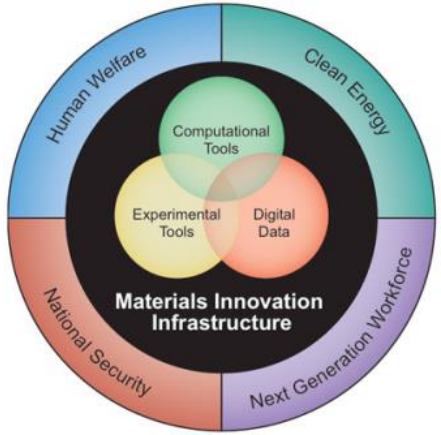
NIST
National Institute of Standards and Technology
U.S. Department of Commerce

JARVIS-ML: Physics inspired AI for fast and accurate screening of materials

Crystals, Surfaces, Grain-boundaries, Molecules, Proteins

Kamal Choudhary, James Hickman, Brian DeCost, Francesca Tavazza
NIST, Gaithersburg
Aalto University, May 11



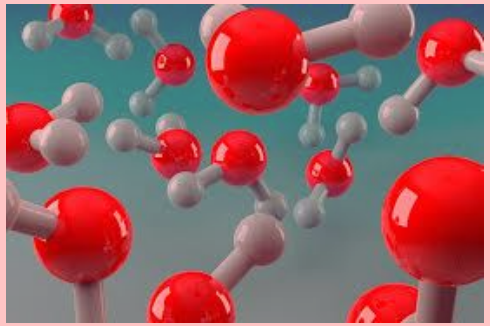
JARVIS

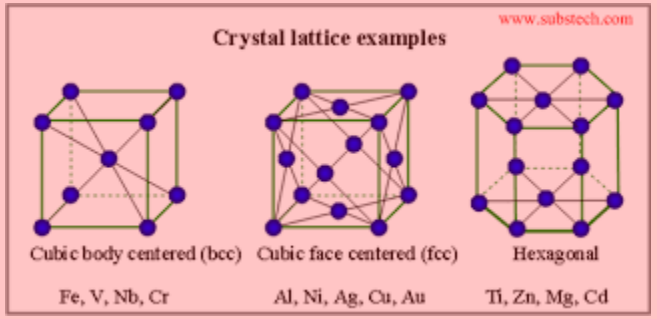
Motivation

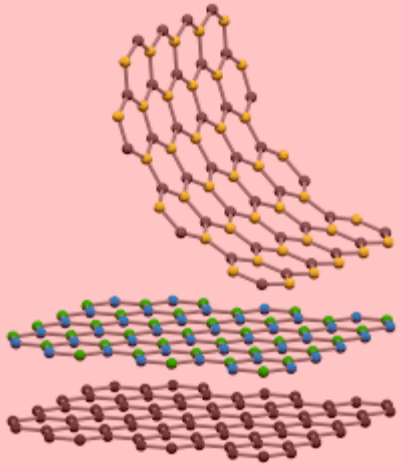
Materials Genome Initiative

National Quantum Initiative

Unification with ML models







Experiments **Computation** **ML?**

JARVIS-DFT, FF and ML datasets and tools

>30000 bulk, 900 monolayer materials

SCIENTIFIC REPORTS

Article | OPEN | Published: 12 July 2017

High-throughput Identification and Characterization of Two-dimensional Materials using Density functional theory

Kamal Choudhary, Irina Kalish, Ryan Beams & Francesca Tavazza

Scientific Reports **7**, Article number: 5179 (2017) | Download Citation

SCIENTIFIC DATA

Data Descriptor | OPEN | Published: 08 May 2018

Computational screening of high-performance optoelectronic materials using OptB88vdW and TB-mBJ formalisms

Kamal Choudhary, Qin Zhang, Andrew C.E. Reid, Sugata Chowdhury, Nhan Van Nguyen, Zachary Trautt, Marcus W. Newrock, Faical Yannick Congo & Francesca Tavazza

Scientific Data **5**, Article number: 180082 (2018) | Download Citation

ELSEVIER

Computational Materials Science

Volume 161, 15 April 2019, Pages 300-308

Convergence and machine learning predictions of Monkhorst-Pack k-points and plane-wave cut-off in high-throughput DFT calculations

PHYSICAL REVIEW B

covering condensed matter and materials physics

Highlights Recent Accepted Authors Referees Search Press About

Access by

Elastic properties of bulk and low-dimensional materials using van der Waals density functional

Kama

Phys.

SCIENTIFIC DATA

Data Descriptor | OPEN | Published: 31 January 2017

Evaluation and comparison of classical interatomic potentials through a user-friendly interactive web-interface

Kamal Choudhary, Faical Yannick P. Congo, Tao Liang, Chandler Becker, Richard G. Hennig & Francesca Tavazza

PHYSICAL REVIEW MATERIALS

Highlights Recent Authors Referees Search Press About Staff

Access by N

Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape

Kamal Choudhary, Brian DeCost, and Francesca Tavazza

Phys. Rev. Materials **2**, 083801 – Published 3 August 2018

IOPscience Journals Books Publishing Support Login Search IOPscien

Journal of Physics: Condensed Matter

PAPER

High-throughput assessment of vacancy formation and surface energies of materials using classical force-fields

arXiv.org > cond-mat > arXiv:1810.10640

Condensed Matter > Materials Science

High-throughput discovery of topological materials using spin-orbit spillage

Kamal Choudhary, Kevin F. Garrity, Francesca Tavazza

(Submitted on 24 Oct 2018)

arXiv.org > cond-mat > arXiv:1903.06651

Search or Ar

(Help | Advanced

Condensed Matter > Materials Science

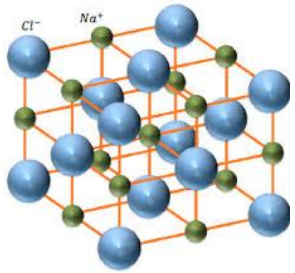
Accelerated Discovery of Efficient Solar-cell Materials using Quantum and Machine-learning Methods

Kamal Choudhary, Marnik Bercx, Jie Jiang, Ruth Pachter, Dirk Lamoen, Francesca Tavazza

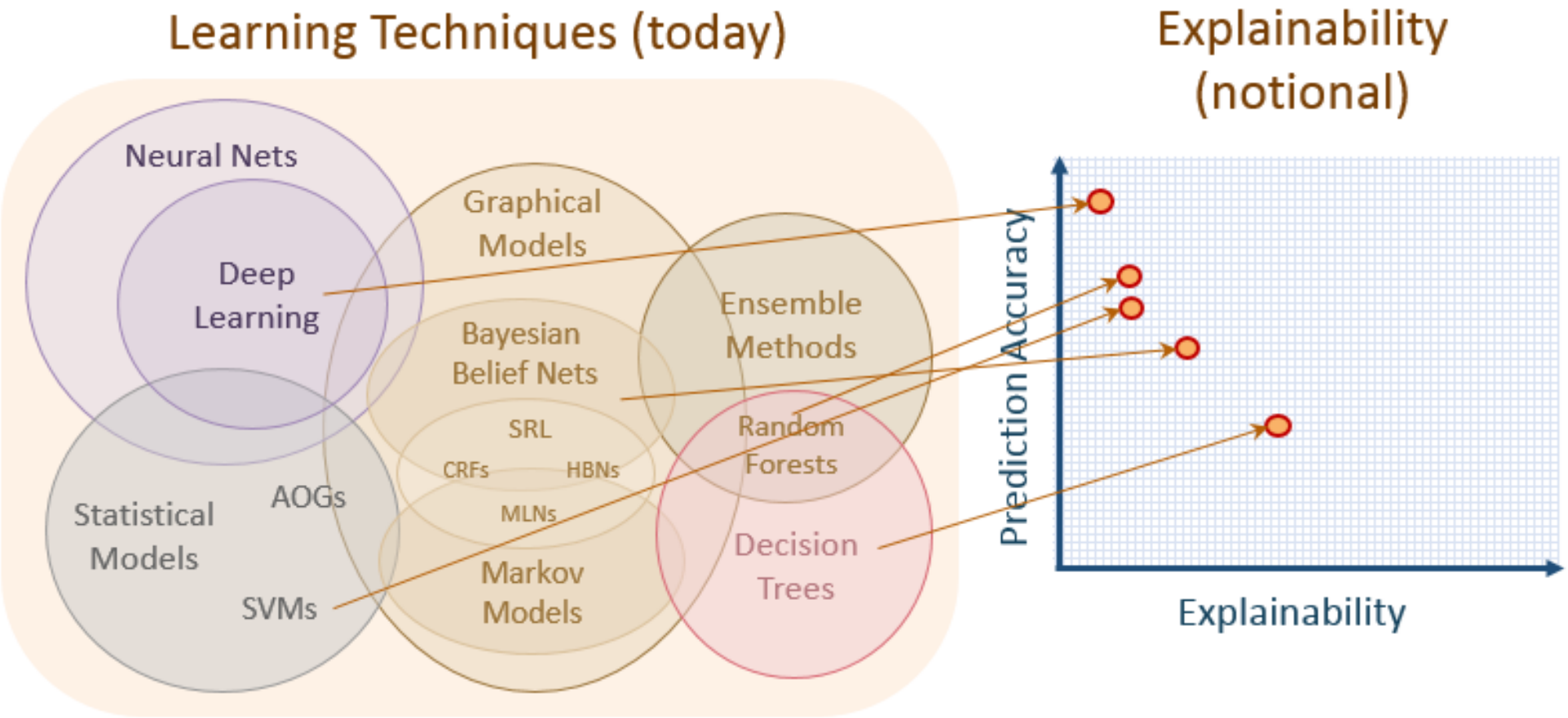
Artificial Intelligence

- Deep Neural network (many layers of NN, less domain knowledge)
- Neural network (feedforward, backward propagation)
- Statistical/Machine learning
  - Supervised (classification, regression) Input-output
  - Unsupervised (clustering/association) Input-no output
- Uncertain reasoning (Bayesian net, game theory)
- Logic (Fuzzy logic, product control system)
- Optimization problem (swarm, genetic algorithm)

Cat/dog images vs Materials data



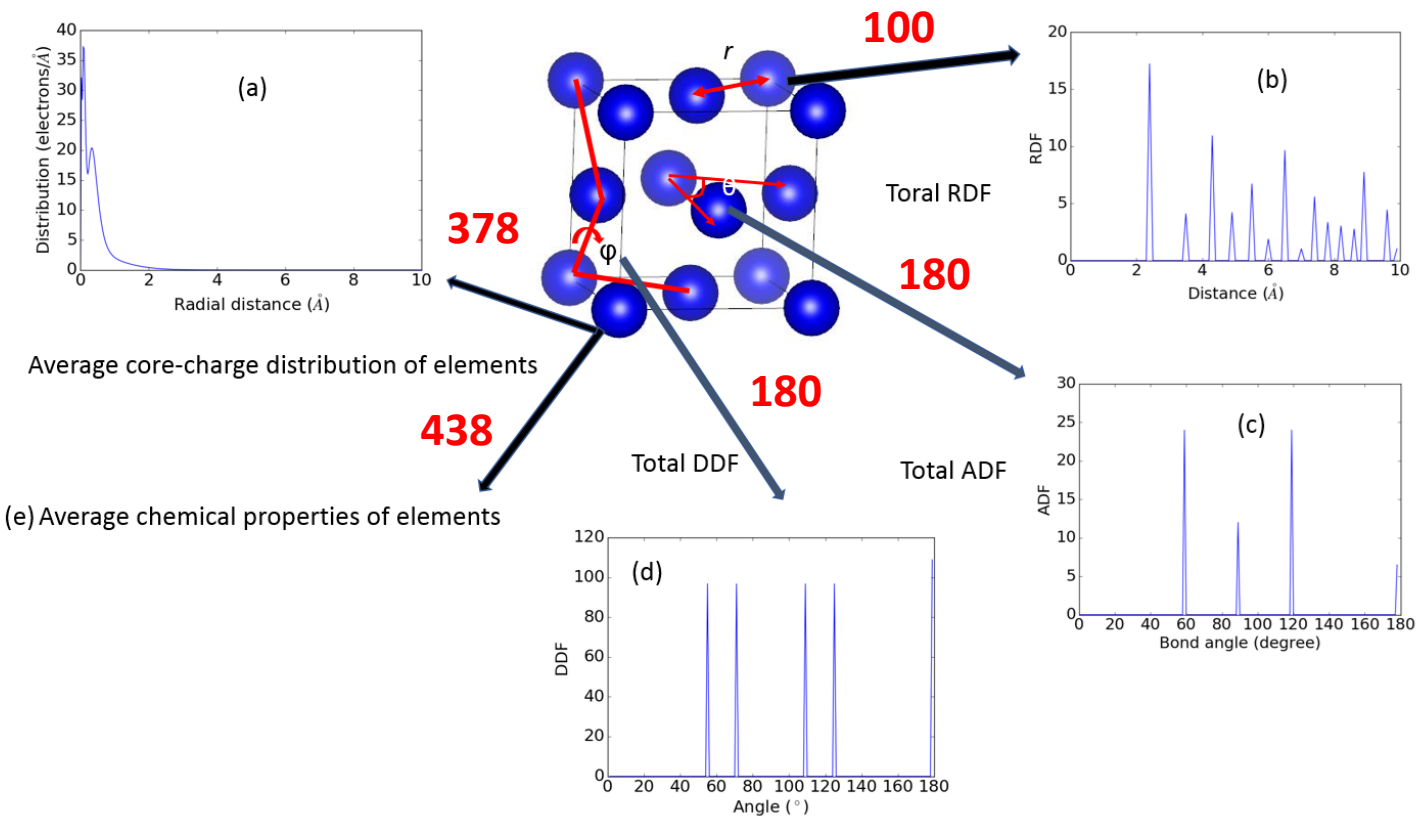
Explainable AI



https://www.darpa.mil/program/explainable-artificial-intelligence

CFID descriptors

1557 descriptors/features for one material



- Classical force-field inspired descriptors
- Arithmetic operations (mean, sum, std. deviation…) of **electronegativity, atomic radii, heat of fusion**,….

of atoms at each site

(example: Electronegativity of Mo+Mo+S+S+S+S)/6 = 0.15

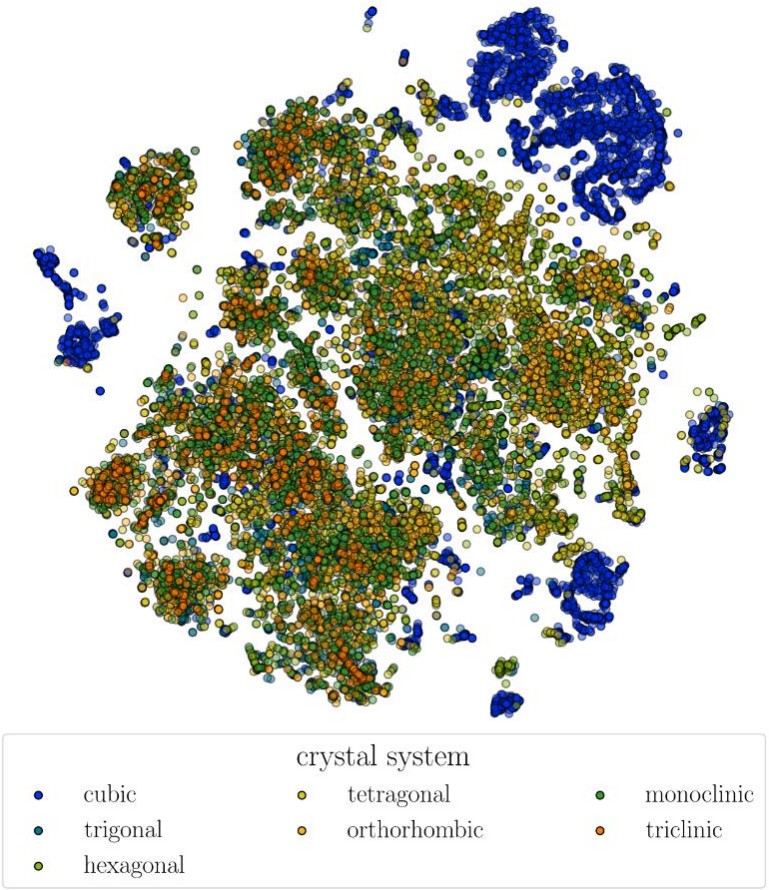
- Atomic bond distance based descriptors
- Angle based descriptors

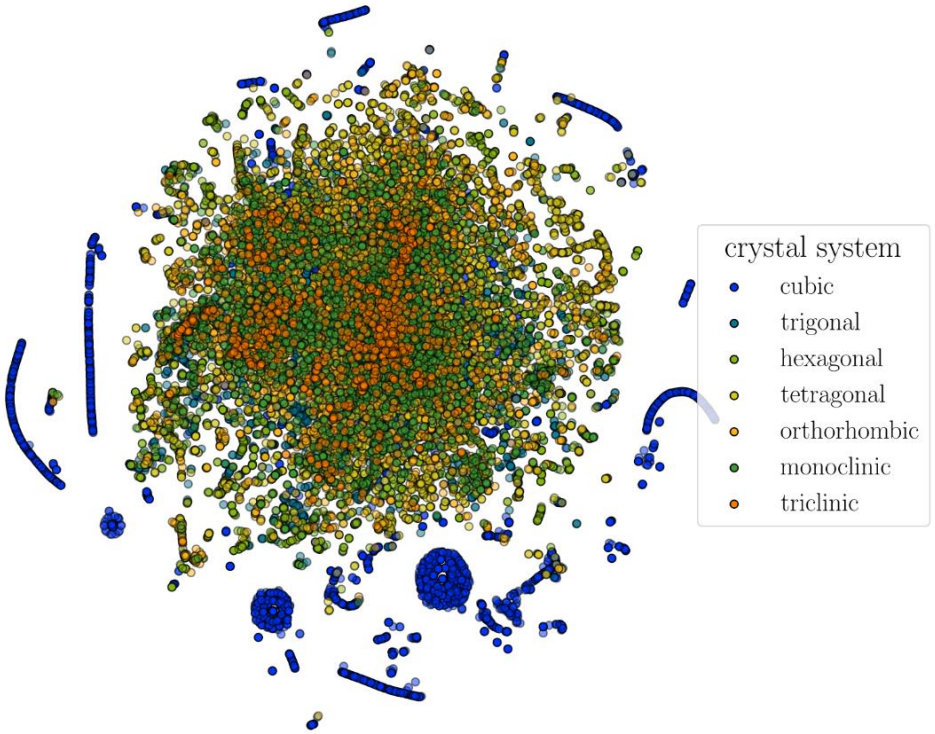
1.5 % unary, 26% binary, 56 % ternary, 13 % quaternary, 2 % quinary and 1% senary compounds, 1-96 atoms

https://github.com/usnistgov/jarvis

https://hackingmaterials.github.io/matminer/index.html

Visualizing multi-dimensional data with t-SNE

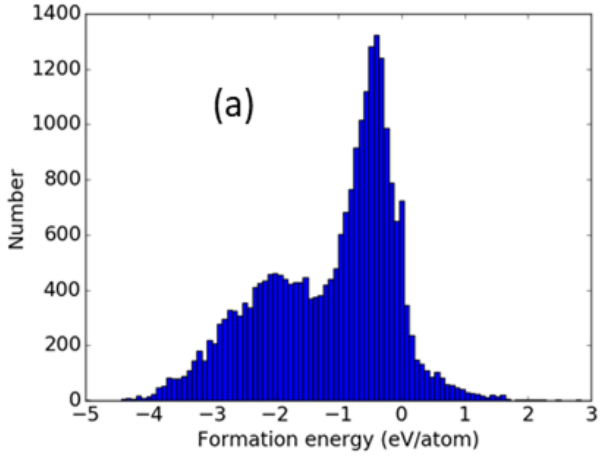


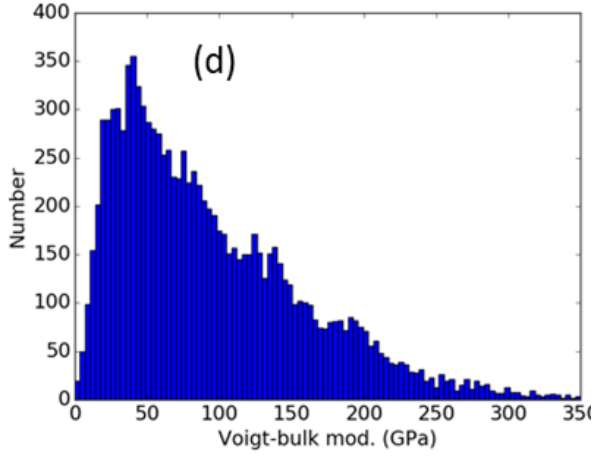


- Converts similarities between data points to joint probabilities
- Visualization with t-SNE for ~25000 materials

http://holoviews.org/
http://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html

Properties of interest & histogram plots

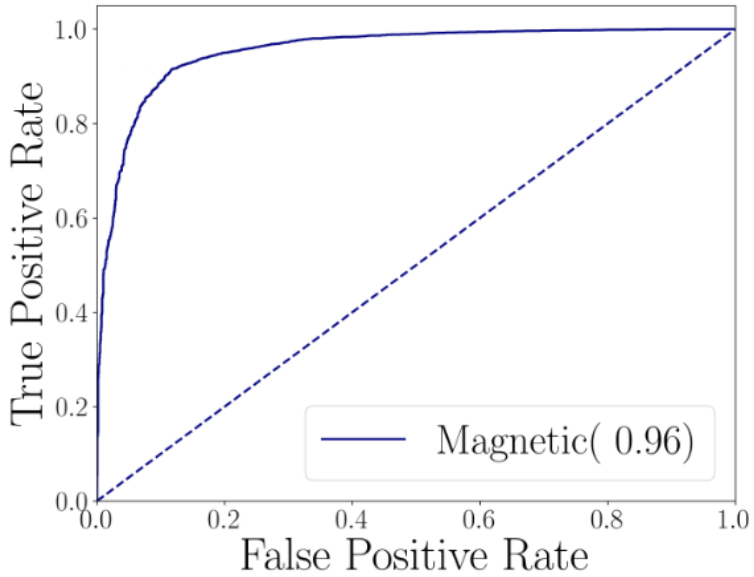


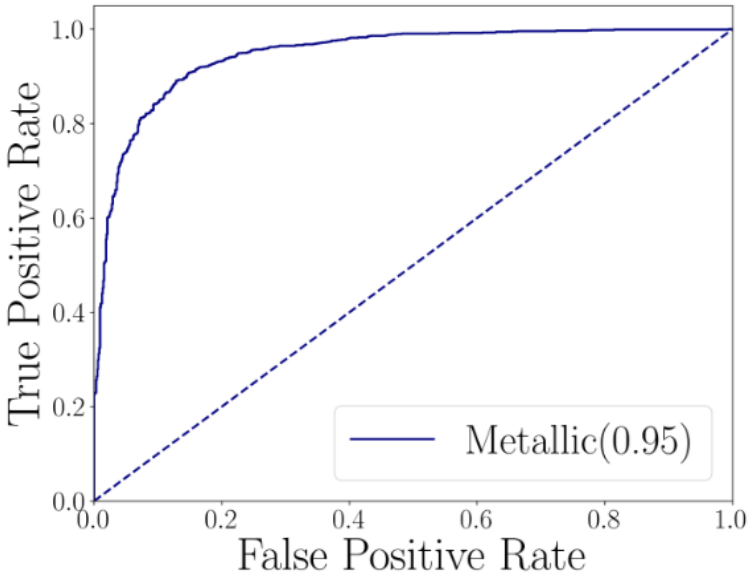


- Formation energy
- Bandgap
- Bulk/shear modulus
- K-points, cut-off
- Thermoelectric metrics
- Solar-cell efficiency
- Refractive index
- 2D Exfoliation energy
- Surface energy
- Grain boundary energy
- Topological spillage

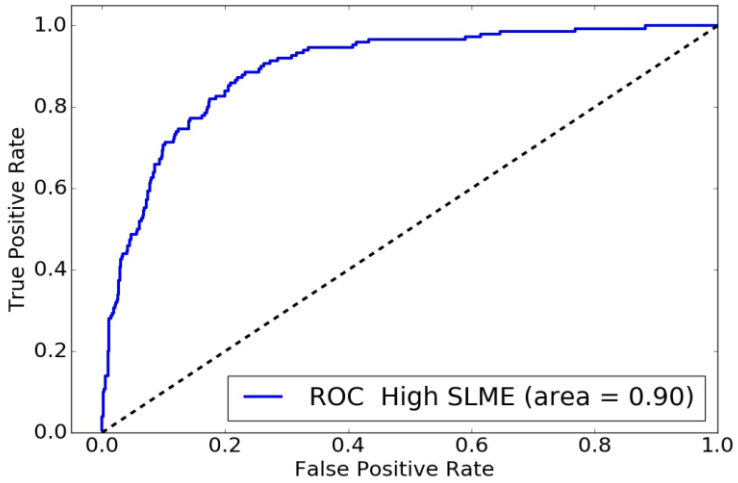
New

Classification: ROC curves



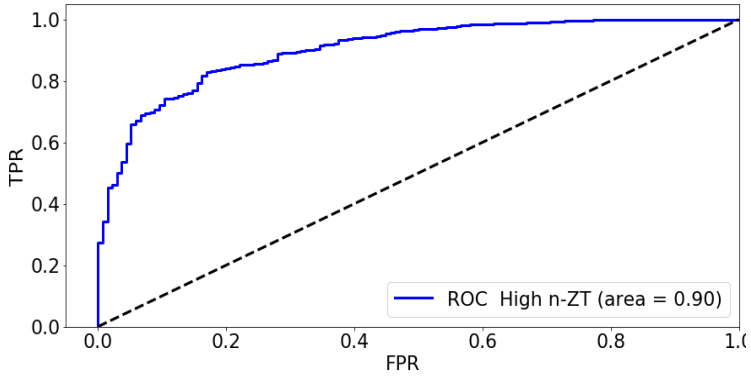


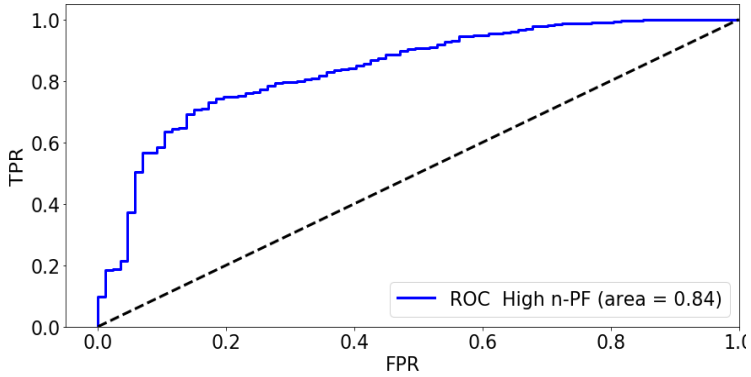
Solar cell efficiency

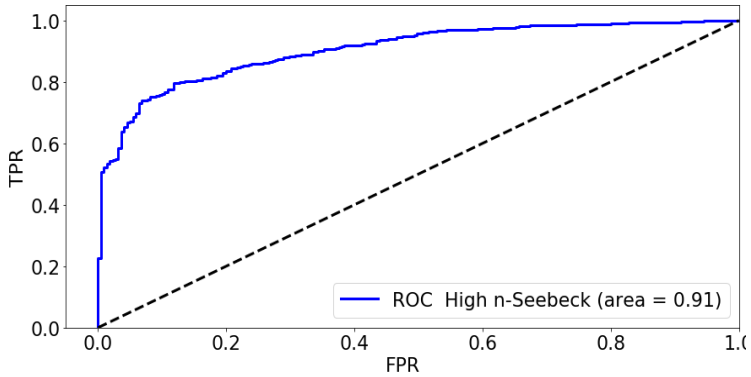


Perfect model area: 1
Random guess: 0.5

Thermoelectric

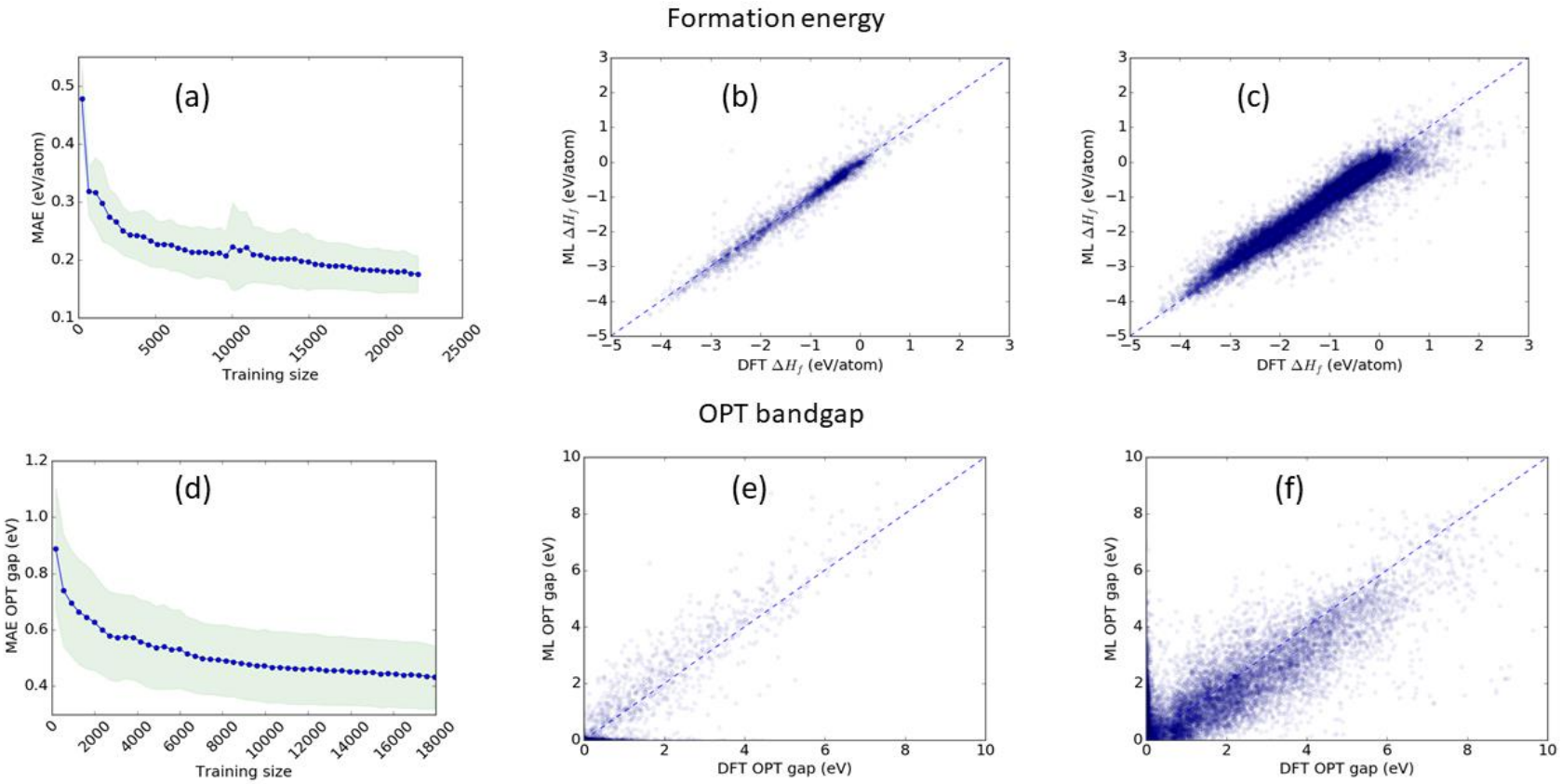






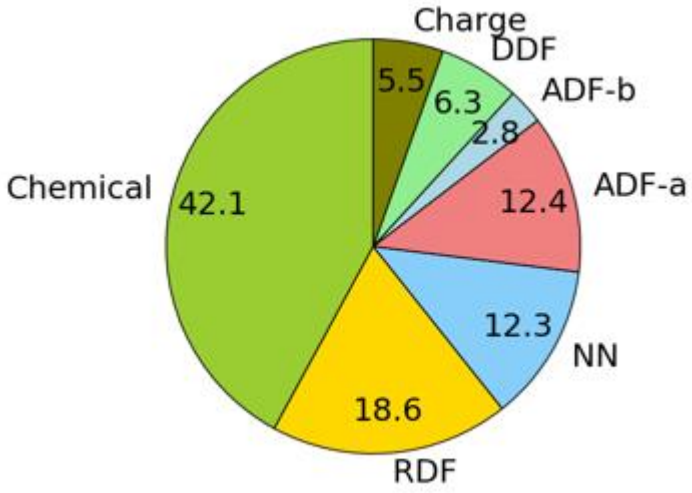
Email: kamal.choudhary@nist.gov

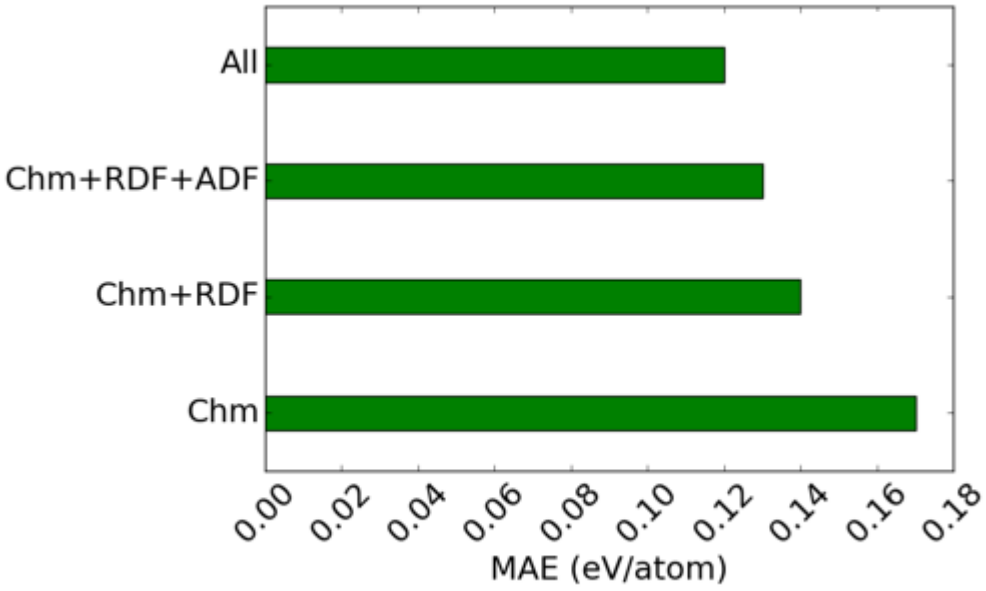
Regression models: formation energy and bandgap model



Learning curve shows scope of further improvement

Explainability: feature importance

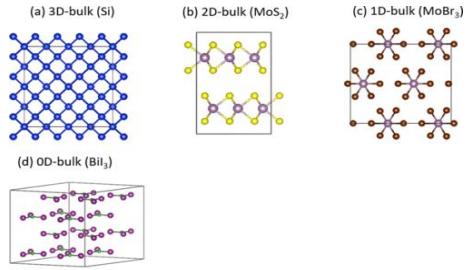




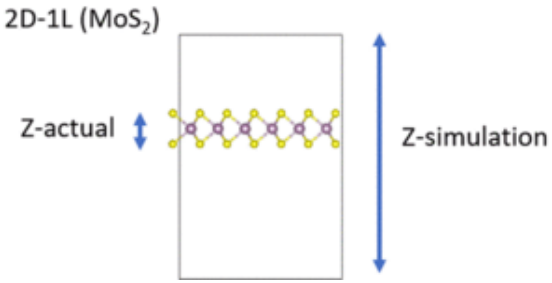
- Chemical features most important followed by RDF and NN
- Incrementally adding structural features decreases MAE

Regression

Property	#Data-points	$MAE_{CFID-DFT}$	$MAE_{CFID-DFT}$ (CV)	$MAE_{DFT-Exp}$
Formation energy (eV/atom)	24549	0.12	0.17±0.05	0.136
OPT-bandgap (eV)	22404	0.32	0.37±0.24	1.33
MBJ-bandgap (eV)	10499	0.44	0.56±0.26	0.51
Bulk modulus (GPa)	10954	10.5	12.63±3.3	10.0
Shear modulus (GPa)	10954	9.5	11.55±3.15	10.0
OPT-n_x (no unit)	12299	0.54	0.65±0.15	1.78
OPT-n_y (no unit)	12299	0.55	0.65±0.16	-
OPT-n_z (no unit)	12299	0.55	0.70±0.18	-
MBJ-n_x (no unit)	6628	0.45	0.55±0.14	1.6
MBJ-n_y (no unit)	6628	0.50	0.51±0.15	-
MBJ-n_z (no unit)	6628	0.46	0.54±0.14	-
Exfoliation energy (meV/atom)	616	37.3	60.13±10.41	-



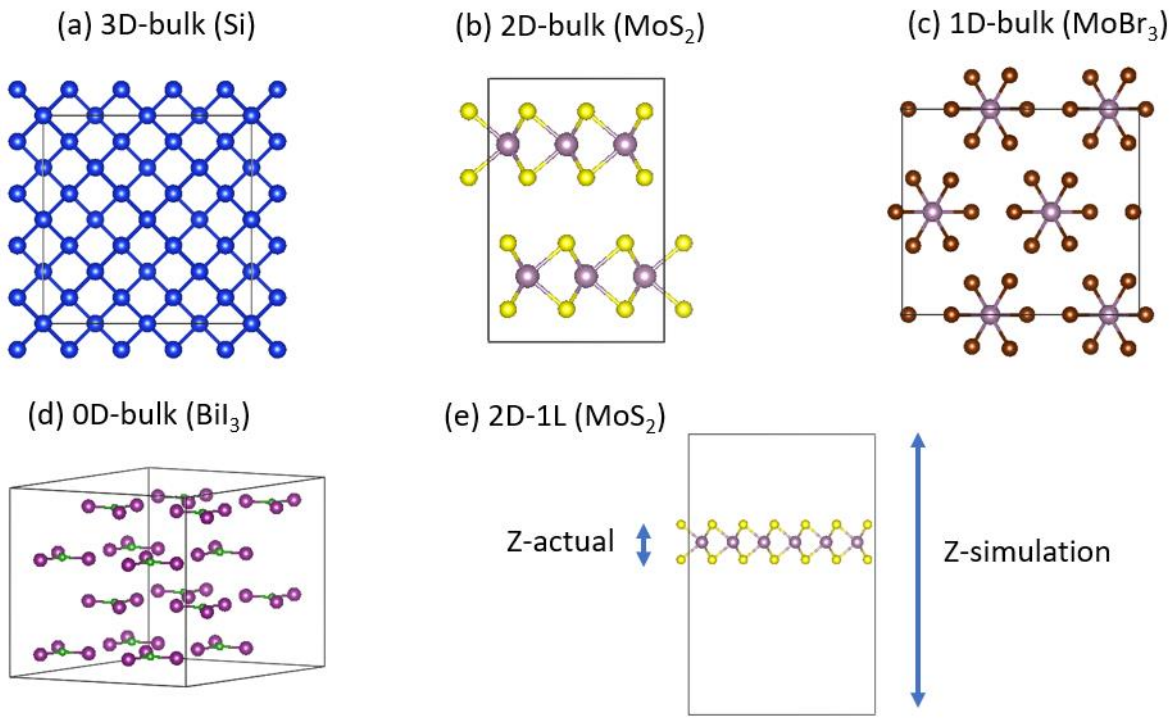
3D



2D

Email: kamal.choudhary@nist.gov

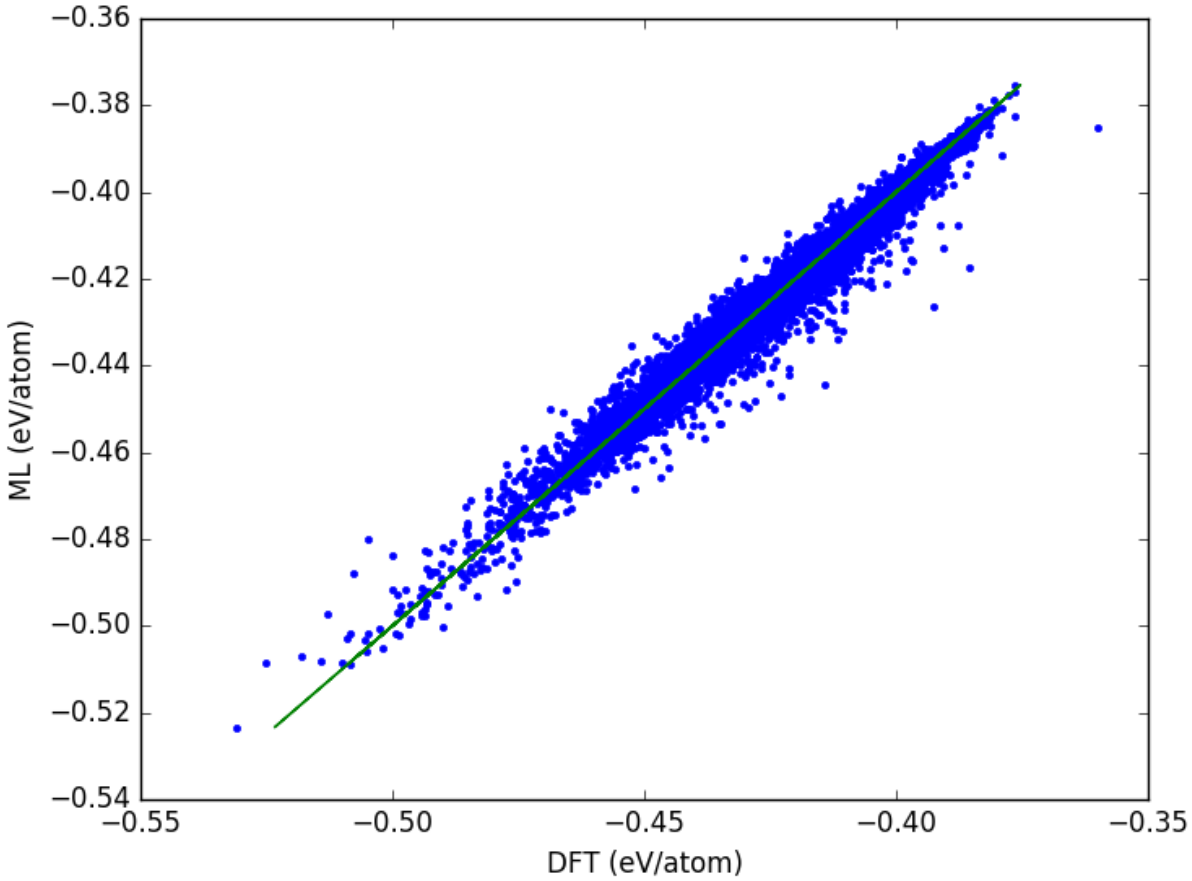
2D materials-screening example



- ~5000 2D materials predicted
- Requires expensive DFT calculations for predicting properties such as bandgap, exfoliation energy etc.
- Use of ML drops down the time to a few seconds
- Using this technique we identified new 2D materials such as CuI, InS etc.
- Validated using DFT

Molecules

MAE internal energy (eV/atom): 0.002
r^2: 0.97
(On-going work)



MENU SCIENTIFIC DATA

Data Descriptor | OPEN | Published: 05 August 2014

Quantum chemistry structures and properties of 134 kilo molecules

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp & O. Anatole von Lilienfeld

Scientific Data **1**, Article number: 140022 (2014) | Download Citation

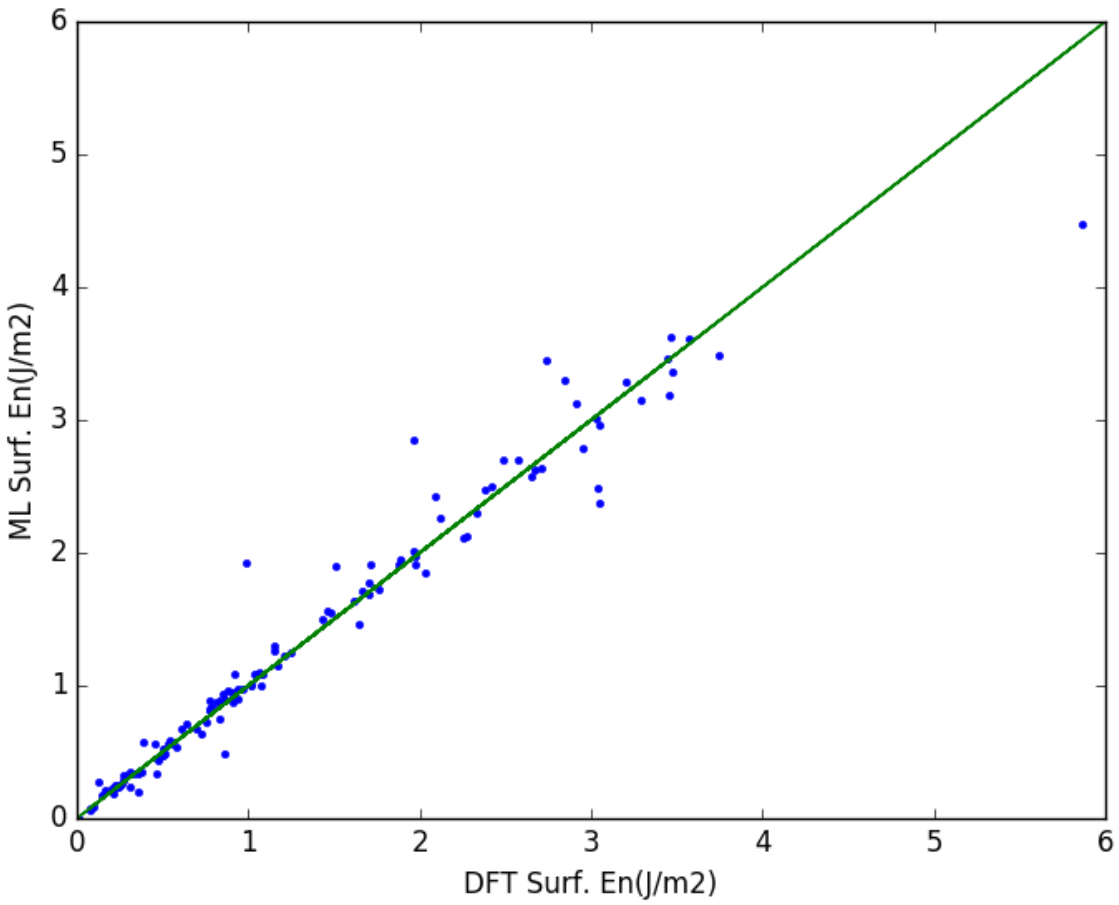
Abstract

Computational *de novo* design of new drugs and materials requires rigorous and unbiased exploration of chemical compound space. However, large uncharted territories persist due to its size scaling combinatorially with molecular size. We report computed geometric, energetic, electronic, and thermodynamic properties for 134k stable small organic molecules made up of CHONF. These molecules correspond to the subset of all 133,885 species with up to nine heavy

Email: kamal.choudhary@nist.gov

Surfaces

MAE: 0.13 J/m^2
r^2:0.94
(On-going work)



SCIENTIFIC DATA

Data Descriptor | OPEN | Published: 13 September 2016

Surface energies of elemental crystals

Richard Tran, Zihan Xu, Balachandran Radhakrishnan, Donald Winston, Wenhao Sun, Kristin A. Persson & Shyue Ping Ong

Scientific Data **3**, Article number: 160080 (2016) | Download Citation

Abstract

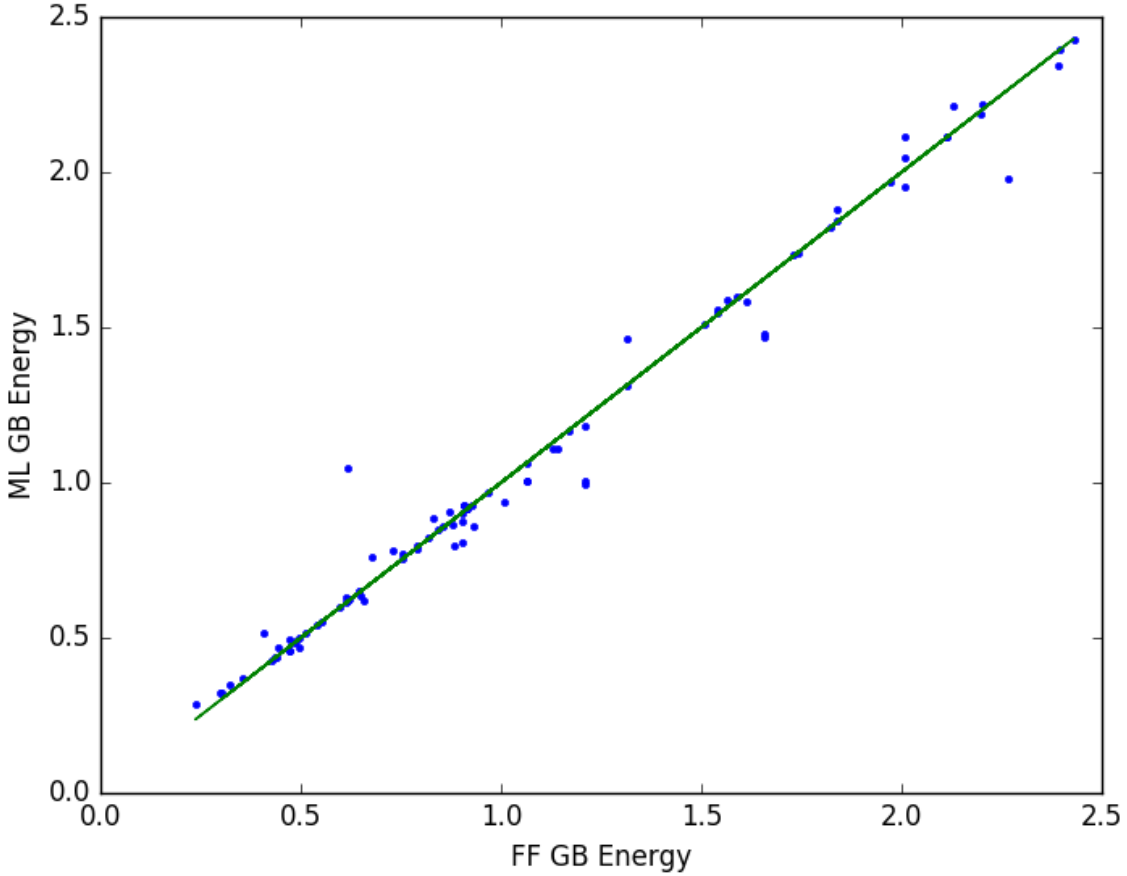
The surface energy is a fundamental property of the different facets of a crystal that is crucial to the understanding of various phenomena like surface segregation, roughening, catalytic activity, and the crystal's equilibrium shape. Such surface phenomena are especially important at the nanoscale, where the large surface area to volume ratios lead to properties that are significantly different from the bulk. In this work, we present the largest database of calculated surface energies for elemental crystals to date. This database contains the surface energies of more than 100 polymorphs of about 70 elements, up to a maximum Miller index of two and three for non-cubic and cubic crystals, respectively. Well-known reconstruction schemes are also accounted

Email: kamal.choudhary@nist.gov

Grain boundaries

MAE: 0.04 J/m^2
r^2:0.98
(On-going work)

Symmetric tilt GBs
FCC: Al, Ni, Cu, Ag, Au, Pd, Pt
BCC: Fe, W, Ta, Mo
Diamond: Si

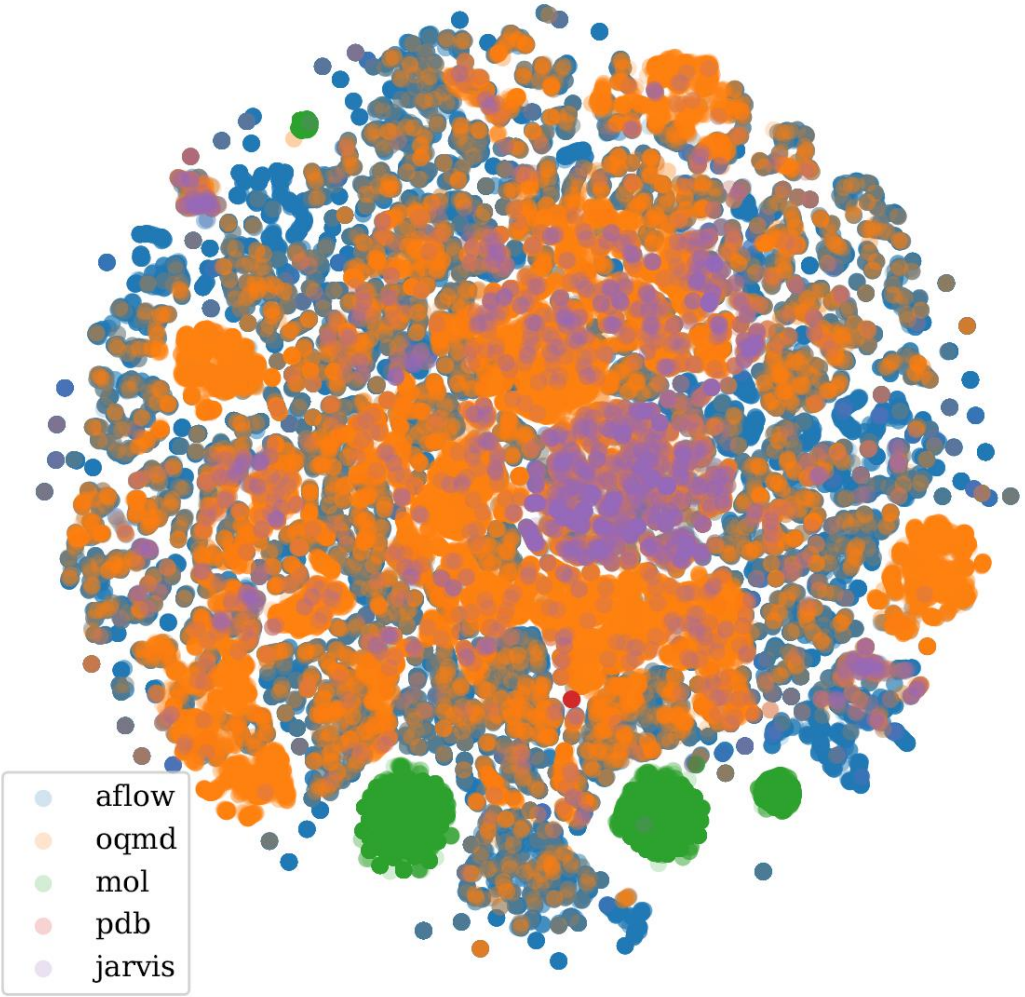


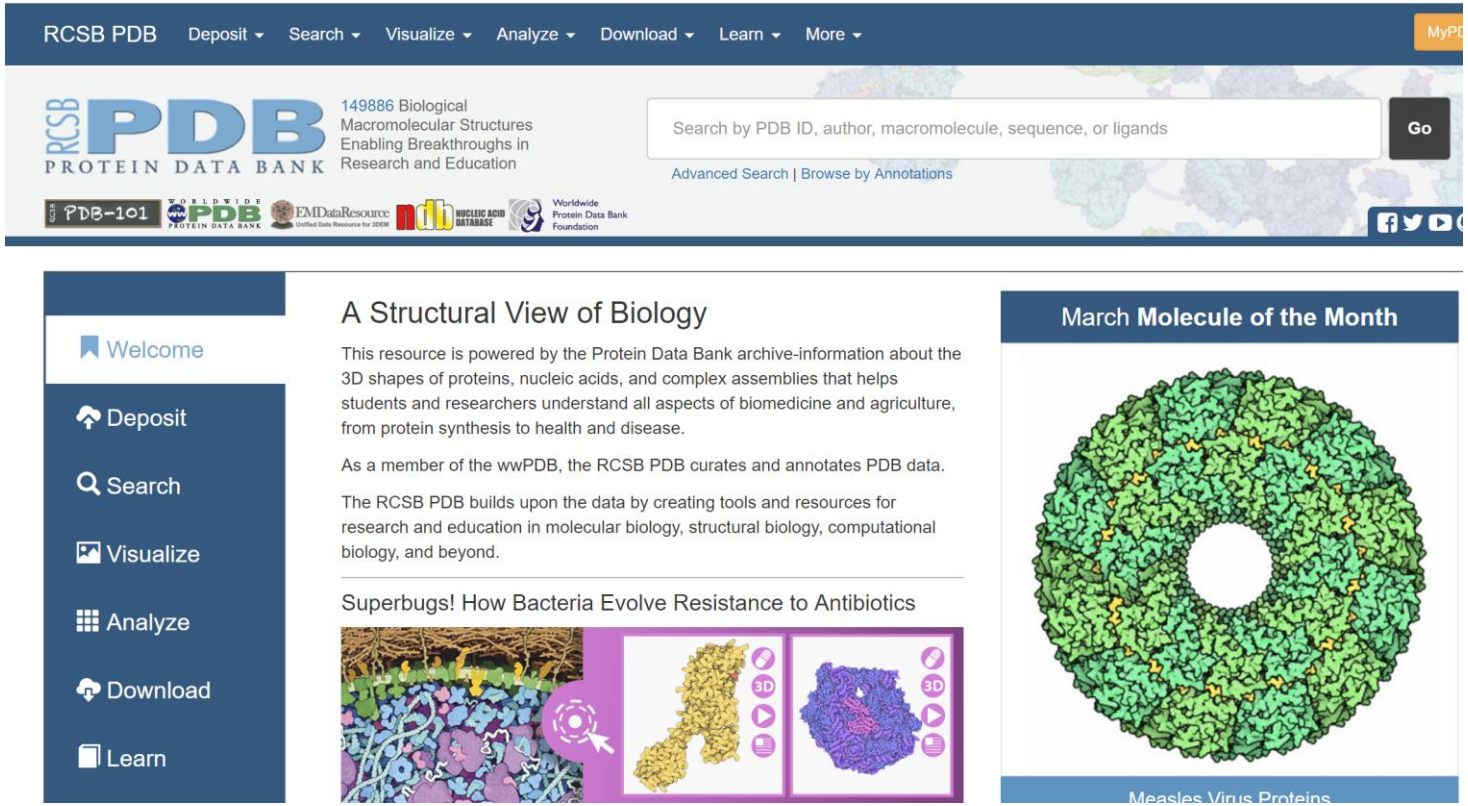
Email: kamal.choudhary@nist.gov

Proteins

t-SNE visualization
(On-going work)

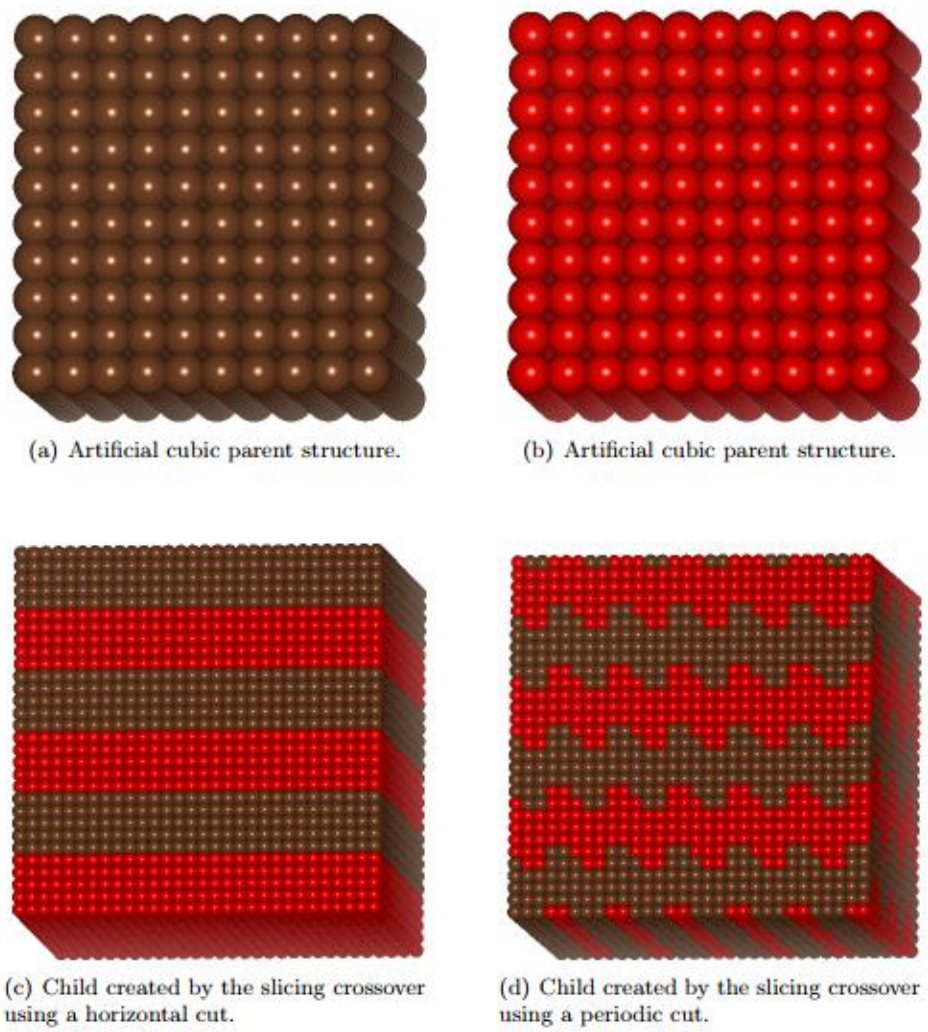
>10000 proteins
>630000 AFLOW
>360000 OQMD
>111000 COD
>820000 MP
> 30000 JV





Email: kamal.choudhary@nist.gov

Stringent validation: Genetic Algorithm with ML



(a) Artificial cubic parent structure.

(b) Artificial cubic parent structure.

(c) Child created by the slicing crossover using a horizontal cut.

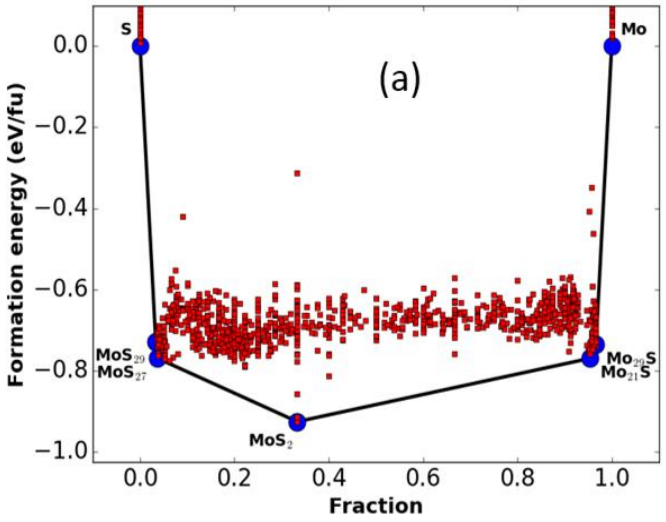
(d) Child created by the slicing crossover using a periodic cut.

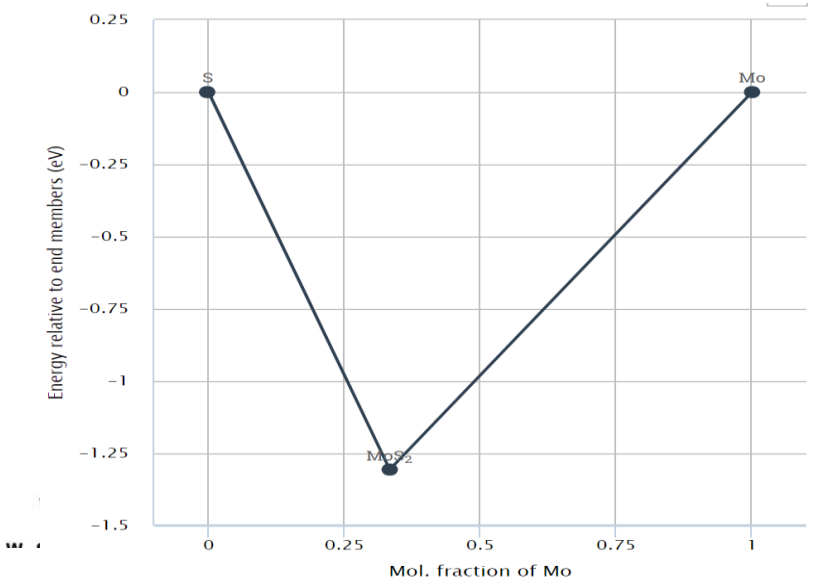
Picture from GASP manual

- Based on **'Survival of the fittest'** theory: fitness of crystal structure based on energy of structure
- Parents to offspring crystal structure
- Generally energy is obtained from DFT, MD…let's try ML ?

D. M. Deaven, Molecular geometry optimization with a genetic algorithm, Physical Review Letters, 75 (1995)
G. Ceder, Data-mining-driven quantum mechanics for the prediction of structure, MRS Bulletin, 31 (2006)
https://github.com/henniggroup/GASP-python/

Genetic algorithm with ML



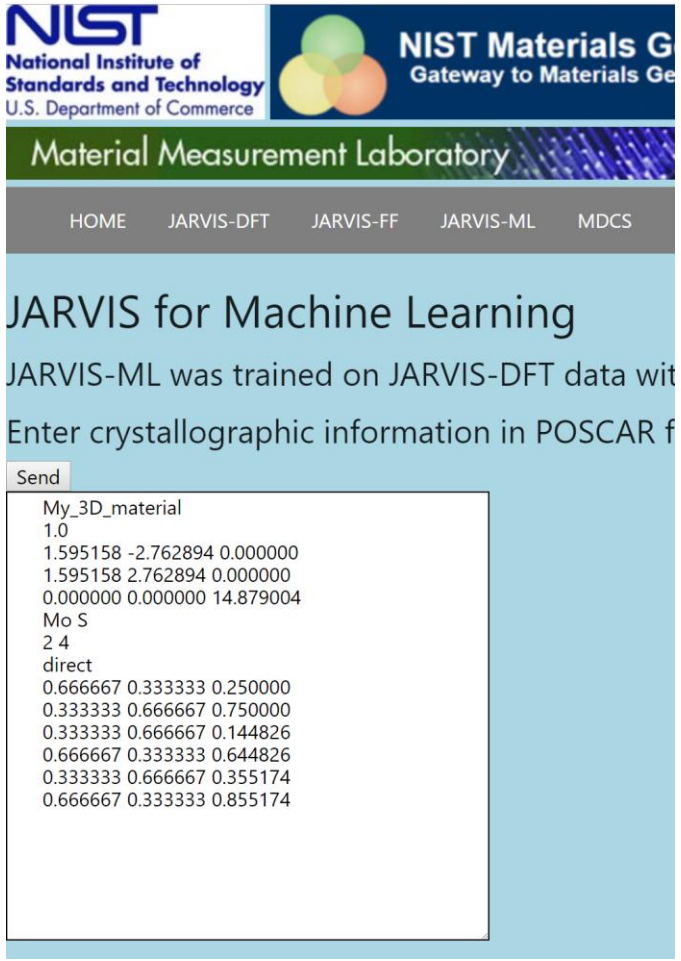


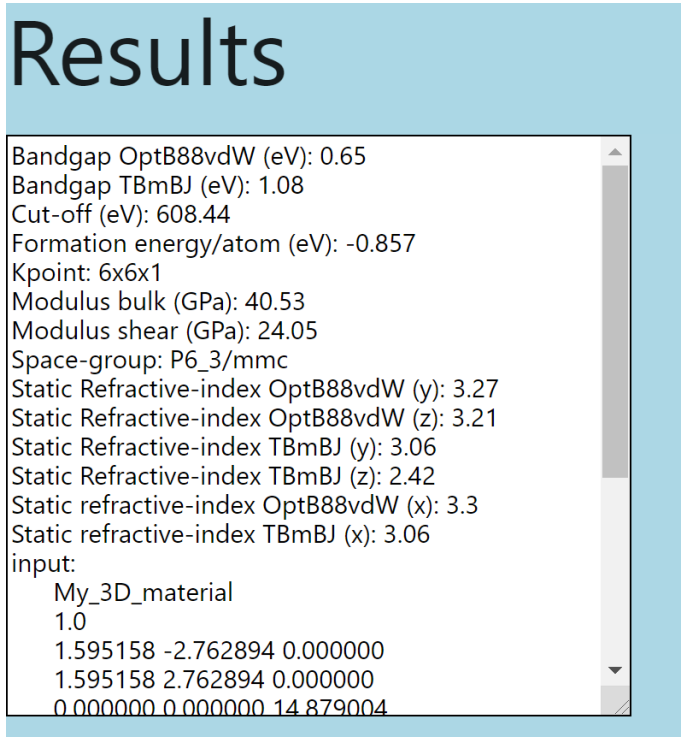
- New way of validating ML model for materials
- MoS_2, WS_2 indeed stable as in DFT and experiments
- Need further verification for low-lying energy structures with DFT

Web-app: DEMO

- **How to use the model?**

https://www.ctcms.nist.gov/jarvisml/





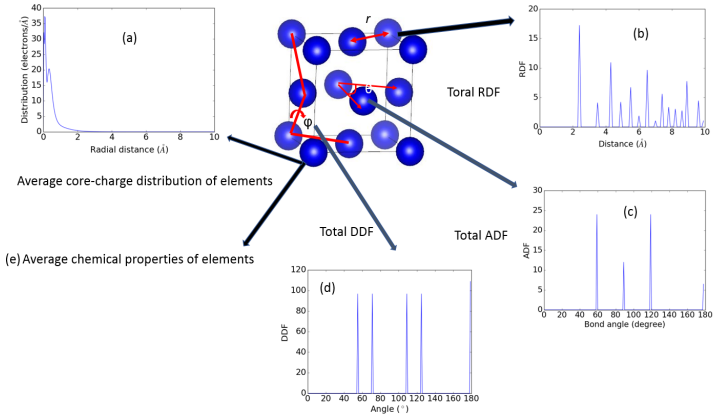
Summary

- Unified machine learning descriptors for various classes of materials
- All the code and data publicly available
- Formation energy convex hull test, beyond data-science metric
- Web-app for on-the fly prediction of properties
- **AIMS workshop**: August 1-2, 2019, Registration open
- More data and tools on the way
- Important links:

✓ https://jarvis.nist.gov/

✓ https://github.com/usnistgov/jarvis

✓ Slides available at: https://www.slideshare.net/KAMALCHOUDHARY4/



Thank you for your time!

Email: kamal.choudhary@nist.gov