Scripta Materialia 111 (2016) 10–15

Contents lists available at ScienceDirect

Scripta Materialia

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

Viewpoint Paper

Data mining our way to the next generation of thermoelectrics

Taylor D. Sparks [a,*], Michael W. Gaultois [b], Anton Oliynyk [c], Jakoah Brgoch [d], Bryce Meredig [e]

[a] Materials Science and Engineering Department, University of Utah, Salt Lake City, UT 84112, USA
[b] Materials Research Laboratory and the Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106, USA
[c] Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada
[d] Department of Chemistry, University of Houston, Houston, TX 77204, USA
[e] Citrine Informatics, Redwood City, CA 94061, USA

ARTICLE INFO

Article history:
Received 2 April 2015
Revised 28 April 2015
Accepted 29 April 2015
Available online 15 May 2015

Keywords:
Materials discovery
Thermoelectric materials
Rapid screening
Data mining
Machine learning

ABSTRACT

In this article we provide an overview of data mining, informatics, and machine learning approaches for thermoelectrics. We describe how the initial development of a thermoelectric materials database has enabled the creation of a recommendation engine governed by machine learning and how this engine introduces a new paradigm in thermoelectric materials development. Performance probability is generated based on training models. A demonstration of the data mining approach is set forth in a ternary intermetallic system, where we report new materials.

1. Introduction

High efficiency thermoelectric materials offer great potential for the development of energy harvesting and solid-state refrigeration devices. The field is also rich in opportunities to explore fundamental condensed matter structure–property relationships. This interest has led to exponential growth in the number of articles written each year (~10^4 this year!) on thermoelectrics, and yet most experiments continue to focus on a well-known set of chemistries and crystal structures, including chalcogenides, skutterudites, and Zintl phases [1–3]. Occasionally an unexpected material class is discovered fortuitously, such as Na_xCoO_2 derived thermoelectrics [4].

The reason for this incremental approach is the high risk associated with searching through chemical whitespace for entirely new materials. Given the heavy focus on traditional, intermetallic thermoelectrics, the development of new high-performance materials that also incorporate low-cost, abundant, and sustainable materials may require different design principles. One possible approach we describe here leverages the wealth of data in the field by incorporating data mining, informatics, and machine learning in conjunction with traditional experimental investigation to lower the risk barrier to entry in exploring new materials and compositions.

2. Data mining visualizations in thermoelectrics

Materials science is on the cusp of being transformed through the use of data and materials genomics in integrated computation and materials design. Some of the challenges and potential impact that open and accessible materials and engineering data can have on future research have recently been detailed [5]. Experimental thermoelectric research is at times guided and accompanied by computation. This has led to the development of *ab initio*, high-throughput, computational screening techniques [6–9]. Only recently in the field of thermoelectrics have efforts been made to begin cataloging data in order to take advantage of informatics and machine learning to rapidly accelerate materials discovery. For example, Gaultois et al. assembled a database allowing simultaneous analysis of performance and resource considerations based on materials availability and geological as well as geopolitical supply risks [10]. This approach has served as a model for developing data-driven approaches to other applications [11].

To maximize the dimensionality and information content, the scatter plots employed in these works provide a third dimension

* Corresponding author.
E-mail addresses: sparks@eng.utah.edu (T.D. Sparks), mgaultois@mrl.ucsb.edu (M.W. Gaultois).

http://dx.doi.org/10.1016/j.scriptamat.2015.04.026
1359-6462/

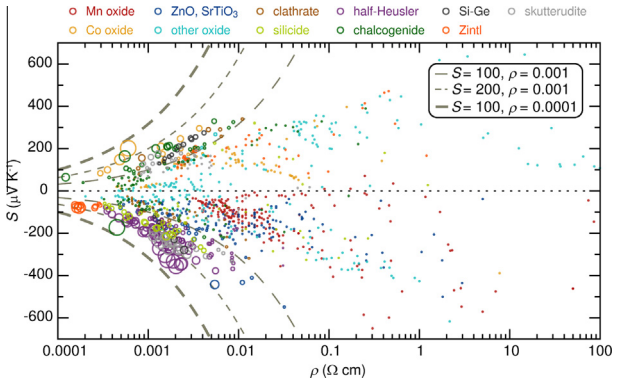

Figure 1. Bird's-eye-view plot of thermoelectric materials featuring four dimensional information visualization. Reprinted with permission from [10]. Copyright (2013) American Chemical Society.

of marker size, which indicates a performance metric (in this case zT), and a fourth dimension of marker color, which indicates the family of related materials (Fig. 1). By incorporating both performance and resource considerations, a unique set of tools for researchers and design criteria were used to outline a parameter space of interest to guide the discovery of new thermoelectric materials with favorable performance.

For example, it is observed that materials must have an electrical resistivity below $10^{-2}\,\Omega\,\mathrm{cm}$ and Seebeck coefficient above 100 μV/K for any reasonable thermoelectric performance, i.e. high zT. This dramatically restricts the design space of oxide thermoelectric materials, which is important to guide future studies. A few other important observations in existing thermoelectric materials include:

1. Many state-of-the-art materials including chalcogenides, skutterudites and Zintl phases rely on costly, scarce materials with geological and geopolitical supply risk whereas lower performing oxides, silicides, and half-Heuslers do not.
2. The Lorenz number, L_0, can be used to determine the fraction of the total conductivity arising from electrical carriers, κ_e/κ_{total}, and this value can be multiplied by S^2/L_0 to yield zT for a given material and zT_{max} occurs when $\kappa_e/\kappa_{total} \rightarrow 1$
3. The properties of materials tend to cluster by material family, rather than change drastically from regions of poor to high performance.
4. Improving the electrical properties (i.e., power factor) is non-trivial in even the most-studied systems (e.g., band engineering, resonant-level doping, quantum wells), and unfathomable in many oxide systems.

Taken together, these observations provide additional thermoelectric design criteria. A balance of moderate performance and sustainability in materials highlights the merits of half-Heusler, oxides and silicides. By examining the limits of zT given thermal conductivity considerations, i.e. κ_e/κ_{total}, some classes of materials, such as some chalcogenides, are approaching the limit of zT whereas other materials with higher Seebeck coefficients have significant room for improvement via reducing the lattice thermal conductivity [12]. Finally the clustering of materials from similar families in parameter space suggests that rather than trying to improve the performance of materials that exhibit low zT, new materials should be sought out with better electrical properties.

3. Predicting new, high-performance materials

In the search for new materials, researchers have turned to the development of computational methodologies to predict the physical properties of materials. Identifying the most promising materials using first principles level calculations is becoming more common due to improvements in computational codes, computing power, and the support of extensive research efforts such as the Materials Project [13]. These advances have undoubtedly supported the development of new materials ranging from intercalation batteries [14] to phosphors for solid state lighting [15]. Often the properties of these functional materials can be accurately reproduced by electronic structure calculations, making high-throughput screening viable for materials development. In terms of thermoelectrics, however, the calculated figure of merit requires a combination of electronic structure and phonon calculations to reproduce the properties necessary for materials prediction.

While an accurate estimations of band-gaps can now be achieved using advanced hybrid functionals, calculating transport properties such as thermal conductivity (κ) at the *ab initio* level remains computationally expensive. Although classic expressions to approximate thermal conductivity have been developed by Slack and others, these calculations require estimations often require the Grüneisen parameter, which is a demanding calculation even for simple crystal structures. Subsequent solutions based on semi-classical transport theory [16,17] and first principles theory[18,19] have also proved tremendously useful to predict materials properties. However, considering the complexities of current thermoelectric materials, the ability to use these calculations in a materials screening or datamining capacity tends to vanish.

Rather than explicitly calculating κ, an alternative approach more conducive to materials screening is to use proxies to estimate thermal conduction. In this case, the minimum thermal conductivity (κ_{min}) in the high temperature limit is an ideal proxy because it indicates compounds that have an inherent potential for extremely low thermal conductivity. This lower-limit of κ provides an indication of what can be achieved experimentally through engineered phonon scattering mechanisms. κ_{min} can be estimated following:

$$\kappa_{\mathrm{min}} \rightarrow 0.87 k_B N_A^{2/3} \frac{N^{2/3}\rho^{1/6}E^{1/2}}{M^{2/3}}$$

where, E is Young's modulus, ρ is the density of the crystal structure, and N and M are the number of atoms and molecular mass of the unit cell [20]. The crystal structure information can be easily determined experimentally while Young's modulus can be readily predicted using density functional perturbation theory in conjunction with the Voigt–Reuss–Hill approximation to determine the elastic moduli [21]. Modern implementations of DFT readily allow this analysis with little post-processing, providing an efficient calculation of the minimum thermal conductivity.

Plotting κ_{min} calculated using the DFT-determined Youngs modulus against the experimentally measured κ for a variety of known thermoelectric materials such as PbTe, SnTe, Mo_3Te_4, TiNiZn, among others, at 300 K (Fig. 2a) in general shows positive agreement between experiment and computation even though it is in poor agreement on an absolute scale. The κ measured at 1000 K (Fig. 2b) is in much better agreement with κ_{min} with a majority of the calculated data points falling near the measured values. The obvious reason for this discrepancy stems from the κ_{min} corresponding to the high temperature limit where Umklapp scattering decreases the lattice thermal conductivity in even defect-free materials. Interestingly, the thermal conductivity is still underestimated indicating that data beyond 1000 K are necessary to achieve the high-temperature limit for many of these compounds. Data at such high temperatures are not