Machine Learning and Global Optimization for Materials Discovery

Logan Ward

Postdoctoral Scholar

Computation Institute, University of Chicago

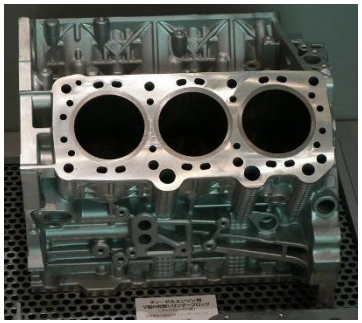
3 October 2017

Machine Learning and Global Optimization for Materials Discovery

Logan Ward, Chris Wolverton
Graduate Student
Northwestern University
3 October 2017

Materials Engineering: A narrow introduction

Examples of “Aluminum”

Application				
Key Properties	Conductivity	Low cost Shapeable	Castable Strong	???
Alloy	**Al + Mg**	**Al + Mg,Mn**	**Al + Si,Cu**	???

How do we tailor materials for new technologies?

How can we do this quickly?

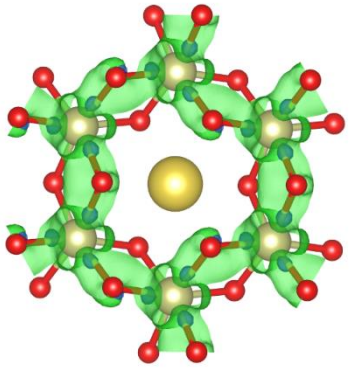
Computational Materials Engineering

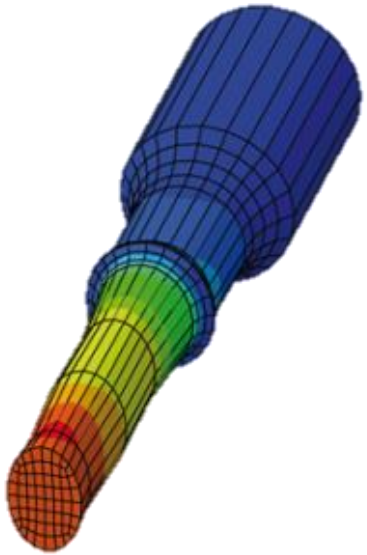
Goal: Accelerate design of materials

Method: Replace experiments with computers

Many Established Tools:

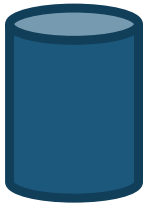
- Density Functional Theory
- Phase Field
- Finite Element Analysis
- Computational Thermodynamics





Emerging Field: **Data-driven Models**

Materials Data



Machine Learning

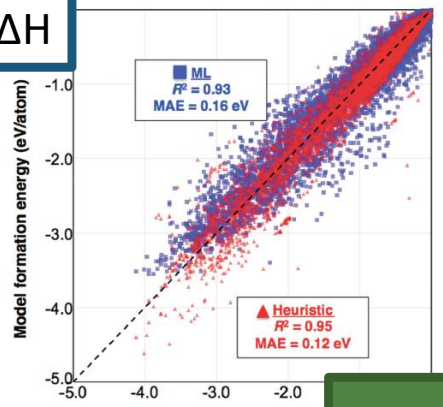
$$\sigma_Y = f(x)$$

Predictive Model

FEA Image: http://www.icams.de/content/research/index.html

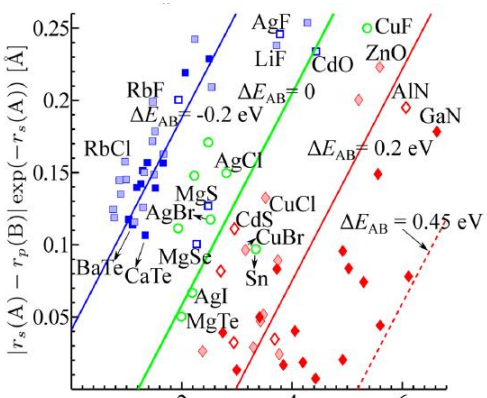
ML + Materials = “Materials Informatics”

ΔH

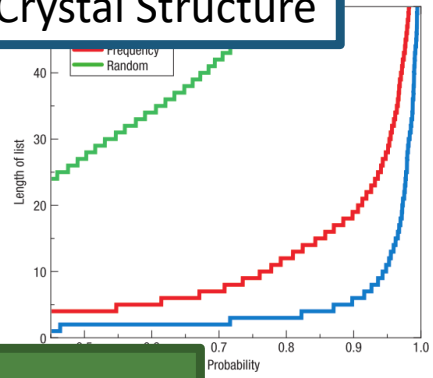


Meredig *et al*. PRB (20

ΔH



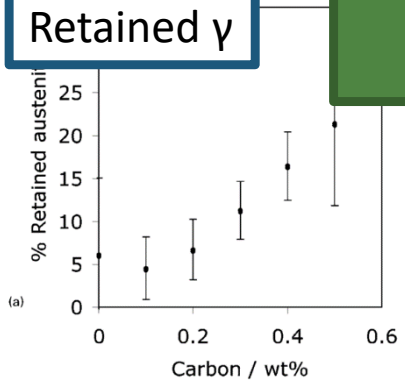
Crystal Structure



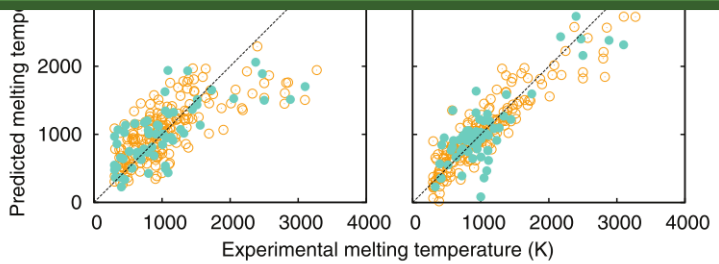
Nat. Mat. (2006), 641

Question: How can I do this, but for many more properties?

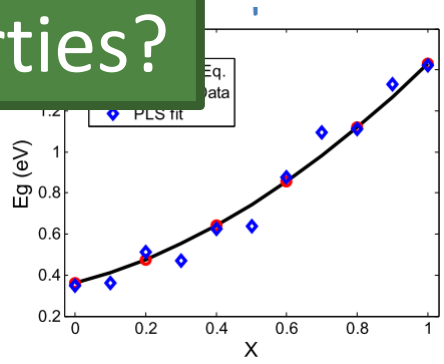
Retained γ



Chatterjee *et al*. MS&T (2007), 819



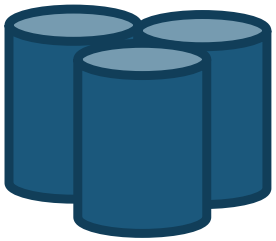
Seko *et al*. PRB (2014), 054303

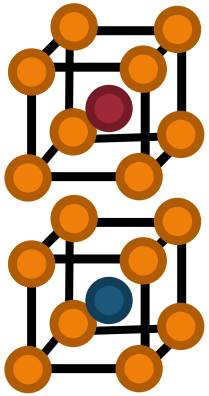


Srinivasan, Rajan. *Materials* (2013), 279

Materials Informatics Workflow

Collect → Process → Represent → Learn





$\Delta H_f = -1.0$

$\Delta H_f = -0.5$

$\vec{X}$ $\vec{y}$

Z_A	Z_B	ΔH_f
3	4	-1.0
3	5	-0.5

$$\Delta H_f = f(Z_A, Z_B)$$

Materials Informatics Challenges

Need: High-quality data

Question: How to simplify using ML in MSE?

Part 1: Automated Crystal Structure Solution

Part 2: General-Purpose Representations for Materials

Part 3: Design of New Materials

Part 1: Automated Structure Solution

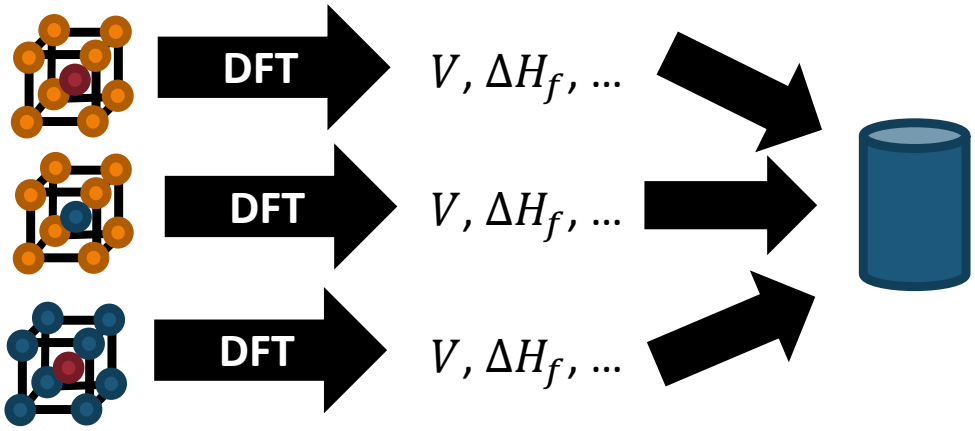
Acknowledgements: Kyle Michel

Why is this important?

One Reason: High-Throughput DFT

Automated calculation of properties

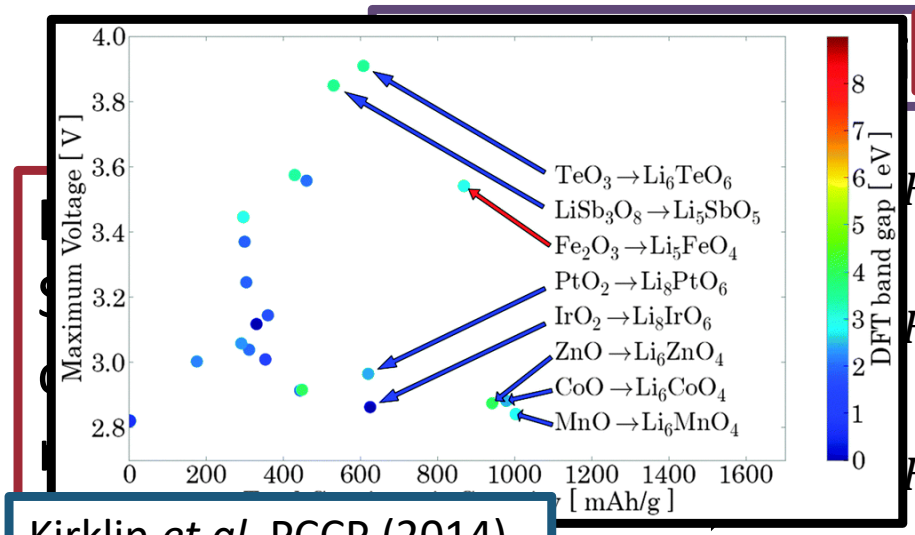
Input:
Structures of known materials



Output:
Large collection of materials data

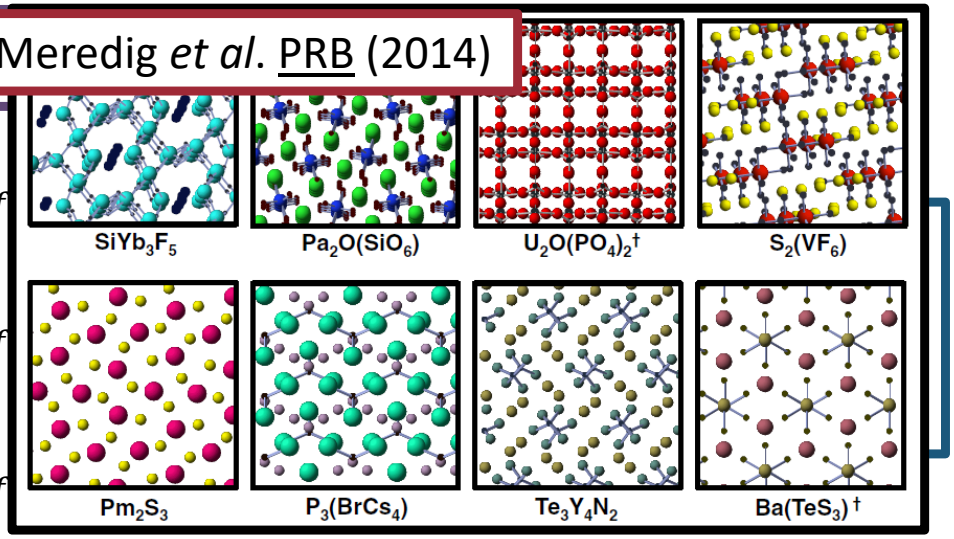
Why is this important?

One Reason: High-Throughput DFT



Kirklin *et al*. PCCP (2014)

Meredig *et al*. PRB (2014)



Any HT search is limited by the availability of high-quality crystallographic data!

Problem: Structures unknown for many materials

Crystal Structure Databases

FIZ Karlsruhe

ICSD

Inorganic Crystal Structure Database

Collection of the crystal structures of compounds

Total entries: 148818

Missing atoms: 13705 (~9%)!

Powder Diffraction File

Database of Powder X-Ray Diffraction patterns

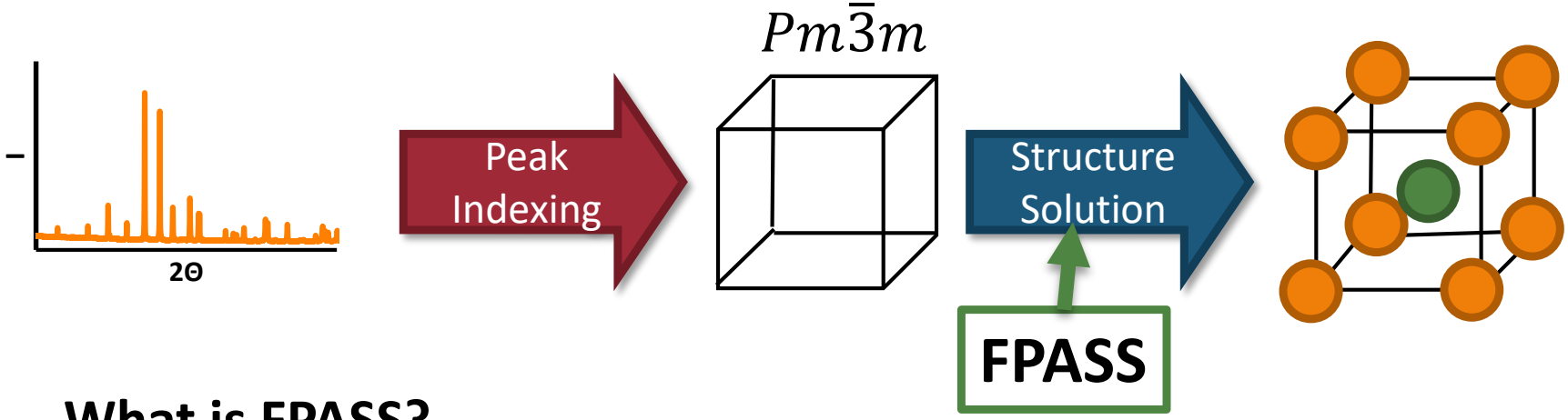
Total entries: 384613

Without structure: 113164

10^3 – 10^5 of materials not included in HT DFT Databases!

Proposed Solution: Solve them with FPASS

First-Principles-Assisted Structure Solution



What is FPASS?

Simple: Single-step method to solve structure

Accurate: Uses both energy and diffraction pattern match

Fast: Constrains search with known symmetry

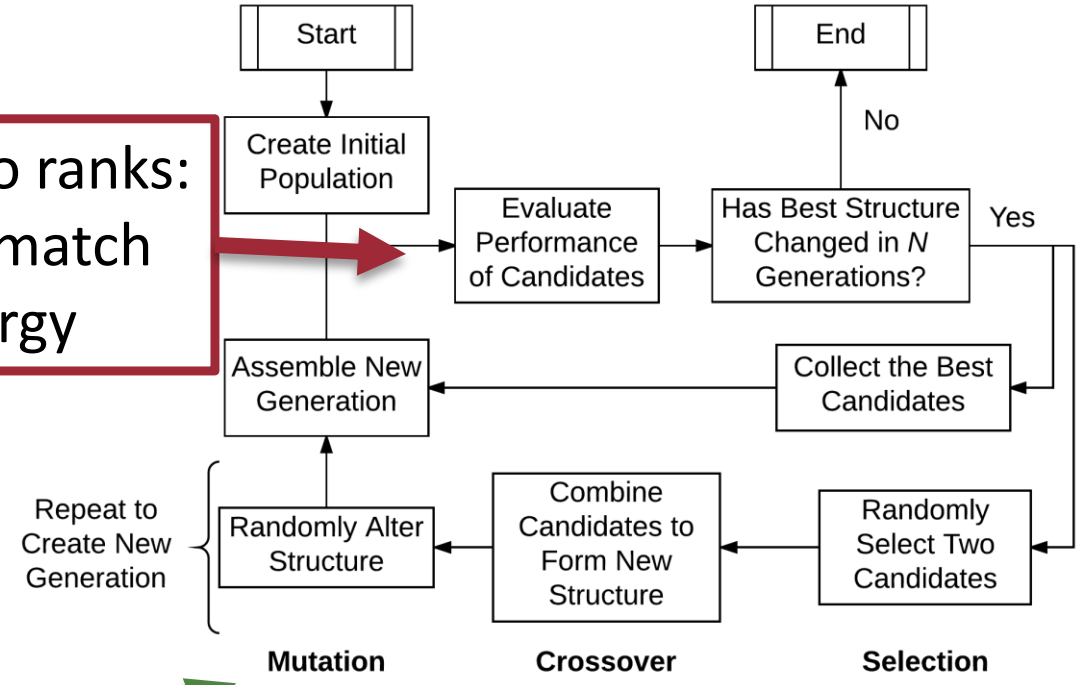
Problem: We had not used FPASS for *new* solutions

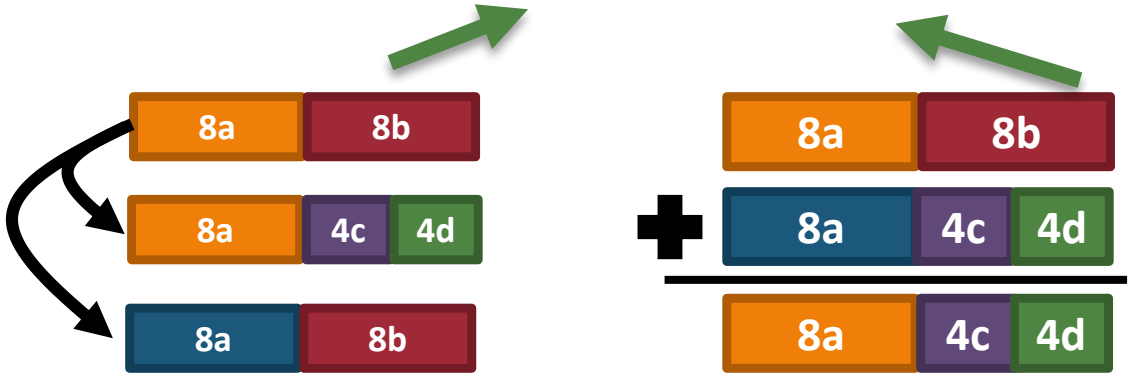
- Can we trust it in an automated scheme?
- Can we automate it?

Meredig and Wolverton. Nature Materials. (2013)

FPASS Uses a Genetic Algorithm

Sum of two ranks:
- Pattern match
- DFT energy





By maintaining Wyckoff sites, we preserve symmetry

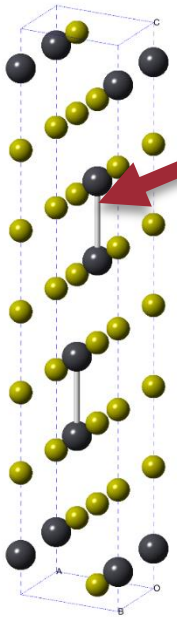
Solving Na_xPb_y Structures

Our starting point: Abstract from 1957

Phase	Ideal formula	Space group	a_0 (A)	c_0 (A)	Pb atom positions
I	Na_5Pb_2(?)	$R\bar{3}m$	5·54	23·15	6 (c): 0, 0, z; ...; z = 0·07
II	Na_9Pb_4	$P6_3/mmc$	5·47*	30·41	4 (e): 0, 0, z; ...; z = 0·05 4 (f): ⅓, ⅔, z; ...; z = 0·30
III	$Na_{13}Pb_5$	$P6_3/mmc$	5·51	40·39	2 (b): 0, 0, ¼; 0, 0, ¾ 4 (f): ⅓, ⅔, z; ...; z = 0·05 4 (f): ⅓, ⅔, z; ...; z = 0·13

indicating a true cell with $a_0' = \sqrt{3}.a_0$, $c_0' = c_0$.

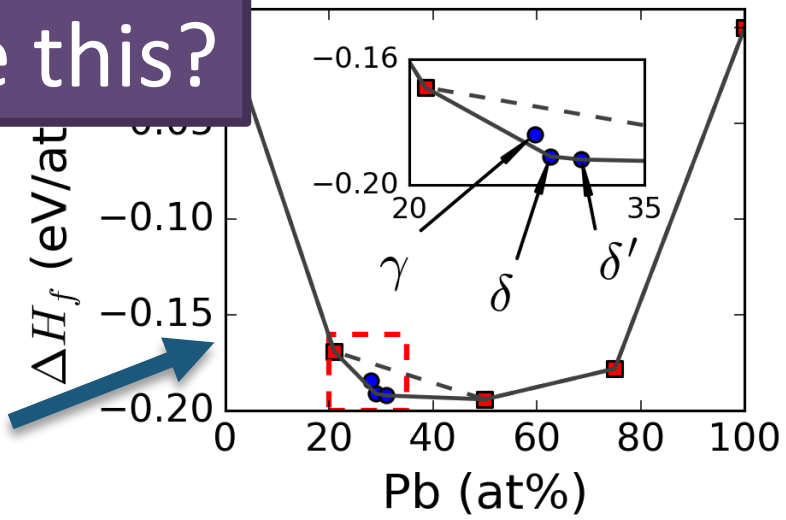
Only missing feature: Na positions



Can we automate this?

Na positions with FPASS

New structure improve computed Na-Pb diagram



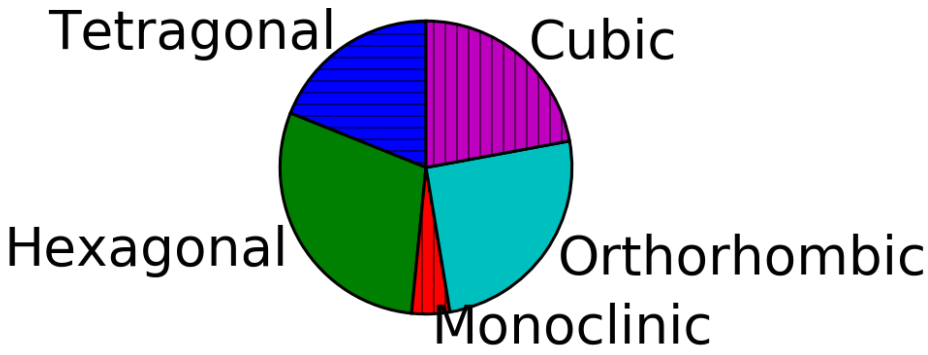
Computed T=0K Phase Diagram

Ward, Michel, Wolverton. Acta Crystal. A17 (2015), 542

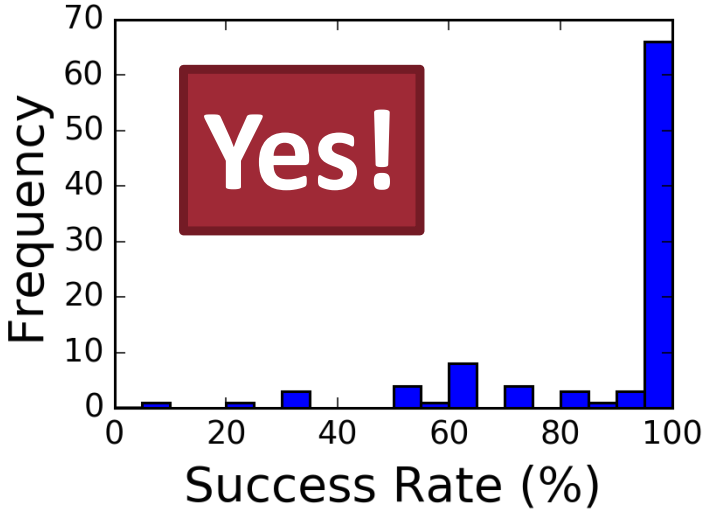
Can we trust automated FPASS?

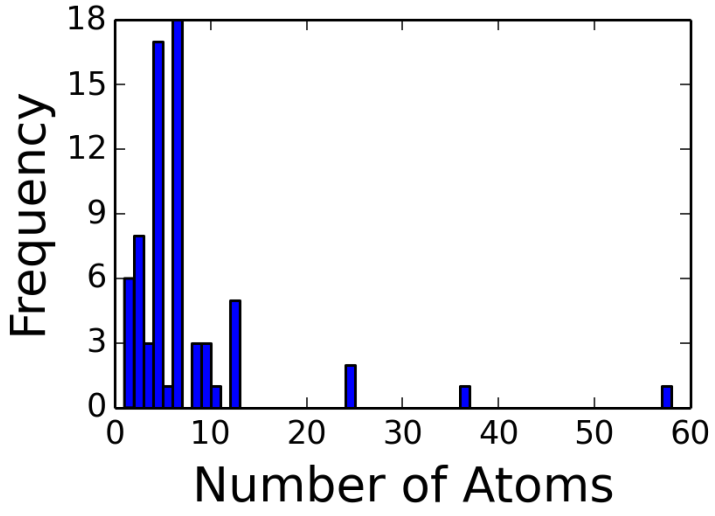
Test Cases: Strukturbericht

- Common structures
- ~100 compounds
- Variety symmetries, compositions, sizes

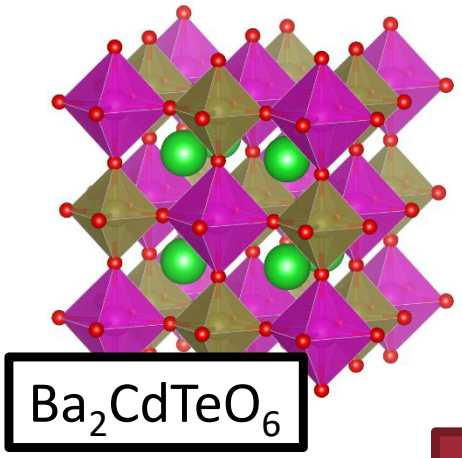


Can we solve them?

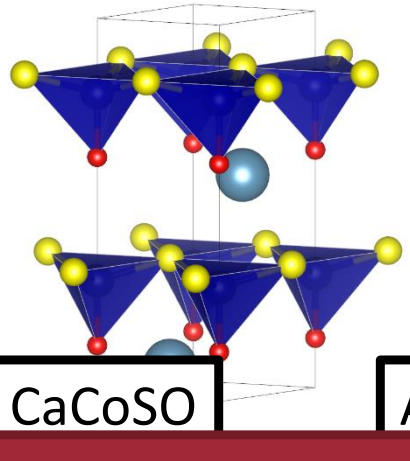




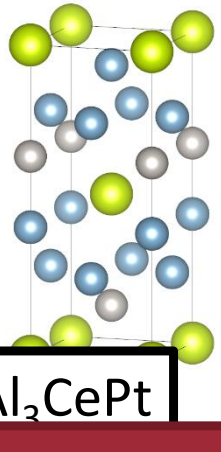
New Solutions



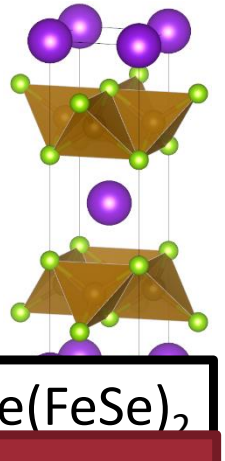
Ba_2CdTeO_6



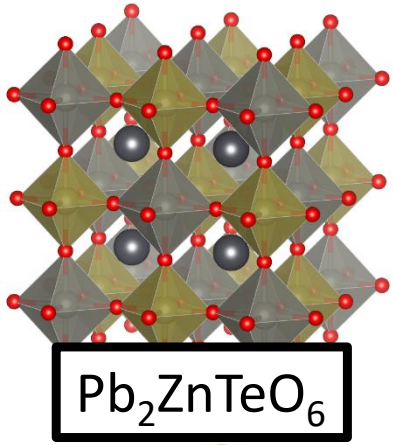
CaCoSO



Al_3CePt

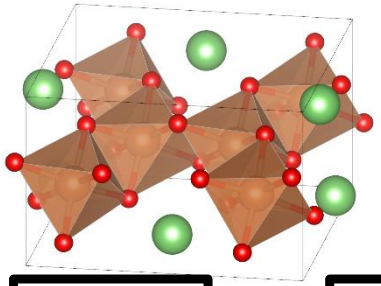


$Ke(FeSe)_2$

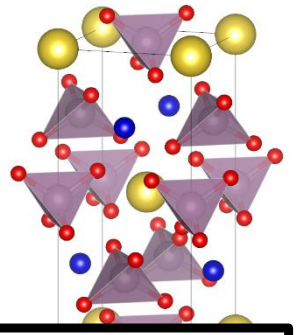


Pb_2ZnTeO_6

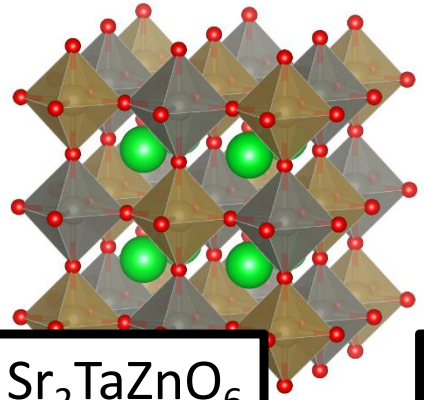
And there are more on the way!



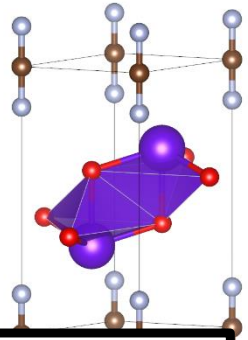
$LiSbO_3$



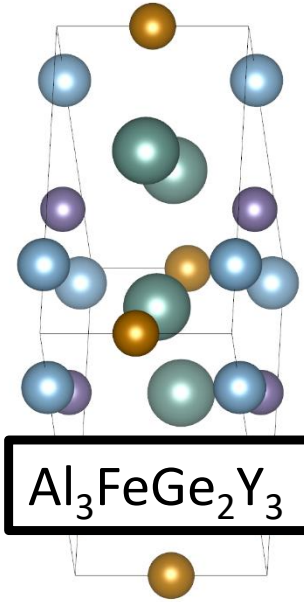
$NaTm(MoO_4)_2$



Sr_2TaZnO_6



$Tb_2O_2CN_2$



$Al_3FeGe_2Y_3$

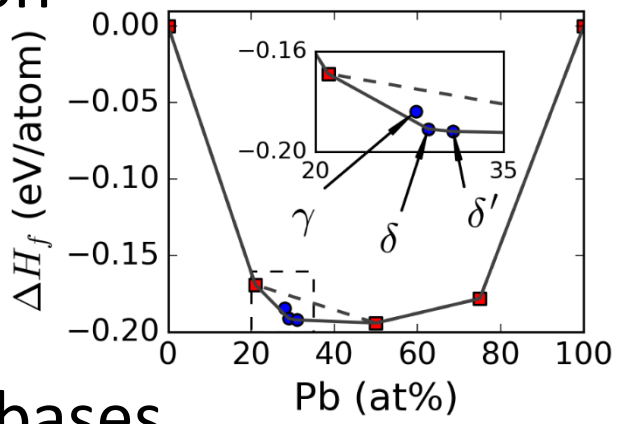
Ward, Michel, Wolverton, *in review*

Summary

Goal: Reduce the number of unsolved compounds

Method: First-Principles-Assisted Structure Solution

1. Created automated implementation
2. Validated against ~100 structures
3. Solved 13 crystal structures



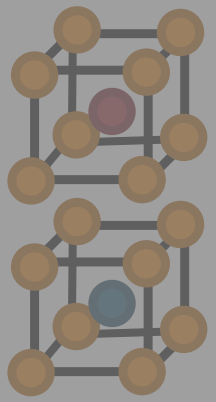
Result: Improved accuracy of HT-databases

Ward, Michel, Wolverton. Physical Review Materials, *in review*

Part 2: General-Purpose ML Methods for Materials

Materials Informatics Workflow

Collect → Process → Represent → Learn



$\Delta H_f = -1.0$

$\Delta H_f = -0.5$

$\vec{X}$ $\vec{y}$

Z_A	Z_B	ΔH_f
3	4	-1.0
3	5	-0.5

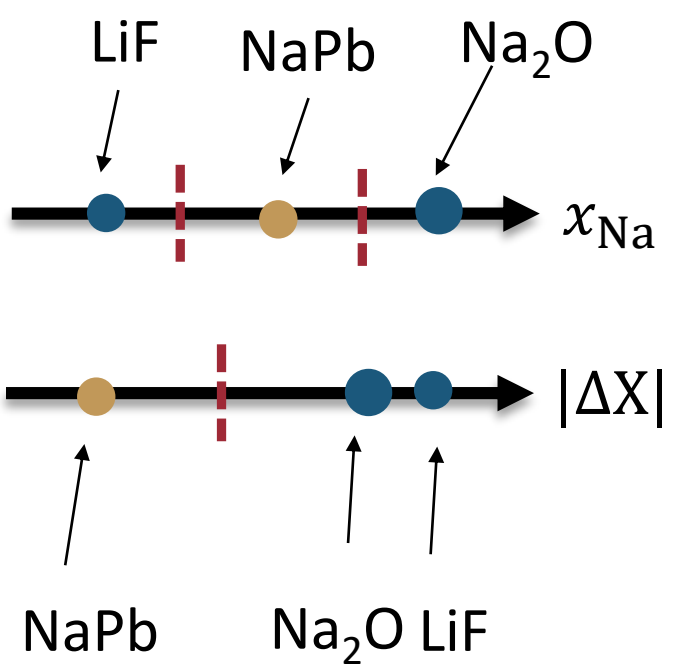
$\Delta H_f = f(Z_A, Z_B)$

How can one create problem-independent representations?

What is a representation?

Set of quantitative attributes that describe a material

$$Property = \boldsymbol{f}(Attributes)$$



Representation of material

Ex: *Attributes* = $\boldsymbol{g}(x_H, x_{He}, \ldots)$

What does a representation need?

Completeness: Differentiate materials

Efficiency: Quick to compute

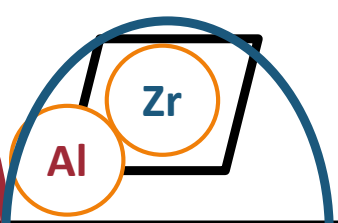
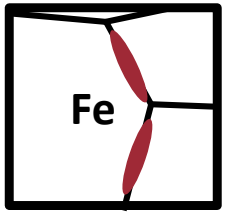
Accuracy: Capture important effects

Diversity: Many possible properties

How do we create "general-purpose" representations?

Variety of Types of Materials Data

How to differentiate materials?

	Zr	Al,Zr	Al_4Zr		
Available Information	Element	Phase Diagram	Composition	Crystal Structure	μ-Structure
Example Properties	Dilute ΔH_{mix}	# Eutectics	Glass-Forming Ability	ΔH_f	σ_Y

Need: A suite of general-purpose representations

Focus #1: Composition

$$Property = f(Composition)$$

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, …*

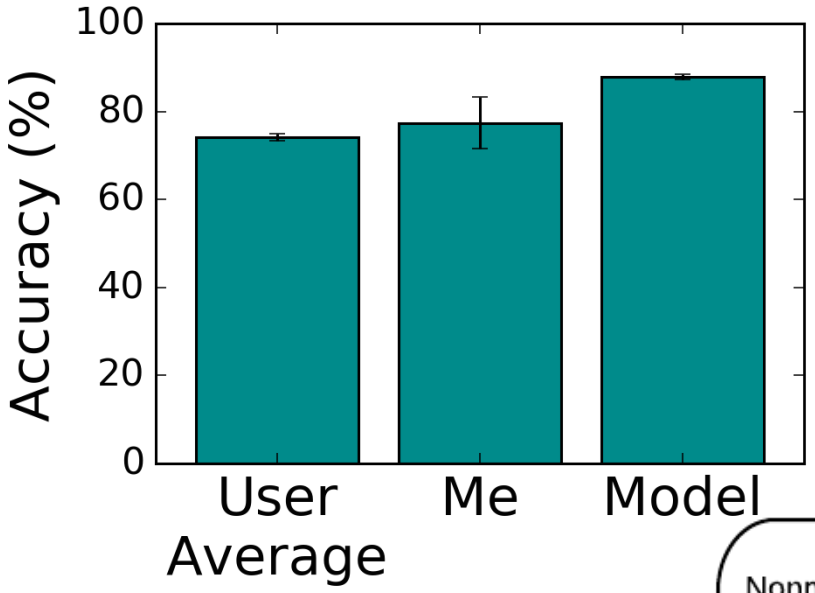
Ref: Ward *et al*. npj Comp. Mat., (2016) 28

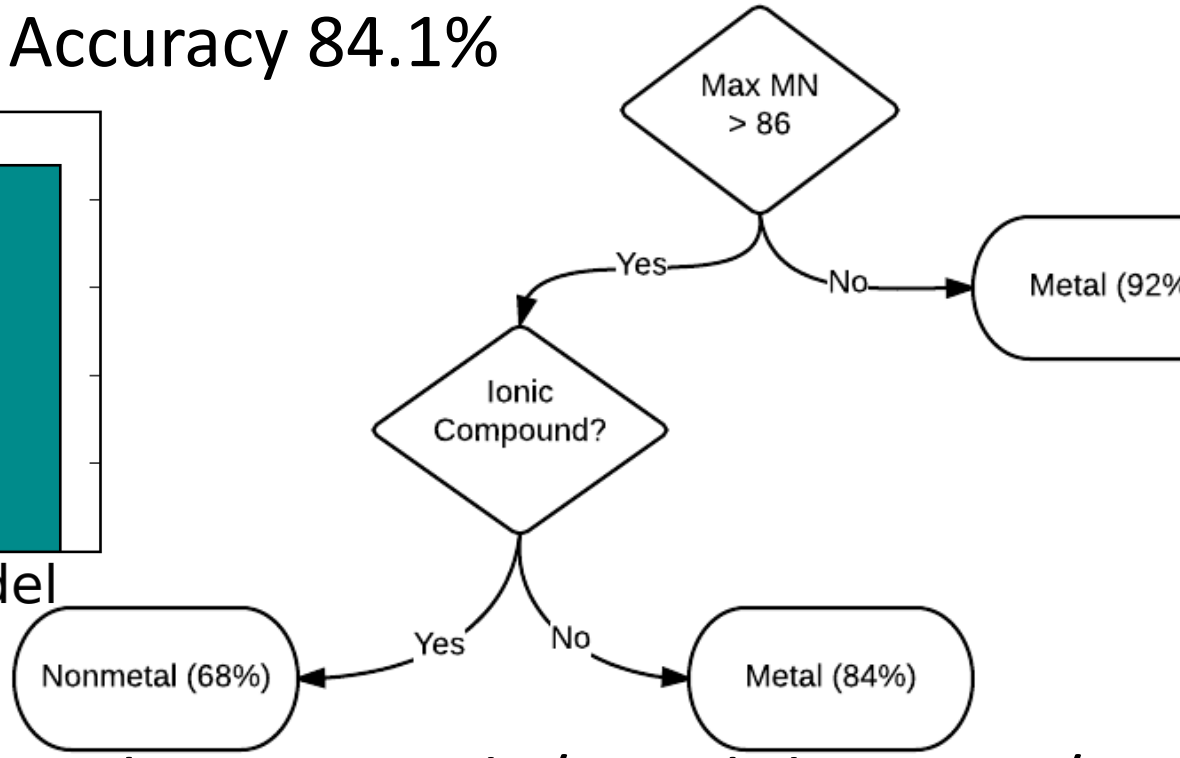
Simple Example: Is it a Metal?

Task: Given composition, $E_g > 0$?

Training Set Dataset: 3000 entries from the OQMD

Simple ML Model: Accuracy 84.1%



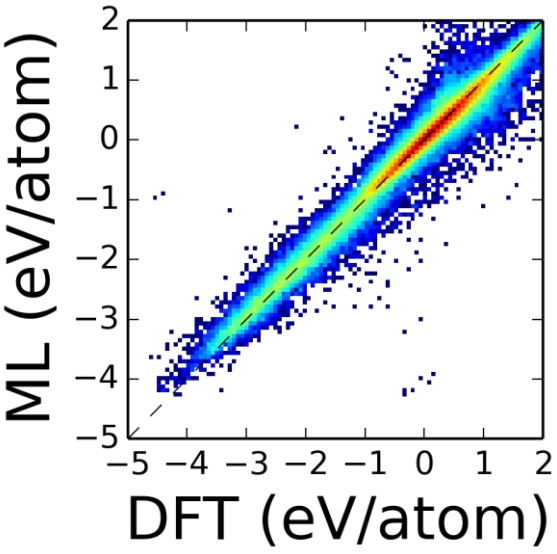


Game: palestrina.northwestern.edu/metal-detection/

Application to the OQMD

Dataset: 240000 DFT Calculations (OQMD.org)

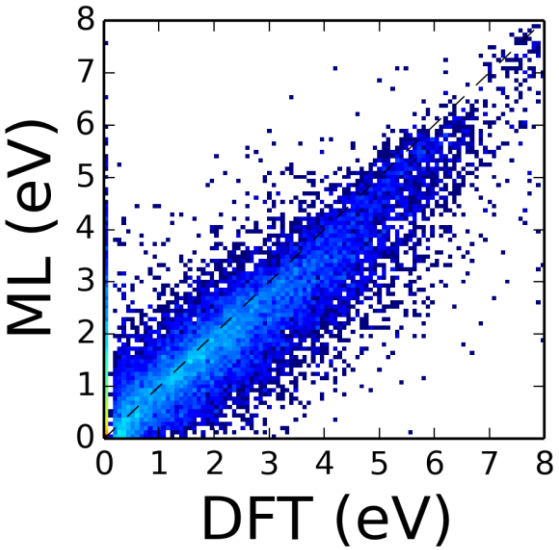
ΔH_f



R: 0.944

MAE: 80.5 meV/atom

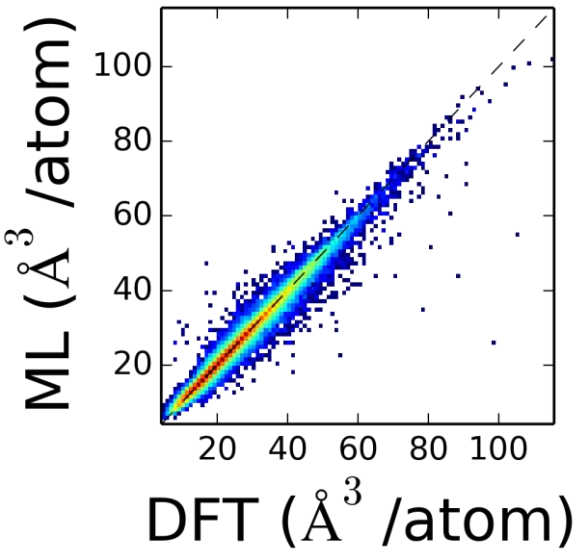
E_g



R: 0.924

MAE: 0.21 eV

V



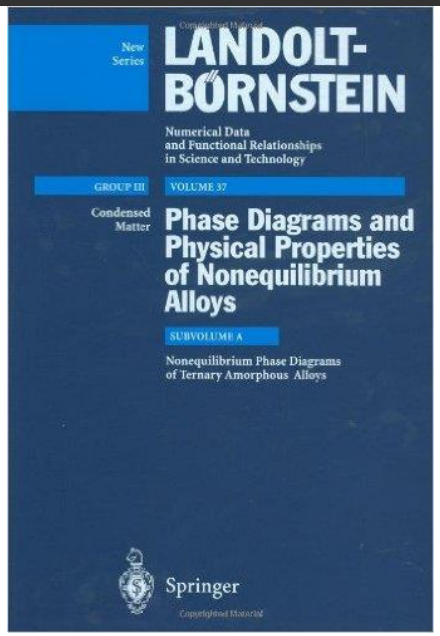
R: 0.993

MAE: 0.452 Å^3/atom

Ref: Ward *et al*. npj Comp. Mat., (2016) 28

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

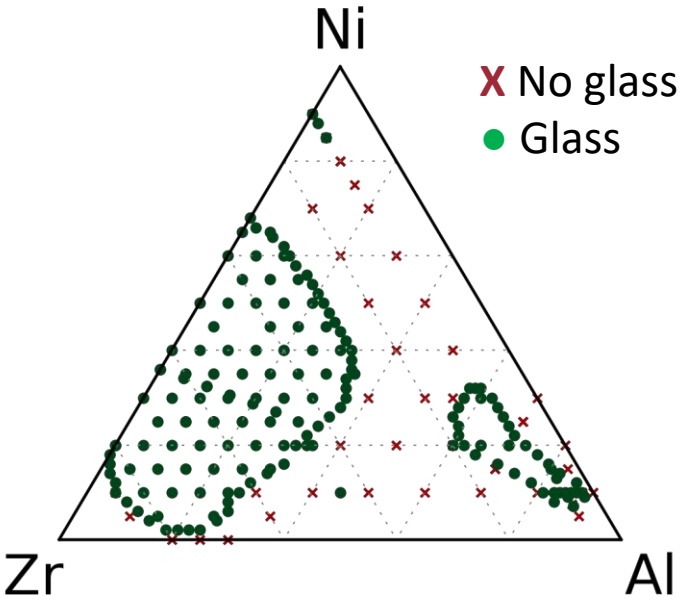
- 90% accurate in 10-fold cross-validation

Ref: Ward *et al*. npj Comp. Mat., (2016) 28

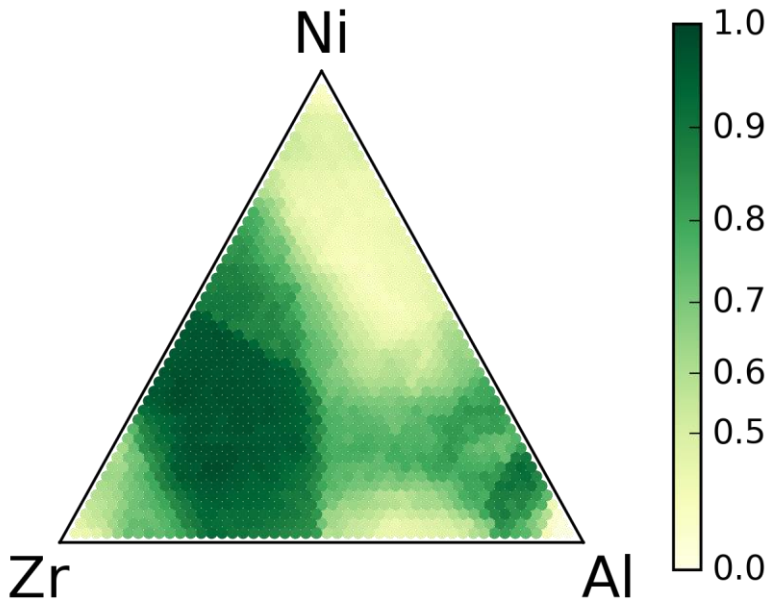
Predicting Glass-Forming Ability

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

Measured



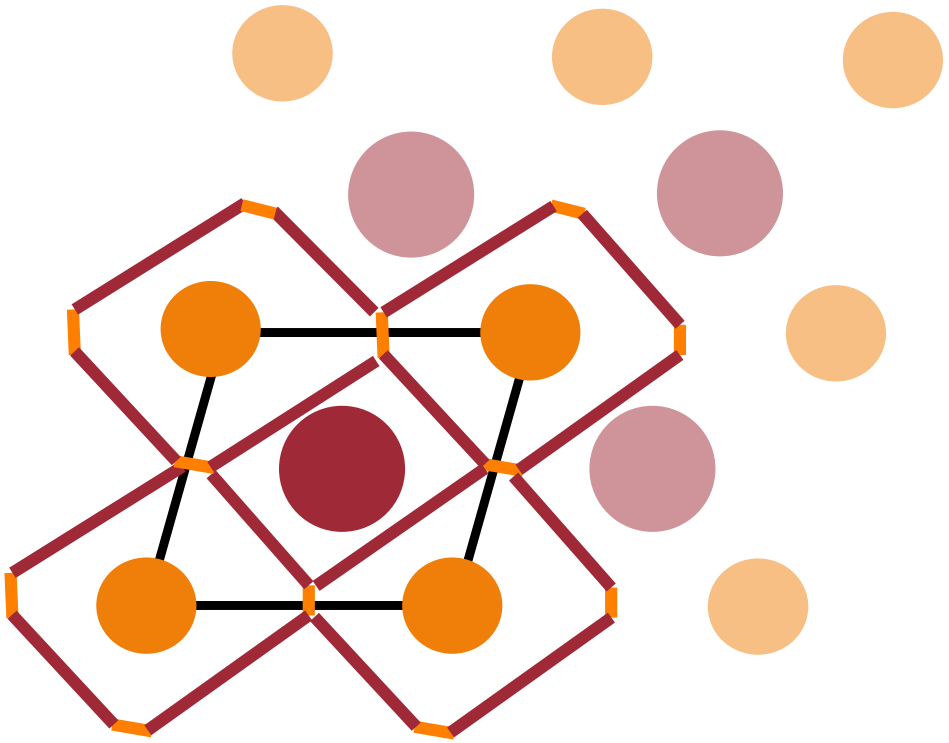
Predicted



Same representation, very different material.

Ref: Ward *et al*. npj Comp. Mat., (2016) 28

Focus #2: Crystal Structure



Our Approach:

Voronoi-tessellation-based attributes

Atomic Characteristics:

1. Element identity
2. Coordination number
3. Bond length
4. Cell size

…

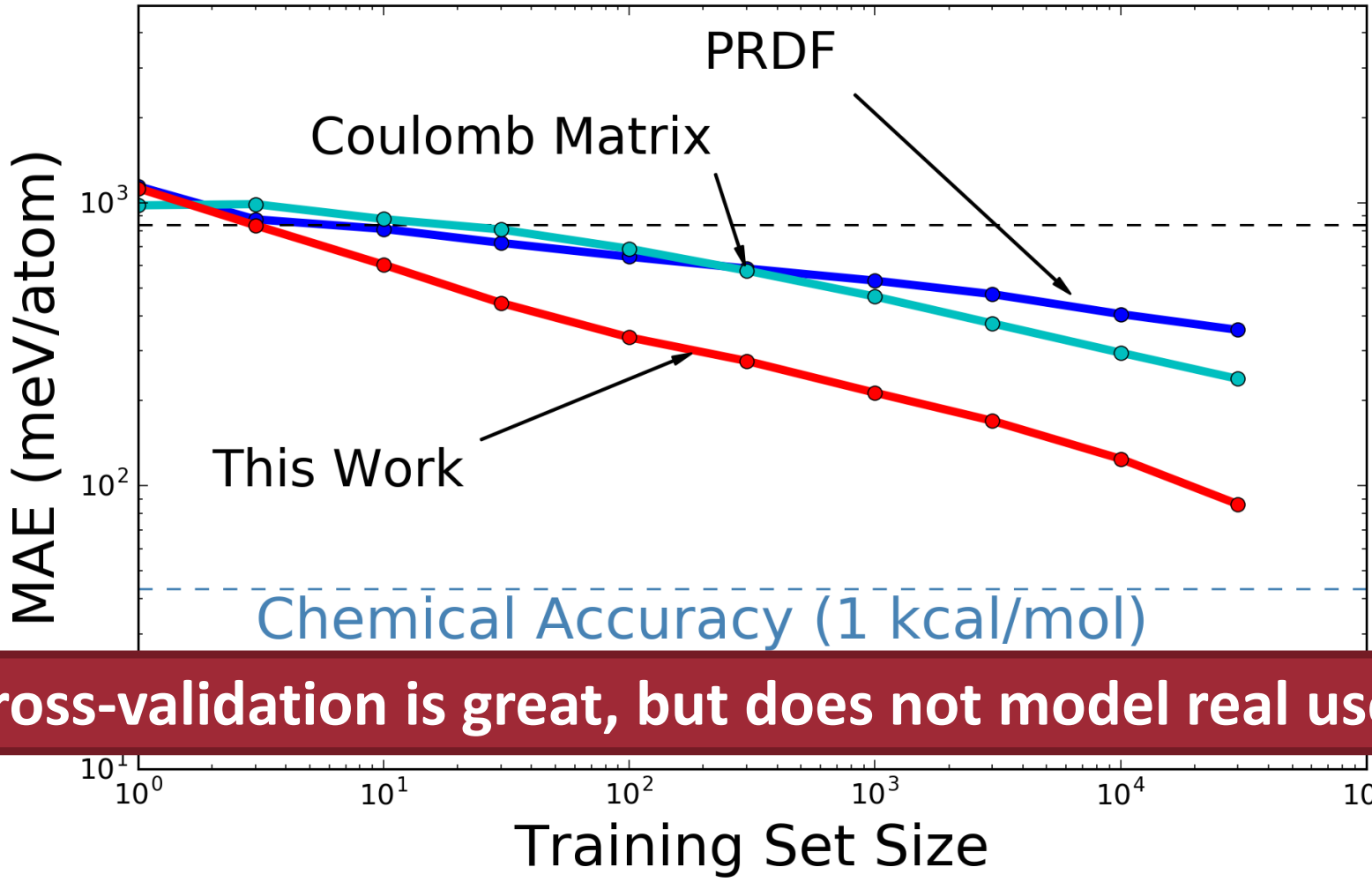
Atomic Characteristics + Descriptive Statistics = 275 Attributes

Ref: Ward *et al*. PRB. (2017) 024104

Learning Rate Comparison

Dataset: 32k DFT ΔH_f from the OQMD

Test: Remove 1000, train on *N* remaining



Cross-validation is great, but does not model real use

CM: Faber *et al*. Int J Quantum Chemisty. 2015; PRDF: Schutt *et al.* PRB. (2014)

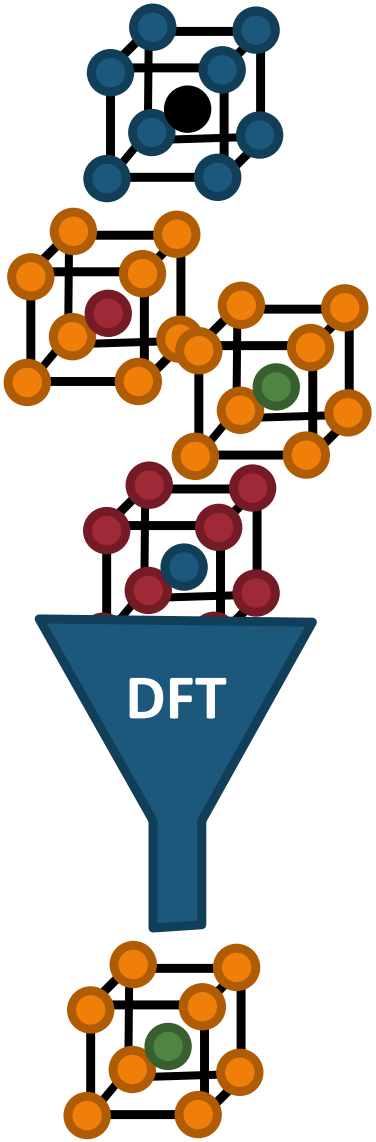
Application: The Prototype Search

Common Method: Prototype Search

1. Select a crystal structure
2. Evaluate **all** possibilities with DFT
3. Select only stable ones

Challenge: Computational cost

Possible Solution: Guide with ML

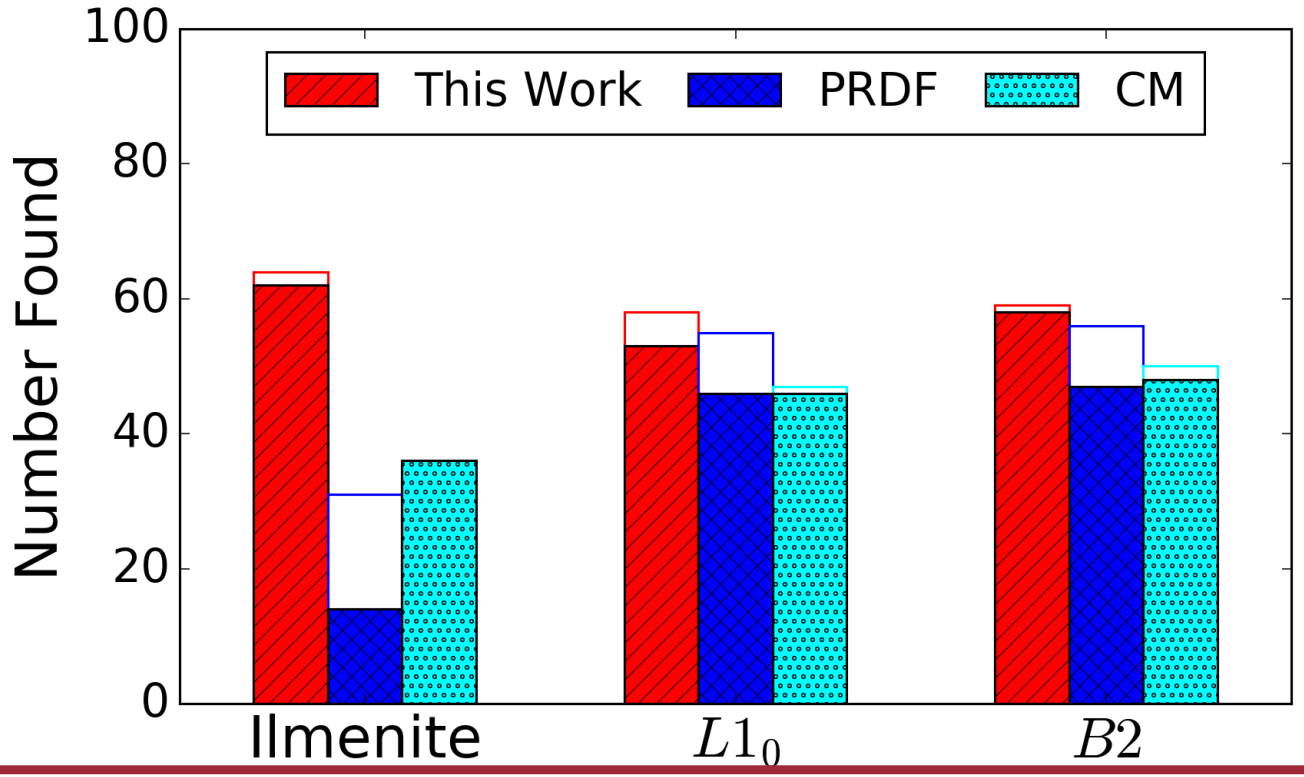


Ref: Ward *et al*. PRB. (2017) 024104

Ranking Candidate Materials

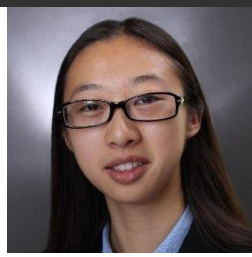
Training Set: 32k entries from OQMD

Test: Select top 100 entries from test set



Ranks entries better than existing methods, good choice for accelerating combinatorial searches

Future Steps: Open Benchmarks



Jiming Chen (UIUC)

Challenge: Which method is best for a certain dataset?

Our proposal: Develop a set of benchmark challenges

Approach: Host data, models using MDF and WholeTale

Current Progress: Compared several composition-based models

WholeTale.org

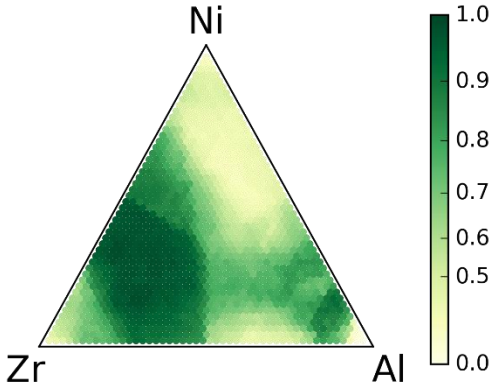
MATERIALS DATA FACILITY

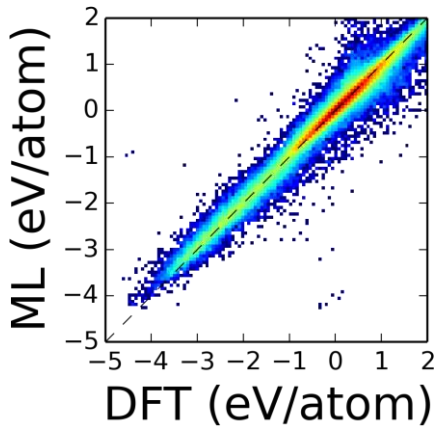
Datasets	Methods			
	Ward 2016	**Deml 2016**	**Meredig 2014**	**deJong 2016**
Full OQMD	**0.161**	0.180	0.214	0.187
Deml 2016	0.170	0.160	**0.144**	0.176
Meredig 2014	**0.072**	0.140	0.087	0.091

Summary

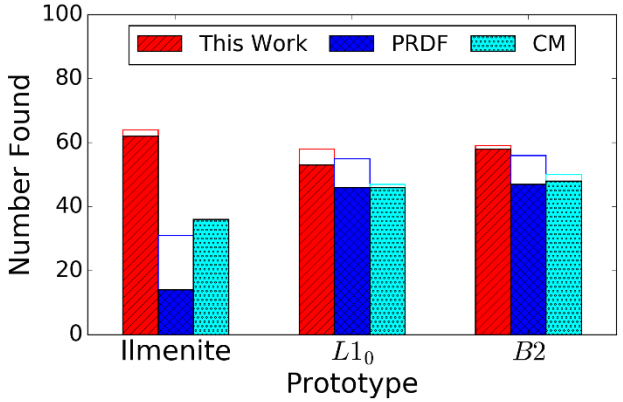
Collect → Process → Represent → Learn

Composition-Based Attributes





Crystal Structure Attributes



How do I use this to find a new material?

Part 3: Using ML To Find Materials

Acknowledgments:

Liquidmetal: S. O'Keeffe, J. Stevik, G. Jelbert

NU: M. Aykol, K. Kim, J. He

Example Design Problems

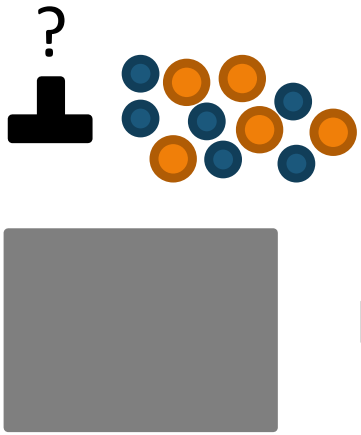
Designing Metallic Glasses Alloys

1. Optimizing commercial alloys
2. Locating new glasses with sputtering

Discovering Quaternary Heuslers

Application: Bulk Metallic Glasses

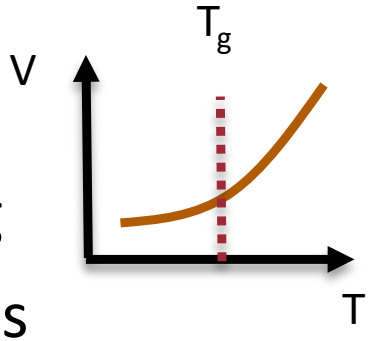
Metals with Amorphous Structures



No Dislocations -> High σ_Y

Glass Transition -> Net Shape Casting

No Grain Boundaries -> Low Hysteresis



Applications: Surgical tools, flight control surfaces, …

Source: liquidmetal.com

What are the tradeoffs?

Source: Wikipedia.org

Problems with BMGs

Main Problem: Metastability of glassy phase

GFA *Few Alloy Systems Known to Form Glasses*

<2% possible ternaries in Landolt-Börnstein handbook

D_{max} *Small Critical Casting Thicknesses*

Typical thickness < 10 mm

ΔT_x *Small Time/Temperature Processing Window*

BMGs may crystallize during processing

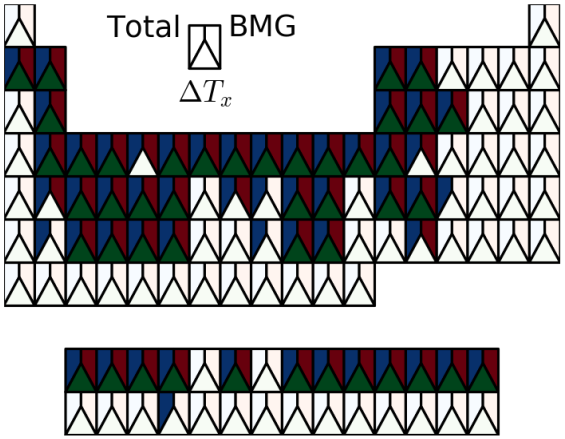
Design Issue: Lack of Composition -> Properties Links

Our Proposed Solution: Data analytics

Goal: Optimize D_{max} and ΔT_x

Materials Informatics Workflow

Collect → Process → Represent → Learn

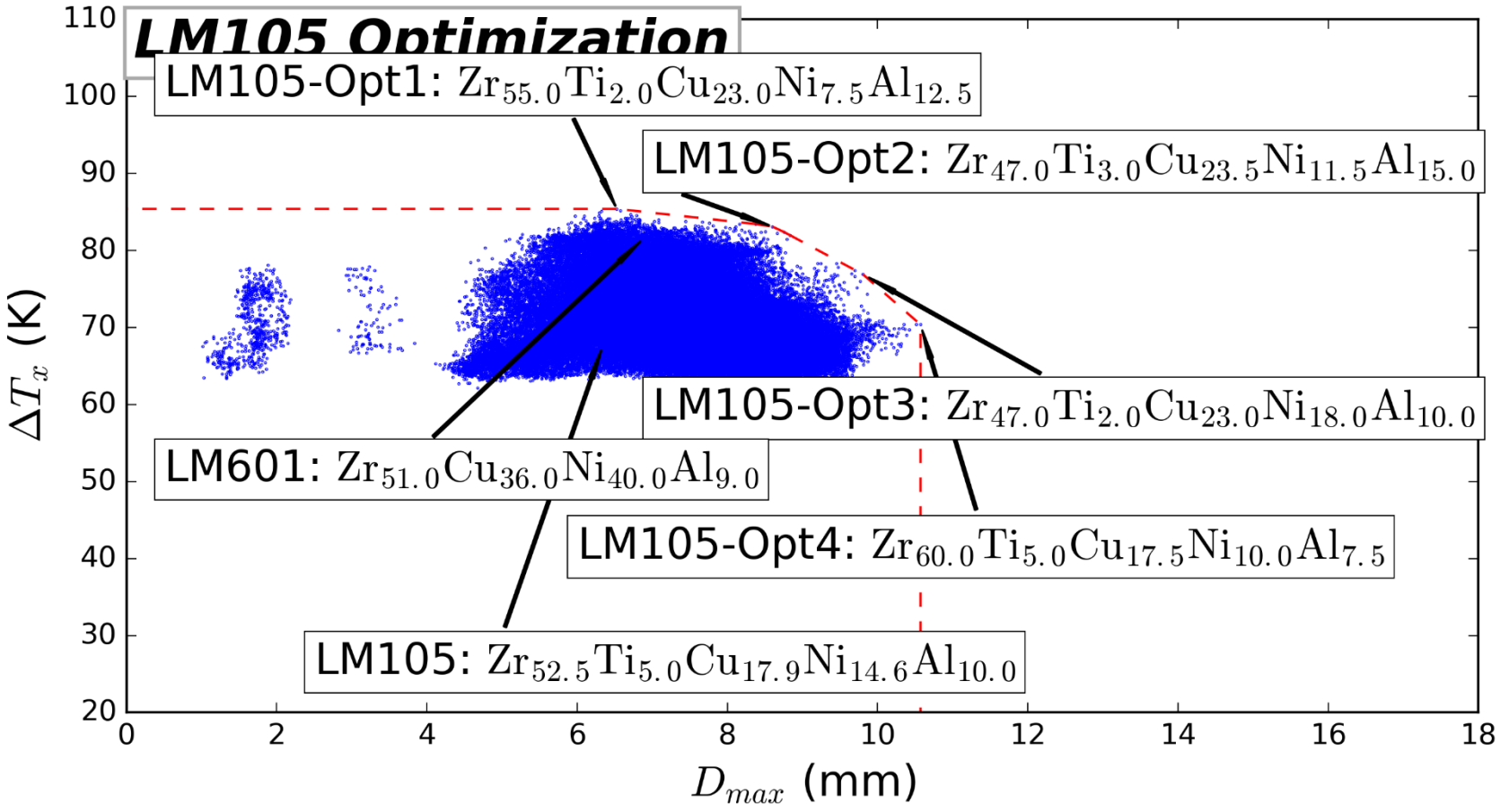


$$\vec{X} = f(Composition)$$

"General Purpose" Attributes
+ Cluster Packing Models
+ DFT ΔH_f from OQMD

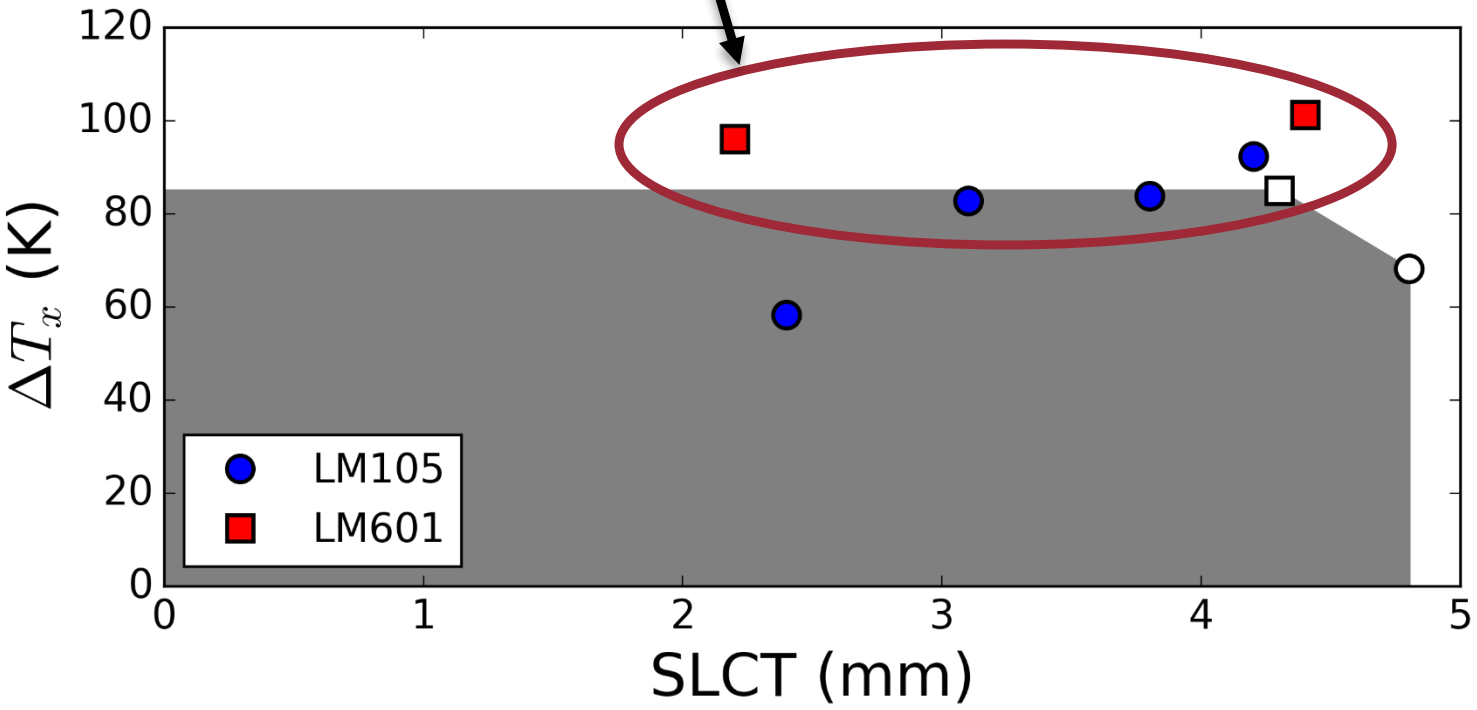
GFA	6315 entries	GFA	Random Forest
D_{max}	5916 entries	D_{max}	Random Forest + Additive Regression
ΔT_x	621 entries	ΔT_x	

Task #1: Tuning Known Alloys



Tuned Alloys: Pareto Analysis

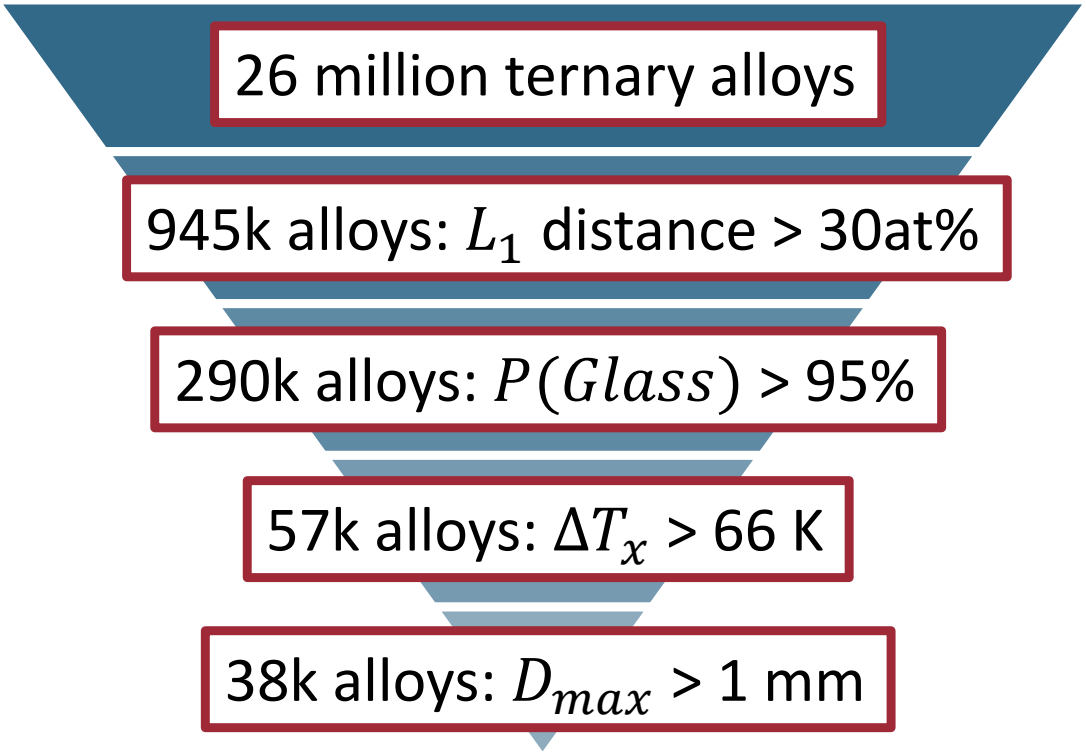
Predicted alloys have superior properties



Results: Improved design space

Task #2: Discovering New Alloys

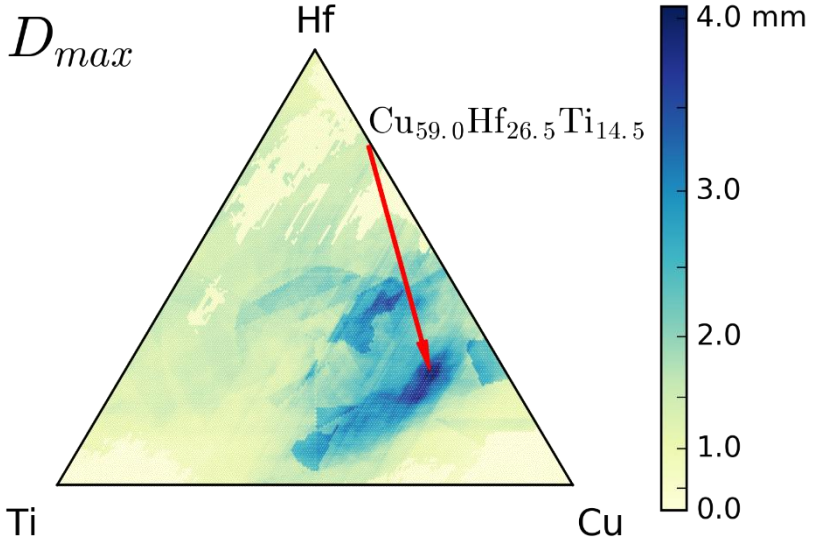
Search Space: Ternary alloys of 53 elements

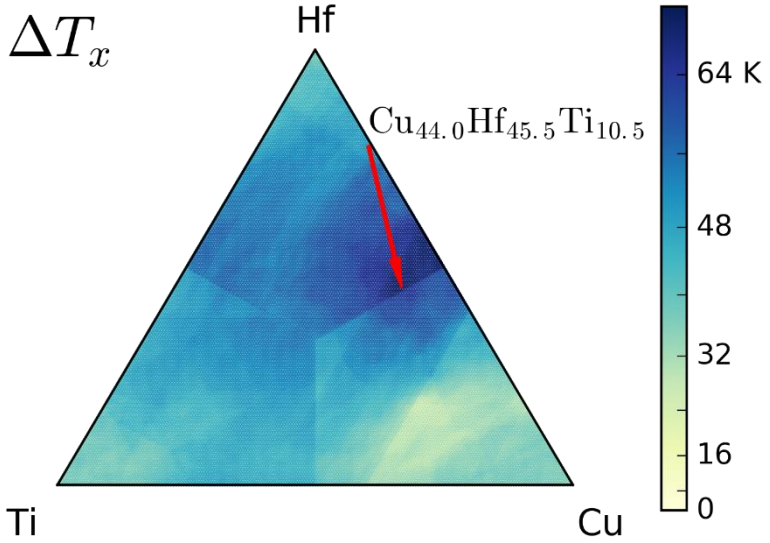


Run Time: ~ 2 days

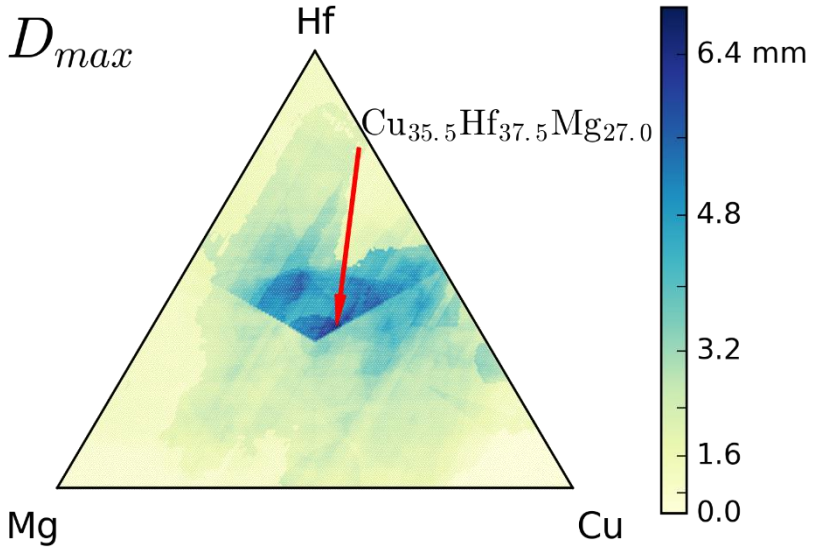
Chosen System: Cu-Hf-[Mg,Ti]

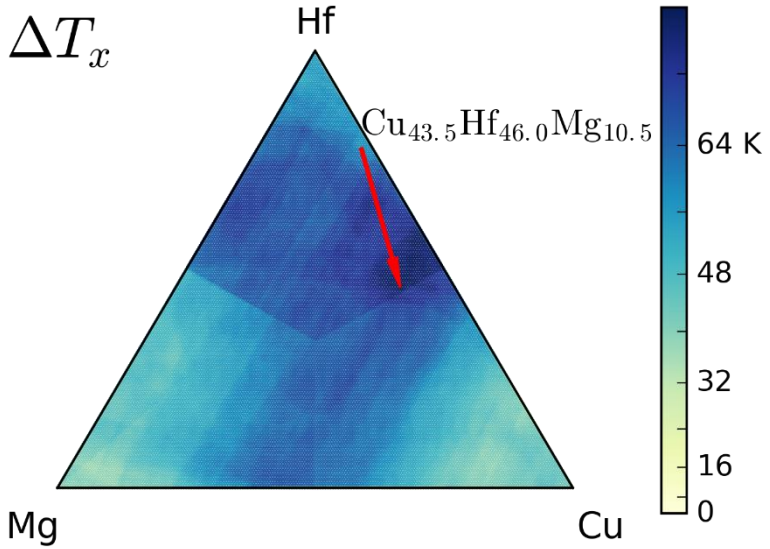
Cu-Hf-Ti





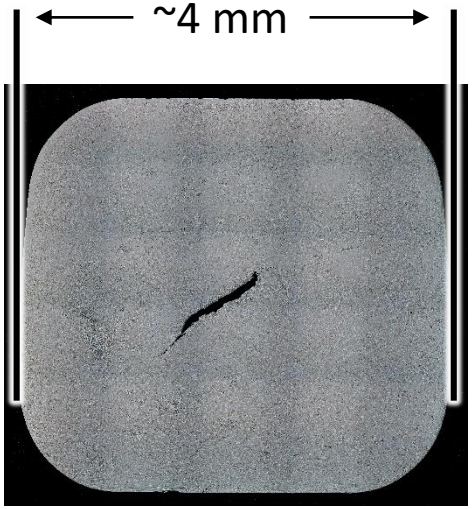
Cu-Hf-Mg





Cu-Hf-[Ti/Mg]: Current results

Cu-Hf-Ti



Sample crystallized and/or shattered during injection molding

- $Cu_{60}Hf_{25}Ti_{15}$ could be formed with copper mold casting*

Cu-Hf-Mg

Mg ignited during arc melting

Lessons Learned:

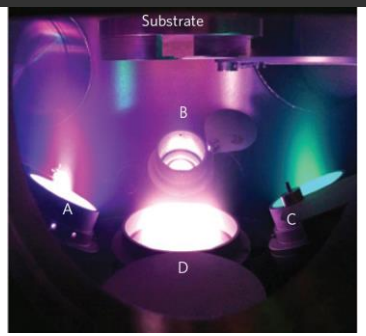
- Integrate data from target process
- Screen based on processing technique

*Inoue *et al*. Acta Mat. (2001), 2645

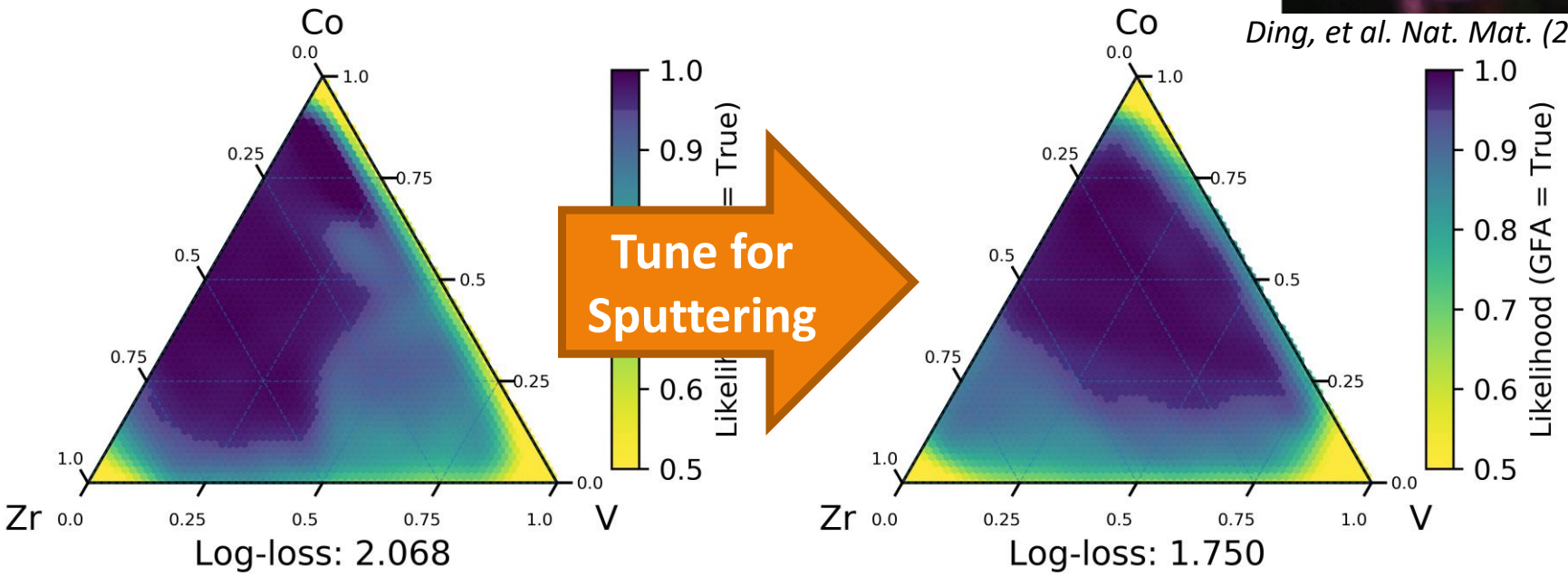
Example #2: Finding New Glasses

Goal: Find new sputtering glasses

Step 1: Include processing information

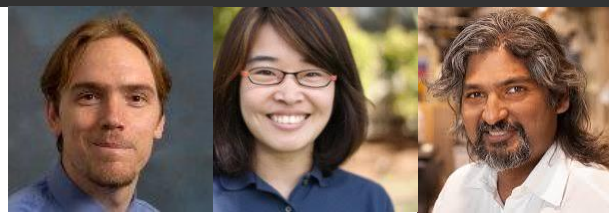


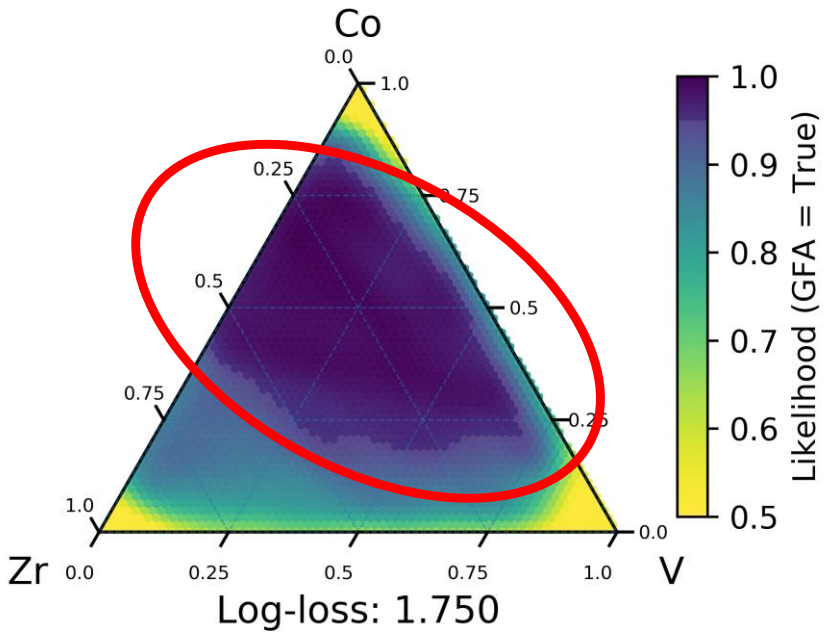
Ding, et al. Nat. Mat. (2014)

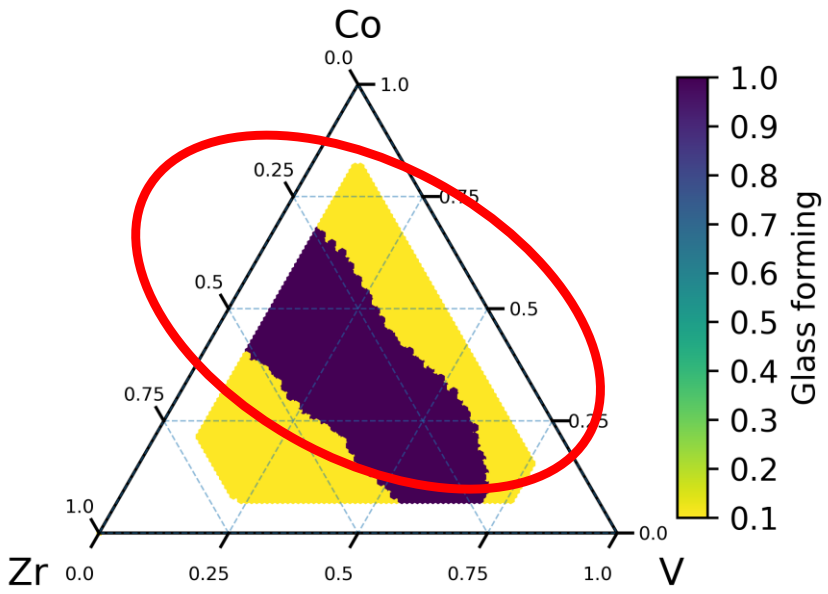


Step 2: Scan 2.4M alloys (~5 min) -> Selected Co-V-Zr

Comparison to Experiment

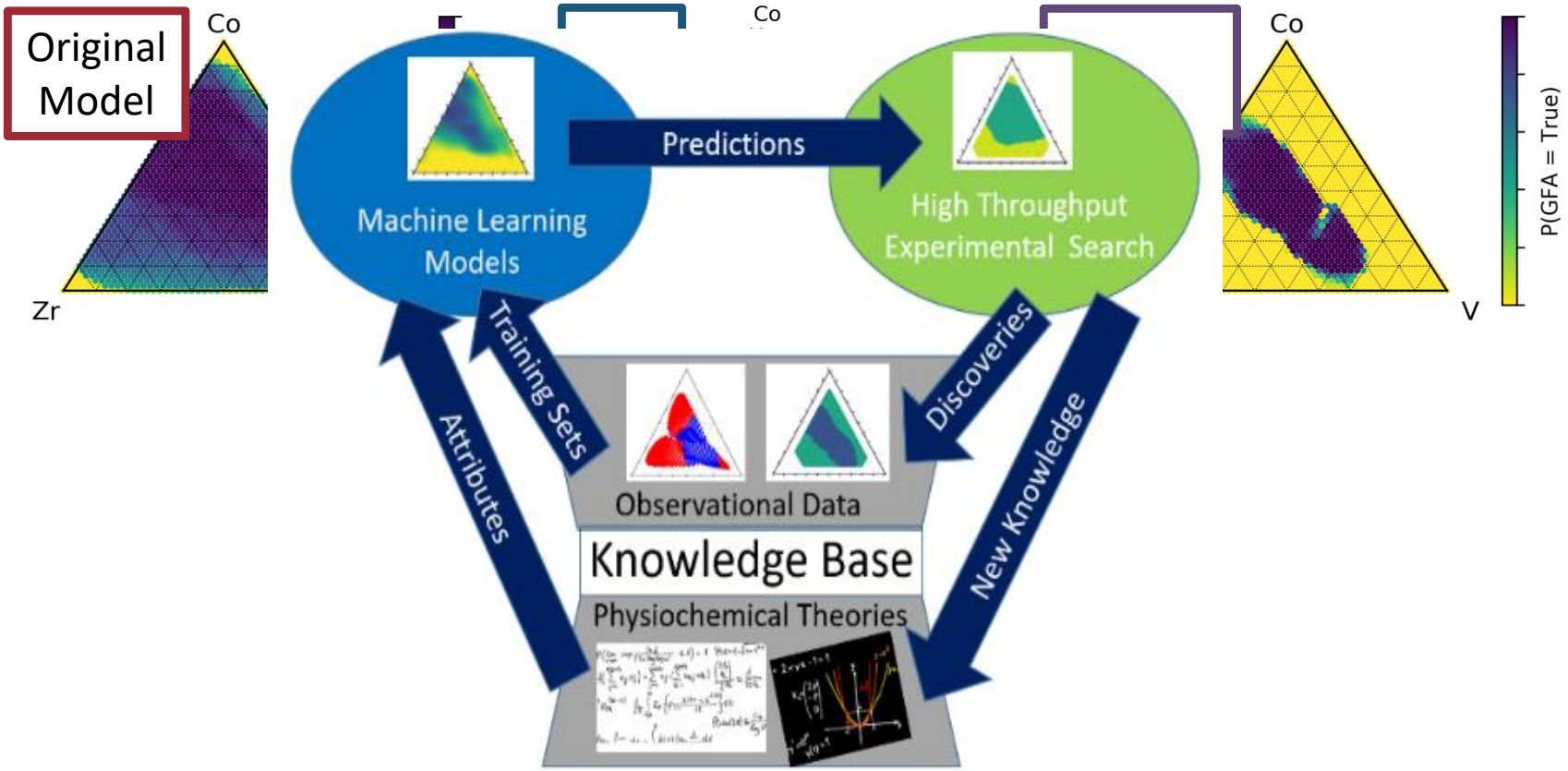






Decent agreement. ML model “Zr-lean”, but close enough for success

Repeat, with Improved Model



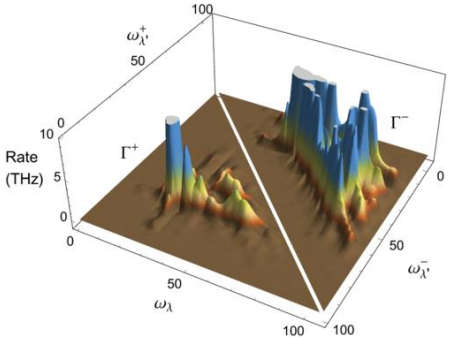
Initial Model → Add Co-V-Zr → Improved Accuracy

Example #2: Finding Compounds

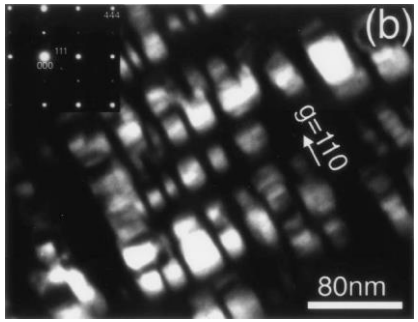
Goal: Find more Quaternary Heuslers (QHs)

Why quaternary Heuslers?

Ternary Heuslers have great properties,



He *et al*. PRL. (2016), 046602

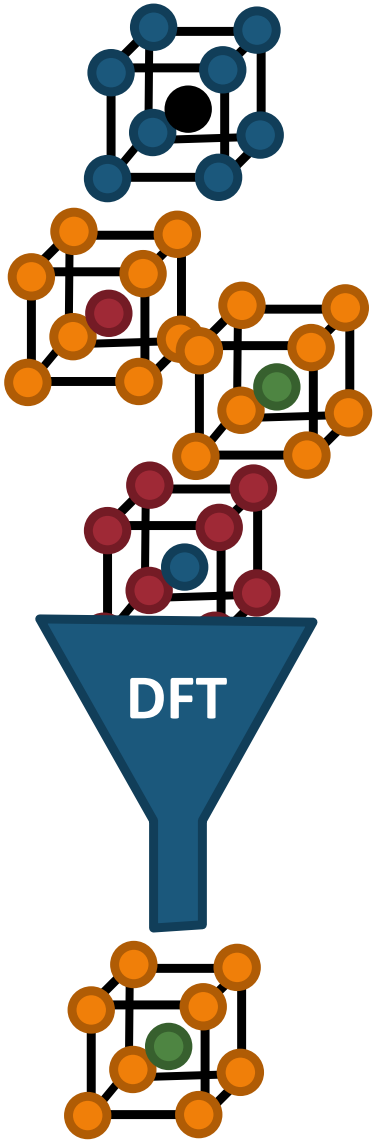


Jung *et al*. MMTA. (2003), 1221

…so a 4th degree of freedom could be better?

Problem: ~3M possible combinations

Solution: Guide search with ML



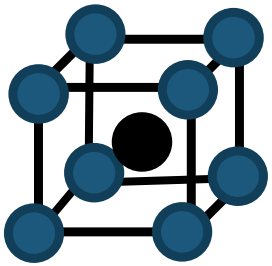
Ref: Kim *et al*., *in preparation*

Materials Informatics Workflow

Collect → Process → Represent → Learn

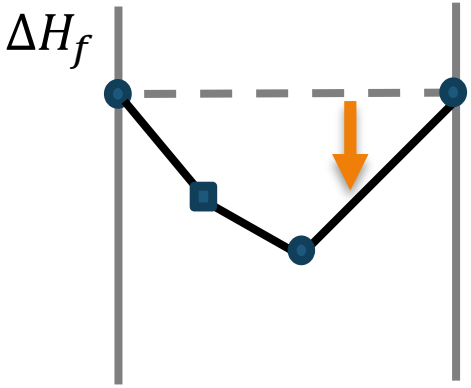
OQMD

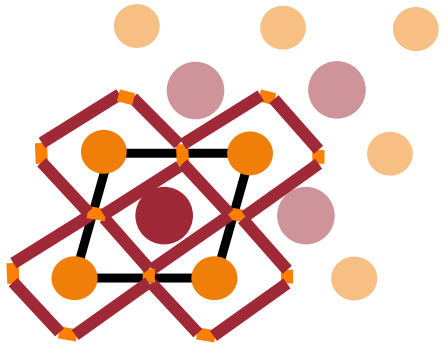
Input: Crystal structure
Output: Stability



Training Data: OQMD

Entries: 410981
96160 QHs





Voronoi-Tessellation Method

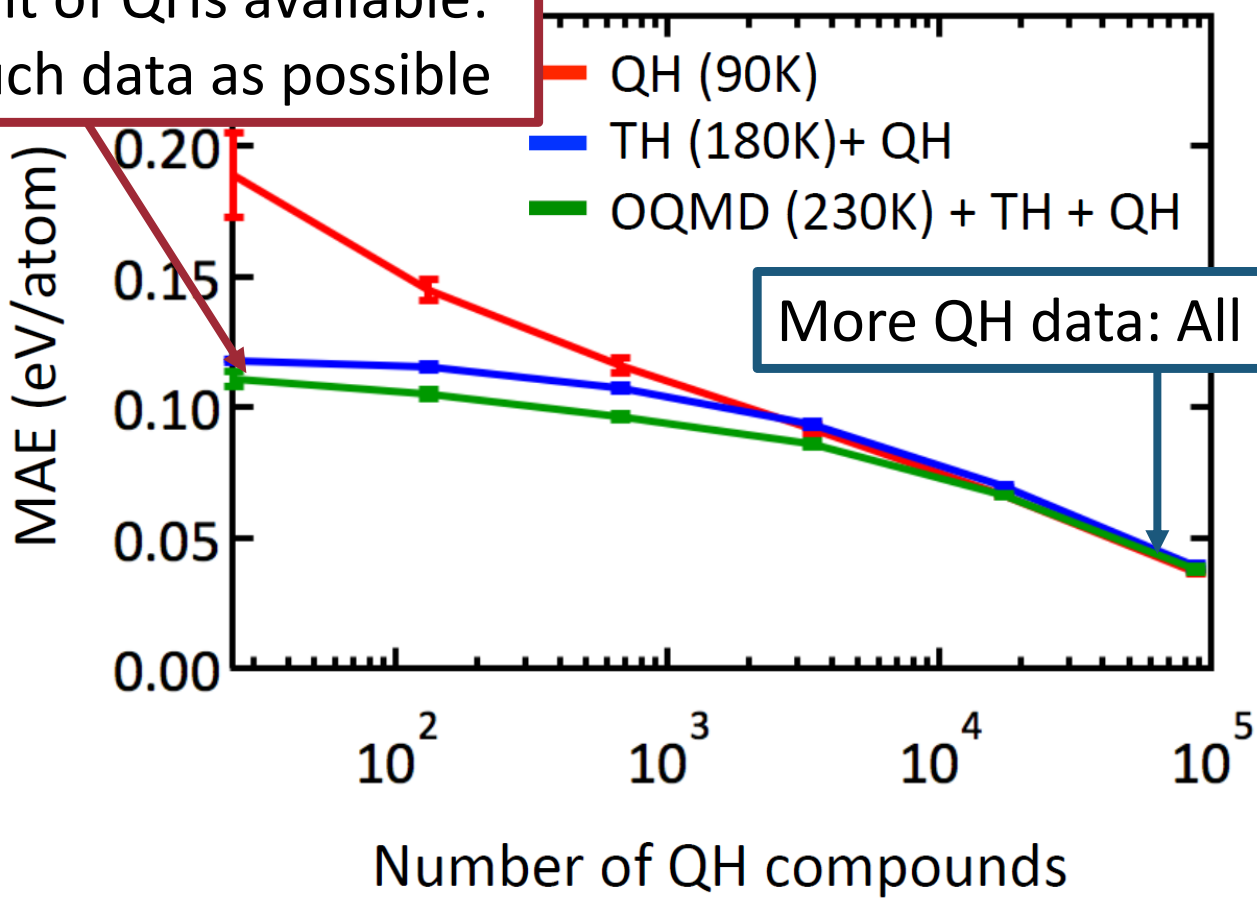
275 attributes

Random Forest

Step 1: Select Training Set, Validate

Training data: All data available, or only similar materials?

Small amount of QHs available:
Select as much data as possible



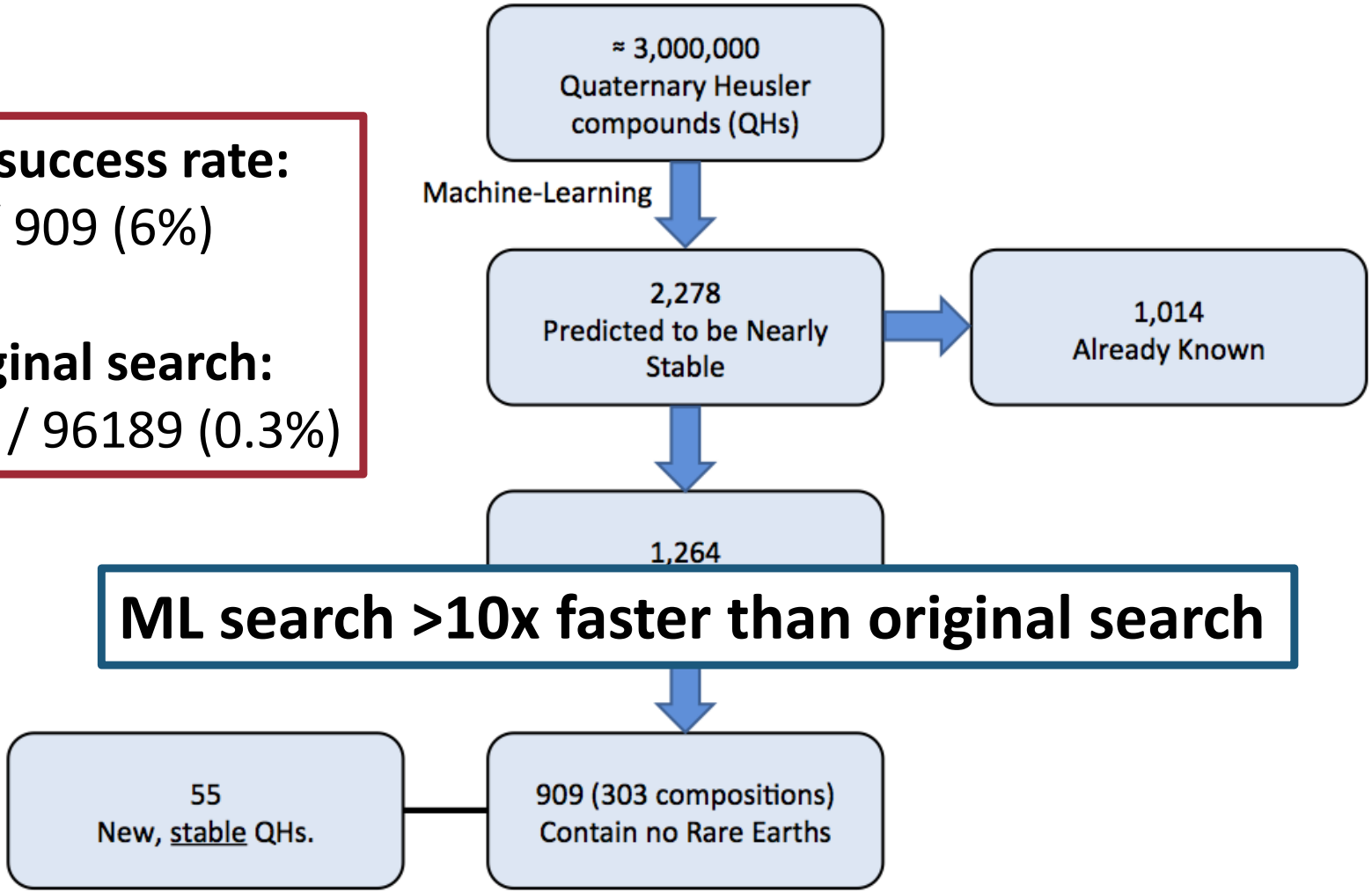
More QH data: All equivalent

Bottom Line: Training on *all available data* is advantageous

Step 2: Find New Heuslers

ML success rate:
55 / 909 (6%)

Original search:
353 / 96189 (0.3%)



Summary

Automated Crystal Structure Solution

- ✓ Validated FPASS method, created automatic framework
- ✓ Solved 13 crystal structures

General-Purpose Representations

Using Composition: DFT ΔH_f, E_g, V; Glass-forming ability

Using Crystal Structure: 2x better than existing approaches

Example applications:

- ✓ Improved ΔT_x of 2 commercial BMGs alloys
- ✓ Discovered many new metallic glasses
- ✓ Found dozens of new quaternary Heuslers

Open Source Software: Magpie, Matminer

http://bitbucket.org/wolverton/magpie

http://github.com/hackingmaterials/matminer

Acknowledgements

Funding:

- CHiMaD/NIST
- NDSEG Fellowship Program
- Weertman Fellowship, Ryan Fellowship

Wolverton Group

Vinay Hedge, Murat Aykol, Max Amsler, Scott Kirklin, Yang Yu, Kyoungdoc Kim

Choudhary Group

Ankit Agrawal, Rosanne Liu, Amar Krishna

Liquidmetal Technologies

Stephanie O'Keeffe, Joseph Stevik, Glenton Jelbert

NIST/USC/SLAC

Apurva Mehta, Fang Ren, Jason Hattrick-Simpers, Travis Williams