Ab Initio Crystal Structure Prediction: High-throughput and Data Mining

Dane Morgan

Massachusetts Institute of Technology

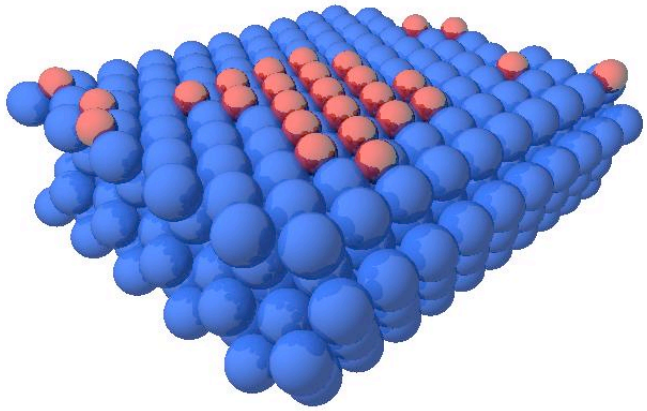
NSF Division of Materials Research

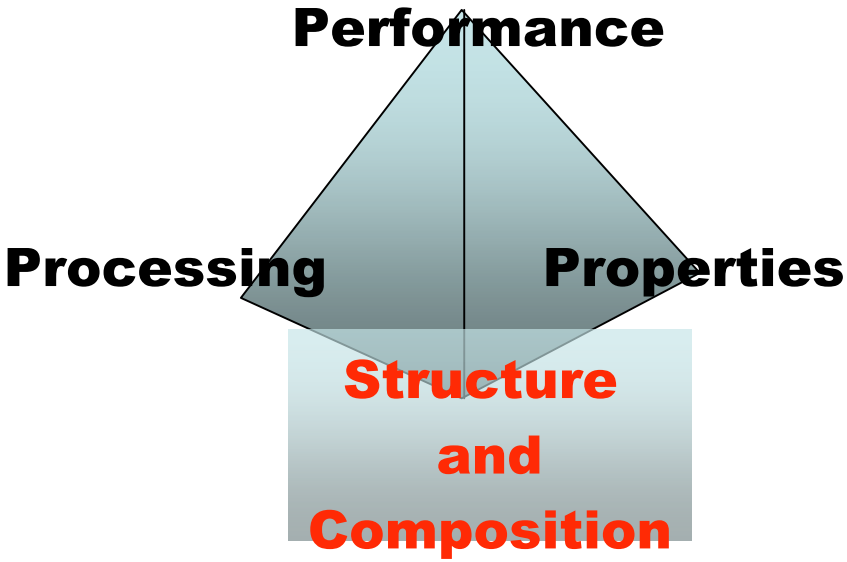
ITR Computational Workshop

June 17-19, 2004

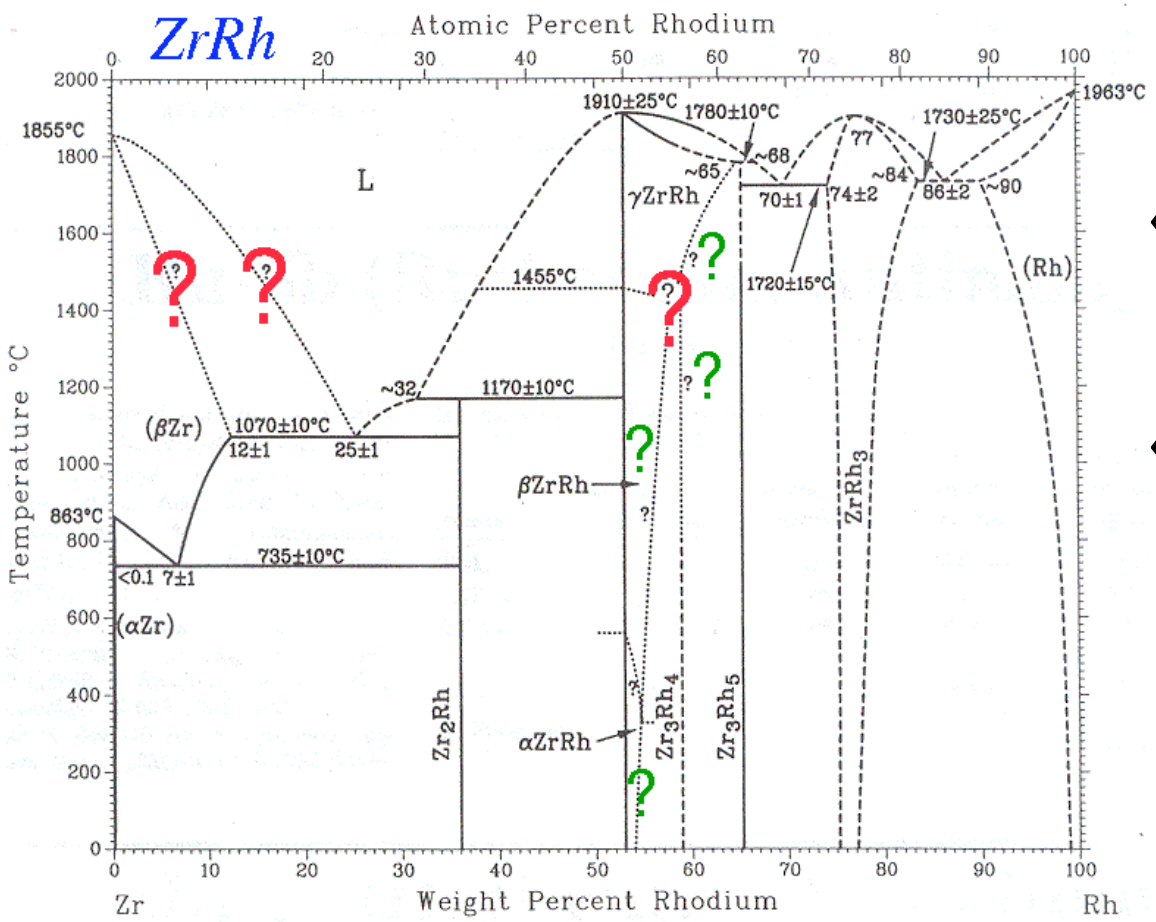
Why Does Structure Matter?
Essential for Rational Materials Design

- Structure key to understand properties and performance
- Key input for property computational modeling





Why Do We Need Structure Predictions? Structural Information is Often Lacking



Massalski, *Binary Alloy Phase Diagrams* '90

- Binary alloys incomplete
- Multi-component systems largely unknown

The Structure Prediction Problem

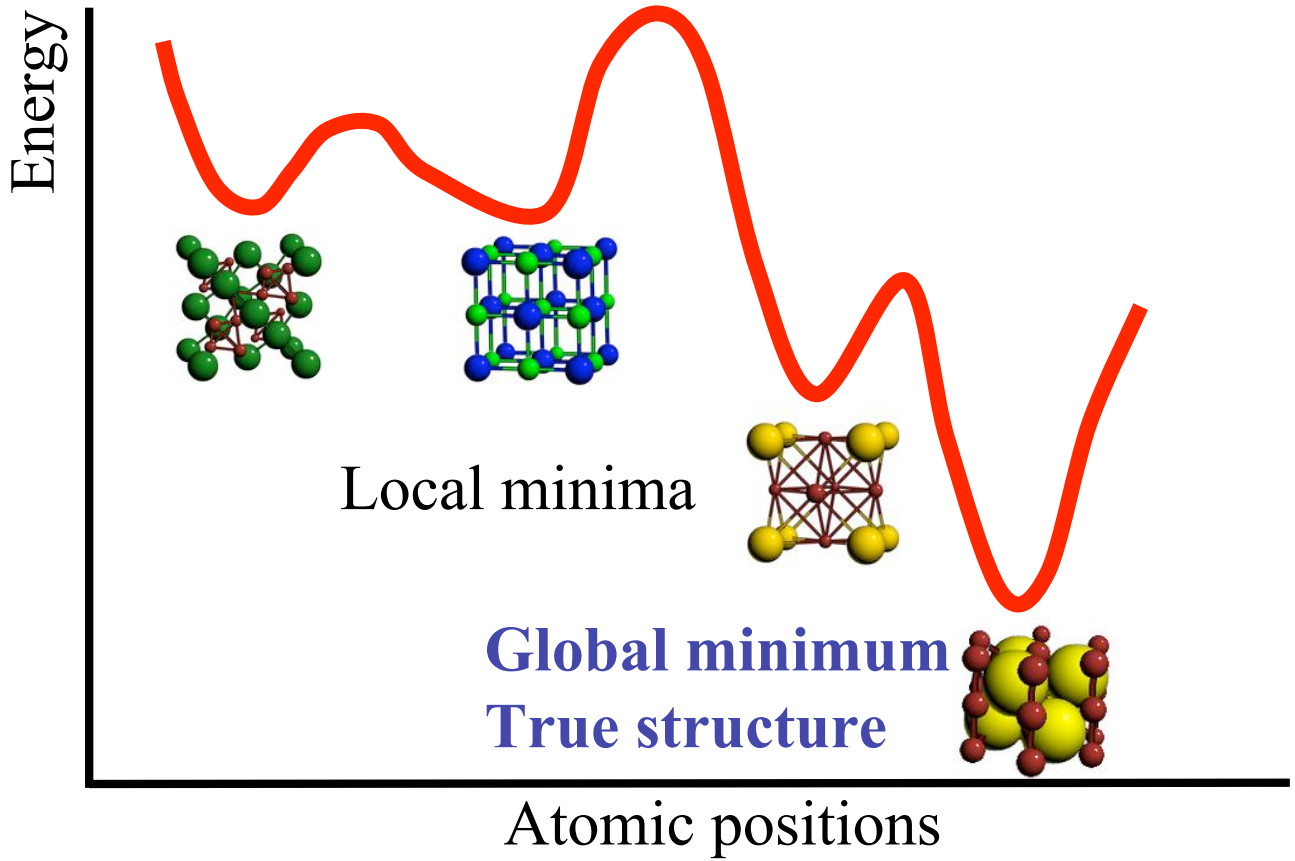
Given elements A, B, C, …
predict the stable low-temperature phases

Present focus

Crystalline phases
***Ab initio* methods**

Why is Structure Prediction Hard?

Ab initio methods give accurate energies, but …



- Infinite structural space
- Rough energy surface – many local minima

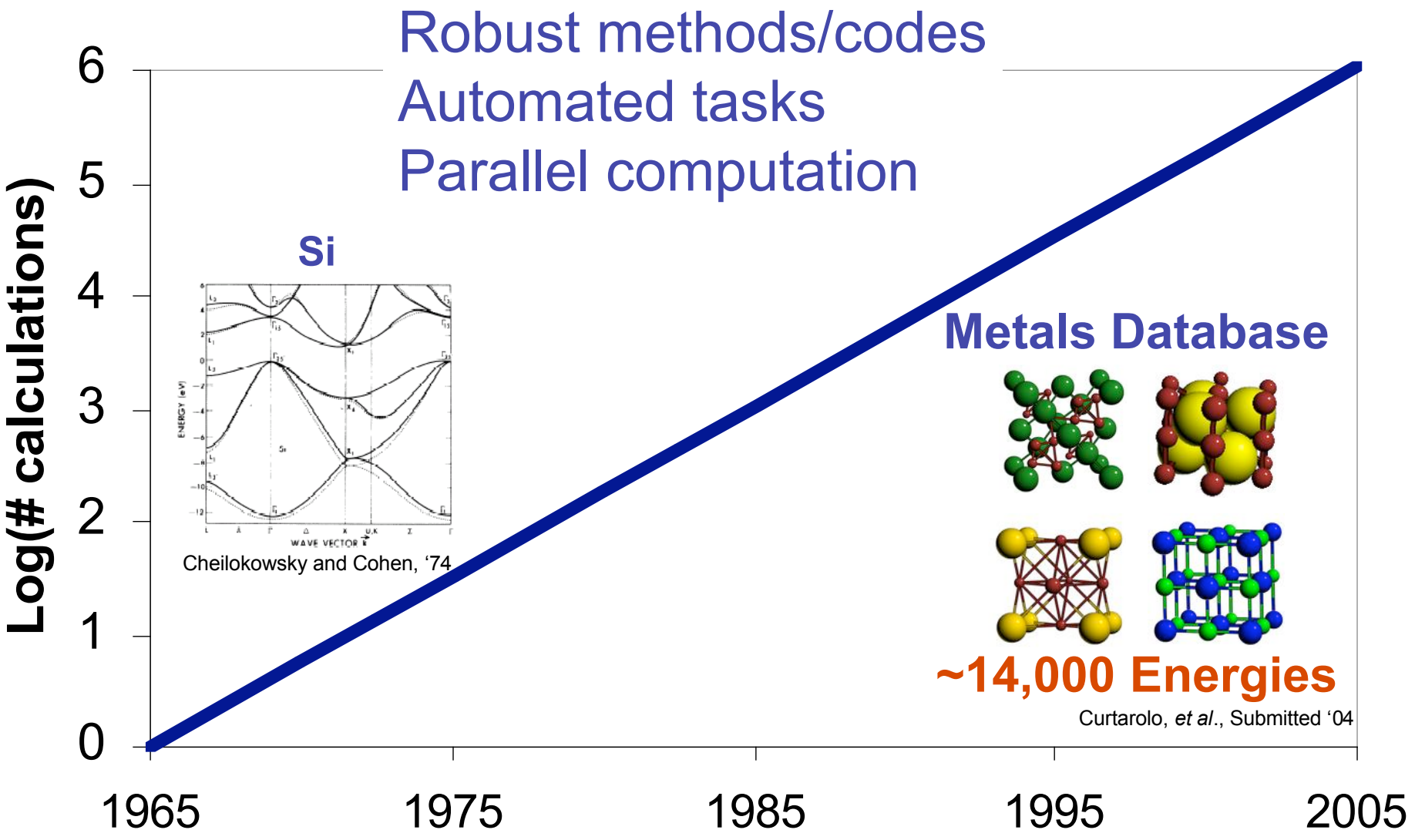
Two New Tools

High-Throughput *Ab Initio*

υ

Data Mining
Calculated/Experimental
Databases

High-Throughput *Ab Initio*



Ab Initio Structure Prediction

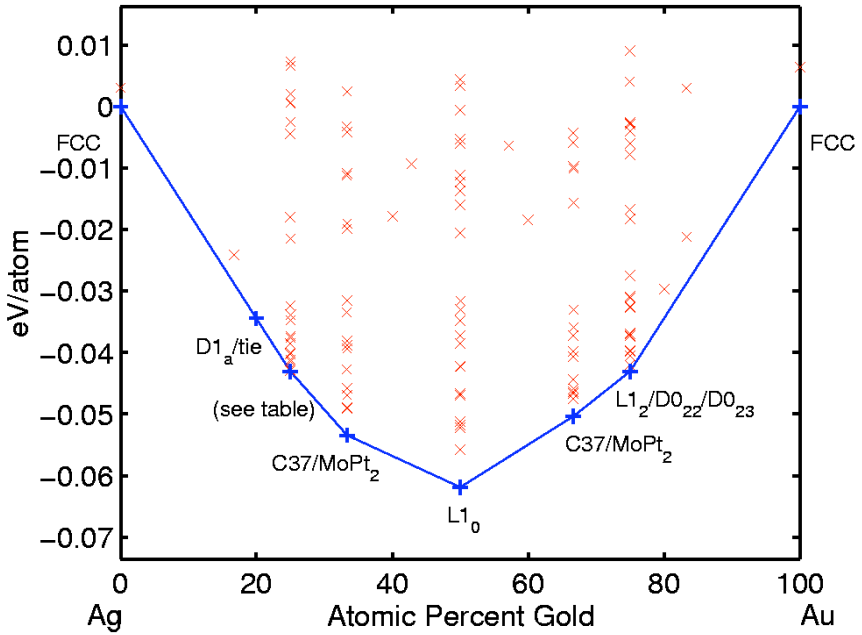
Obtain a manageable list of likely **candidate structures** for high-throughput calculation

- **Directly optimize *ab initio* Hamiltonian** with Monte Carlo, genetic algorithms, *etc*. (too slow)
- **Simplified Hamiltonians** – potentials, cluster expansion (fitting challenges, limited transferability/accuracy)
- **Intelligent guess** at good candidates

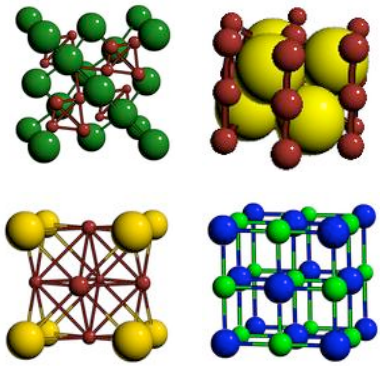
How good can this be?

“Usual Suspects” Structure List

80 binary intermetallic alloys
176 “usual suspects” structures
(“usual suspects” = Most frequent in CRYSTMET, hcp, bcc, fcc superstructures)



Metals Database

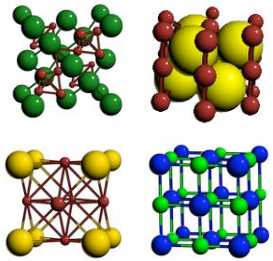


~14,000 Energies

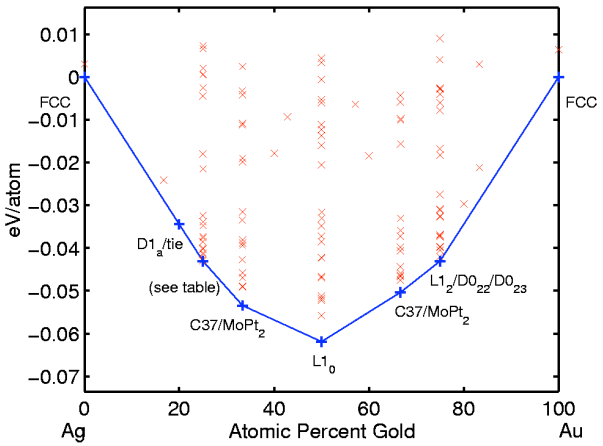
Calculate energies
Construct convex hulls
Compare to experiment

High-Throughput Predictions

Metals Database



~14,000 Energies

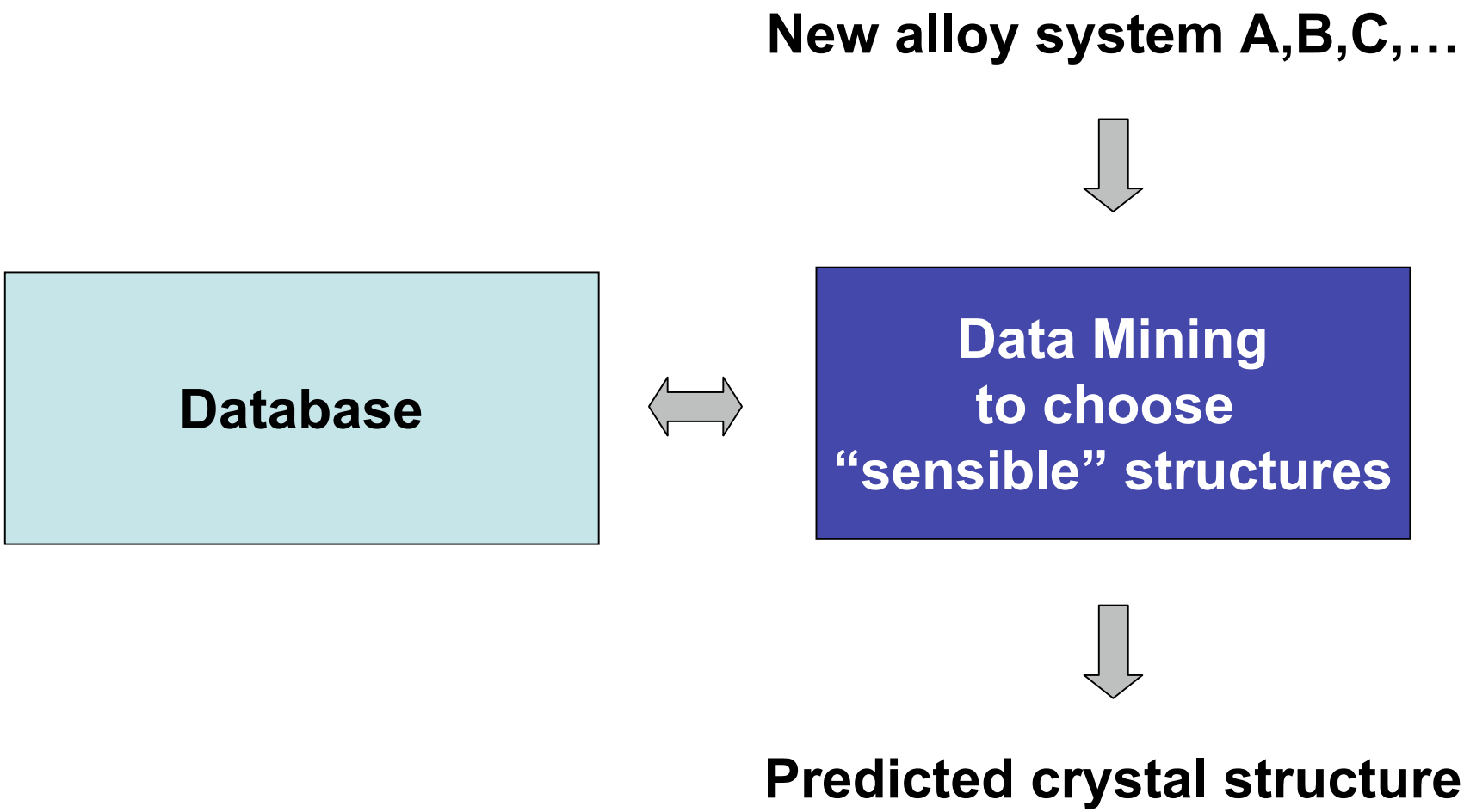


- 95 predictions of new compounds
- 21 predictions for unidentified compounds
- 110 agreements
- 3 unambiguous errors

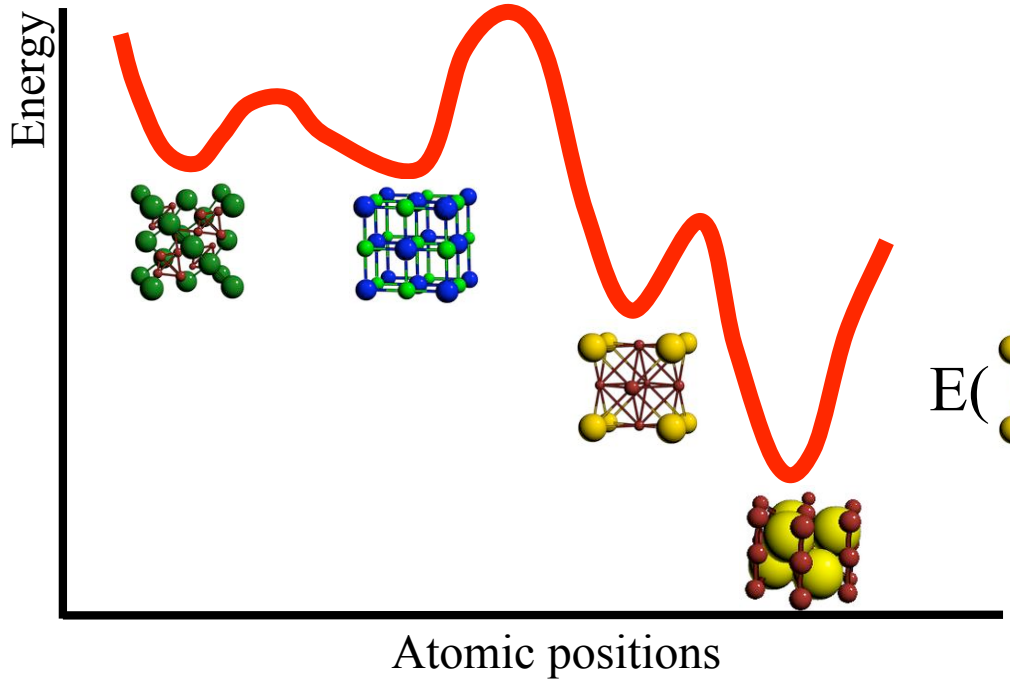
Curtarolo, *et al*., Submitted '04

But far too many structures + alloys to explore!!
Need smart way to choose "sensible" structures!!

Data Mining


New alloy system A,B,C,…
Database
Data Mining
to choose
“sensible” structures
Predicted crystal structure

Data Mining with Correlations



$$E(\ ) = E(\ )$$

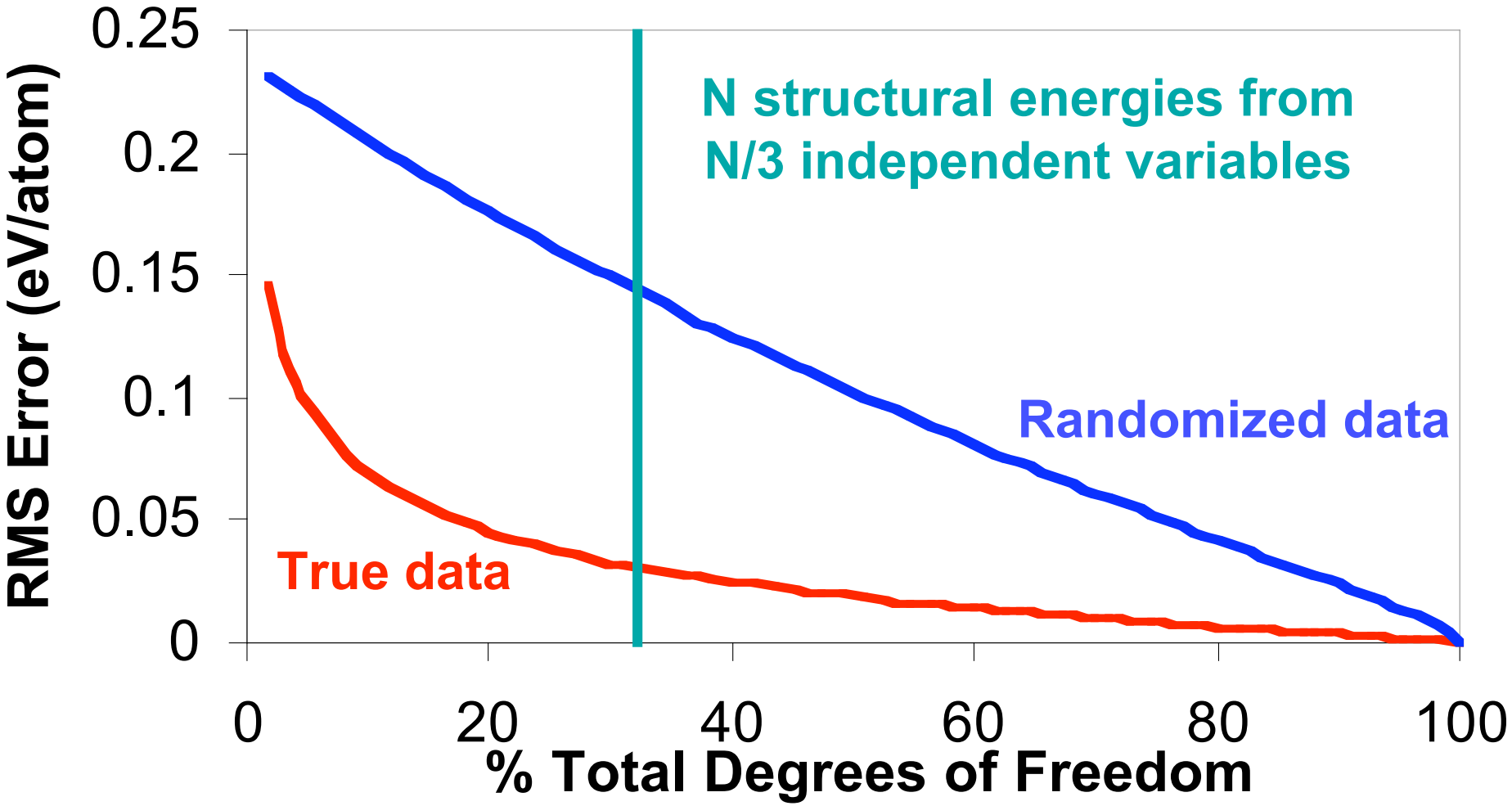
$$E(\ ) = \frac{1}{2}[E(\ ) + E(\ )]$$

Linear correlations between energies ⇨ All energies do not need to be calculated ⇨ Faster to find low energies

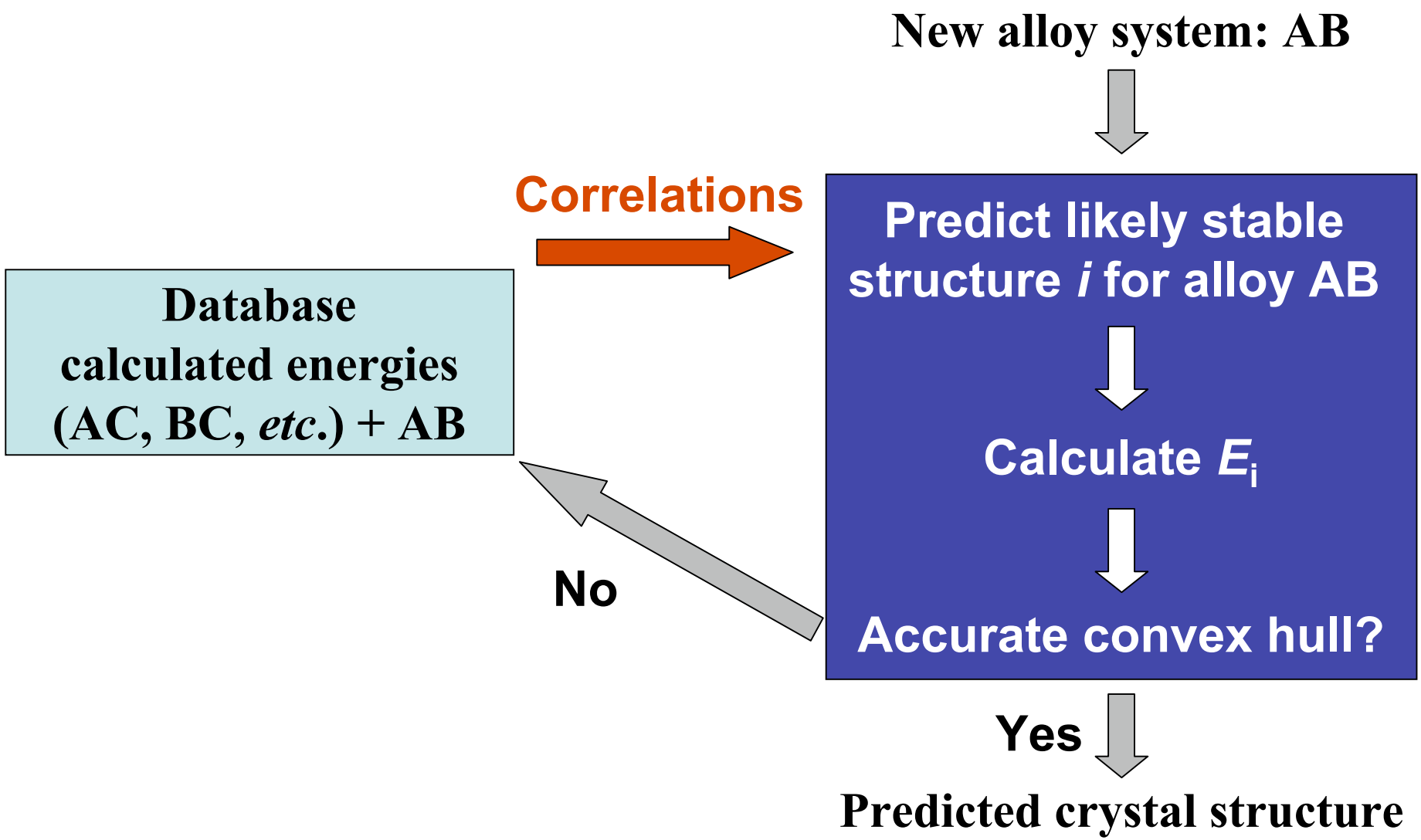
Do linear correlations exist between structural energies across alloys?

Structural Energy Correlations Exist!

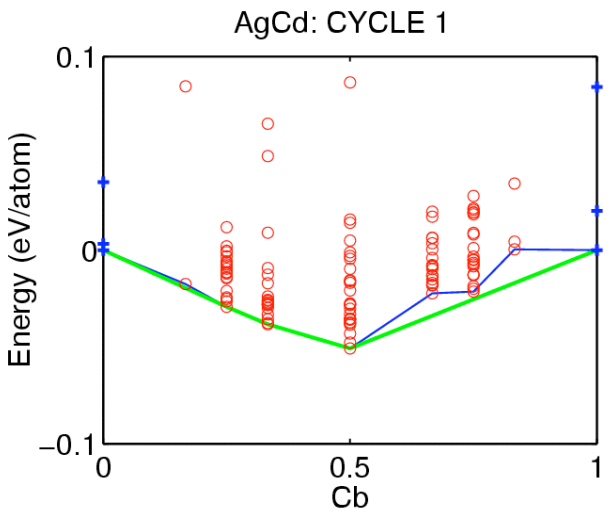
Principal Component Analysis identifies correlations

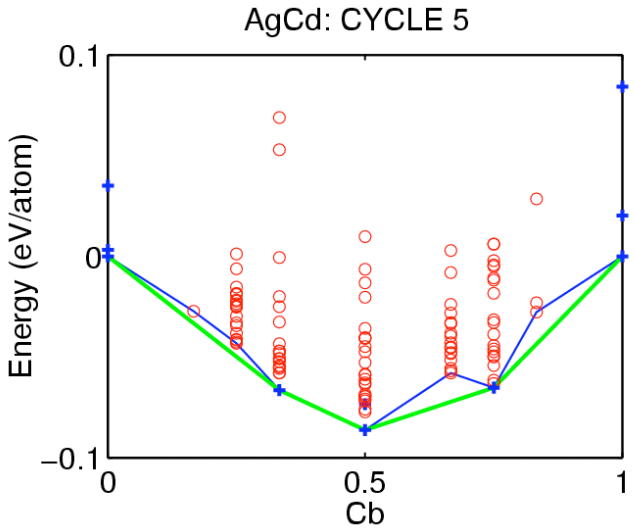


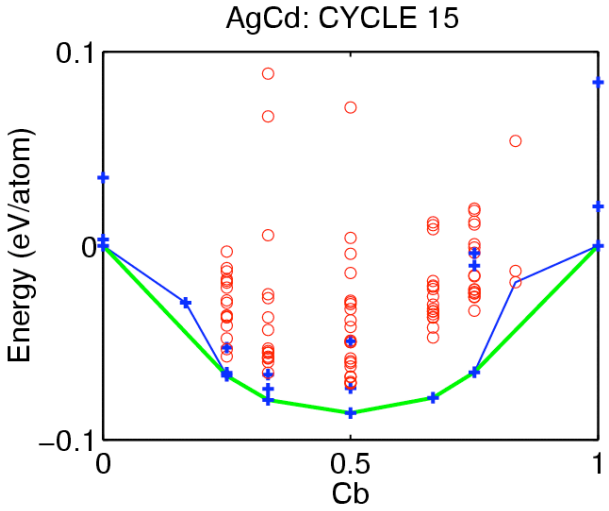
Using Correlations for Structure Prediction

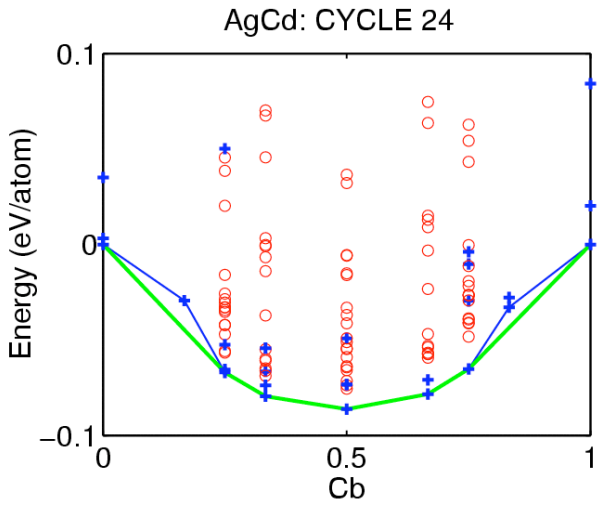


Data Mining Example: AgCd

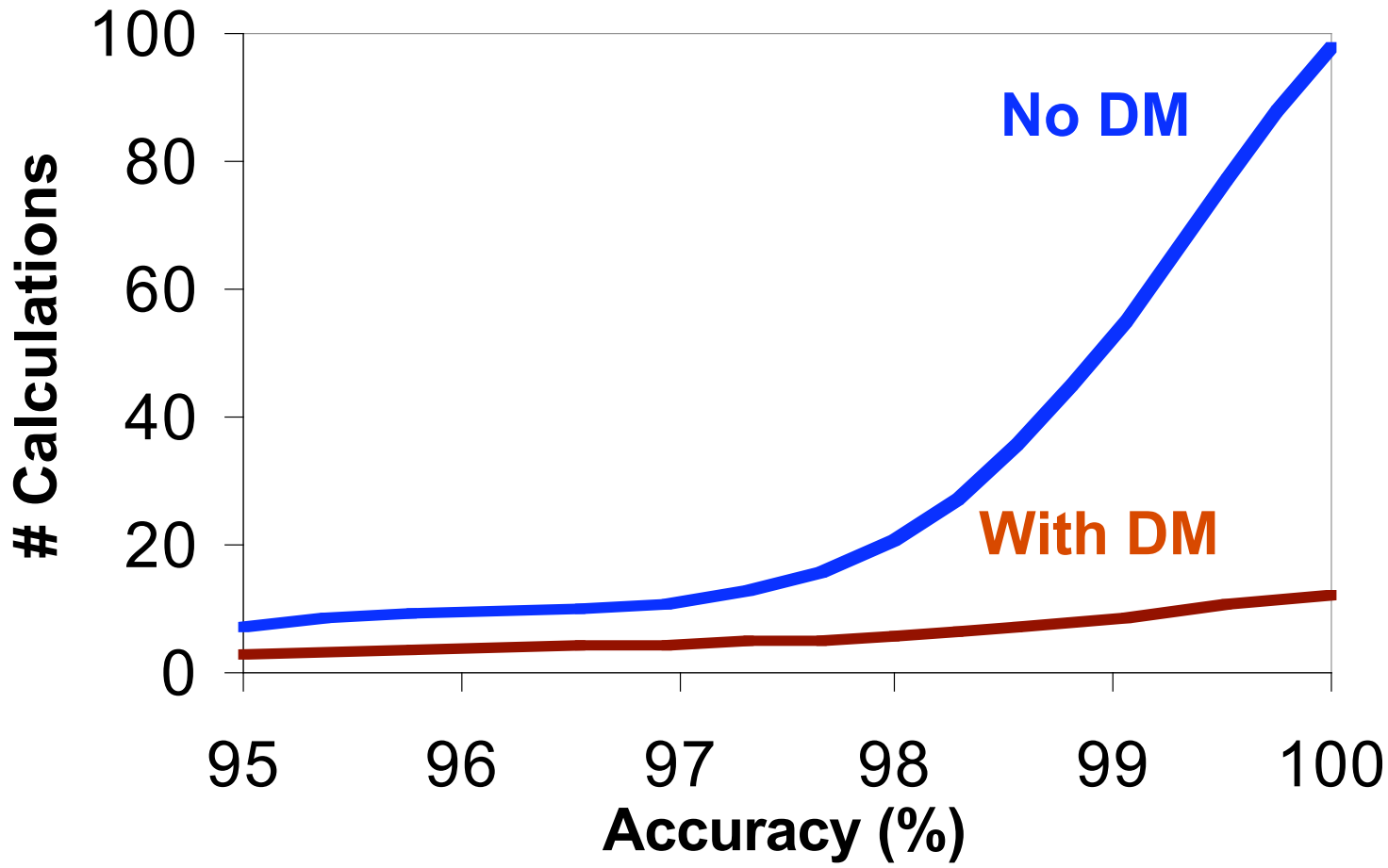






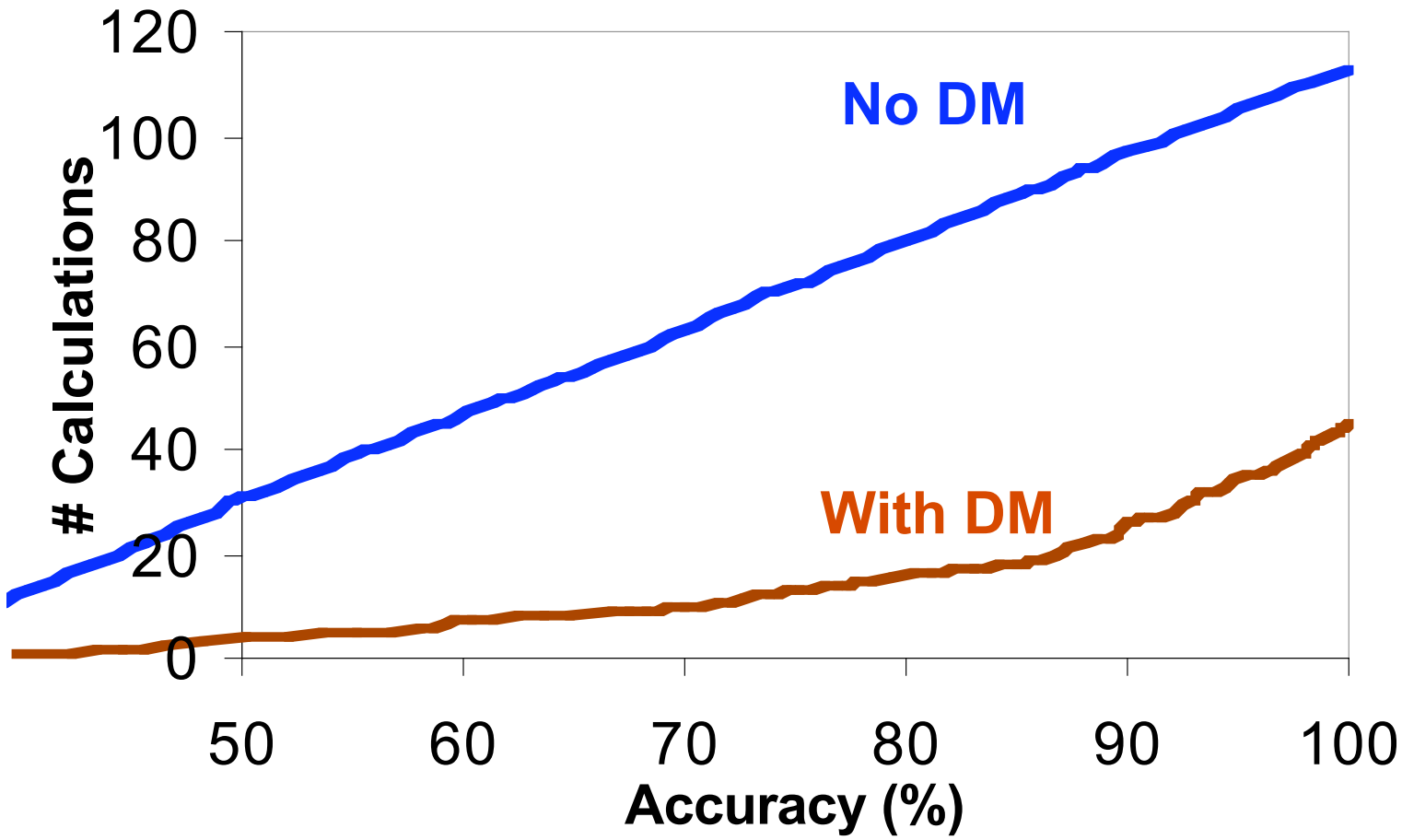


Compound Forming Vs. Phase Separating



~2-8x speedup from Data Mining

Ground State Prediction



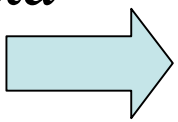
~4x speedup from Data Mining

Conclusions

- **High-throughput *ab initio*** approaches are a powerful tool for crystal structure prediction.
- **Data Mining of previous calculations** can create significant speedup when studying new systems.

Future work

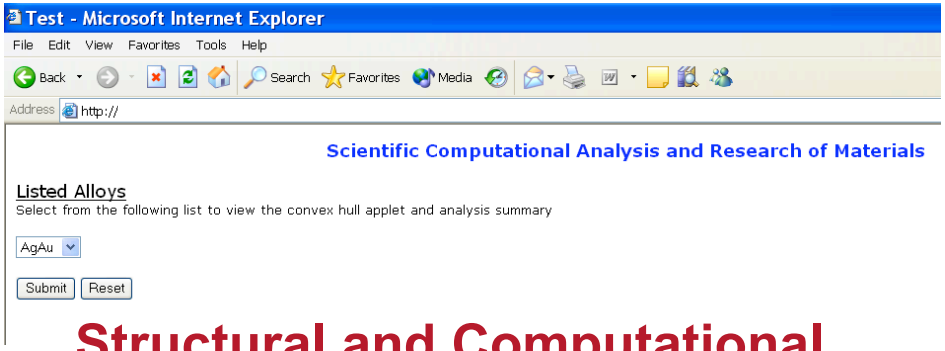
More experimental/computed data
More data mining tools
Web interface



Practical tool to predict crystal structure

Web Access to Database

Easy Interface

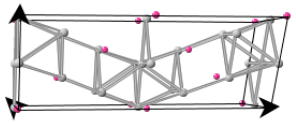


Structural and Computational Data, Visualization

System Name: AgAu

Structure ID	3
Pearson ID	tP4
Spacegroup Symbol	P4/mmm
Strukturbericht Symbol	L1_0
Prototype Name	AuCu
Spacegroup Name	#123
Descriptor File	
default description for str_id 3 AB along [001]	
INCAR	
SYSTEM = AgAu.3 ISPIND=2 ISPIN=2 PREC=med ISMEAR=1 SIGMA=0.2 IBRION=2 NSW=51 ISIF=3 ENMAX=270.903 NBAND=26 MAGMOM= 5 5	
KPOINTS	
KPOINTS file 0 Monkhorst-Pack 12 12 8 0 0 0	
POSCAR (initial)	Visualize
AgAu.3 -33.0807 0.000000000000 -0.500000000000 -0.500000000 0.000000000000 -0.500000000000 0.500000000 -1.000000000 0.000000000000 0.000000000 1 1 Direct 0.000000000000 0.000000000000 0.000000000000 Ag+5 0.500000000000 0.500000000000 0.500000000000 Au+5	
CONTCAR (final)	Visualize
AgAu.3 4.044631561899543 0.000000000000000000 -0.498176558240203 4 -0.498176558240203 4 0.000000000000000000 -0.498176558240203 4 0.498176558240203 4 -1.002414929106388 0 0.000000000000000 0.000000000000000 1 1 Direct 0.000000000000000 0.000000000000000 0.000000000000000 0.500000000000000 0.500000000000000 0.500000000000000	

Initial Poscar File Visualization for
System Name: AgAu
Structure ID: 239

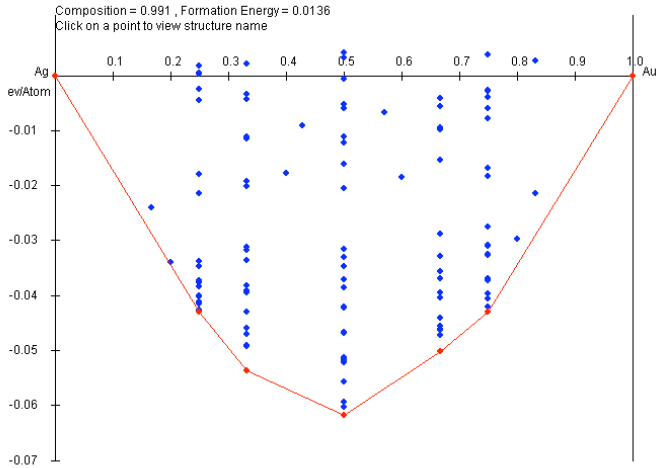


Please drag mouse to change orientation.
Right click on applet to choose visual settings

Analysis: Convex hull, Ground States

Analysis Summary

If you do not see the applet showing a convex hull, please download Java Runtime Environment



Ground States for AgAu

Description	Structure Number	Structure Name	Composition	Total Energy (ev/atom)	Formation Energy (ev/atom)	Energy above hull (ev/atom)	Cutoff Energy (ev/atom)	Pseudopotential	xc_approx	xc_func
Show	2	2	0.0	-3.743	0.0	0.0	270.9	US	LDA	CA
Show	286	286	0.2	-3.907	-0.034	0.0	270.9	US	LDA	CA
Show	279	279	0.25	-3.948	-0.043	0.0	270.9	US	LDA	CA
Show	26	26	0.25	-3.948	-0.043	0.0	270.9	US	LDA	CA

Collaborators/Acknowledgements

Collaborators

- Mohan Akula (*MIT*)
- Stefano Curtarolo (*Duke*)
- Chris Fischer (*MIT*)
- Kristin Persson (*MIT*)
- John Rodgers (*NRC Canada, Toth, Inc.*)
- Kevin Tibbetts (*MIT*)

Funding

National Science Foundation Information Technology Research (NSF-ITR) Grant No. DMR-0312537

END