Citrine Informatics

The data analytics platform for the physical world

Analyzing Large-Scale Data to Solve Applied Problems in Materials R&D

Bryce Meredig, Citrine Informatics

Artificial Intelligence Applied to Materials Discovery and Design

9 August 2017

"Centaur" Materials Science

Human <u>and</u> Machine > Human <u>or</u> Machine

Things Scientists Do With Data: Conventional View of Informatics

Machine Learning Predicts Properties

Collect | Process | Visualize | Search | Model | Optimize | Interpret & Learn | Curate & Evaluate

Things Scientists Do With Data: Citrine's Vision

Machine Learning Will Be a Behind-the-Scenes Copilot for All of These

Collect | Process | Visualize | Search | Model | Optimize | Interpret & Learn | Curate & Evaluate

Covered in today's talk

Citrine also working in these areas—ask me

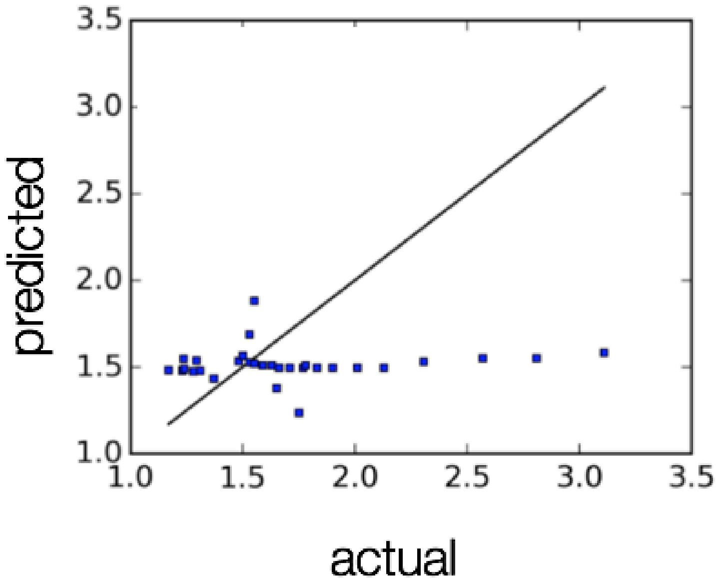
Citrine Has a Unique Business Model

- Open Citrination’s data & infra available to everyone
  - Open data (one of world’s largest collections of free materials data)
  - Open PIF data standard
  - Open-source lolo machine learning library
  - Open-source API tools and tutorials
- Enterprise Citrination users pay to leverage Open data while keeping internal data proprietary

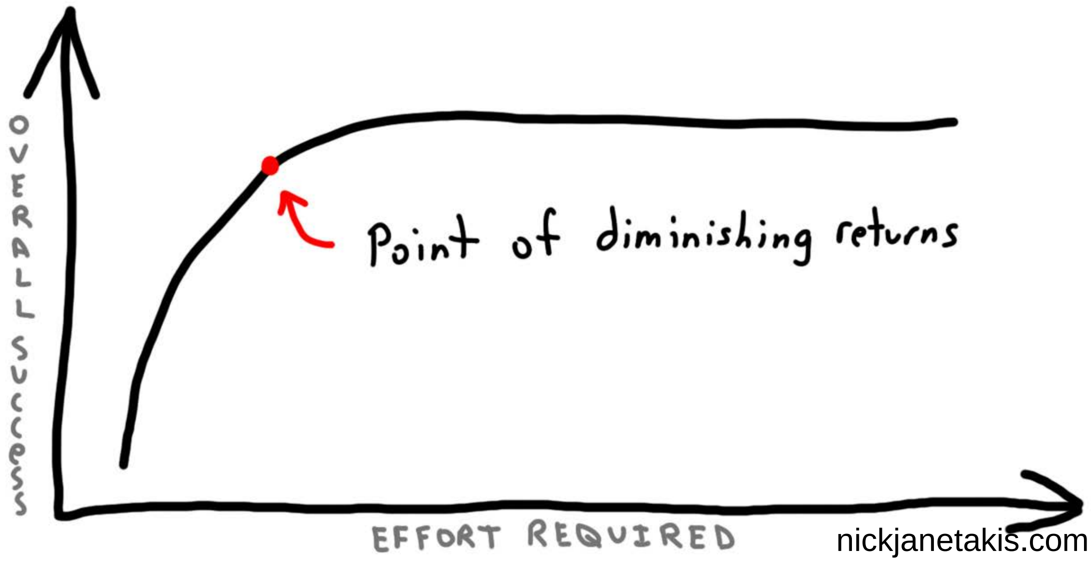
Industrial Applications & Research

Industrial Case Study #1

Isolated dataset too small
for machine learning



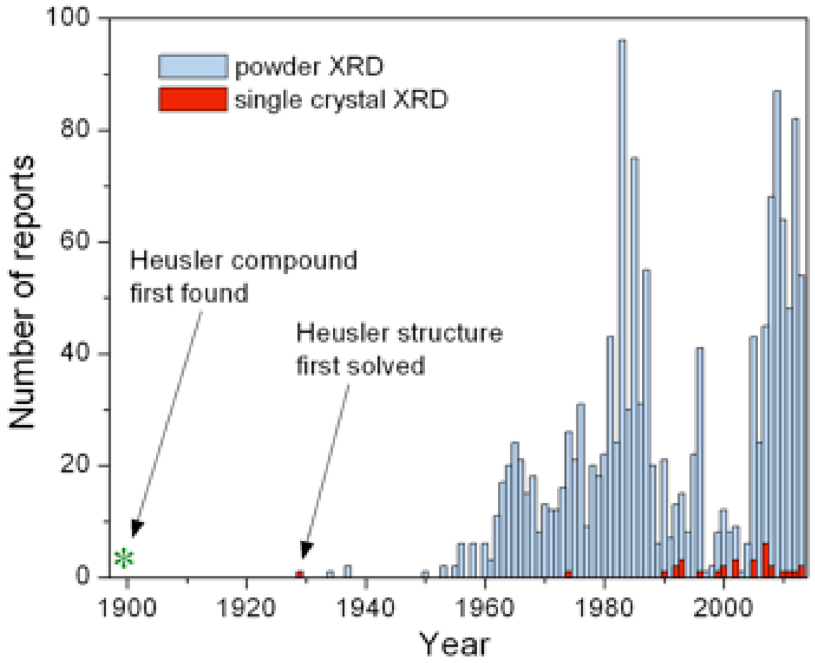
Industrial Case Study #2

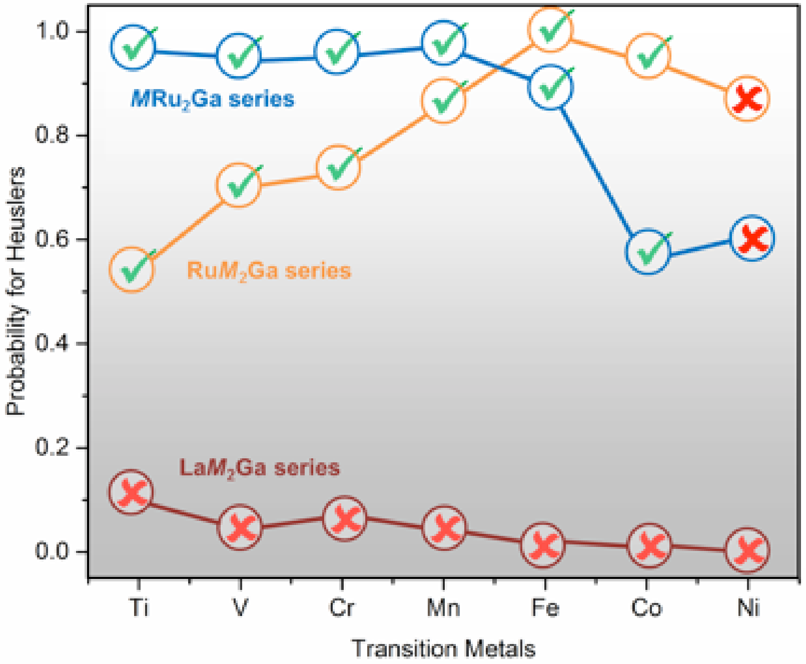


Machine learning helped save a customer >$30m on materials characterization by identifying the point of diminishing returns

Citrine Informatics

Data-Driven Methods Give >10x Yield Boost



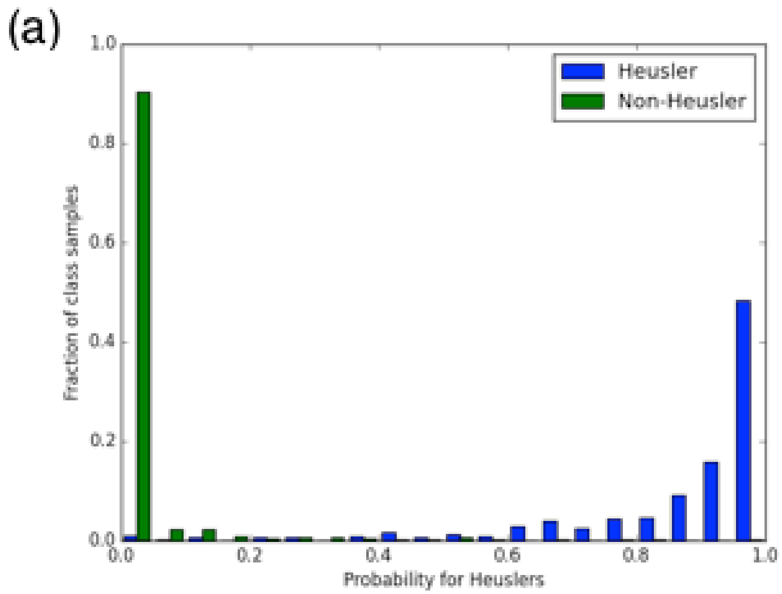


Scientific community discovers ~50 Heuslers/year

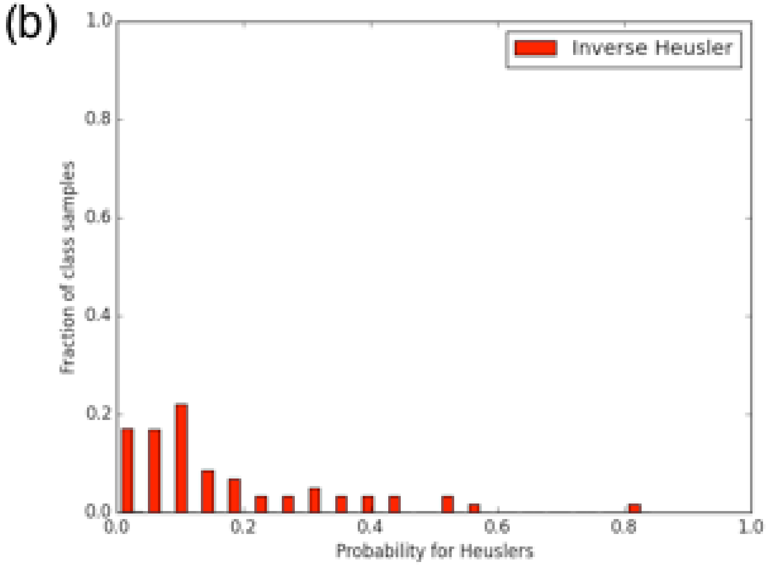
Our collaborators made 12 in one paper

Oliynyk et al., *Chem. Mater.*, 2016, **28** (20), pp 7324–7331

Models Discriminate Similar Structures



Heusler vs other

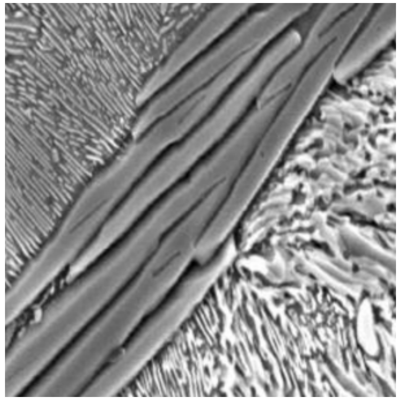


Heusler vs inverse

Oliynyk et al., *Chem. Mater.*, 2016, **28** (20), pp 7324–7331

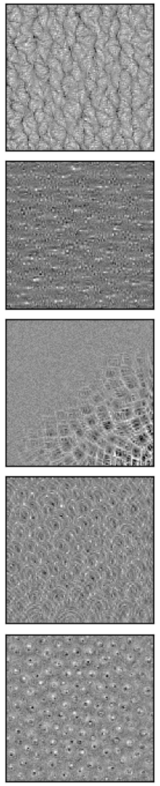
Deep Learning for Microstructure Featurization

SEM Image

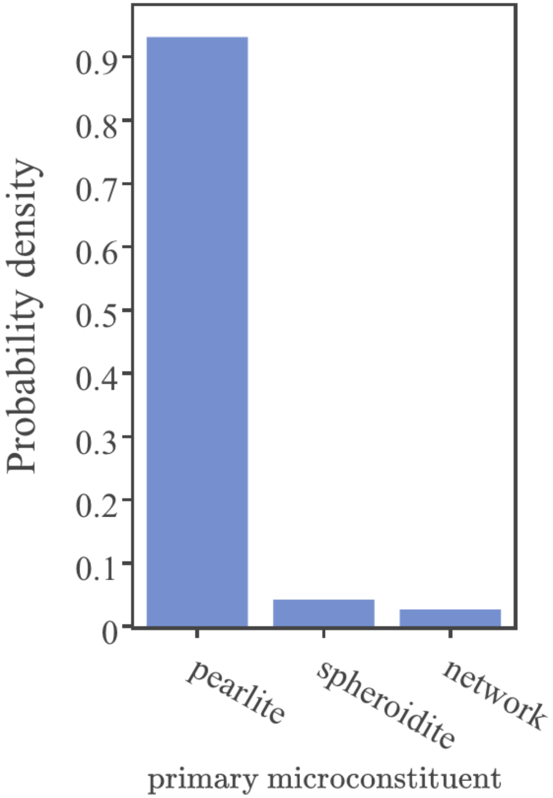


Deep Learning Featurization

Vector of Textures



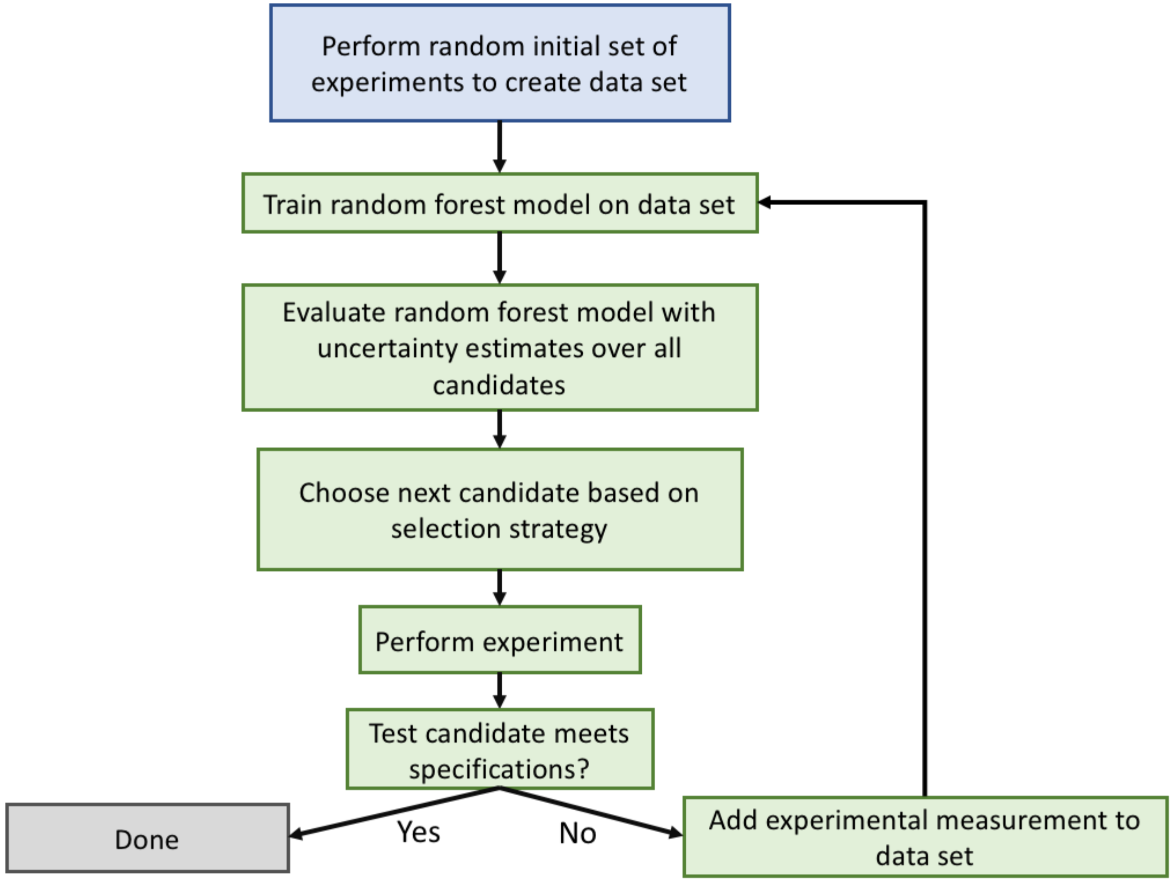
Predictions



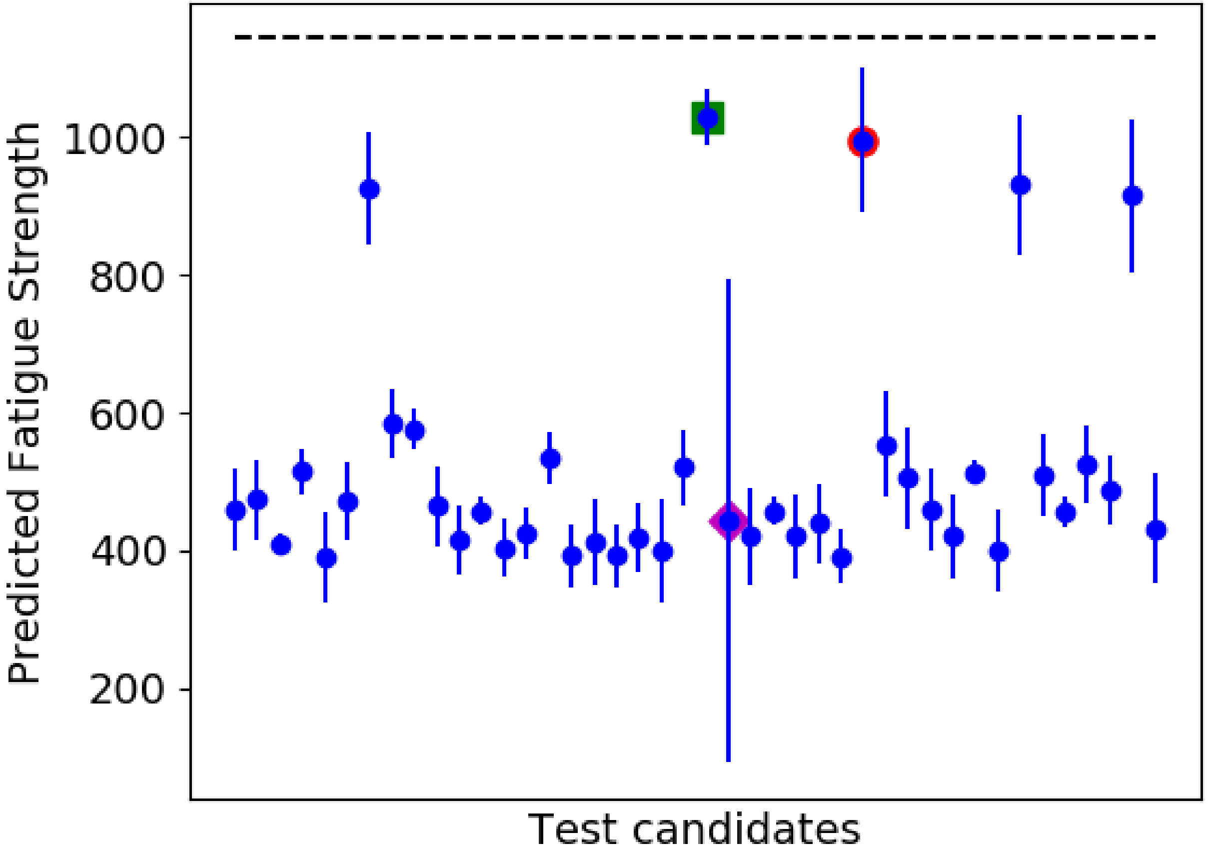
The Holm group at CMU is leading the way on open data and methods in machine learning for microstructure; see e.g. DeCost et al., UHCSDB: UltraHigh Carbon Steel Micrograph DataBase, *IMMI* (2017).

Emerging Methods to Accelerate Materials Discovery

Forest with Uncertainty Estimates for Learning Sequentially (FUELS)



Candidate Selection Strategy



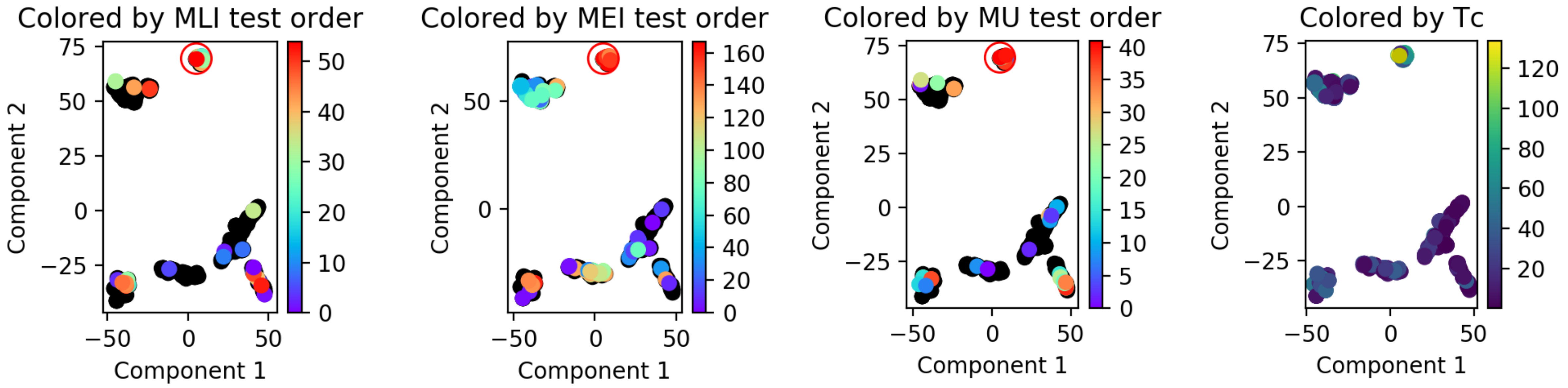
Strategies:

- Maximum Likelihood of Improvement (MLI
- Maximum Expected Improvement (MEI)
- Maximum Uncertainty (MU)

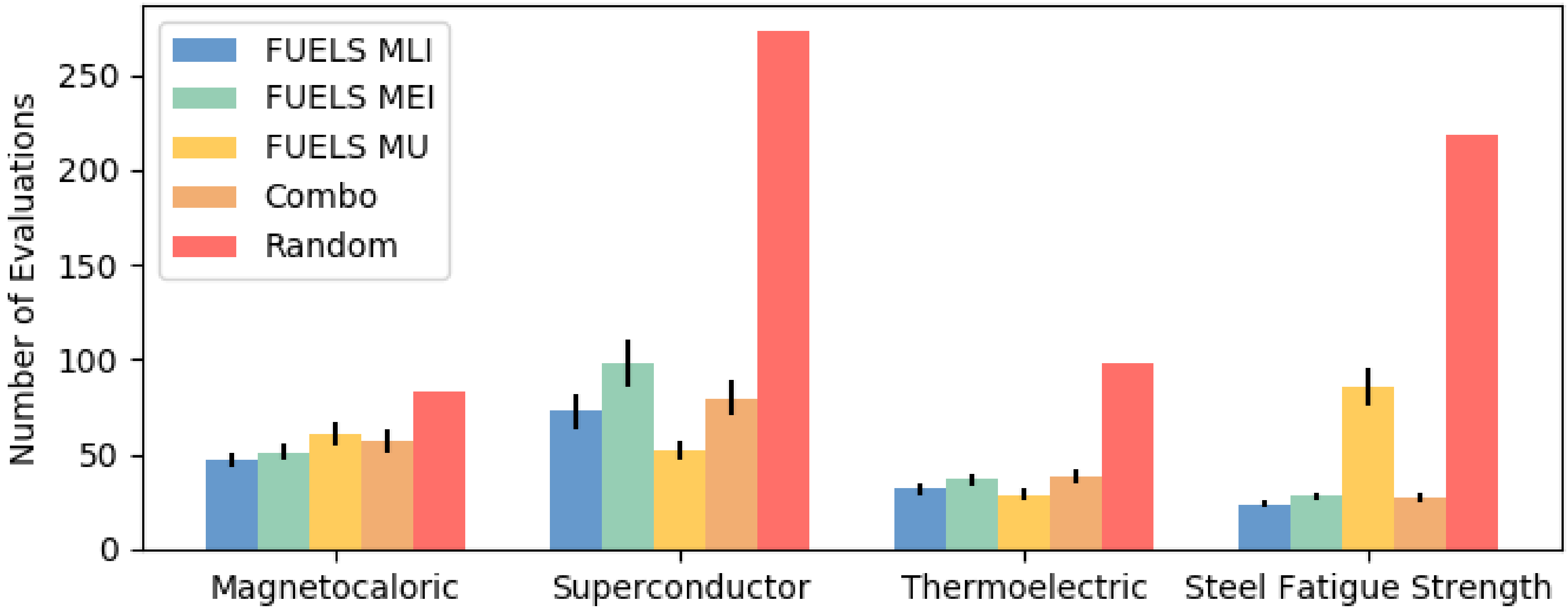
Test Cases

- Four materials problems
    1. Magnetocalorics (maximize magnetic deformation)
    2. Superconduction (maximize critical temperature)
    3. Thermoelectrics (maximize ZT figure of merit)
    4. Steel Fatigue Strength (maximize fatigue strength)
- Five candidate selection strategies:
    1. Maximum Expected Improvement
    2. Maximum Likelihood of Improvement
    3. Maximum Uncertainty
    4. COMBO Bayesian Optimization (Ueno et al. 2016)
    5. Random guessing
- Inputs: 20-60 features based on composition and processing
- Goal: Find optimal candidate after fewest number of “measurements”

Visualizing the Optimization Path



Benchmark Results

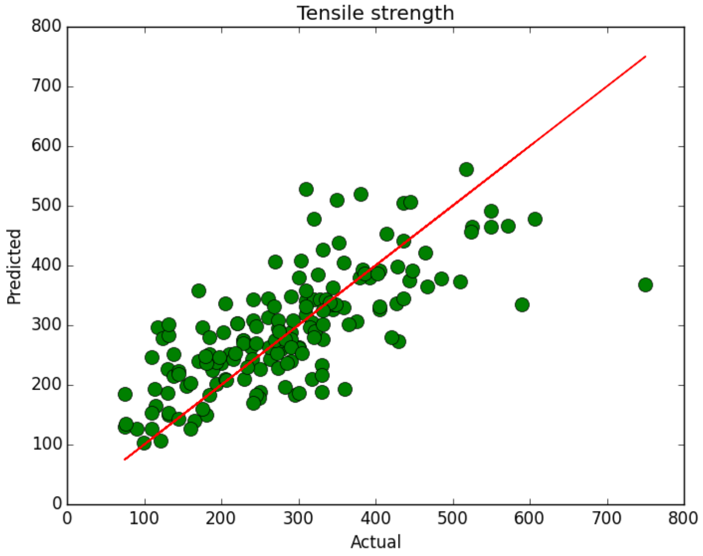


Integrating Known Physics in ML

Collaboration with Computherm to demonstrate benefits of CALPHAD data in training ML to predict Al alloy mech properties

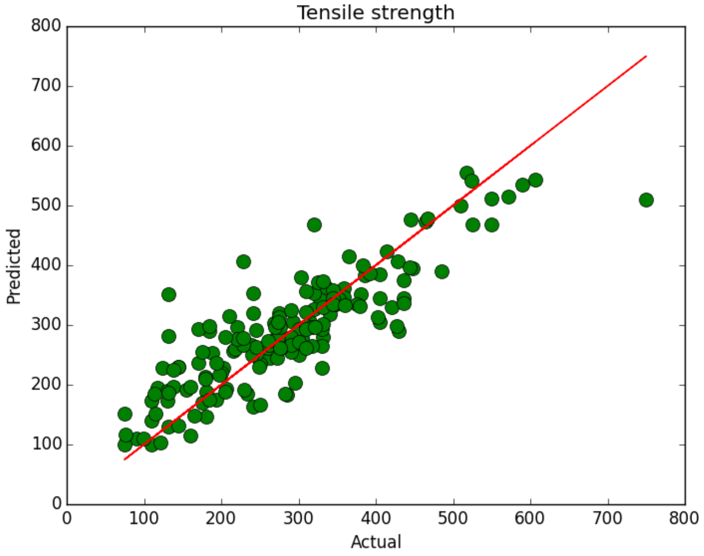
ML without CALPHAD

RMSE = 82 MPa

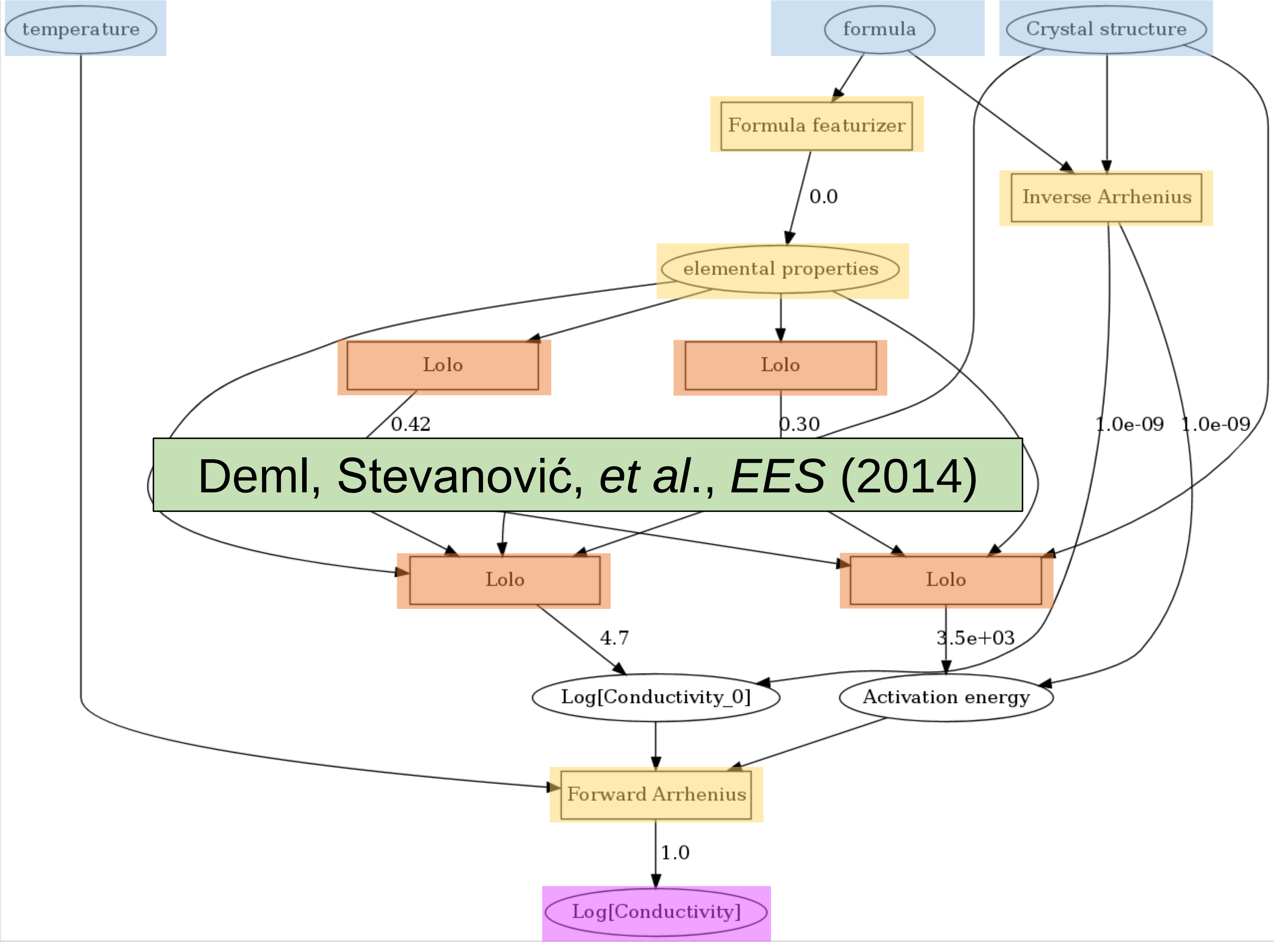


ML with CALPHAD

RMSE = 61 MPa

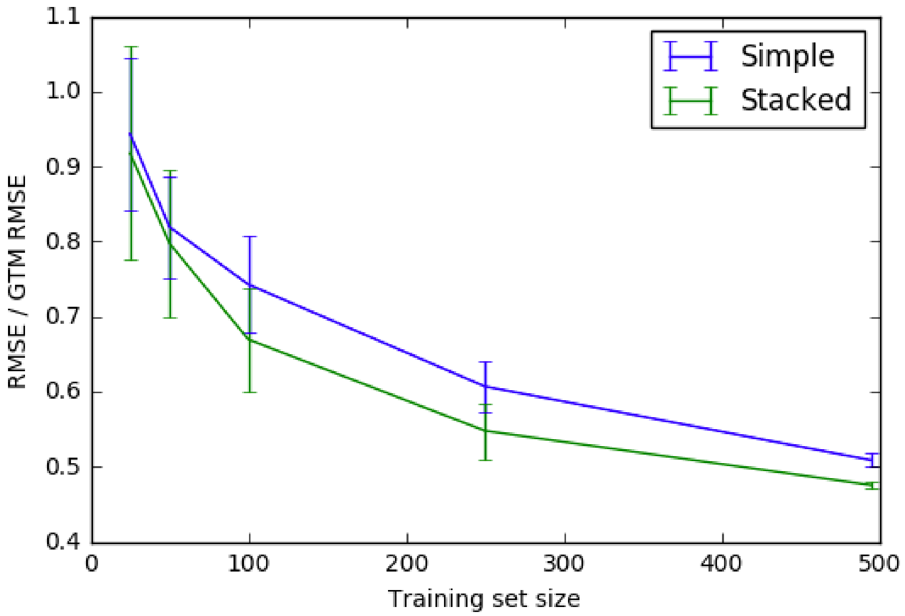


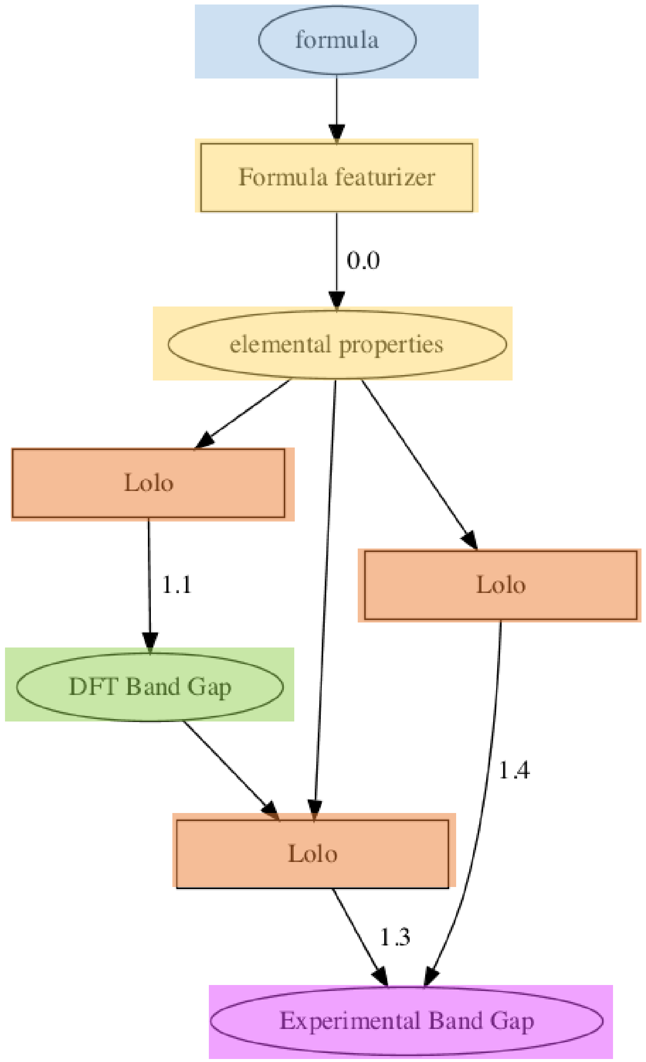
Citrine Informatics



- Experimental measurement (data)
- Experimental inputs
- Computational data
- Theory and known relationships
- Machine learning glue

Using even small number (~2k) calculations from Materials Project as a node on the relation graph reduces error in band gap model



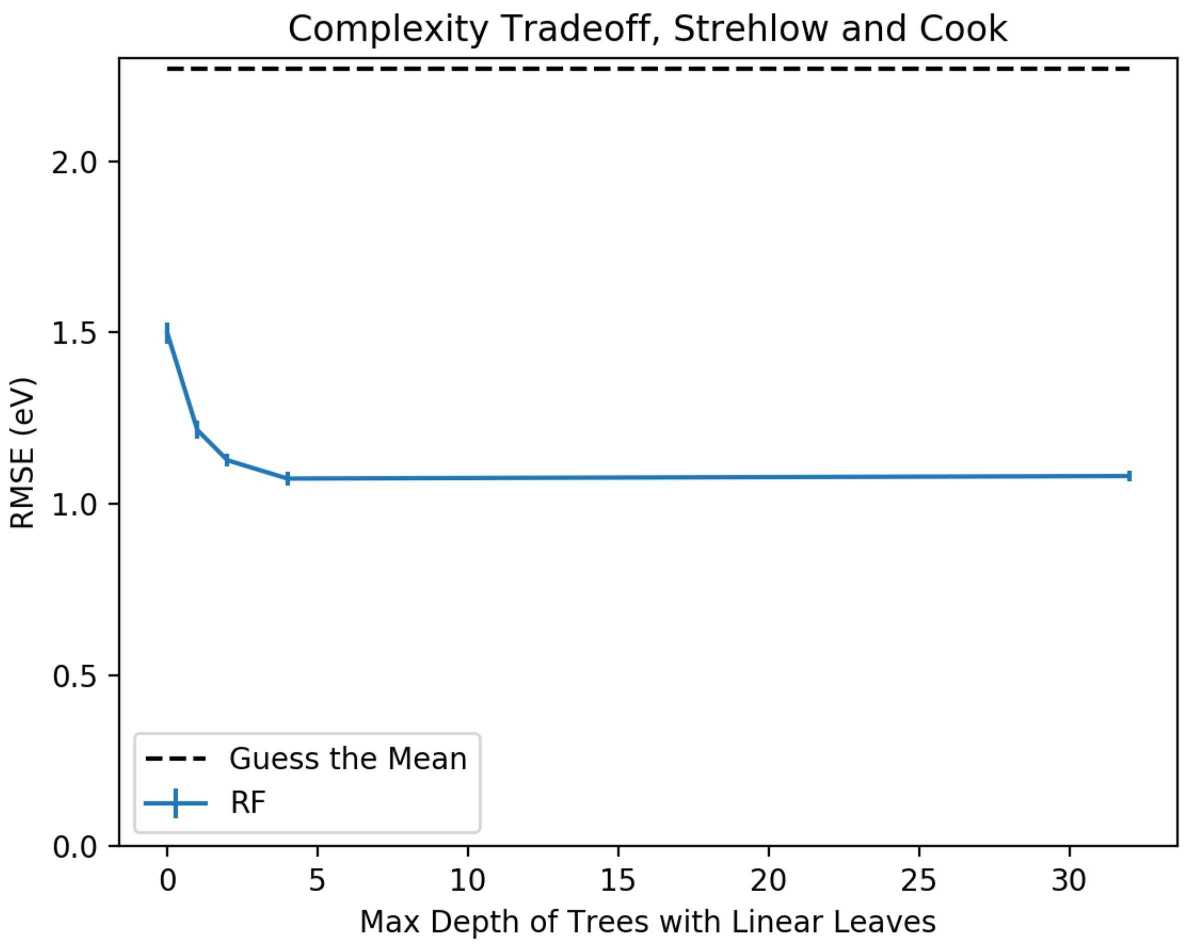


Barriers to Materials Discovery Nirvana

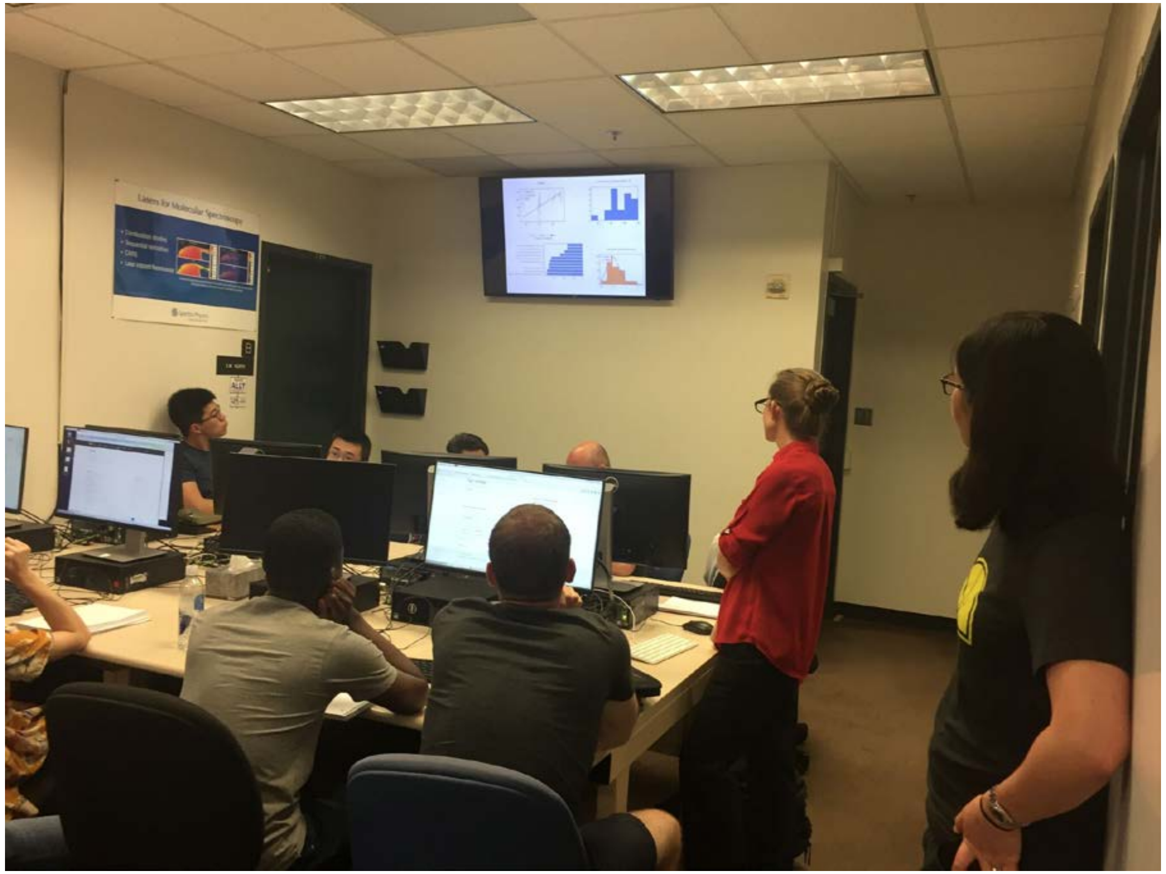
Democratized (free, easy-to-use) infrastructure, data, ML

Education for physical scientists in machine learning

Physical Interpretability vs. Accuracy




UW-Madison MSE
@UWMadisonMSE
Follow
We are excited to announce that department undergrad Vaness Meschke has won the @Citrine_io NextGen Fellowship! skunkworks.engr.wisc.edu/vaness-meschke ...
9:54 AM - 13 Jun 2017
4 Retweets 4 Likes
Citrine
4
4



2017 Summer School on Computational Materials Science Across Scales, Texas A&M

Summary

Machine learning has significantly accelerated materials discovery across a variety of application areas

In our experience, lack of tools and education prevent such acceleration from becoming ubiquitous

We strive to give materials scientists **both the platform and knowledge** to accelerate their work with machine learning