

Designing Halide Perovskites for Solar Absorption using Machine Learning

Arun Mannodi Kanakkithodi

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



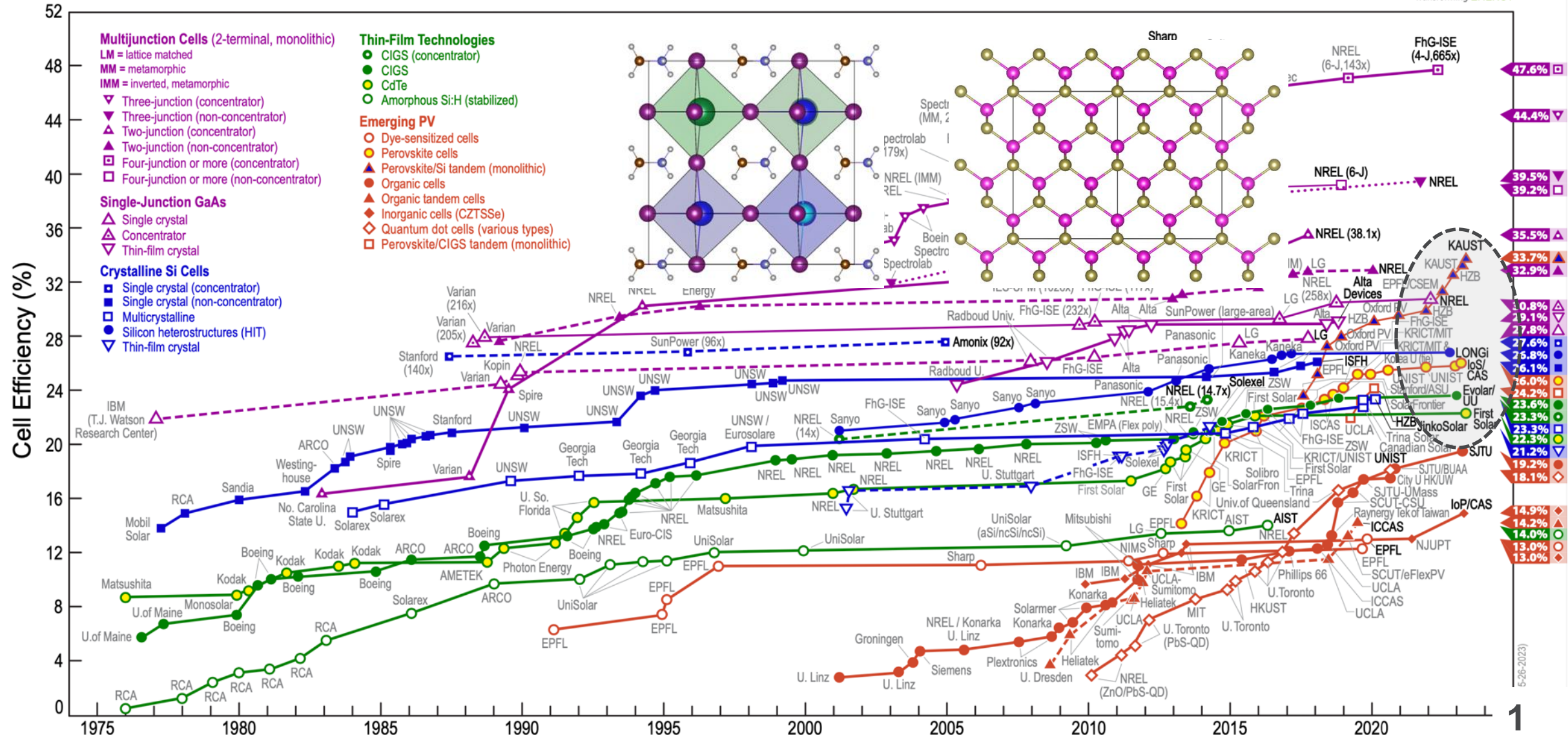
PURDUE
UNIVERSITY®

School of Materials Engineering

Research: Absorbers for Next-Gen Solar Cells

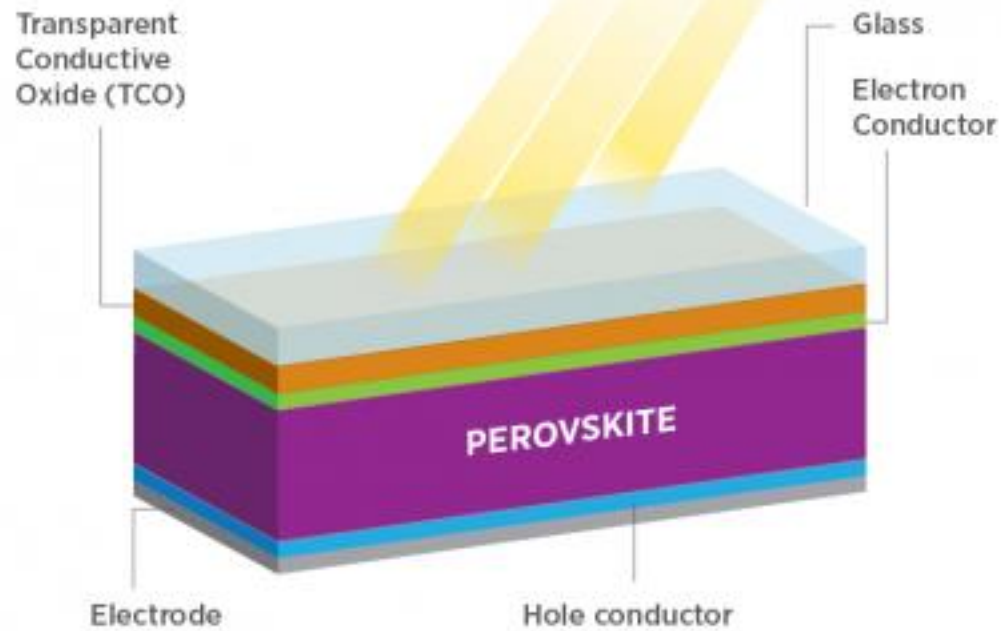
Best Research-Cell Efficiencies

<https://www.nrel.gov/pv/cell-efficiency.html>

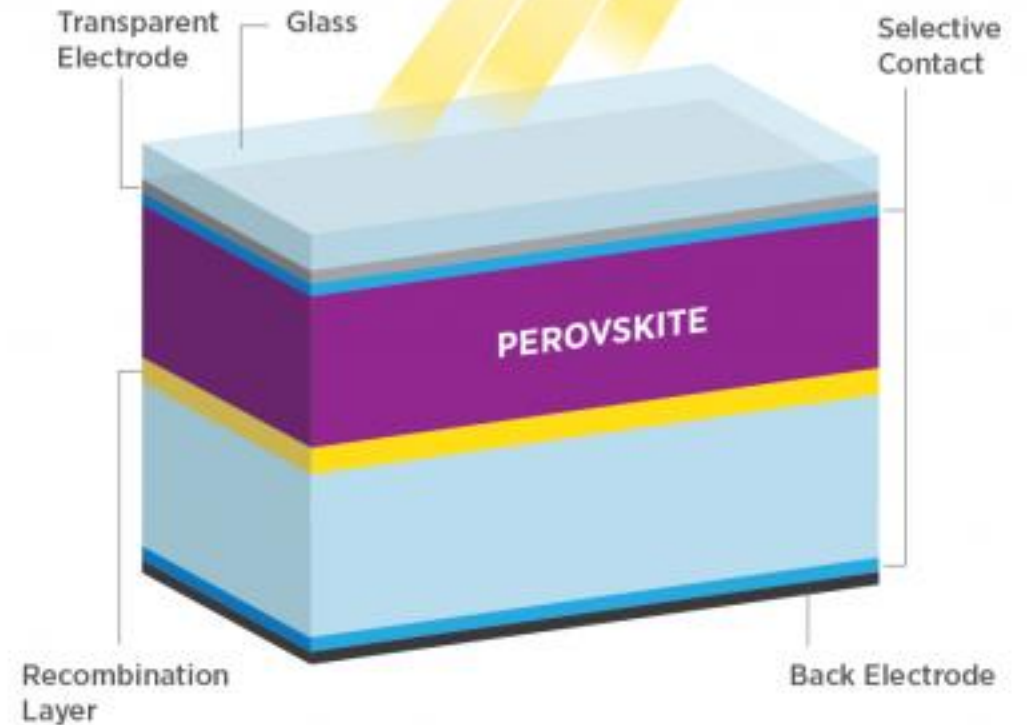


Perovskite Solar Cells

**THIN FILM
PEROVSKITE SOLAR CELL**

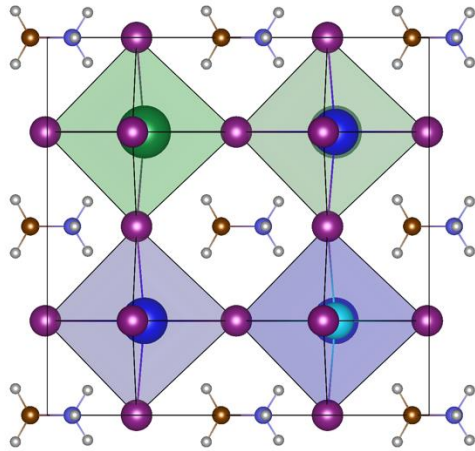


**PEROVSKITE ON SILICON
TANDEM SOLAR CELL**



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

Halide Perovskites

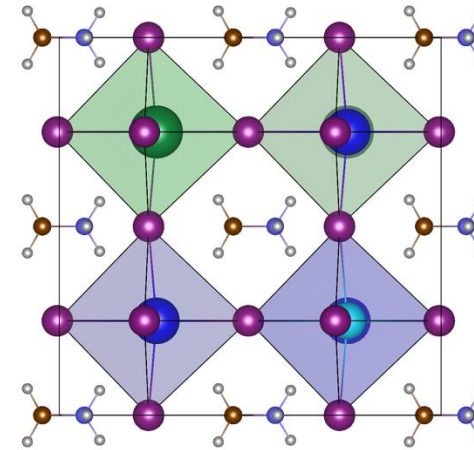
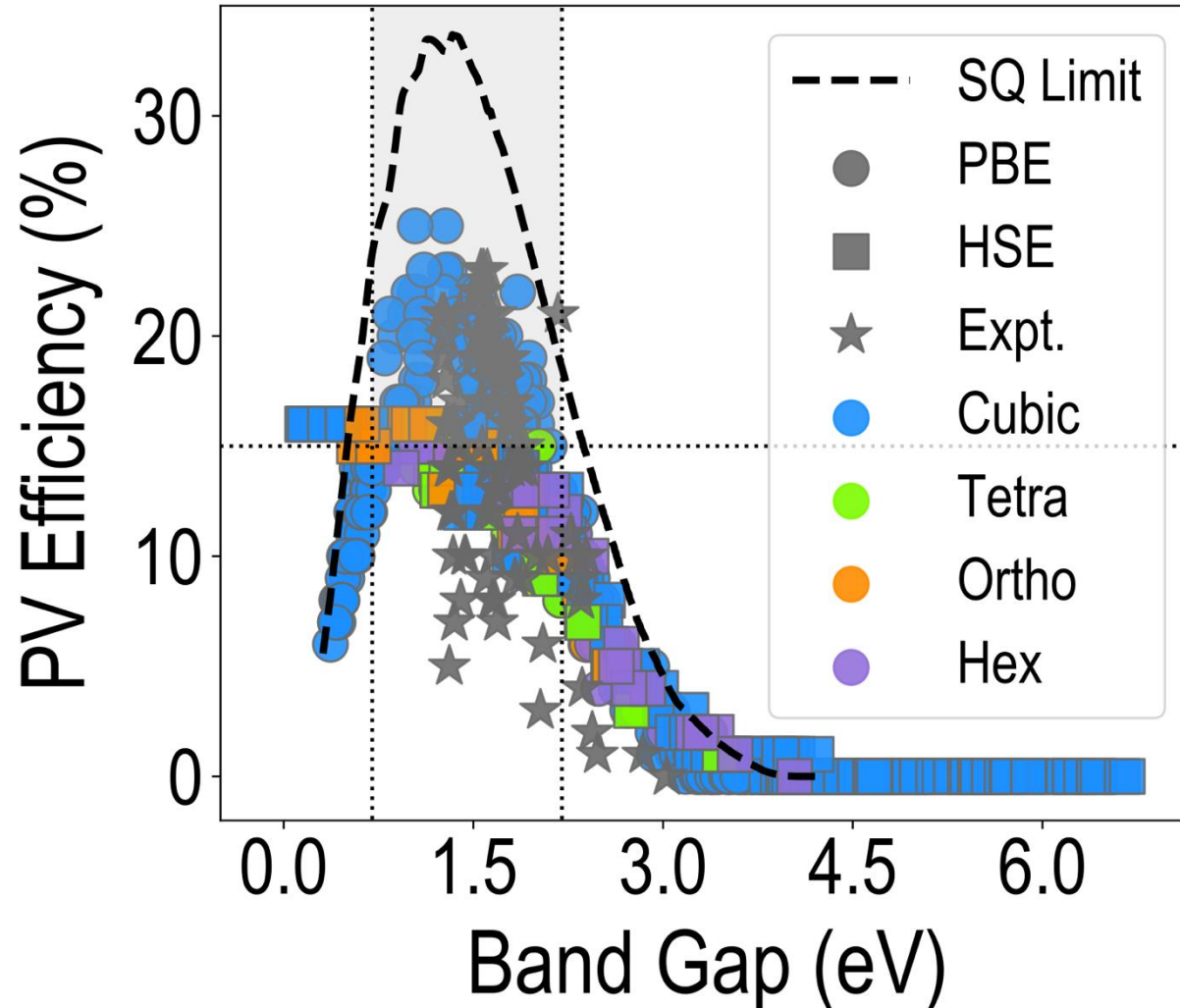


General formula: ABX_3

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 \rightarrow data-driven design based on high-throughput density functional theory (HT-DFT) and machine learning (ML).

High-Throughput DFT Dataset of Perovskite Properties

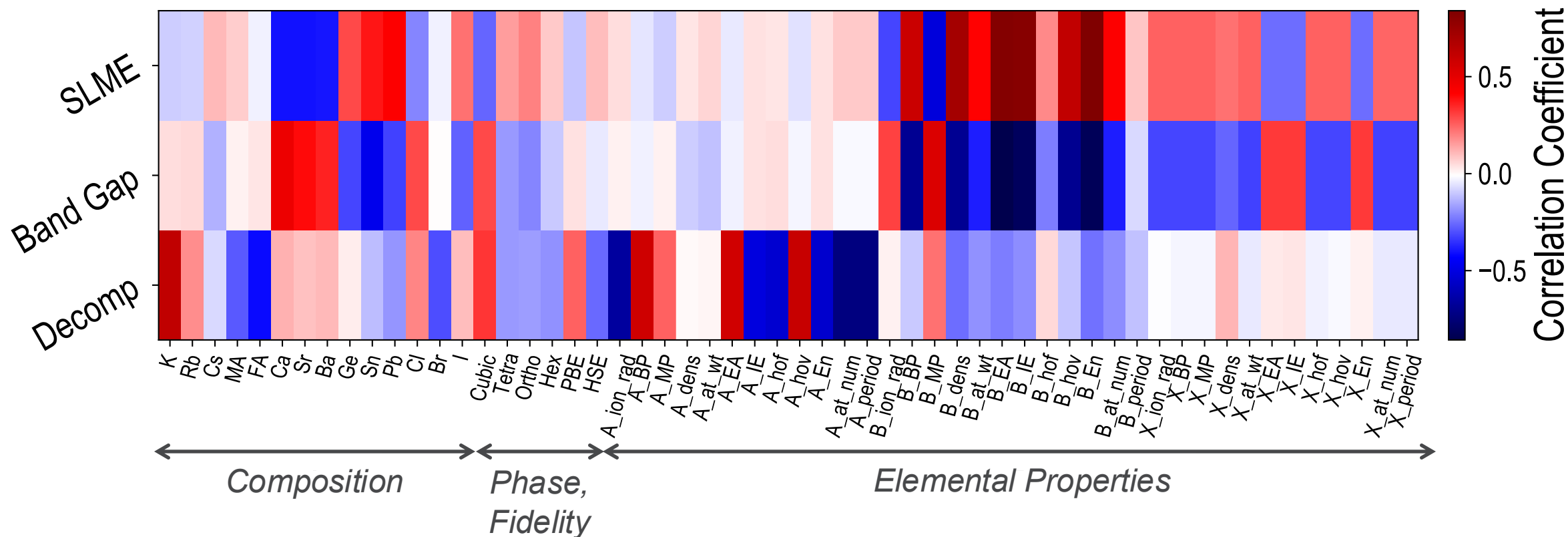


A	B	X
MA	Pb	I
FA	Sn	Br
Cs	Ge	Cl
Rb	Ba	
K	Sr	
	Ca	

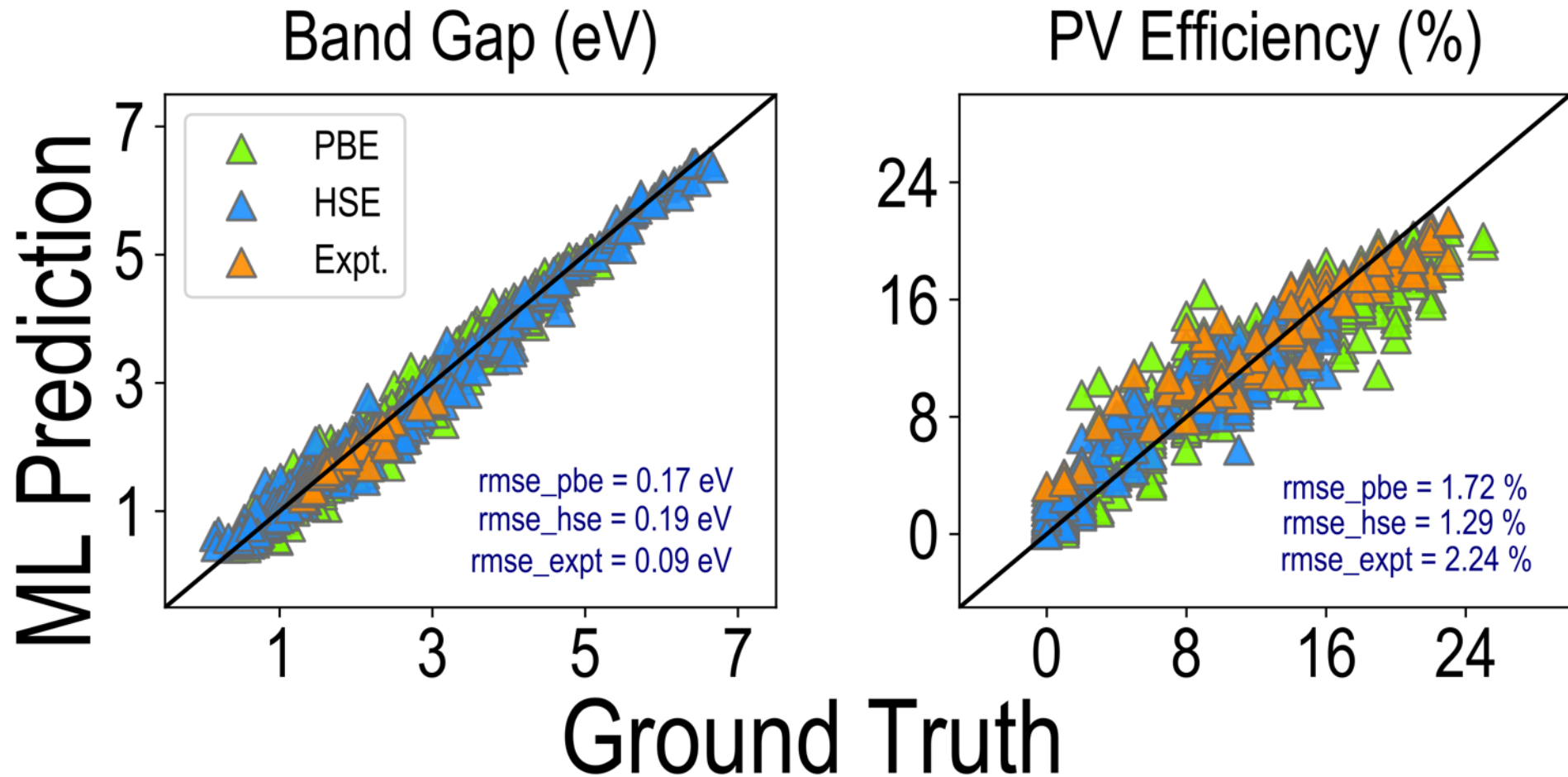
- DFT dataset: > 1000 compounds, multiple functionals and properties.
 - Descriptors for ML: composition & elemental properties of A/B/X species.
- A.M.K. et al., *Energy Environ. Sci.*, 15 (5), 1930-1949 (2022).
 - J. Yang et al., *Digital Discovery*, 2, 856-870 (2023).
 - J. Yang et al., *J. Chem. Phys.* 160, 064114 (2024).
 - M. Biswas et al., *Phys. Chem. Chem. Phys.* 26, 23177-23188 (2024).

Linear Correlation between DFT Properties and Input Descriptors

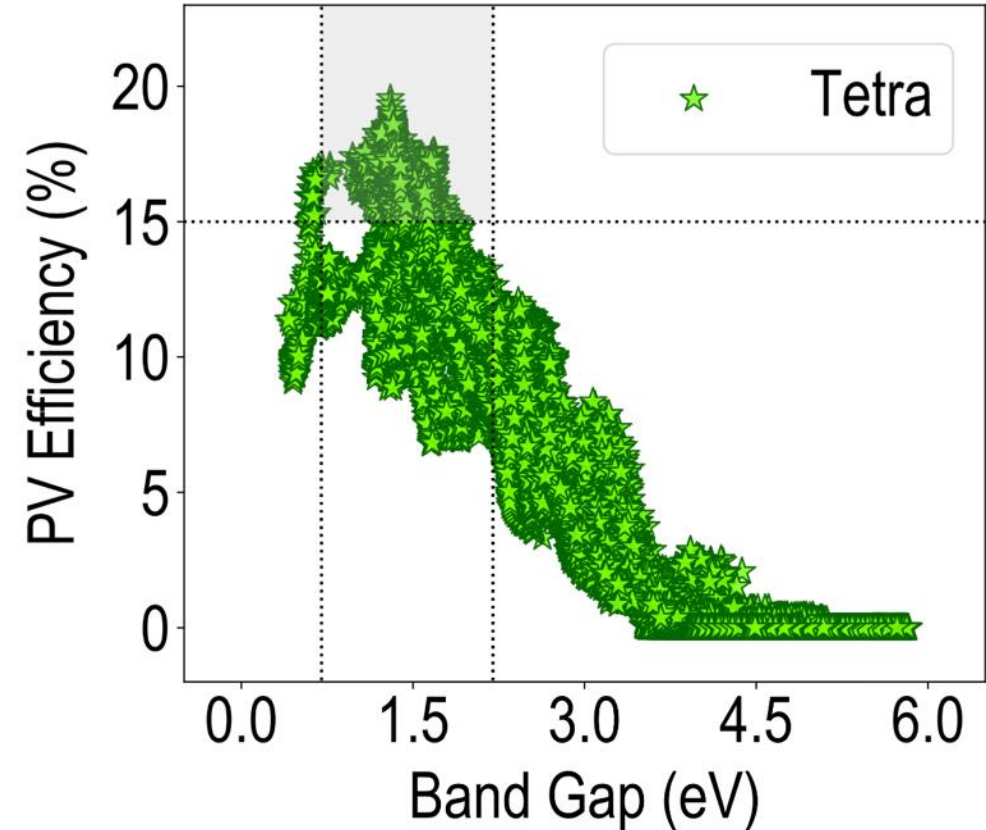
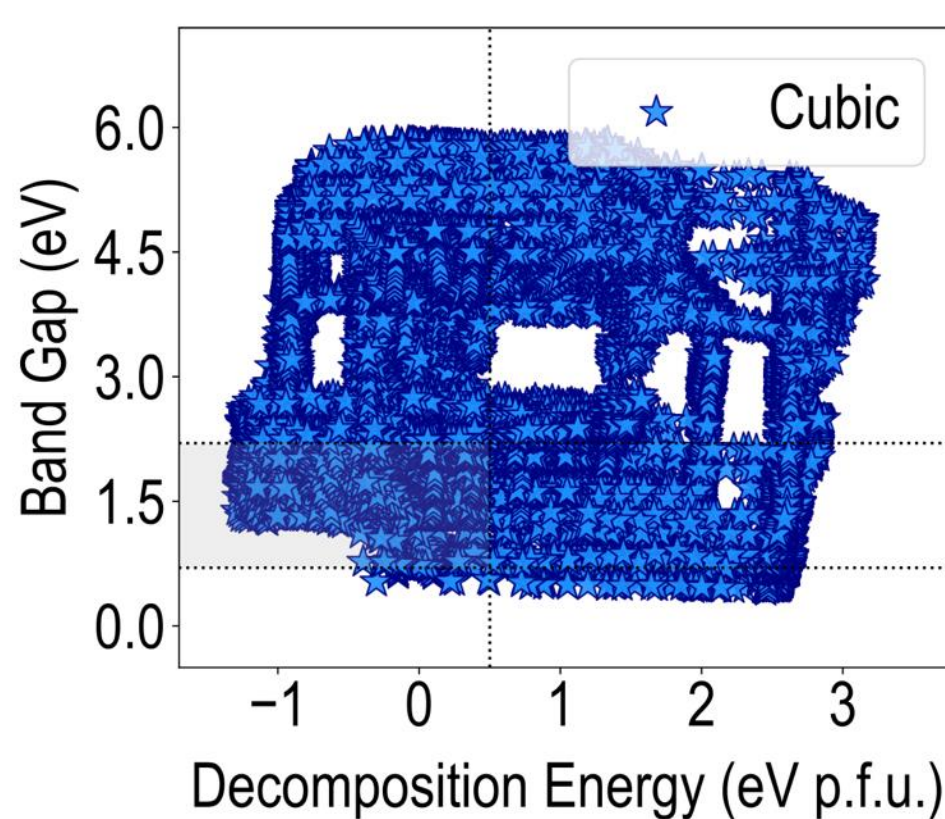
Correlation: Descriptors vs DFT Properties



Descriptor-Based Multi-Fidelity Random Forest Regression Models

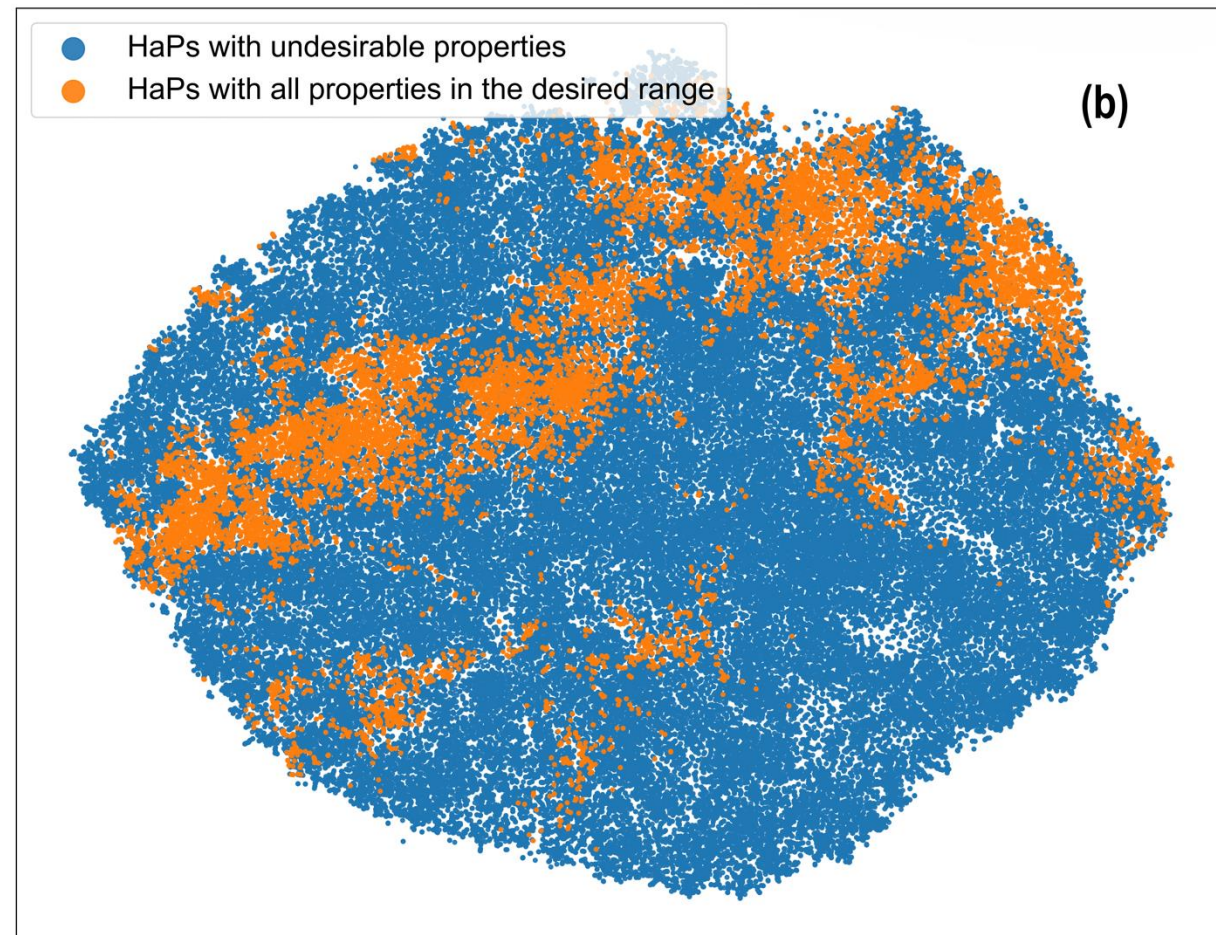
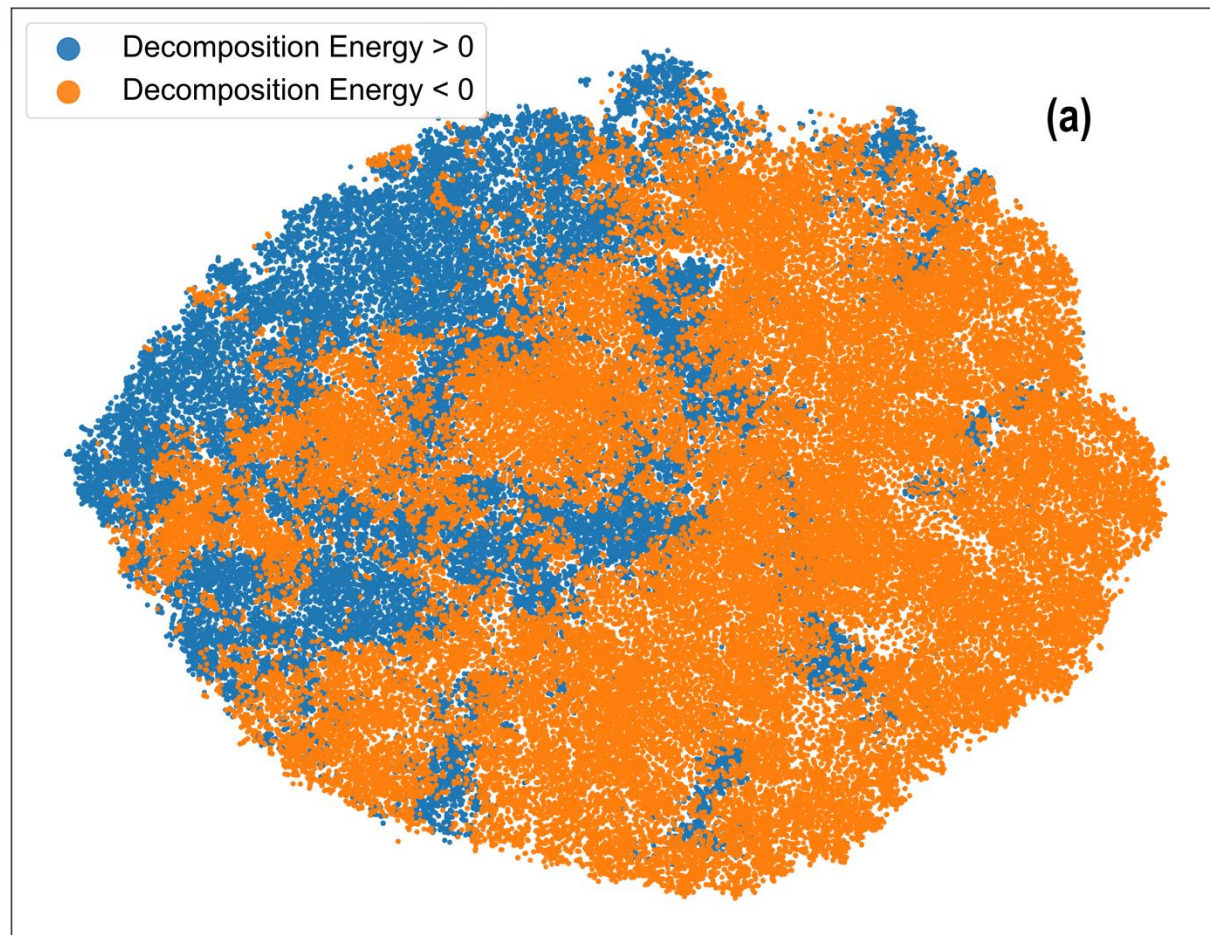


Enumeration → Expt. Prediction → Screening



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%

2D Projection example: t-SNE



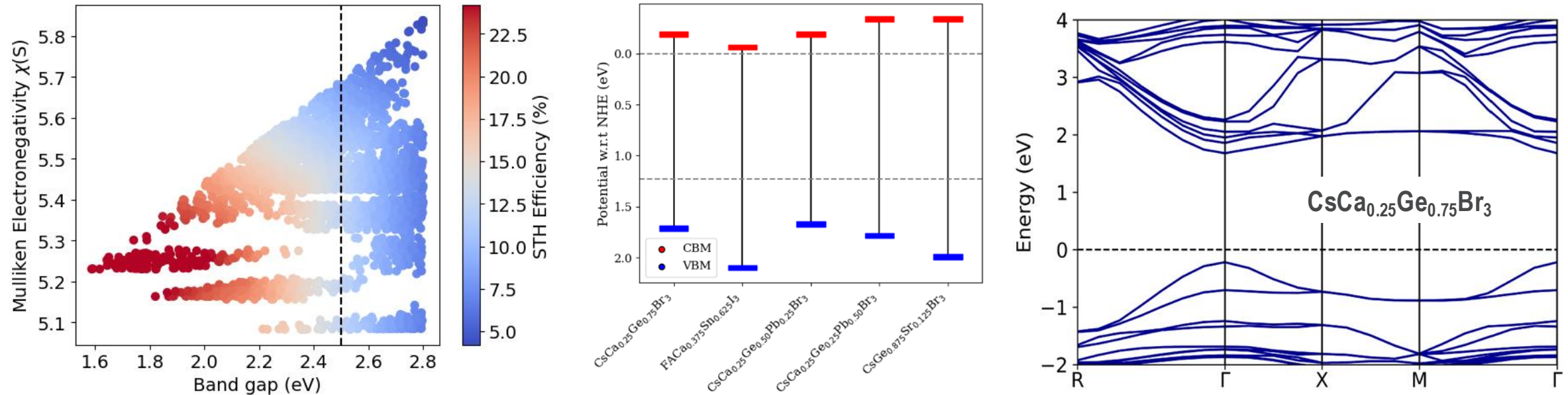
I. Agrawal, M. Biswas, and A. Mannodi-Kanakkithodi*, "Discovering Novel Halide Perovskites using Generative Adversarial Networks and Random Forest Regression", accepted, Computational Materials Science. (2025).

ML Predictions over 150,000 compounds

→ screening of promising compounds

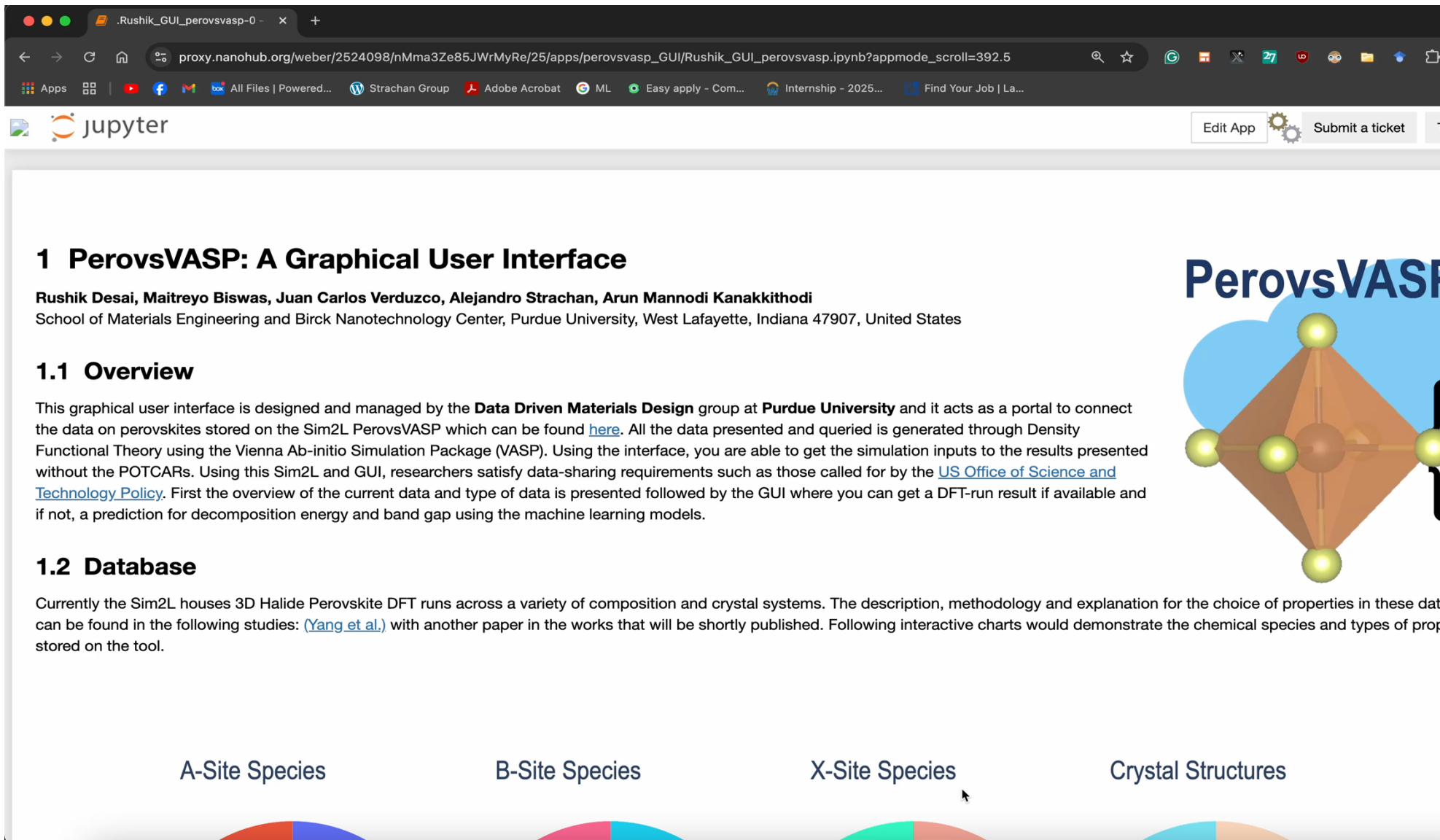
Compound	ML-HSE ΔH (eV p.f.u.)	ML-Expt. E_{gap}	ML-Expt. PV efficiency (%)
FA _{0.75} MA _{0.25} PbI ₃	−1.07	1.65	20
Cs _{0.625} FA _{0.125} Rb _{0.25} PbI ₃	−0.55	1.66	19
Cs _{0.5} FA _{0.375} Rb _{0.125} PbI ₃	−0.23	1.66	19
Cs _{0.875} Rb _{0.125} PbI ₃	−0.38	1.38	19
Cs _{0.625} MA _{0.25} Rb _{0.125} 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

Interface with the data and models



The screenshot shows a web browser window with the URL `proxy.nanohub.org/weber/2524098/nMma3Ze85JWrMyRe/25/apps/perovsvasp_GUI/Rushik_GUI_perovsvasp.ipynb?appmode_scroll=392.5`. The browser's address bar and tabs are visible. Below the browser window, the Jupyter logo is present. The main content area displays the title "1 PerovsVASP: A Graphical User Interface" followed by the authors' names: "Rushik Desai, Maitreyo Biswas, Juan Carlos Verduzco, Alejandro Strachan, Arun Mannodi Kanakkithodi". Below the names is the affiliation: "School of Materials Engineering and Birk Nanotechnology Center, Purdue University, West Lafayette, Indiana 47907, United States". The section "1.1 Overview" follows, containing a paragraph about the GUI's design and purpose. To the right of the text is a diagram of a perovskite crystal structure. Below the overview section is "1.2 Database", which describes the data stored in the Sim2L PerovsVASP database. At the bottom of the page, there are four labels: "A-Site Species", "B-Site Species", "X-Site Species", and "Crystal Structures", each with a corresponding colored semi-circle below it.

1 PerovsVASP: A Graphical User Interface

Rushik Desai, Maitreyo Biswas, Juan Carlos Verduzco, Alejandro Strachan, Arun Mannodi Kanakkithodi
School of Materials Engineering and Birk 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



PerovsVASP

