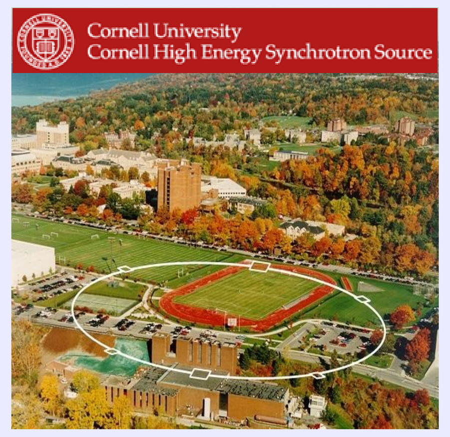
Human Computation for Combinatorial Materials Discovery

Ronan Le Bras   Richard Bernstein   Carla P. Gomes   Bart Selman   R. Bruce van Dover

Motivation



[Goals]

- Discover **new materials** with improved catalytic activity for **fuel cell** applications.
- **Reduce** the processing **time** of the **data analysis** to dynamically optimize expensive Synchrotron experiments.

[Characteristic of Combinatorial Materials Discovery]

- Complex local **physical constraints**, that require a **sophisticated optimization approach**
- Interpretation of complex high-intensity X-ray **diffraction data**, that appears to be well-suited for a **human computation** approach

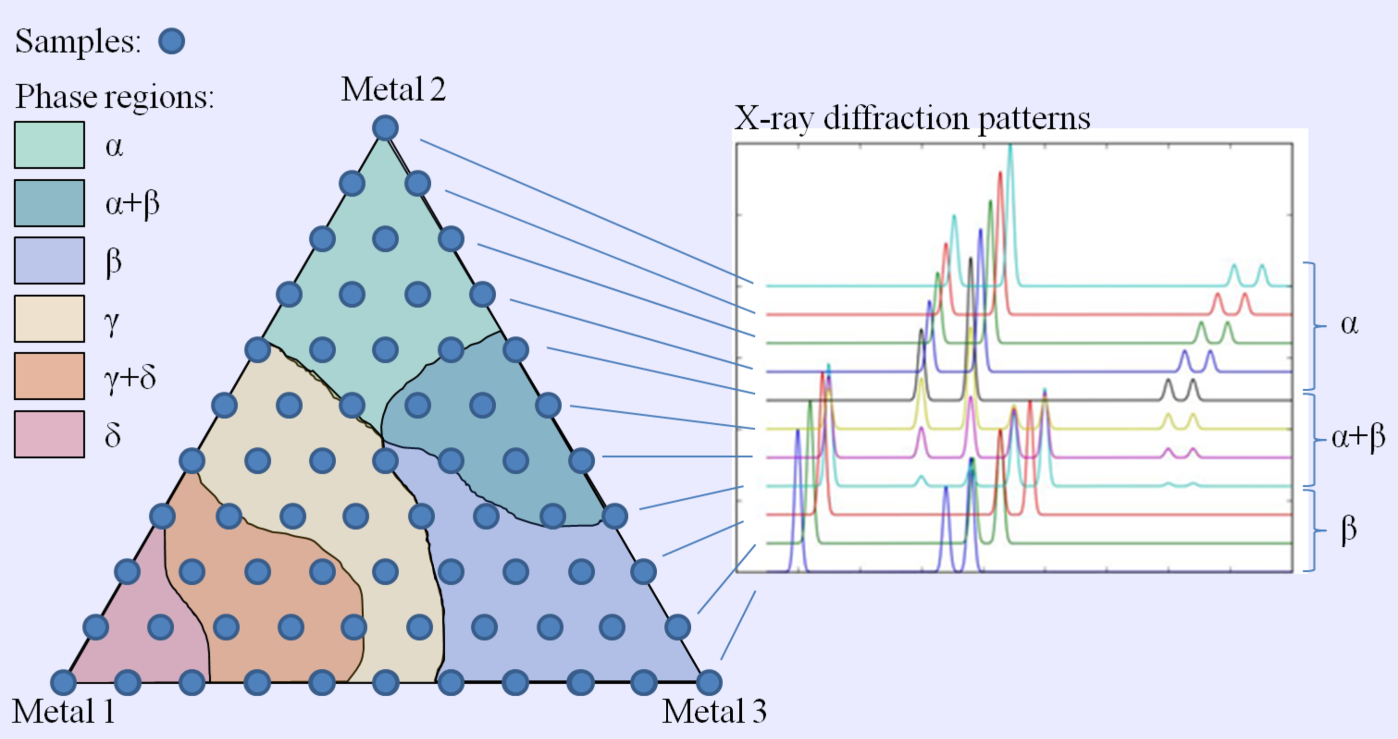
[Research Question]

- Can **human input** significantly **boost** the performance of combinatorial **reasoning** and **optimization** methods?

Phase-Map Identification Problem

Combinatorial Method: sputtering 3 metals (or oxides) onto a silicon wafer (which produces a *thin-film*) and using x-ray diffraction to obtain structural information about crystal lattice.

Input: Diffraction patterns $D_1,\ldots, D_n$ of n sample points on the thin-film.



Output: Set of K basis patterns (or *phases*) $B_1,\ldots,B_K$ (along with weights a_{ij} and shifts s_{ij} of basis pattern j in point i).

Physical Characteristics

Each diffraction point D_i is explained by the basis patterns:

$$D_i = a_{i1}B_1 + \ldots + a_{iK}B_K$$

There is experimental noise:

$$\min \|D_i - a_{i1}B_1 + \ldots + a_{iK}B_K\|$$

Non-negative basis patterns and coefficients:

$$B_i \geq 0\ ,\ a_{ij} \geq 0$$

At most M non-zero coefficients per point:

$$|\{j \mid a_{ij} > 0\}| \leq M$$

Basis patterns appear in contiguous locations on the silicon wafer:

The subgraph induced by $|\{i \mid a_{ij} > 0\}|$ *is connected*

Basis patterns can be shifted: (Shift operator; Shift coefficients)

$$\|D_i - a_{i1}S(B_1,s_{i1}) + \ldots + a_{iK}S(B_K,s_{iK})\|$$

Shifts coefficients are bounded, continuous and monotonic:

$$s_{11} \leq s_{12} \leq s_{13} \leq s_{14}$$

$$|s_{12} - s_{11}| \leq c$$

Satisfiability-Modulo-Theories Formulation

Integer variables e_{ij} for the **peak locations** in each B_i

Integer variables for the shift coefficients s_{ij}

An observed peak p is "*explained*" if there exists s_{ij}, e_{il} s.t. $|p-(s_{ij} + e_{il})| \leq \varepsilon$

Every observed peak must be "*explained*"

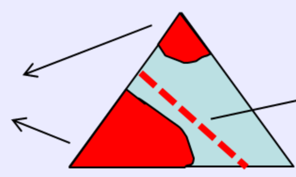
Bound the number of missing peaks $\leq T$

Minimization by (binary) search on T

Linear phase usage constraint (up to M basis patterns per point)

Linear constraint for shift monotonicity and continuity ($s_{ij} \leq s_{lm}$)

Lazy connectivity: add a cut if current solution is not connected

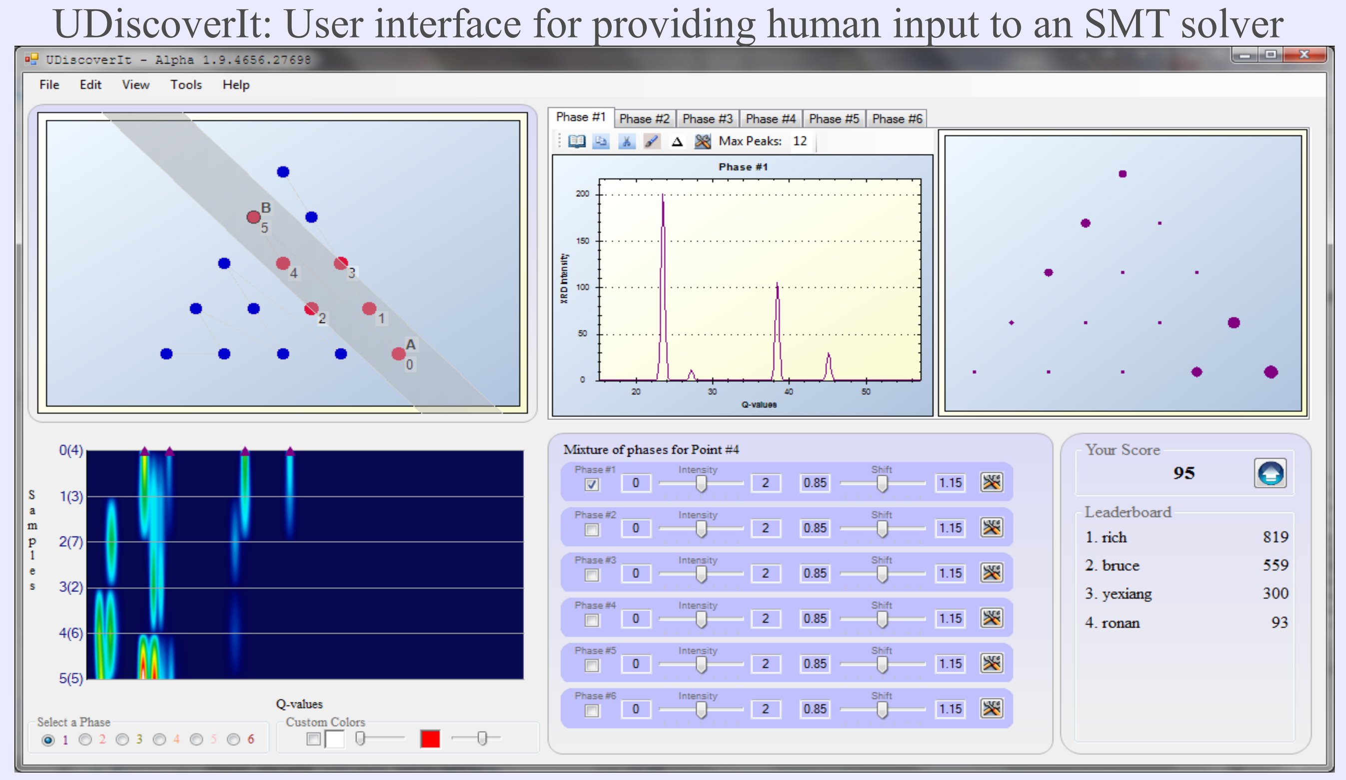


Symmetry breaking: Renaming of pure phases, ordering the peaks location e_{ij} (per phase)

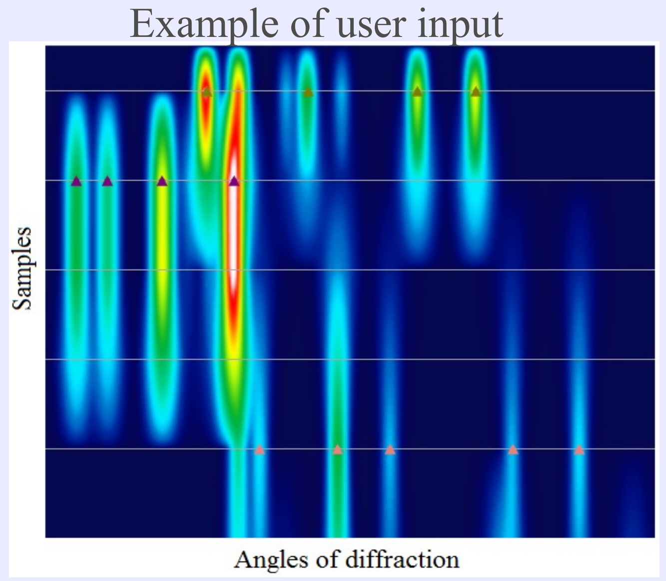
⇨ *{quantifier-free linear integer arithmetic theory}*

Experimental Results

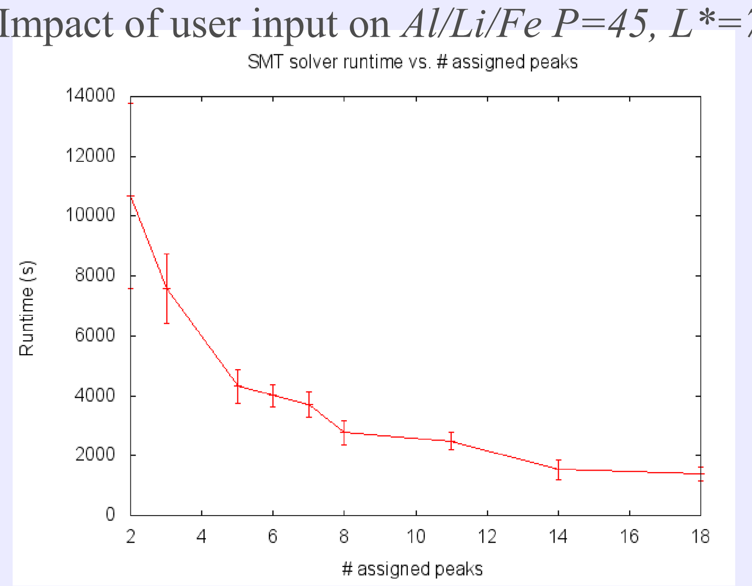
UDiscoverIt: User interface for providing human input to an SMT solver



Example of user input



Impact of user input on *Al/Li/Fe P=45, L*=7*



	Dataset					Time w/o	Time w/	# assigned
System	*P*	*L**	*K*	*#var*	*#cst*	user input *(s)*	user input *(s)*	var. by user
A/B/C	36	8	4	408	2095	3502	**150**	19 (4.6%)
A/B/C	60	8	4	624	3369	17345	**261**	18 (2.9%)
Al/Li/Fe	15	6	6	267	1009	79	**27**	6 (2.2%)
Al/Li/Fe	28	6	6	436	1864	346	**83**	12 (2.7%)
Al/Li/Fe	28	8	6	490	2131	10076	**435**	26 (5.3%)
Al/Li/Fe	28	10	6	526	2309	28170	**188**	23 (4.3%)
Al/Li/Fe	45	7	6	693	3281	18882	**105**	28 (4.0%)
Al/Li/Fe	45	8	6	711	3410	46816	**74**	30 (4.2%)

- **Human computation** can dramatically **speed up** the performance of combinatorial **optimization methods**
- Our approach leverages the **complementary strength** of **human input**, providing **global insights** into problem structure, and the power of combinatorial **solvers** to exploit complex **local constraints.**

Acknowledgments

The authors gratefully acknowledge the support of the **National Science Foundation**, award number 0832782.