TE Design Lab: A Virtual Laboratory for Thermoelectric Material Design

Prashun Gorai^{a,b}, Duanfeng Gao^c, Brenden Ortiz^a, Sam Miller^d, Scott A. Barnett^d, Thomas Mason^d,
Qin Lv^c, Vladan Stevanović^{a,b}, Eric S. Toberer^{a,b}

^aColorado School of Mines, Golden, CO 80401 USA

^bNational Renewable Energy Laboratory, Golden, CO 80401, USA

^cUniversity of Colorado Boulder, Boulder, CO 80309, USA

^dNorthwestern University, Evanston, IL 60208, USA

Abstract

The discovery of advanced thermoelectric materials is the key bottleneck limiting the commercialization of solid-state technology for waste heat recovery and compression-free refrigeration. Computationally-driven approaches can accelerate the discovery of new thermoelectric materials and provide insights into the underlying structure-property relations that govern thermoelectric performance. We present TE Design Lab (www.tedesignlab.org), a thermoelectrics-focused virtual laboratory that contains calculated thermoelectric properties as well as performance rankings based on a metric [29] that combines *ab-initio* calculations and modeled electron and phonon transport to offer a reliable assessment of the intrinsic material properties that govern the thermoelectric figure of merit zT . Another useful component of TE Design Lab is the suite of interactive web-based tools that enable users to mine the raw data and unearth new structure-property relations. Examples that illustrate this utility are presented. With the goal of establishing a close partnership between experiments and computations, TE Design Lab also offers resources to analyze raw experimental thermoelectric data and contribute them to the open access database.

Keywords: Thermoelectrics, High-throughput, Materials Genome Initiative, TE Design Lab

1. Introduction

Thermoelectric (TE) materials, which have the ability to harness the flow of heat to produce electricity, can potentially revolutionize the energy sector by recovering waste heat [5], producing solar-derived electricity [2], and providing compression-free refrigeration [11]. Despite the need for new thermoelectric materials, experimental efforts to date have been slow in identifying new materials with good thermoelectric performance because of the inherent challenges associated with optimization of contradicted charge carrier and thermal transport prop-

erties as well as the lack of robust chemical design strategies. Experimental efforts driven by intuition and serendipity have produced promising materials [5, 32, 13, 28, 31, 27, 23, 30, 24, 25, 33] but rational design and discovery of new thermoelectric materials are still in their early stages. Identification of functional relations between real (structural) and reciprocal (transport) space properties and thermoelectric performance is a key aspects in rationally designing thermoelectric materials.

A data-driven approach, wherein large chemical spaces containing of the order of 10^4 compounds are ranked for their thermoelectric performance based on simple yet reliable metrics, in conjunction with data visualization and data mining tools can reveal

Email address: etoberer@mines.edu (Eric S. Toberer)

Figure 1: Only tens of materials (33 shown) among the $\sim 40,000$ known stoichiometric and crystalline metal-non metal solids (gray circles) from the ICSD have so far been considered as thermoelectric materials (colored squares). Known thermoelectric materials span a diverse chemical space represented by the standard deviation in electronegativity and average atomic mass, with the majority of them being relatively “heavy” and covalent.

relations between tunable real and reciprocal space parameters and thermoelectric performance. Such an approach is also powerful in the discovery of new materials. Figure 1 shows a subset ($\sim 40,000$) of the known stoichiometric and crystalline structures from the Inorganic Crystal Structure Database (ICSD) [3], with known thermoelectric materials indicated by colored squares. It is evident from Fig. 1 that known thermoelectric materials are found over a large span of the known chemical space that have emerged through empirical efforts rather than rational design. An even larger chemical space comprising currently unknown compounds and metastable polymorphs remains to be systematically explored for new thermoelectric materials. Several materials databases are currently available, the Materials Project [14], AFLOWLIB [8], and OQMD [18] being the most prominent ones. In general, these open

access materials databases provide computed structure and electronic properties of mostly known materials in the ICSD, with some offering modules geared towards specific applications such as energy storage [14], radiation detection [8] and thermoelectricity [8]. However, functionalities to identify quantitative relations between relevant properties and material functionalities is currently missing, especially for thermoelectric materials.

In a notable example of a thermoelectrics-focused database, experimental thermoelectric data from over 100 publications have been curated and made available to the public along with tools to visualize and analyze the data [10, 20]. Measured properties (e.g. electrical resistivity, Seebeck coefficient, thermal conductivity, zT), a limited set of structural parameters (e.g. unit cell volume, atoms per cell), and indices related to the crustal abundance and production of the constituent elements can be interactively plotted to reveal trends and structure-property relations. While still useful, developing design strategies from experimental data *alone* has its challenges: (1) small data sets lead to predictions within the same class of materials [19, 1] and by design, do not incorporate the structure-property relations unique to materials outside the dataset, (2) more explored chemical systems skew trends by introducing historical biases, and (3) the strong dependence of thermoelectric properties on carrier concentration leads to ineffective descriptors. A computationally-driven approach can address these limitations.

In this paper, we describe TE Design Lab, a thermoelectric-focused virtual laboratory that offers open access computational and experimental data and tools to accelerate the discovery of new thermoelectric materials and enable the identification of useful structure-property correlations. The development of design principles from structure-property relations adds a new dimension to the computationally-driven Materials Genome Initiative (MGI) approach. We have recently demonstrated a computational methodology [29] to gauge the potential thermoelectric performance of materials. Through TE Design Lab, we provide open access to our raw data as well as ranking of materials based on our descriptor [29]. As the name suggests, TE Design Lab is a laboratory

where users can utilize interactive web-based tools to test design hypotheses on a large body of computational data. With the goal of establishing a feedback loop between computations and experiments, TE Design Lab also offers resources for users to analyze their experimental thermoelectric data, compare them with computational predictions and contribute them to the open access experimental databank.

2. User-Driven Resources on TE Design Lab

As a user-focused portal, TE Design Lab offers tools that enables users to screen for thermoelectric materials and identify structure-property relations. Specifically,

- A metric that quantifies the potential of materials for thermoelectric performance.
- Sufficiently accurate computational data, augmented by experimental data, available for data exploration and mining that enables the discovery of new structure-property relations.
- Visualization and analysis tools to explore and mine the data repository.
- TE Design Lab also offers the flexibility to export data, which can be further analyzed by the users with their own offline tools.
- TE Design Lab also allows users to contribute their own data to the open access databank.

3. Metric, Methodology, and Data Handling

In this section, we describe the metric we have developed to assess the thermoelectric performance of materials. The computational methodology for determining the performance metric and the high-throughput infrastructure for automated handling of large numbers of first-principles calculations are explained. Finally, the mechanics of the cloud-based repository is briefly discussed.

3.1. Assessment of Thermoelectric Performance

The performance of thermoelectric materials is typically quantified by a figure of merit, zT :

$$zT = \frac{\alpha^2 \sigma T}{(\kappa_L + \kappa_e)}, \quad (1)$$

where α is the Seebeck coefficient, σ the electrical conductivity, and κ_L and κ_e the lattice and electronic components of thermal conductivity κ . In the relaxation time approximation [7], the Boltzmann transport equations for α , σ and κ_e yield an alternative expression for zT :

$$zT = \frac{u\beta}{(v\beta + 1)}, \quad (2)$$

where u and v depend on charge carrier chemical potential (η) and scattering mechanisms. β is a material-dependent parameter. In Eq. (2), maximizing zT requires simultaneously optimizing the intrinsic material properties to tune β and doping to tune u and v . The parameter β depends on the electronic structure, phonon properties and temperature through the following relationship:

$$\beta = \frac{2e}{\hbar^3} \left(\frac{k_B}{e} \right)^2 \left(\frac{k_B}{2\pi} \right)^{3/2} \frac{\mu_0 m_{DOS}^*}{\kappa_L} T^{5/2} \quad (3)$$

where μ_0 is the intrinsic charge carrier mobility, m_{DOS}^* the density of states (DOS) effective mass, κ_L the lattice thermal conductivity, k_B the Boltzmann constant, and e the electronic charge. For materials that are optimally doped to maximize zT , we find that β calculated from measured room temperature μ_0 , m_{DOS}^* and κ_L is a good metric of the maximal zT [29].

Direct calculation of transport properties (μ_0 , κ_L) entering Eq. 3 is challenging, often unreliable and computationally expensive. To facilitate accelerated assessment of the thermoelectric performance, we have recently developed semi-empirical models for μ_0 and κ_L by combining first-principles calculations and measured room temperature μ_0 and κ_L for a range of materials [29]. Utilizing these semi-empirical models, β (now referred to as β_{SE}) can be evaluated from simple first-principles calculations in a high-throughput fashion.

To determine μ_0 and κ_L from the semi-empirical models (see Ref. [29]), the conduction (n -type) and valence (p -type) band degeneracies (N_b) and density of states (DOS) effective masses (m_{DOS}^*) and the bulk modulus (B) need to be determined. Other parameters that enter the semi-empirical models are readily accessible from the same calculations. For each material, two values of β_{SE} are calculated depending on the assumed doping type. β_{SE} is used to rank and screen materials for their thermoelectric performance. For example, β_{SE} for n -doped PbTe is 15.3 and for p -doped 13.4, which can be used as a reference β_{SE} value for gauging the performance of other materials relative to a known thermoelectric. We have demonstrated the efficacy of β_{SE} in identifying known thermoelectrics in Ref. 29.

3.2. Computational Methodology

We describe here our current implementation of the computational methodology based on density functional theory (DFT) to compute N_b , m_{DOS}^* , and B . DFT calculations are performed with the VASP code [15], with the exchange correlation in the Perdew Burke Ernzerhof (PBE) functional form within the projector augmented wave (PAW) formalism [6].

For structure relaxations, a procedure similar to that employed in Ref. 21 is used, with plane wave cut-offs of 340-500 eV. An on-site correction in the form of Hubbard U in the rotationally invariant form, introduced by Dudarev and coworkers [9], is applied for transition metals [21]. In case of compounds containing transition metals, we perform a limited search for the magnetic ground state by enumerating over magnetic configurations on a primitive unit cell. To determine the lowest energy magnetic state, the volume and ionic relaxations are performed with the magnetic moments initialized in a limited number of configurations. Because the number of initial configurations scale as 2^N with the number of magnetic atoms (N), we limit the total number of calculations. We consider up to 32 possible magnetic configurations for each compound, i.e., for compounds with $N \leq 6$, all configurations are considered and for $N \geq 6$, 32 magnetic configurations are randomly chosen (including ferromagnetic). Higher-energy magnetic configurations are not included in the following analysis.

TE Design Lab contains calculated transport properties for materials that can be found in the ICSD. Each entry in TE Design Lab specifies the associated ICSD number. The thermodynamic stability of ICSD structures has been already calculated and the data made available by several open-access databases (Materials Project, AFLOW, OQMD), including our sister database NRELMatDB (materials.nrel.gov) from where we inherit the relaxed structures.

Calculation of m_{DOS}^* and N_b is undertaken on a dense $\mathbf{k}$ -point grid with a fixed number of $\mathbf{k}$ -points per atom, as determined by the equation $N_{atoms} \times N_{kpts} = 8000$, where N_{atoms} is the number of atoms in the primitive cell and N_{kpts} the number of $\mathbf{k}$ -points. This $\mathbf{k}$ -point density is equivalent to a $14 \times 14 \times 14$ $\mathbf{k}$ -mesh for diamond Si, which provides sufficiently converged electronic structures. B is calculated by fitting the Birch-Murnaghan equation of state [4] to a set of total energies computed at different volumes. m_{DOS}^* is determined from the electronic DOS within a parabolic band approximation, such that the parabolic band reproduces the same number of states as the computed DOS within a 100 meV (adjustable) energy window from the relevant band edge. When the Fermi level lies close to the band edge, as in many thermoelectrics, the energy window is a measure of the extent to which the bands are populated with carriers at a given temperature. Note that β_{SE} does not explicitly depend on the band gap; the optimal working temperature of thermoelectric materials does as well as its dopability. In our analysis, we include all materials with DFT-calculated band gaps larger than 10 meV because of the well-known DFT underestimation of band gaps.

3.3. High-Throughput Infrastructure

The high-throughput DFT calculations are handled within PyLada [17], a modular Python framework to control first-principles calculations, which not only creates the required VASP files and storage folder structure but also interfaces with the job queuing system on parallel processing supercomputing clusters. PyLada offers a variety of useful in-built tools for constructing and manipulating periodic crystal structures, for manipulating functionals

Figure 2: The landing page (“Materials”) for TE Design Lab website offers a search field to query the material database. The search field can interpret Boolean operations as indicated in the tool tip displayed when hovering the mouse over the field. The default search field also displays the current number of materials in the database.

and extracting their output, and for checking the results from thousands of calculations simultaneously.

We use Google Cloud SQL, a MySQL database hosted by the Google Cloud Platform, to manage all the thermoelectric data that are available through TE Design Lab. MySQL is an open source relational database that has been widely used in many domains. It supports various types of data queries with high efficiency, as well as flexible data insertion. Furthermore, by using Google Cloud SQL, all data are replicated across multiple distributed zones thus allowing for greater availability and durability. Therefore, Google Cloud SQL is suitable for storing and managing the data that has been (and will be) generated and added to TE Design Lab.

3.4. Structure-Property Data

The ICSD contains duplicate structures of many compounds. Therefore, we determine the equivalence between different crystal structures of a given composition by matching the space group and Wyckoff position multiplicities. For the computed unique struc-

tures, TE Design Lab offers a repository of: (1) structural data (e.g., lattice constants, space group, average coordination number), which are extracted from the ICSD Crystallographic Information Files (CIFs) or obtained by processing the extracted data, (2) relevant reciprocal space (e.g. N_b , m_{DOS}^* , m_b^*) and real space (e.g. B , d) parameters, (3) thermoelectric properties such as μ_0 , κ_L calculated from reciprocal and real space parameters using the semi-empirical models, and (4) value of β_{SE} for both doping (n - and p -) types, which can be used to rank and screen materials. In total, ~ 50 different properties are available. As with our continued efforts in improving our metric and methodology, we are also regularly expanding the parameter set that is available on TE Design Lab.

4. TE Design Lab.ORG

In this section, we describe the mechanics of TE Design Lab and the various tools and functionalities available to users, along with labeled snapshots of the webpages. The TE Design Lab website

(www.tedesignlab.org) has four main pages (Figure 2), namely: Materials, Visualization, Resources, and Contribute.

4.1. “Materials”: Search Database and Download

The landing page of the website presents the user a search field to query the database (Figure 2). The total number of materials available in the database is displayed as the default text in the search field, which is updated in real time. Typically, several hundred calculations are added on a monthly basis. To display the entire database, the user can hit the **Search** button without entering a search condition.

The search field is capable of interpreting Boolean expressions (hover the mouse over the search field for Boolean syntax help). For example, a Boolean search expression “**Se Pb -Ga**” will return all Se-containing materials that also contain Pb but not Ga. Similarly, an expression “**Pb or Se**” will return all materials that contain either Pb or Se. The search can be further customized by restricting the number of elements (or range), specifying crystal structures (by specifying a space group number or range), or choosing a crystal system. We also have a limited database of experimental thermoelectric properties that can be queried by choosing the appropriate “**Data Type**” from the drop-down menu.

Once a search condition has been established and queried, the results are displayed in the table below. The data columns can be displayed by choosing the appropriate option from the drop down menu. All the search results or a subset created by using the checkboxes can be downloaded in a comma separated values (csv) file format. A user interested in even more details, beyond the data columns displayed in the table, can click the material **Formula**, which is hyperlinked to a corresponding page containing tabulated data for several structural and computed properties.

4.2. “Visualization”: Tools for Data Visualization

TEDesignLab contains web-based visualization tools, which are a key component in materials screening and identification of structure-property relations. Currently, three visualization tools are implemented, which are explained with examples in the following

sections. For each of these visualization schemes, the user can (i) search material datasets from the repository, (ii) display them on multi-dimensional plots by choosing the dimensions from a set of ~ 30 structural and thermoelectric parameters, (iii) customize the plots, (iv) compare trends between data subsets, and (v) download the plots as images or the associated data as csv files for plotting with the user’s preferred tools or for further analysis.

4.2.1. Plot 4 dimensions

The first visualization scheme allows the user to create a 4-dimensional plot (abscissa, ordinate, marker radius, and heatmap) for a *single* dataset, as shown in Figure 3. This form of visualization is useful when searching for correlations in material properties in specific chemical spaces. The search field below the plot functions the same way as the one on the **Materials** page. Searching “**Se**” returns all Se-containing compounds present in the database. The parameters to display can be chosen from the **Parameters** list on the left. In Figure 3, the abscissa (x-axis) is lattice thermal conductivity (κ_L) and the ordinate (y-axis), hole mobility ($\mu_{0,h}$). The marker radius represents the value of the metric, β_{SE} for valence band transport $\beta(p)$ and the heatmap the valence band degeneracy ($N_{b,VB\bar{M}}$). The plot can be customized by selecting the appropriate options in the left panel. Hovering the mouse over a displayed data point brings up a tool tip that contains the compound name, space group and the value of the 4 parameters chosen. For example, SnSe (space group 63) is shown in Figure 3. Dynamic zooming is another useful tool that allows the user to visualize a subset of the displayed data. The data displayed in the plot can be downloaded either as an image file or in tabular form as a csv file.

4.2.2. Plot 2 or more datasets in 3 dimensions

The user may want to compare the properties of *two or more* datasets that represent different chemical spaces, structure types or other criteria. This form of visualization is helpful in discovering structure-property relations unique to each material dataset rather than across all materials. The user can plot 3-dimensional data, chosen from the **Parameter**

Figure 3: One of the visualization options available on TE Design Lab allows the user to plot 4 dimensions - abscissa, ordinate, marker radius, and heatmap, with the ability to plot on logarithmic axes. There are currently 30 structural and calculated parameters available to plot.

list, for two or more datasets. An example is shown in Figure 4, where the hole mobility ($\mu_{0,h}$) versus lattice thermal conductivity (κ_L) is plotted for Se-, S- and Te-containing compounds in the database. The third dimension, the marker radius can be scaled or made uniform (Figure 4) in the left panel. The search field below the plot can be used to query the database. Once a search has been performed, the dataset can be added to the current plot. This action can be performed multiple times to visualize two or more datasets. The corresponding legends are displayed in the floating box on the right. The displayed datasets can be deleted or hidden using the options in the floating legend box. Legend names and colors can also be customized by the user. This plot also can be dynamically zoomed and the data downloaded as an image or csv datafile.

4.2.3. Plot n dimensions on parallel coordinates

One of the tools to visualize correlations across an n -dimensional parameter space is the parallel coordinate plot (PCP), which is implemented on TE Design Lab. The PCP comprises of n parallel axes; a data point is represented by a line connecting the corresponding points on each of the n parallel axes, as shown for an example dataset in Figure 5. A set of lines bunched together on m parallel axes suggests that those m parameters are correlated, either through a causal or functional relation. The main advantage of this visualization scheme is that, in principle, it can be utilized to identify correlations across several dimensions simultaneously. Figure 5 illustrates an example of a PCP, for S-containing compounds. The search field below the plot can be used to create a set of criteria that determines the dataset displayed on the PCP, similar to the other visualization schemes. In addition, several functionalities to

Figure 4: The second visualization option available on TE DesignLab allows the user to plot more than one dataset simultaneously. The plot, legend names and legend colors are all customizable offering a great degree of control over the datasets displayed.

customize and analyze the PCP are available: (1) the axis columns to display can be selected from the parameter list below the plot, (2) the display of columns can be toggled (show/hide) and optionally converted to logarithmic scale, (3) data lines that fall within a subrange of the full-scale values on each axis can be highlighted, for one or more axes simultaneously as shown in (Figure 5), and the selection bar can then be dynamically slid on the axis and the subrange resized without creating a new selection, and (4) the order of the axis columns can be changed by dragging the column title to an appropriate position.

4.3. “Resources”: Recipes and Tools for Data Extraction and Analysis

In addition to the materials search and visualization tools, users are provided with recipes and tools to reproduce our computational data and extract the relevant parameters (e.g., band degeneracy) from their own calculations. Also, tools to analyze experimental thermoelectric data are provided under

the “Resources” tab. The *Mathematica*[®] notebooks available under “Data Analysis” are helpful in obtaining useful derived parameters such as DOS effective mass, by analyzing experimental thermoelectric data. Additional *Mathematica*[®] notebooks will be provided for pedagogical purposes to graphically introduce Boltzmann transport theorem in the context of thermoelectrics and to mathematically understand trends in experimentally-measured transport properties. In the near future, we also intend to add offline tools for performing data clustering analysis that have the capability of revealing material subgroups with commonalities.

4.4. “Contribute”: Submit User Data

To expedite and standardize integration of experimental data into TE DesignLab, we provide online forms and downloadable template files (to be filled out and uploaded by users) as a means for users to contribute data to this database. Contributors are expected to provide the structure information

Figure 5: The parallel coordinate plot is an effective tool for visualization of n -dimensional data on parallel axes (columns). The interactive plot allows users to select and highlight data lines that fall within a subrange of the full-scale axes values for one or more axes simultaneously (gray selection bar). The subrange selection can then be dynamically resized or slid on the axes without having to create a new selection. Reordering the arrangement of the columns allows grouping of columns.

(chemistry, space group, ICSD number or CIF file), room temperature transport properties and if available, the peak zT and the associated temperature. By offering the option to input only room temperature measurements, we ensure that most contributors can submit their data even if zT has not been optimized. We also encourage contributors to provide the DOI(s) of the publication(s) that contain these experimental data so that the contributor(s)/author(s) are appropriately credited when displaying the data on TE Design Lab. Alternatively, a CSV template can be downloaded and filled out offline and submitted through the same portal. Once the experimental data is successfully submitted, we manually examine the data for potential errors and subsequently, allow it to be displayed as part of the searchable main database. At present, we do not offer a similar platform for submission of computational data because of practical challenges associated with: (1) gathering information on the details of the methods

employed, and (2) the volume and format of the submitted data. However, we do provide the option of working directly with the contributor(s) to facilitate computational data submission to TE Design Lab.

5. TE Design Lab.ORG in Action

Three examples that demonstrate how the tools on TE Design Lab can be used for screening of materials and identification of structure-property relations are discussed in this section. Example 1 and 2 pertain to thermoelectric materials. Some of the structural and reciprocal space properties are also relevant to applications beyond thermoelectrics, such as photovoltaics and power electronics. Example 3 illustrates the utility of the database in screening for transparent conductors (TCs), which are used in photovoltaic devices as contact layers.

Figure 6: Plot of β_{SE} as a function of the space group number for (a) n - and (b) p -doped materials. The heatmap represents the corresponding band degeneracies. The dotted line indicates the value of β_{SE} of PbTe for n -type (15.3) and p -type (13.4) doping.

5.1. Screening: New Thermoelectric Materials

One of the primary objectives of TE Design Lab is to screen for promising new thermoelectric materials. The performance metric β_{SE} discussed previously can be used as a screening parameter. Using β_{SE} of PbTe as a reference, we can screen for promising new thermoelectric materials. In this example, we utilize the visualization scheme described in Section 4.2.1 to plot β_{SE} as the ordinate, space group as abscissa, and band degeneracy as the heatmap for n - and p -type materials, as shown in Figures 6(a) and 6(b), respectively. The reference value of β_{SE} for n -

Figure 7: Plot of bulk modulus (B) as a function of unit cell volume (V) on a log-log scale for binary halides (Cl, Br, I), oxides and chalcogenides (S, Se, Te), and pnictides (N, P, As, Sb, Bi) belonging to the space group 225. The power law dependence $B \propto V^{-x}$, where x is a positive number, has been previously identified from empirical data for tetrahedrally-bonded semiconductors with diamond, zincblende and wurtzite structures [16]. Here we see that the dependence is also valid for rocksalt structures derived from a diverse chemical space.

type PbTe (15.3) and for p -type (13.4) are marked by dashed lines on the corresponding plots. The screening procedure correctly identifies many known thermoelectric materials, some of which (PbTe, PbS, PbSe, SnTe, GeTe and In_2Te_5) are labeled in Figures 6(a) and 6(b). Many new thermoelectric materials, including higher-energy polymorphs of known thermoelectric materials, can be reliably screened from large chemical spaces. A few materials predicted to be potentially good n - and p -doped thermoelectric materials are shown in Figures 6(a) and 6(b), respectively. CdSb has been recently demonstrated to be good thermoelectric with a peak zT of 1.3 [26] at 560K, in agreement with our prediction that CdSb is a good n -type thermoelectric material. AgBiTe_2 is predicted to be a good n - and p -type (Figure 6) material. AgBiS_2 , which belongs to the same ternary chalcogenide family as AgBiTe_2 , was shown to exhibit a $zT \sim 1.1$ [22] at 775K when alloyed with SnTe. We expect AgBiTe_2 to demonstrate a similarly high zT .

An observation from Figure 6 is that the conduc-

tion band edges are predominantly singly degenerate as indicated by the overwhelming majority of blue circles ($N_b=1$) in Figure 6a, while valence band exhibits a more diverse spectrum of band degeneracies (Figure 6b). This observation can be understood in the tight-binding picture - in s - p systems, which is represented by a majority of the points in Figure 6, the conduction band is typically composed of cation s orbitals that are singly degenerate at the Γ point while the valence band is composed of anion p orbitals that are three fold degenerate and appear off of Γ , resulting in larger valence N_b . Exceptions to this simplistic picture occur in systems containing ions in a low oxidation state (PbTe, SnSe) or transition metals.

5.2. Structure-Property Relations: Bulk Modulus Dependence on Unit Cell Volume

Figure 7 shows the power law dependence of bulk modulus (B) on the unit cell volume (V) for binary halides (Cl, Br, I), oxides and chalcogenides (S, Se, Te), and pnictides (N, P, As, Sb, Bi) belonging to the space group 225 ($Fm\bar{3}m$). This plot was generated using the visualization discussed in Section 4.2.2. The dotted line in Figure 7 provides a guide to the eye, the slope of which determines the exponent (x) of the power law dependence $B \propto V^{-x}$. The emergence of this structure-property relation within a specific class of materials derived from only *computational data* and visualized using online tools demonstrates the utility of TE Design Lab. In fact, this power law dependence of B on V has been previously established from *empirical data* for tetrahedrally-bonded semiconductors with diamond, zincblende, and wurtzite structures [16]. Figure 7 suggests that the power law relationship between B and V is also applicable to rocksalt structures (space group 225) derived from a large and diverse chemical space of halides, oxides, chalcogenides, and pnictides. This example shows that TE Design Lab can also be used to extend known structure-property relations to larger or different chemical spaces.

Figure 8: Screening for transparent conducting oxides (TCO) based on large charge carrier (electron) mobilities and moderate band gaps. SnO_2 (space group 136), a known n -type TCO is “correctly” identified as a good TCO material. The marker color represents the space group number.

5.3. Beyond Thermoelectrics: Screening for Transparent Conducting Oxides

The utility of TE Design Lab can be extended beyond thermoelectrics, as demonstrated in this example. Transparent conductors (TCs), which are typically used in optoelectronic devices as the window contact layers, should fulfill *at least* two criteria: (a) a large band gap so that the material is transparent and can function as a window for the sunlight to pass through to the underlying active absorber layer, and (b) high charge carrier mobility to be an effective conductor. TCs are majority carrier materials that are heavily doped to make them n - or p -type, similar to thermoelectric materials. A distinguishing feature of thermoelectric materials is that they typically exhibit small rather than large band gaps.

Oxides are one of the most common materials used as n -type TCs because they generally exhibit large band gaps i.e. transparent. However, the electron mobilities are also typically low in oxides. Currently only a few n -type oxides, such as In_2O_3 , ZnO and SnO_2 , are known to offer a balance between these antagonistic properties (optical transparency, high conductivity). The number of p -type TCs are even more limited.

Among the computed properties available on TEDesignLab are charge carrier mobilities (μ_0) and DFT band gaps (E_g). Notwithstanding the common underestimation of band gaps computed within the GGA+U approximation of DFT, a plot of μ_0 vs. E_g for oxide materials can be utilized to screen for potential candidate TC materials. In Figure 8, all known n -type TCOs (space group number in parantheses) such as ZnO (186), SnO₂ (136), In₂O₃ (206), Ga₂O₃ (167) are correctly identified. Several candidate n -type TCOs are also identified (space group number in parantheses): BaSnO₃ (221), CdGeO₃ (62), GeO₂ (205), Zn₂GaInO₅ (194), MnO (225). However, we would like to stress that many oxides are prone to be polaronic rather than band conductors. The calculated mobilities on TEDesignLab assume band conduction, which are typically much larger than polaron hopping mobilities. The band conduction mobilities are useful for *screening* promising candidates. Once screened, detailed follow up calculations may be necessary to determine the possibility of polaronic conduction. The fact that the promising candidates in Figure 8 were also identified through a different high-throughput approach [12] dedicated to the search for n -type TCOs further bolsters our confidence in the calculated property data available on TEDesignLab that could be useful in areas beyond thermoelectrics.

6. Closing Remarks

We expect that our efforts in developing TEDesignLab will not only advance the field of thermoelectric materials design but also create a new paradigm in MGI-inspired open access materials databases by offering data visualization and mining tools for targeted applications. As with any large scale project, the development is an ongoing effort. We are continually improving the tools available on TEDesignLab as well as expanding the materials database and improving our methodologies. We look forward to feedback from the user community to drive the improvements. In addition, we welcome users, both experimentalists and computationalists, to contribute data and collaborate on this open source database.

7. Acknowledgments

The development of TEDesignLab is supported by the National Science Foundation (NSF) under Grants 1334713, 1334351 and 1333335. Computational infrastructure for first-principles calculations has been enabled by the Department of Energy (DOE), through the National Renewable Energy Laboratory (NREL).

- [1] Balachandran, P. V., Theiler, J., Rondinelli, J. M., Lookman, T., 2015. Materials prediction via classification learning. Sci. Rep. 5, 13285. URL <http://dx.doi.org/10.1038/srep13285>
- [2] Baranowski, L., Warren, E., Toberer, E., 2014. High-temperature high-efficiency solar thermoelectric generators. J Electron. Mat. 43 (6), 2348–2355. URL <http://dx.doi.org/10.1007/s11664-014-3063-z>
- [3] Belsky, A., Hellenbrandt, M., Karen, V. L., Luksch, P., 2002. Acta Crystallogr. B 58, 364. URL <http://dx.doi.org/10.1107/S0108768102006948>
- [4] Birch, F., 1952. Elasticity and constitution of the earth’s interior. J. Geophys. Res. 57 (2), 227–286. URL <http://dx.doi.org/10.1029/JZ057i002p00227>
- [5] Biswas, K., He, J., Blum, I. D., Wu, C.-I., Hogan, T. P., Seidman, D. N., Dravid, V. P., Kanatzidis, M. G., 2012. High-performance bulk thermoelectrics with all-scale hierarchical architectures. Nature 489, 414–418. URL <http://dx.doi.org/10.1038/nature11439>
- [6] Blöchl, P. E., Dec 1994. Projector augmented-wave method. Phys. Rev. B 50, 17953–17979. URL <http://dx.doi.org/10.1103/PhysRevB.50.17953>
- [7] Chasmar, R. P., Stratton, R., 1959. The thermoelectric figure of merit and its relation to

- thermoelectric generators. J. Electron. Control 7, 52–72.
URL <http://dx.doi.org/10.1080/00207215908937186>
- [8] Curtarolo, S., Setyawan, W., Wang, S., Xue, J., Yang, K., Taylor, R. H., Nelson, L. J., Hart, G. L., Sanvito, S., Buongiorno-Nardelli, M., Mingo, N., Levy, O., 2012. Aflowlib.org: A distributed materials properties repository from high-throughput ab initio calculations. Comp. Mat. Sci. 58, 227 – 235.
URL <http://dx.doi.org/10.1016/j.commatsci.2012.02.002>
- [9] Dudarev, S. L., Botton, G. A., Savrasov, S. Y., Humphreys, C. J., Sutton, A. P., Jan 1998. Electron-energy-loss spectra and the structural stability of nickel oxide: An lsda+u study. Phys. Rev. B 57, 1505–1509.
URL <http://link.aps.org/doi/10.1103/PhysRevB.57.1505>
- [10] Gaultois, M. W., Sparks, T. D., Borg, C. K. H., Seshadri, R., Bonificio, W. D., Clarke, D. R., 2013. Data-driven review of thermoelectric materials: Performance and resource considerations. Chem. Mat. 25 (15), 2911–2920.
URL <http://dx.doi.org/10.1021/cm400893e>
- [11] Goldsmid, H., 1964. Thermoelectric Refrigeration. Temple Press Books LTD, London.
URL <http://www.springer.com/us/book/9781489957252>
- [12] Hautier, G., Miglio, A., Waroquiers, D., Rignanese, G.-M., Gonze, X., 2014. How does chemistry influence electron effective mass in oxides? a high-throughput computational analysis. Chem. Mat. 26 (19), 5447–5458.
URL <http://dx.doi.org/10.1021/cm404079a>
- [13] Heremans, J. P., Jovic, V., Toberer, E. S., Saramat, A., Kurosaki, K., Charoenphakdee, A., Yamanaka, S., Snyder, G. J., 2008. Enhancement of thermoelectric efficiency in pbte by distortion of the electronic density of states. Science 321, 554–557.
URL <http://dx.doi.org/10.1126/science.1159725>
- [14] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., Persson, K. A., 2013. Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL Materials 1, 011002 1–10.
URL <http://dx.doi.org/10.1063/1.4812323>
- [15] Kresse, G., Furthmüller, J., Oct 1996. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. Phys. Rev. B 54, 11169–11186.
URL <http://dx.doi.org/10.1103/PhysRevB.54.11169>
- [16] Neumann, H., 1987. Bulk modulus–volume–ionicity relationship in tetrahedrally bonded semiconductors. Crystal Res. Tech. 22 (1), 99–104.
URL <http://dx.doi.org/10.1002/crat.2170220122>
- [17] Pylada, ???? A python framework for high-throughput first-principles calculations. <https://github.com/pylada>.
- [18] Saal, J., Kirklin, S., Aykol, M., Meredig, B., Wolverton, C., 2013. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (oqmd). JOM 65, 1501–1509.
URL <http://dx.doi.org/10.1007/s11837-013-0755-4>
- [19] Seko, A., Maekawa, T., Tsuda, K., Tanaka, I., Feb 2014. Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids. Phys. Rev. B 89, 054303.
URL <http://link.aps.org/doi/10.1103/PhysRevB.89.054303>
- [20] Sparks, T. D., Gaultois, M. W., Oliynyk, A., Brgoch, J., Meredig, B., 2015. Data mining our

- way to the next generation of thermoelectrics. *Scripta Materialia*, –.
URL <http://dx.doi.org/10.1016/j.scriptamat.2015.04.026>
- [21] Stevanović, V., Lany, S., Zhang, X., Zunger, A., Mar 2012. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Phys. Rev. B* 85, 115104.
URL <http://dx.doi.org/10.1103/PhysRevB.85.115104>
- [22] Tan, G., Shi, F., Sun, H., Zhao, L.-D., Uher, C., Dravid, V. P., Kanatzidis, M. G., 2014. SnTe as an efficient thermoelectric material with low thermal conductivity. *J. Mater. Chem. A* 2, 20849–20854.
URL <http://dx.doi.org/10.1039/C4TA05530F>
- [23] Wang, H., Gibbs, Z. M., Takagiwa, Y., Snyder, G. J., 2014. Tuning bands of PbTe for better thermoelectric efficiency. *Energy Environ. Sci.* 7, 804–811.
URL <http://dx.doi.org/10.1039/C3EE43438A>
- [24] Wang, H., Pei, Y., LaLonde, A. D., Snyder, G. J., 2011. Heavily doped p-type PbTe with high thermoelectric performance: an alternative for PbTe. *Adv. Mat.*
URL <http://dx.doi.org/10.1002/adma.201004200>
- [25] Wang, H., Schechtel, E., Pei, Y., Snyder, G. J., 2013. High thermoelectric efficiency of n-type PbTe. *Adv. Energy Mater.* 3 (4), 488–495.
URL <http://dx.doi.org/10.1002/aenm.201200683>
- [26] Wang, S., Yang, J., Wu, L., Wei, P., Yang, J., Zhang, W., Grin, Y., 2015. Anisotropic multi-center bonding and high thermoelectric performance in electron-poor CdSb. *Chem. Mat.* 27 (3), 1071–1081.
URL <http://dx.doi.org/10.1021/cm504398d>
- [27] Wu, D., Zhao, L.-D., Hao, S., Jiang, Q., Zheng, F., Doak, J. W., Wu, H., Chi, H., Gelbstein, Y., Uher, C., Wolverton, C., Kanatzidis, M., He, J., 2014. Origin of the high performance in GeTe-based thermoelectric materials upon Bi₂Te₃ doping. *J. Amer. Chem. Soc.* 136 (32), 11412–11419.
URL <http://dx.doi.org/10.1021/ja504896a>
- [28] Xiong, D.-B., Okamoto, N. L., Inui, H., 2013. Enhanced thermoelectric figure of merit in p-type Ag-doped ZnSb nanostructured with Ag₃Sb. *Scripta Materialia* 69 (5), 397 – 400.
URL <http://dx.doi.org/10.1016/j.scriptamat.2013.05.029>
- [29] Yan, J., Gorai, P., Ortiz, B., Miller, S., Barnett, S. A., Mason, T., Stevanovic, V., Toberer, E. S., 2015. Material descriptors for predicting thermoelectric performance. *Energy Environ. Sci.* 8, 983–994.
URL <http://dx.doi.org/10.1039/C4EE03157A>
- [30] Yu, B., Zebbarjadi, M., Wang, H., Lukas, K., Wang, H., Wang, D., Opeil, C., Dresselhaus, M., Chen, G., Ren, Z., 2012. Enhancement of thermoelectric properties by modulation-doping in silicon germanium alloy nanocomposites. *Nano Lett.* 12 (4), 2077–2082.
URL <http://dx.doi.org/10.1021/nl3003045>
- [31] Zhai, Y., Zhang, Q., Jiang, J., Zhang, T., Xiao, Y., Yang, S., Xu, G., 2013. Thermoelectric properties of In_{1.3-x}Sn_xSe prepared by spark plasma sintering method. *J. Alloys Compounds* 553, 270 – 272.
URL <http://dx.doi.org/10.1016/j.jallcom.2012.11.139>
- [32] Zhao, L.-D., Lo, S.-H., Zhang, Y., Sun, H., Tan, G., Uher, C., Wolverton, C., Dravid, V. P., Kanatzidis, M. G., 2014. Ultralow thermal conductivity and high thermoelectric figure of merit in SnSe crystals. *Nature* 508 (7496), 373–377.
URL <http://dx.doi.org/10.1038/nature13184>

- [33] Zhou, M., Gibbs, Z. M., Wang, H., Han, Y., Xin, C., Li, L., Snyder, G. J., 2014. Optimization of thermoelectric efficiency in snite: the case for the light band. *Phys. Chem. Chem. Phys.* 16 (38), 20741–20748.
URL <http://dx.doi.org/10.1039/c4cp02091j>

- Materials
- Visualization
- Resources
- Contribute

Number of Elements

Space Group

Crystal System

Data Type

Data Extraction

Data Analysis

Contribute Data

Experimental

Computational