Materials Discovery: New Opportunities at the Intersection of Constraint Reasoning and Learning*

Ronan Le Bras[1] Stefano Ermon[1] Theodoros Damoulas[1] Rich Bernstein[1]
Carla Gomes[1] Bart Selman[1] R. Bruce van Dover[2]

[1] Department of Computer Science, [2] Department of Materials Science and Engr.,
Cornell University, Ithaca NY 14853, USA

Combinatorial materials science involves the rapid, high-throughput synthesis, measurement, and analysis of a large number of different but structurally related materials. In combinatorial materials discovery, materials scientists search for intermetallic compounds with desirable physical properties by obtaining measurements on hundreds of samples from a composition spread. Determining the structure of the materials formed in a composition spread is key to understanding composition and property relations and can potentially result in a breakthrough materials discovery.

This is an important and exciting direction in the emerging field of *computational sustainability* [4] as it aims to achieve the best possible use of our available material resources. One ultimate objective is to help discover the next-generation materials for fuel-cell catalysis, as such materials have the potential of dramatically increasing fuel cell capacity while reducing their cost.

The analysis of composition spreads remains, however, a manual and laborious task. Thus the need for new techniques to automatically analyze and interpret such data. Whereas the data-intensive aspect of the area of materials discovery seems to favor Data-Mining or Machine Learning techniques, the rigorous and highly-structured physical properties that govern the crystallization on the composition spread interestingly suggest that constraint reasoning is key to a physically meaningful analysis. In this paper, we describe two novel approaches to this problem that integrate domain-specific scientific background knowledge about the physical and chemical properties of the materials. Our first approach combines constraint programming (CP) and machine learning (ML), while the second is based on *satisfiability modulo theory* (SMT).

We evaluate the performance of our methods on realistic synthetic measurements, and we show that it provides accurate and physically meaningful interpretations of the data, even in the presence of artificially added noise.

Combinatorial Materials Discovery

In the composition spread approach, three metals (or oxides) are sputtered onto a silicon wafer using guns pointed at three distinct locations, resulting in a so-called *thin film* (see Figure 1). Different locations on the silicon wafer correspond to different concentrations of the sputtered materials, depending on their distance from the gunpoints. During experimentation, a number of locations (samples) on the thin film are examined using an x-ray diffraction technique that provides, for each sample, a diffraction pattern (i.e. the intensity of the electromagnetic waves as a function of the scattering angle of diffraction).

The observed diffraction pattern characterizes the underlying crystal structure, and provides important insights about the chemical and physical properties of the corresponding composite material. As illustrated in Figure 1 (Right), shifting and merging diffraction patterns are the result of the crystallization process.

Kernel-Based Clustering to Boost CP Scalability

This crystallization process can be formulated as a CP encoding. However, this formulation does not scale up to instance sizes of interest, especially in the presence of noise. To improve the scalability of the CP approach, we propose in [6] to leverage ideas from machine learning (ML), specifically kernel-based similarity

*Supported by NSF Expeditions in Computing award on Computational Sustainability (Grant 0832782).

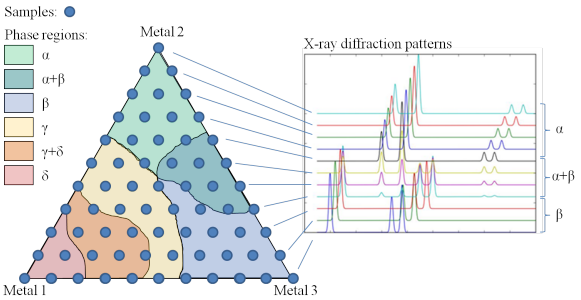

Figure 1: Left: Pictorial depiction of the problem, showing a set of sampled points on a *thin film*. Each sample corresponds to a different composition, and has an associated measured x-ray diffraction pattern. Colors correspond to different combinations of the basis patterns $\alpha, \beta, \gamma, \delta$. Right: Scaling (shifting) and merging of the diffraction patterns as one moves from one point to a neighboring one.

measures and clustering, in order to make the problem solving task easier for the CP formulation. This novel integration of the two approaches is inspired by their complementary strengths: while CP techniques are particularly suited to enforce detailed constraints at a local level, data-driven ML methods are more robust to noise and effective at identifying global patterns of similarity.

SMT for Materials Discovery

We integrate domain-specific knowledge about the physical properties of the materials into an SMT reasoning framework based on linear arithmetic [3]. The problem has a hybrid nature, as it combines continuous measurement data, discrete decision variables and combinatorial constraints, which is particularly suited for an SMT reasoning. Using our novel encoding, state-of-the-art SMT solvers can automatically analyze large synthetic datasets, and generate interpretations that are physically meaningful and very accurate, even in the presence of artificially added noise. Moreover, our approach scales to realistic-sized problem instances. Furthermore, we show that SMT solving outperforms both Constraint Programming and Mixed Integer Programming translations of our SMT formulation. This suggests that the improvements come from the SMT solving procedure rather than from the new arithmetic-based encoding, opening a novel application area for SMT solving technology beyond the traditional verification domains [2].

Empirical Validation

To validate our approaches, synthetic x-ray diffraction data was generated using diffraction patterns from the JCPDS database [1] with parameter reflecting those of a recently developed combinatorial crystallography technique [5]. These experiments show that, as opposed to previous work, our algorithms are able to explain the observed diffraction patterns with high accuracy and in a physically meaningful way.

References

[1] *Powder Diffract. File, JCPDS Internat. Centre Diffract. Data, PA*, 2004.

[2] A. Biere. Sat, smt and applications. *Logic Programming and Nonmonotonic Reasoning*, pp. 1–1, 2009.

[3] S. Ermon, R. Le Bras, C. P. Gomes, B. Selman, and R. B. van Dover. Smt-aided combinatorial materials discovery. In *Proceedings of the Conference on Theory and Applications of Satisfiability Testing (to appear)*, SAT'12, 2012.

[4] C. P. Gomes. Computational Sustainability: Computational methods for a sustainable environment, economy, and society. *The Bridge, National Academy of Engineering*, 39(4), Winter 2009.

[5] J. M. Gregoire, D. Dale, A. Kazimirov, F. J. DiSalvo, and R. B. van Dover. High energy x-ray diffraction/x-ray fluorescence spectroscopy for high-throughput analysis of composition spread thin films. *Rev. Sci. Instrum.*, 80 (12):123905, 2009.

[6] R. Le Bras, T. Damoulas, J. M. Gregoire, A. Sabharwal, C. P. Gomes, and R. B. van Dover. Constraint reasoning and kernel clustering for pattern decomposition with scaling. In *Proceedings of the 17th international conference on Principles and practice of constraint programming*, CP'11, pp. 508–522, 2011.