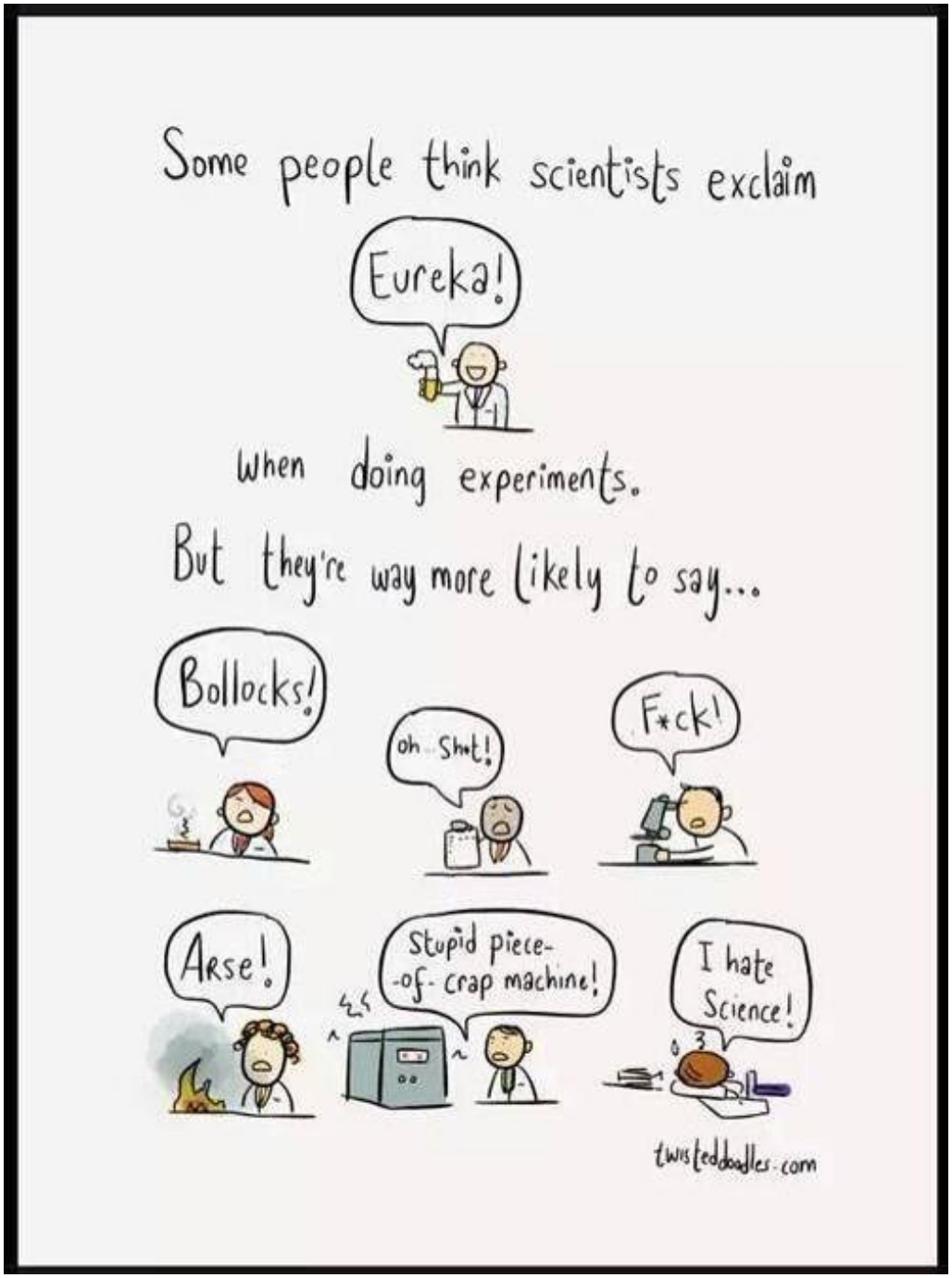

Some people think scientists exclaim
Eureka!
when doing experiments.
But they're way more likely to say...
Bollocks!
oh Shit!
F*ck!
Arse!
Stupid piece-of-crap machine!
I hate Science!
twistedoodles.com

Computationally Guided Materials Discovery

High-throughput Computations, Softwares and Databases, Experimental Feedback

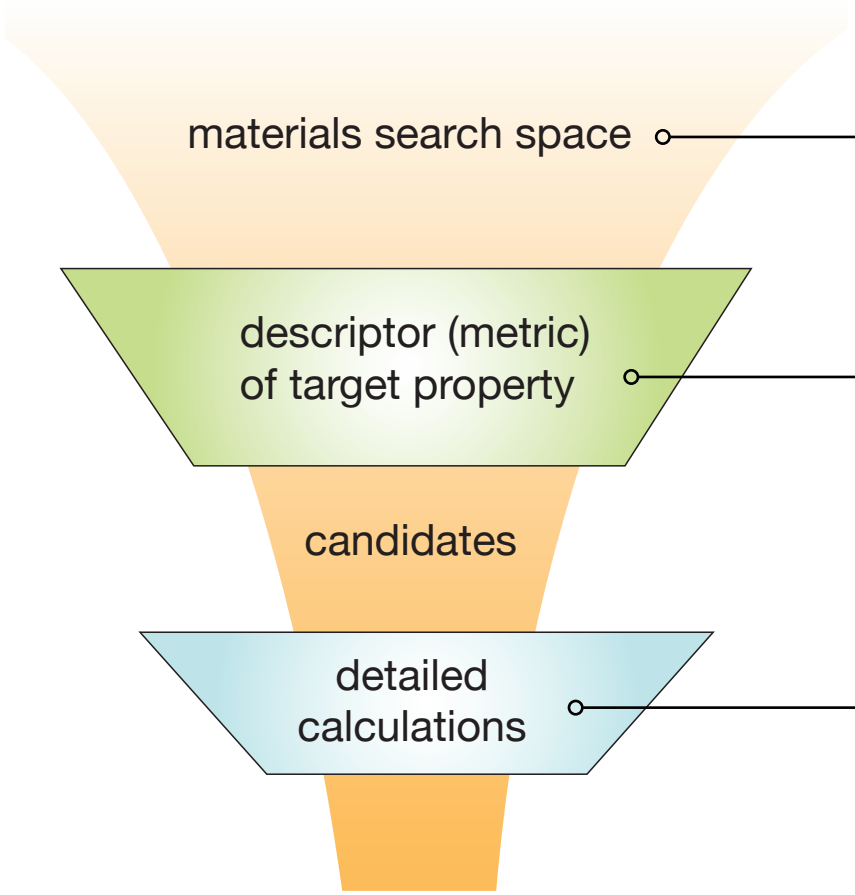
Prashun Gorai

National Renewable Energy Laboratory

Colorado School of Mines

www.prashungorai.org/imrc2017

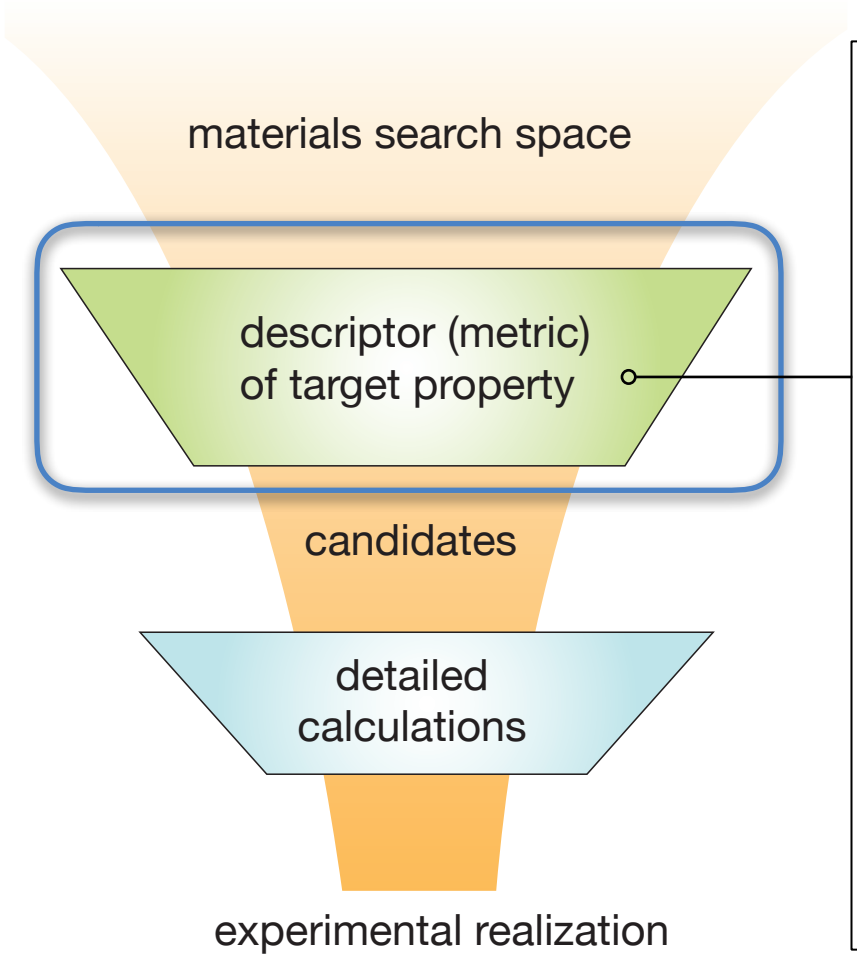
High-throughput Computations: General Down-selection Approach



- Define search space (known materials, hypothetical structures, alloy compositions)
- Develop a descriptor/metric to quantify the target property (e.g. a figure of merit)
  - Approximate but reliable
  - Computationally tractable
- Perform higher accuracy or detailed calculations
  - Phase stability, defect calculations etc.
- Eureka! (hopefully)

Seemingly straightforward, practical implementation non-trivial

Developing a Robust, Quantitative Descriptor is Challenging



- A descriptor is a quantitative metric for the desirable material properties

1. Can be a set of material properties

   example: transparent conducting oxides

   large band gap (transparent), high carrier mobility

2. Can be a figure of merit

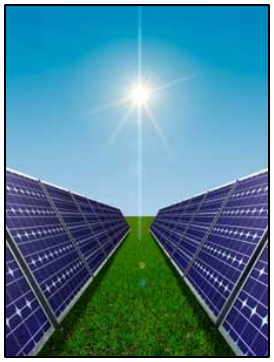
   example: thermoelectric material

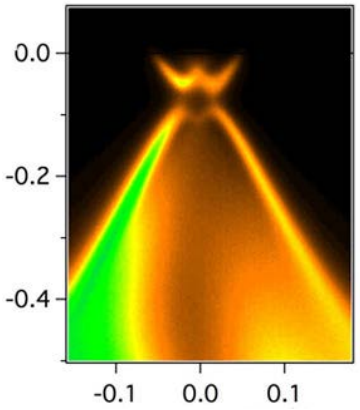
   thermoelectric figure of merit zT

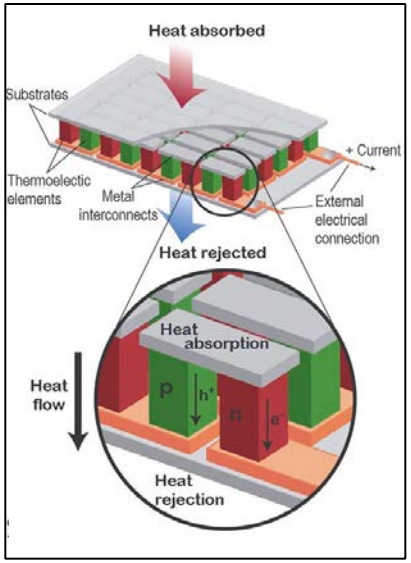
Calculation of the descriptor should also be computationally tractable

Examples of Materials Discovery Enabled by HT Computations

PV, topological insulators, batteries, thermoelectrics, gas storage, water splitting, …







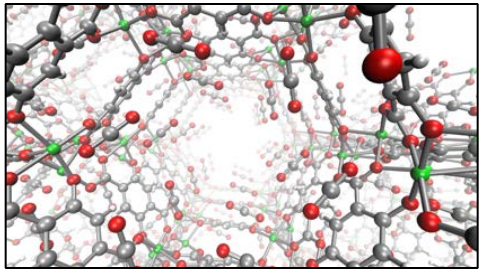
nature materials

REVIEW ARTICLE

PUBLISHED ONLINE: 20 FEBRUARY 2013 | DOI: 10.1038/NMAT3568

The high-throughput highway to computational materials design

Stefano Curtarolo[1,2]*, Gus L. W. Hart[2,3], Marco Buongiorno Nardelli[2,4,5], Natalio Mingo[2,6], Stefano Sanvito[2,7] and Ohad Levy[1,2,8]



Review Article

Computational predictions of energy materials using density functional theory

Anubhav Jain, Yongwoo Shin & Kristin A. Persson

Nature Reviews Materials **1**,
Article number: 15004 (2016)
doi:10.1038/natrevmats.2015.4

Published online: 11 January 2016

High-throughput Computations: Good, Bad and Ugly

The Good

- Large number of materials can be rapidly screened to down select candidates
- Can accelerate materials discovery and learning of new structure-property relationships
- Computational advances (softwares, infrastructure) have been rapid to enable HT computations

The Bad

- Developing a robust descriptor is quite challenging
- Results may depend on the computational method of choice - lack of prediction consistency
- Experimental validation may not be able keep pace with the rate at which predictions are generated

The Ugly

- Without experimental validation, the literature is inundated with nonsensical results
- HT computations can lead to HT errors!

Seemingly straightforward, practical implementation non-trivial

Example: Computationally Guided Discovery of Thermoelectric Materials

1. Urgent need to discover high-performing thermoelectric materials

   computations can accelerate the discovery

2. Edisonian trial-and-error approaches have resulted in limited success

   computationally guided discovery may be successful

3. Descriptor of TE performance is complicated due to contraindicated properties

   computationally guided discovery may be successful

Monday August 21, 10:00 (Tulum F)

Gorai, Stevanovic, Toberer, Nature Reviews Materials, **2**, 17053 (2017)

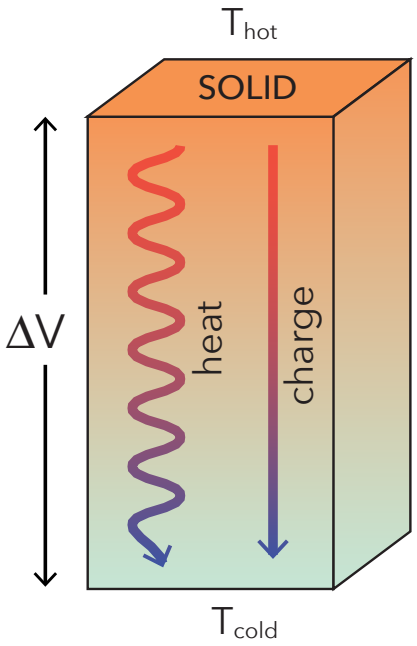
Outline

- Break (2.5 min), Primer on thermoelectrics (1)
- Why is computational search of thermoelectric materials challenging? (2)
- Boltzmann transport theory (1)
- Descriptors of thermoelectric performance (8+break+ 6)
  - Theory, examples, limitations, improvements
- Challenges: dopability (4)
- Break (2.5 min), Softwares for HT computations (2)
- Open-access databases (3)
- Role of data informatics (1)
- Role of experimental validation (1)
- Outlook (1), Q&A (10 min)

Questions?


i wonder if this will bounce
theycantalk.com
maybe that one will

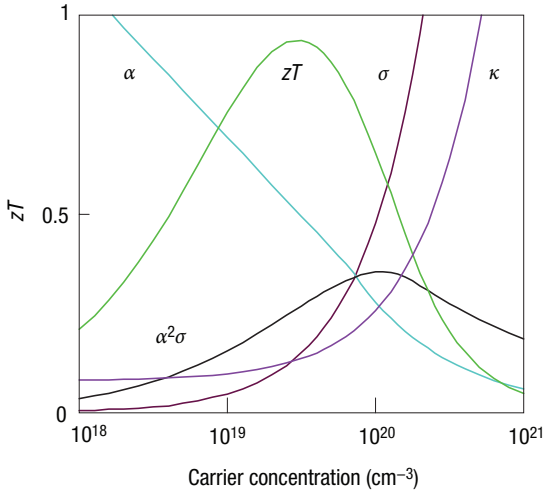
Primer on Thermoelectrics



Thermoelectric Figure of Merit

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

- Complex function of
  - Intrinsic material properties
  - Temperature
  - Carrier concentration (doping)



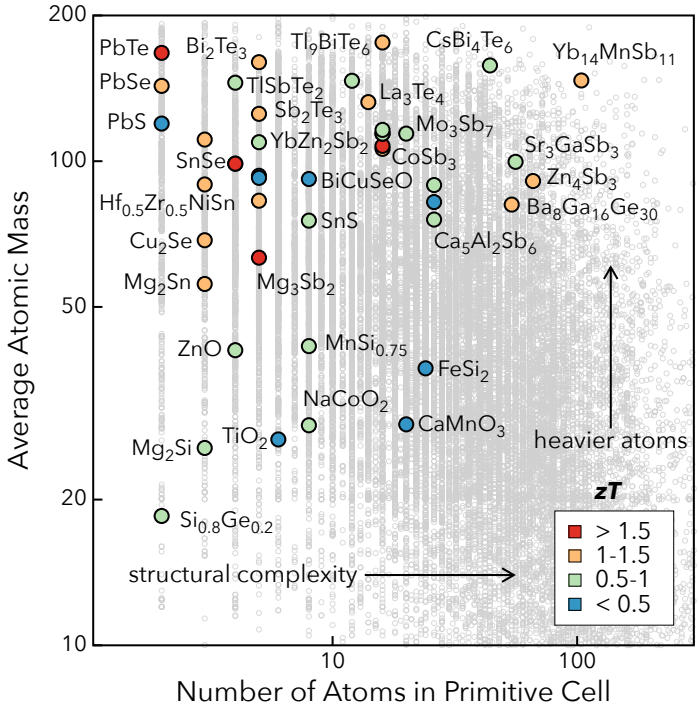
- Advantages
  - Wide range of waste heat sources, enormous potential
  - Solid-state, highly-scalable, reliable
- Challenges
  - Device conversion efficiencies are low
  - Discovery of new materials challenging because of contraindicating properties

Materials discovery key to advancing thermoelectric technology

Gorai, Stevanovic, Toberer, Nature Reviews Materials, **2**, 17053 (2017)

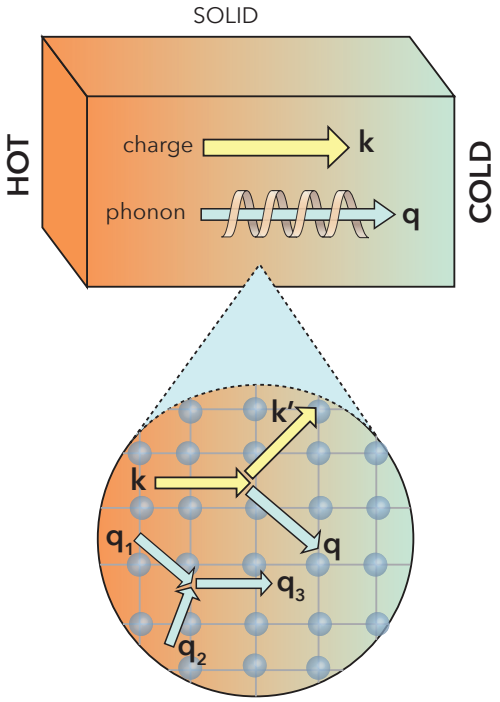
TE Materials Discovery is Complex: Well Suited for Computations

Search space is chemically, structurally diverse



~40,000 metal-non metal compounds from ICSD

Contraindicated transport properties



Computations can guide discovery of new TE materials, identify new structure-property relations

Where do we look for great TE in the vast chemical space?

crystal structure databases

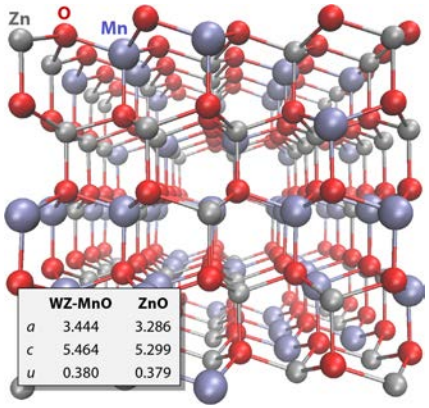
FIZ Karlsruhe ICSD

Cambridge Crystallographic Data Centre

COD

- Inorganic crystal structure database, Crystallography open database, Pearson's crystal data, Cambridge structural database
- Predominantly stoichiometric, ordered compounds
- Majority previously synthesized

alloys



- Varying alloy compositions between known materials
- Represents a continuum search space - infinite possibilities
- Present computational methods not efficient for alloys

hypothetical materials

USPEX Computational Materials Discovery

CALYPSO

- Exciting, unchartered chemical space
- Structure prediction tools are emerging
- Experimental validation is necessary

What is Needed to Predict Thermoelectric Performance?

- Transport properties derived within the Boltzmann transport theory

Electrical Conductivity

$$\sigma = \int \left(-\frac{\partial f_0}{\partial \epsilon}\right) \Sigma(\epsilon) d\epsilon$$

Seebeck Coefficient

$$\alpha = \frac{1}{T\sigma} \int \left(-\frac{\partial f_0}{\partial \epsilon}\right) \Sigma(\epsilon)(\epsilon - \mu) d\epsilon$$

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

Electronic Thermal Conductivity

$$\kappa_0 = \frac{1}{T} \int \left(-\frac{\partial f_0}{\partial \epsilon}\right) \Sigma(\epsilon)(\epsilon - \mu)^2 d\epsilon$$

$$\kappa_e = \kappa_0 - T\sigma\alpha^2$$

Lattice Thermal Conductivity

$$\kappa_L = \sum_{n,q} [v_n(q).i]^2 \tau_n(q) C_n(q)$$

$$\Sigma(\epsilon) = \sum_{n,k} [v_n(k).i]^2 \tau_n(k) \delta(\epsilon - \epsilon(k))$$

Need to compute: dispersions (electron, phonon) and scattering relaxation times

HT Search Strategies and Descriptors of Thermoelectric Performance

- Constant relaxation time approximation (CRTA)
  - Descriptor focuses only on electronic transport
- Constant mean free path approximation (CMFP)
  - Descriptor focuses only on electronic transport
- Semi-empirical descriptor
  - Descriptors accounts for both electron and phonon transport

theory ⟶ examples ⟶ limitations ⟶ improvements

1. Constant Relaxation Time Approximation (CRTA)

Electrical Conductivity

$$\sigma = \int \left(-\frac{\partial f_0}{\partial \epsilon}\right) \Sigma(\epsilon) d\epsilon$$

Seebeck Coefficient

$$\alpha = \frac{1}{T\sigma} \int \left(-\frac{\partial f_0}{\partial \epsilon}\right) \Sigma(\epsilon)(\epsilon - \mu) d\epsilon$$

Transport Distribution Function

$$\Sigma(\epsilon) = \sum_{n,k} [v_n(k).i]^2 \tau_n(k) \delta(\epsilon - \epsilon(k))$$

- In this approach, the relaxation time (τ) is assumed to be constant
  - Electrical conductivity scales linearly with τ, Seebeck is τ-independent

- Reduced power factor ($\alpha\sigma^2/\tau$) is the descriptor for thermoelectric performance
  - Large power factors are desirable for good thermoelectric performance
  - Effect of thermal conductivity is often overlooked

Electronic part of zT is evaluated within CRTA

1. Softwares for Solving Boltzmann Transport Equations

BoltzTraP. A code for calculating band-structure dependent quantities ☆

Georg K.H. Madsen [a,*], David J. Singh [b]

[a] *Department of Chemistry, University of Aarhus, DK-8000 Århus C, Denmark*
[b] *Condensed Matter Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6032, USA*

Received 25 September 2005; accepted 3 March 2006

Available online 2 May 2006

- One of the most widely used softwares for calculation of transport coefficients from first-principles DFT calculations
- Other similar softwares:
  - BoltzWann (use localized Wannier functions)[1]
  - LanTrap (within Landauer formalism)[2]
  - ShengBTE (solves phonon Boltzmann transport equations)[3]

[1] Pizzi *et al.*, Comp. Phys. Comm. **185**, 422 (2014)

[2] Lundstorm *et al*. https://nanohub.org/resources/lantrap (2014)

[3] Li *et al*., Comp. Phys. Comm., **185**, 147 (2014)

1. Example: Search for Antimonide Thermoelectrics

J|A|C|S
ARTICLES
Published on Web 08/25/2006

Automated Search for New Thermoelectric Materials: The Case of LiZnSb

Georg K. H. Madsen

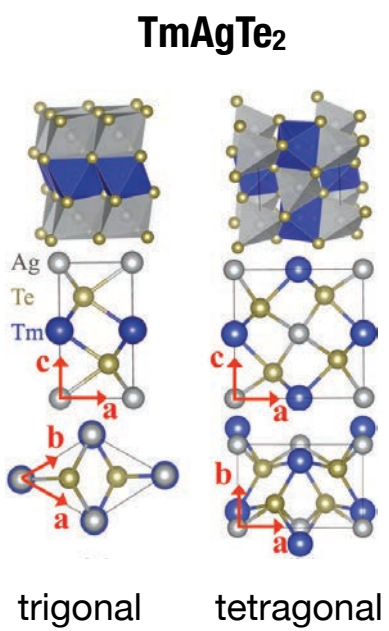
Contribution from the Department of Chemistry, University of Aarhus, DK-8000 Århus C, Denmark

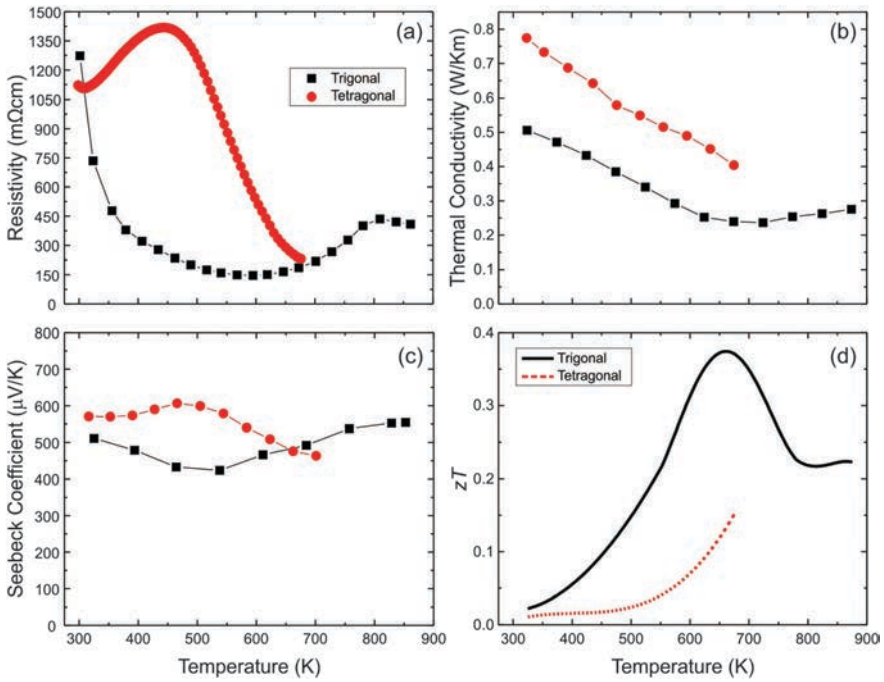
Received April 12, 2006; E-mail: georg@chem.au.dk

- One of the earliest applications of the CRTA approach to search for new TE materials
  - Screened 570 Sb-containing compounds from the ICSD
  - Assumed fixed values of τ and κ_L for all compounds
  - LiZnSb identified as promising candidate - experiments could not confirm
  - Interesting candidates (Zr_5ZnSb_3, NaZnSb) remain unverified

Madsen, J. Amer. Chem. Soc., **128**, 12140 (2006)

1. Example: XYZ_2 Thermoelectrics

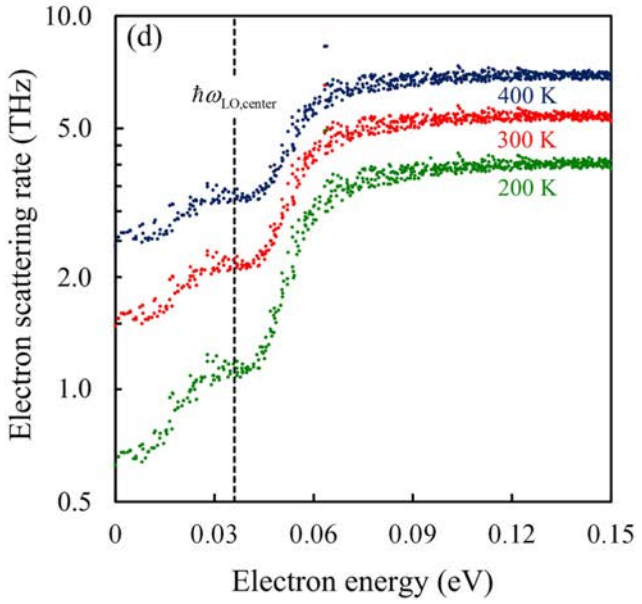


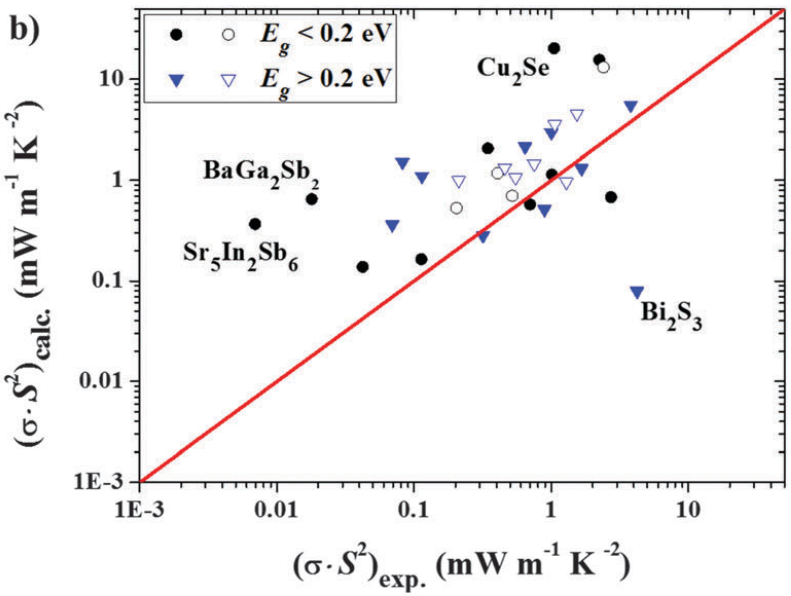


- Search identified XYZ_2 chemistry as a promising family of thermoelectric materials
  - Assumed constant τ for all compounds
  - Calculated minimum κ_L in the amorphous limit
  - zT of ~0.35 achieved for $TmAgTe_2$, carrier concentration tuning was challenging

Zhu *et al.*, J. Mater. Chem. C, **3**, 10554 (2015)

1. Limitations of Constant Relaxation Time Approximation





- Constant relaxation time is a very crude approximation
  - Scattering relaxation time (~ 1/rate) is highly energy- and temperature-dependent
  - In general, power factors are overestimated
  - Thermal conductivity is ignored!

Liu *et al.*, Phys. Rev. B, **95**, 075206 (2017), Chen *et al*., J. Mater. Chem. C, **4**, 4414 (2016)

2. Constant Mean Free Path Approximation (CMFPA)

Electrical Conductivity

$$\sigma = \int \left(-\frac{\partial f_0}{\partial \epsilon}\right) \Sigma(\epsilon) d\epsilon$$

Seebeck Coefficient

$$\alpha = \frac{1}{T\sigma} \int \left(-\frac{\partial f_0}{\partial \epsilon}\right) \Sigma(\epsilon)(\epsilon - \mu) d\epsilon$$

Transport Distribution Function

$$\Sigma(\epsilon) = \sum_{n,k} [v_n(k).i]^2 \tau_n(k) \delta(\epsilon - \epsilon(k))$$

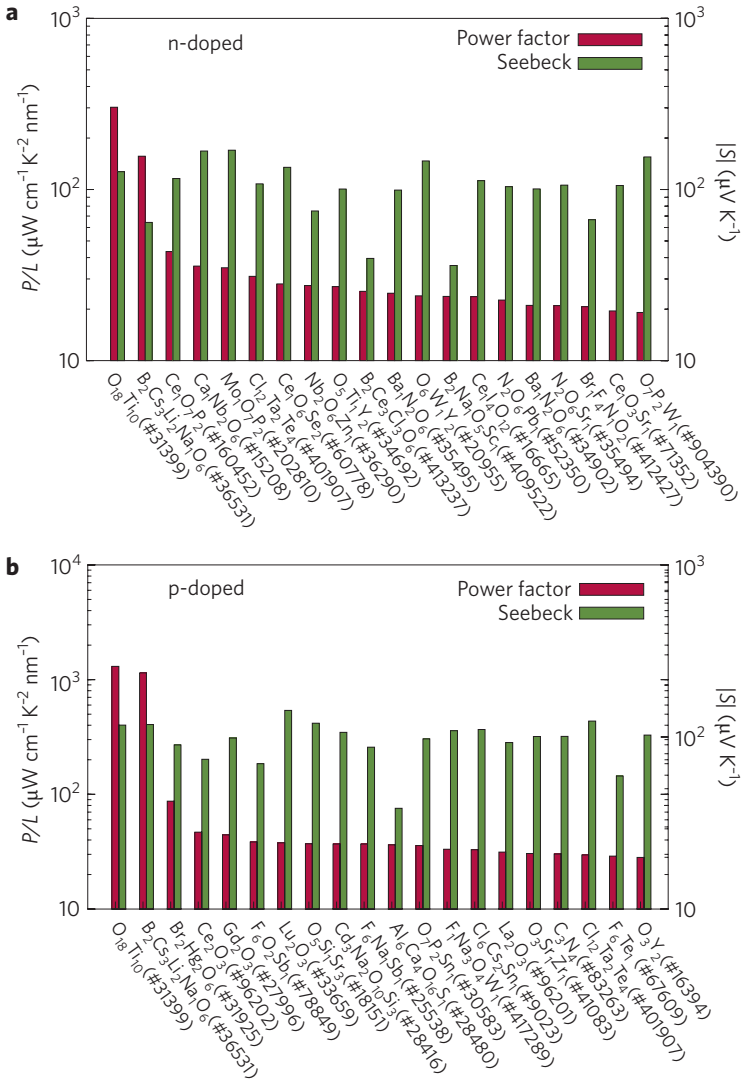
$$v(E)\tau(E) = \lambda$$

- In this approach, the relaxation time (τ) is now energy-dependent
  - Materials have to be nanostructured to ensure λ is limited by grain size

- Reduced power factor ($\alpha\sigma^2/\lambda$) is the descriptor for thermoelectric performance
  - Effect of thermal conductivity is overlooked

Electronic part of zT is evaluated within CMFPA

2. Example and Limitations



- HT screening of ~2500 materials from the ICSD
  - TiO_2 Magneli phases and $Cs_3Li_2NaB_2O_6$ identified as promising candidates
  - Candidates remain experimentally unverified
  - Well-known thermoelectrics not identified

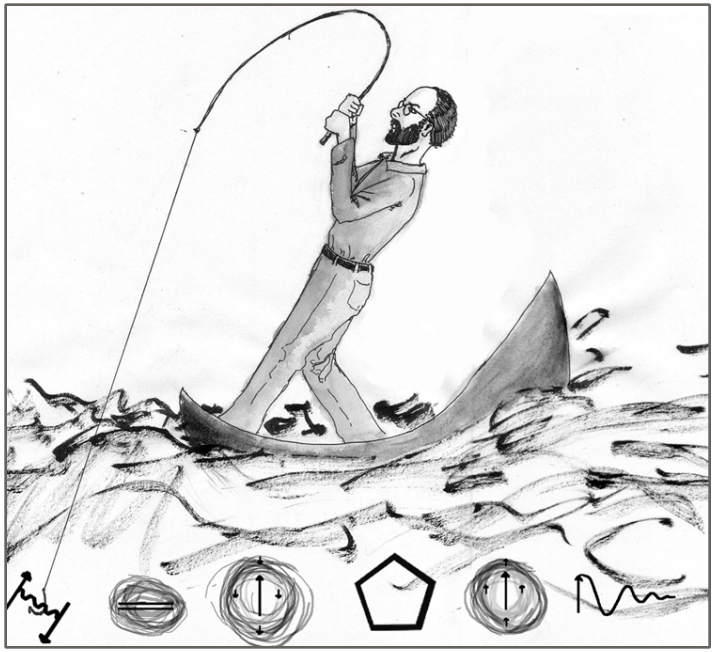
- Limitations: mean free path assumed constant for all materials
  - Some materials have to be nanostructured below 5 nm
  - Thermal stability of nanostructured materials is an issue

Questions?

PERSPECTIVE | FOCUS

Fishing the Fermi sea

PAUL C. CANFIELD



The scientist who is skilled in the art of new materials design and discovery will develop a variety of strategies for identifying promising spots to fish for new materials and ground states.

P. C. Canfield, Nature Physics **4**, 167 (2008)

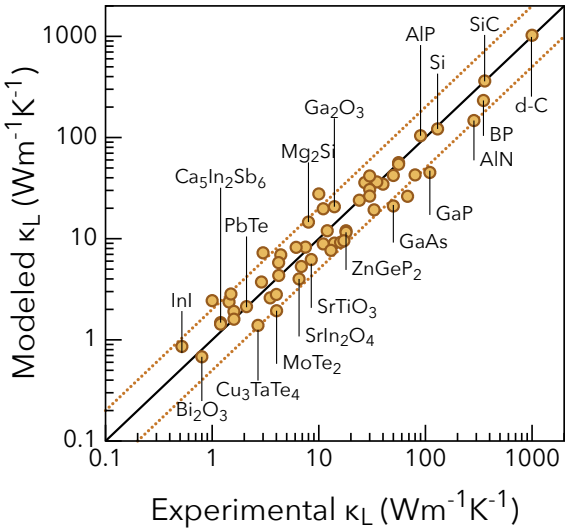
3. Semi-empirical Thermoelectric Quality Factor

$$zT = \frac{\alpha^2 \sigma T}{(\kappa_L + \kappa_e)} \longrightarrow zT = \frac{u\beta}{(v\beta + 1)} \longrightarrow \beta = \frac{2e}{\hbar^3}\left(\frac{k_B}{e}\right)^2 \left(\frac{k_B}{2\pi}\right)^{3/2} \frac{\mu_0 m^*_{DOS}{}^{3/2}}{\kappa_L} T^{5/2}$$

constant | core | T-dep

Lattice thermal conductivity

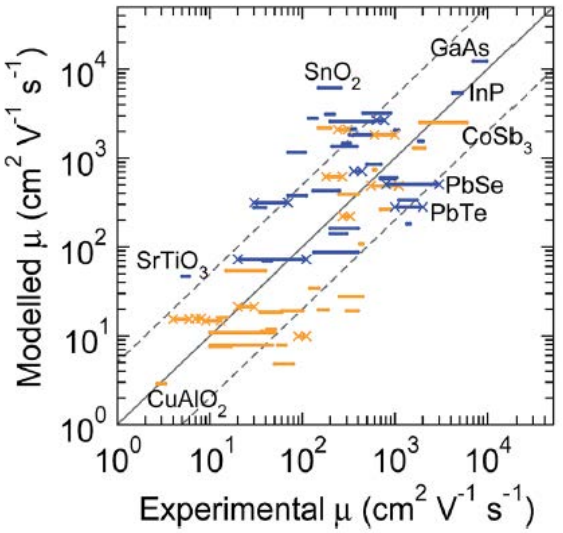
$$\kappa_L = A_1 \frac{\bar{M} v_s^y}{T\gamma^2 V^z n^x} + \frac{3k_B}{2}\left(\frac{\pi}{6}\right)^{1/3} \frac{v_s}{V^z}\left(1 - \frac{1}{n^{2/3}}\right)$$



Miller, Gorai, Stevanovic, Toberer *et al*, *Chem. Mat.* **29**, 2494 (2017)

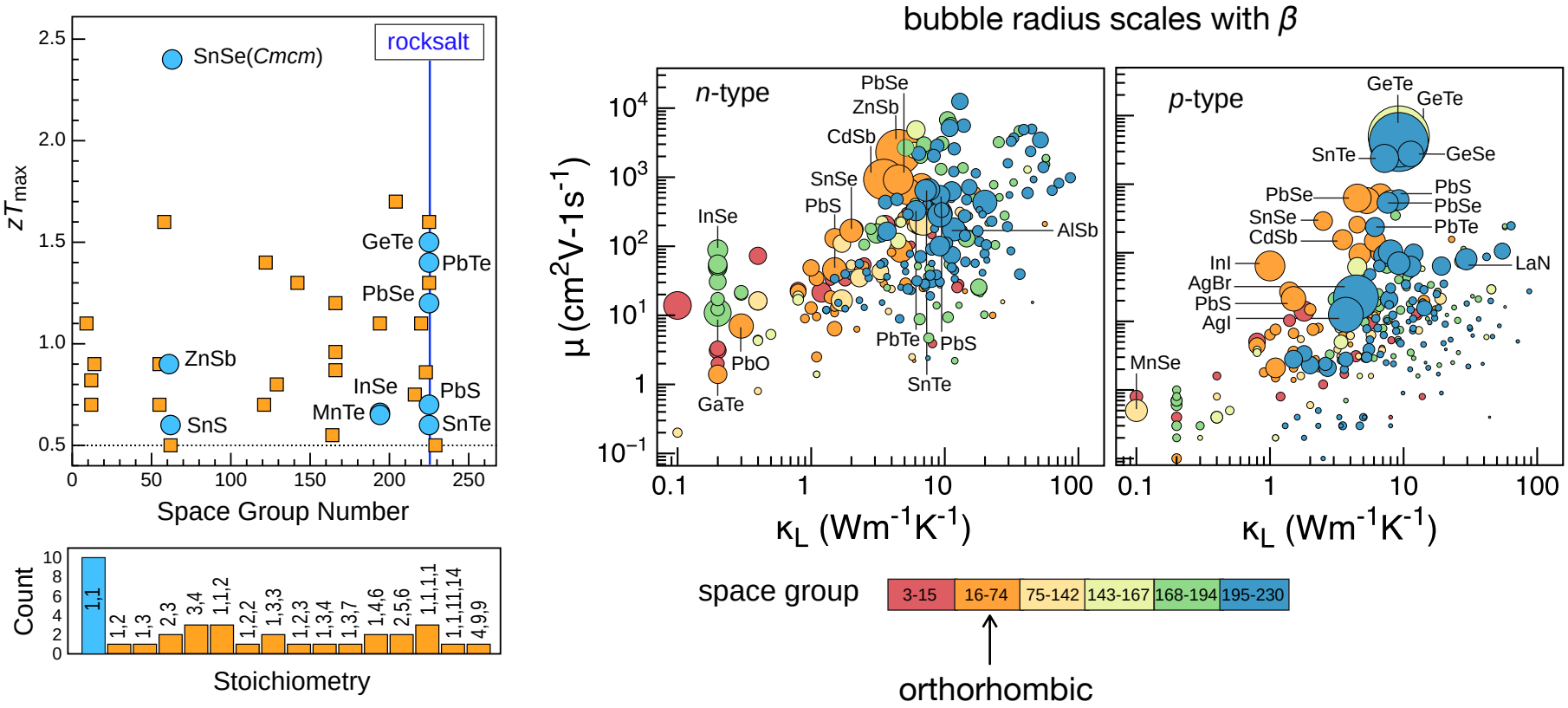
Charge carrier mobility

$$\mu_0 = A_0 (B)^s (m_b^*)^{-t}$$



Yan, Gorai, Stevanovic, Toberer *et al*, Energy Environ. Sci. **8**, 983 (2015)

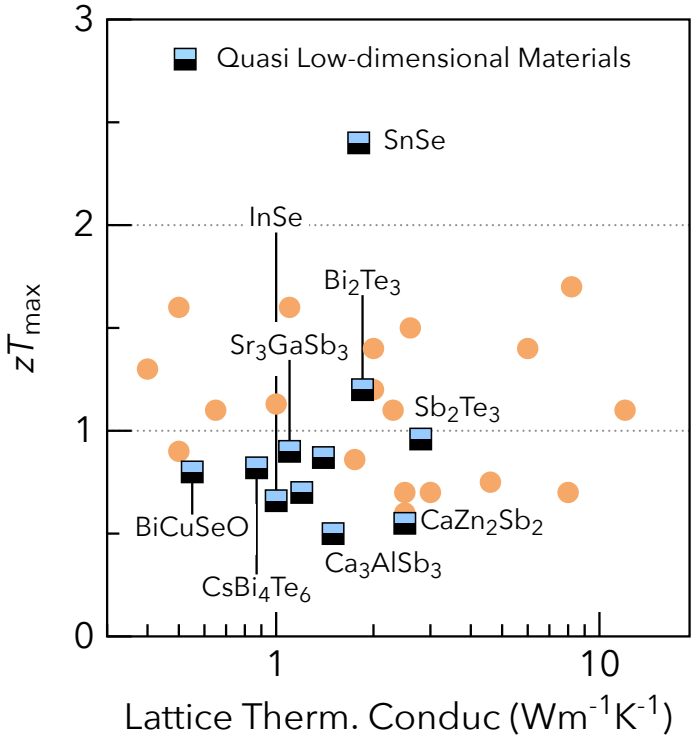
3. Example: Search for New Binary A_1B_1 Thermoelectric Materials



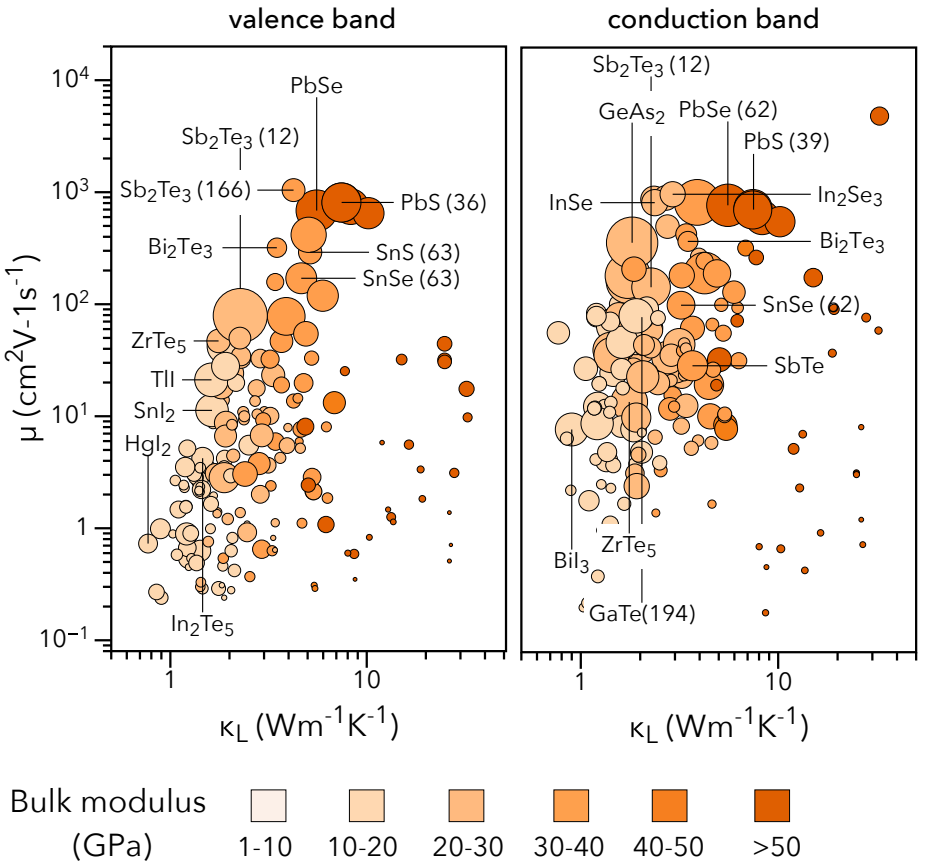
- Nearly 30% of known thermoelectric materials are A_1B_1 compounds
- Known thermoelectric materials are correctly identified with β
- Identified new candidate materials; independent re-discovery (*n*-type SnSe, zT = 2.2)

Gorai, Toberer, Stevanovic *et al.*, Chem. Mater. **18**, 6213 (2015), Cho *et al.* Nature Comm. **7**, 13713 (2016)

3. Example: Computational Identification of Quasi-2D TE Materials



Screening of 427 binary quasi-2D layered materials



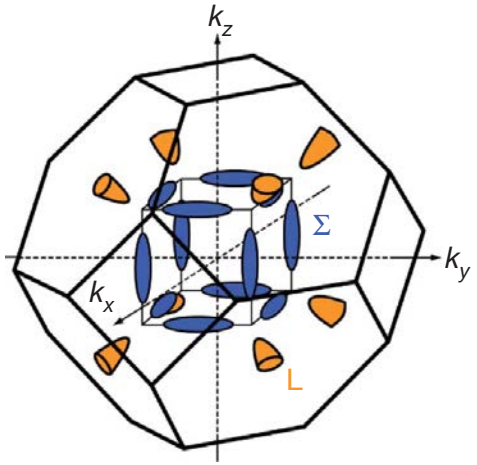
- Quasi low-dimensional structures are abundant among TE materials

Gorai, Toberer and Stevanovic, J. Mater. Chem. A **4**, 11110 (2016)

3. Semi-empirical Descriptor: Limitations

- By design, semi-empirical models will not capture phenomena not represented in the learning dataset
  - Can be improved by including more experimental data in learning dataset
- Assumption: transport is isotropic
- Assumption: phonon-phonon interactions are harmonic (lattice thermal conductivity)

Fermi surface: PbTe

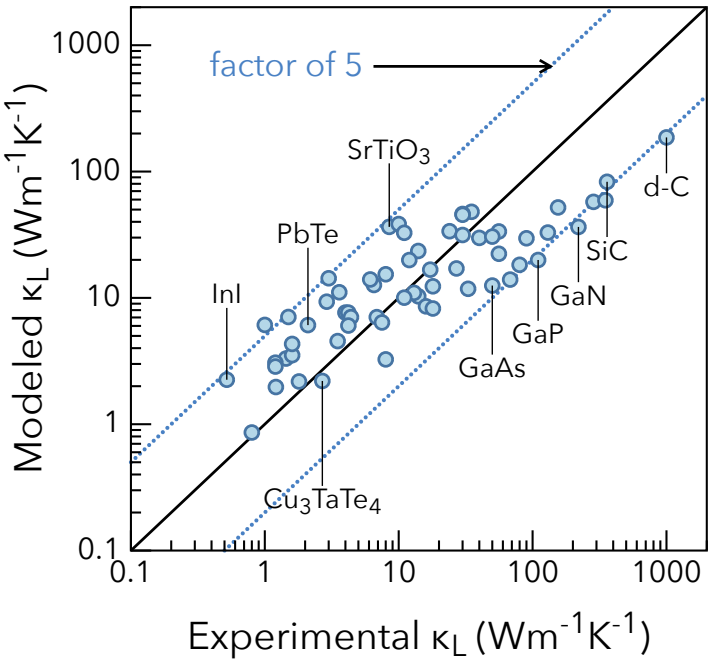


valence band (orange),
conduction band (blue)

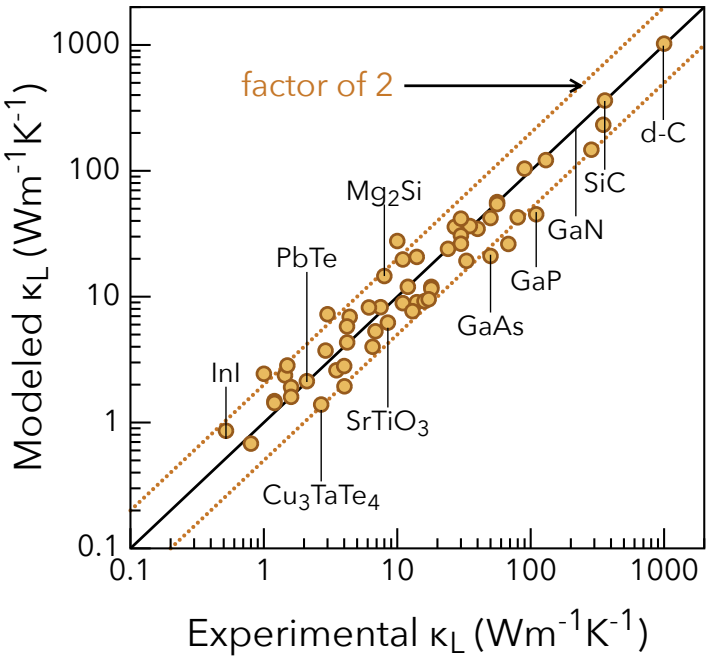
Pei *et al.*, Nature Lett. **473**, 66 (2011)

3. Descriptor Improvements: Incorporating Anharmonicity in κ_L

Original Model



Improved Model



$$\kappa_L = A_1 \frac{\bar{M} v_s^3}{V^{2/3} n^{1/3}} + A_2 \frac{v_s}{V^{2/3}} \left(1 - \frac{1}{n^{2/3}}\right)$$

$$\kappa_L = A_1 \frac{\bar{M} v_s^y}{T \gamma^2 V^z n^x} + \frac{3 k_B}{2} \left(\frac{\pi}{6}\right)^{1/3} \frac{v_s}{V^z} \left(1 - \frac{1}{n^{2/3}}\right)$$

Incorporating coordination as a measure of anharmonicity, improves prediction

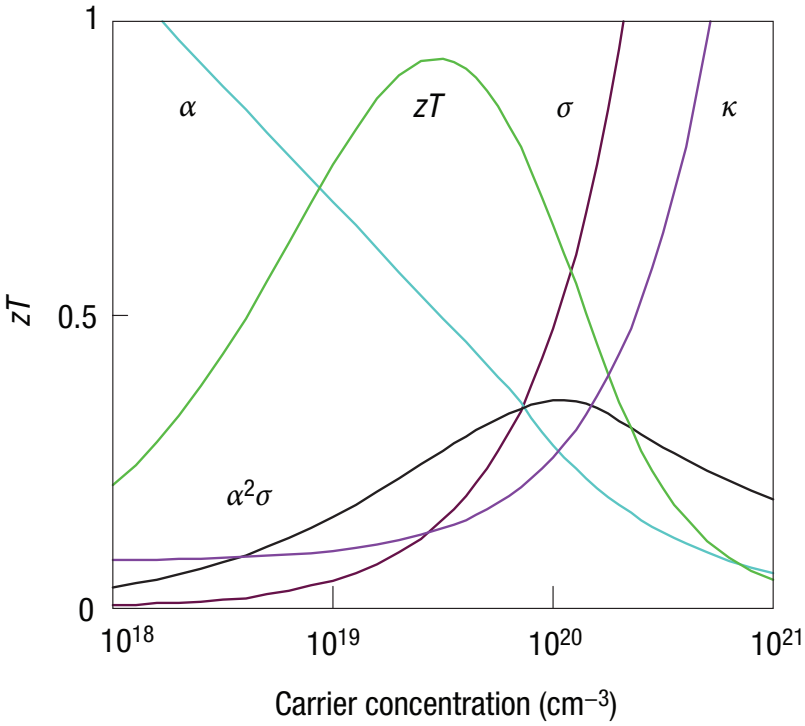
Miller, Gorai, Stevanovic, Toberer *et al*., Chem. Mater., **29**, 2494 (2017)

Something Missing in Descriptors: Dopability

Thermoelectric Figure of Merit

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

- Complex function of
  - ✔ Intrinsic material properties
  - ✔ Temperature
  - ✘ Carrier concentration (doping)



Rapid a priori prediction of material dopability is difficult

Prediction of Dopability is a Bottleneck in TE Materials Discovery

The Curious Case of LiZnSb (and many others)

J|A|C|S
ARTICLES

Published on Web 08/25/2006

Automated Search for New Thermoelectric Materials: The Case of LiZnSb

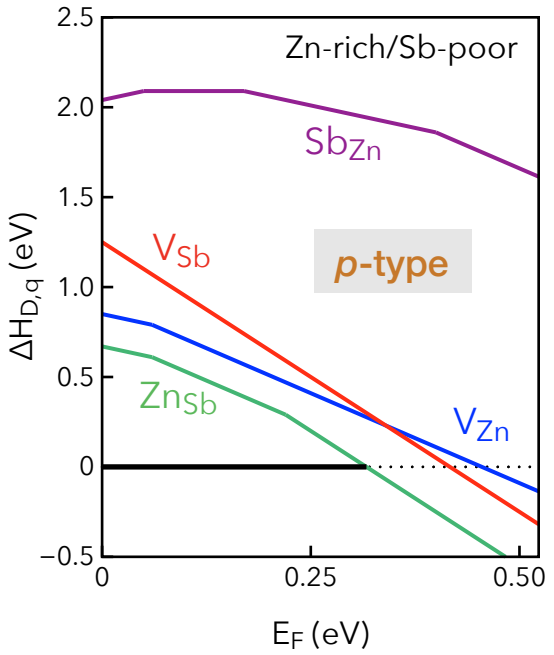
Georg K. H. Madsen

Contribution from the Department of Chemistry, University of Aarhus, DK-8000 Århus C, Denmark

Received April 12, 2006; E-mail: georg@chem.au.dk

***n*-type**

defect calculations



experimental assessment ***p*-type**

JOURNAL OF APPLIED PHYSICS **105**, 063701 (2009)

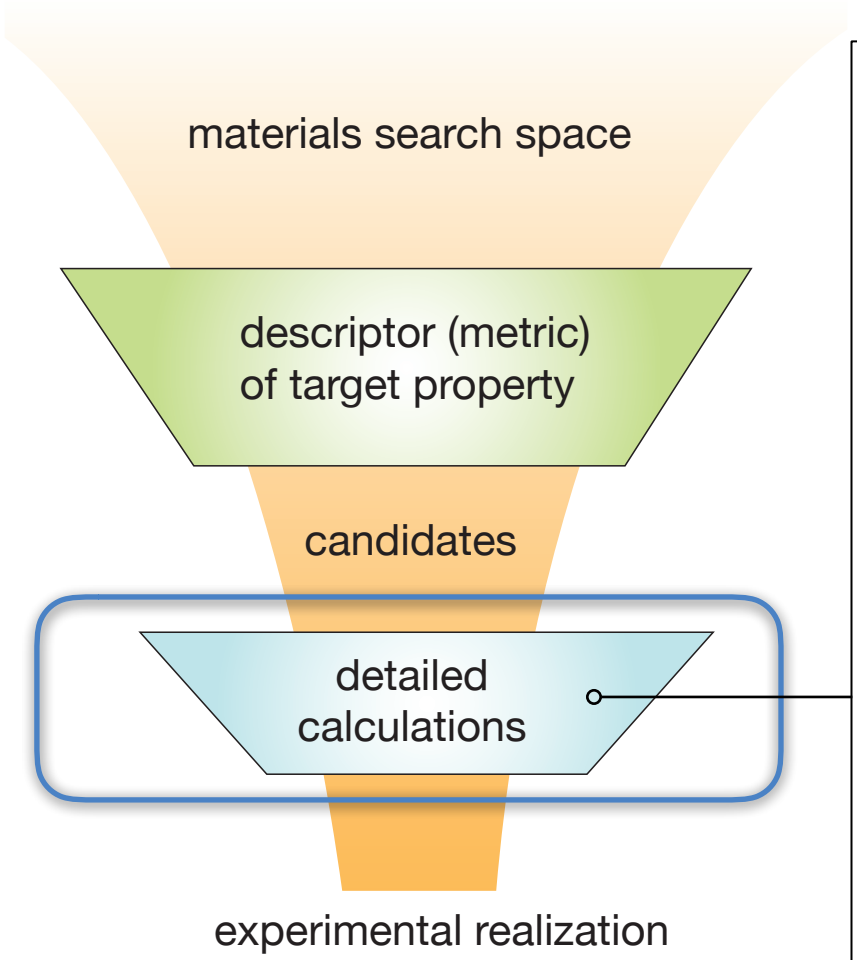
Thermoelectric properties of *p*-type LiZnSb: Assessment of *ab initio* calculations

Eric S. Toberer,[a] Andrew F. May, Cidney J. Scanlon, and G. Jeffery Snyder

Materials Science, California Institute of Technology, Pasadena, California 91125, USA

Material has to be appropriately (type, concentration) doped to realize predicted TE performance

Defect Calculations Can Guide Doping



- First-principles defect calculations performed on down selected materials

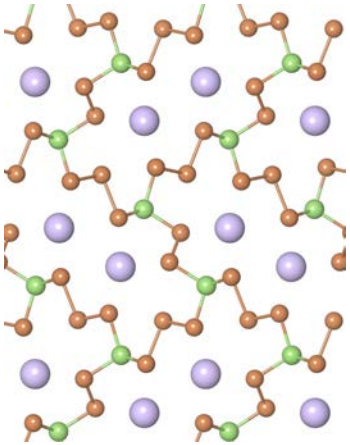
1. Computationally demanding
   - Cannot be performed on large number of materials
   - Analysis requires significant human intervention

2. Softwares to automate defect calculations are emerging
   - Significantly reduces the need for human intervention

Example of Detailed Defect Calculations Guiding Doping

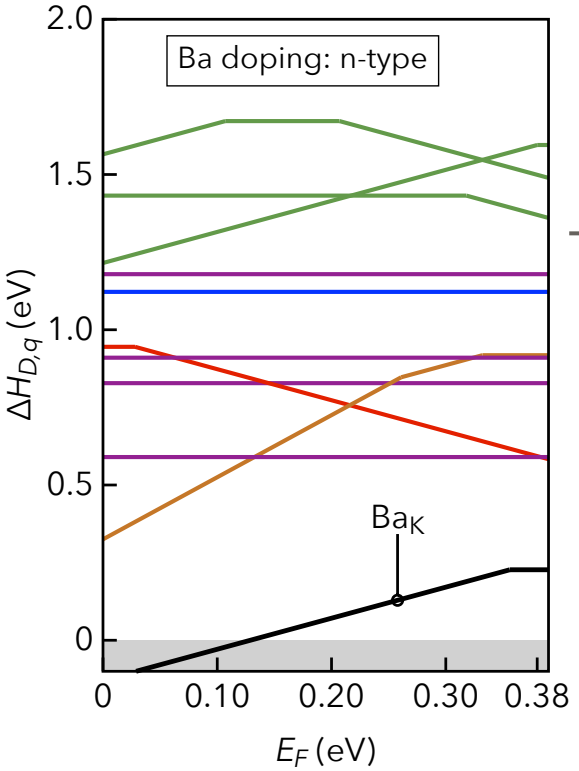
Discovery of new n-type Zintl pnictides

HT search for n-type Zintls

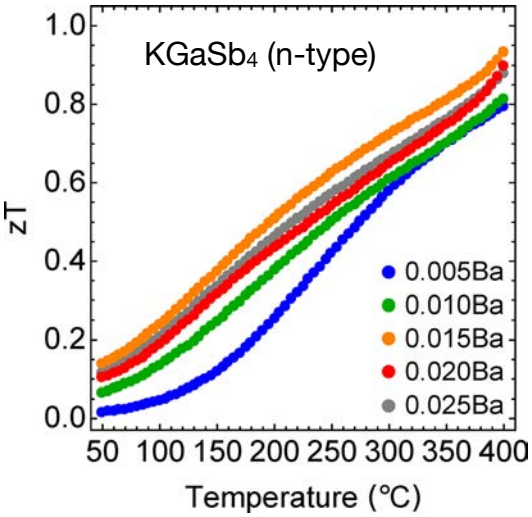


$KGaSb_4$, $KAlSb_4$

Ba is an effective donor

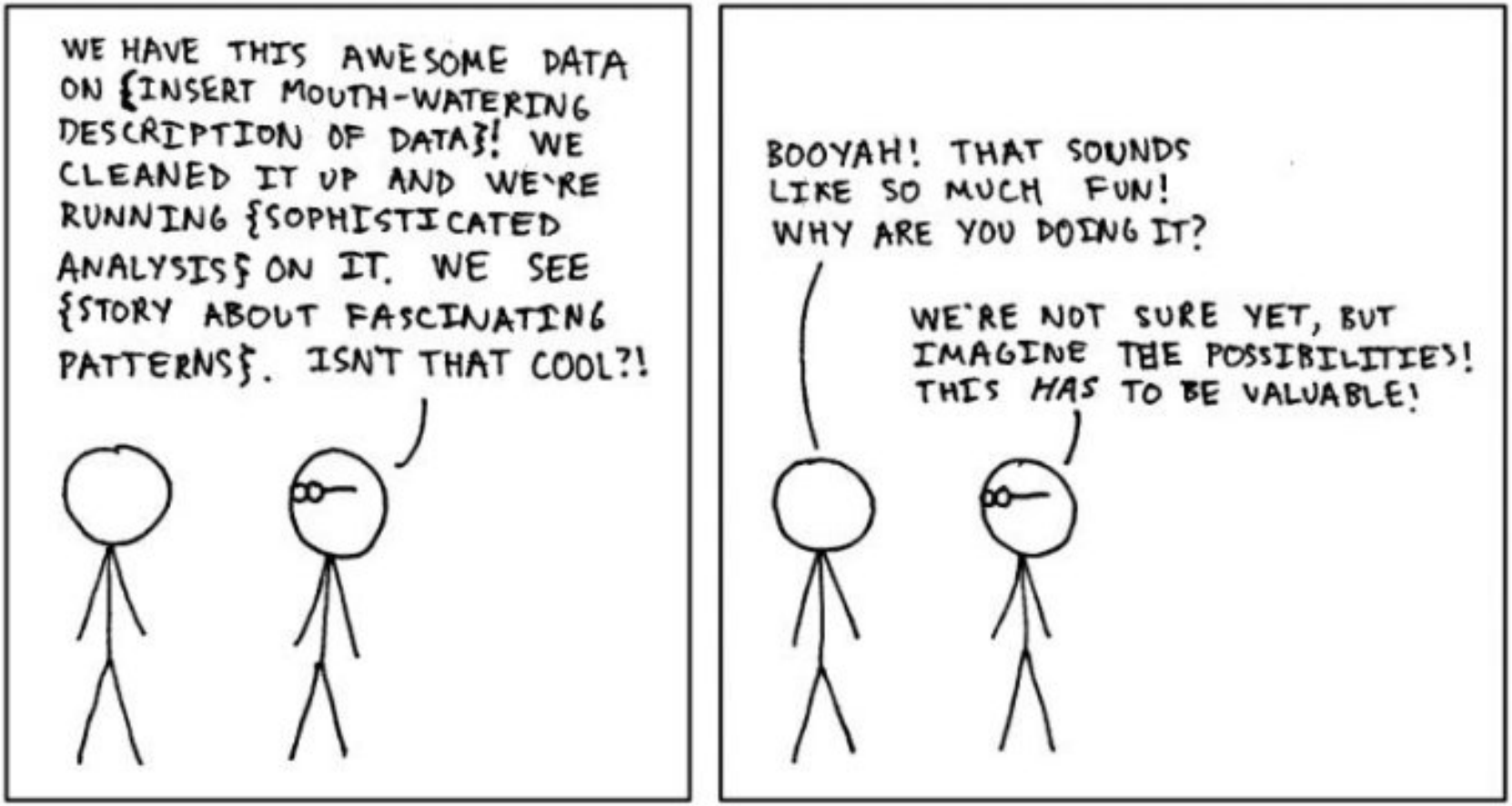


zT~1 realized



Ortiz, Gorai, Stevanovic, Toberer, *J. Mater. Chem. A* **5**, 5036 (2017), *Chem. Mater.* **29**, 4523 (2017)

Questions?


WE HAVE THIS AWESOME DATA ON {INSERT MOUTH-WATERING DESCRIPTION OF DATA}! WE CLEANED IT UP AND WE'RE RUNNING {SOPHISTICATED ANALYSIS} ON IT. WE SEE {STORY ABOUT FASCINATING PATTERNS}. ISN'T THAT COOL?!
BOOYAH! THAT SOUNDS LIKE SO MUCH FUN! WHY ARE YOU DOING IT?
WE'RE NOT SURE YET, BUT IMAGINE THE POSSIBILITIES! THIS HAS TO BE VALUABLE!

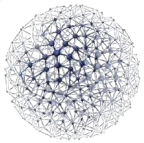
Softwares for High-throughput Computations

pymatgen

www.pymatgen.org

PyLada

github.com/pylada



AFLOW

Automatic - FLOW for Materials Discovery

aflowlib.org

AiiDA

www.aiida.net

- A number of softwares designed for HT computations are now available for free

1. Open source softwares
   - Typically written in Python
   - Highly customizable
   - Thriving community

2. Manage jobs and analyze outputs
   - Interacts with supercomputer job schedulers
   - Error handling
   - Powerful tools for post-calculation analyses

Automated Softwares for Defect Calculations

PyLada Defects

github.com/pylada/pylada-defects

PyCDT

pypi.python.org/pypi/pycdt

www.pythonhosted.org/MAST

- Softwares to automate defect calculations have been developed only recently

1. Set up defect calculations
   - Create defect structures
   - Create input files for running defect calculations

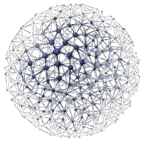
2. Perform post-calculation analyses
   - Finite-size corrections
   - Create phase stability maps
   - Self-consistently calculate carrier concentrations

Role of Open-Access Databases in Accelerating Materials Discovery

MATERIALS PROJECT

NREL NATIONAL RENEWABLE ENERGY LABORATORY

High Performance Computing Center
Materials Database



AFLOW
Automatic-FLOW for Materials Discovery

WOLVERTON
Research Group

AiiDA

MatNavi
NIMS 物質・材料データベース

TE DESIGN LAB

- Several open-access databases provide computational data on:
  1. Phase stability
  2. Electronic properties
  3. Thermal properties
  4. Elastic and piezoelectric properties etc.

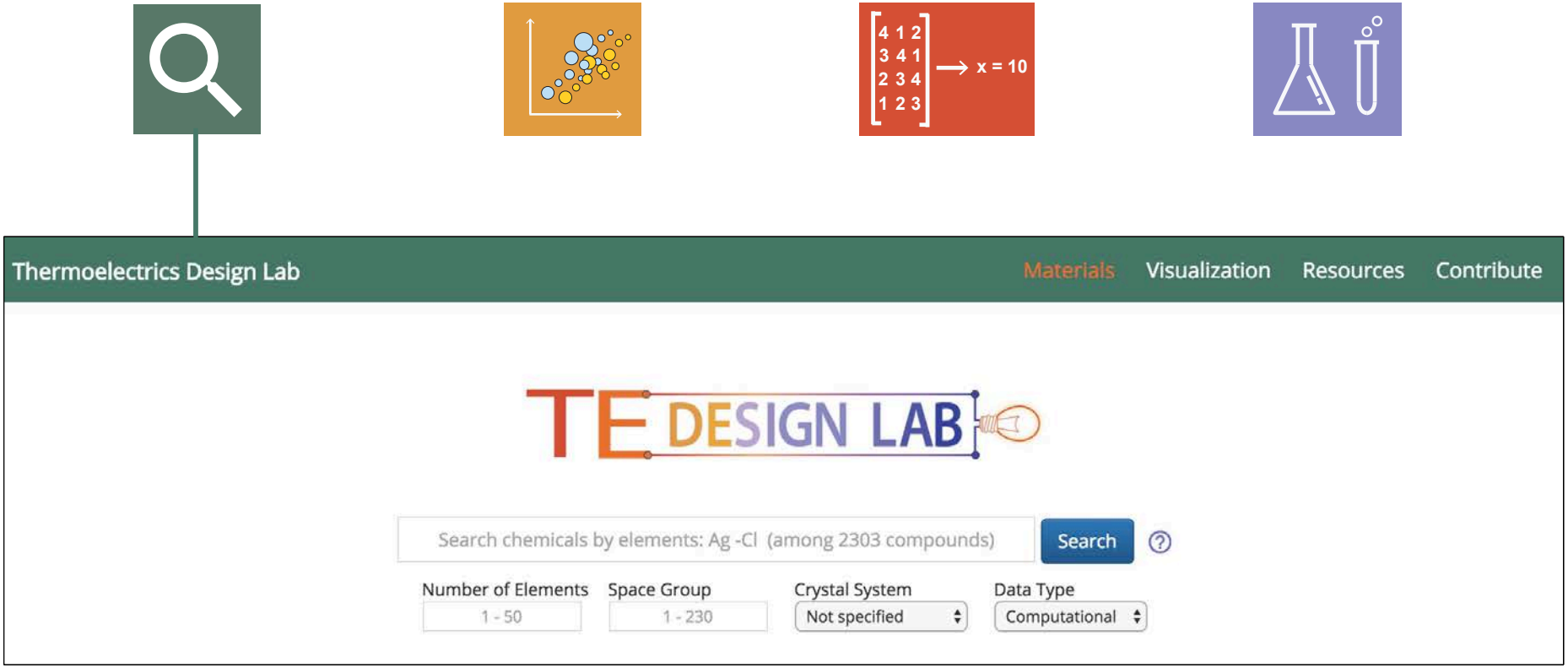
- Data visualization is an important component several databases
  1. Analyze high-dimensional data
  2. Identify new structure-property relations

Example: Thermoelectrics Design Lab

TE DESIGN LAB

www.tedesignlab.org

Materials | Visualization | Resources | Contribute



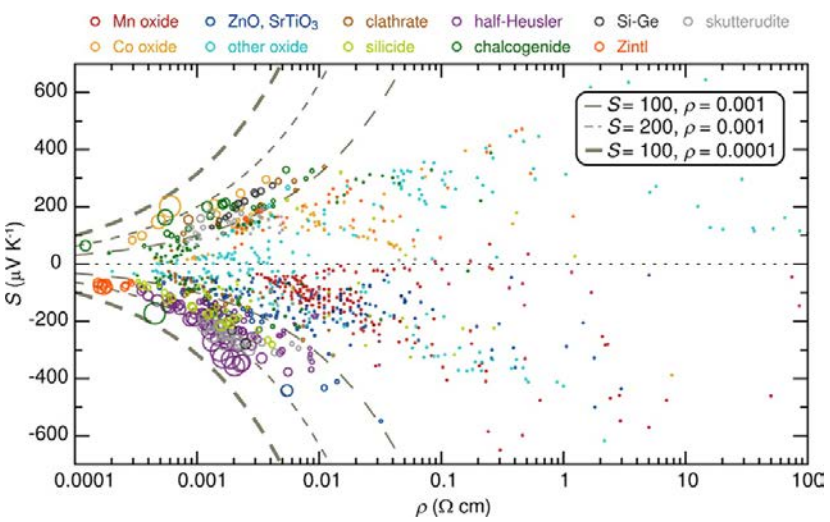
Gorai, Stevanovic, Toberer *et al.*, *Comp. Mat. Sci.* **112A**, 368 (2016)

Live Demonstration of TEDesignLab

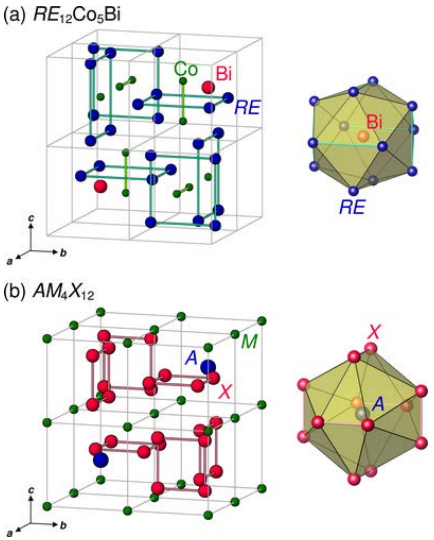
If the internet cooperates …

Role of Data Informatics in Accelerating Materials Discovery

UCSB Experimental Thermoelectrics Database



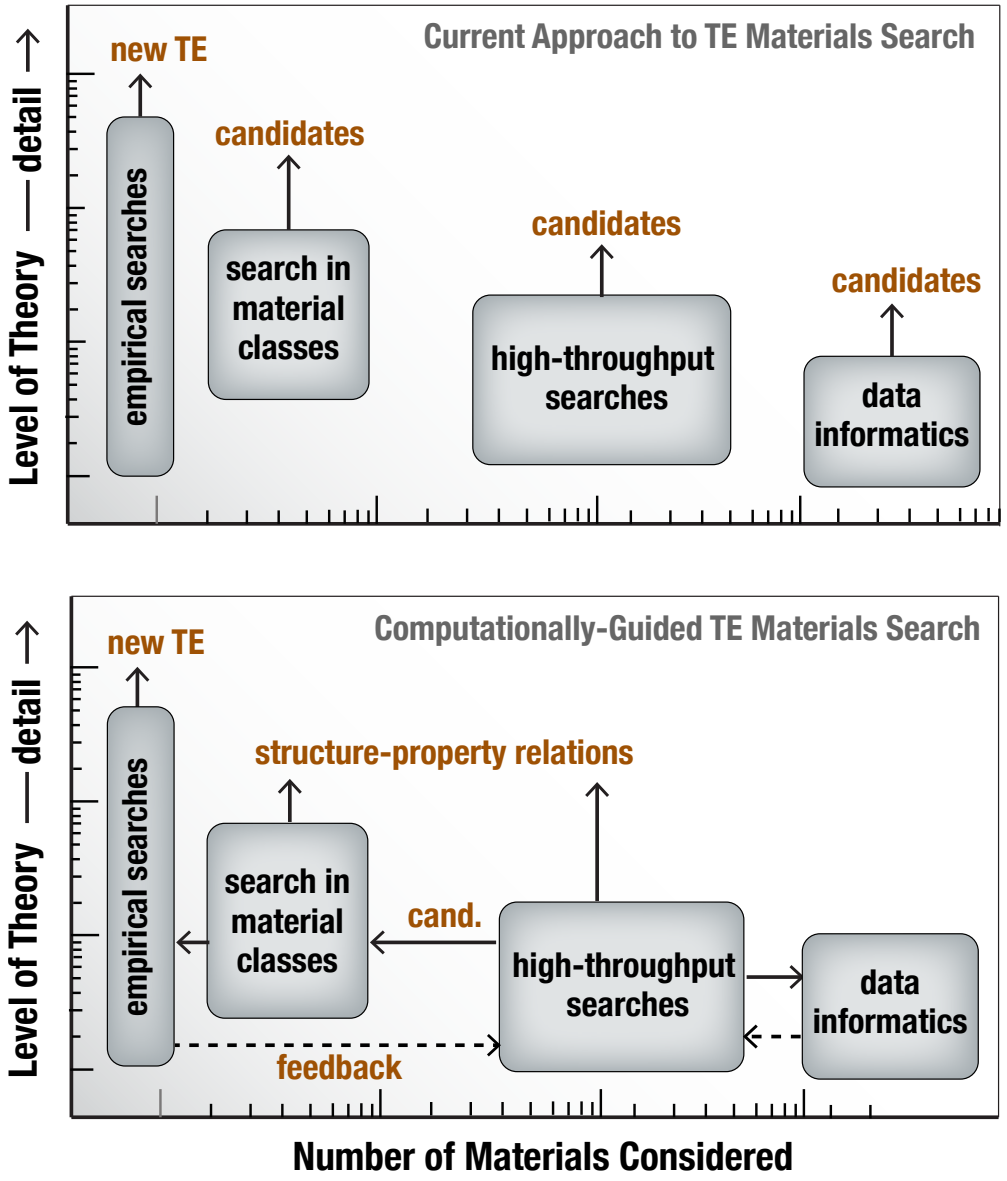
Chemistries identified by recommendation engine



- Large amounts of real and reciprocal space data generated through HT computations
  - Some parameters are experimentally unaccessible
- Data mining can unravel new structure-property relations
- To be robust, even larger amounts of data are needed
  - Chemical and structure phase high dimensional, number of materials small

Gaultois *et al*., Chem. Mater., **25**, 2911 (2013), Gaultois *et al*., APL Mater., **4**, 053213 (2016)

The Need for Experimental Validation, Feedback, and Integration



Outlook

- High-throughput (HT) computations can accelerate materials discovery
- However, HT computations also means there is a chance for HT errors!
- Developing computationally tractable descriptors is the most challenging task
- Present infrastructure is quite sophisticated to enable HT computations
- Experimental validation is critical to the success of computational materials discovery

---

Funding

NSF-DMREF Grant 1334713, 1334351
NREL High Performance Computing (HPC)

Email: prashun.gorai@nrel.gov or pgorai@mines.edu

Please download slides: www.prashungorai.org/imrc2017