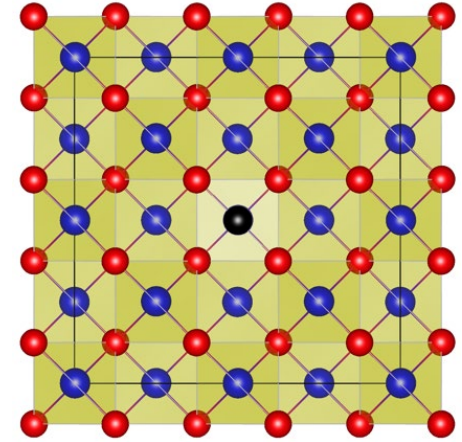# Modeling, machine learning, and characterization of defects and dopants in Cd(Se,Te)

Arun Mannodi-Kannakithodi, Fatih Sen, Eric Schwenker, Chaitanya Kolluru, Davis Unruh, Justin Pothoof, Maria Chan, Argonne National Laboratory
Tadas Paulauskis, Jinglong Guo, Robert Klie, University of Illinois Chicago
Srisuda Rojsatien, Mariana Bertoni, Arizona State University
Jerry Seidler, University of Washington

Part I: DFT Modeling + ML for Defect Property Prediction

Argonne
NATIONAL LABORATORY

Density Functional Theory + Machine Learning

Initial choices

Prototypical Structure — *Combinatorial Chemical Space* — *Choice of Atoms for Compositions*

High-Throughput DFT Computations

Input Scripts for DFT Jobs — *Tools for Automated Job Submission and Processing*

Data representation

Data Visualization — *Dataset for Machine Learning*

10^x possibilities for compounds / compositions / data points in chemical space
→ choose representative points from every region for data generation
→ choose best features of atoms, structure etc. as descriptors

ML Predictive Models: Descriptors → Properties — *Training a regression model using the Python package Scikit-learn* — *Parity Plot ML vs. DFT*

Multi-fidelity Learning

ML predictions on entire combinatorial space → high-throughput screening

Higher Fidelity Data
- Complete convergence
- Meta-GGA, HSE06, GW approx.

Increasing Accuracy & Cost

Lower Fidelity Data
- Relaxed convergence (e.g. smaller cell size, sparse k-points)
- LDA/GGA properties

ML screened candidates

Arun Mannodi-Kanakkithodi & Maria Chan, "Computational Data-Driven Materials Discovery," Trends in Chemistry 3, 79 (2021)   **mchan@anl.gov**

Predicting properties of defects

- Pb Substitution in MAPbBr$_3$: **Chem. Mater** 2019.
- Cd chalcogenides: **npj Comput. Mater** 2020.
- Primer: **Trends in Chemistry** 2021.
- All zinc blende semiconductors: **Patterns** 2022.
- With corrected DFT: **J Chem Phys** 2022
- Halide Perovskites: **Journal of Materials Science** 2022.
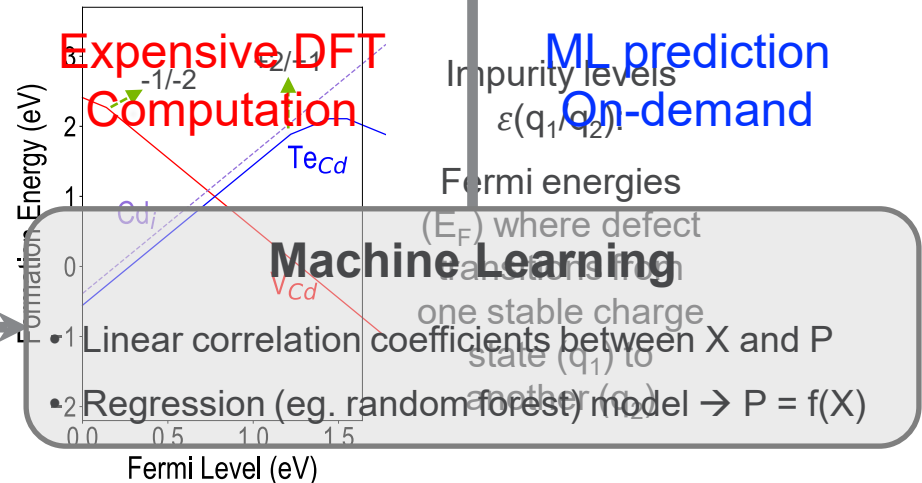
Semiconductor + impurity
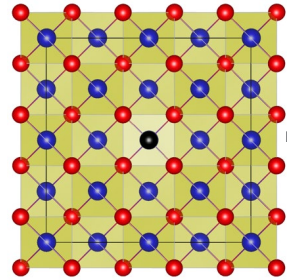
Density Functional Theory (DFT)

- $E^f(q) = E(D^q) - E(bulk) + \sum n_i \mu_i + q(E_F + E_{vbm}) + E_{corr}$
- Impurity levels: $\varepsilon(q_1/q_2) = [E^f(q_1) - E^f(q_2)] / (q_2 - q_1)$

Descriptors (X)

- Elemental properties of M
- Coordination environment
- Unit cell defect calculations

Expensive DFT Computation

ML prediction On-demand

Impurity levels $\varepsilon(q_1/q_2)$!

Fermi energies (E_F) where defect transitions from one stable charge state (q_1) to another (q_2)

Machine Learning

- Linear correlation coefficients between X and P
- Regression (eg. random forest) model → P = f(X)
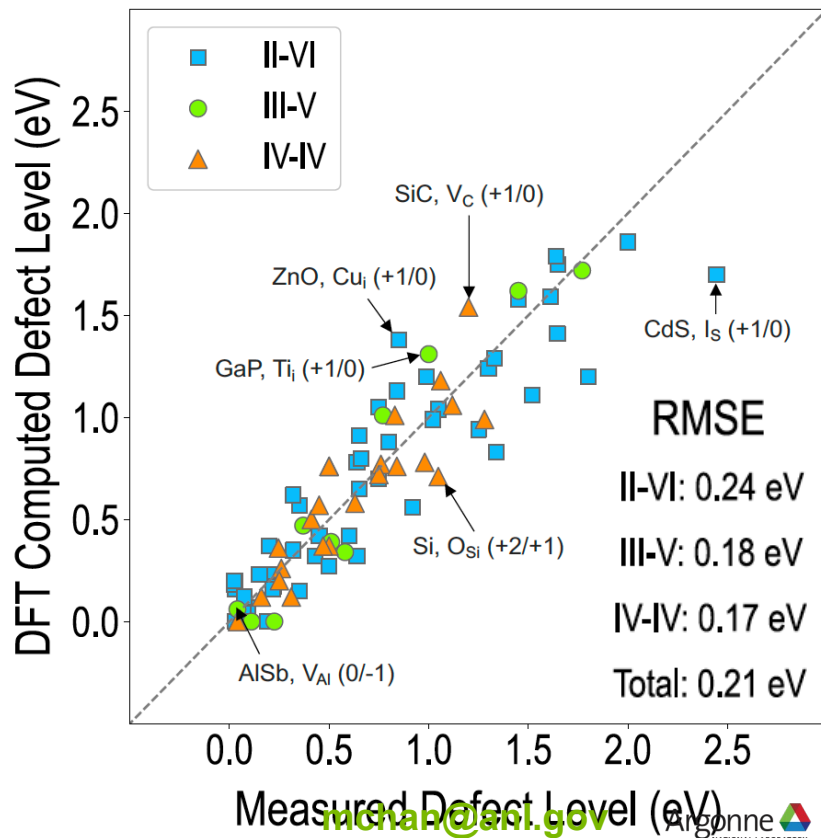

mchan@anl.gov

Accuracy of DFT defect levels?

Experimental and DFT values from literature

CdTe ↓, all zinc blende semiconductors →

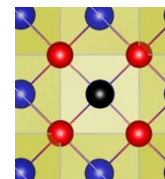
(d)

RMSE: PBE: 0.22 eV, HSE: 0.35 eV

RMSE

II-VI: 0.24 eV

III-V: 0.18 eV

IV-IV: 0.17 eV

Total: 0.21 eV
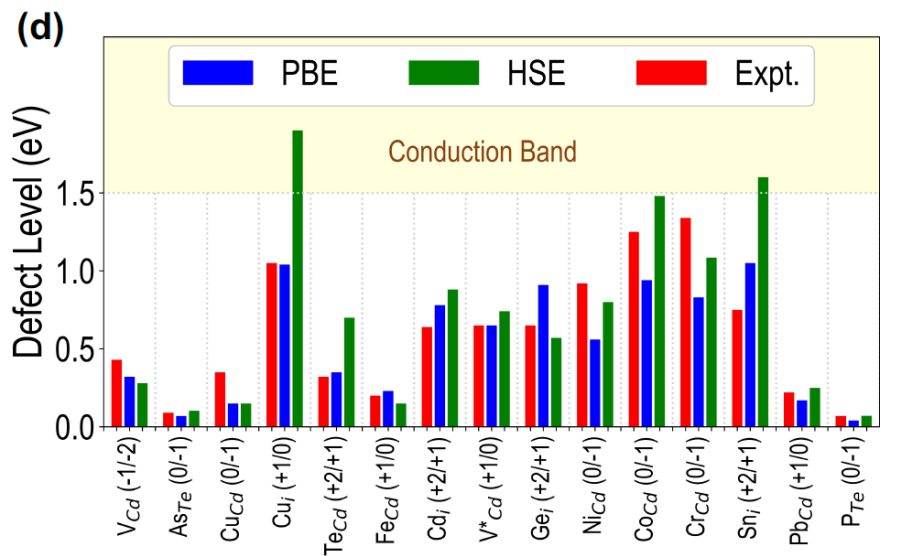
mchan@anl.gov

Descriptors for Machine Learning

+unit cell (8 atom) defect calculations

Computed properties of semiconductor AB

*Impurity/Defect **M** in Semiconductor **AB***

● A
● B
● M

Defect coordination environment (Coulomb Matrix)

$$M_j = Z_X * Z_j / R_{X-j}$$

- Sum over bonds within cutoff
- Sum over N nearest neighbors

Tabulated elemental properties of defect atom

Tabulated elemental properties of atom A

Tabulated elemental properties of atom B

mchan@anl.gov

Argonne
NATIONAL LABORATORY

Correlation between predictions and descriptors

Unit cell defect calculations
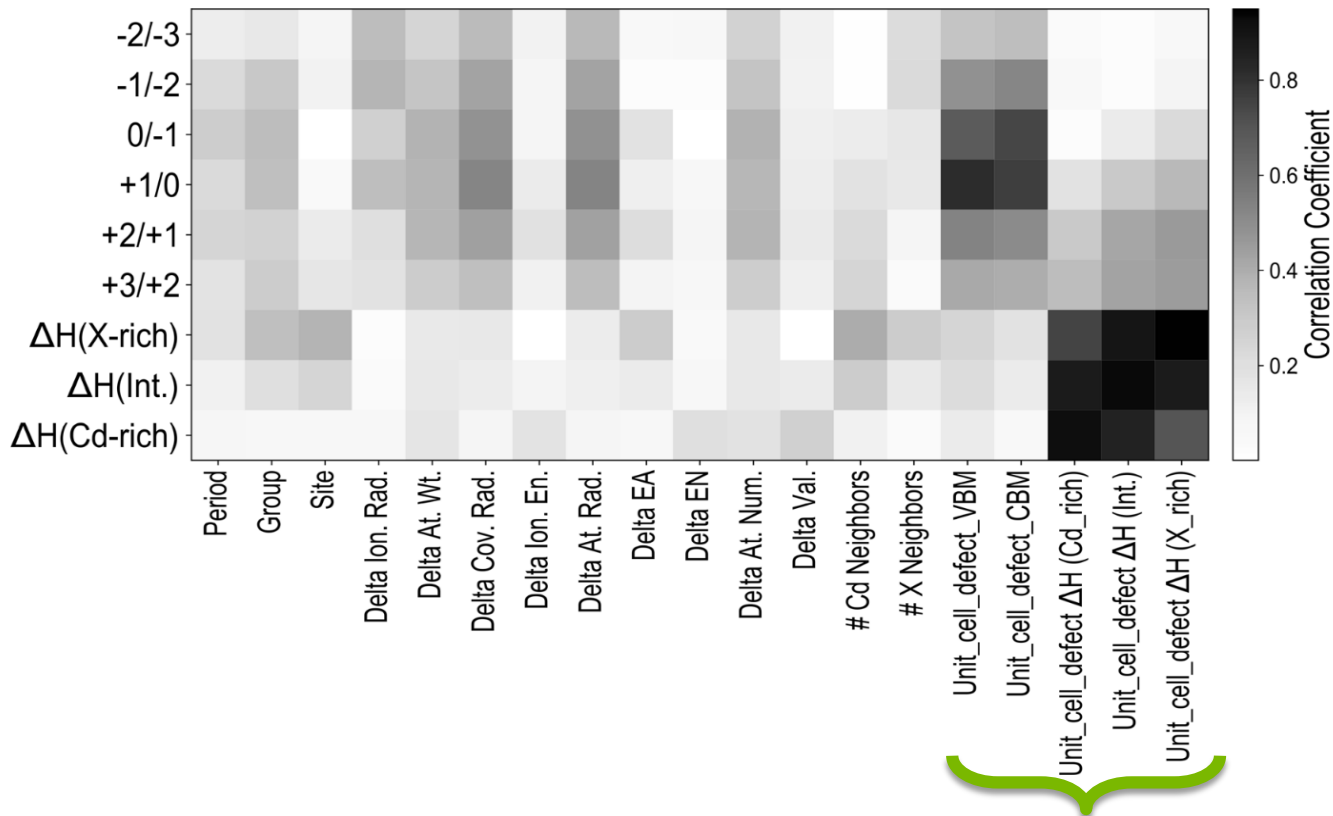
Training curves

Formation energies
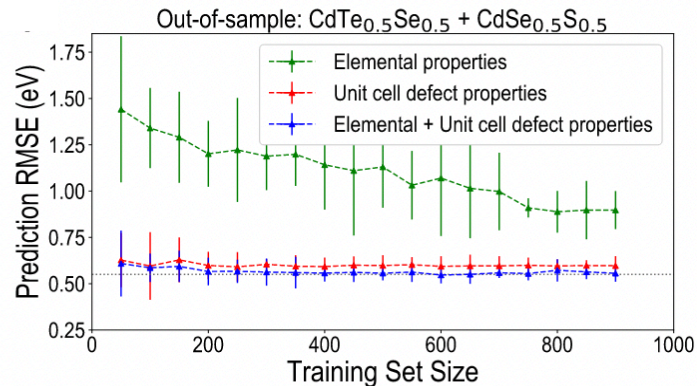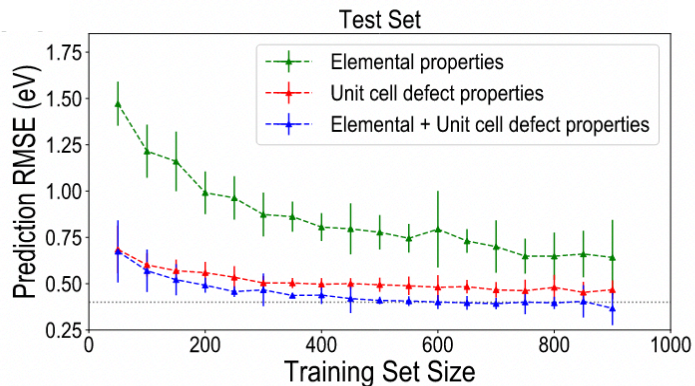
Defect levels
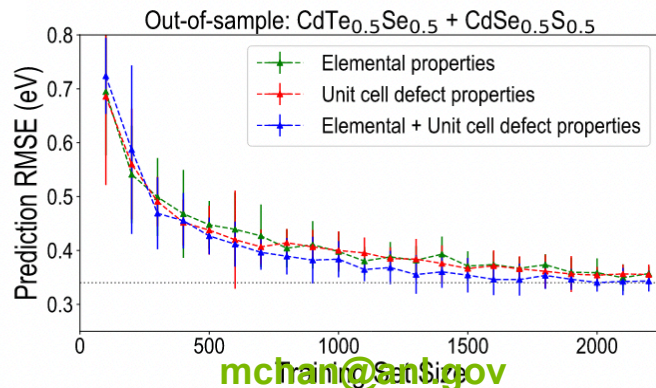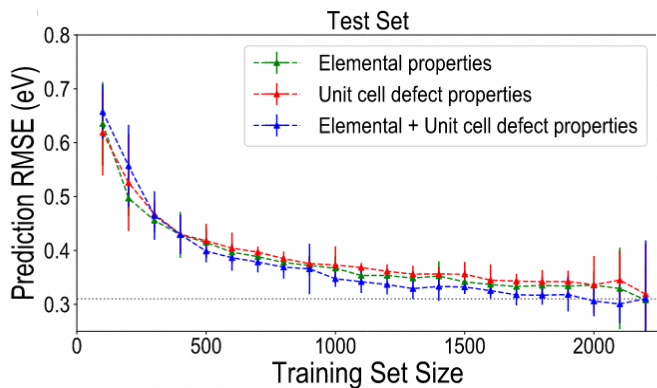
Energies (CdS/Se/Te) – Random Forest Regression

ΔH (Cd-rich) (eV)

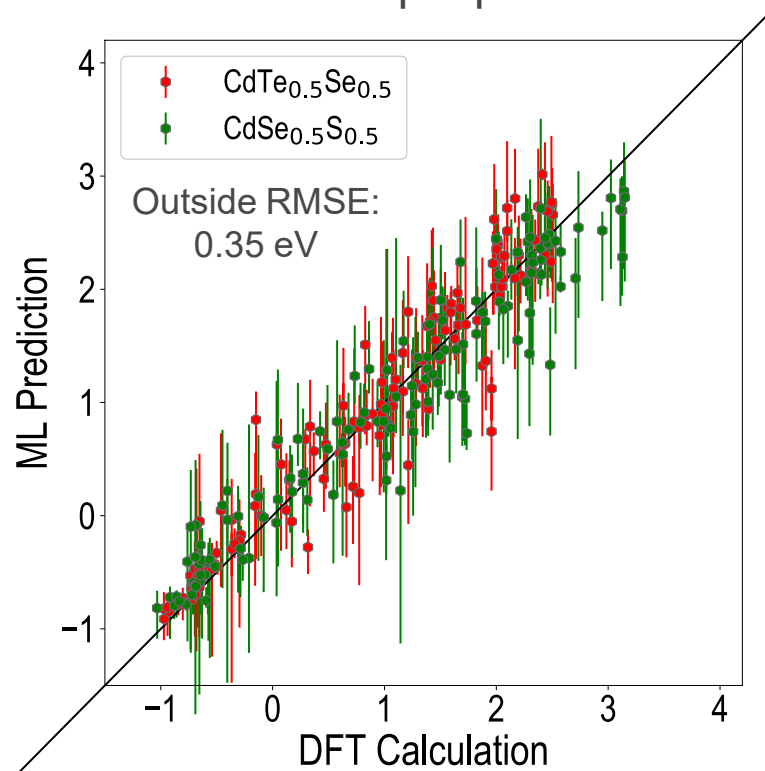ΔH (X-rich) (eV)

mchan@anl.gov

Argonne
NATIONAL LABORATORY
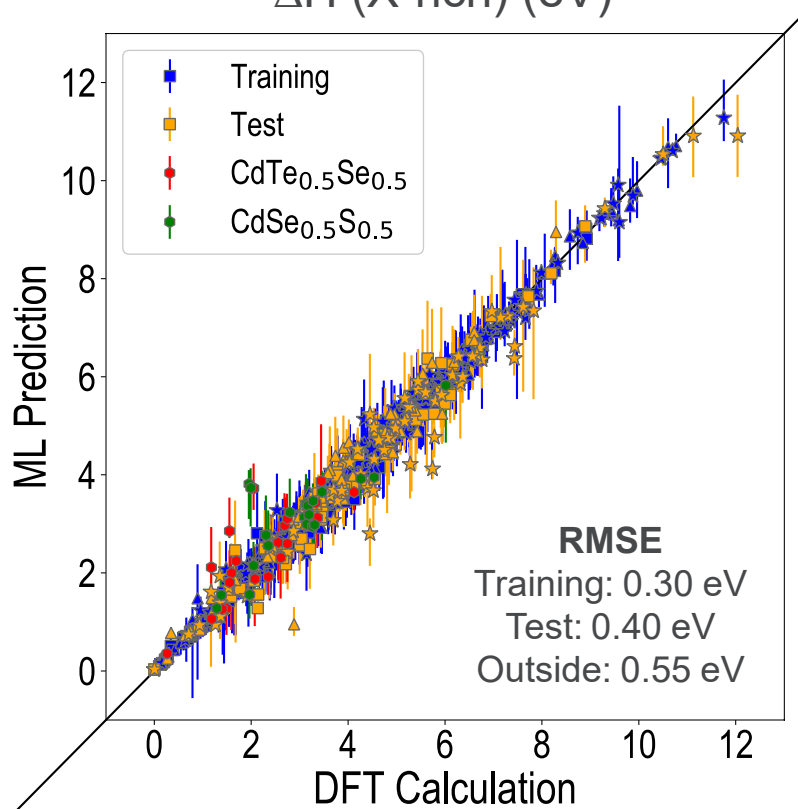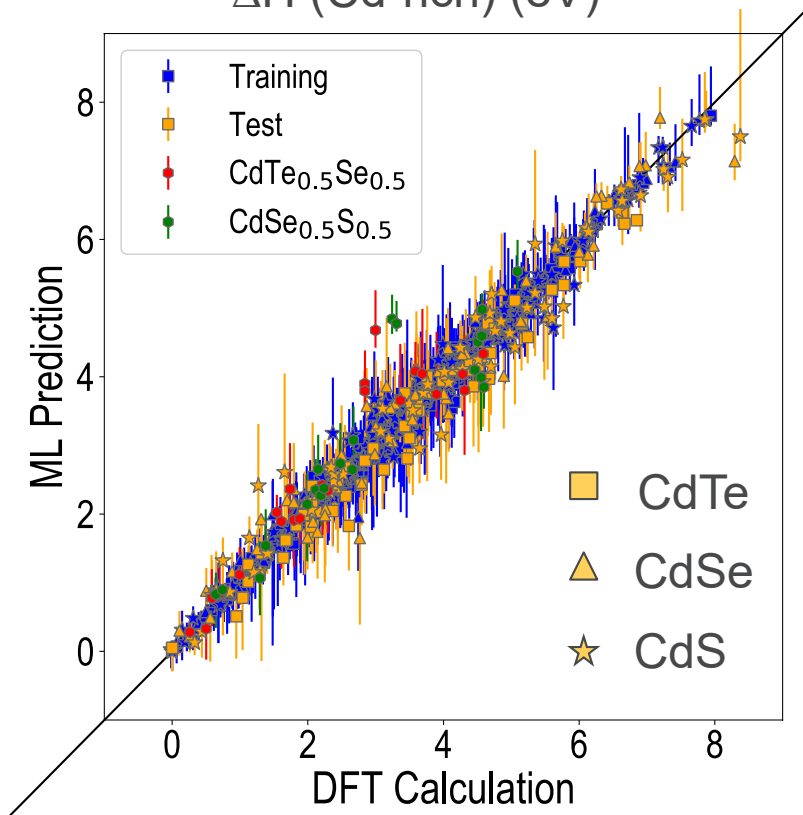
Defect Levels (CdS/Se/Te) – Random Forest Regression

RF Model to predict $\varepsilon(q_1/q_2)$ (eV)

Test RMSE: 0.30 eV

Legend:
- Training
- Test
- CdTe
- CdSe
- CdS
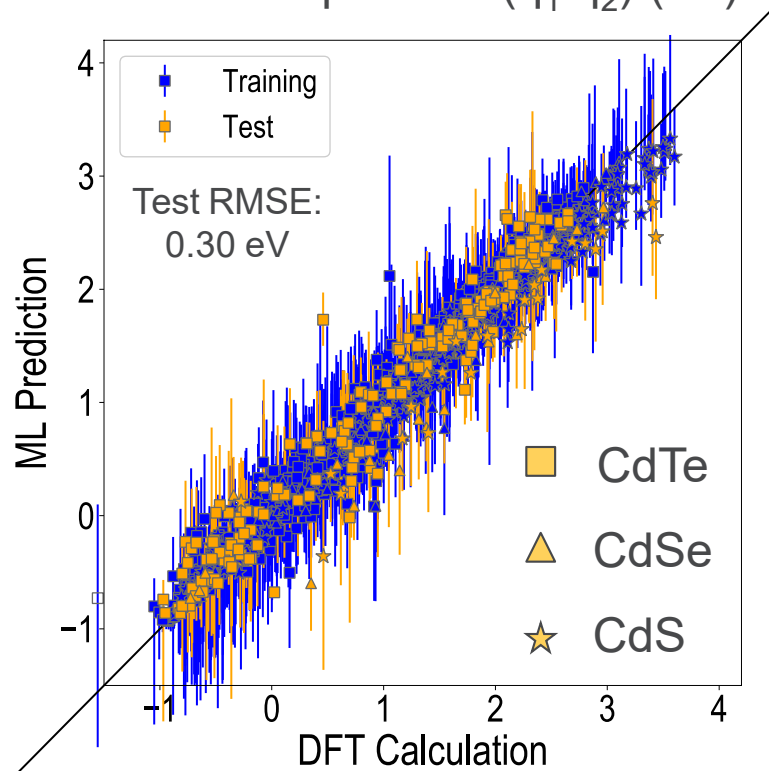
Out-of-sample prediction

Outside RMSE: 0.35 eV

Legend:
- $CdTe_{0.5}Se_{0.5}$
- $CdSe_{0.5}S_{0.5}$

mchan@anl.gov

Argonne
NATIONAL LABORATORY

Screening of dopants

(a) CdTe (b) CdSe (c) CdS (d) CdTe$_{0.5}$Se$_{0.5}$ (e) CdSe$_{0.5}$S$_{0.5}$
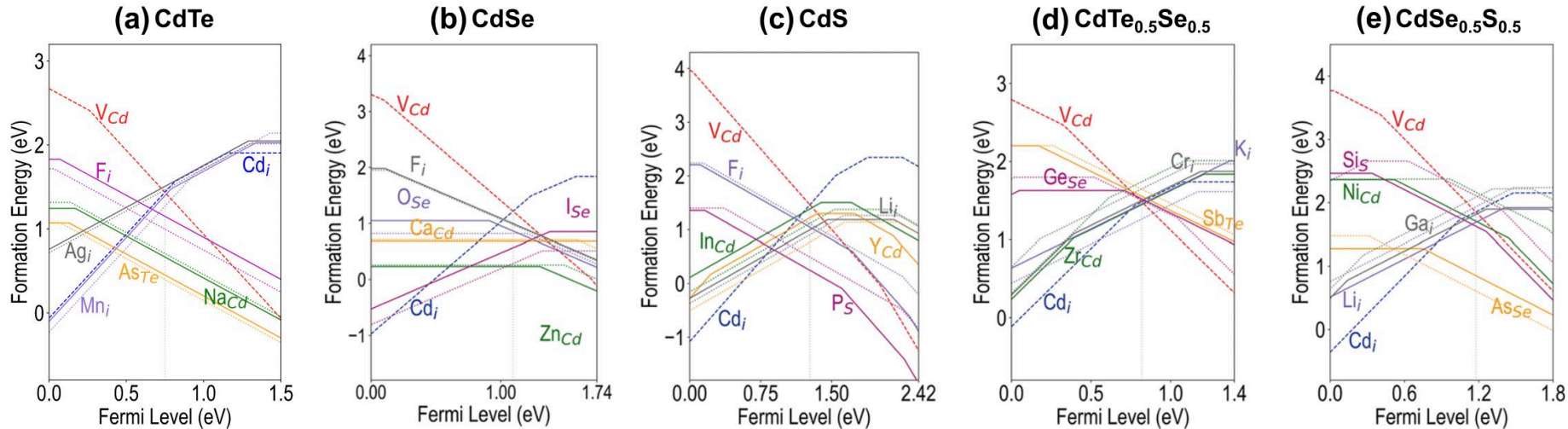
Arun Mannodi-Kanakkithodi, Michael Y. Toriyama, Fatih G. Sen, Michael J. Davis, Robert F. Klie, and Maria K. Y. Chan, "Machine-learned impurity level prediction for semiconductors: the example of Cd-based chalcogenides," npj Computational Materials 6, 39 (2020)

Screening of dopants

(a) CdTe (b) CdSe (e) CdSe$_{0.5}$S$_{0.5}$

(a) CdTe
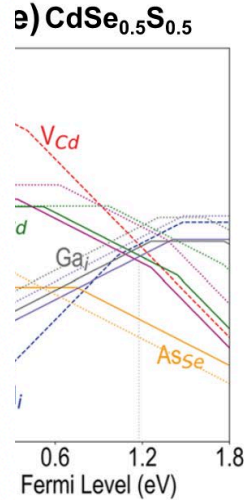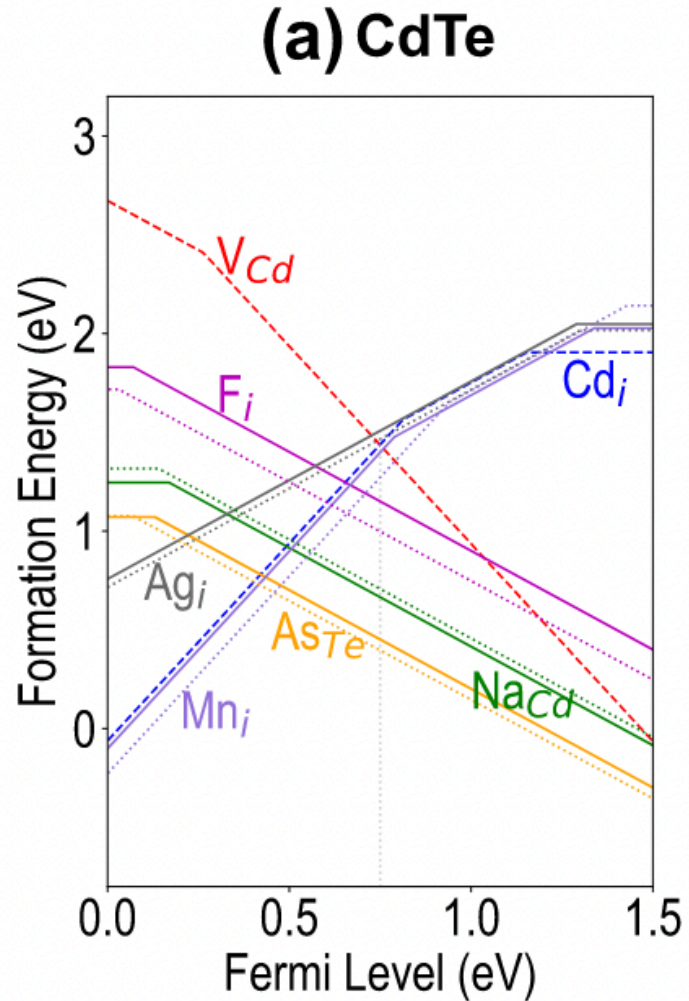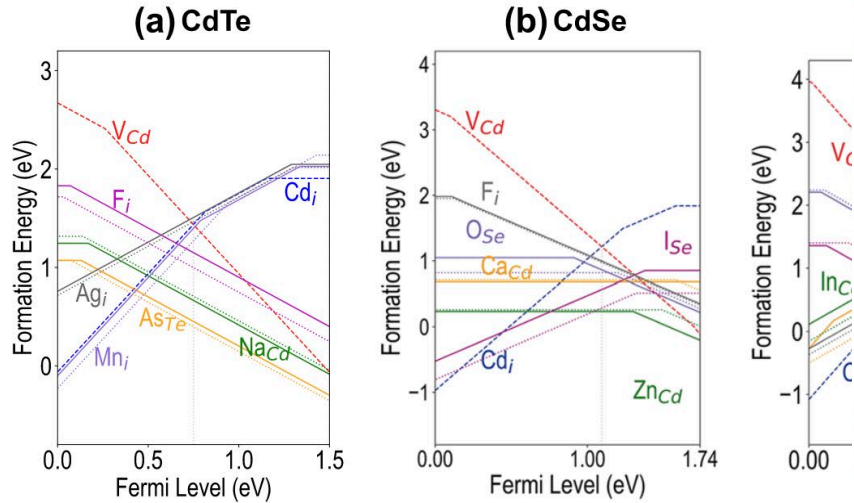
Arun Mannodi-Kanakkithodi, Michael Y. Toriyan... . Klie, and
Maria K. Y. Chan, "Machine-learned impurity lev... nple of
Cd-based chalcogenides," npj Computational M...

Screening of c

(a) CdTe

(b
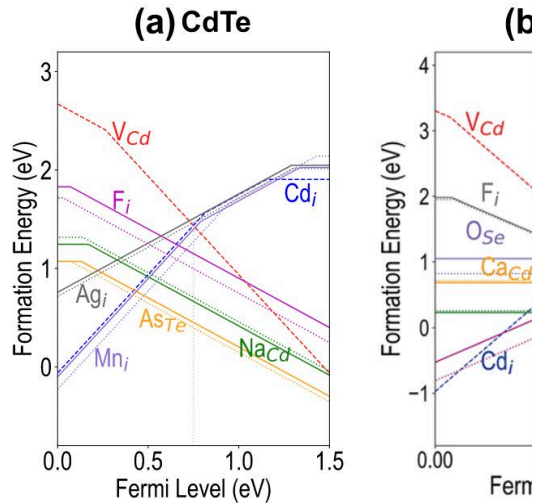
Table 5. A comparison between predictions by DFT and ML of 'dominating impurities' in CdTe.

Verdict	Cd-rich		Te-rich	
	Predicted	% of total	Predicted	% of total
False positives	5	1.59	3	0.95
False negatives	10	3.17	6	1.90
True negatives	272	86.35	275	87.30
True positives	28	8.89	31	9.84

True positives refer to the cases that were predicted to be dominating by both DFT and ML, and true negatives are the cases predicted to be non-dominating by both. False positives were predicted to be dominating by only ML whereas false negatives were predicted to be dominating by only DFT.

Extend to all zinc-blende semiconductors

Total chemical space of impurities: 77 elements * 5 defect sites (or 3) * 34 (II-VI, III-V, IV-IV) semiconductors = ~ 12,000 points.

Zinc Blende structure is used for every compound.

Semiconductors

II-VI

A	B
Cd	O
Zn	S
	Se
	Te

8 candidates

III-V

A	B
B	N
Al	P
Ga	As
In	Sb

16 candidates

IV-IV

A	B
C	C
Si	Si
Ge	Ge
Sn	Sn

10 candidates

Defect Sites

Impurity Atoms

mchan@anl.gov
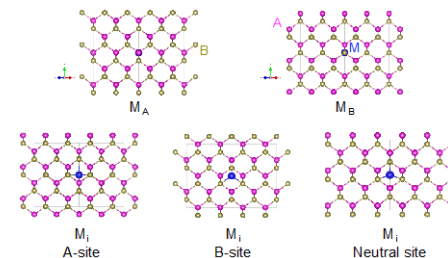
Descriptor Function and Reduction

- Descriptors show different correlations to quantities of interest

- Functional dependence on descriptors is unknown

- SISSO (sure independence screening and sparsifying operator) technique used (Ouyang et al, Phys. Rev. Mat. 2018)

- Form functions (power, exponents, etc) of descriptors, and then use LASSO to reduce dimensions

Pearson's Coefficient of 'overall' dataset

mchan@anl.gov
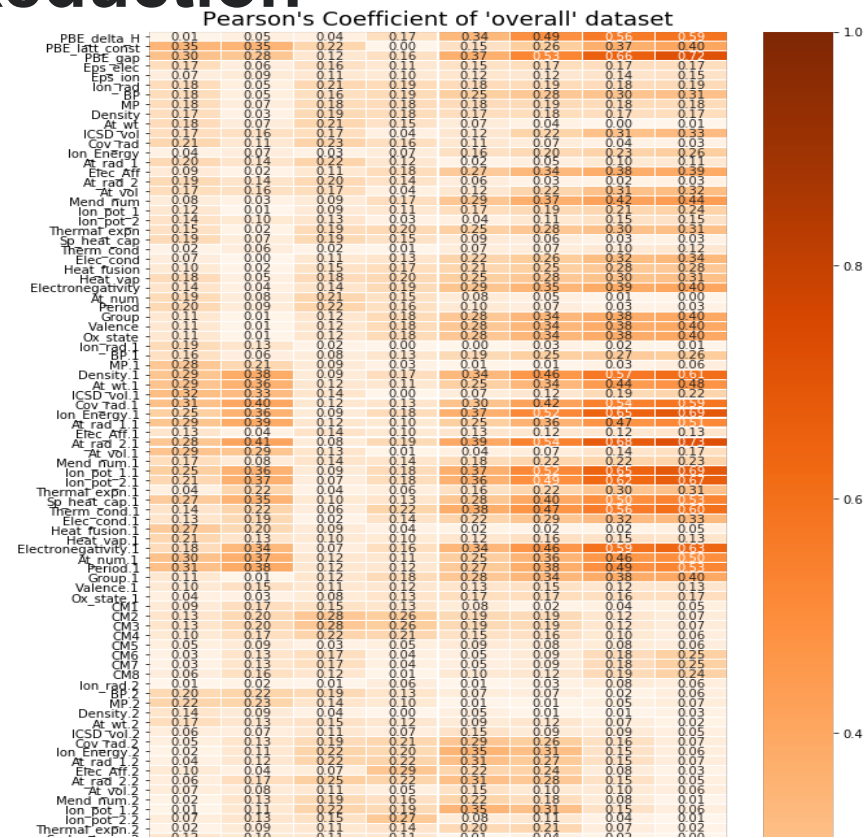
All Chemistries: Shallow Neural Networks

mchan@anl.gov
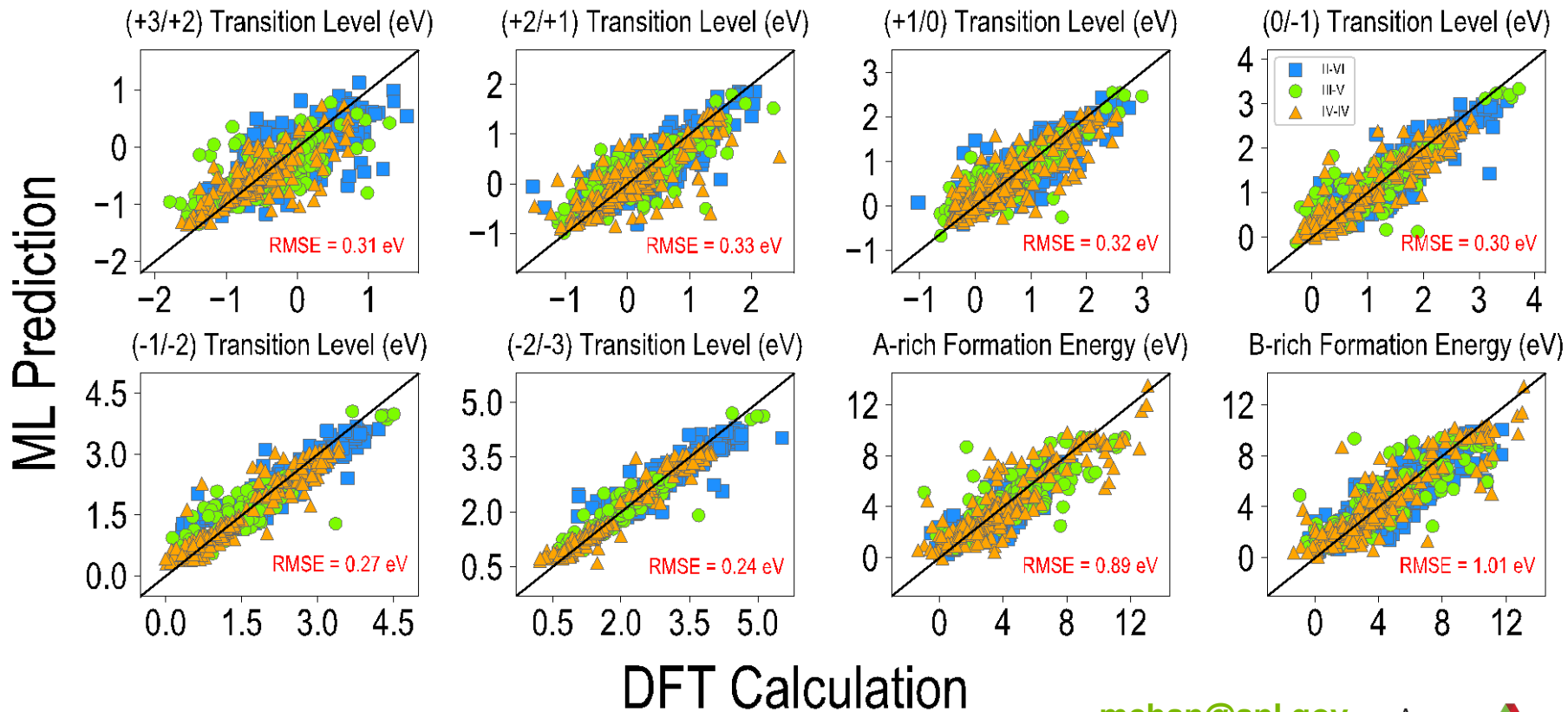
Online Tool for Prediction from ML Models
Allows rapid evaluation of different combinations
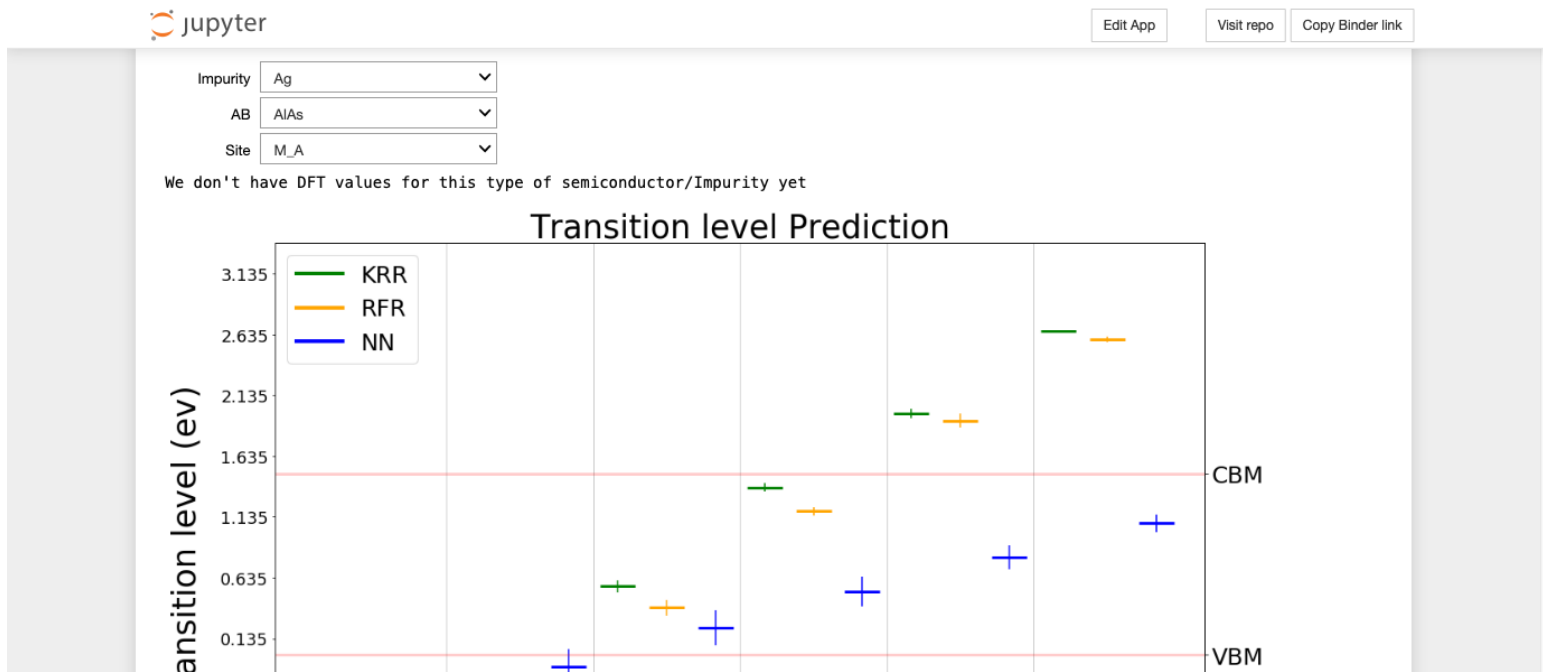
Arun Mannodi-Kanakkithodi, Xiaofeng Xiang, Laura Jacoby, Robert Biegaj, Scott Dunham, Daniel Gamelin, Maria Chan, "Universal Machine Learning Framework for Defect Predictions in Zinc Blende Semiconductors", Patterns 3, 100450 (2022).

Conclusion: DFT+ML for Defects

- DFT gives a reasonable prediction of defect levels in CdTe.
- ML models trained on DFT calculations used to predict formation energies and charge transition levels of impurities in Cd(Te,Se,S) and later to all zinc blende semiconductors.
- Such models can be used to screen dopants
- Expensive+more accurate DFT calculations can be reasonably approximated by cheap DFT calculations + ML.

(d)

mchan@anl.gov
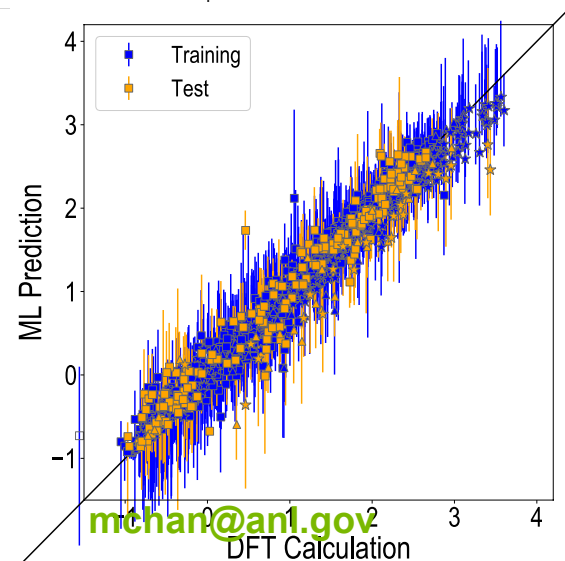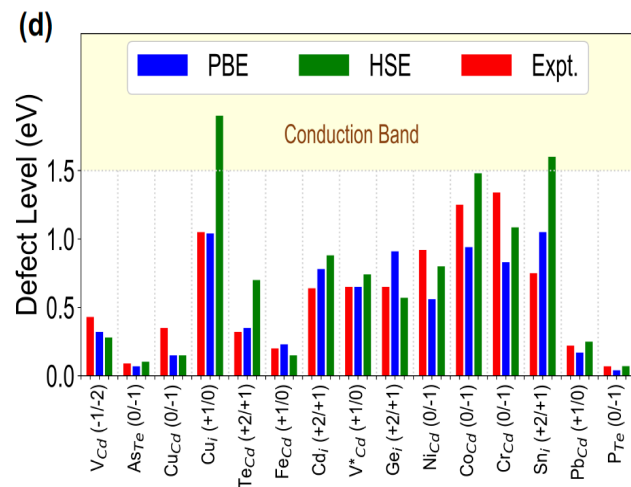
Part II: AI/ML+modeling guided interpretation of microscopy & spectroscopy data

Understanding grain boundaries in CdTe

Moon Kim
Univ.
Texas
Dallas
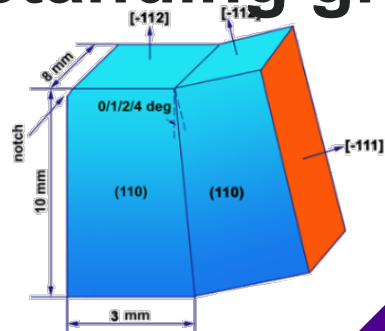
Bicrystal fabrication

Robert Klie
Univ. Illinois
Chicago

Overview
SunShot
PVRD
CdTe
GBs

Opto-electrical measurements

STEM imaging
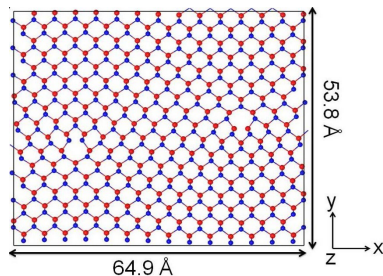
DFT modeling

3d model

Maria Chan
Argonne Natl Lab

2d image
mchan@anl.gov

Argonne
NATIONAL LABORATORY
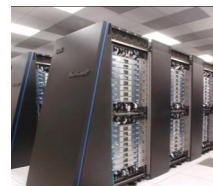
Inversion of characterization data

Structure:
Bulk crystals
Interfaces
Surfaces
Nanostructures
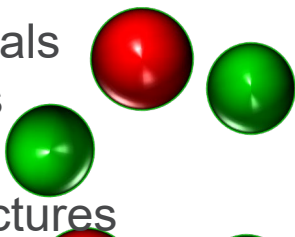
Electron microscopy

signal or 2d images

**??? **

Synchrotron x-ray

Atom probe
Laser
etc.
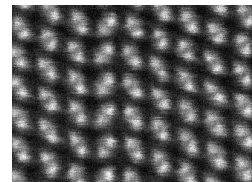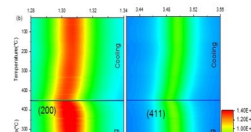
← **Invert characterization data to get structures (with theoretical modeling as guidance/constraints)**

7105 7115 7125 7135
Energy (eV)
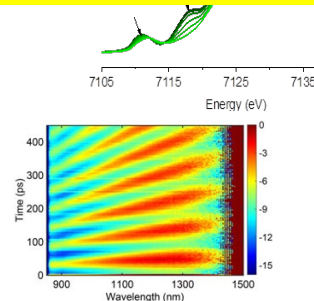
INGRAINED – simulated and experimental STEM/TEM/STM image matching

- **Have:** microscopy image**; Want:** matching simulated image
- **Scope**: electron microscopy, scanning probe microscopy
- **Challenge**: image varies with simulation parameters
- **Approach**: use computer vision measure for image
  similarity and automated searching

- Eric Schwenker
- Chaitanya Kolluru

Ingrained – STEM results

Examples of initial structures generated

Optimal correspondence between boundaries of bulk segments. FANTASTX will optimize structure of the interface!

CdTe (100)-(110)

NMC battery materials

1 nm
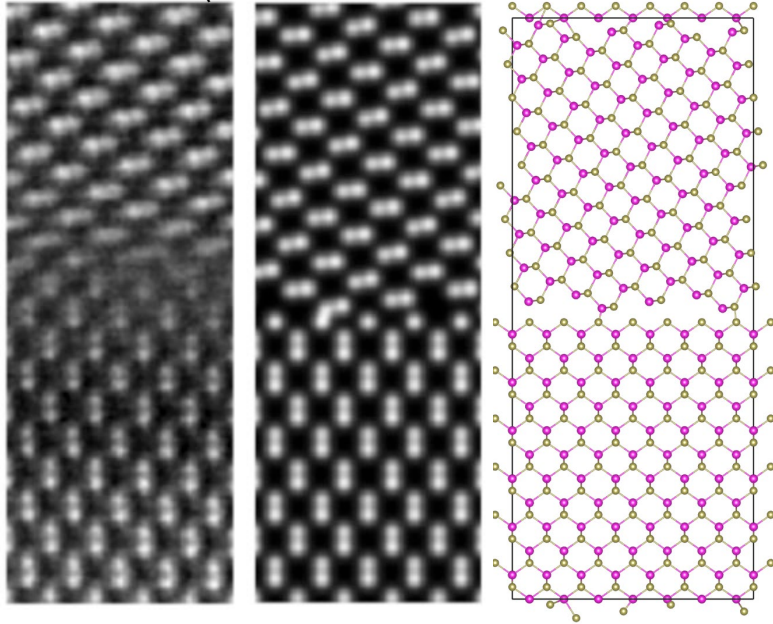
1-VIFP: 0.799

Appl. Phys. Lett. **115**, 153901 (2019)
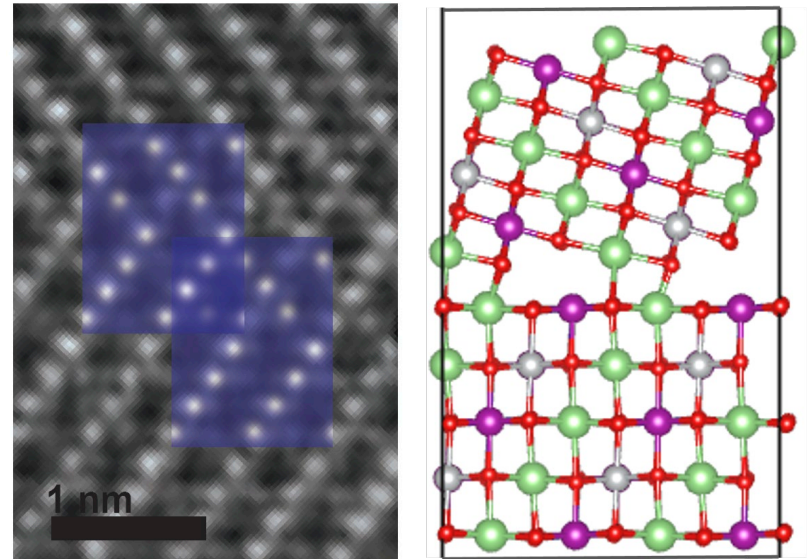
Nature Energy, 2022

mchan@anl.gov

Argonne
NATIONAL LABORATORY

FANTASTX – characterization data inversion through structure search
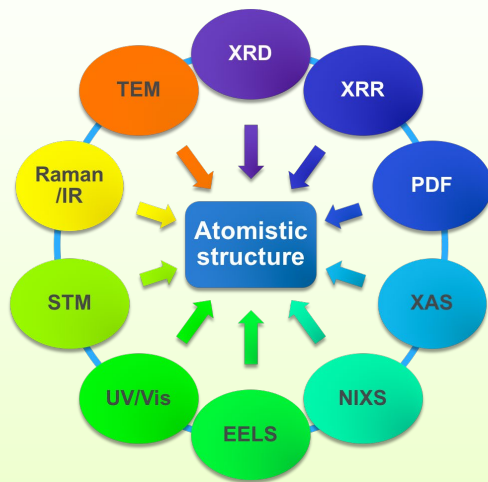
- **Have:** characterization data**; Want:** nanoscale atomistic structure
- **Scope**: X-ray, electron microscopy, scanning probe
- **Challenge**: non-unique mapping, underconstrained space
- **Approach**: use physics (Hamiltonian) to constrain solutions

FANTASTX – Fully Automated Nanoscale To Atomistic Structure from Theory and eXperiment

- Chaitanya Kolluru
- Davis Unruh
- Zisheng Zhang
- Previous: Spencer Hills, Eric Schwenker, Fatih Sen

24

mchan@anl.gov

FANTASTX
Fully Automated Nanoscale To Atomistic Structure from Theory & eXperiment

Computer vision image comparison

STEM | Simulated

1 nm

$H\psi = E\psi$

Match to experiments
S/TEM
XRD/PDF/XR
XAS/EELS

Energies
DFT
Interatomic potentials

Potential parameters fit to DFT data

(a) $\delta = 0.07 \pm 0.04$ eV/at

HyBOP Energy (eV/at)

DFT Energy (eV/at)

Globular
Planar

Genetic algorithms
Basin hopping
Monte Carlo
Iterative sampling
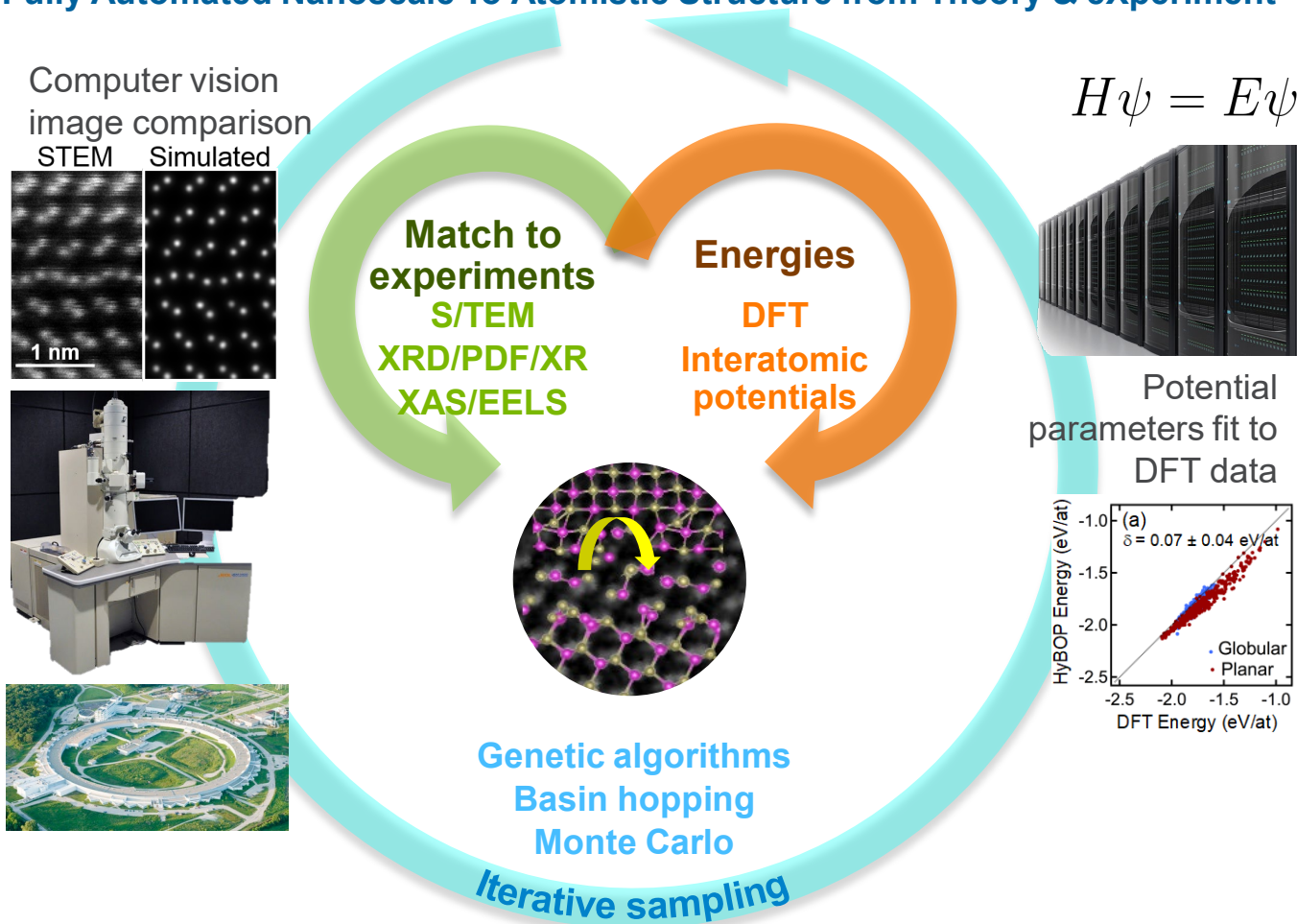
Structural models allow materials design

STEM image
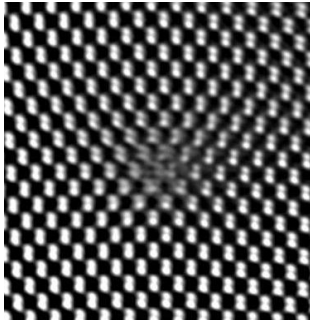
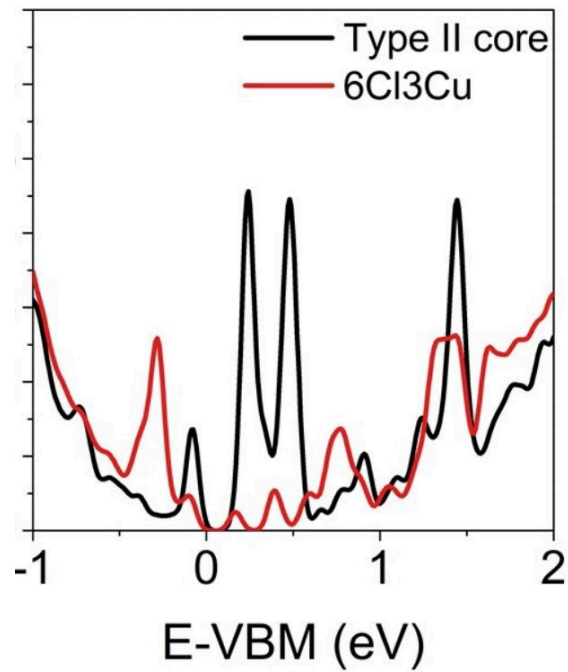**Dislocation core model**
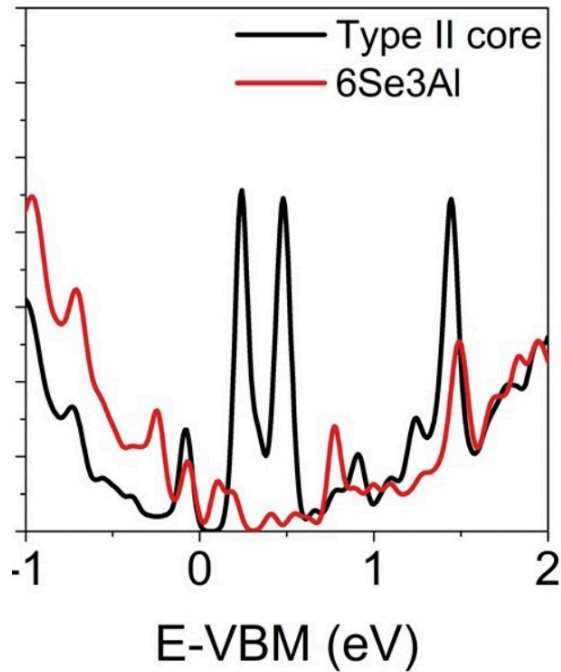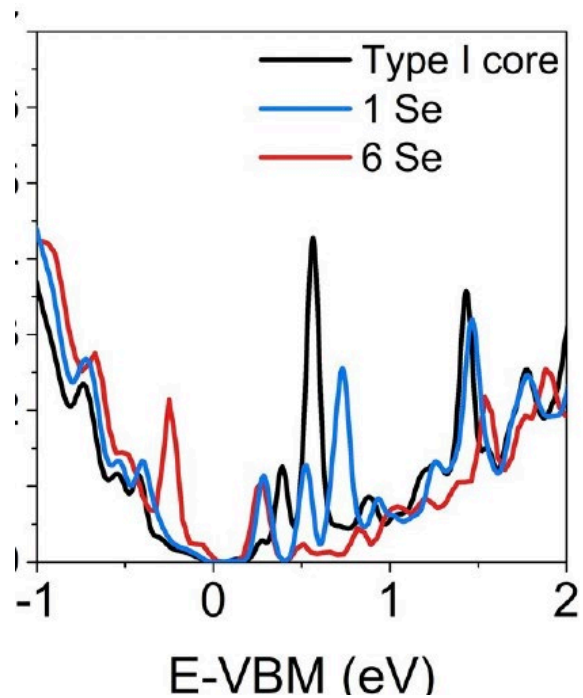
Electronic structure (DFT)

Two types of dislocation cores from model
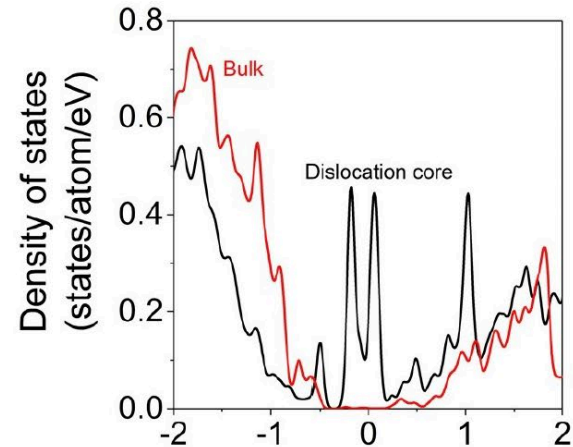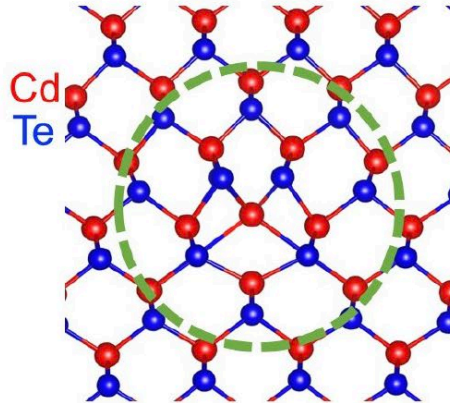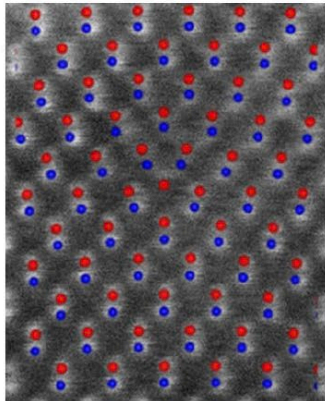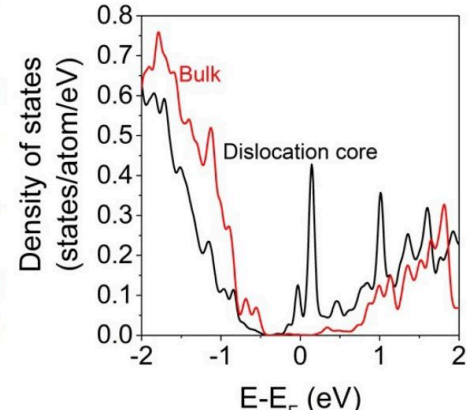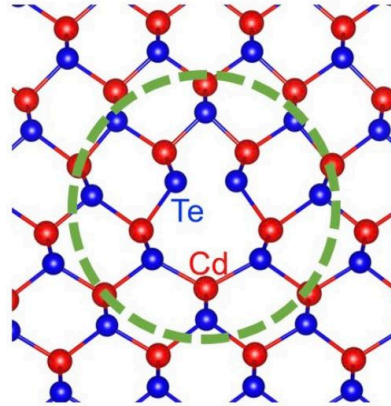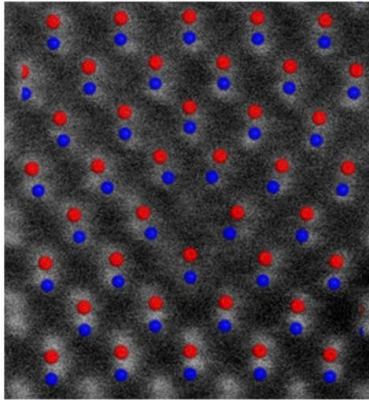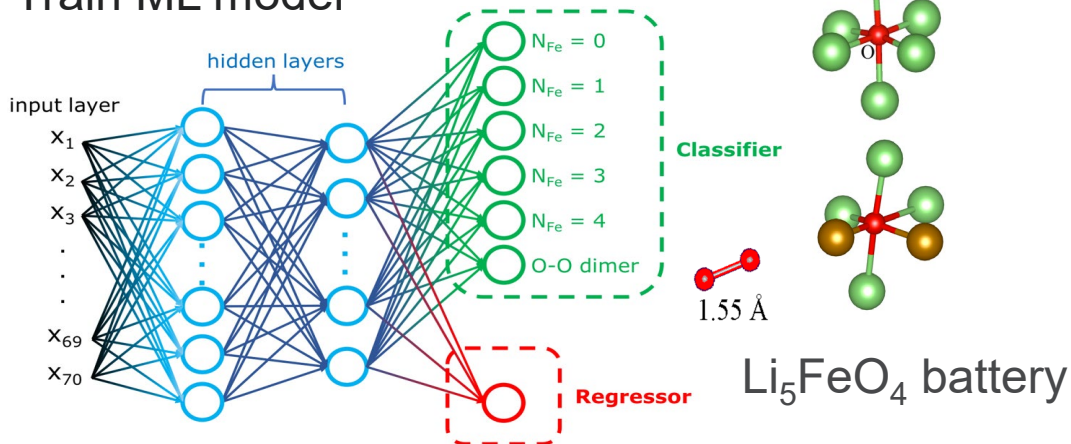
Different origin of mid-gap states
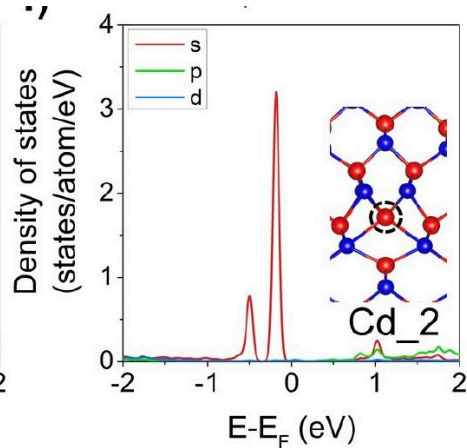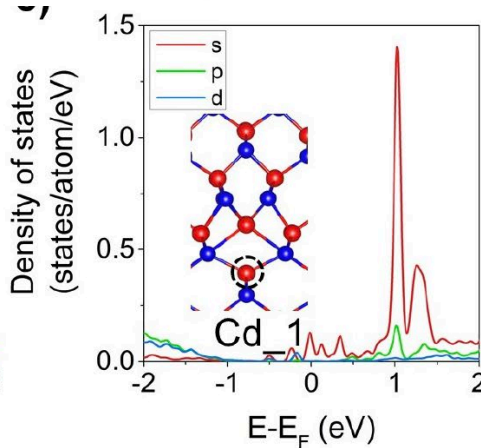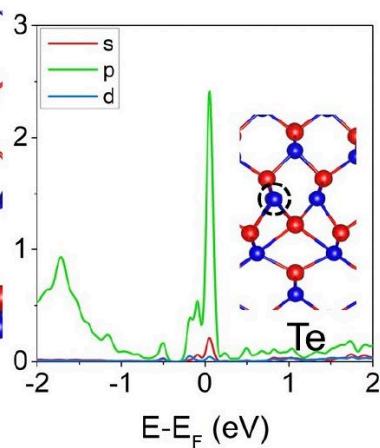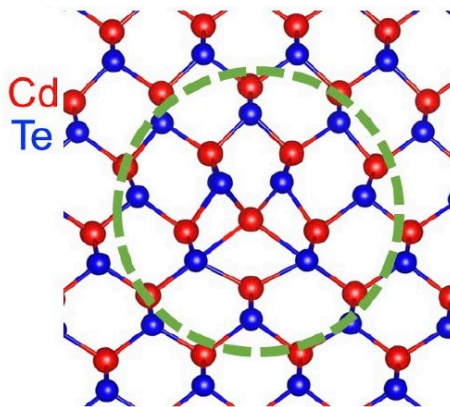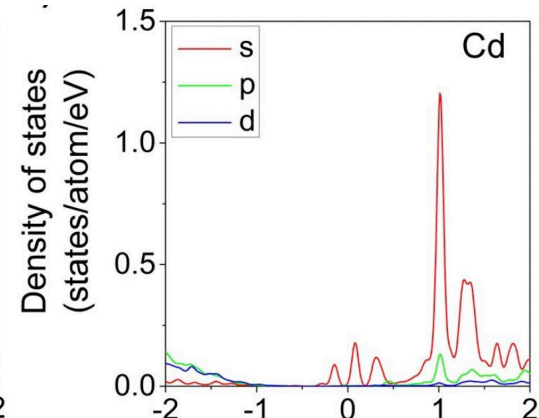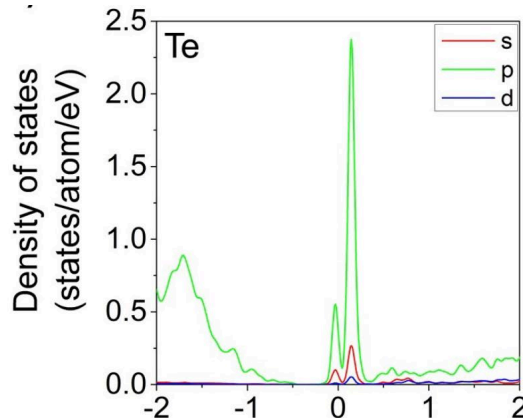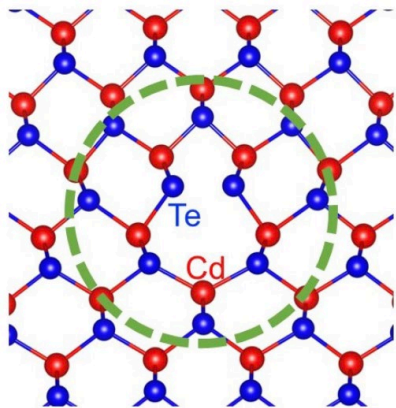
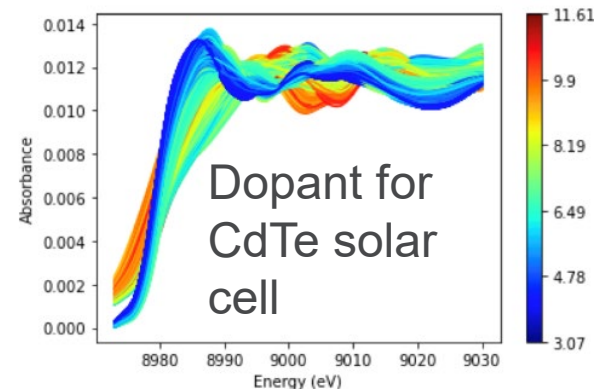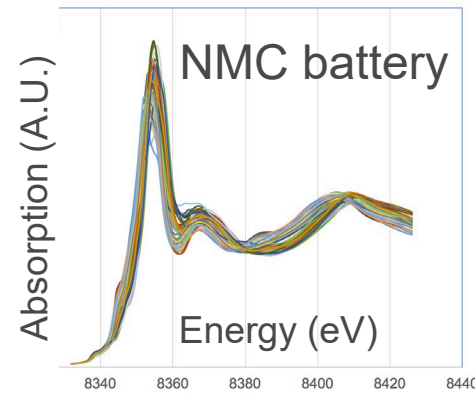# Different passivation/co-passivation

X-ray absorption/emission spectra

- **Goal:** determine structural/electronic descriptors from spectra, specifically coordination number

- **Approach:**
  - compute different configurations using DFT
  - calculate spectra using different first principles- and multiplet-based approaches and codes
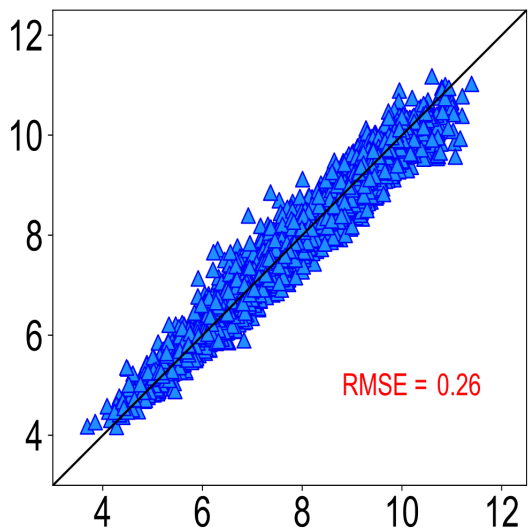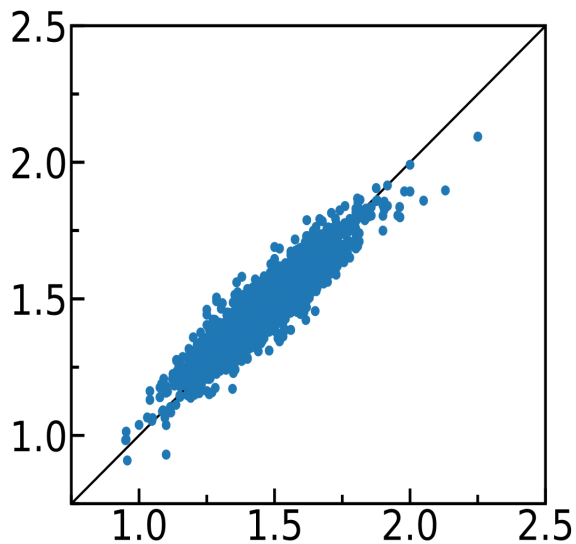 - Train ML model

X-ray absorption spectra

NMC battery

Dopant for CdTe solar cell

Mariana Bertoni

Li_5FeO_4 battery

Liang Li, Arun Mannodi Kannakithodi, Justin Pothoof, Yiming Chen   **mchan@anl.gov**
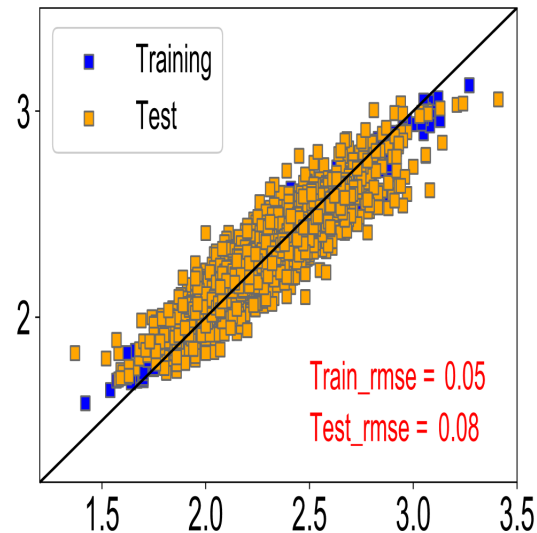
Argonne NATIONAL LABORATORY

Predicted vs ground truth coordination numbers

Cu-Te for CdTe solar cell
# of neighbors of Cu
FEFF simulations
Neural network

Li_5FeO_4 battery
# of Fe neighbors of O
OCEAN simulations
Neural network

NMC battery
# of Li neighbors of Ni
FEFF simulations
Random forest

mchan@anl.gov

Conclusion – AI/ML+modeling guided interpretation of microscopy & spectroscopy data

- Need atomic structures to enable understanding/optimization of point and extended defects, processing, interfaces
- Atomic structures also allow design of passivants
- Used combination of modeling, AI/ML, and detailed characterization to determine structures
- Developed general purpose software codes

mchan@anl.gov

MaterialEyes
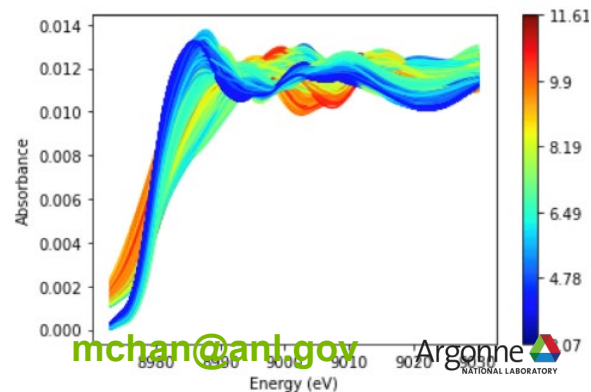
https://github.com/MaterialEyes

Software tools
- **exsclaim** – collect images from literature (Trevor Spreadbury, Eric Schwenker, Weixin Jiang)
- **fantastx** – determine nanoscale structures from theory and experiments (Chaitanya Kolluru, Spencer Hills, Eric Schwenker)
- **ingrained** – match simulated & experimental TEM/STM images (Eric Schwenker, Chaitanya Kolluru)
- **manipulatt** – manipulate lattice structures (Joydeep Munshi)

Dataset
- **atomagined** – Simulated STEM dataset (Eric Schwenker)
https://doi.org/10.18126/szeq-yde5

mchan@anl.gov

Acknowledgements

mchan@anl.gov

Argonne
NATIONAL LABORATORY

The DOE Nanoscale Science Research Centers:
User Facilities for Creating, Characterizing, and Understanding Nanomaterials

About the Center for Nanoscale Materials:

- Free access to staff expertise and equipment (if intent is to publish)

- Three annual proposal calls; short-term projects are accepted continuously

- Simple 2-page proposal

- Co-located with Advanced Photon Source and Argonne Leadership Computing Facility

Center for Nanoscale Materials Capabilities:

- **Synthesis –** 2D materials, biofunctional structures, diamond/CNT, colloidal chemistry, self-assembly, quantum and energy materials

- **Nanofabrication –** 11,000 sq. ft. cleanroom, direct-write fabrication, nonlinear phenomena, nanoscale interaction manipulation, device integration, NEMS/MEMS, flat lens optics

- **Advanced Microscopy –** synchrotron hard x-ray nanoprobe, ptychography, AFM/STM (LT, SP, UHV, optical, synchrotron x-ray), aberration-corrected and in-situ TEM/STEM

- **Nanophotonics –** ultrafast transient absorption and emission spectroscopy and microscopy; UV-to-THz

- **Theory/Modeling –** electronic structure calculations, atomistic studies, electrodynamics, multiscale approaches, machine learning

www.anl.gov/cnm

mchan@anl.gov