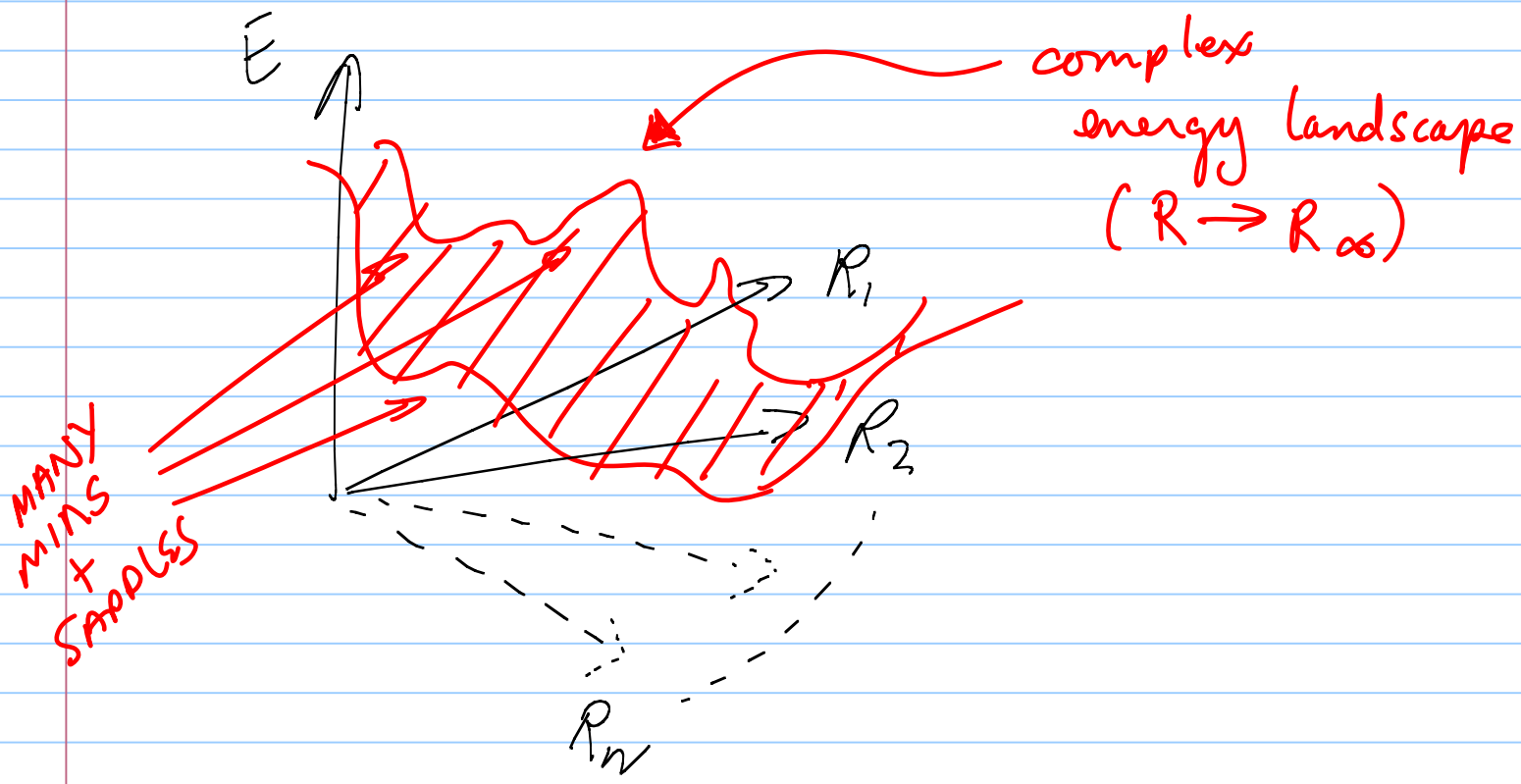
DATA MINING - DRIVEN QM PREDICTION OF CRYSTAL STRUCTURES

* Gerd Ceder (MIT)

- burgaz.mit.edu
- datamine.mit.edu

14,000 DFT calcⁿ

? CAN WE PREDICT ALL STRUCTURES ?



* CAN WE SEARCH THIS CONFIGURATIONAL SPACE TO FIND GROUND STATE STRUCTURES

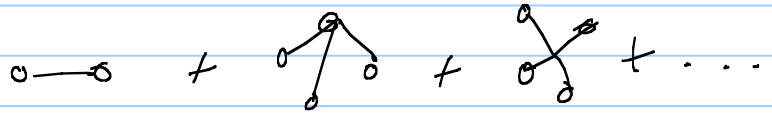
(consider $\emptyset$-K limit, extend to finite-T later)

APPROACHES

(1) MODEL $\hat{H}$: need correct description of energetics; fit data

(2) OPTIMIZATION PROBLEMS : optimize non-linear eqt^{n}'s w/ 3N positions for N atoms / unit cell.

~ Genetic Algorithms
~ Cluster Expansions } somewhat successful but tend to fail at saddle points



(3) INTUITION → works well, but we have limited data set in our heads

↓

2. can we find a more systematic approach to build a list of highly probable XII' structs

A: Yes!

- Databases of XU-structures exist (80,000) Pauling File / ICSD (100,000+)
- NATURE IS NOT RANDOM, these datasets have intrinsic correlation, which we build our chemical intuition on...

DATAMINING APPROACH

KNOWLEDGE

Computed data

XPTL data

$\{S_{XU}\}$

DATAMINING FIND CORRELATIONS

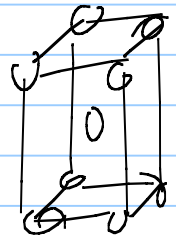
PREDICT

Synobable-XUs

find $|G.S\rangle$

$\hat{H}|\psi\rangle = E|\psi\rangle$

STRUCTURE SOLUTION



FOUNDATION

XII-structures have well-defined energy dependencies
(* but we are not parameterize E *)

c.f./ two ionic Lattices:

L_1 L_2

Energy L_1 v. L_2

depends on ionic system is.

D.M. captures these correlations purely mathematically
↳ w/o $E(\alpha, \beta \ldots)$ model.
↳ LESS BIAS

How do we do this...

consider: A-B alloy

look at Database of AB compounds
* ∃ some frequency of XII to appear
* ∃ " " " - pairs to appear

(due to underlying QM interactions { ionic, covalent, metallic)

define. correlation between two XU-structures

$$f(X_i, X_j) = \frac{p(X_i, X_j)}{p(X_i)\,p(X_j)}$$

$p(X_i, X_j)$ ← probability both XU-structures occur in some AB system at random

$f(X_i, X_j)$ represents XU-struct at composition i occurring in alloy

$p(X_i)\,p(X_j)$: probability of structure occurring at composition i

nature not random → $f > 1$ ~ similar stabilizing forces

$f \to 0$ ~ very different chemistries

eg/ STRONG CORRELATION B/W

Fe_3C structure at AB_3 composition

(WITH)

$MgCu_2$ structure at A_2B

52/87 alloys structures occur together when Fe_3C is found

$\Rightarrow f \sim 8.50$

→ for Fe_3C present at AB_3 composition, it is ~ 8.5 X more likely $MgCu_2$ structures will form A_2B compositions.

(size-effect stabilized structures).

* this information can be expressed mathematically with a "mutual information parameter", which in principle is sufficient to predict the structure of an unknown composition from known compositions

$$\underline{\underline{I}}_{i,j} = \sum_{\underline{\chi}_i, \underline{\chi}_j} p(x_i, x_j) \ln\left[f\langle i,j\rangle\right]$$

$\forall$ struct combinations

* define. probability function

$$P\{\underline{\underline{\chi}} = (A, B, \ldots; x_i, x_j \ldots)\}$$

- fills space of all structures and compositions

structures found by evaluating a conditional probability:

$$\mathbb{P}\{\underline{\chi} \mid \underline{e}\}$$

$\underline{e}$ ← available info about system

* need form of $\mathbb{P}$

cluster-like expansion (MORITA)

$$\mathbb{P}\{\chi\} = \prod_i p(x_i) \prod_{jck} g^2(x_j, x_k) \prod_{lcmcn} g^3(x_l, x_m, x_n)$$

cumulant fn^cts

$p(x_i)$: independent variations

$g^2(x_j, x_k)$: pairwise correlations

$g^3(x_l, x_m, x_n)$: triplet correlations

* in practice, truncate to $O(2)$

and $g^2 \rightarrow f(x_i, x_j)$

* anti-correlations are just as important.

---

* to setup $p(x_i)$ implement Bayes' rule + Dirichlet dist. (cf. Nat. Mater 5 641 (2006)

EXAMPLE

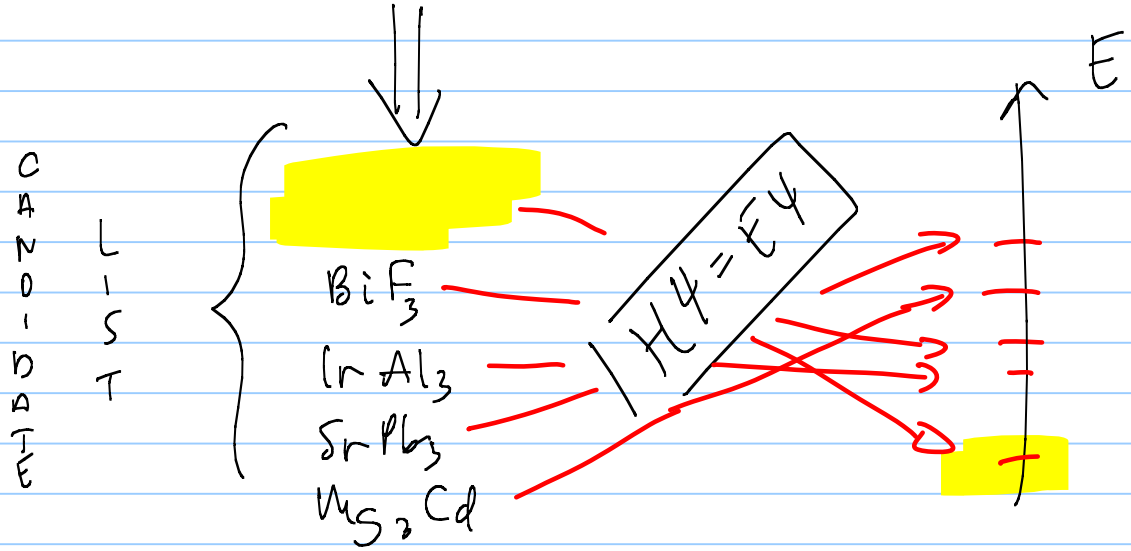
Ag-Mg alloy at ~75% Mg ($AgMg_3$)

= The structure is not know = → somewhat stable

Setup conditional $\mathbb{P}$

$$\underline{e} = \begin{cases} X_A = Ag, \; X_0 = fcc, \\ X_{0.25} = ZrAl_3 \\ X_{0.50} = CsCl \\ X_{1.0} \; hcp, \; X_B = Mg \end{cases}$$

evaluate $\mathbb{P}\{\underline{\chi}_{0.75} \mid \underline{e}\}$



MOST STABLE STRUCTURE IS $Cu_{2.82}P$ (24 atom/unc)

→ not intuitive

$Cu_{2.82}P$

SG 185 : $P6_3cm$

FORD, G.W. "INVESTIGATION OF THE EFFECTS OF IMPURITIES IN COPPER"
J. of Inst. of Metals, London 43 41 (1930)