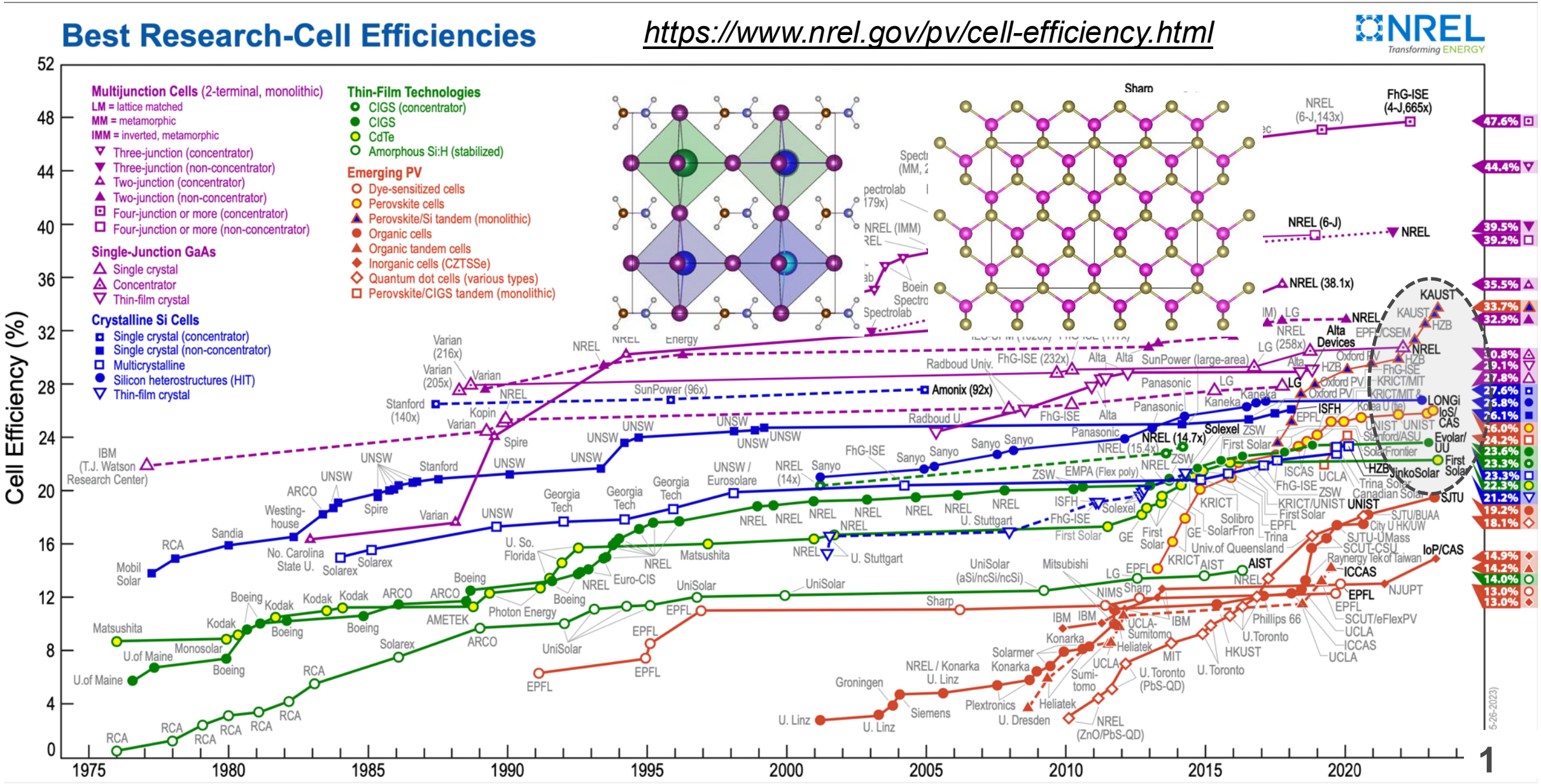
Designing Halide Perovskites for Solar Absorption using Machine Learning

Arun Mannodi Kanakkithodi

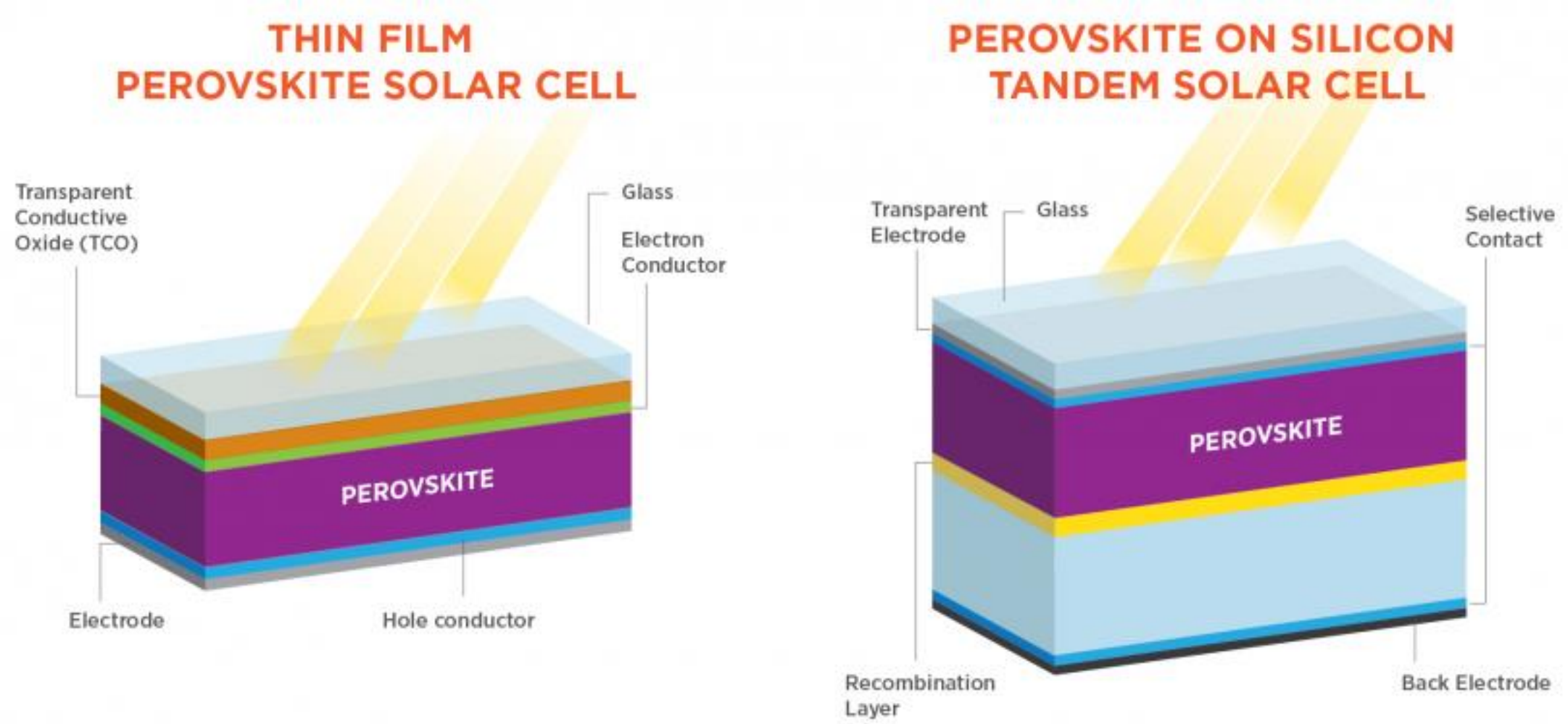
Assistant Professor, School of Materials Engineering
Purdue University, West Lafayette, IN

PURDUE UNIVERSITY | School of Materials Engineering

Research: Absorbers for Next-Gen Solar Cells

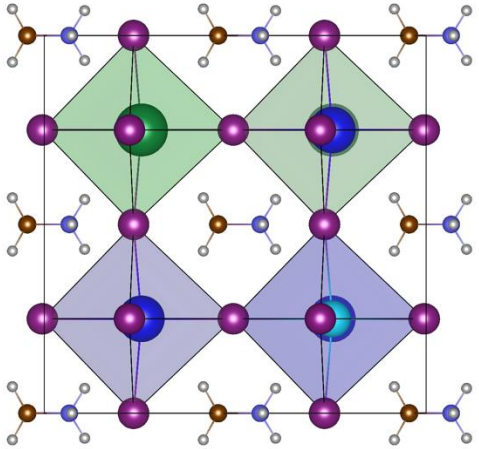


Perovskite Solar Cells



https://www.energy.gov/eere/solar/perovskite-solar-cells

Halide Perovskites

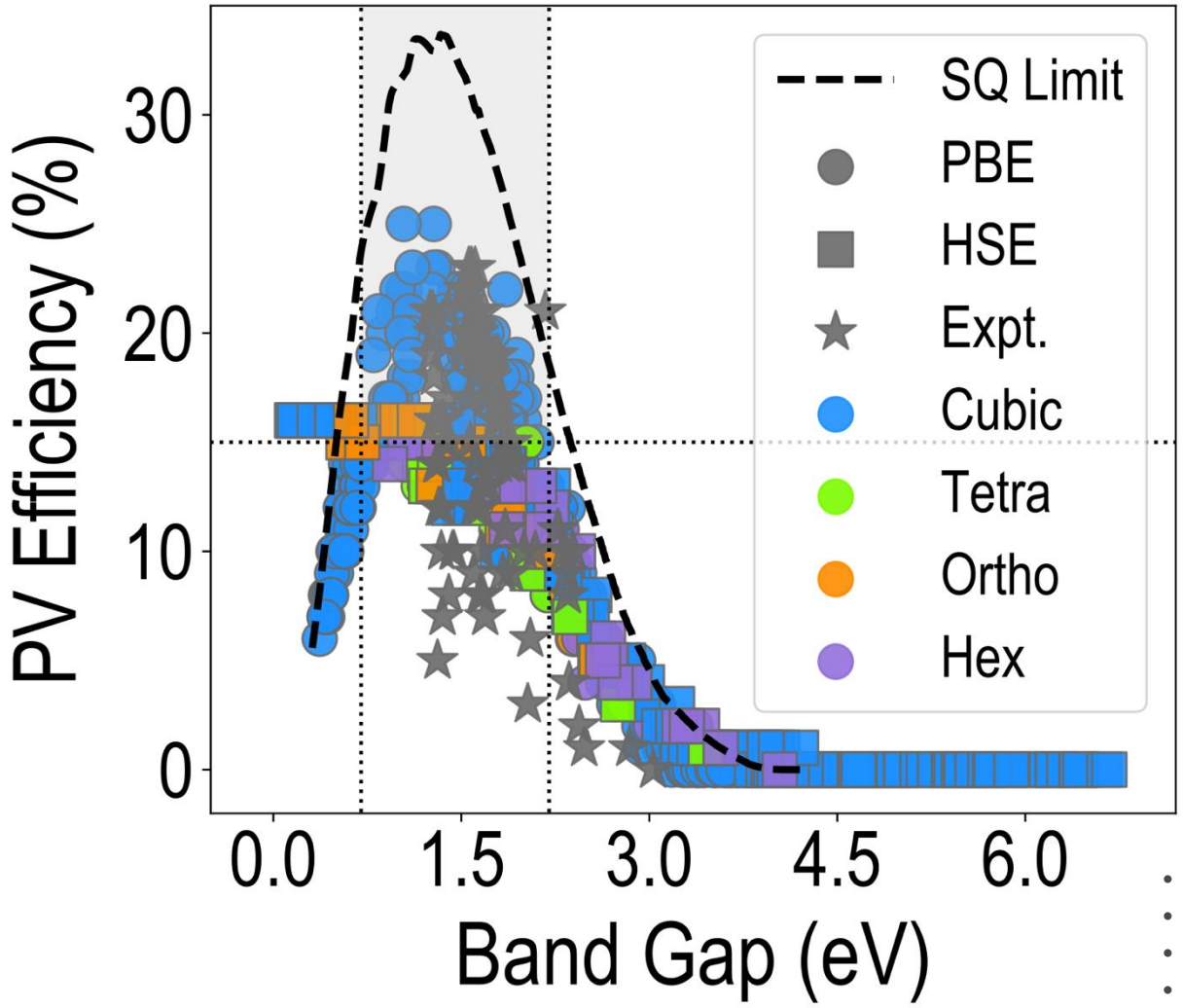


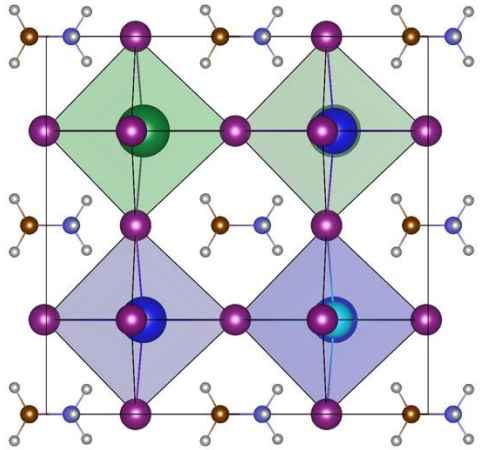
General formula: ABX_3

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

- ABX_3 perovskites, both purely inorganic (e.g., $CsPbI_3$) and hybrid organic-inorganic (e.g., $MAPbBr_3$, MA = methylammonium), are sought after for photovoltaic (PV) applications (PCE > 25%).
- Other applications: photodiodes, infrared sensors, electronics, LEDs, lasers, transistors, quantum information sciences.
- Issues: inherent instability, toxicity of Pb, defects.
- Stability and electronic, optical, and defect properties of ABX_3 compounds can be tailored by changing atoms, composition, structure, octahedral arrangements, point defects, interfaces, etc.
- Atom-composition-structure space is combinatorial → data-driven design based on high-throughput density functional theory (HT-DFT) and machine learning (ML).

High-Throughput DFT Dataset of Perovskite Properties



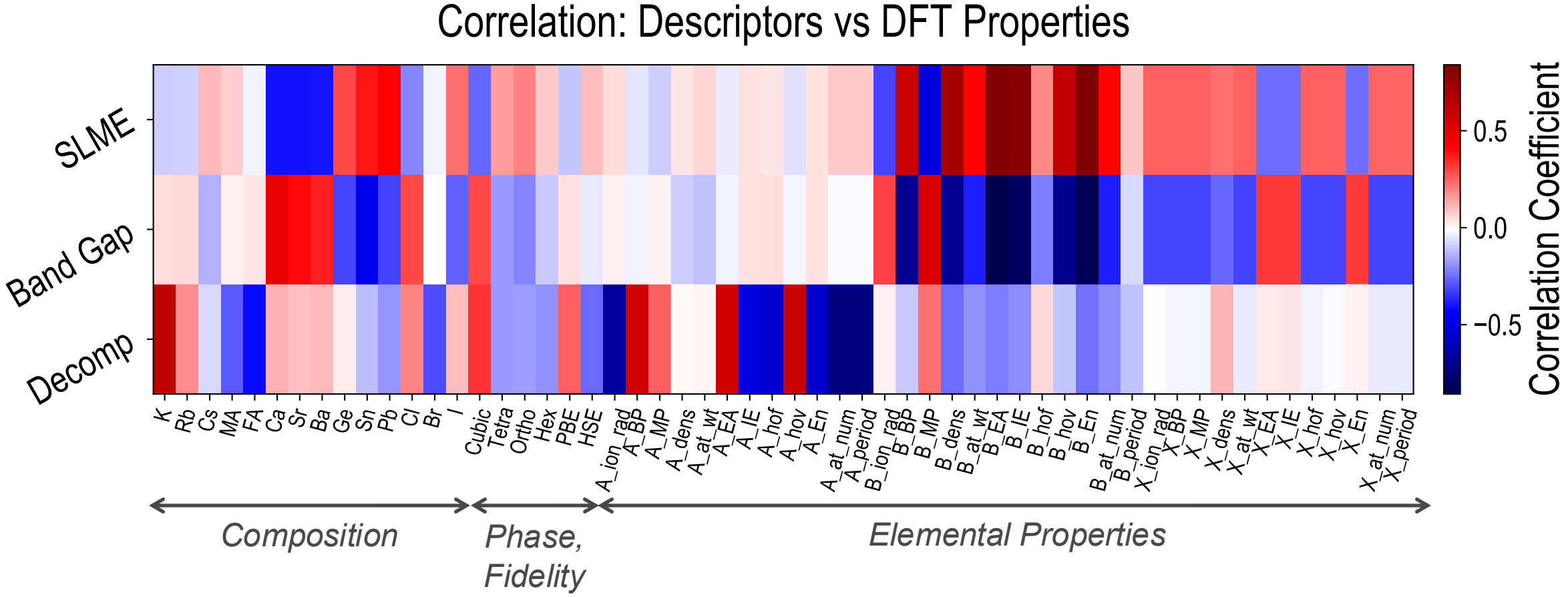


A	B	X
MA	Pb	I
FA	Sn	Br
Cs	Ge	Cl
Rb	Ba	
K	Sr	
	Ca	

- DFT dataset: > 1000 compounds, multiple functionals and properties.
- Descriptors for ML: composition & elemental properties of A/B/X species.

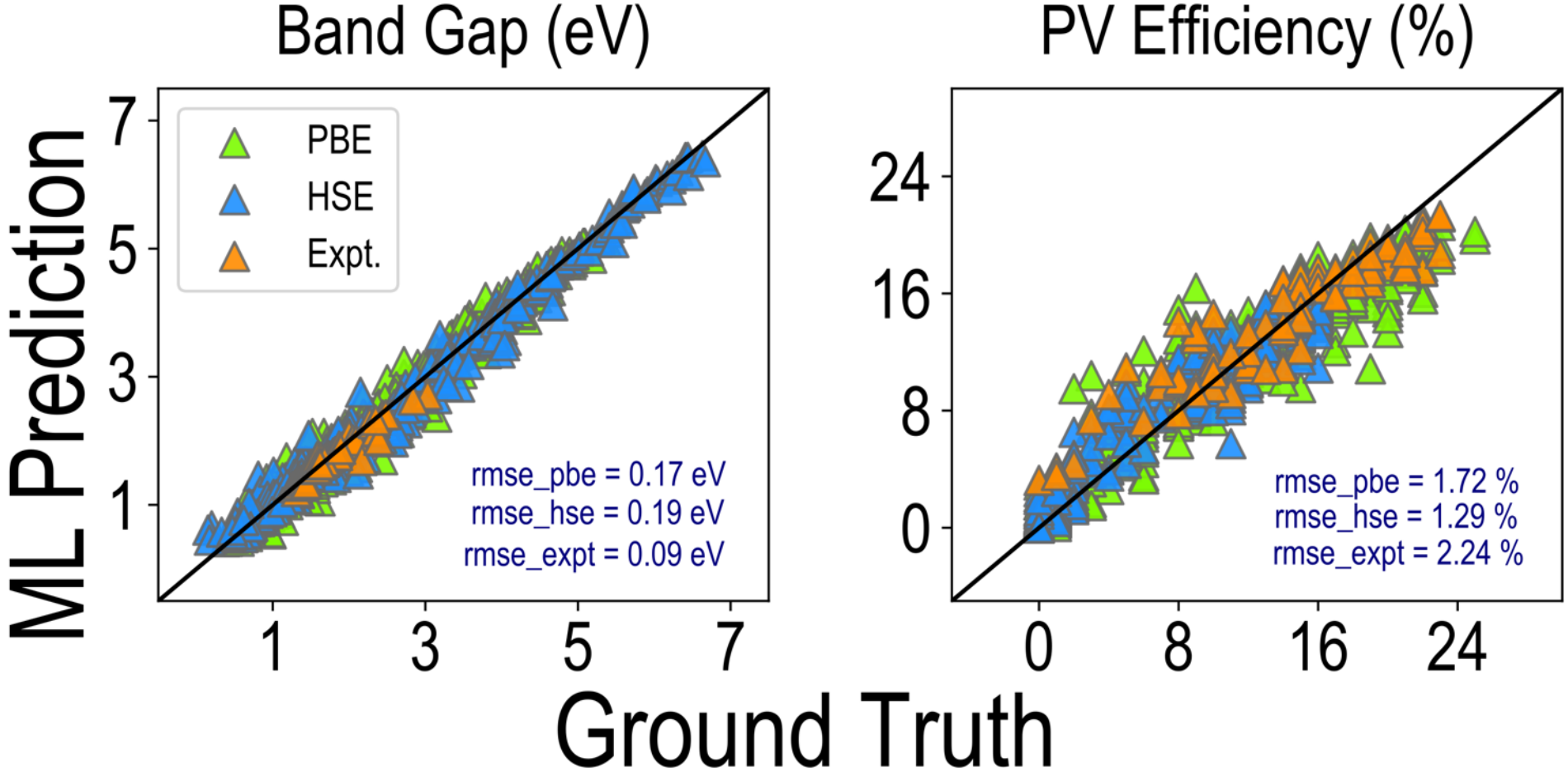
- *A.M.K. et al., Energy Environ. Sci., 15 (5), 1930-1949 (2022).*
- *J. Yang et al., Digital Discovery, 2, 856-870 (2023).*
- *J. Yang et al., J. Chem. Phys. 160, 064114 (2024).*
- *M. Biswas et al., Phys. Chem. Chem. Phys. 26, 23177-23188 (2024).*

Linear Correlation between DFT Properties and Input Descriptors



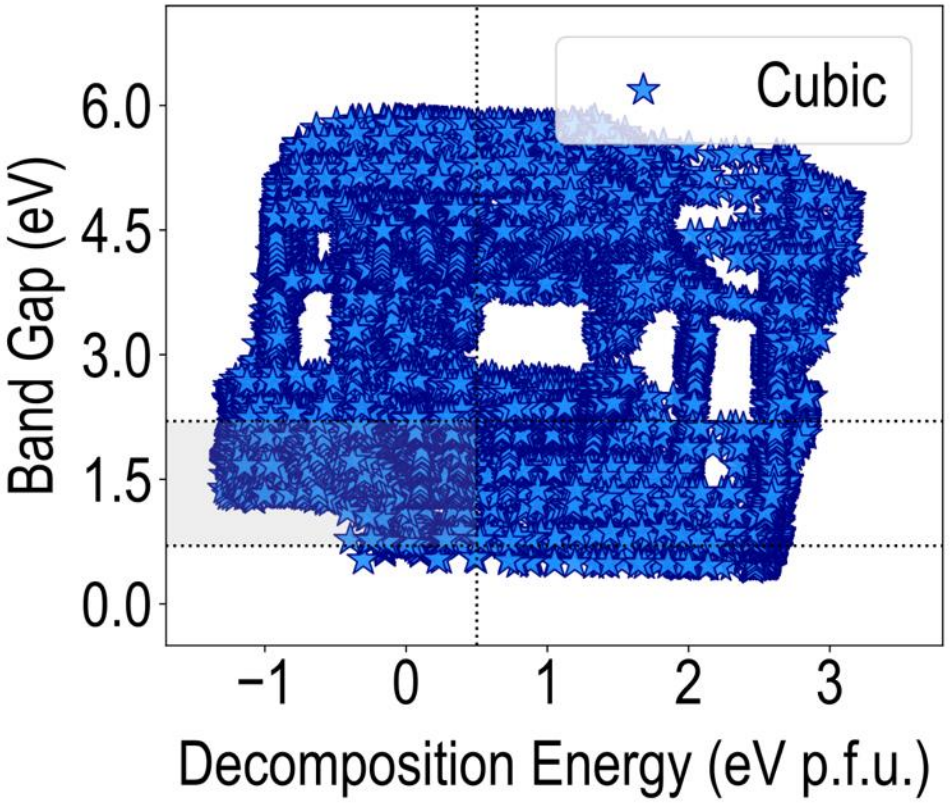
- *A.M.K. et al., Energy Environ. Sci., 15 (5), 1930-1949 (2022).*
- *J. Yang et al., MRS Bulletin., 47, 940–948 (2022).*
- *J. Yang et al., Digital Discovery, 2, 856-870 (2023).*
- *J. Yang et al., J. Chem. Phys.* 160, 064114 (2024)

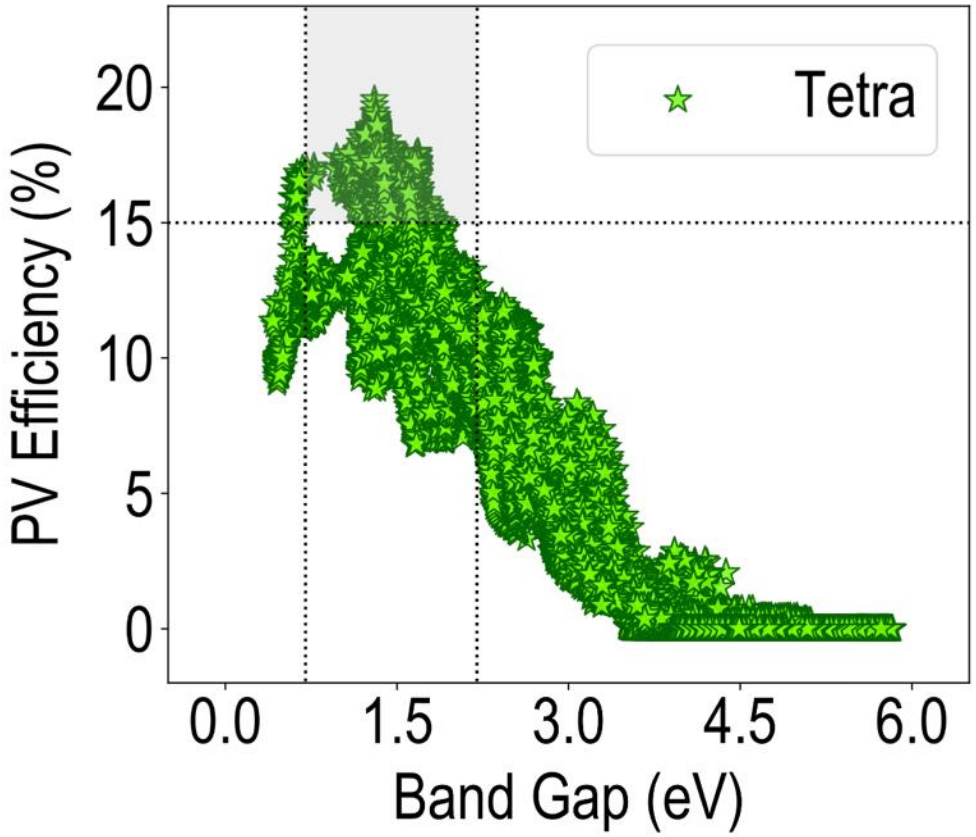
Descriptor-Based Multi-Fidelity Random Forest Regression Models



- *J. Yang et al., J. Chem. Phys. 160, 064114 (2024).*
- *M. Biswas et al., Phys. Chem. Chem. Phys. 26, 23177-23188 (2024).*

Enumeration → Expt. Prediction → Screening

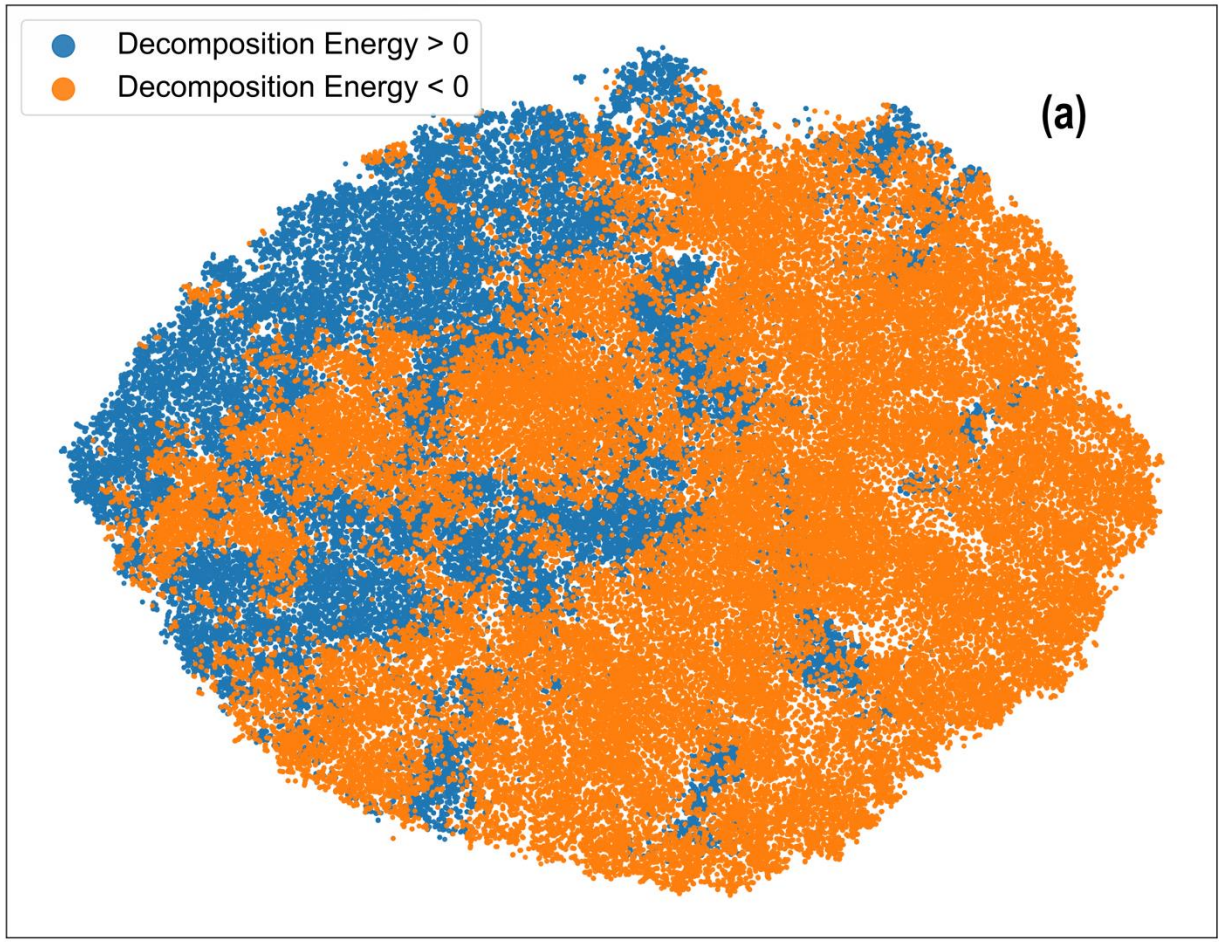


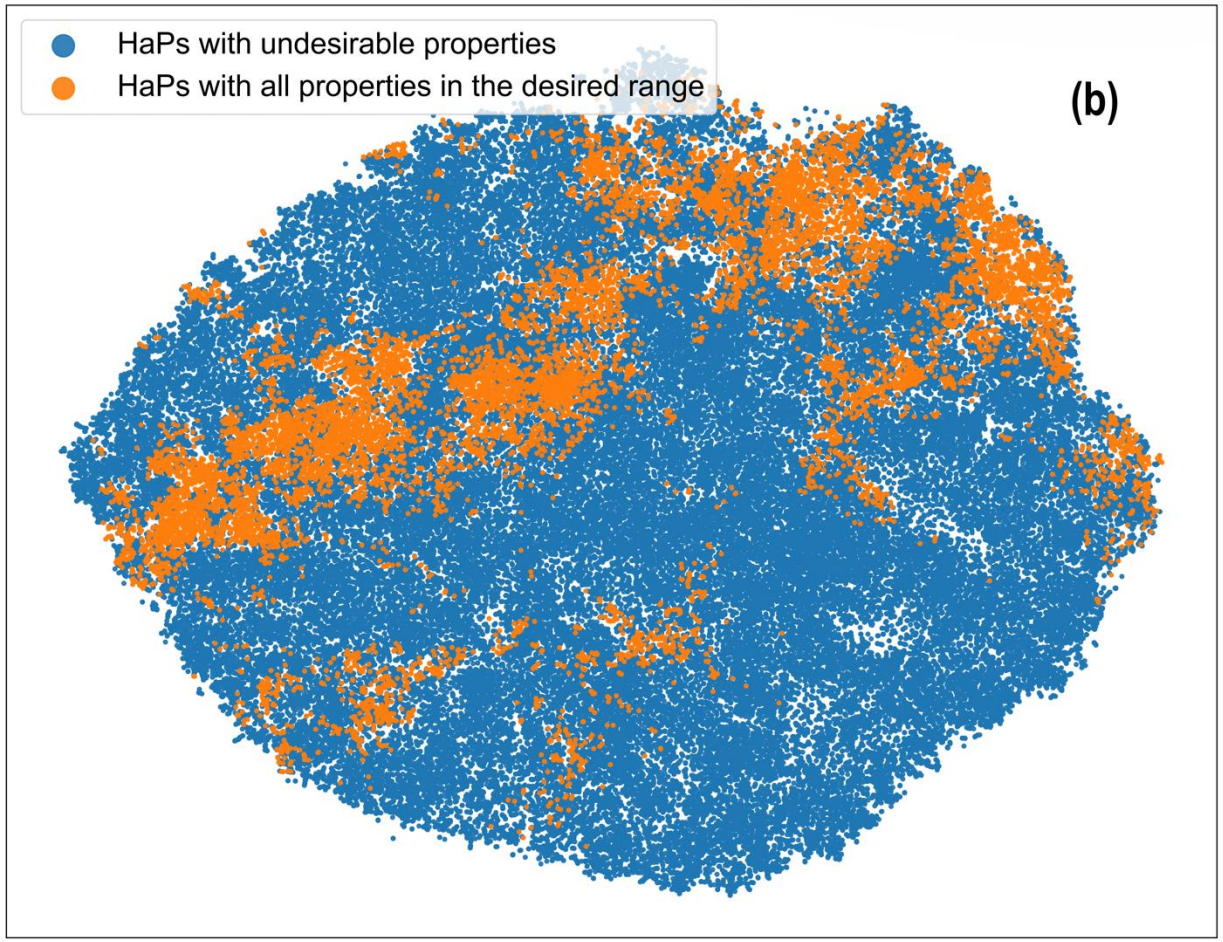


37,775 compositions in n/8 mixing fractions → 4 phases each = ~ 151,000 compounds → 3610 compounds with low ΔH, E_{gap} between 1 and 2 eV, PCE > 15%

- *J. Yang et al., Digital Discovery, 2, 856-870 (2023).*
- *J. Yang et al., J. Chem. Phys.* 160, 064114 (2024)

2D Projection example: t-SNE



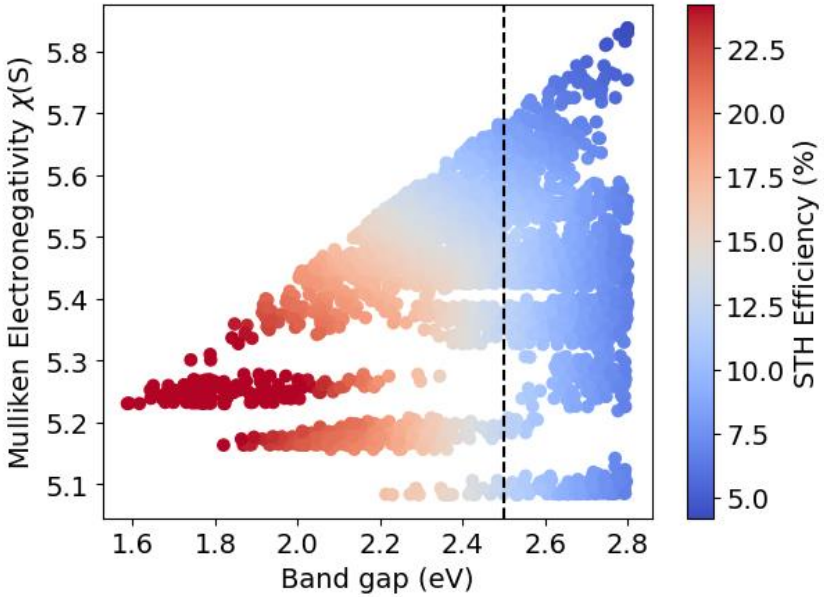


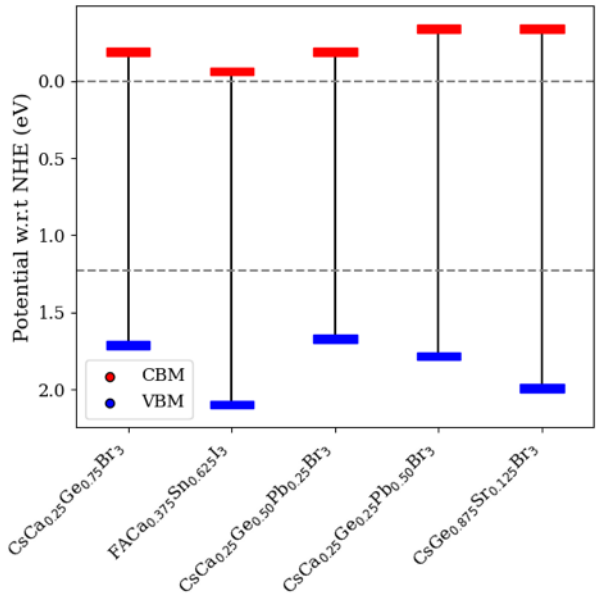
I. Agrawal, M. Biswas, and A. Mannodi-Kanakkithodi*, "Discovering Novel Halide Perovskites using Generative Adversarial Networks and Random Forest Regression", accepted, Computational Materials Science. (2025).

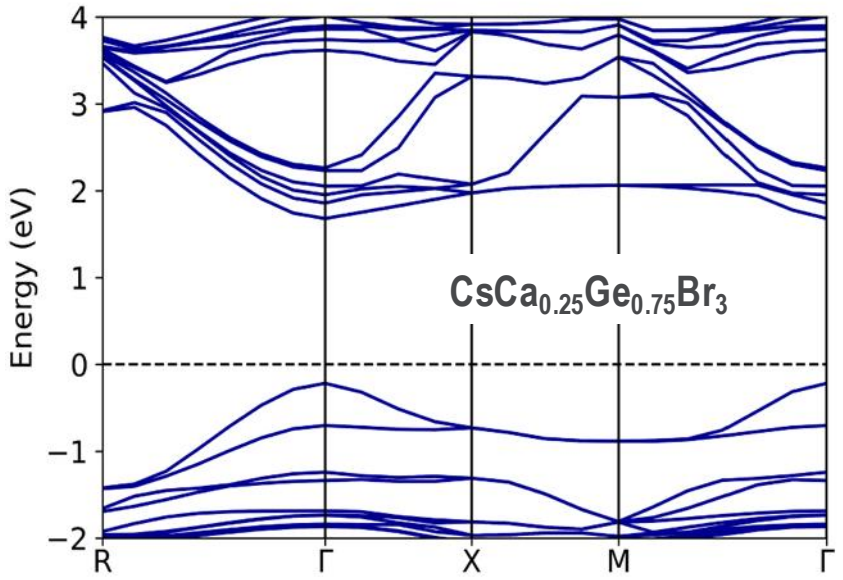
ML Predictions over 150,000 compounds → screening of promising compounds

Compound	ML-HSE ΔH (eV p.f.u.)	ML-Expt. E_{gap}	ML-Expt. PV efficiency (%)
$FA_{0.75}MA_{0.25}PbI_3$	−1.07	1.65	20
$Cs_{0.625}FA_{0.125}Rb_{0.25}PbI_3$	−0.55	1.66	19
$Cs_{0.5}FA_{0.375}Rb_{0.125}PbI_3$	−0.23	1.66	19
$Cs_{0.875}Rb_{0.125}PbI_3$	−0.38	1.38	19
$Cs_{0.625}MA_{0.25}Rb_{0.125}PbI_3$	−0.35	1.68	18
$FABa_{0.125}Ge_{0.875}I_3$	−0.99	1.52	20
$FAGe_{0.875}Sr_{0.125}I_3$	−0.98	1.57	19
$FACa_{0.125}Ge_{0.75}Sn_{0.125}I_3$	−0.98	1.54	19
$FACa_{0.125}Ge_{0.875}I_3$	−0.98	1.56	19
$FABa_{0.125}Ge_{0.75}Sn_{0.125}I_3$	−0.99	1.50	19

Using the same framework 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

M. Biswas et al., "Screening of Novel Halide Perovskites for Photocatalytic Water Splitting using Multi-Fidelity Machine Learning", Phys. Chem. Chem. Phys. 26, 23177-23188 (2024).

Interface with the data and models