Accelerating Materials Discovery with Data-Driven Atomistic Computational Tools

Chris Wolverton

Dept. of Materials Science and Eng.

Northwestern University

AToM

Acknowledgements

Photo credit: Yongli Wang

Scott Kirklin
James Saal
Bryce Meredig
Logan Ward
Vinay Hegde
Kyle Michel
Jeff Doak
Alex Thompson

How to discover new materials?

DATA CREATION!

COLLECTION &! CLASSIFICATION!

DATA MINING & PREDICTION!

Calculating many known materials!

Solving unknown materials structures!

Databases of materials properties!

Materials discovery!

- Open Quantum Materials Database (OQMD)
- Machine Learning models to accelerate Materials Discovery

The Open Quantum Materials Database (OQMD)

- **Open** – An online (oqmd.org), freely available database…
- **Quantum** – … of self-consistently DFT-calculated properties (VASP, PAW, PBE)…
- **Materials** – … for >40,000 experimentally observed and >400,000 hypothetical structures (decorations of commonly occuring crystal structures)…
- **Database** – … built on a standard and extensible database framework.

Saal, Kirklin, Aykol, Meredig, and Wolverton "*Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)*", JOM **65**, 1501 (2013)

oqmd.org

OQMD:

An Open Quantum Materials Database

Home Materials Computing Analysis Documentation Download

Welcome to the Open Quantum Materials Database

The OQMD is a database DFT calculated thermodynamic and structural properties. We are providing this online interface for convenient, small scale access; however for more powerful utilization we recommend downloading the entire database and the API for interfacing with it, detailed in the link below.

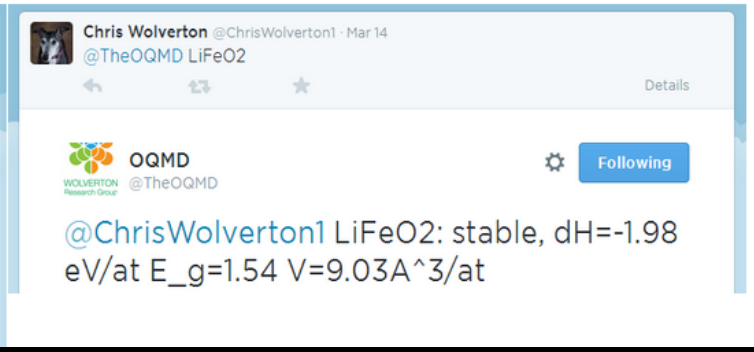
Current status

Database contains DFT calculations for **326628** compounds. In addition, calculations are constantly on-going, there are 267 new structures currently being run!
Recently completed compounds include: BeFeCuGe BeAlVCu TiBeGaCu CrCoCuBi ScBeCuGe

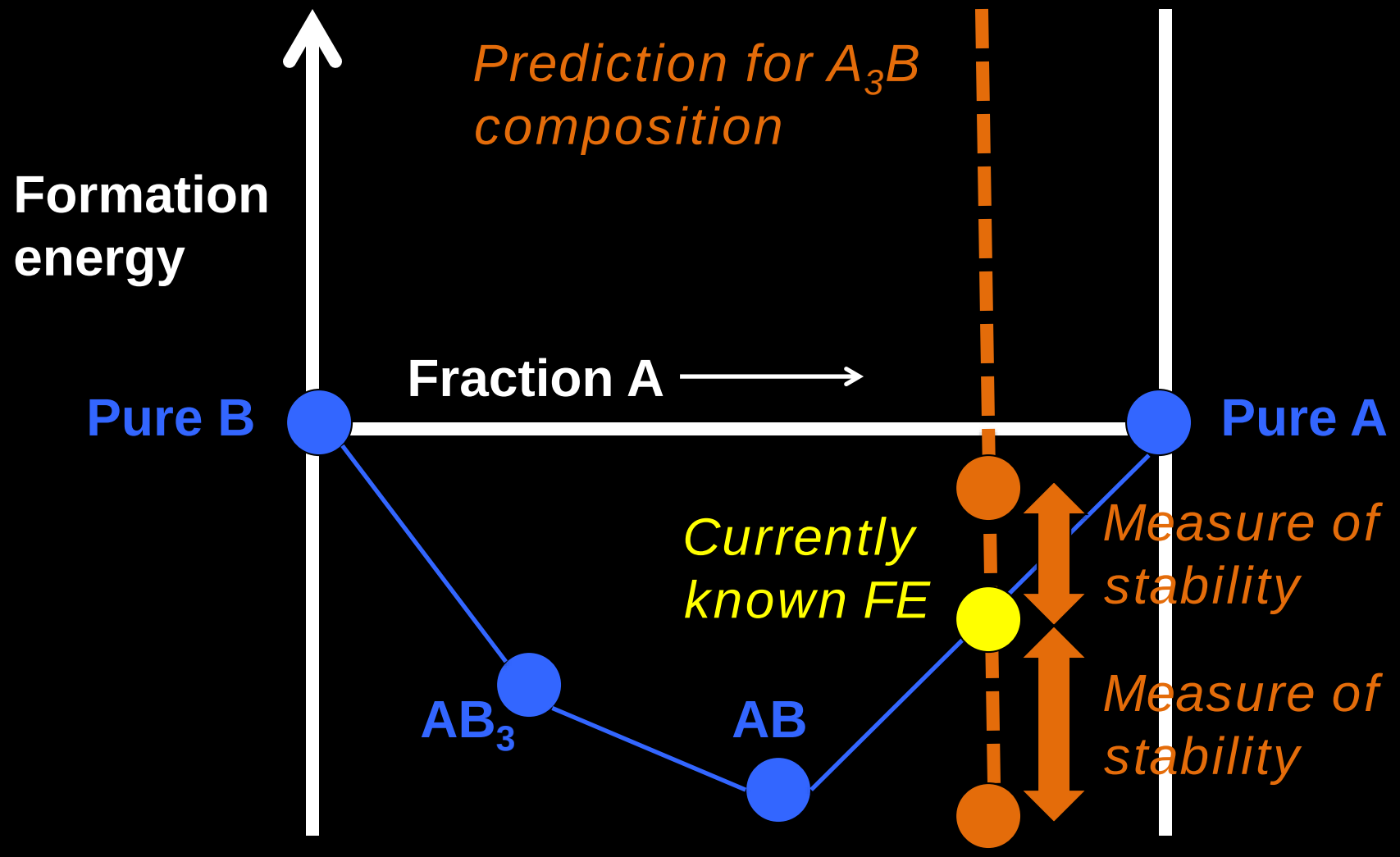
You can...

Search for materials by composition,
Create phase diagrams using database thermochemical data,
Determine ground state compositions,
Visualize crystal structures, or
Download the entire database for your own use!

Tweet @TheOQMD to ask what is stable at a composition, or to get a simple phase diagram!



Formation Energy→Stability

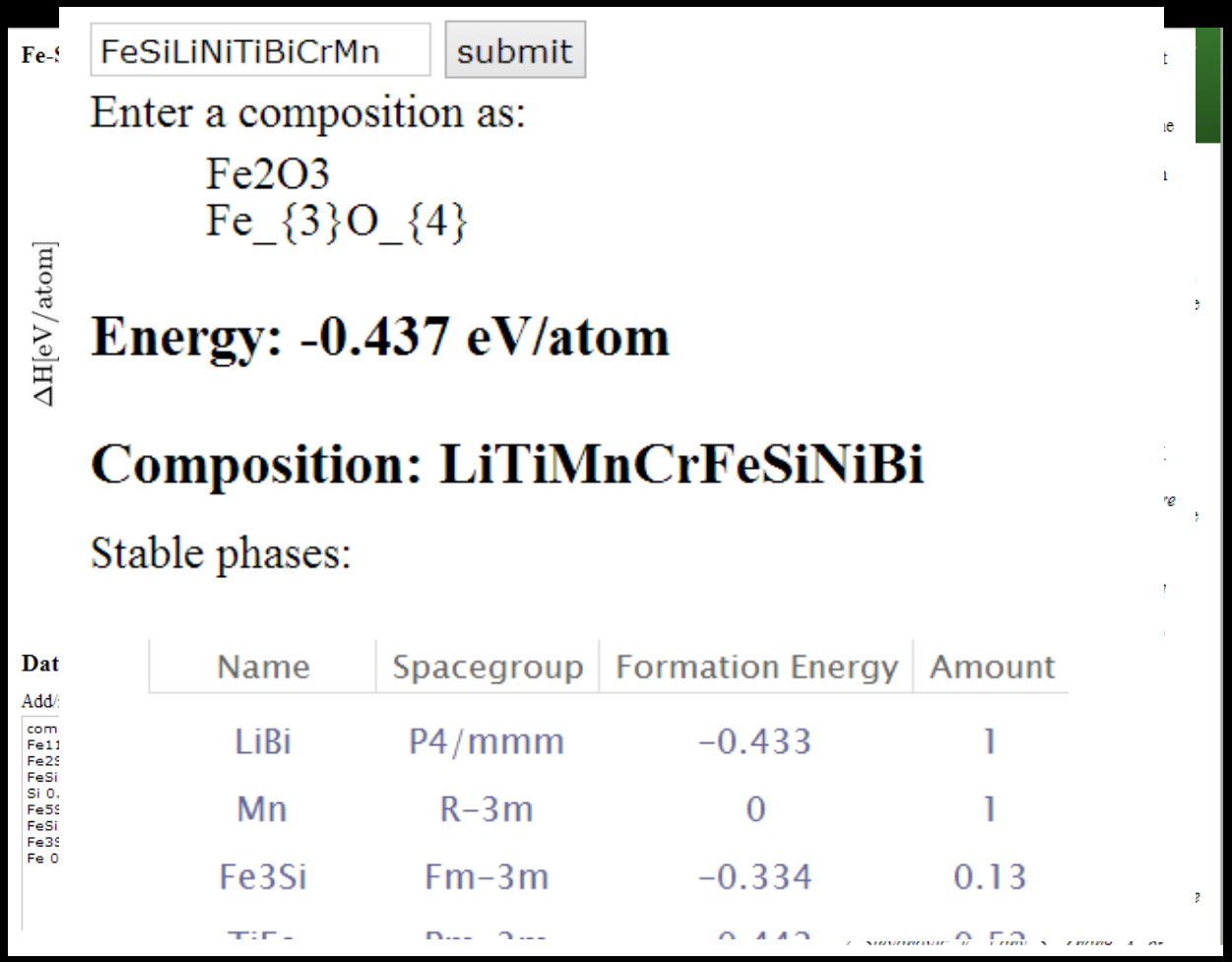


oqmd.org

Phase Diagrams (T=0K)

- binary
- ternary
- quaternary
- higher

GCLP[1]

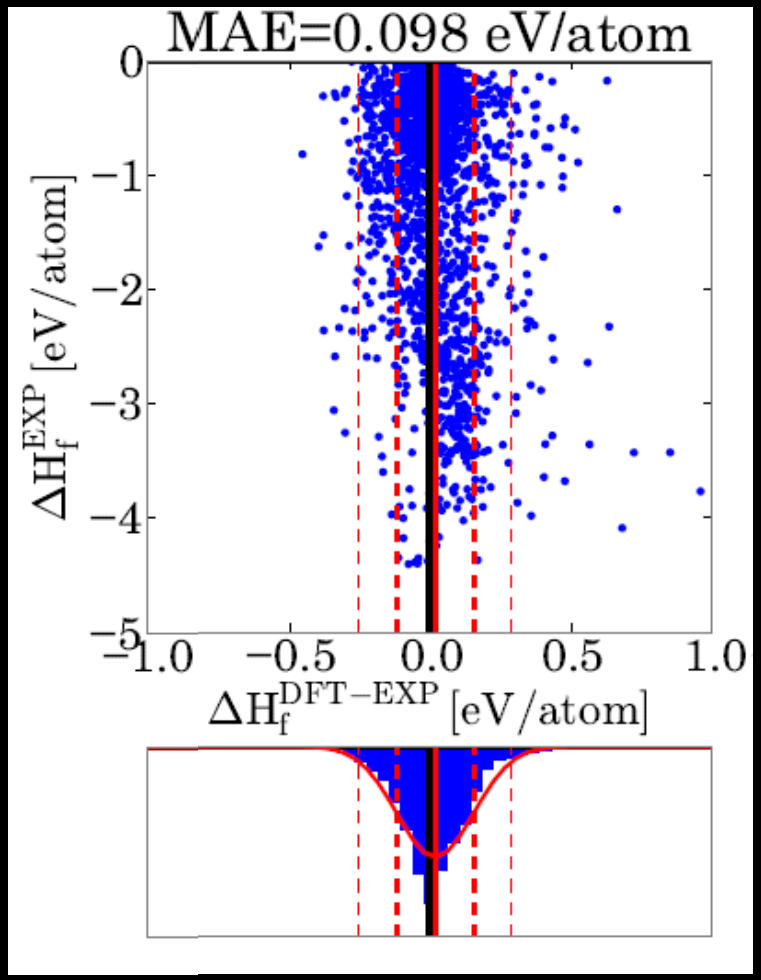


[1]R. Akbarzadeh, A., Ozoliņš, V. & Wolverton, C.. Advanced Materials 19, 3233–3239 (2007).

Accuracy of DFT Formation Energies

(comparison with a large number of ~1670 experimentally measured points)

$$\Delta H_f(\sigma) = E(\sigma) - \Sigma x_i E_i$$

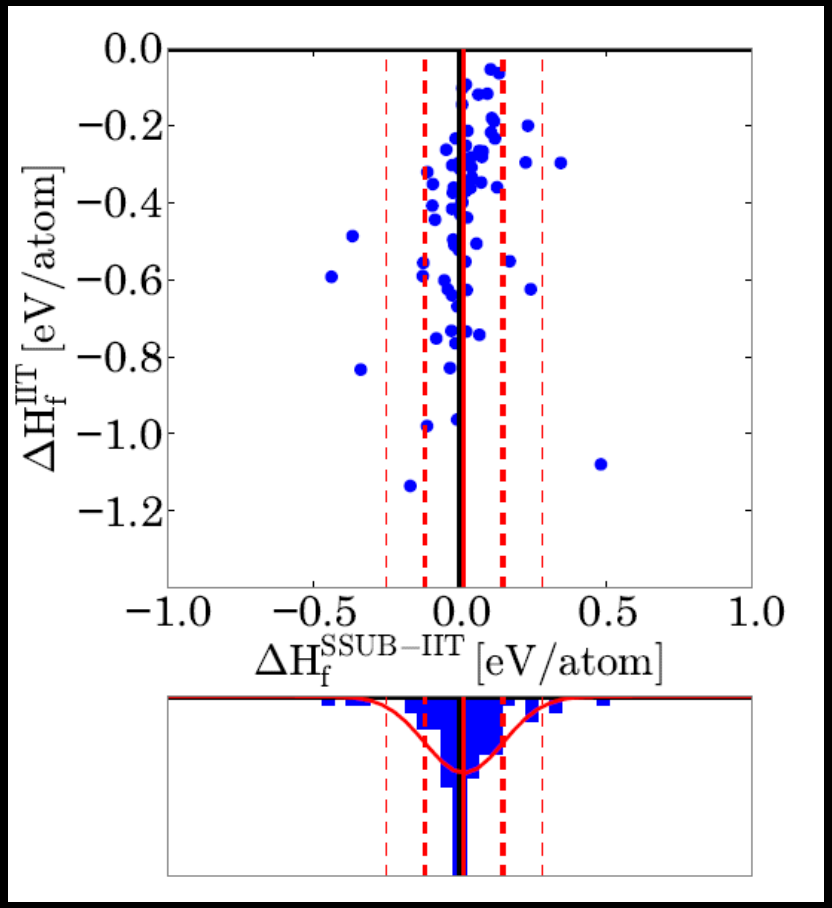


J. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, JOM 65, 1501 (2013).
FERE: V. Stevanovic, S. Lany, X. Zhang, and A. Zunger, Phys. Rev. B 85, 115104 (2012).
Mixing GGA/GGA+U: A. Jain et al., Comput. Mater. Sci. 50, 2295 (2011).

What about accuracy of experimental data?

(75 cases where two experimental measurements exist for the same compound)

Comparison of data in SSUB and IIT databases (curated)



MAE (one experiment vs. another) = 0.082 eV

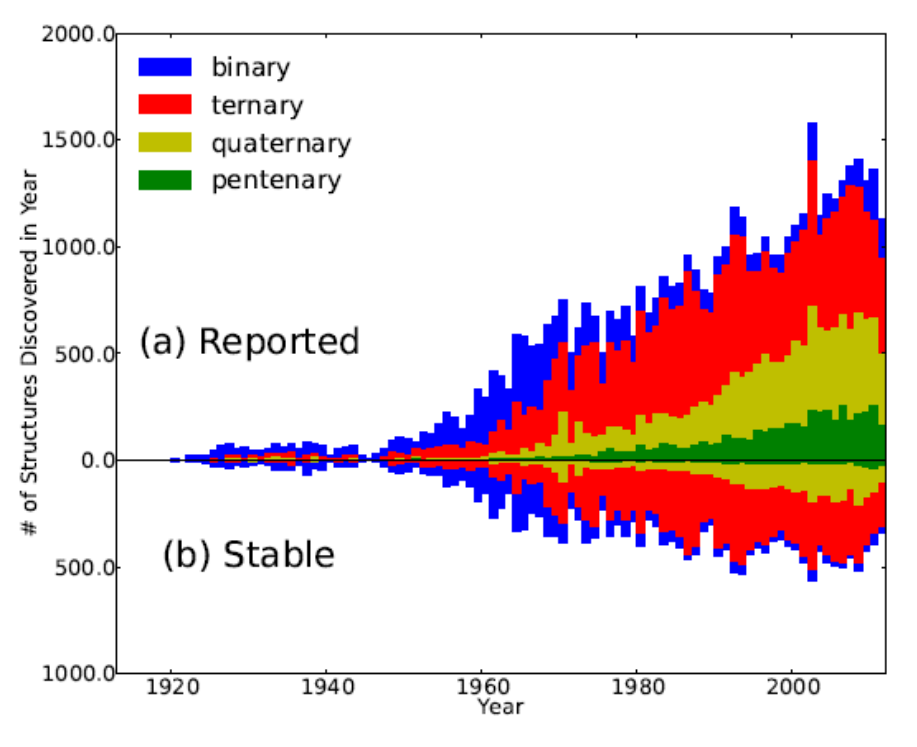
OQMD vs. experiment (for intermetallics) = 0.071 eV

We need more, high quality experimental thermochemical data!!!

How many compounds in our database are stable?

OQMD database (as of 2014):

- Total 297,099 compounds
- 19,757 T=0K stable
- 16,118 from ICSD
- 3487 “prototype” structures



The rate of compound discovery (total and stable) within the ICSD by year.

Each of these cases represents a prediction of a system where new compounds should exist! Many gaps in our current knowledge of phase stability…

High-Throughput DFT Calculations: OQMD

Can search through database to "screen" materials for various applications

- Heusler phase precipitates
- High strength Mg alloys
- Li-ion battery coatings
- Li-O_2 materials
- High-efficiency Thermoelectrics

Saal, Kirklin, Aykol, Meredig, and Wolverton "*Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Mechanical Database (OQMD)*", JOM **65**, 1501 (2013)

How to discover new materials?

DATA CREATION!

COLLECTION &! CLASSIFICATION!

DATA MINING & PREDICTION!

Calculating many known materials!

Solving unknown materials structures!

Databases of materials properties!

Materials discovery!

- Open Quantum Materials Database (OQMD)
- Machine Learning models to accelerate Materials Discovery

Data Mining in Real Life: Netflix

BellKor's Pragmatic Chaos wins the Netflix Prize

Machine Learning Strategy

- Recall basic calculation recipe:
  - Composition
  - Structure
- People focus on predicting/solving structure, but what if we could predict properties without it?
- Application: **Discovery of new ternary compounds $\mathbf{A}_x\mathbf{B}_y\mathbf{C}_z$**

Structure-Independent Model

Instead of mapping an atomic configuration to properties, i.e.,

$$C(r_1, r_2, \ldots r_n) \rightarrow P$$

we instead train a formation energy model on composition only:

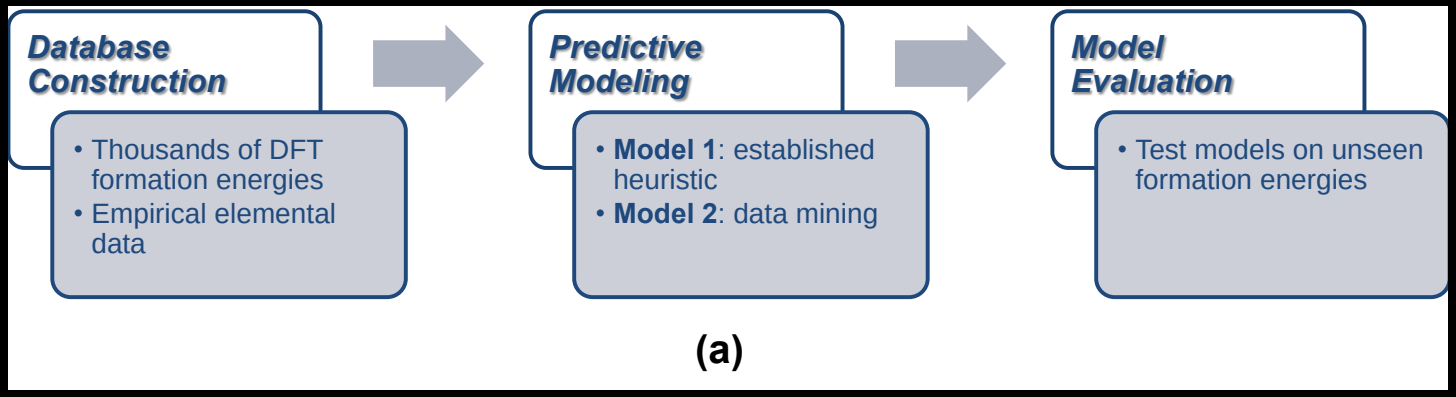
$$M(x_H, x_{He}, x_{Li} \ldots x_{Pu}) \rightarrow \mathrm{D}E_f$$

ML Model

- Data: ~15,000 DFT energies – stable ternary compounds + all binary convex hulls
- Fit to 4000 ternaries (and all binaries) / withhold ~8600 ternaries
- Descriptive Attributes (19, all functions of elements, no DFT outputs):
  - Average atomic mass
  - Average column/row on periodic table
  - Maximum difference / Average in atomic number
  - Maximum difference / Average in atomic radii
  - Maximum difference / Average in electronegativity
  - Average number of *s, p, d, f* valence electrons
  - *s, p, d, f* fraction of valence electrons
- Rotation forest ensembling technique with reduced error pruning trees as the underlying regression model

Meredig et al., Phys. Rev. B **89**, 094104 (2014).

Discovery Machinery



Meredig et al., Phys. Rev. B **89**, 094104 (2014).

Predictions for Discovery:
4500 new stable compounds

Machine learning model can predict the thermodynamic stability of arbitrary compositions without any other input (i.e., without the structure).

Six orders of magnitude less computer time than DFT.

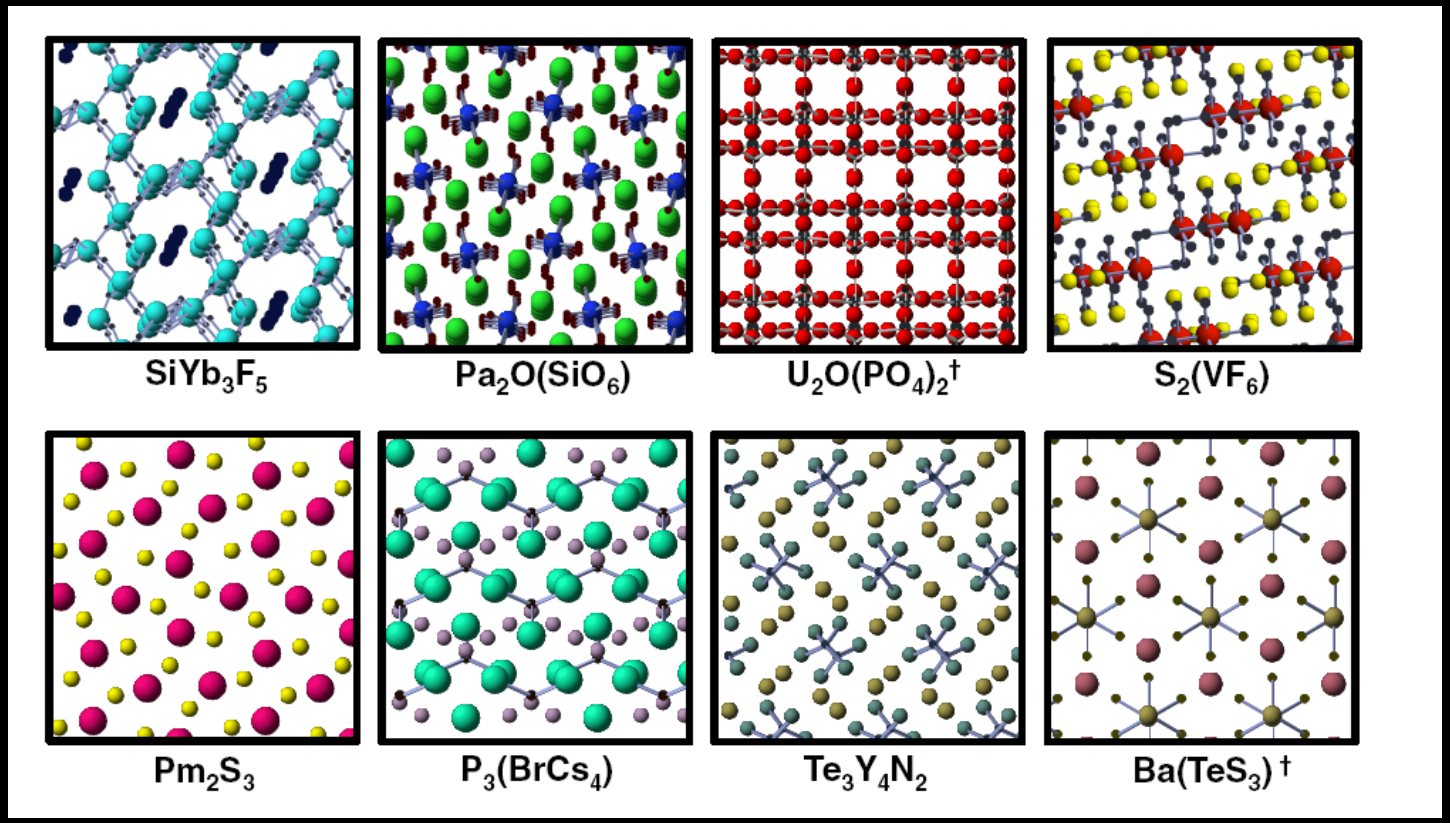
We scan ~1.6 million candidate compositions for novel ternary compounds ($A_xB_yC_z$),

Predict 4500 new stable materials (would represent a ~10% increase in the total number of known ternary compounds).

Complete list of predicted compounds:
http://journals.aps.org/prb/supplemental/10.1103/PhysRevB.89.094104/predictions_dat.pdf

Meredig et al., Phys. Rev. B **89**, 094104 (2014).

Validating high-ranking compositions with crystal structure prediction



Tested 9 predicted stoichiometries. In 8 cases, crystal structure prediction methods found a structure with DFT energy lower than all combinations of existing known phases.

Composition-based Attributes

Property	Attributes	Reference
Crystal Structure	VE, ΔX, n_{av}, $\Delta n_{ws}^{1/3}$	Kong et al., 2012
Band Gap	ΔX, Z, T_m, R, n_{av}	Srinivasan & Rajan, 2013
Formation Energy	ΔX, Z, $n_{s\|p\|d\|f}$, row, col	Meredig *et al.*, 2014
Melting Point	Z, m, n, r^{cov}, I, X, …	Seko *et al.*, 2014
ΔH_f: Rocksalt – Wurtzsite	IP, EA, r_s, r_p, …	Ghiringhelli *et al.*, 2015

Observations:

- Different properties, different attributes
- All based on elemental property statistics

Our Strategy: Create set that includes all of these and more

General-Use Attributes

Elemental Property Stats.: *Mean T_m, Range Z, …*

6 Statistics: Mean, variance, max, min, range, mode

22 Elemental Properties: Z, EN, Row, Column, Radius, …

Stoichiometric: *# Components,* $\|x_Z\|_p$

Electronic Structure Based: *Fraction p Electrons, …*

Ionicity: *Can form Ionic, % Ionic Character, …*

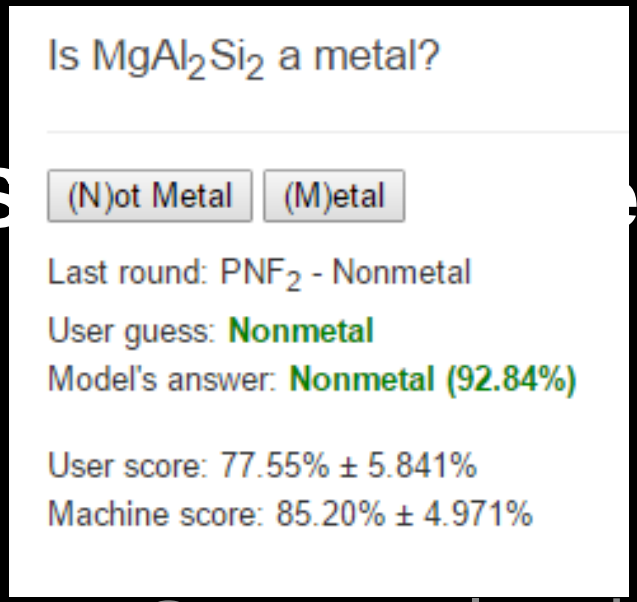
Ward, Agrawal, Choudhary, Wolverton, npj Computational Materials 2, 16028 (2016).

Simple Example: Is it a Metal?

Task: Given composition, $E_g > 0$?

Training Set Dataset: 3000 entries from the OQMD

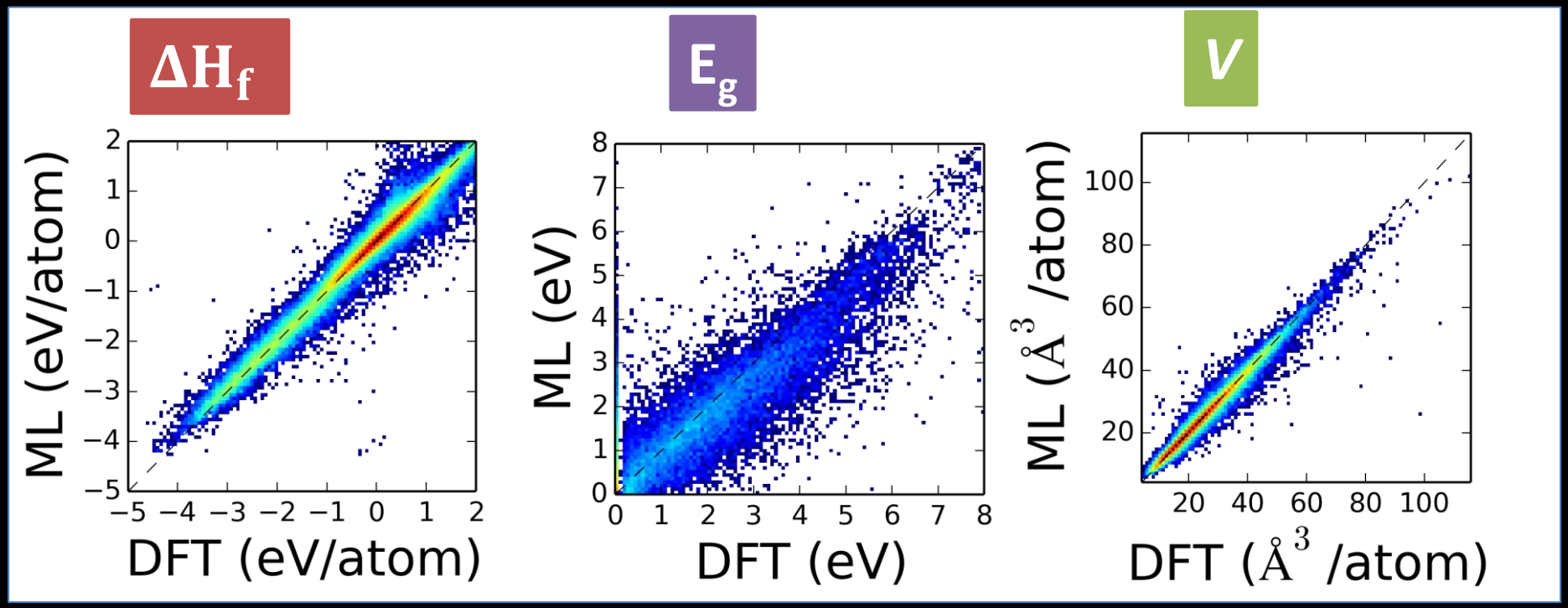
S **el:** Accuracy ~90%



Game: palestrina.northwestern.edu/metal-detection/

Application to the OQMD

Dataset: 240000 DFT Calculations (OQMD.org)



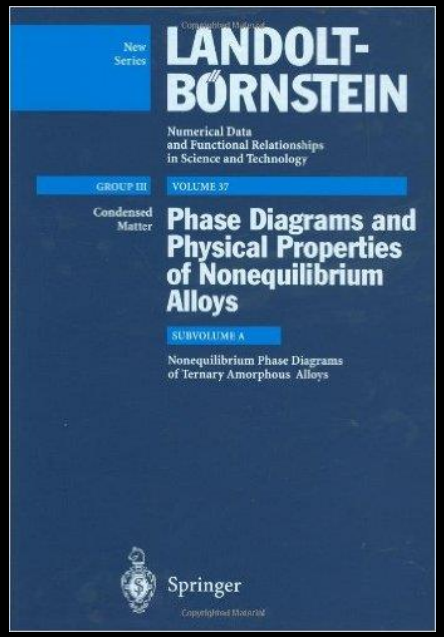
R: 0.944
MAE: 80.5 meV/atom

R: 0.924
MAE: 0.21 eV

R: 0.993
MAE: 0.452 $Å^3$/atom

Ward, Agrawal, Choudhary, Wolverton, npj Computational Materials 2, 16028 (2016).

Predicting Glass Forming Ability



Application: Metallic Glasses

Goal: Predict glass-forming ability

Dataset: Landolt-Börnstein

- 6836 experimental measurements
- 295 ternary systems
- Binary property: **[Can Form Glass]** | **[Cannot Form]**

Model: Random Forest

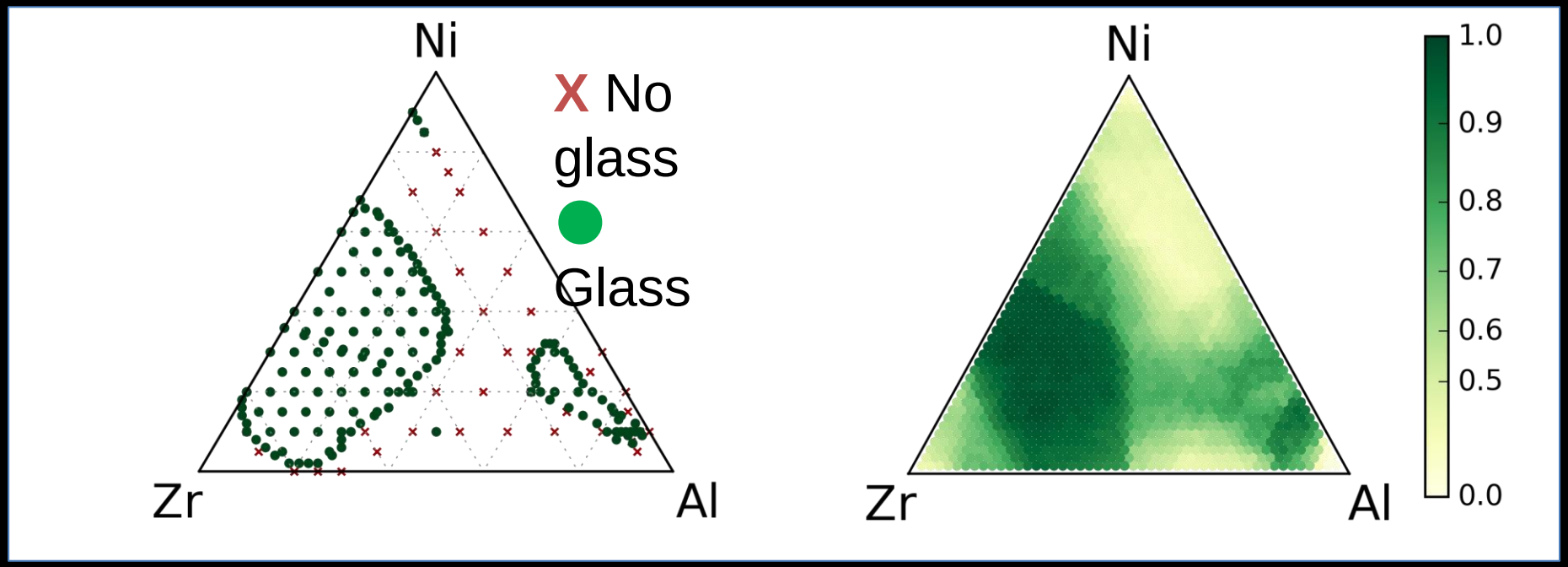
- 90% accurate in 10-fold cross-validation

Ward, Agrawal, Choudhary, Wolverton, npj Computational Materials 2, 16028 (2016).

Predicting Glass-Forming Ability

Test: Remove Al-Ni-Zr data from training data, try to predict

Measured **Predicted**



Same representation, very different material property

Ward, Agrawal, Choudhary, Wolverton, npj Computational Materials 2, 16028 (2016).

Doing this Yourself: Magpie

Materials-Agnostic Platform for Informatics and Exploration

https://bitbucket.org/wolverton/magpie

Main Features:

- Attribute calculation (145 General-Purpose Attributes)
- Built-in partitioning schemes
- Access to modern ML algorithms (Weka, Scikit-Learn)
- Simple text interface
- Integration with other codes: Apache Thrift

Why this Is Crucial:

- Replicate and share results
- Use ML without much training

Ward, Agrawal, Choudhary, Wolverton, npj Computational Materials 2, 16028 (2016).

How to discover new materials?

DATA CREATION!

COLLECTION &! CLASSIFICATION!

DATA MINING & PREDICTION!

Calculating many known materials!

Solving unknown materials structures!

Databases of materials properties!

Materials discovery!

- Open Quantum Materials Database (OQMD)
- Machine Learning models to accelerate Materials Discovery

More information…

- OQMD (high-throughput DFT database)
  - oqmd.org
  - @TheOQMD
  - J. Saal et al., "*Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Mechanical Database (OQMD)*", JOM **65**, 1501 (2013)
  - S. Kirklin et al., "*The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies*", npj Computational Materials 1, 15010 (2015).

- Machine Learning models
  - MAGPIE https://bitbucket.org/wolverton/magpie
  - B. Meredig et al., "*Combinatorial screening for new materials in unconstrained composition space with machine learning*", Phys. Rev. B **89**, 094104 (2014).
  - L. Ward et al., "*A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials*" npj Computational Materials 2, 16028 (2016).
  - L. Ward, C. Wolverton, "*Atomistic calculations and materials informatics: A review*", Curr. Opin. Solid State Mater. Sci. 21, 167 (2017).
  - L. Ward et al., "*Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations*", Phys. Rev. B (in press, 2017).