MATERIALS ENGINEERING SEMINAR

"Multi-fidelity Machine Learning for Perovskite Band Gap Predictions"

Panayotis Manganaris

Purdue MS Final Exam

Advisor: Professor Arun Kumar Mannodi Kanakkithodi

ABSTRACT

A wide range of optoelectronic applications demand semiconductors optimized for purpose. My research focused on data-driven identification of ABX_3 Halide perovskite compositions for optimum photovoltaic absorption in solar cells. I identified mixtures of candidate constituents at the A, B or X sites of the perovskite supercell which leveraged how mixed perovskite band gaps "bow" away from the linear interpolations predicted by Vegard's law of mixing to obtain a selection of stable perovskites with band gaps in the ideal range of 1 to 2.5 eV for visible light spectrum absorption.

I trained machine learning models on previously reported datasets of halide perovskite band gaps based on first principles computations performed at different fidelities. The primary objective of these models was to predict the perovskite band gap using the composition and inherent elemental properties as descriptors, eventually leading to accurate prediction and screening across the much larger chemical space from which the data samples were drawn.

I utilized a recently published density functional theory (DFT) dataset of more than 1300 perovskite band gaps from four different levels of theory, added to an experimental perovskite band gap dataset of ~100 points, to train random forest regression (RFR), Gaussian process regression (GPR), and Sure Independence Screening and Sparsifying Operator (SISSO) regression models, with data fidelity added as one-hot encoded features. I found that RFR yields the best model with a band gap root mean square error of 0.12 eV on the total dataset and 0.15 eV on the experimental points. SISSO provided compound features and functions for direct prediction of band gap, but errors were larger than from RFR and GPR. Additional insights gained from Pearson correlation and Shapley additive explanation (SHAP) analysis of learned descriptors suggest the RFR models performed best because of (a) their focus on identifying and capturing relevant feature interactions and (b) their flexibility to represent nonlinear relationships between such interactions and the band gap. The best model was deployed for predicting experimental band gap of ~40,000 hypothetical compounds, based on which we screened ~3000 stable compounds with band gap predicted to be between 1 and 2.5 eV at experimental accuracy.

Date: Thursday, June 8, 2023
Time: 2:00 P.M.
Place: ARMS 1028 or via the link: https://purdue.webex.com/join/amannodi

PURDUE UNIVERSITY | School of Materials Engineering