

Rational Computational Design of Next-Generation Semiconductors

Arun Mannodi Kanakkithodi

Assistant Professor, Materials Engineering
Purdue University, West Lafayette, Indiana, USA

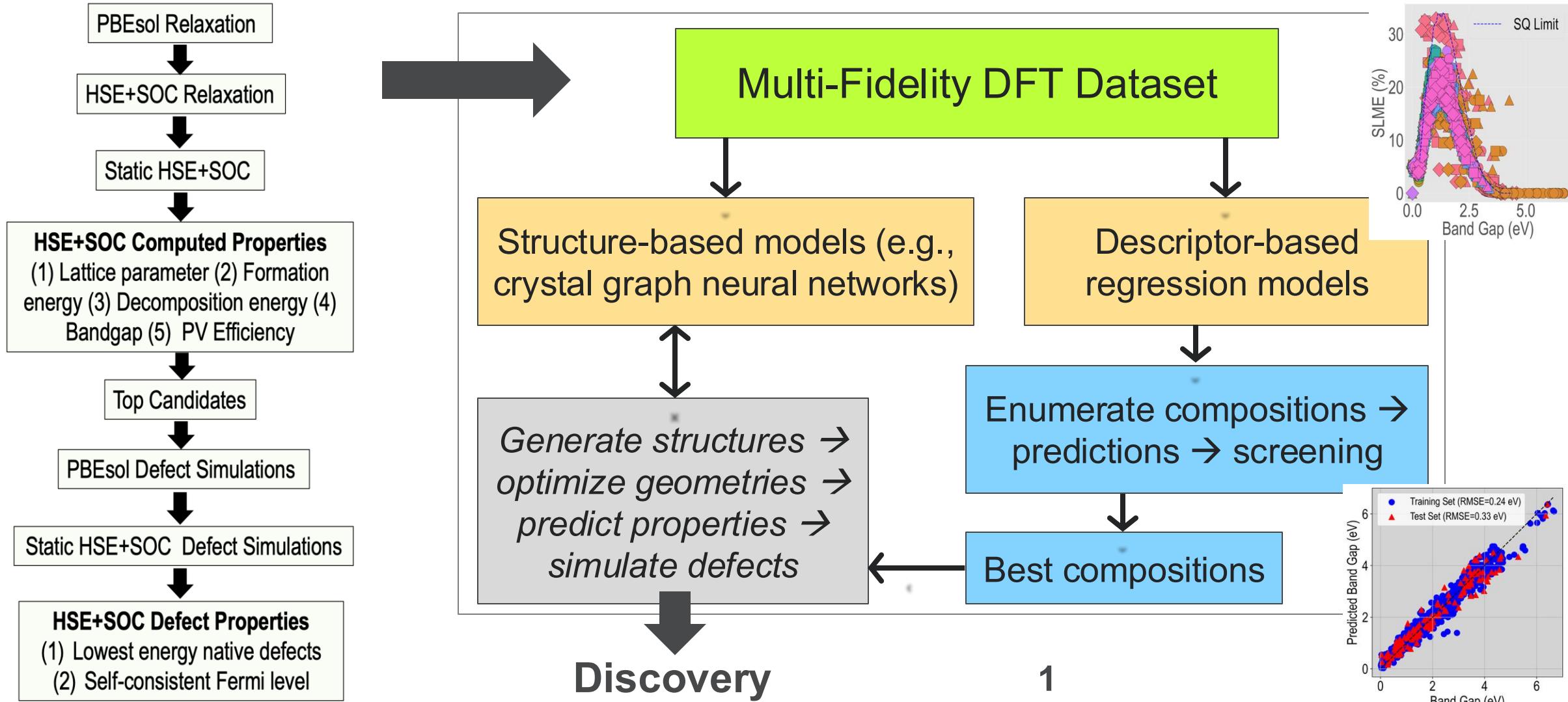
NIST AIMS Workshop
Wednesday, July 9, 2025



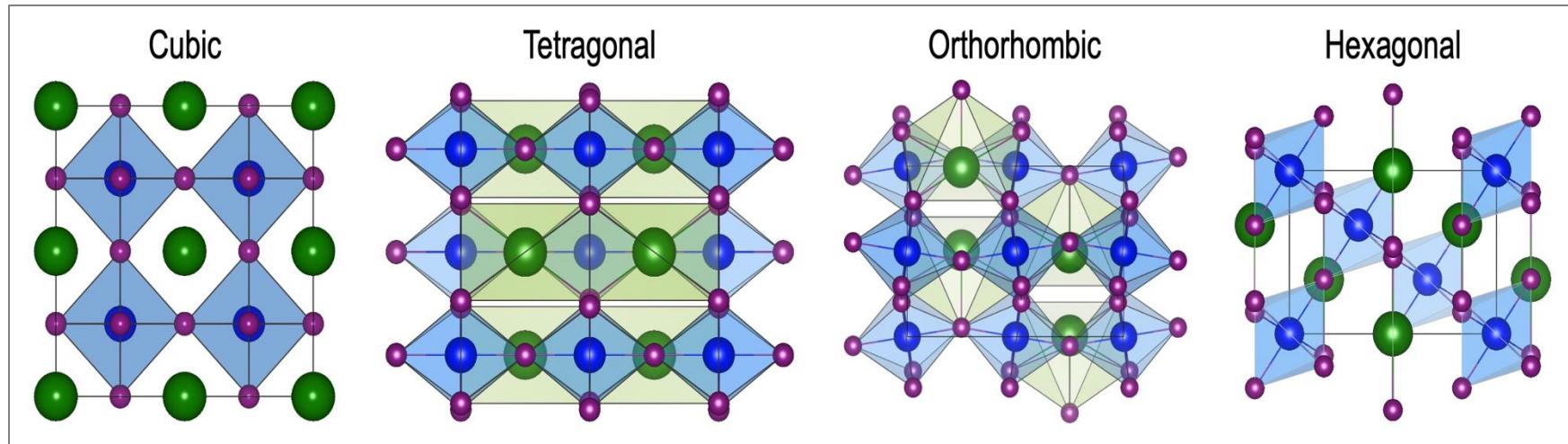
PURDUE
UNIVERSITY®

School of Materials Engineering

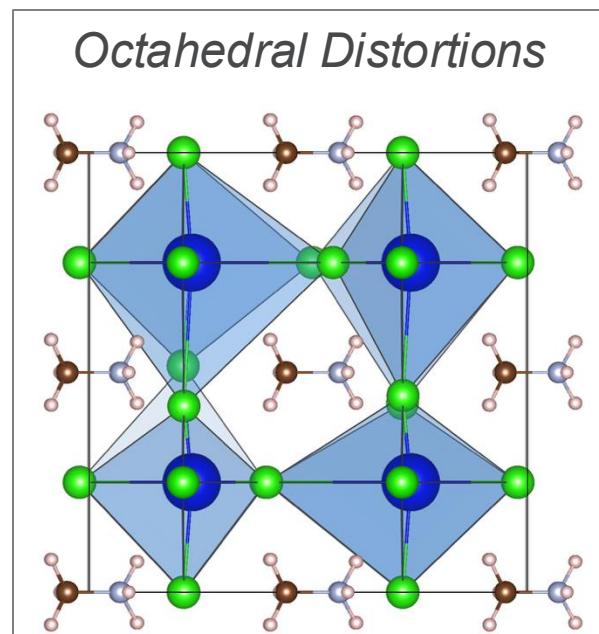
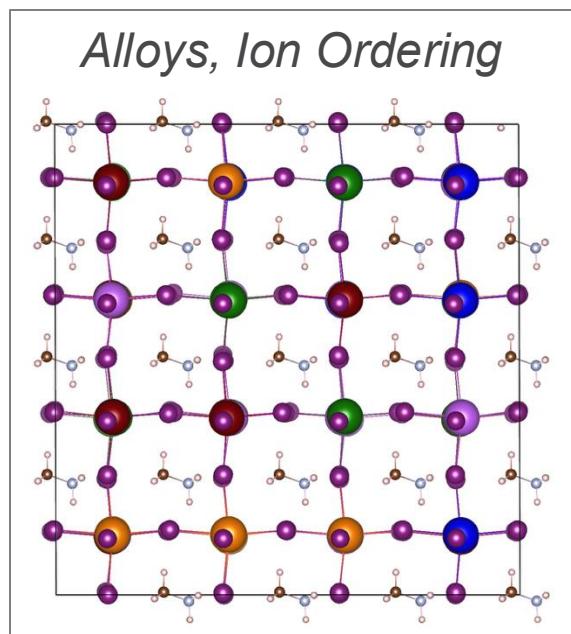
DFT+ML workflow for designing new materials with tailored electronic and defect properties



ABX_3 Halide and Chalcogenide Perovskites

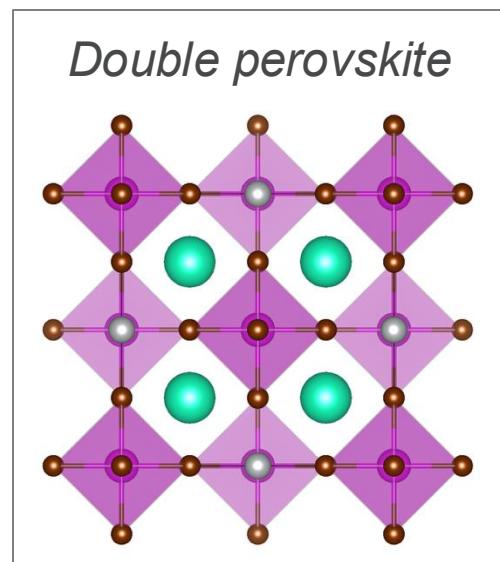


Halide perovskites		
A	B	X
MA	Pb	I
FA	Sn	Br
Cs	Ge	Cl
Rb	Ba	
K	Sr	
		Ca



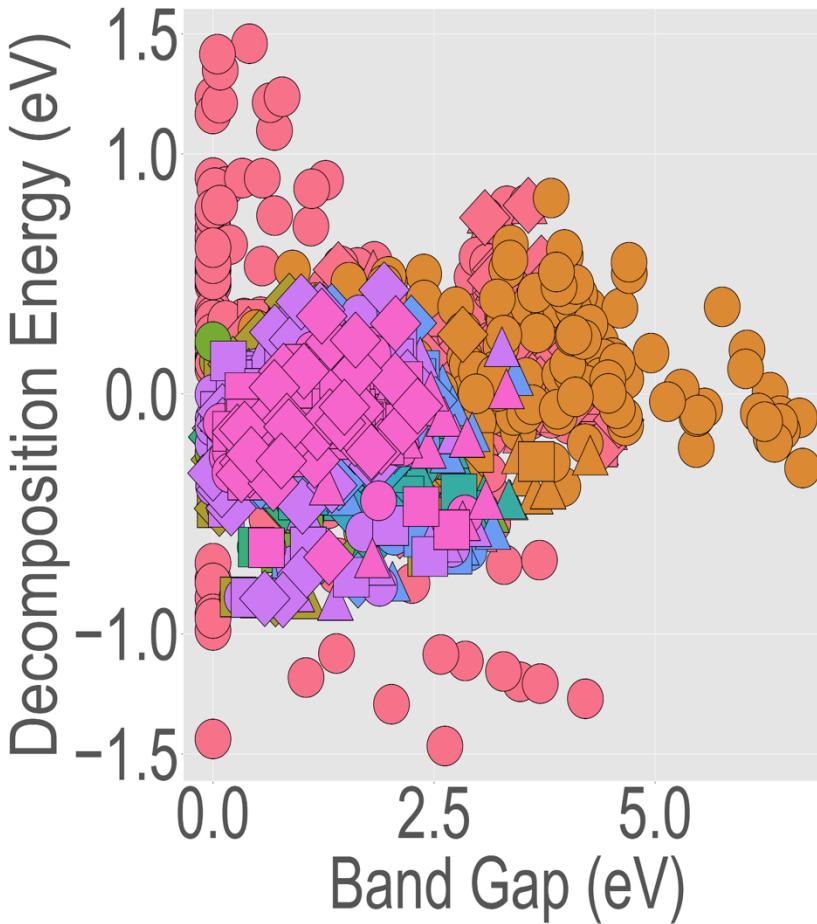
Chalc. perovskites

A	B	X
Ba	Ti	S
Sr	Zr	Se
Ca	Hf	Te
	Sn	
	Ge	
La	Al	S
Y	Sc	Se
Ce	Sb	Te

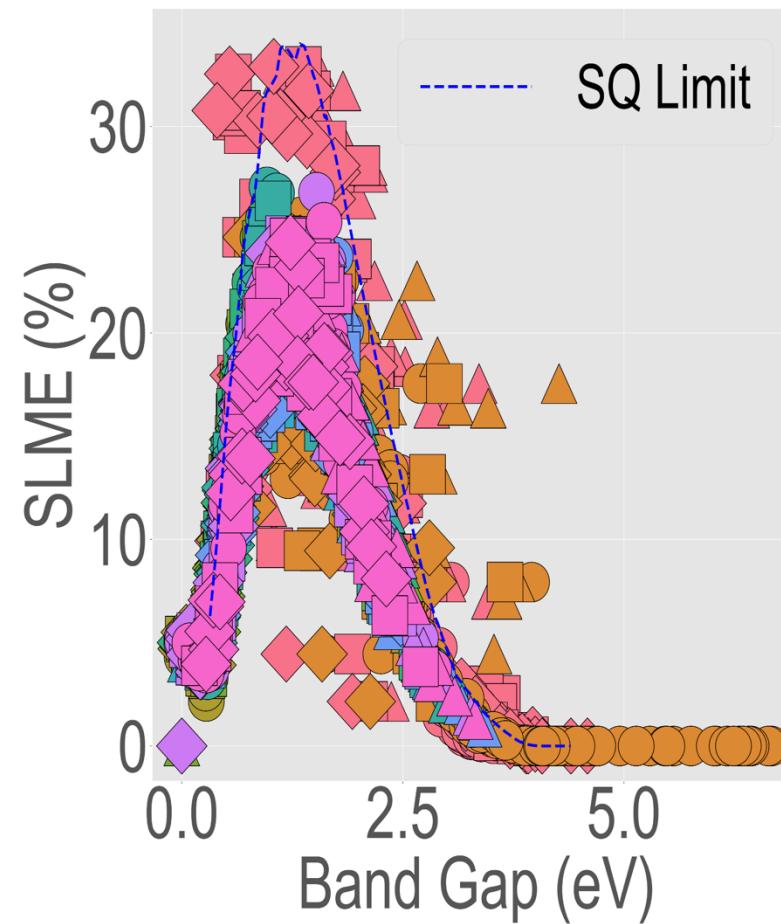


Our Perovskite DFT Dataset

- Cubic (49.3%)
- ▲ Hexagonal (12.4%)



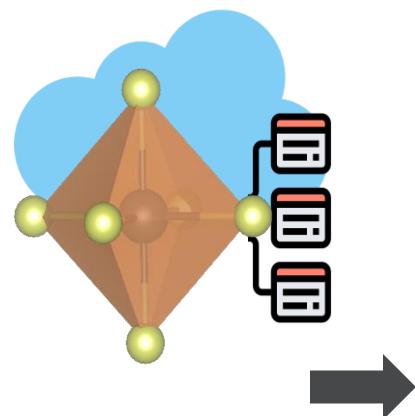
- Tetragonal (14.3%)
- ◆ Orthorhombic (15.3%)



~3000 Data Points

- PBE (44.5%)
- HSE06-PBE+SOC (13.2%)
- PBESol+D3+SOC (5.8%)
- PBE+D3+SOC (5.5%)
- PBESol+SOC (5.9%)
- PBE+SOC (5.5%)
- HSE06-PBESol+SOC (2.0%)
- PBE+D3 (5.9%)
- PBESol+D3 (5.8%)
- PBESol (5.9%)

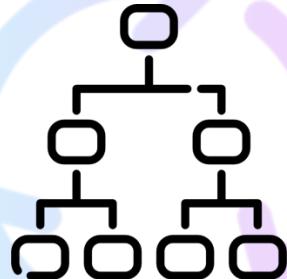
Composition-Based ML Models



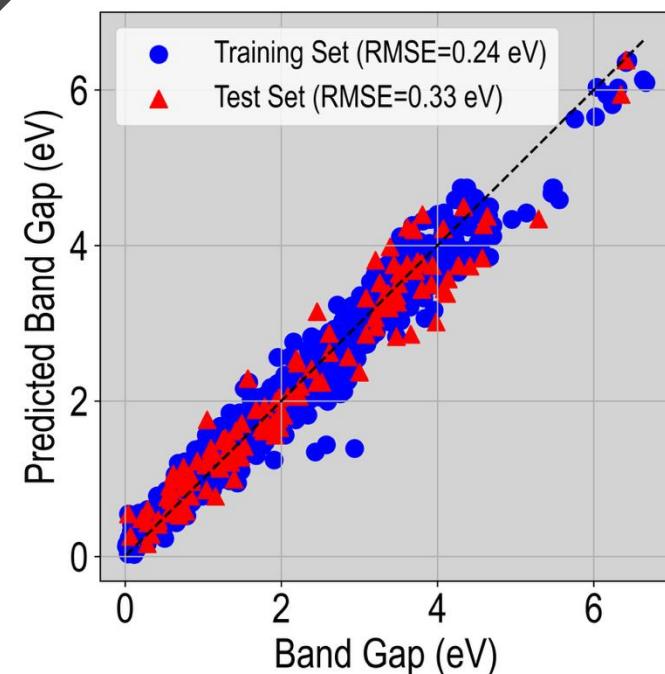
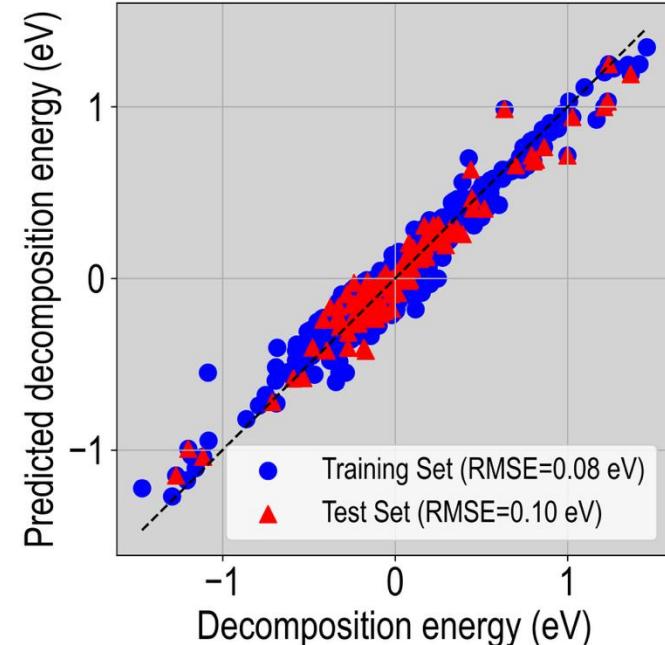
Properties and Descriptors

Band Gap (eV)
Decomposition Energy (eV)

Composition Vector
Elemental Properties
One-hot encode phase
One-hot encode theory

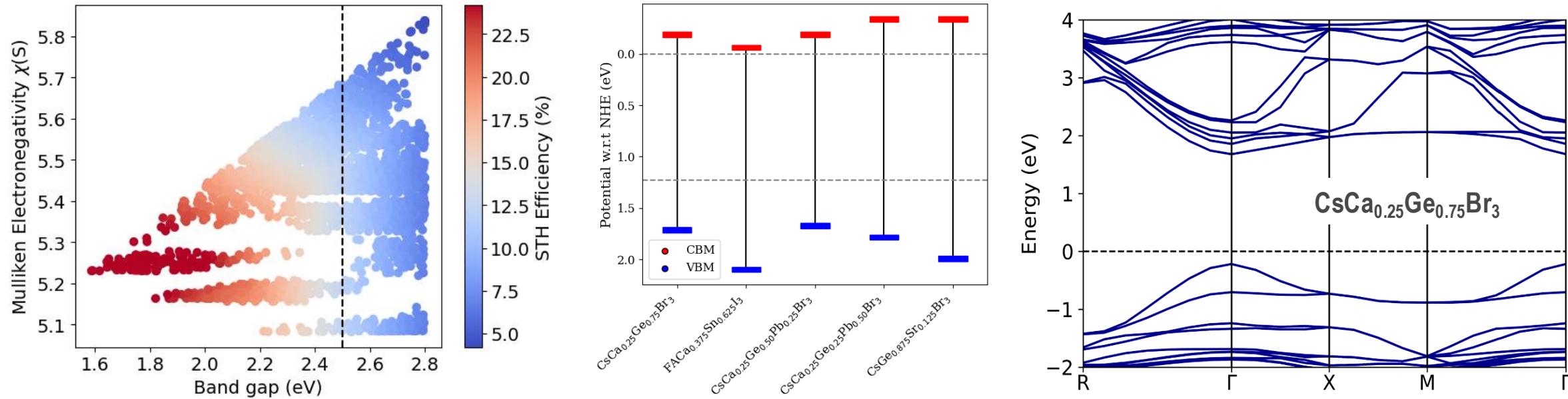


Multi-fidelity
Regularized
Greedy Forest
Regression



- Biswas, M.; Desai, R.; Mannodi-Kanakkithodi, A. Screening of Novel Halide Perovskites for Photocatalytic Water Splitting Using Multi-Fidelity Machine Learning. *Phys. Chem. Chem. Phys.* 2024, 26 (35), 23177–23188. <https://doi.org/10.1039/D4CP02330G>.
- Yang, J.; Manganaris, P.; Mannodi-Kanakkithodi, A. Discovering Novel Halide Perovskite Alloys Using Multi-Fidelity Machine Learning and Genetic Algorithm. *The Journal of Chemical Physics* 2024, 160 (6), 064114. <https://doi.org/10.1063/5.0182543>.

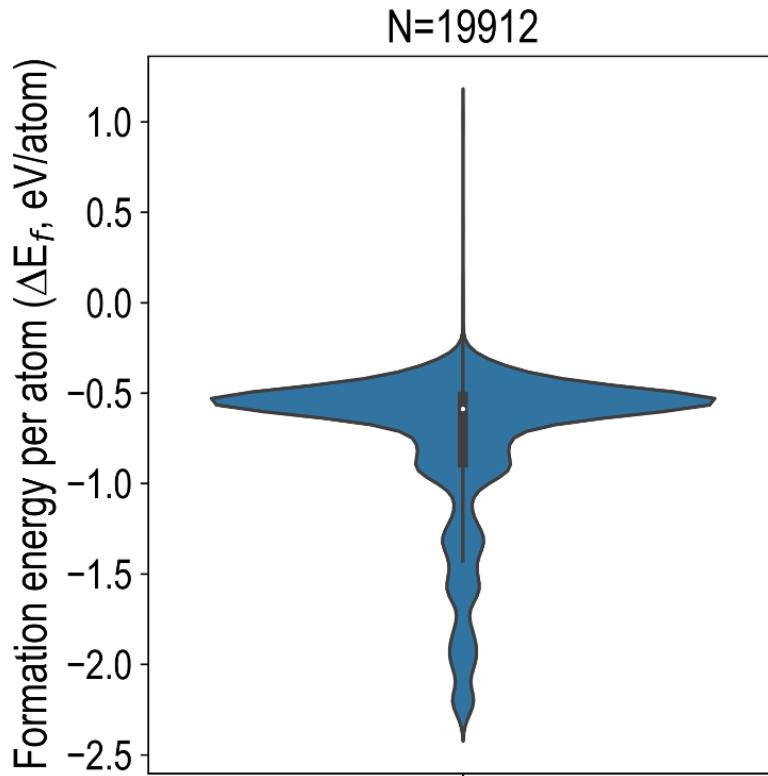
Enumeration and Prediction to discover materials for photocatalytic water splitting



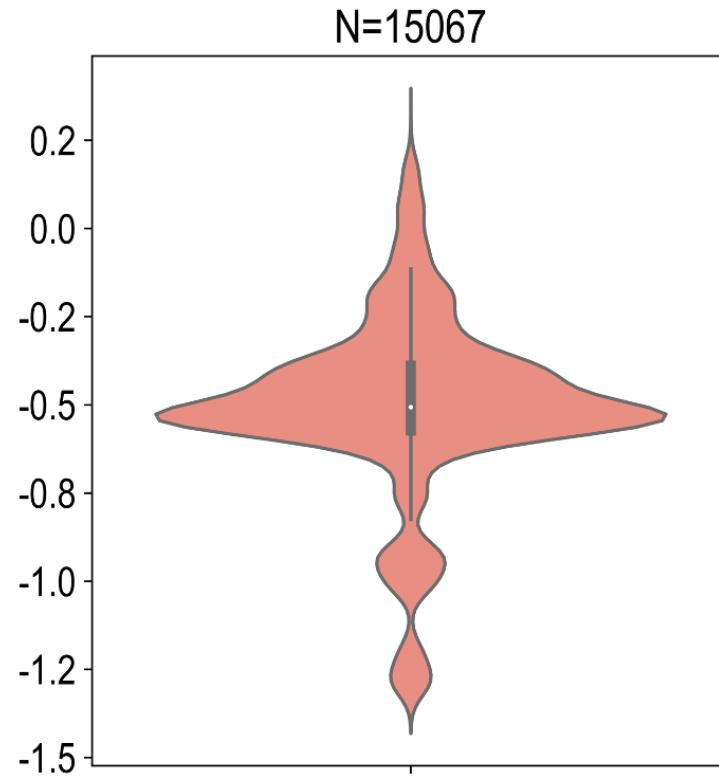
Predictions on enumerated compositions → suitable decomposition energies
and band gaps → empirical band edges to straddle the H_2O redox potentials
→ high solar-to-hydrogen conversion efficiency → validation DFT

Training Structure-Based Models

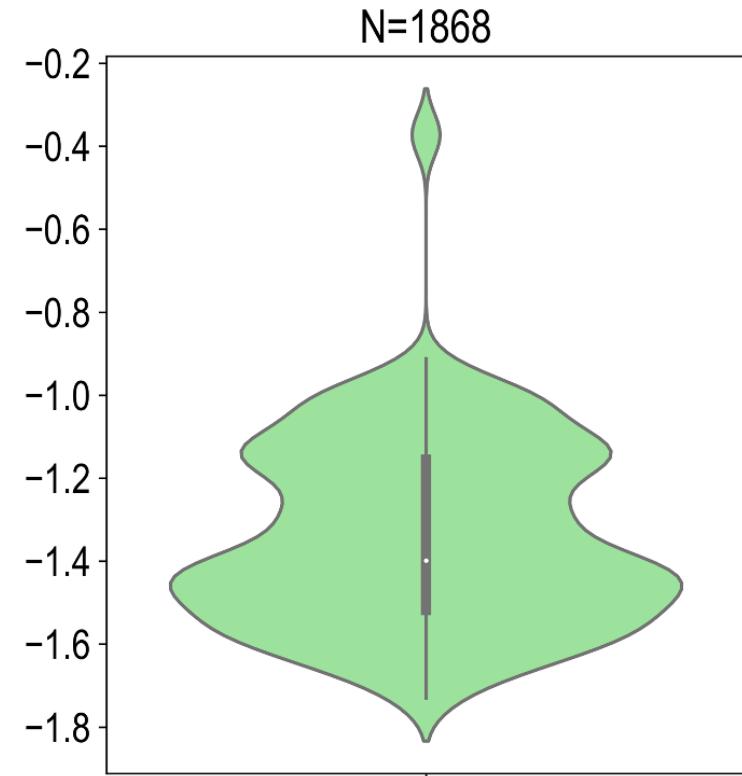
Bulk (Pure+Alloys)



Defects and Dopants

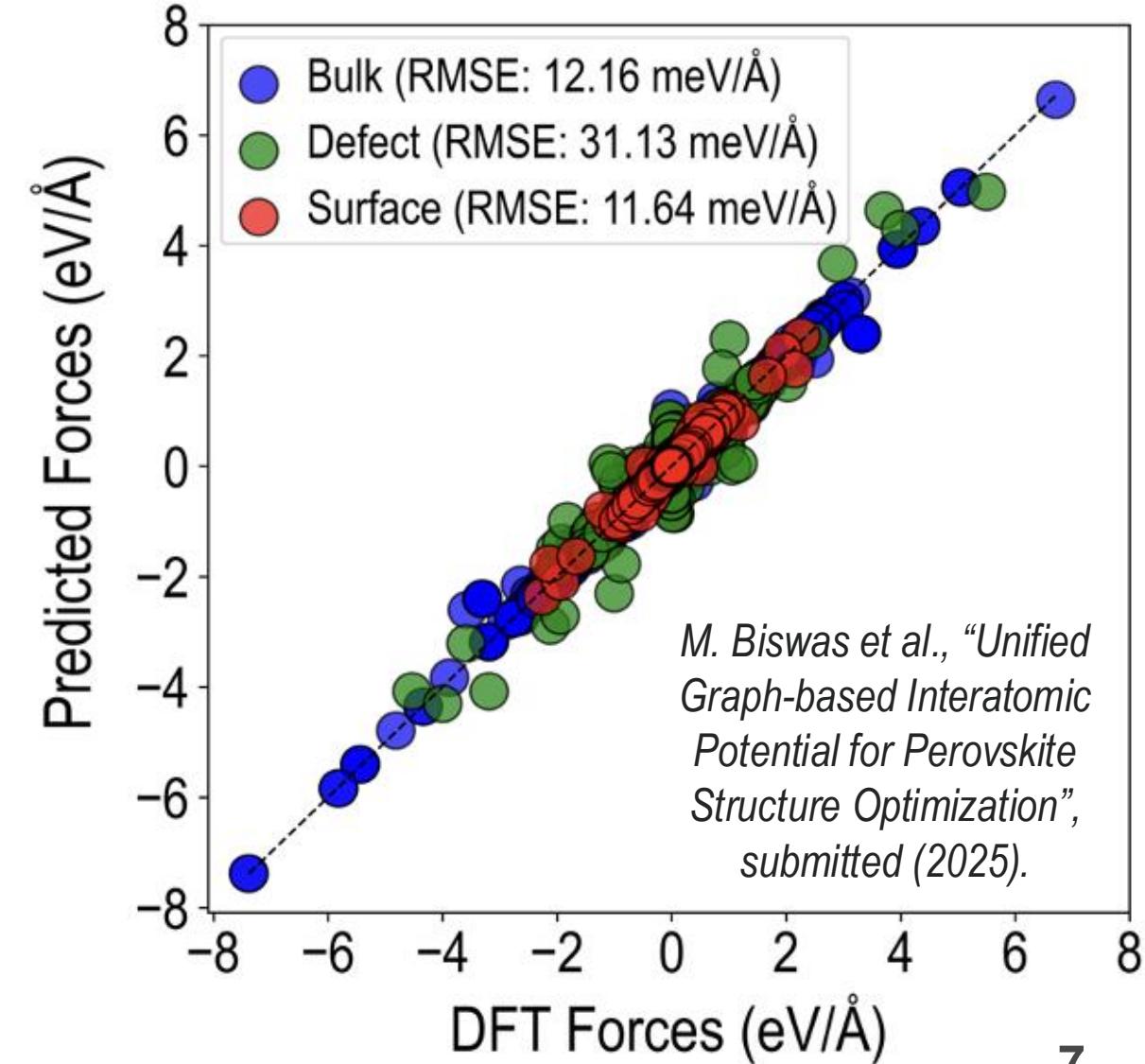
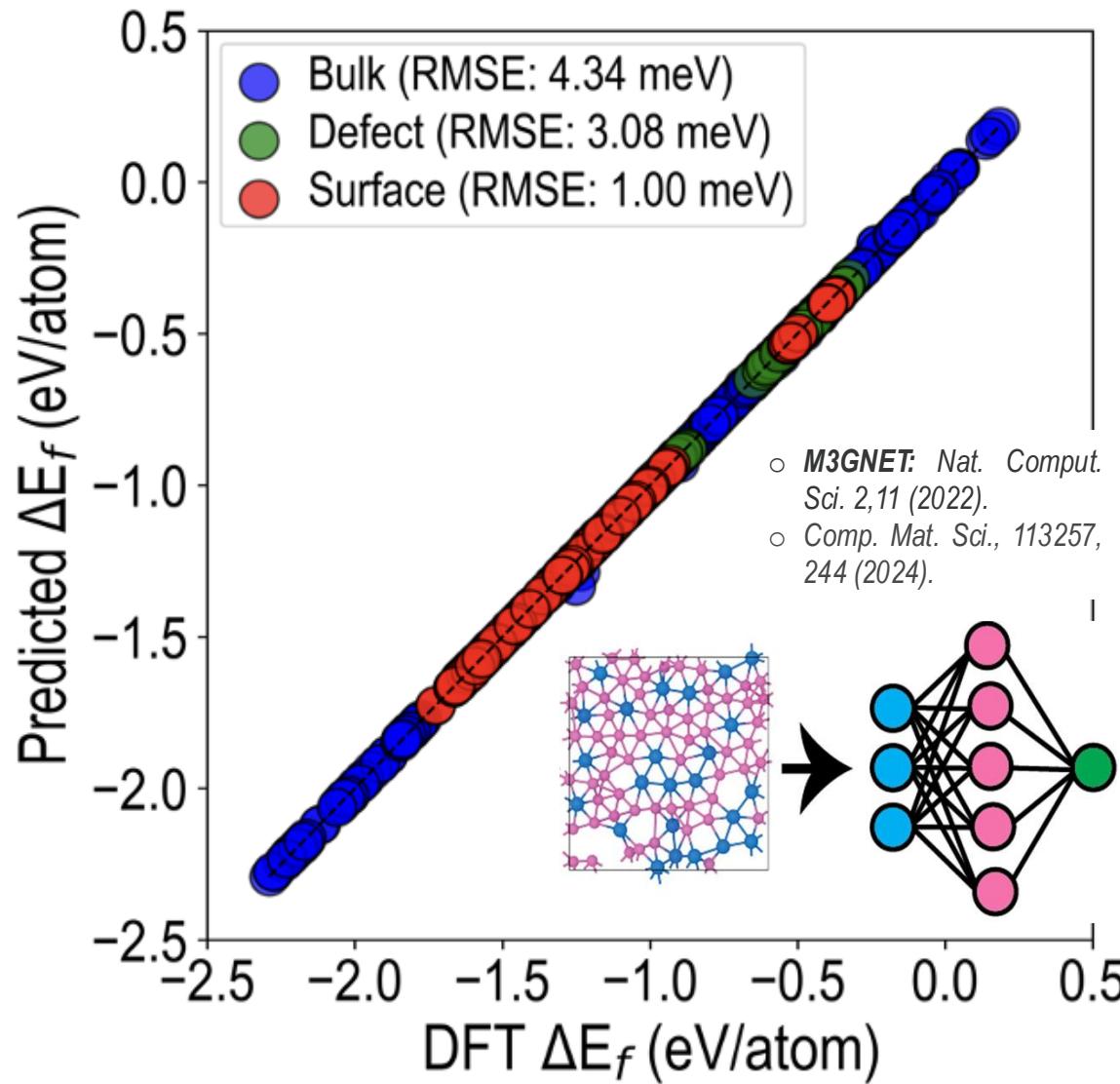


Surface Slabs



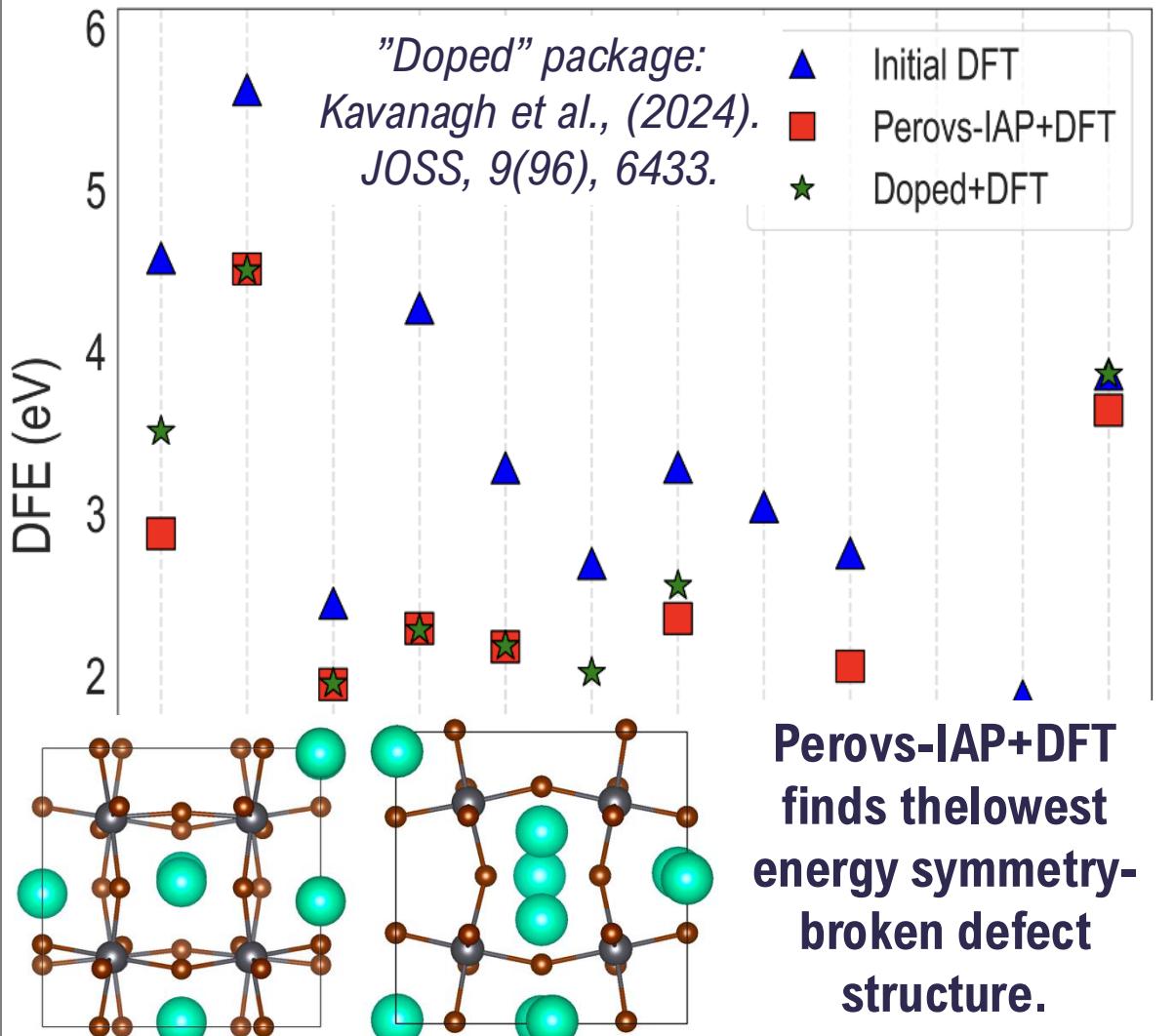
- (1) *Chem. Mater.* 31 (10), 3599–3612 (2019). (2) *J. Phys. Chem. C.* 124, 31, 16729–16738 (2020). (3) *J. Mater. Sci.* 57, 10736–10754 (2022). (4) *Energy. Environ. Sci.* 15, 1930–1949 (2022). (5) *Digital Discovery.* 2, 856-870 (2023). (6) *J. Chem. Phys.* 160, 064114 (2024). (7) *Phys. Chem. Chem. Phys.* 26, 23177-23188 (2024). (8) Biswas et al., “Unified Graph-based Interatomic Potential for Perovskite Structure Optimization”, submitted.

Structure-Based Graph Neural Network (GNN) Model

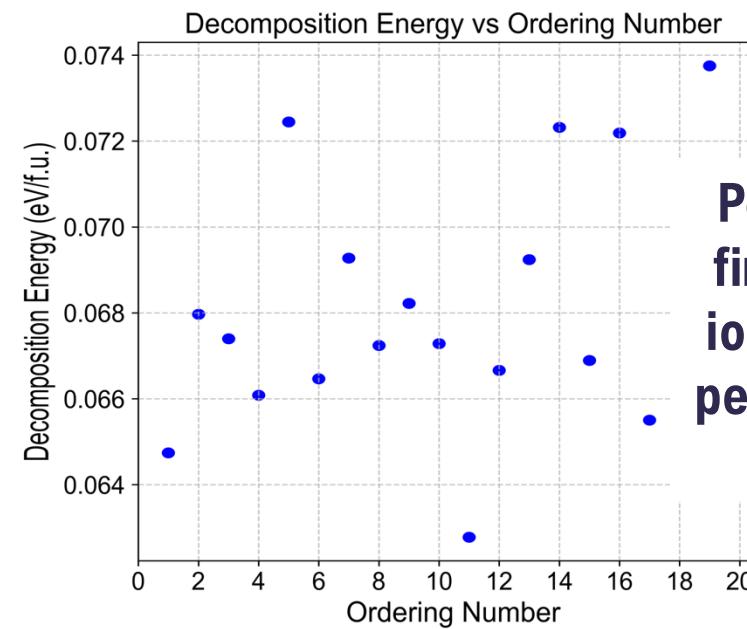
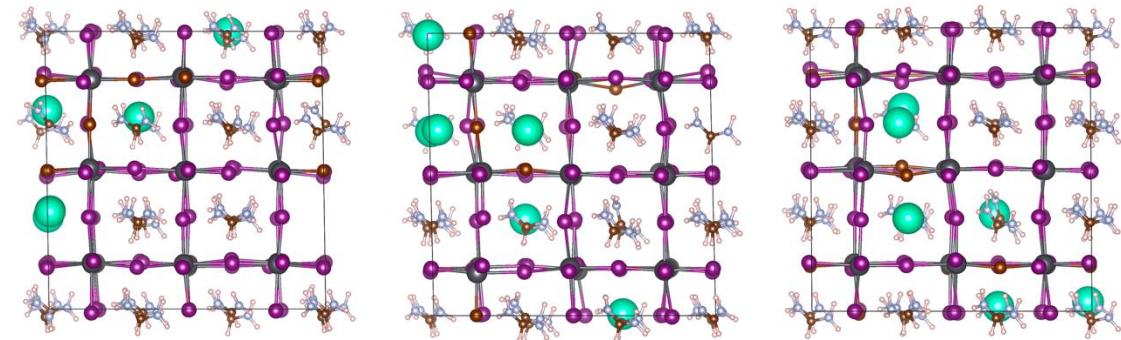


Optimizing Defects and Alloy Structures

Native Point Defects in CsPbBr_3

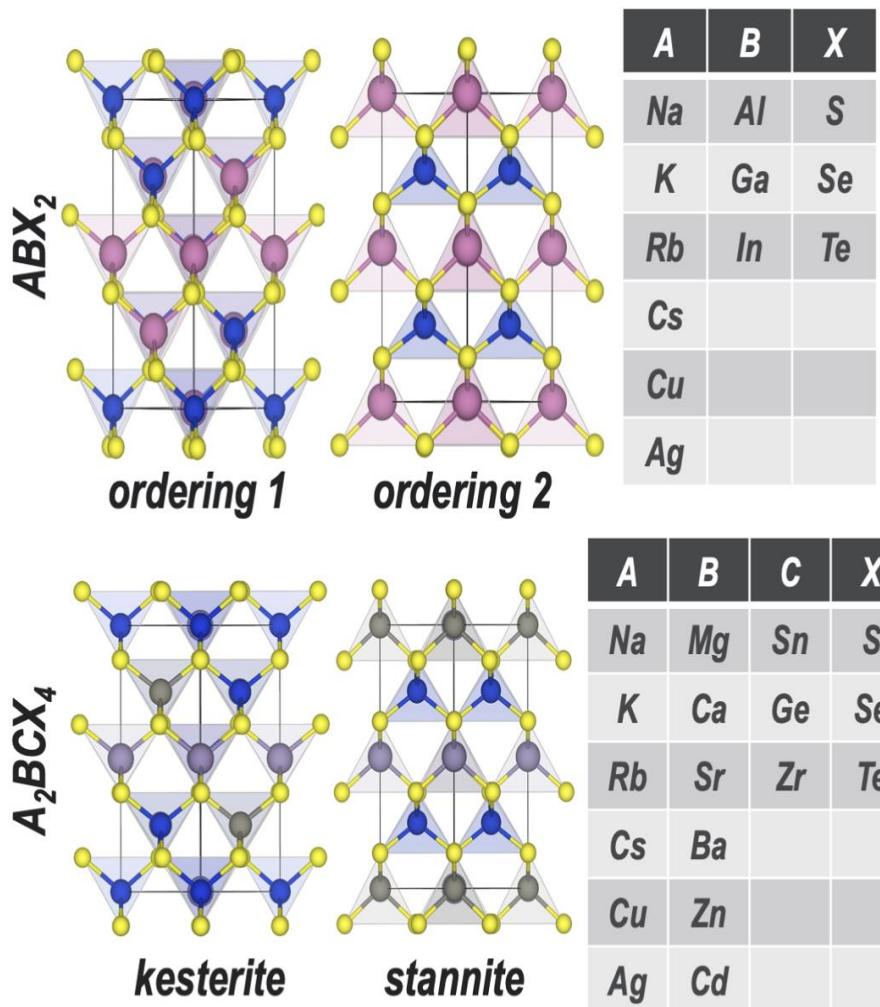


Ionic Ordering in Cubic $\text{Cs}_{0.22}\text{FA}_{0.78}\text{Pb}(\text{Br}_{0.15}\text{I}_{0.85})_3$

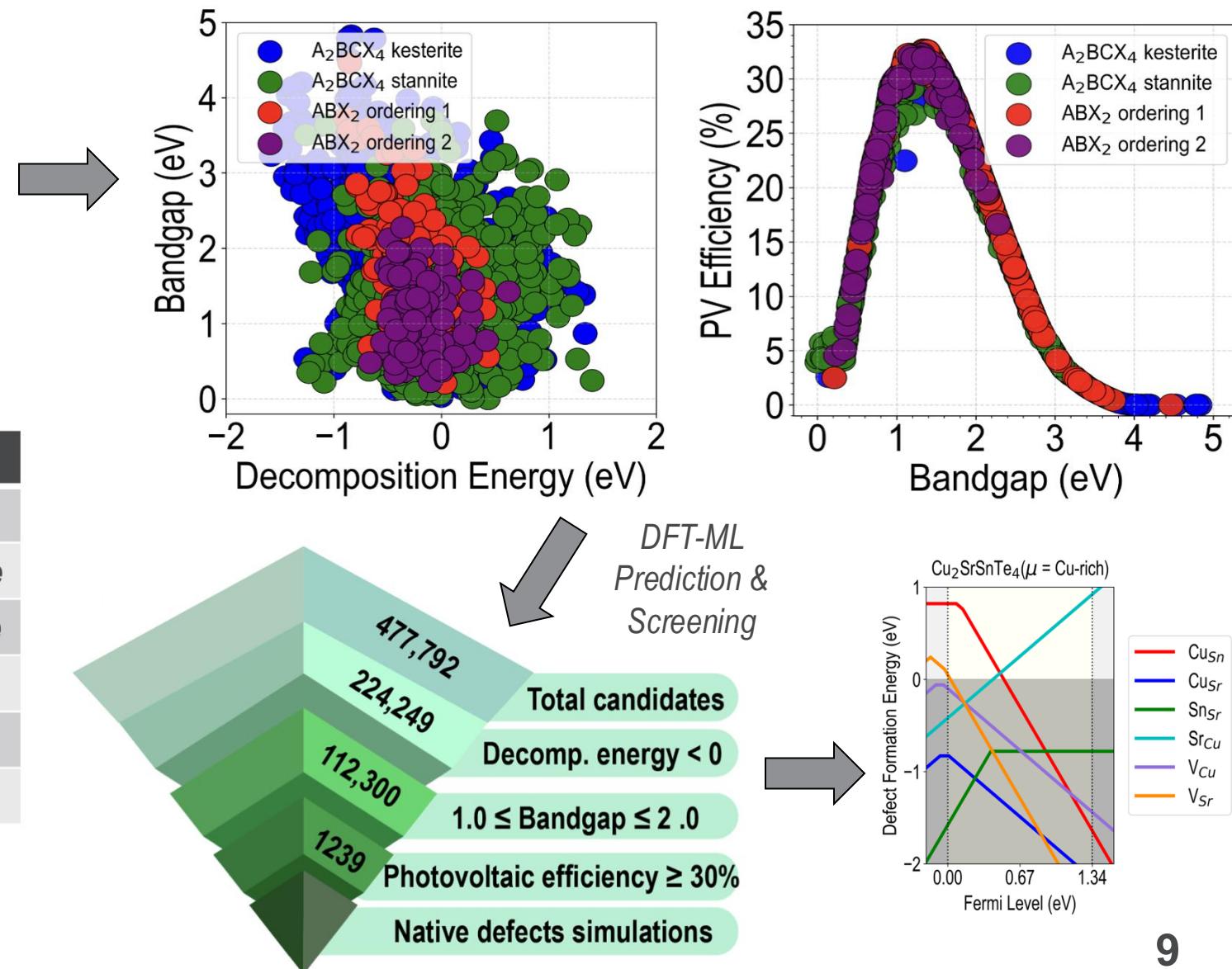


Perovs-IAP+DFT finds low energy ionic ordering for perovskite alloyed structures.

Discovering novel chalcogenide semiconductors

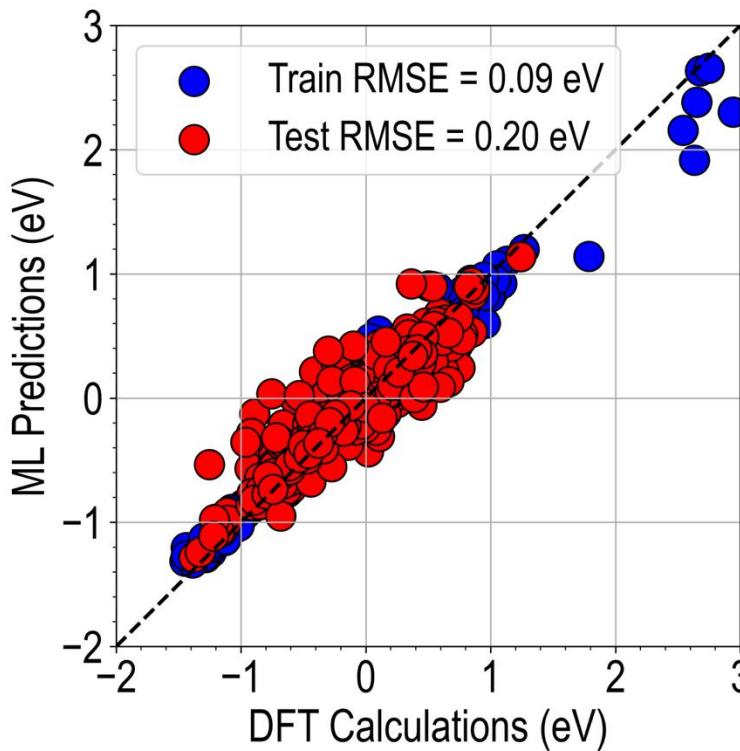


M.H. Rahman et al., Comput.
Mater. Sci. 249, 113654 (2025).

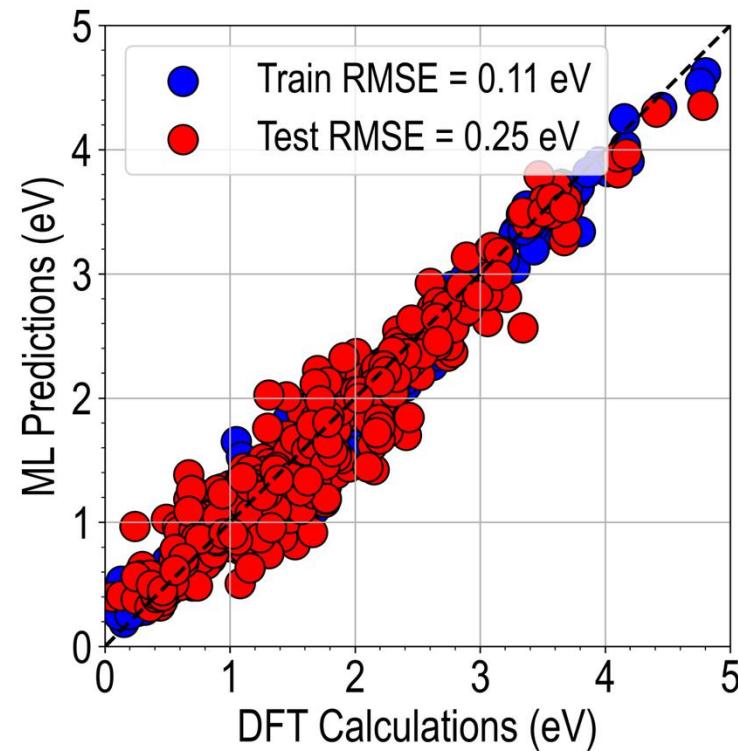


ML Models for Chalcogenides

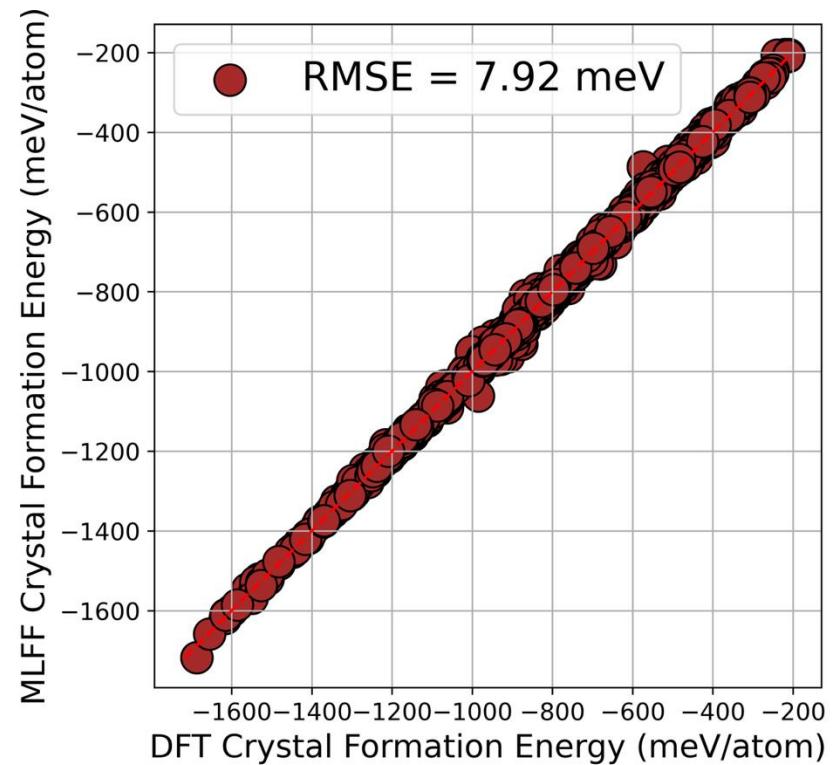
*Random Forest Model
for Decomp. Energy*



*Random Forest Model
for Bandgap*

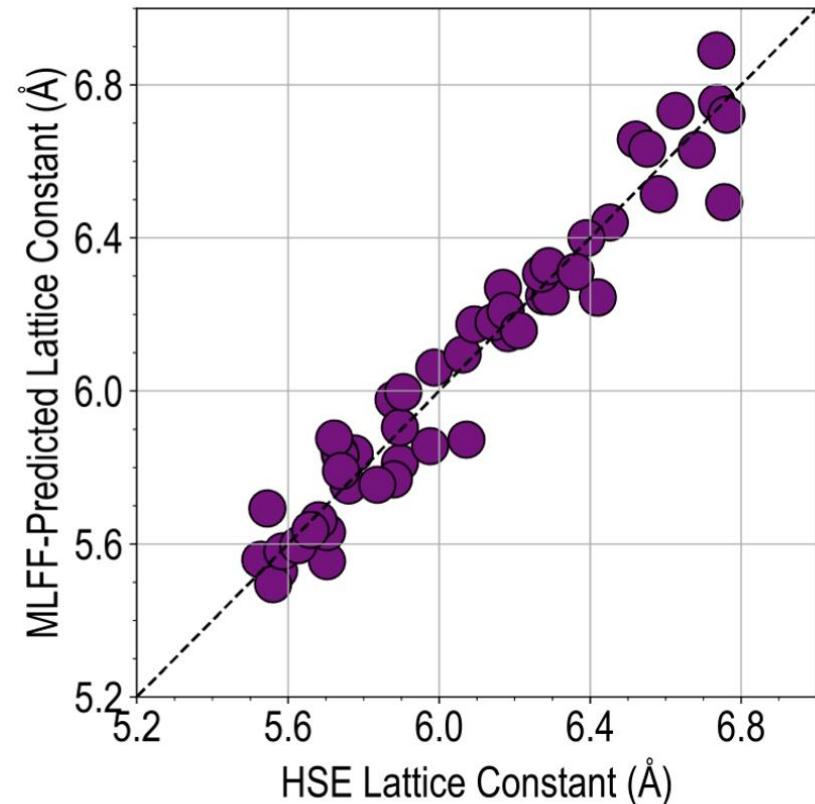
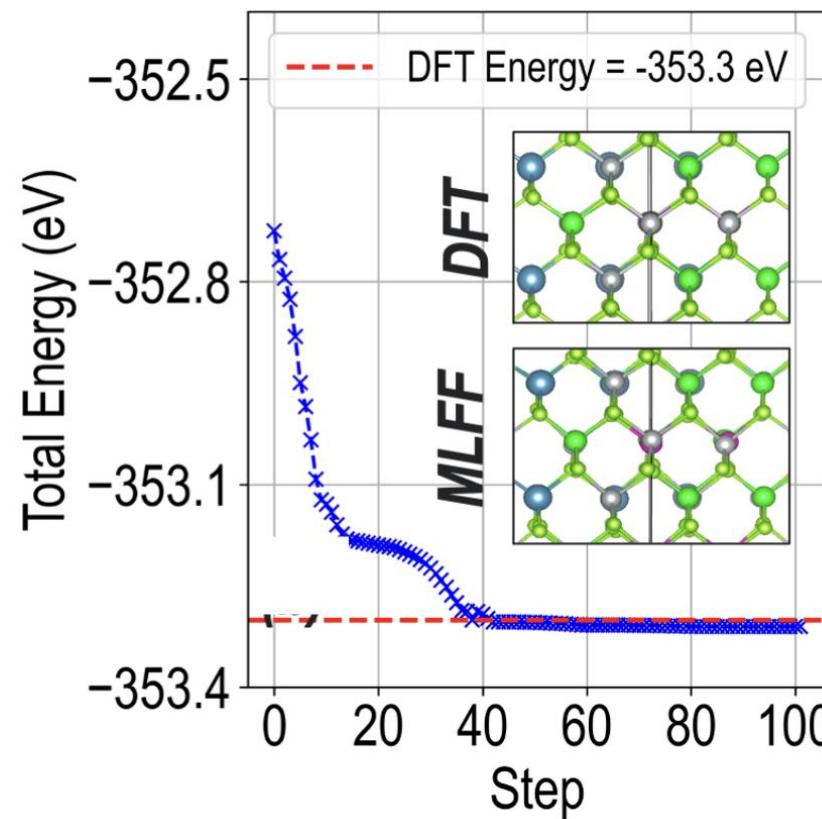
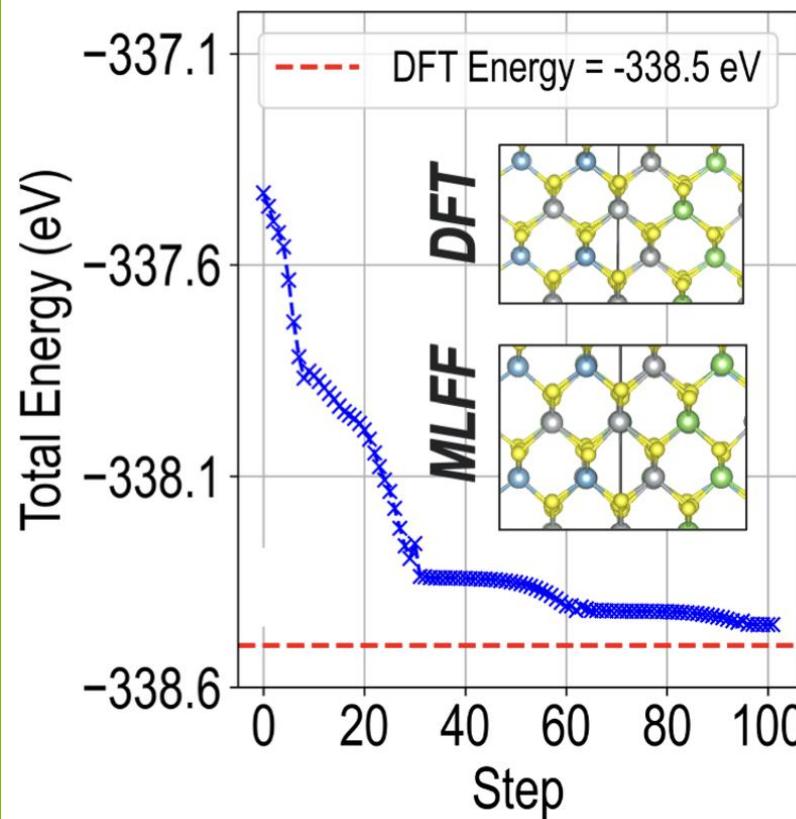


*M3GNET-MLFF Predicted
Formation Energy*



Populate ABX_2 and A_2BCX_4 compositions → RFR predictions → Generate structures for promising compositions → MLFF predictions of bulk and defect energies → Discovery

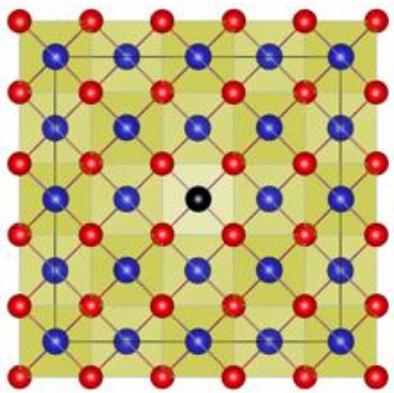
MLFF Optimization of New Compounds



The model can optimize chalcogenide bulk and defect configurations at HSE06 accuracy; currently being expanded and deployed for new predictions.

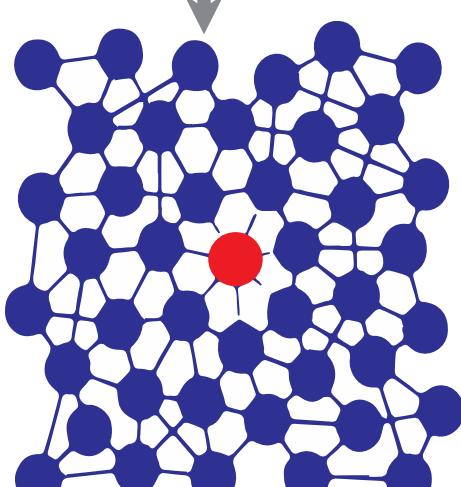
DFT-GNN Framework for Defects

Crystal with defect



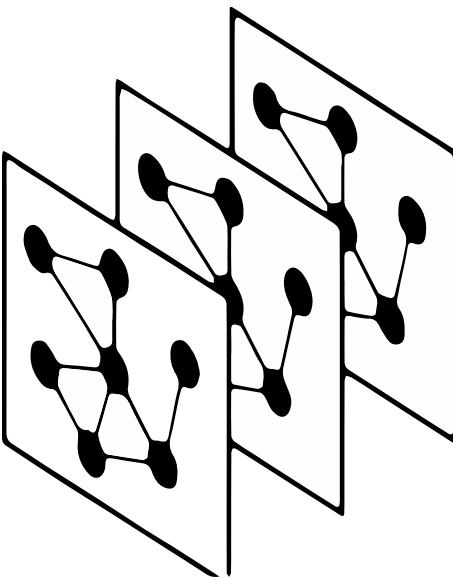
Density Functional Theory

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

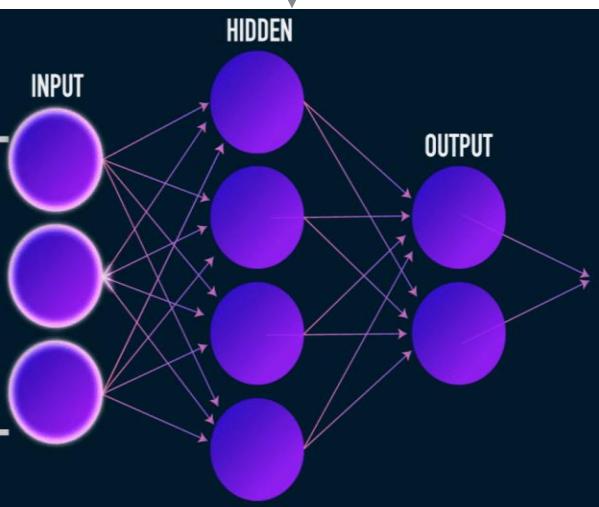


Graph Convolution Networks (GCN)

Message passing among nodes and edges



Pooling Layer



Crystal Graph

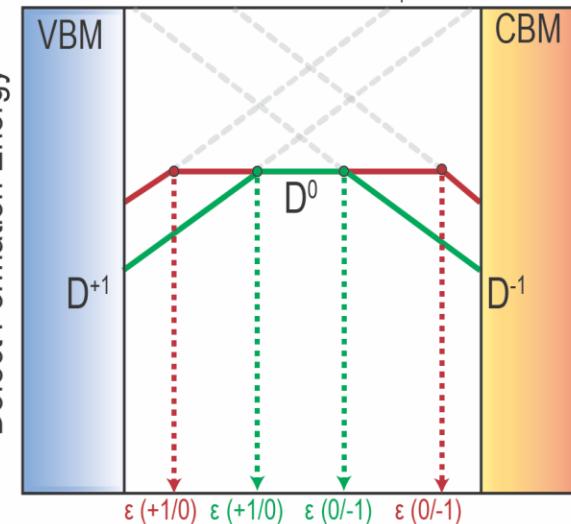
New Defects →

ML Models

→ $E^f(q, E_F), \varepsilon(q_1/q_2)$

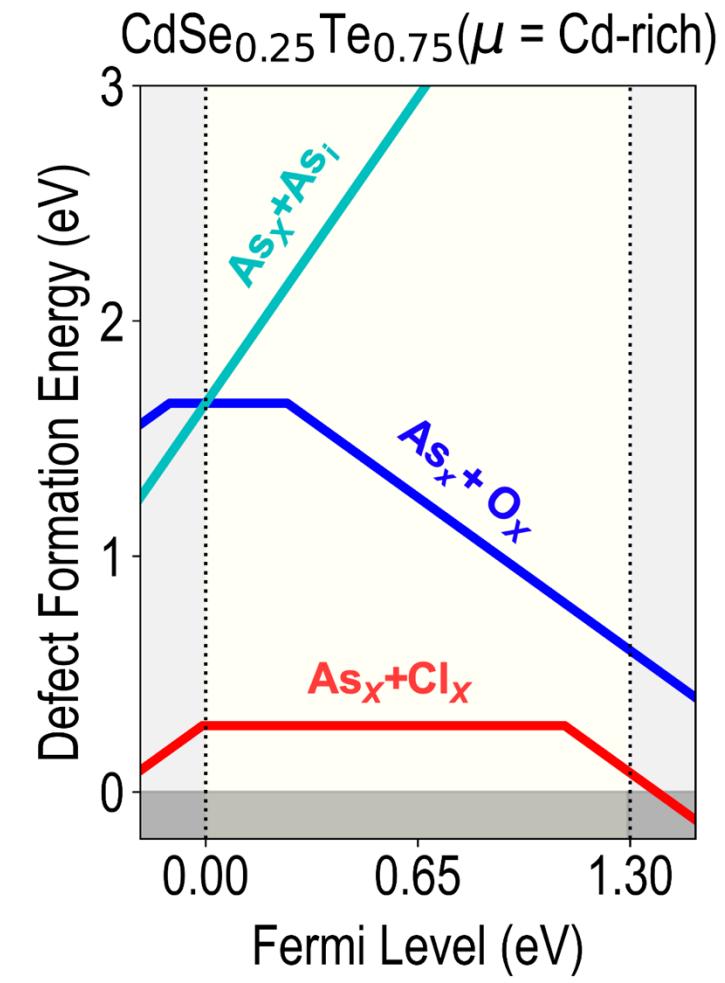
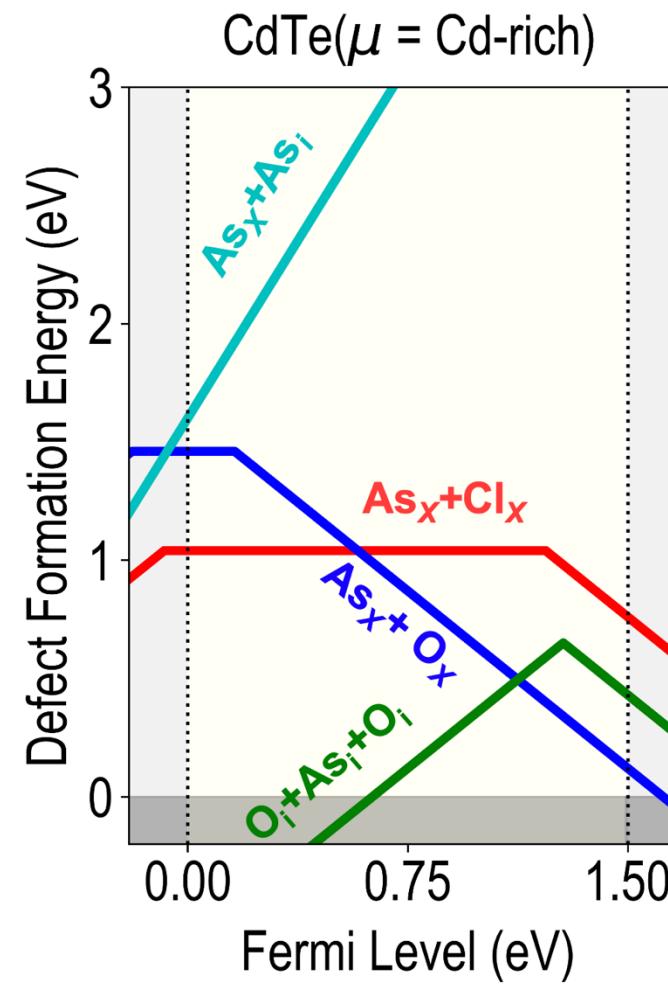
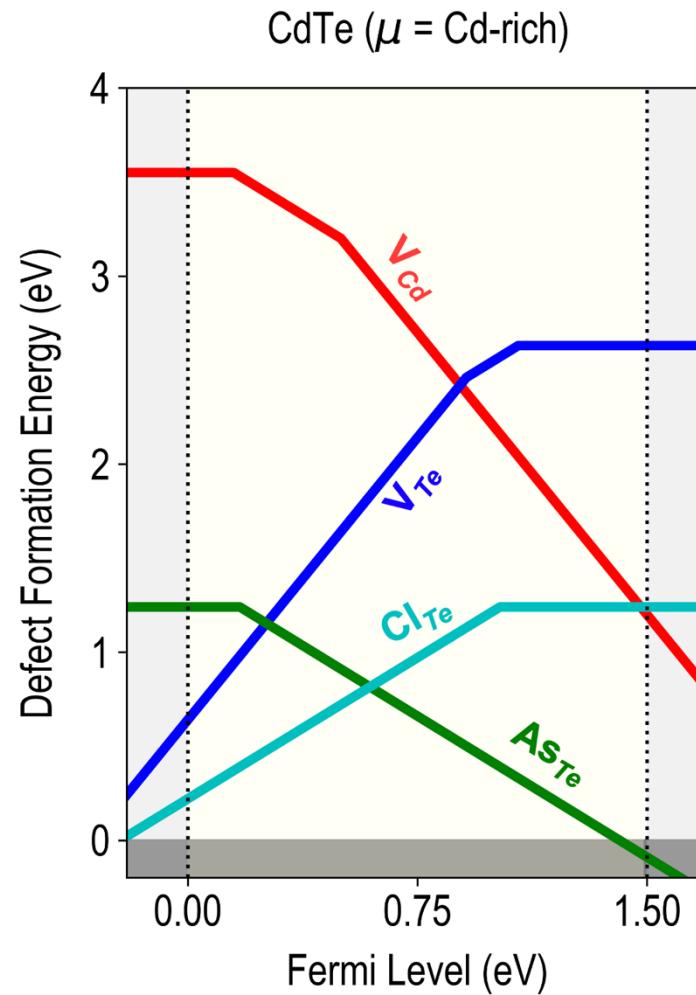
→ Accelerated Discovery

Fermi Level E_F



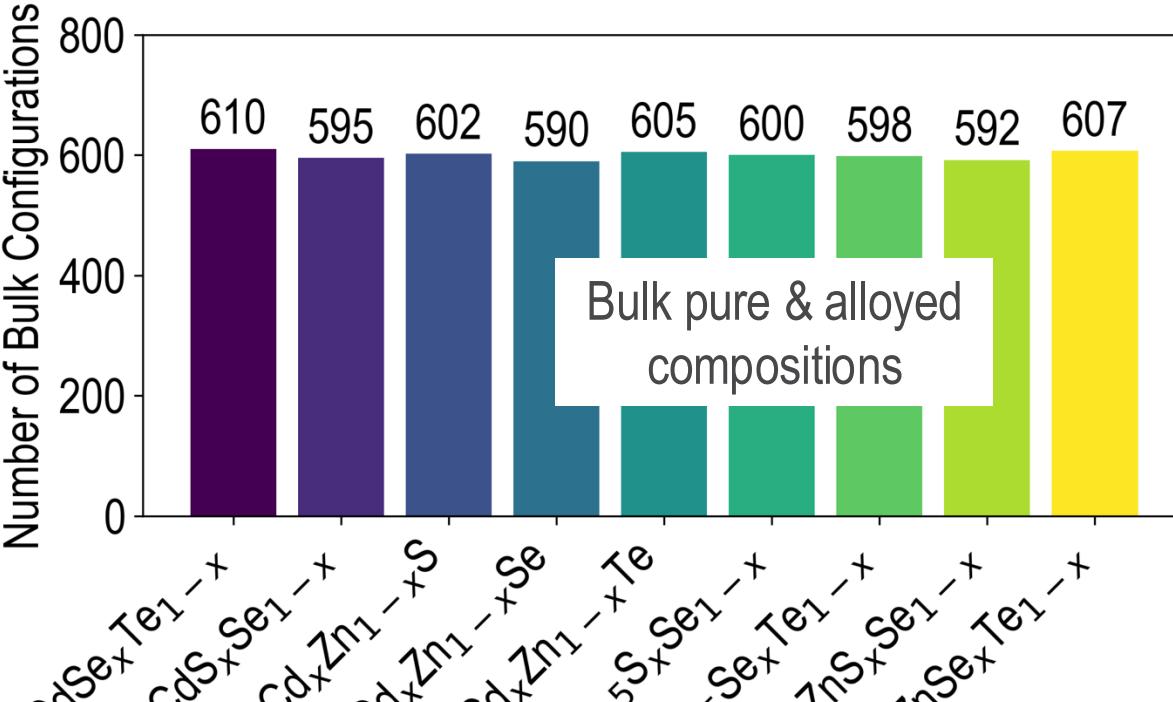
$E^f(q, E_F), \varepsilon(q_1/q_2)$

Defect Formation Energies in CdSeTe

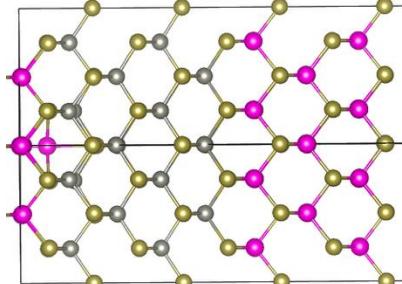


[1] A. Mannodi-Kanakkithodi et al., Patterns. 3, 3, 100450 (2022). [2] A. Mannodi-Kanakkithodi, Modelling Simul. Mater. Sci. Eng. 30 044001 (2022). [3] M.H. Rahman et al., APL Machine Learning. 2, 016122 (2024). [4] M.H. Rahman et al., “First Principles Investigation of Dopants and Defect Complexes in $\text{CdSe}_x\text{Te}_{1-x}$ ”, under review. [5] M.H. Rahman et al., “Learning Defect Thermodynamics in CdSeTe Solar Cells Using Atomistic Simulations and Graph-based Force Fields”, in preparation.

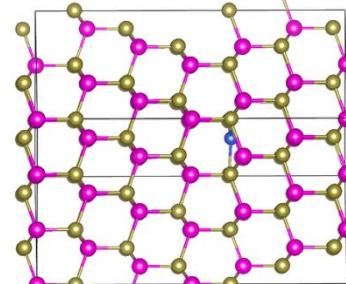
Defect Dataset for Training GNN Models



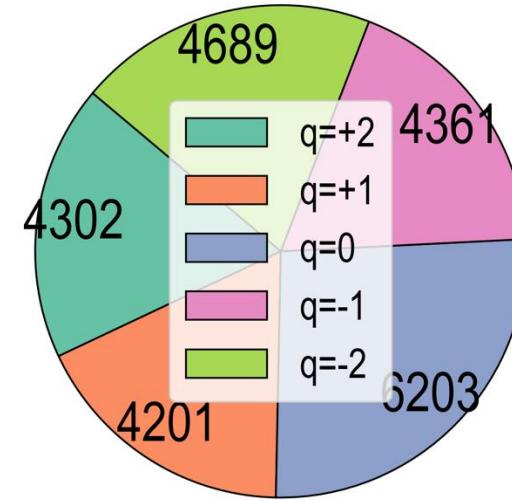
Cd_i in CdTe/ZnTe



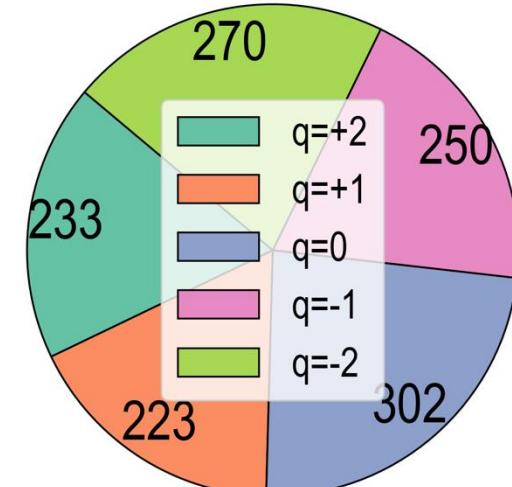
Cu_i in CdTe grain



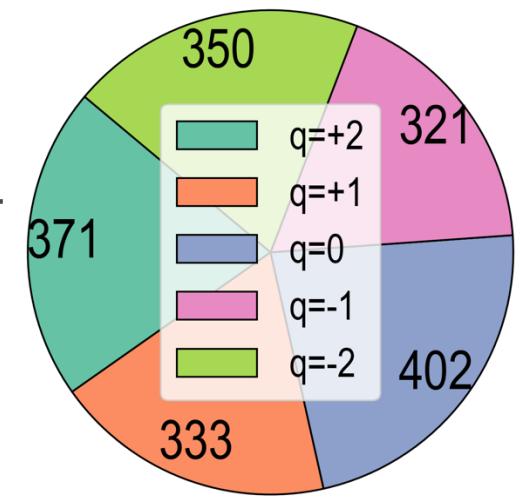
64-atom supercells



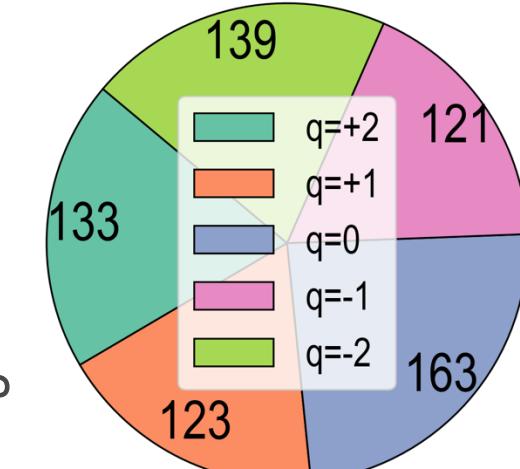
CdTe/ZnTe interface



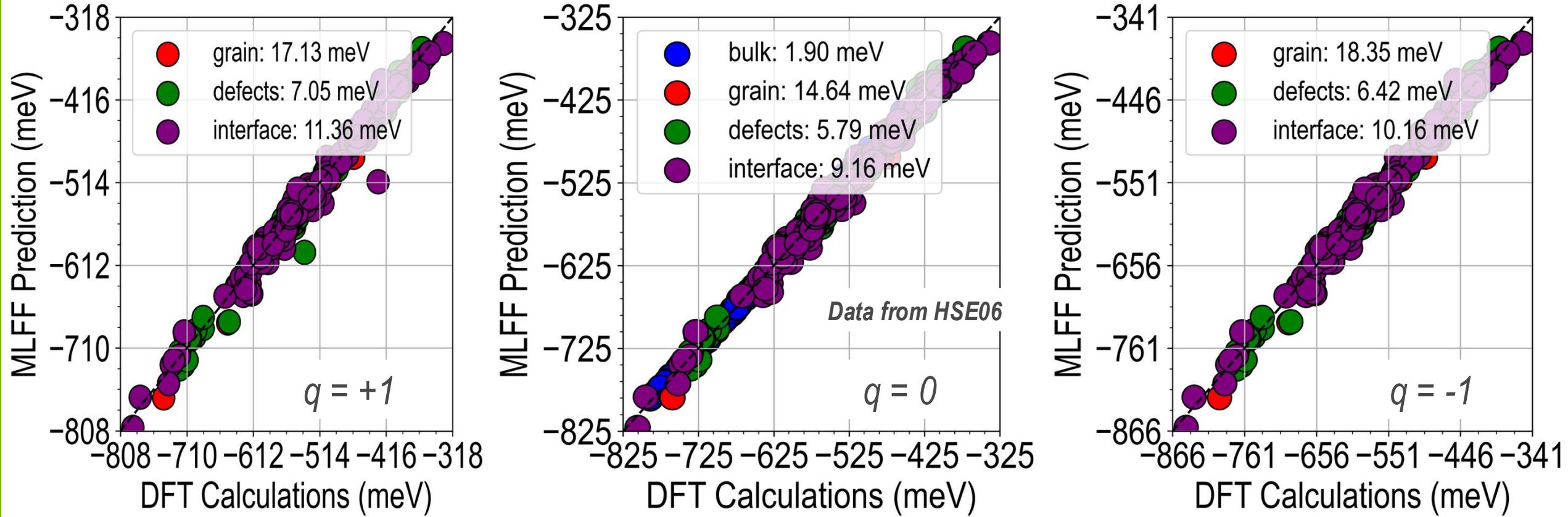
216-atom supercells



CdTe grain boundaries

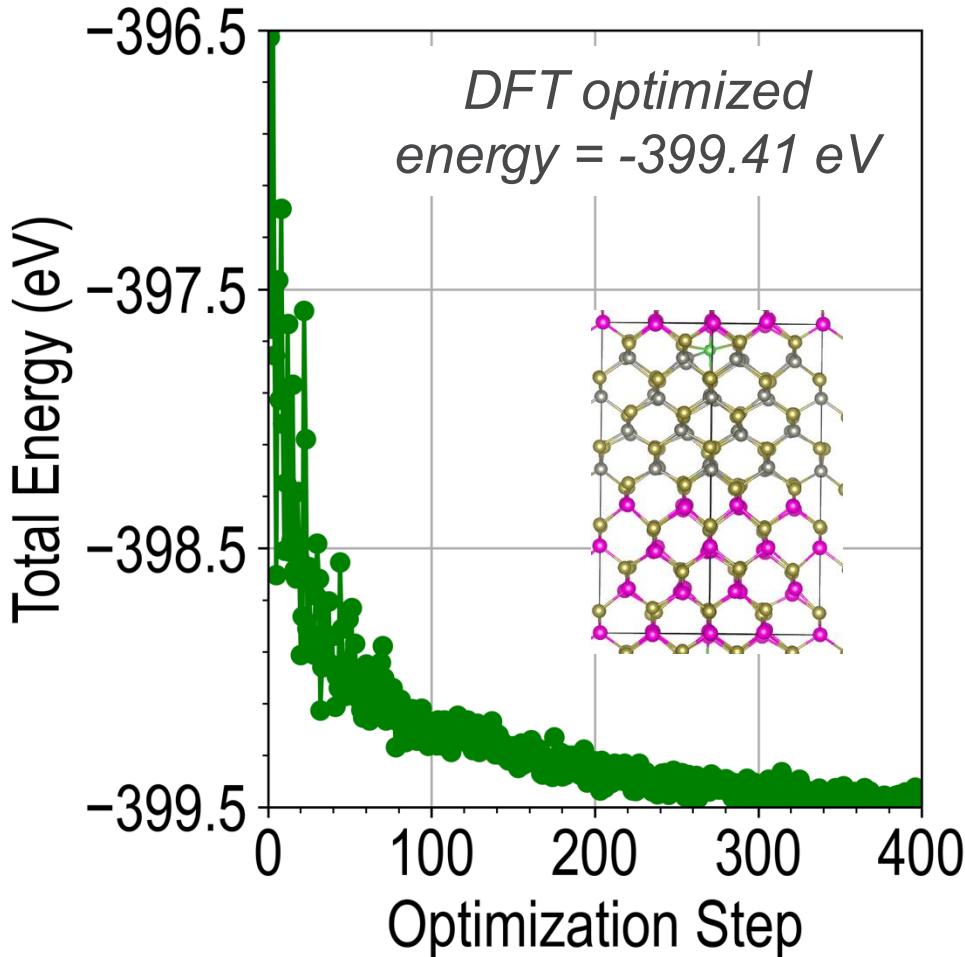


M3GNET-MLFF Models for Defect Structures

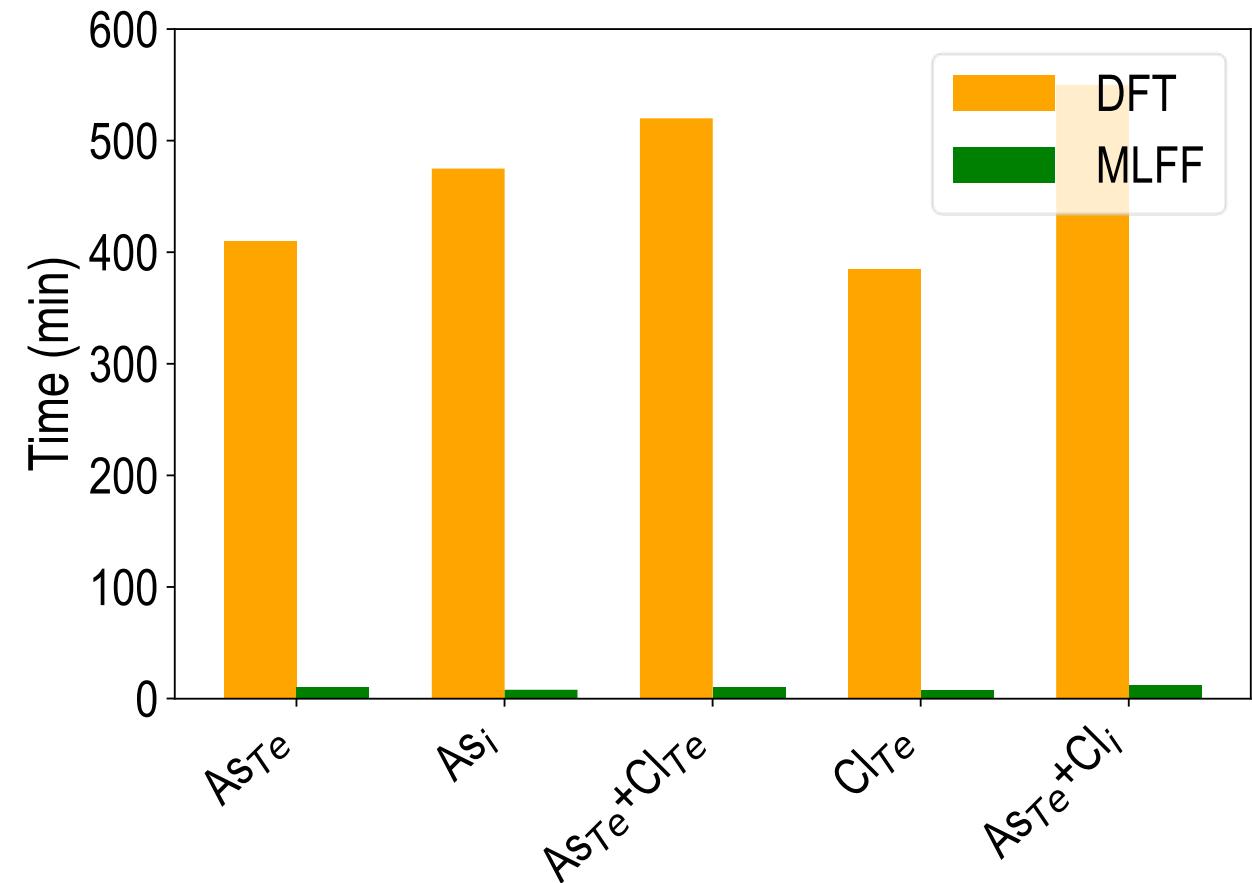


- M3GNET-MLFF models trained for different charge states across the entire space of Cd/Zn-Te/Se/S compositions (bulk alloys, interfaces, GBs, etc.).
- These models are “surrogates” for DFT: quick energy prediction, optimization, and screening of defects.

Rapid Geometry Optimization



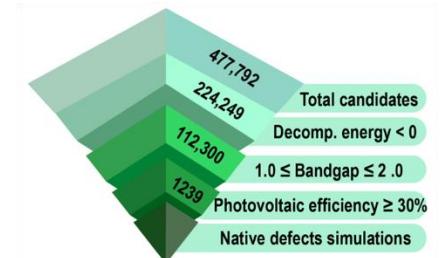
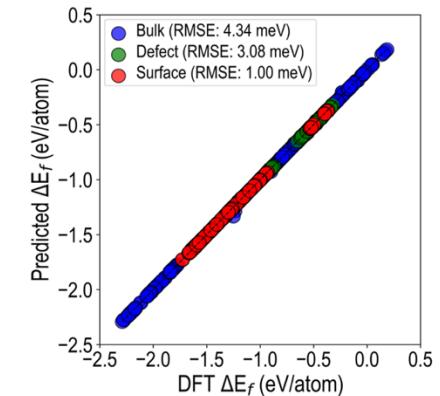
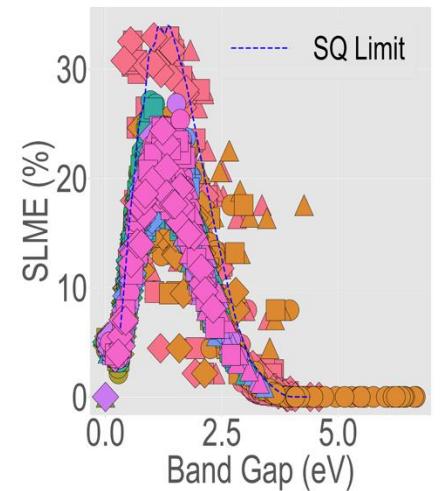
GNN-MLFF used to optimize **As_i** defect at the **CdTe/ZnTe** interface



DFT (HSE06) takes around 8 hours per defect, MLFF takes a few minutes

Summary and Ongoing Work

- Our DFT+ML framework predicts thermodynamic stability, electronic structure, and defect properties of halide and chalcogenide semiconductors.
- GNN interatomic potential rapidly optimizes structures before expensive DFT is performed for properties of interest.
- GNN models are being extended to: (a) universally apply to all halides and chalcogenides, (b) predict directly at HSE06 accuracy, (c) predict defect structures in the presence of charge, and (d) evaluate synthesis likelihood of compounds.



<https://www.mannodigroup.com/>

Interface with the data and models



Input the composition string, phase, and prediction fidelity (PBE/HSE)

```
▶ # Input composition as a list
#A_site = ['K', 'Rb', 'Cs', 'MA', 'FA',]
#B_site = ['Ca', 'Sr', 'Ba', 'Ge', 'Sn', 'Pb']
#X_site = ['Cl', 'Br', 'I']

compositions = ["Cs0.5K0.5PbBr1.5Cl1.5", "CsCa0.25Ge0.75Br3", "FAPbI3"] # Input composition string like this
phase = ["tetra", "cubic", "cubic"] # phases options: "cubic"/"tetra"/"ortho"/"hex"
prediction_fidelity = ["HSE"] # or "PBE"
```

Featurize the input data to generate descriptors

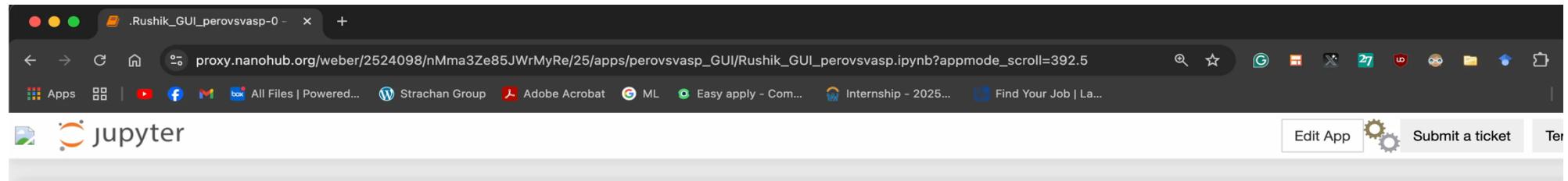
	Composition	K	Rb	Cs	MA	FA	Ca	Sr	Ba	Ge	...	X_Heat_vap	X_Electronegativity	X_At_num	X_Period	Cubic	Tetra	Ortho	Hex	PBE	HSE	
0	Cs0.5K0.5PbBr1.5Cl1.5	0.5	0	0.5	0	0.0	0.00	0	0	0.00	...	12.4625		3.06	26.0	3.5	0	1	0	0	0	1
1	CsCa0.25Ge0.75Br3	0.0	0	1.0	0	0.0	0.25	0	0	0.75	...	14.7250		2.96	35.0	4.0	1	0	0	0	0	1
2	FAPbI3	0.0	0	0.0	0	1.0	0.00	0	0	0.00	...	20.9000		2.66	53.0	5.0	1	0	0	0	0	1

3 rows × 57 columns

Check for pre-existing DFT/ML predicted data, otherwise provide on-demand ML predictions

	Composition	x(S)	Phase	Source	Functional	Band_Gap	CBM	VBM	Suitable_for_WS	STH_Efficiency	Decomposition_Energy_HSE	Decomposition_Energy_PBE
0	Cs0.5K0.5PbBr1.5Cl1.5	5.477136	Tetragonal	ML-new	HSE	1.765362	0.154455	1.919816	False	NaN	-0.127706	NaN
1	CsCa0.25Ge0.75Br3	5.201399	Cubic	ML-ensemble	HSE	2.221220	-0.349211	1.872009	True	18.873539	-0.187585	-0.233319
2	FAPbI3	5.670442	Cubic	DFT	PBE	1.942000	0.259442	2.201442	False	NaN	NaN	0.136705
3	FAPbI3	5.670442	Cubic	DFT	HSE	1.687100	0.386892	2.073992	False	NaN	0.147921	NaN

Interface with the data and models



The screenshot shows a web browser window titled ".Rushik_GUI_perovsasp-0". The URL is proxy.nanohub.org/weber/2524098/nMma3Ze85JWrMyRe/25/apps/perovsasp_GUI/Rushik_GUI_perovsasp.ipynb?appmode_scroll=392.5. The browser interface includes a navigation bar with back, forward, and search buttons, and a toolbar with various icons. Below the toolbar, there's a "jupyter" logo and a "Edit App" button. On the right side of the browser window, there's a QR code.

1 PerovsVASP: A Graphical User Interface

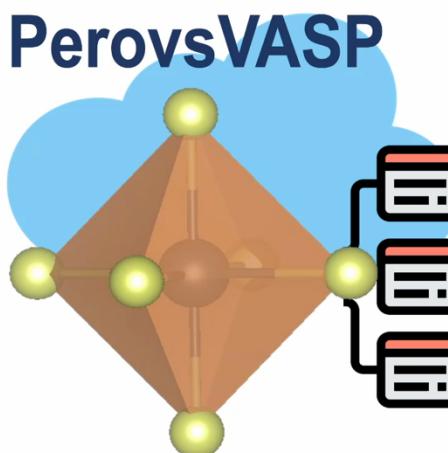
Rushik Desai, Maitreyo Biswas, Juan Carlos Verduzco, Alejandro Strachan, Arun Mannodi Kanakkithodi
School of Materials Engineering and Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana 47907, United States

1.1 Overview

This graphical user interface is designed and managed by the **Data Driven Materials Design** group at **Purdue University** and it acts as a portal to connect the data on perovskites stored on the Sim2L PerovsVASP which can be found [here](#). All the data presented and queried is generated through Density Functional Theory using the Vienna Ab-initio Simulation Package (VASP). Using the interface, you are able to get the simulation inputs to the results presented without the POTCARs. Using this Sim2L and GUI, researchers satisfy data-sharing requirements such as those called for by the [US Office of Science and Technology Policy](#). First the overview of the current data and type of data is presented followed by the GUI where you can get a DFT-run result if available and if not, a prediction for decomposition energy and band gap using the machine learning models.

1.2 Database

Currently the Sim2L houses 3D Halide Perovskite DFT runs across a variety of composition and crystal systems. The description, methodology and explanation for the choice of properties in these databases can be found in the following studies: [\(Yang et al.\)](#) with another paper in the works that will be shortly published. Following interactive charts would demonstrate the chemical species and types of properties stored on the tool.



A-Site Species B-Site Species X-Site Species Crystal Structures

Interface with the data and models



```
Defects-DFT-GNN.ipynb
File Edit View Insert Runtime Tools Help
+ Code + Text Copy to Drive

Data Visualization

[ ] import pandas as pd
      import matplotlib.pyplot as plt

file_url = 'https://raw.githubusercontent.com/msehabibur/defect\_GNN\_gen\_1/main/Codes/id\_prop.csv'

data_no_header = pd.read_csv(file_url, header=None, on_bad_lines='skip')

data_no_header.columns = ['Filename', 'Property Value']

data_no_header['Category'] = data_no_header['Filename'].apply(lambda x: 'Bulk' if 'bulk' in x else 'Defect')

category_counts = data_no_header['Category'].value_counts()

colors = ['#66c2a5', '#fc8d62']
```

Interface with the data and models

▼ Relaxation on Perovskite Structures using Perovs-IAP

This notebook would aid you in loading our fine-tuned M3GNET model Perovs-IAP, and relax a perovskite structure from the Materials Project. We also plot the relaxation trajectory which shows how the relaxation proceeds.

```
✓ [13] #ignore warnings command
    import warnings
    warnings.filterwarnings('ignore')

✓ [14] #Uninstall conflicting libraries
    !pip uninstall -y torch torchaudio torchvision torchdata torcht
    # Install matgl and its dependencies
    !pip install matgl
```



Acknowledgements

Purdue students: Jiaqi Yang, Panayotis Manganaris, Habibur Rahman, Maitreyo Desai, Rushik Desai, Anika Bhoopalam, Gavin Bidna, Benji Posey, Tyriek Craigs, Yi Yang, Kushal Samanta



Collaborators:
Mariana Bertoni,
Srisuda Rojsatien
(ASU), Clemens Heske
(UNLV), Dmitry Krasikov
(FS), Kamal Choudhary,
Brian DeCost (NIST),
David Fenning (UCSD),
Letian Dou, Shubhra
Bansal, Rakesh Agrawal
(Purdue), Maria Chan
(Argonne).

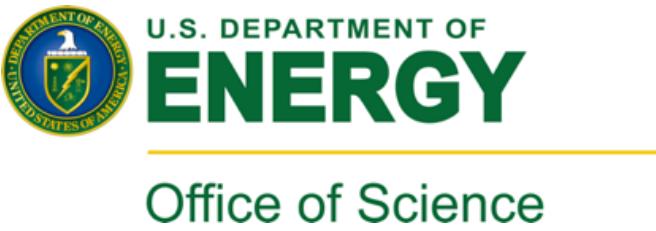
Acknowledgements

Purdue MSE: Startup account

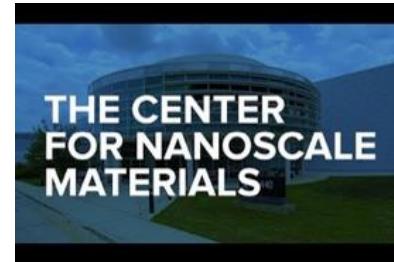
DOE EERE: SETO-SIPS #DE-EE0011158

NSF-CHE 1955336, NSF-CHE 2328176

DOE EERE: SETO Award Number 37989



School of Materials Engineering

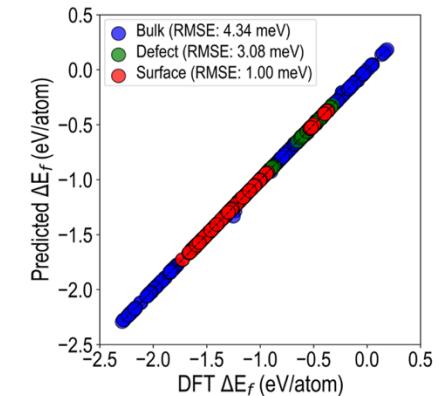
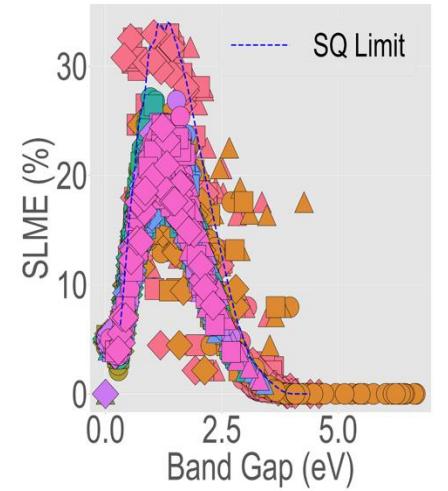


National Energy Research
Scientific Computing Center



Summary and Ongoing Work

- Our DFT+ML framework predicts thermodynamic stability, electronic structure, and defect properties of halide and chalcogenide semiconductors.
- GNN interatomic potential rapidly optimizes structures before expensive DFT is performed for properties of interest.
- GNN models are being extended to: (a) universally apply to all halides and chalcogenides, (b) predict directly at HSE06 accuracy, (c) predict defect structures in the presence of charge, and (d) evaluate synthesis likelihood of compounds.



<https://www.mannodigroup.com/>

