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

- Binary alloys incomplete
- Multi-component systems largely unknown

Massalski, *Binary Alloy Phase Diagrams* '90
The Structure Prediction Problem

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

Present focus

Crystalline phases

\textit{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*

\[ \nu \]

Data Mining
Calculated/Experimental Databases
High-Throughput Ab Initio

Robust methods/codes
Automated tasks
Parallel computation

Log(# calculations)


Metals Database
~14,000 Energies

Curtarolo, et al., Submitted '04
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

- 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

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?
Principal Component Analysis identifies correlations

N structural energies from N/3 independent variables

RMS Error (eV/atom)

% Total Degrees of Freedom
Using Correlations for Structure Prediction

New alloy system: AB

Predict likely stable structure $i$ for alloy AB

Calculate $E_i$

Accurate convex hull?

Yes

Predicted crystal structure

No

Database calculated energies (AC, BC, etc.) + AB

Correlations
Data Mining Example: AgCd
Compounds Forming Vs. Phase Separating

![Graph showing the comparison between Compound Forming and Phase Separating with and without Data Mining. The graph indicates a significant speedup of ~2-8x 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

Analysis: Convex hull, Ground States

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

Scientific Computational Analysis and Research of Materials

Listed Alloys
Select from the following list to view the convex hull applet and analysis summary

- AgAu

Submit Reset

Structural and Computational Data, Visualization

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