A data-driven review of thermoelectric materials: Performance and resource considerations

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Gaultois, Michael W., Taylor D. Sparks, Christopher K. H. Borg, Ram Seshadri, William D. Bonificio, and David R. Clarke. Forthcoming. Data-driven review of thermoelectric materials: Performance and resource considerations. Chemistry of Materials.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1021/cm400893e</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:10665230">http://nrs.harvard.edu/urn-3:HUL.InstRepos:10665230</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
A data-driven review of thermoelectric materials: Performance and resource considerations

Michael W. Gaultois,*†‡ Ram Seshadri,*¶†‡ William D. Bonificio,§ and David R. Clarke§

Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106, Materials Research Laboratory, University of California, Santa Barbara, CA 93106, Materials Department, University of California, Santa Barbara, CA 93106, and School of Engineering and Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, MA 02138

E-mail: mgaultois@mrl.ucsb.edu; sparks@mrl.ucsb.edu; seshadri@mrl.ucsb.edu

*To whom correspondence should be addressed
†Department of Chemistry and Biochemistry, UCSB
‡Materials Research Laboratory, UCSB
¶Materials Department, UCSB
§School of Engineering and Applied Sciences, Harvard University
Abstract

In this review, we describe the creation of a large database of thermoelectric materials prepared by abstracting information from over 100 publications. The database has over 18,000 data points from multiple classes of compounds, whose relevant properties have been measured at several temperatures. Appropriate visualization of the data immediately allows certain insights to be gained with regard to the property space of plausible thermoelectric materials. Of particular note is that any candidate material needs to display an electrical resistivity value that is close to 1 m$\Omega$ cm at 300 K, i.e., samples should be significantly more conductive than the Mott minimum metallic conductivity. The Herfindahl-Hirschman index, a commonly accepted measure of market concentration, has been calculated from geological data (known elemental reserves) and geopolitical data (elemental production) for much of the periodic table. The visualization strategy employed here allows rapid sorting of thermoelectric compositions with respect to important issues of elemental scarcity and supply risk.

**Keywords:** Thermoelectrics, datamining, Herfindahl-Hirschman Index, elemental abundance
Introduction

The nature of thermoelectric phenomena and materials — competing and contraindicated
properties, the complexity and variety of the material systems involved — make it some-
what difficult to develop rational strategies that can lead to significant improvements in
performance. Notwithstanding these difficulties, creative approaches have yielded highly
promising materials.\textsuperscript{1–8} The guiding principle behind the design of thermoelectric mate-
rials, and indeed, any functional material, is to completely understand the causal physics
and use such knowledge to rationally optimize material properties. However, even without
knowledge of causality, progress across numerous fields is enabled by correlation, without
\textit{a priori} understanding of the drivers. Particularly when large datasets are available, robust
correlation can be found between seemingly disparate observables. This ability to extract
meaningful information from large pools of data has been somewhat under-utilized in the
search for new materials.

Thermoelectrics and their historical development have been surveyed by Snyder and
Toberer,\textsuperscript{9} Nolas \textit{et al.},\textsuperscript{10} Goldsmid,\textsuperscript{11} and the two CRC handbooks on thermoelectrics by
Rowe.\textsuperscript{12,13} The volume and complexity of research on thermoelectric materials makes the
field fertile for a data-driven review — sometimes referred to as data mining, or mate-
rials informatics. Informatics-based approaches have been successfully used for estimat-
ing some physical properties\textsuperscript{14} and the relative stability of selected material systems,\textsuperscript{15}
and high-throughput methods are becoming increasingly helpful in materials design.\textsuperscript{16}
Given the volume of information that has been published, knowing the right information
to abstract is the first step to such an approach. Furthermore, developing an appropriate
visualization strategy to explore the space of thermoelectric materials is crucial. Accord-
ingly, we have reduced the problem of reviewing thermoelectric performance to several
key properties at four temperatures of interest, and have created an interactive frame-
work to visualize the large amounts of information. Trends in materials properties emerge
from such visualization, and guiding principles for the development of high-performance
thermoelectric materials emerge. Perhaps more importantly, guidelines are suggested for
where not to look in the parameter space of candidate materials.

Since the efficiency of a thermoelectric device is related to the thermoelectric figure
of merit, \( zT = S^2T/(\rho\kappa) \), with \( S \) the Seebeck coefficient, \( \rho \) the electrical resistivity, and \( \kappa \)
the thermal conductivity, these different measured parameters were the obvious choice for
extraction from publications into a database. A particularly useful format for displaying
the large amount of property-based data is based on a modification of the well-known
Jonker plot,\(^\text{17}\) in which the Seebeck coefficient is usually graphed as a function of the elec-
trical conductivity. The use of the Jonker plot has traditionally been limited to optimizing
the carrier concentration in a single thermoelectric material. In contrast, we display all
materials in a single plot, with the different material classes grouped by marker color. In
the modified version employed by us, we prefer to use the electrical resistivity (on a log-
arithmic scale) \( \rho \) as abscissa, and the Seebeck coefficient \( S \) as ordinate. The radius of the
circular marker represents the performance, which is commonly either \( zT \) or the power
factor, \( S^2/\rho \).

There is more to a material than just performance. Given the proposed widespread
application of thermoelectrics and the potential for high-volume use of materials if certain
performance targets are reached, they must be composed of elements that are accessible
and not in danger of a supply risk. Crustal abundance of elements, their global production,
reserves, and use, are some of the factors that determine supply risk. The criticality of
elements in the context of metals that are crucial to energy conversion has been described
by Graedel.\(^\text{18}\) Homm and Klar\(^\text{19}\) have specifically raised these issues for thermoelectric
materials. Following the approach of Graedel,\(^\text{18}\) we have used our database to calculate
several criticality indices for the thermoelectric materials featured here. Resources that
are produced almost entirely in a particular region can provide a single entity leverage
in determining supplies and prices. This geopolitical influence over materials supply and
price can be measured by market concentration, often quantified through the Herfindahl-
Hirschman Index (HHI). The HHI is a financial tool commonly used to measure the monopoly of entities over a commodity or product, and has been previously used as a measure of geopolitical influence on elemental production of a select few elements. Here we calculate the HHI, based on available 2011 data, of almost all of the first 83 elements in the periodic table. Only H, the noble gases, Tc, and Pm are excluded. The HHI is calculated both for elemental production (HHI_P), reflecting the specific geopolitics of the element, as well as for elemental reserves (HHI_R), based on known deposits that could be processed. In conjunction with composition data entered into the database, we employ HHI indices and elemental scarcity values to determine practical issues that influence the likelihood a particular thermoelectric material will find widespread use.

Methods

Nature and source of data

We have abstracted data from representative publications on a wide variety of thermoelectric compounds so that they can be accessed and compared easily with other compounds. The data correspond to four different temperatures in the different regimes of interest for high-temperature thermoelectric devices: 300 K, 400 K, 700 K, 1000 K. There are currently over 1100 database entries (rows), each with 17 associated components: temperature, electrical resistivity, Seebeck coefficient, thermal conductivity, power factor, κzT, zT, chemical composition, material family, preparatory route, material form (whether single crystal or polycrystalline), author, year of publication, DOI link, and comments. Additionally, we generate new metadata from the chemical composition of materials, such as HHI_P, HHI_R, scarcity, and average atomic weight, \( \bar{M} \). To facilitate comparison of materials, we limit our choice of systems in the database to be bulk, polycrystalline samples. In many instances, we employ the power factor or κzT rather than than the actual figure of merit zT for the visualization, because zT requires κ, the thermal conductivity to be measured. Thermal
conductivity is not reported for many materials. Furthermore, of all the thermoelectric parameters it is the least reliable and/or reproducible because it is so sensitive to processing conditions.

Data from published work was extracted manually from digital publications using free software such as PlotDigitizer\textsuperscript{24} and DataThief.\textsuperscript{25} In general, data was extracted from plots of a physical property vs. temperature. If data was not explicitly reported at a temperature of interest, values were interpolated, or extrapolated when appropriate. If property traces (curves) were found to rapidly or unpredictably changing in the region of extrapolation, the point was omitted, or data was taken from the nearest reported temperature. In these cases, the temperature of an extrapolated data point is mentioned in the metadata comment. Because data was entered by hand, the power factor and figure of merit ($\mathit{zT}$) were calculated from the extracted data and checked against the reported values to ensure the data was self-consistent.

HHI values based on production and reserves for each element were calculated from 2011 USGS commodity statistics following the approach used by others.\textsuperscript{18,22,23} When 2011 data was unavailable, data from 2010 or 2009 was used. This has generated a set of HHI values presented in table 1 for much of the periodic table, a significant expansion from previous studies, which focused on eight parent metals.\textsuperscript{18,23} For elements where reserves are seen as adequate or extremely large (e.g., C, O, F, Na, Al, Ca, S, etc.) quantitative reserves are not available but their use is unlikely to reduce their availability to critical levels. For elements such as sodium, where multiple commodities are reported separately (carbonates, sulphates, chloride . . . ), the reports were combined and the HHI values were generated from the aggregate. Another important consideration is that the production and/or reserve values for a country may be unknown or are withheld. In these instances, where the number of producers and/or the amount of reserves are low, these omissions introduce considerable uncertainty. In these cases, the HHI values were estimated based on the general information provided in the USGS report. Estimates of these elements are
denoted by an asterisk in table 1.

Figure 1: Periodic table indicating elemental scarcity and the HHI (production, reserves) indices for most elements. Values were calculated using USGS statistics from 2011, or if unavailable, from statistics no earlier than 2009.

HHI values were calculated using the following expression, $\text{HHI} = \sum_{i} s_i^2$, where $s_i$ is the percent market share of country $i$ in the world production or reserves of a given element. The US Department of Justice and the Federal Trade Commission have designated markets as unconcentrated when HHI < 1500, moderately concentrated when the HHI lies between 1500 and 2500, and highly concentrated when HHI > 2500. If a single country controlled the entire market, HHI = 100$^2$. Elemental HHI values were then used to calculate weighted HHI production and reserve values based on the weight fraction of each element in the chemical formula. Atomic weights were taken from the CRC Handbook. The crustal abundance of elements were obtained from the CRC Handbook and used to generate elemental scarcity values, $\zeta$ (crustal abundance in inverse ppm). These scarcity values were used to calculate the effective scarcity of materials based on the weight frac-
Table 1: Herfindahl-Hirschman Index (HHI) calculated for much of the periodic table, using recent USGS data, typically 2011 for most materials, but no older than 2009. Asterisks against numbers indicate some uncertainty (see text).

<table>
<thead>
<tr>
<th>Element</th>
<th>HHIP</th>
<th>HHIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>3200</td>
<td>3900</td>
</tr>
<tr>
<td>Li</td>
<td>2900</td>
<td>4200</td>
</tr>
<tr>
<td>Be</td>
<td>8000</td>
<td>4000*</td>
</tr>
<tr>
<td>B</td>
<td>2900</td>
<td>2000</td>
</tr>
<tr>
<td>C</td>
<td>500*</td>
<td>500*</td>
</tr>
<tr>
<td>N</td>
<td>1300</td>
<td>500*</td>
</tr>
<tr>
<td>O</td>
<td>500*</td>
<td>500*</td>
</tr>
<tr>
<td>F</td>
<td>1500*</td>
<td>1500*</td>
</tr>
<tr>
<td>Na</td>
<td>1100</td>
<td>500*</td>
</tr>
<tr>
<td>Mg</td>
<td>5300</td>
<td>500*</td>
</tr>
<tr>
<td>Al</td>
<td>1600</td>
<td>1000*</td>
</tr>
<tr>
<td>Si</td>
<td>4700</td>
<td>1000*</td>
</tr>
<tr>
<td>P</td>
<td>2000</td>
<td>5100</td>
</tr>
<tr>
<td>S</td>
<td>700</td>
<td>1000*</td>
</tr>
<tr>
<td>Cl</td>
<td>1500*</td>
<td>1500*</td>
</tr>
<tr>
<td>K</td>
<td>1700</td>
<td>7200</td>
</tr>
<tr>
<td>Ca</td>
<td>3900</td>
<td>1500*</td>
</tr>
<tr>
<td>Sc</td>
<td>5500*</td>
<td>4500*</td>
</tr>
<tr>
<td>Ti</td>
<td>1100</td>
<td>1600</td>
</tr>
<tr>
<td>V</td>
<td>3300</td>
<td>3400</td>
</tr>
<tr>
<td>Cr</td>
<td>3100</td>
<td>4100</td>
</tr>
<tr>
<td>Mn</td>
<td>1600</td>
<td>1800</td>
</tr>
<tr>
<td>Fe</td>
<td>2400</td>
<td>1400</td>
</tr>
<tr>
<td>Co</td>
<td>3100</td>
<td>2700</td>
</tr>
<tr>
<td>Ni</td>
<td>1000</td>
<td>1500</td>
</tr>
<tr>
<td>Cu</td>
<td>1600</td>
<td>1500</td>
</tr>
<tr>
<td>Zn</td>
<td>1600</td>
<td>1900</td>
</tr>
<tr>
<td>Ga</td>
<td>5500*</td>
<td>1900*</td>
</tr>
<tr>
<td>Ge</td>
<td>5300</td>
<td>1900*</td>
</tr>
<tr>
<td>As</td>
<td>3300</td>
<td>4000*</td>
</tr>
<tr>
<td>Se</td>
<td>2200</td>
<td>1900</td>
</tr>
<tr>
<td>Br</td>
<td>3300</td>
<td>6900</td>
</tr>
<tr>
<td>Rb</td>
<td>6000*</td>
<td>6000*</td>
</tr>
<tr>
<td>Sr</td>
<td>4200</td>
<td>3000*</td>
</tr>
<tr>
<td>Y</td>
<td>9800</td>
<td>2600</td>
</tr>
<tr>
<td>Zr</td>
<td>3400</td>
<td>2600</td>
</tr>
<tr>
<td>Nb</td>
<td>8500</td>
<td>8800</td>
</tr>
<tr>
<td>Mo</td>
<td>2400</td>
<td>5300</td>
</tr>
<tr>
<td>Ru</td>
<td>3200*</td>
<td>8000*</td>
</tr>
<tr>
<td>Rh</td>
<td>3200*</td>
<td>8000*</td>
</tr>
<tr>
<td>Pd</td>
<td>3200</td>
<td>8000*</td>
</tr>
<tr>
<td>Ag</td>
<td>1200</td>
<td>1400</td>
</tr>
<tr>
<td>Cd</td>
<td>1700</td>
<td>1300</td>
</tr>
<tr>
<td>In</td>
<td>3300</td>
<td>2000*</td>
</tr>
<tr>
<td>Sn</td>
<td>2600</td>
<td>1600</td>
</tr>
<tr>
<td>Sb</td>
<td>7900</td>
<td>3400</td>
</tr>
<tr>
<td>Te</td>
<td>4900</td>
<td>4800</td>
</tr>
<tr>
<td>I</td>
<td>6000*</td>
<td>6000*</td>
</tr>
<tr>
<td>Ba</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>La</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Ce</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Pr</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Nd</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Pm</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Sm</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Eu</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Gd</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Tb</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Dy</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Ho</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Er</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Tm</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Yb</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Lu</td>
<td>9500</td>
<td>3100</td>
</tr>
<tr>
<td>Hf</td>
<td>3400*</td>
<td>2600*</td>
</tr>
<tr>
<td>Ta</td>
<td>2300</td>
<td>4800</td>
</tr>
<tr>
<td>W</td>
<td>7000</td>
<td>4300</td>
</tr>
<tr>
<td>Re</td>
<td>3300</td>
<td>3300</td>
</tr>
<tr>
<td>Os</td>
<td>5500*</td>
<td>9100*</td>
</tr>
<tr>
<td>Ir</td>
<td>5500*</td>
<td>9100*</td>
</tr>
<tr>
<td>Pt</td>
<td>5500</td>
<td>9100*</td>
</tr>
<tr>
<td>Au</td>
<td>1100</td>
<td>1000</td>
</tr>
<tr>
<td>Hg</td>
<td>5500</td>
<td>3100</td>
</tr>
<tr>
<td>Tl</td>
<td>6500*</td>
<td>6500*</td>
</tr>
<tr>
<td>Pb</td>
<td>2700</td>
<td>1800</td>
</tr>
<tr>
<td>Bi</td>
<td>5300</td>
<td>6000</td>
</tr>
</tbody>
</table>
tion of elements in the chemical formula, \( \zeta = \sum_i^N (\zeta_i \times \frac{m_i}{m_{tot}}) \) where \( \zeta_i \) and \( m_i \) are the scarcity and weight of an element \( i \) in a material. The scarcity and Herfindahl-Hirschman Index values (based on production and reserves) for much of the periodic table is shown schematically in figure 1.

**Mechanics of visualization**

![Flowchart for extraction, organization, and visualization of data](image)

Figure 2: Flowchart for extraction, organization, and visualization of data. Note that the first step of data extraction is manual.

At the initial stages of this work, it became evident that gathering the large amount of data would be futile in the absence of an appropriate framework that permitted the gathered data to be visualized in a flexible manner. In this section we describe a website we have developed and hosted at [www.mrl.ucsb.edu:8080/datamine/thermoelectric.jsp](http://www.mrl.ucsb.edu:8080/datamine/thermoelectric.jsp) as an example of a plausible framework for organizing and visualizing the results of such data mining. The flowchart describing the data mining, database formation, and visualization process is
Figure 3: Screenshot of the web-based visualization tool, that permits the simultaneous visualization of four parameters: abscissa, ordinate, marker size and color. Several variables can be chosen as abscissa and ordinate, and measures of thermoelectric performance can be represented by the radius of the data points. To simplify navigation, families of related materials can be displayed or hidden by clicking their legend marker. Further, hovering over a data point intuitively reveals a tooltip with pertinent information: the names and values for the abscissa and ordinate, the chemical composition, the sample form (e.g., polycrystalline, single crystal, nanoparticles), the preparatory route (e.g., ceramic method, arc melting), the author and year, and either the power factor, $\kappa zT$, or $zT$ value. Clicking on a data point leads the web-browser directly to the source publication via the document object identifier (DOI).
summarized in figure 2. A screenshot of the website is shown in figure 3. Because several physical parameters were collected and tabulated, visualizing any number of combinations along the abscissa and ordinate is possible using the website. While not all combinations will yield insightful relationships, the website allows using as axes choices, the electrical resistivity, Seebeck coefficient, thermal conductivity, average atomic weight, elemental scarcity, and HHI based on production or reserves of a material. To increase the information density that can be visualized, a third dimension is also plotted: the size of the data point (the radius of a circle) is proportional to material performance. The marker size can be either power factor ($S^2/\rho$), $\kappa zT$, or $zT$. $\kappa zT$ allows a rough comparison with $zT$, but does not require the thermal conductivity to have been measured or reported. Many materials have a thermal conductivity between $1 \text{ W m}^{-1}\text{K}^{-1}$ to $10 \text{ W m}^{-1}\text{K}^{-1}$, and in the case that $\kappa = 1 \text{ W m}^{-1}\text{K}^{-1}$, the numerical value of $\kappa zT$ is the same as the numerical value of $zT$.

The website allows interactive exploration of all the data. Hovering over a data point reveals a tooltip with pertinent information: the values for the abscissa and ordinate, the chemical composition, the sample form (e.g., polycrystalline, single crystal, nanoparticles), the preparatory route (e.g., ceramic method, arc melting), the author and year, and either the power factor, $\kappa zT$, or $zT$ value. Additionally, clicking on a data point leads the web-browser directly to the publication via the document object identifier (DOI). To enable sorting of the large number of datasets, the user may choose to sort by material family or temperature regime. Although the material families are shown with different marker color the ability to hide or show a given family can make direct comparison more clear. This is accomplished by clicking the name of the dataset in the legend. Finally, option is available to visualize the results of the thermoelectric database employed here, or for users to upload their own data, following an Excel™ template file available on the website. In this way a user can look for trends in their data or use the website’s Java™ code to generate additional non-performance-related data such as scarcity, average atomic weight.
or HHI.

Results and discussion

Examples of useful visualization schemes

Table 2: The flexibility of the visualization framework allows users to examine the relation between any number of different parameters. Several examples are tabulated here, along with what can potentially be learned.

<table>
<thead>
<tr>
<th>Abscissa</th>
<th>Ordinate</th>
<th>Size (radius)</th>
<th>Use</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>$S$</td>
<td>$S^2/\rho$, $\kappa zT$, $zT$</td>
<td>General trends, insight</td>
<td>Compound must be metallic for high performance</td>
</tr>
<tr>
<td>$\bar{M}$</td>
<td>$\kappa$</td>
<td>$\kappa zT$, $zT$</td>
<td>Effect of $\bar{\omega}$</td>
<td>$\kappa$ decreases with increasing $\bar{M}$</td>
</tr>
<tr>
<td>$\rho_{300,K}$</td>
<td>$S_{300,K}$</td>
<td>$\kappa zT$, $zT$ at different temperatures</td>
<td>Estimate high-$T$ performance</td>
<td>Rapid screening of high-$T$ materials</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$\kappa$</td>
<td>$\kappa zT$, $zT$</td>
<td>Effect of $\kappa_e$</td>
<td>Slope of $\kappa$ vs. $\rho$ changes linearly with $T$ of dataset</td>
</tr>
<tr>
<td>HHIP</td>
<td>HHIR</td>
<td>$\kappa zT$, $zT$</td>
<td>Material choice and criticality</td>
<td>Intensive use of rare-minerals and/or Sb may strain markets</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>HHIP</td>
<td>$\kappa zT$, $zT$</td>
<td>Material choice and criticality</td>
<td>Some elements are abundant, yet will have volatile prices. Many state-of-the-art high-performance materials rank poorly in criticality indices</td>
</tr>
</tbody>
</table>

In table 2, we list those combinations of parameters that we have found particularly useful to plot. The various combinations and the findings are described more thoroughly in the following sections.

Rapid screening of materials

We begin by visualizing the general trend in Seebeck coefficient as a function of the electrical resistivity (figure 4): These two properties jointly contribute to $zT$. Higher information density is achieved by encoding the power factor as the marker size. This style of plot
Figure 4: The Seebeck coefficient of a wide variety of materials grouped by material family is plotted against electrical resistivity. The marker size (radius) here is proportional to the the power factor. Materials with high performance exhibit an high Seebeck coefficient at a given electrical resistivity, and lie outside the conical envelope defined by most thermoelectric materials. This envelope can be described by lines of constant power factor ($S^2/\rho$), which are shown to give lower and upper bounds of performance. With few exceptions, all investigated materials with reasonable thermoelectric performance are well described as metals ($\rho_{300 K} << 0.01 \Omega \text{cm}$).
was first introduced in a previous report. Here we enhance it by assigning a color to a family of related materials. Although the axes are similar to those for a Jonker plot (where Seebeck coefficient is plotted vs. electrical conductivity) the application is quite distinct. Jonker plots are usually employed to examine the effect of changing the carrier concentration in a single material.

Looking at the data as a whole, general envelope of materials is cone-like, widening at higher electrical resistivities (figure 4). The best materials clearly define the left-most edges of the envelope. All the high-performance materials have interesting physics that make them special, even when looking only at these two properties. Specifically, materials with metallic behavior generally have low Seebeck coefficients, but high-performance thermoelectric materials violate this principle. For example, band asymmetry or high band degeneracy near the Fermi level in BiTe- and PbTe-based systems leads to an unusually high Seebeck coefficient, despite their metallic behavior. Likewise, even though Na$_x$CoO$_2$ is metallic, with polycrystalline samples having $\rho_{300K} \approx 0.002 \Omega \text{cm}$, the correlated behavior of electrons and spin contribution to thermopower–or, arguably, the unique band structure–lead to a remarkably high Seebeck coefficient.

Looking at the ensemble of data (figure 4) also provides insight about particular families of materials. For example, we find that optimizing non-metallic moderate-performance materials is unlikely to lead to dramatic gains in performance, as the electrical resistivity is too high. For example, CaMnO$_3$-based systems remain a topic of intense study, but when viewed in the context of all materials, their placement in the map suggests it is unlikely that high performance will ever be reached.

It was previously shown that room-temperature thermoelectric properties (Seebeck coefficient and electrical resistivity) of several high-performance thermoelectric materials clustered in one area of the thermoelectric map (figure 4) when compared with low-performance materials. Here we suggest that the relationship between the properties at room-temperature and the properties at the temperature of highest $zT$ for a small set of
Figure 5: Several high-performance thermoelectric materials were chosen to examine how their room temperature properties relate to their properties at the temperature of highest $zT$. Arrows show the change from room temperature (solid circles) to the temperature of highest $zT$ (dashed circles). The properties of several different material families appear to cluster at their temperature of highest $zT$, with the electrical resistivity between 0.001 $\text{W cm}$ and 0.01 $\text{W cm}$ and the absolute value of the Seebeck coefficient between 150 $\mu \text{V K}^{-1}$ and 300 $\mu \text{V K}^{-1}$. Additionally, nearly all high-performance materials have a room-temperature electrical resistivity below the Mott maximum metallic resistivity ($\rho_{300K} < 0.01 \Omega \text{ cm}$). The materials shown are $\text{Zn}_{0.98}\text{Al}_{0.02}\text{O}$, $\text{Na}_{2}\text{CoO}_{2}$, $\text{Ca}_{2}\text{Co}_{2}\text{O}_{5}$, $\text{(Zr}_{0.5}\text{Hf}_{0.5})_{0.5}\text{Ti}_{0.5}\text{NiSn}$, $\text{Mg}_{2}\text{Si}_{0.99}\text{Bi}_{0.001}$ (p-type), $\text{(Mg}_{2}\text{Si})_{0.97}\text{Bi}_{0.03}$ (n-type), $\text{Si}_{0.8}\text{Ge}_{0.2}$ (p-type), $\text{Si}_{0.8}\text{Ge}_{0.2}$ (n-type), $\text{PbTe}_{0.75}\text{Se}_{0.25}$, $\text{CsBi}_{4}\text{Te}_{6}$, $\text{Ba}_{8}\text{Ga}_{18}\text{Ge}_{28}$ (p-type), $\text{Ba}_{8}\text{Ga}_{16}\text{Ge}_{30}$ (n-type), and $\text{In}_{0.25}\text{Co}_{4}\text{Sb}_{12}$.\(^{33}\)\(^{34}\)\(^{35}\)\(^{36}\)\(^{37}\)\(^{38}\)\(^{39}\)\(^{40}\)\(^{41}\)\(^{42}\)\(^{43}\)
mid- and high-temperature thermoelectric materials (figure 5). Interestingly, the properties of several different material families appear to cluster at their temperature of highest $zT$, with the electrical resistivity between $0.001 \Omega \text{cm}$ and $0.01 \Omega \text{cm}$ and the absolute value of the Seebeck coefficient between $150 \mu \text{V K}^{-1}$ and $300 \mu \text{V K}^{-1}$. Although the small dataset prevents any strong conclusions, examining more materials may provide some predictive ability of high-temperature properties, using those measured at room temperature. Even some hints with regard to high-temperature properties would be powerful: measurement of high temperature properties is time-consuming and requires specialized instruments. Furthermore, there is much more room-temperature data available in the literature and a single room-temperature measurement could facilitate combinatorial testing of a large phase space.\textsuperscript{44}

The thermoelectric survey shown in figure 4 and the high-temperature trends observed in figure 5 suggest that all thermoelectric materials with any appreciable performance are all metallic, with an electrical resistivity well below the Mott maximum metallic resistivity at room temperature, i.e. $\rho_{300K} < 0.01 \Omega \text{cm}$. This provides a valuable thermoelectric screening criterion and guiding direction for future studies. Unlike intermetallic compounds, which are generally metallic or semi-metallic, transition metal oxides span the gamut of electrical resistivities, ranging from insulating (e.g. TiO$_2$) to metallic (e.g. ReO$_3$). A common strategy to seeking effective thermoelectric oxides is to examine the metal/non-metal border, i.e. proximal to the Mott minimum metallic conductivity, where Seebeck coefficients are often substantial. However, the utility of this approach appears somewhat questionable, given that the good thermoelectrics appear well on the metallic side.

**Thermal conductivity**

Over the past several decades, the majority of improvements in thermoelectric materials have resulted from decreasing the lattice contribution to thermal conductivity. A number of materials selection guidelines have been identified in order to reduce the thermal
Figure 6: Thermal conductivity plotted against average atomic weight for a variety of thermoelectric materials grouped by material family, with marker indicating $zT$. The dashed line represents a best fit regression. Notwithstanding the combination of pure phase materials and heavily nano-structured materials, the thermal conductivity generally decreases with increasing average atomic weight. However, at fixed average atomic weight, there is considerable tunability.

One such strategy is to lower the vibrational frequency, and thus the thermal conductivity, by using materials with a large average atomic weight. This has frequently been touted as a reason for the higher thermal conductivity in oxides relative to compounds with heavier anions such as Te, Se, and Sb. Examination of the relationship between thermal conductivity, $\kappa$, and average atomic weight, $\bar{M}$, confirms the general reduction of thermal conductivity in heavier compounds (figure 6). For example, compounds with $\bar{M} \approx 25\,\text{g/mol}$ have an average $\kappa \approx 4\,\text{Wm}^{-1}\text{K}^{-1}$ whereas those with $\bar{M} \approx 105\,\text{g/mol}$ have an average $\kappa \approx 2\,\text{Wm}^{-1}\text{K}^{-1}$. In the absence of point defects or other scattering mechanisms, the reduction in thermal conductivity should scale as $\bar{M}^{-1/2}$. However, many of the materials contained in this plot rely on additional techniques to reduce thermal conductivity (e.g., increased phonon scattering from defects, alloying, grain boundaries, interfaces, nano-bulk compounds, complex crystal structures, etc.). On the other hand, the lower vibrational frequency achieved in compounds with large average...
atomic weight has only a small impact on the electrical resistivity.

Figure 7: Thermal conductivity plotted against electrical resistivity for a variety of thermoelectric materials grouped by dataset temperature with marker size indicating $zT$. Lines are best fit linear regressions of the data plotted on log-log plot. The slopes are then plotted as a function of dataset temperature in the inset. Despite the large scatter in the data that has been fitted, the general trend that emerges is consistent with the $T^{-1}$ dependence of Umklapp scattering of phonons.

Finally, we examine the relationship between total thermal conductivity and electrical resistivity (figure 7); the thermal conductivity decreases with increasing electrical resistivity. This behavior is expected for the electronic contribution from the Wiedemann-Franz law, $\kappa_e/\sigma = \pi^2 k_B^2/3e^2T = L_0T$ where $L_0$ is the Lorenz number ($2.44 \times 10^{-8} W \Omega K^{-2}$) and $\sigma$ is the electrical conductivity. There is another consideration that could be important. For a given composition, the same structural features that give rise to low resistivity, — for example structures that are highly connected in three-dimensions — are the same features that often lead to high lattice thermal conductivity.

Examining the broad correlation across all material families in the metallic region ($\rho < 0.01 \Omega \text{cm}$) reveals the slope of the log-log plot $\kappa$ vs. $\rho$ decreases linearly with increasing temperature (inset, figure 7). The power-law behavior revealed from the fit is consistent with Umklapp scattering, explained below.
Examining correlations across many material families with disparate properties will only provide general trends rather than precise values for a specific material family. Nevertheless, there is insight in comparing the correlation (i.e., slope) at different temperatures. For example, while there is considerable scatter in the data at each temperature, the temperature dependence of the correlation is strongly linear \( R^2 = 0.997 \), in good agreement with the expected behavior as Umklapp scattering begins to dominate at higher temperatures. The changing slope as a function of dataset temperature results from increased phonon scattering lowering the phonon contribution to thermal conductivity at higher temperatures. When the lattice \( \kappa_l \) is minimized, the electronic \( \kappa_e \) dominates, so changes in electrical resistivity have a larger impact on thermal conductivity. The power dependence of temperature on \( \kappa \) vs. \( \rho \) slope may result from the Umklapp scattering, a three-phonon scattering process. According to the Bose-Einstein expression, \( < n > = 1/\exp[\hbar \omega/k_B T] - 1 \approx k_B T/\hbar \omega \), the average phonon population, \( < n > \), increases with temperature which increases the likelihood of phonon-phonon scattering. In fact, Grimvall\(^{47}\) has predicted a \( T^{-1} \) temperature dependence for three-phonon scattering at moderate temperatures (near the Debye temperature).

**Resource considerations**

In this section we review parameters beyond thermoelectric performance. We analyze resource considerations such as scarcity and supply risk of thermoelectric materials based on their elemental composition. The implications of the analysis on the choice of thermoelectric materials is discussed below. A combined analysis of performance and resource considerations is necessary for identifying thermoelectric materials with the greatest promise of widespread application. Scare elements could be employed in efforts to make high-impact discoveries and to obtain better understanding of materials trends. However, when materials are meant for widespread deployment, the incorporation of scarce elements becomes
an important point to consider during the design phase.

We begin by discussing scarcity which places a fundamental limit on the amount of material available for use. The following section addresses HHI, a measure of geopolitical influence. Abundance of raw materials is a key factor that determines cost, in addition to other factors such as actuarial costs, packaging, transportation, and assembly, to name a few. Technologies relying on scarce elements such as Re, Te, or Pt group metals, are susceptible to large materials costs. For example, turbine engines consume 70% of the world’s Re production, which drives the cost of Re.\(^4\)\(^8\) The cost of scarce materials can be exacerbated when multiple technologies use the same element, such as the use of In for transparent conducting oxides (TCO) and copper indium gallium selenide (CIGS) photovoltaics. Cost can also be high for relatively abundant materials where both the demand and production are low, such as low purity Si.\(^1\)\(^9\)

Figure 8: Nearly all high-performance thermoelectric materials pose either a scarcity or HHI risk, which will limit any intensive deployment for energy applications. The scarcity (inverse crustal abundance) of familiar elements is shown to give context. Although not immediately apparent, incorporation of Bi, Sb, Te, or rare-earth elements leads to significant risk in either scarcity or HHI. Materials that pose a lower risk of criticality are SiGe, silicides, half-Heuslers, and cobaltates.
Plotting scarcity against HHI\textsubscript{P} for thermoelectric materials is a useful tool to identify technologies where materials cost can be a concern. About half of the highest efficiency materials contain a large weight fraction of Te. (e.g. Be\textsubscript{2}Te\textsubscript{3}, PbTe, La\textsubscript{3}Te\textsubscript{4}, TAGS, Ag\textsubscript{9}Tl\textsubscript{5}Te\textsubscript{5}, SbTe, or Tl\textsubscript{6}BiTe\textsubscript{6}.) Given that Te has one of the highest scarcity values, these materials may pose problems for widespread deployment. Te may be rare, but its HHI\textsubscript{P} is not critical because it is produced in many countries as a byproduct of Cu, reducing supply risks. Additionally, with proper recycling and waste processing policies, Te could be recovered and reprocessed, a practice already carried out by some photovoltaic companies.

An unanticipated finding from analysis of figure 8 was that incorporation of even small amounts of scarce elements in otherwise Earth-abundant compounds may be problematic. Mg\textsubscript{2}Si has a low scarcity of $\zeta = 27.3$, but when doped with trace amounts of Bi to Mg\textsubscript{2}Si\textsubscript{0.993}Bi\textsubscript{0.007}, $\zeta$ increases to $1.1 \times 10^6$ because of the extremely low crustal abundance (0.025 ppm) and high atomic weight of Bi. Even 0.1% substitution of Bi on the Si site results in a four orders of magnitude increase in $\zeta$.

A compromise between performance and resource considerations is seen in oxide materials. Cobalt oxides, such as NaCo\textsubscript{2}O\textsubscript{4}, Ca\textsubscript{2}Co\textsubscript{2}O\textsubscript{5}, and Ca\textsubscript{3}Co\textsubscript{4}O\textsubscript{9}, as well as the reduced ferroelectric oxide Sr\textsubscript{0.61}Ba\textsubscript{0.39}Nb\textsubscript{2}O\textsubscript{6}, have competitive $\zeta T$ values with scarcity values four orders of magnitude lower than state-of-the-art chalcogenides, two orders of magnitude lower than skutterudites (ACo\textsubscript{4}Sb\textsubscript{12}, A = caged atom) and ZintlS (Zn\textsubscript{4}Sb\textsubscript{3}, Yb\textsubscript{14}MnSb\textsubscript{11}), and one order of magnitude lower than clathrates (Ba\textsubscript{8}Ga\textsubscript{18}Ge\textsubscript{28}) and half-Heuslers (TiNiS). ZnO and SrTiO\textsubscript{3} are even more abundant, but have lower performance. It is worth noting that, as with Mg\textsubscript{2}Si, if Bi is added to improve performance (e.g., Bi\textsubscript{2}Sr\textsubscript{2}Co\textsubscript{2}O\textsubscript{8}), a dramatic increase in $\zeta$ is observed.

Determining the Herfindahl-Hirschman Index (HHI) for all elements has shown the following elements to be most at risk for market abuse: Be, K, Br, Rb, Nb, Sb, Cs, W, the rare-earth (RE) elements including Y, and the platinum group elements (PGE): Ru, Rh, Pd, Os, Ir, Pt. These elements have HHI values of greater than 6000 in either production or
reserves. As with scarcity, introduction of even small amounts of elements with high HHI\textsubscript{P} or HHI\textsubscript{R} such as REs, Sb, or Tl, may not be justified for a marginal gain in performance. Materials like Zintl or clathrate antimonides and rare-earth containing chalcogenides are poor outliers compared to other thermoelectric materials in terms of HHI\textsubscript{P} risk.

The importance of HHI when considering thermoelectric materials can be illustrated through recent price spikes in antimony. In 2011, a single country that was responsible for nearly 90\% of Sb mining halted much of its production. As a result, Sb prices rose over 15\% in less than a year. While Sb is particularly concentrated, over 70\% of elements exceed the 2500 threshold to be considered “highly concentrated” according to Federal Trade Commission definition. This makes many thermoelectric materials sensitive to changing market conditions. This analysis signals that emphasis should be placed on developing materials with a lower HHI. Examples of materials that have reasonable performance and low HHI include S- and Se-containing chalcogenides, the half-Heusler materials low in Hf- and Nb-content, and sodium cobalt oxides. Practical consideration of these criticality parameters, although not performance related, would benefit both material and experimental design.

**Future directions**

Thermoelectric data mining provides insights, enabling rational material design. There are, in addition to the properties described, a number of additional materials properties and parameters that were not included that are nevertheless highly desired in comparing different thermoelectric materials. For example, crystallographic data would allow for the calculation of average atomic volume, unit cell volume, and number of atoms in the unit cell. This could be valuable in looking at changes in thermal conductivity. Mechanical properties, coefficients of thermal expansion, and high temperature stability indices would be valuable in thermoelectric device design. Furthermore, additional non-performance re-
lated parameters such as toxicity are of great interest in environmental and life-cycle analyses of thermoelectric components. Some of these properties could be extracted from other databases, measured, or approximated. Other properties, such as toxicity, are fundamentally more challenging to obtain. A comprehensive approach to comparing toxicity across compounds is not well established, since it depends on environmental transformations and persistence, exposure route, form (e.g., oxidation state, presence of counter-ions), and many other factors. Finally, adding a dimension of time to the plots we have shown may increase their utility. For example, looking at the evolution of HHI over time may provide a better ability to estimate risk and market volatility of particular materials.

The data-driven approach and visualization can be readily extended to other materials functions. For example, photovoltaics, thermal barrier coatings, dielectrics, fuel cells, transparent conducting oxides, gas turbine superalloys, and batteries are all areas of research where progress depends on optimizing several competing requirements concurrently, that could benefit from a comprehensive, data-driven approach.

**Closing remarks**

The framework demonstrated here allows researchers obtain a birds-eye view of a large domain of thermoelectrics research, and compare new materials of interest. The thermoelectric design space is large, and this overview of the field, including important non-performance related parameters, allows researchers to focus on property regions and material families best suited for a given application. In the case of power generation with thermoelectrics, analysis of HHI, scarcity, and materials properties can be done visually. Using the methods shown here, we highlight several families of materials that combine good performance-related properties with non-critical resource availability: highly conductive (electrically) early transition metal oxides, silicides with low bismuth content, and half-Heuslers. We also observe that all high-\(\varepsilon T\) materials are found in the metallic con-
duction region 0.001Ω cm to 0.01Ω cm. This visualization framework is also valuable from a design perspective, where it serves as a useful guide to identifying promising new materials.

Acknowledgement

We thank the National Science Foundation for support of this research through NSF-DMR 1121053. RS acknowledges useful discussions with Ambuj Singh. MWG thanks the Natural Sciences and Engineering Council of Canada for support through a NSERC Postgraduate Scholarship and the US Department of State for an International Fulbright Science & Technology Award. The research at Harvard was supported by the Office of Naval Research through contract N000141110894 and by the Harvard University Center for the Environment.
Appendix

Table 3: General class of material family and publications from which data was extracted. Representative publications are chosen for each class of materials.

<table>
<thead>
<tr>
<th>Material family</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mn oxide</td>
<td>49–60</td>
</tr>
<tr>
<td>ZnO, SrTiO3</td>
<td>33, 61–72</td>
</tr>
<tr>
<td>Co oxide</td>
<td>34, 35, 73–77</td>
</tr>
<tr>
<td>other oxide</td>
<td>28, 63, 78–96</td>
</tr>
<tr>
<td>Si-Ge</td>
<td>39, 97</td>
</tr>
<tr>
<td>clathrate</td>
<td>41, 42, 98–103</td>
</tr>
<tr>
<td>half-Heusler</td>
<td>36, 104–107</td>
</tr>
<tr>
<td>skutterudite</td>
<td>1, 7, 43, 108</td>
</tr>
<tr>
<td>chalcogenide</td>
<td>5, 30, 40, 109–122</td>
</tr>
<tr>
<td>silicide</td>
<td>37, 38, 123–126</td>
</tr>
<tr>
<td>Zintl</td>
<td>127–132</td>
</tr>
</tbody>
</table>
References


[30] Heremans, J. P.; Jovovic, V.; Toberer, E. S.; Saramat, A.; Kurosaki, K.; Charoen-


4647.


[37] Bux, S. K.; Yeung, M. T.; Toberer, E. S.; Snyder, G. J.; Kaner, R. B.; Fleurial, J. P. J.


4375–4380.
A data-driven review of thermoelectric materials, based around “mining” data from around 100 publications, and then devising an appropriate visualization scheme to view the mined data yields insights into performance metrics (indicated by the size of the circle) and economic realities.