

APS Scientific Computation Seminar Series

Speaker: Apurva Mehta
Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory

Title: Data-Driven Approaches to Accelerated Discovery of Complex and Metastable Materials

Date: Monday, November 2, 2020

Time: 2:00 p.m. (Central Time)

Location: <https://bluejeans.com/934501005>

Hosts: Nicholas Schwarz and Mathew Cherukara

Abstract:

The complex material systems are beginning to play a growing role in addressing some of our current technological challenges. They include multi-principal element alloys and oxides (often also called high entropy alloys/oxides), metastable materials like metallic glasses and carbon nanotubes, solid-state energy storage devices, heterogeneous catalysts, and increasingly complex devices for power electronics and quantum computing. However, there are easily tens of million and possibly even a billion compositions out of which a very small fraction are potential candidates. Many of the desired materials are metastable and therefore processing paths play a crucial role in synthesizing them. The composition-processing combinatorial search space is too vast for a dependence on serendipitous discoveries, to provide urgently need novel materials and devices; and a brute-force survey of the combinatorial space is too slow and expensive.

We need guidance to navigate the vast composition-processing space. Our physiochemical understanding of complex material systems is still too nascent to provide it. Recent advances in machine-learning (ML) and other AI-tools, however, suggests that a data-driven approach that builds up on insights from physiochemical theories, and bridges the gaps in the insights by experimental observations can provide the needed navigation. Application of ML to large experimental observations can highlight hidden and complex trends that theories may miss and help direct us to the next high impact experiment to perform, leading to accelerated discoveries and new physiochemical insights. Iteration of the ML-guided cycle, starting from predictions, to smart experiments, to new discoveries and insights, resulting in next generation of predictions, is an approach that would not only significantly accelerate the pace of discoveries but will also change the way we do science.

The data-driven needs a large amount of data, which fortunately, the last 30 years of public investment in national scientific user facilities are primed to produce. The rate and complexity of measurements at

these facilities has been increasing exponentially over the last two decades. However, for the data-driven discoveries we do not need raw measurements, we need new scientific information (knowledge) contained in them. Unfortunately, the rate of new knowledge lags the rate of measurement because currently it is extracted by humans, who are struggling to keep up as the pace and complexity of measurements is accelerating. We need new data-tools to produce knowledge from measurements in real-time. We also need tools that utilize the newly produced knowledge to make experimentation smarter. Operation of the data-driven discovery cycle also needs a supporting cyber-infrastructure that provides on-demand computation and seamlessly movement data where and when needed.

Here, I will use the search for new Multi-Principal-Element Alloys, and specifically wear-resistant metallic glasses, to illustrate the data-driven discovery paradigm. I will also use it to highlight the analytical and cyber-infrastructure challenges that must be addressed for this paradigm to produce the urgently needed new materials and technologies.