High-throughput Materials Discovery through Materials Genomics: Machine-learning discoveries in solid state chemistry

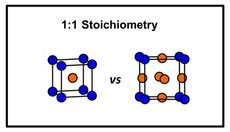
A.O. Oliynyk <oliynyk@ualberta.ca>, L.A. Adutwum, H. Pisavadia, M. Macklin, A. Mar <amar@ualberta.ca>

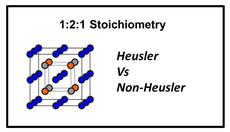
BACKGROUND

The ***Materials Genome Initiative*** has an ambitious goal to discover and develop new materials at greatly accelerated rates at a much lower cost. One way of achieving this is by taking advantage of breakthroughs in modeling and data-mining methods for the analysis of "big data," and by integrating theory (design) and experiment (synthesis, property measurement) more closely. Rational materials discovery is extremely difficult, with most materials uncovered and optimized via empirical methods.

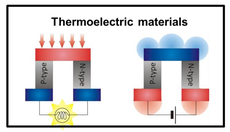
What we already can do with machine learning?

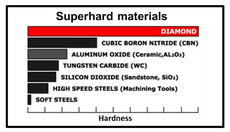
Classify Crystal Structures





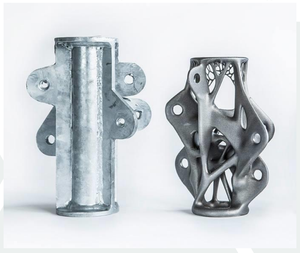
Predict Physical Properties





SHORT-TERM OBJECTIVES

We will implement machine-learning methods to guide researchers in the solar research group to "think outside the box." Subject to optimized parameters (e.g., improved performance, earth-abundant materials, low cost), we will identify new candidates for solar cell absorber and materials to provide leads for the Buriak and Shankar groups for photovoltaics, and catalyst for solar fuels to provide leads for the Bergens group. As long-term objectives, experimental data will be fed back to the growing database of materials structures and properties to improve the quality of predictions and eventually lead to the design of better solar materials.



http://www.arup.com/news/2015_05_may/11_may_3d_makeover_for_hyper-efficient_metalwork

The same structural loads and forces, 75% weight reduction

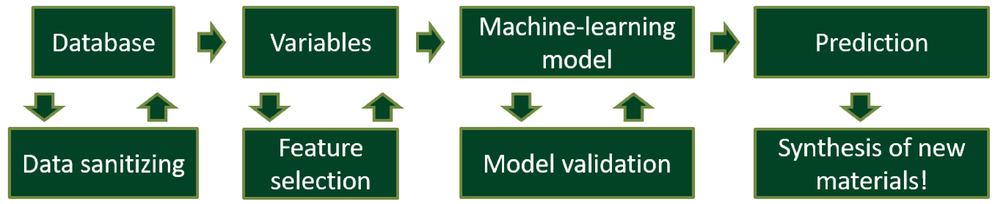
Synthesis condition optimization is one way to apply machine learning techniques. It allows researchers increase the success rate in experimental work, as well as, find an optimal way to make known materials. In our case, we will focus on photovoltaic and catalysis applications.

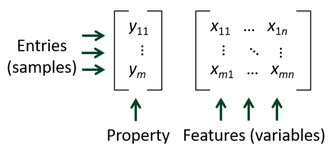
PROJECT OVERVIEW

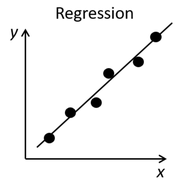
The **Mar Research Group** leads the effort within UAlberta in implementing machine-learning tools using scientific data that are publically available, and through feedback provided by other members of the solar group (Buriak, Shankar, Bergens). Also, the Mar Research Group provides expertise in solid state synthesis and characterization. The **Buriak and Shankar Research Groups** are experts in semiconductor nanoparticles, specifically designed for photovoltaic applications; they build prototype devices and characterize the optical and electrical properties. The **Bergens Research Group** has expertise in photocatalysis of water splitting and investigating catalytic behavior. Together, the team benefits from complementary expertise and the resulting synergies will prove fruitful in accelerating materials research in alternative energies.

Our projects will include:

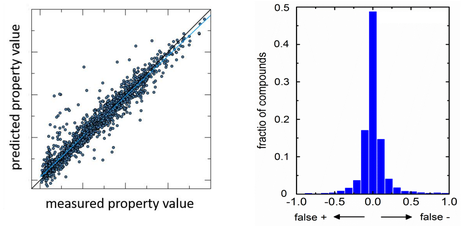
- Optimization of synthesis conditions for existing solar cell systems
- Proposing novel solar cell candidates via machine learning prediction
- Improving catalytic activity of the materials for photocatalysis
- Machine-learning design of novel solid-state intermetallic semiconductor materials
- Fundamental classification and rationalization of existing database of inorganic materials







Example of model statistics



THEME OVERVIEW

Solar

The sun powers the entire world, providing warmth, light, and sustenance for countless forms of life. Technologies have made it possible to use some of the sun's energy to produce electricity and fuels, but new refinements may allow us to diversify the ways in which solar energy can be generated, stored, and utilized. By identifying lower-cost materials for use in the construction of solar cells, finding new catalysts to enable different types of production, identifying more efficient methods for market integration, and considering the possibility of solar-derived hydrogen fuels, it may be possible to develop vast energy resources from the most abundant source in our lives.

EXPECTED OUTCOMES

We anticipate building a platform in which the other partners in the solar group will be able to specify the desired requirements of a "good" material (e.g. limits on band gap size, restriction to certain earth-abundant starting materials, or avoidance of strategically critical elements). Subsequently, a list of proposed new materials will be generated with quantifiable probabilities indicating their likelihood to exhibit those properties.

Method	Example	How fast is this discovery?	How expensive is this discovery?
Serendipitous discoveries	Trial and error	Fast	Expensive
Scientific intuition (traditional approach)	Data-driven discoveries	Slow	Inexpensive
First principles ab initio calculations	DFT computations	Slow	Expensive
Machine-learning prediction	Recommendation engines	Fast	Inexpensive

Imagine targeting new material as search for a needle in a haystack

Here is how we would picture the problem

http://blog.a2zinc.net/2012/04/20/finding-the-needle-in-the-haystack/

Here is how machine learning would look like

https://hackingmaterials.com/2013/11/11/why-hack-materials/

EXTERNAL PARTNERS

Citrine Informatics (Redwood City, California)
Help us with advanced informatics tools

Sparks Research Group (Salt Lake City, Utah)
Help us with physical property measurements

Michael Gaultois (Cambridge, UK)
Help us with materials characterization

Department of Chemistry, University of Alberta, Edmonton, AB Canada T6G 2G2

UNIVERSITY OF ALBERTA
FUTURE ENERGY SYSTEMS

CANADA FIRST RESEARCH EXCELLENCE FUND

This research has been undertaken thanks in part to funding from the Canada First Research Excellence Fund.