. Snyder, *J. Mater. Chem.*, 2011, **21**, 15843.
74 D. Clarke, *Surf. Coat. Technol.*, 2003, **163–164**, 67–74.
75 D. G. Cahill and R. O. Pohl, *Annu. Rev. Phys. Chem.*, 1988, **39**, 93–121.
76 D. G. Cahill, P. V. Braun, G. Chen, D. R. Clarke, S. Fan, K. E. Goodson, P. Keblinski, W. P. King, G. D. Mahan, A. Majumdar, H. J. Maris, S. R. Phillpot, E. Pop and L. Shi, *Appl. Phys. Rev.*, 2014, **1**, 011305.
77 M. de Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. K. Ande, S. Van Der Zwaag, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson and M. Asta, *Sci. Data*, 2015, **2**, 150009.
78 R. Hill, *Proc. Phys. Soc., London, Sect. A*, 2002, **65**, 349–354.
79 C. Chiritescu, D. G. Cahill, N. Nguyen, D. Johnson, A. Bodapati, P. Keblinski and P. Zschack, *Science*, 2007, **315**, 351–353.
80 J.-H. Pohls, M. B. Johnson and M. A. White, *Phys. Chem. Chem. Phys.*, 2015, **18**, 1185–1190.
81 M. Ester, H. P. Kriegel, J. Sander and X. Xu, *Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226–231.
82 J. A. Hartigan and M. A. Wong, *Appl. Stat.*, 1979, **28**, 100.
83 L. Van Der Maaten and G. Hinton, *J. Mach. Learn. Res.*, 2008, **9**, 2579–2605.

84 L.-D. Zhao, S.-H. Lo, Y. Zhang, H. Sun, G. Tan, C. Uher, C. Wolverton, V. P. Dravid and M. G. Kanatzidis, *Nature*, 2014, **508**, 373–377.
85 X. Lu, D. T. Morelli, Y. Xia, F. Zhou, V. Ozolins, H. Chi, X. Zhou and C. Uher, *Adv. Energy Mater.*, 2013, **3**, 342–348.
86 K. Suekuni, F. S. Kim, H. Nishiate, M. Ohta, H. I. Tanaka and T. Takabatake, *Appl. Phys. Lett.*, 2014, **105**, 132107.
87 Y. He, T. Day, T. Zhang, H. Liu, X. Shi, L. Chen and G. J. Snyder, *Adv. Mater.*, 2014, **26**, 3974–3978.
88 A. P. Gonçalves and C. Godart, *Eur. Phys. J. B*, 2014, **87**, 42.
89 H. Wang, W. Chu and H. Jin, *Comput. Mater. Sci.*, 2012, **60**, 224–230.
90 J. Yang, H. Li, T. Wu, W. Zhang, L. Chen and J. Yang, *Adv. Funct. Mater.*, 2008, **18**, 2880–2888.
91 J. Heyd, J. E. Peralta, G. E. Scuseria and R. L. Martin, *J. Chem. Phys.*, 2005, **123**, 174101.
92 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2006, **124**, 219906.
93 G. Pizzi, D. Volja, B. Kozinsky, M. Fornari and N. Marzari, *Comput. Phys. Commun.*, 2014, **185**, 422–429.
94 C. Toher, J. J. Plata, O. Levy, M. de Jong, M. Asta, M. B. Nardelli and S. Curtarolo, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **90**, 174107.
95 M. A. Blanco, E. Francisco and V. Luaña, *Comput. Phys. Commun.*, 2004, **158**, 57–72.
96 G. K. H. Madsen and G. Santi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2005, **72**, 220301.
97 A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput and I. Tanaka, *Phys. Rev. Lett.*, 2015, **115**, 205901.
98 F. Zhou, W. Nielson, Y. Xia and V. Ozoliņš, *Phys. Rev. Lett.*, 2014, **113**, 185501.
99 S. Ohta, T. Nomura, H. Ohta, M. Hirano and H. Hosono, *Appl. Phys. Lett.*, 2005, **87**, 092108.
100 S. Lee, R. H. T. Wilke, S. Trolier-McKinstry, S. Zhang and C. A. Randall, *Appl. Phys. Lett.*, 2010, **96**, 031910.
101 R. Funahashi, I. Matsubara, H. Ikuta, T. Takeuchi, U. Mizutani and S. Sodeoka, *Jpn. J. Appl. Phys.*, 2000, **39**, L1127.
102 R. Funahashi and M. Shikano, *Appl. Phys. Lett.*, 2002, **81**, 1459–1461.
103 M. Shikano and R. Funahashi, *Appl. Phys. Lett.*, 2003, **82**, 1851–1853.
104 K. Fujita, T. Mochida and K. Nakamura, *Jpn. J. Appl. Phys.*, 2001, **40**, 4644.
105 B. Meredig and C. Wolverton, *Chem. Mater.*, 2014, **26**, 1985–1991.

