Defense Materials Manufacturing and its Infrastructure (DMMI)

Workshop on:
Data analytics and what it means to the Materials Community

Dane Morgan

(*University of Wisconsin – Madison, WI USA*)

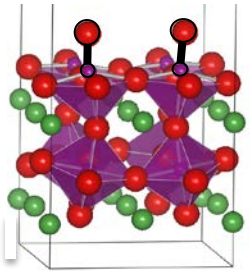
THE UNIVERSITY of WISCONSIN MADISON

National Academy of Sciences building
Room NAS120 at 2101 Constitution Avenue, NW.
July 16, 2019

Computational Materials Group

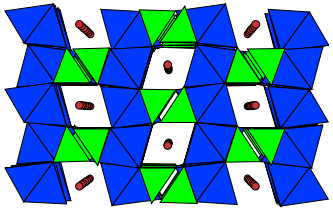
SOFC catalysts

(*oxygen catalysis, ABO_3 perovskites*)



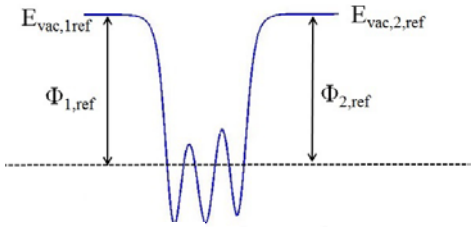
Battery Electrodes

(*ab initio cathode/anode design*)



Electron Emitter Cathodes

(*low work function*)



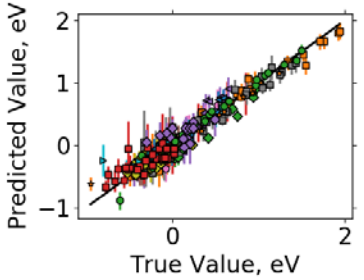
Nuclear Materials

(*radiation effects, fuel and cladding properties)*

Ab initio foundation
Thermokinetics/Data Science
Understanding/Prediction

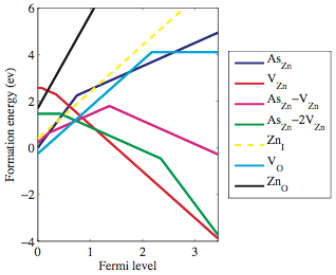
Machine Learning

(*Property prediction, Image analysis*)



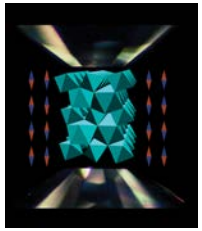
Doped Semiconductors

(*Doped ZnO*, GaAs)



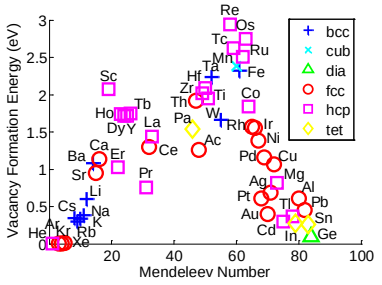
High-Pressure Materials

(*Lower mantle phases and spin*)



Modeling Tools

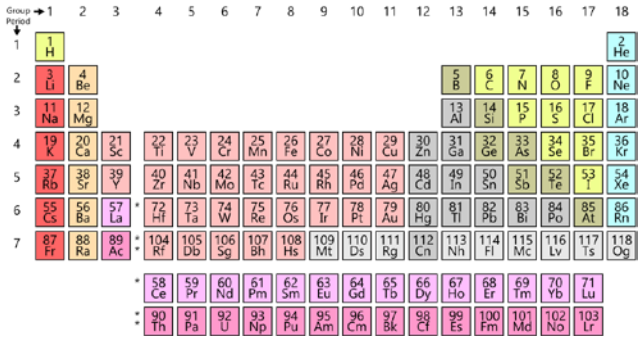
(*High-throughput, machine learning*)



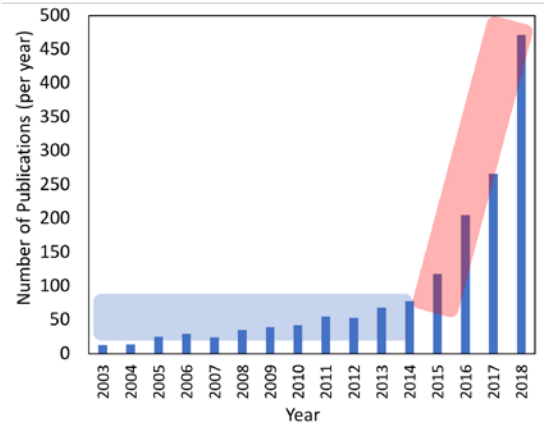
Present Machine Learning Uses in MS&E

- Using existing data to make predictions about new data, e.g., to guide future experiments or increase database size.
- Replacing physical models with ”correlative” data-centric models (e.g. machine learned catalytic activity vs. composition/processing).
- Interpreting and extracting data from images and text.

Not new, but now aided by new data, computation, and algorithms



Rapid growth in machine learning related materials research



Opportunities/Challenges for Machine Learning in MS&E

- Developing model dissemination infrastructure (living part of the materials innovation infrastructure - DOI, discoverable, persistent, reproducible, reusable, extensible, API, “FAIR” tools)
- Continuing software/data/infrastructure support (in MGI spirit) to enable machine learning in MS&E, e.g., automated data extraction and model development tools.
- Supporting engagement with data scientists to import best practices and methods.
- Integrating with robotics for closed loop design at speed of machines.
- Making more data available to solve open problems, e.g., standardized formats in papers, research paper libraries for text mining. ~90% of ML problems is typically database development.
- Exploring the “hard” problems (innovative design/discovery) vs. “easy” problems (more, faster), e.g. find halide perovskites for solar photovoltaics vs. find materials with 1.4 eV band gap. New methods (e.g., generative adversarial networks (GANs) and related for innovating new compounds) and new design environments (e.g., AI driven project dashboards) are likely enabling.