AI SEMICONDUCTORS

Laura Jacoby, Xiaofeng Xiang, Robert Biegaj

Introduction

Goal: Use DFT data to train ML models that predict the formation energy and charge transition levels of any atomic impurity at any site in any group II-VI, III-V or IV-IV semiconductor.

Techniques: Neural Network, Random Forest Regression, Kernel Ridge Regression

Original descriptors and properties of interest
- 8 properties of interest
  - Formation energy for dopant in (A-rich) and (B-rich site)
  - Transition energy levels: (+3/+2), (+2/+1), (+1/0), (0/-1), (-1/-2), (-2/-3)
- 94 original descriptors
  - 5 properties of AB compound
  - 27 elemental properties of atom A
  - 27 elemental properties of atom B
  - 27 elemental properties of atom impurity atom
  - 8 descriptors to characterize the defect coordination environment using Coulomb matrix elements
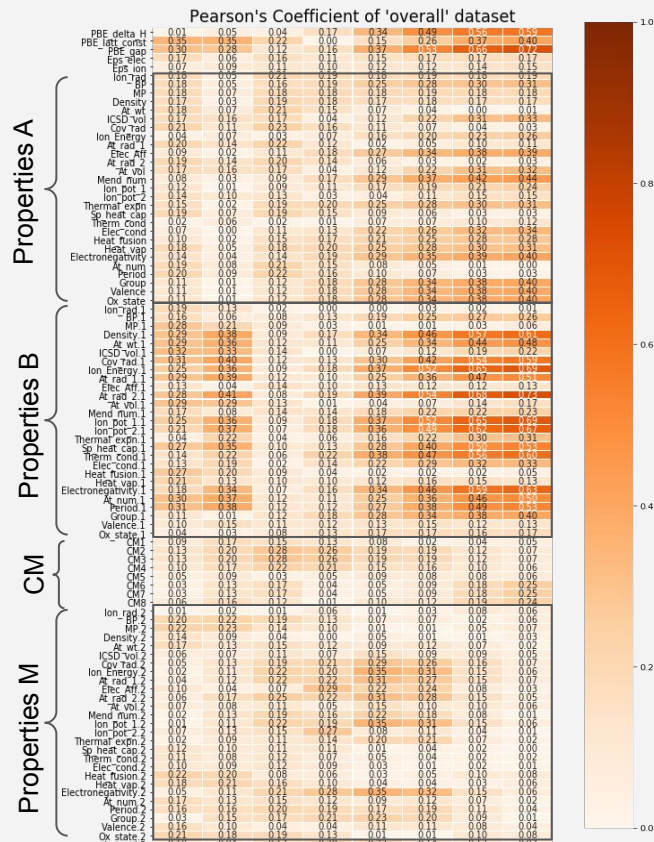
Training Dataset
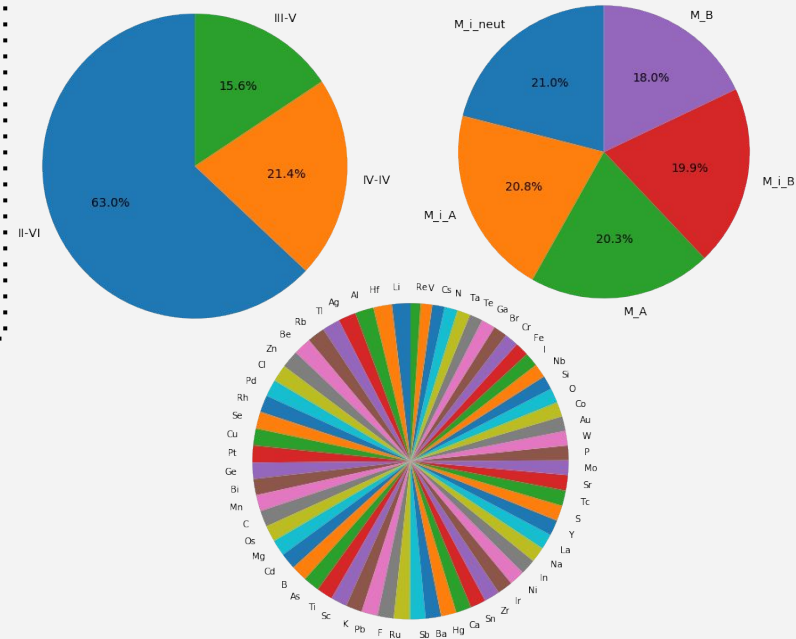
Predicting dataset (~12,500 points).
- 32 semiconductors (AB)
  - 3 types
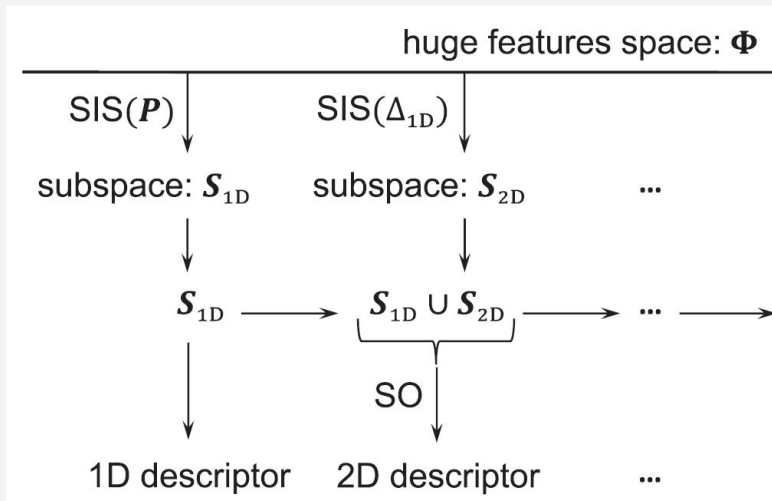- 63 impurity dopants
- 5 dopant defect sites
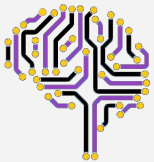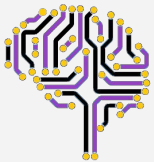
Descriptor Selection

Pearson Correlation

Pearson's Coefficient of 'overall' dataset

SISSO/LASSO

- SISSO Feature Expansion and selection[1].

- Operator: ''(+)(-)(*)(/)(exp)(log)(^-1)(^2)(^3)(sqrt)(cbrt)(|-|)'

- Lasso filters top 1200 SISSO features (80 ~ 130 features left)

[1]Ouyang, Ruhai, et al. *Phys. Rev. Mat.* 2.8 (2018): 083802

Model exploration, optimization, and training

Outlier Detection

Primary Component Analysis (PCA)

PCA Outliers
- 51% III-V
- 38% IV-IV
- 11% II-VI
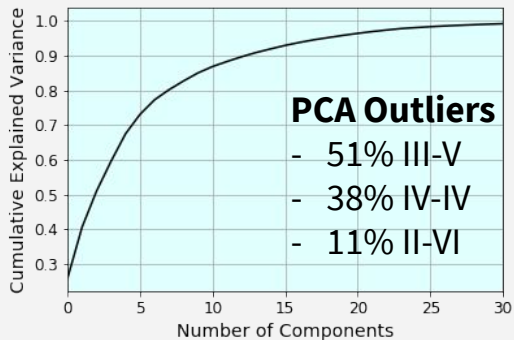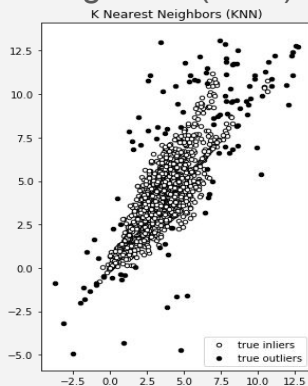
K Nearest Neighbors (KNN)
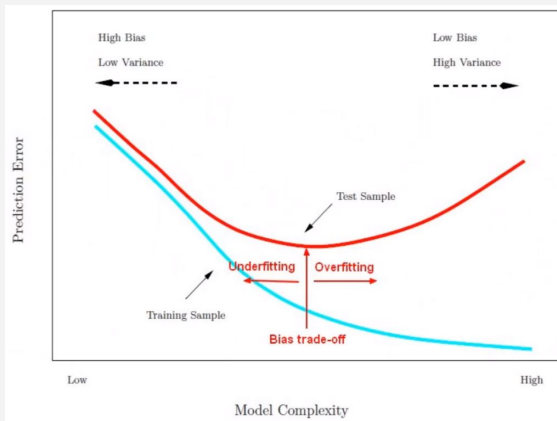

KNN Outliers
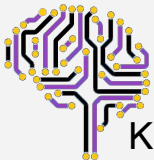- 42% III-V
- 24% IV-IV
- 34% II-VI

Hyperparameter Tuning

- Random Search CV
- Grid Search CV
- Bayesian Optimization
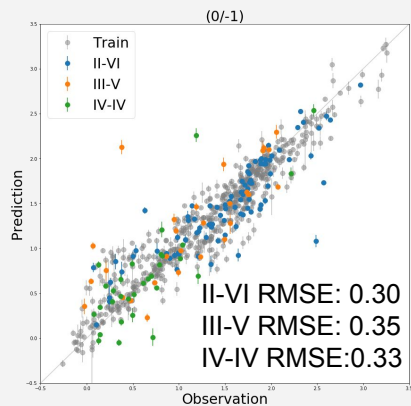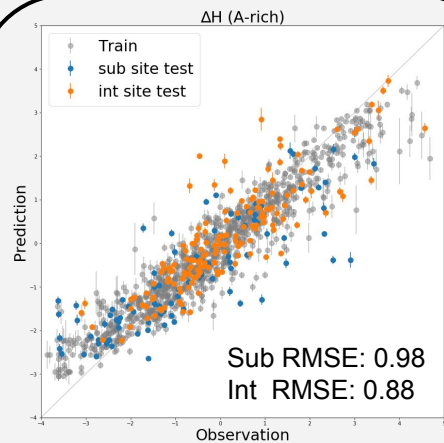 - Minimize loss for test RMSE *and* difference between test/train rmse

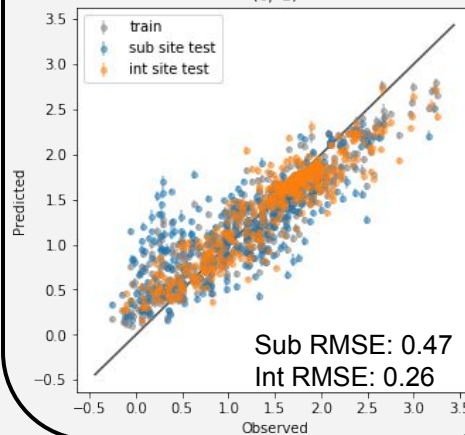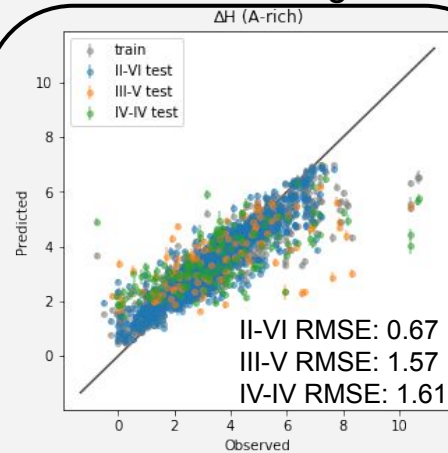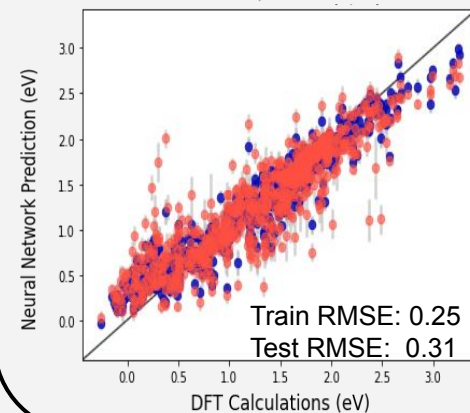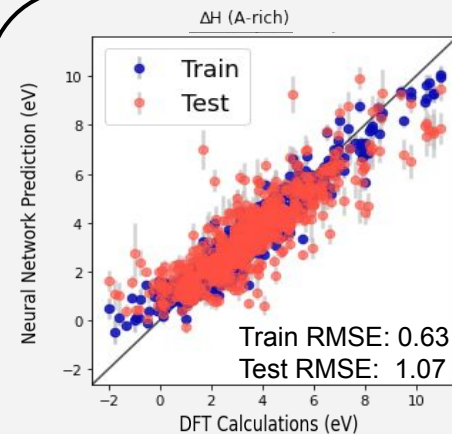## Balancing bias/variance and underfitting/overfitting

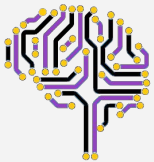# Model performance

Kernel Ridge Regression

Random Forest Regression
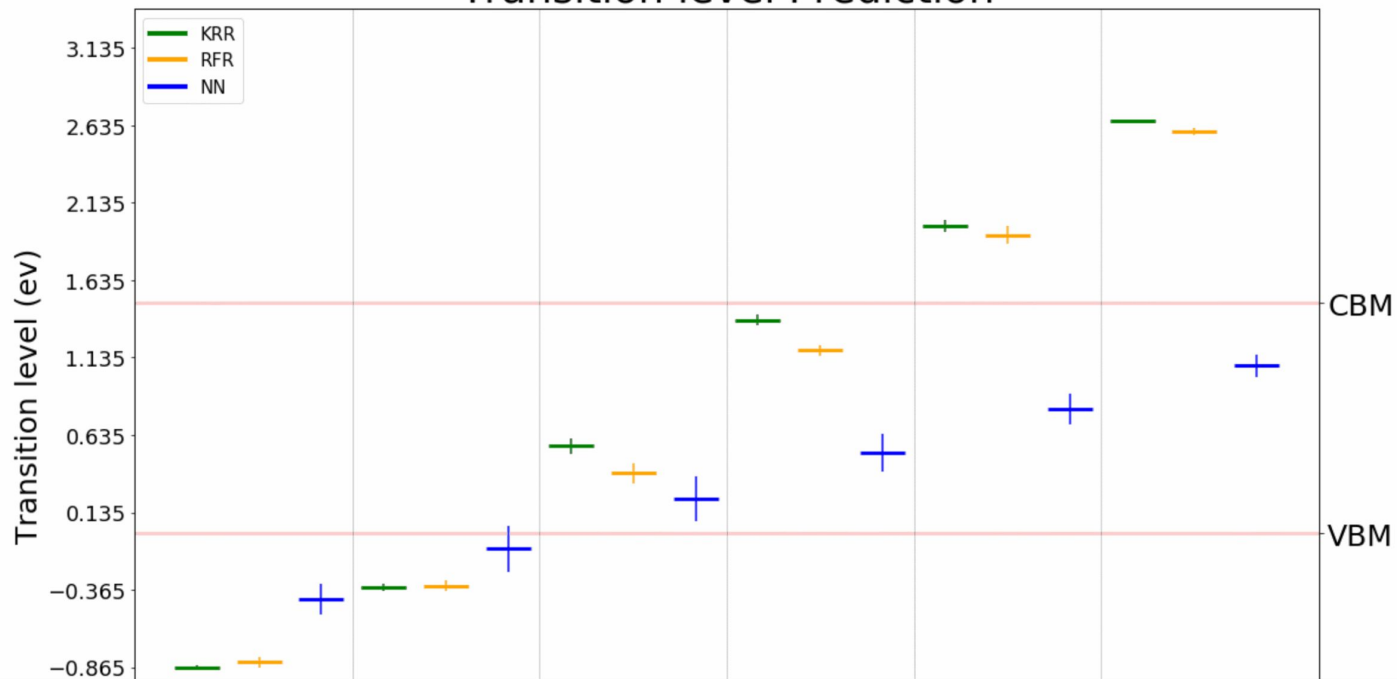

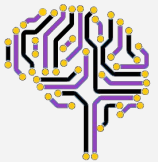## Neural Network

Prediction Tool

Impurity | Ag ▾
AB | AlAs ▾
Site | M_A ▾

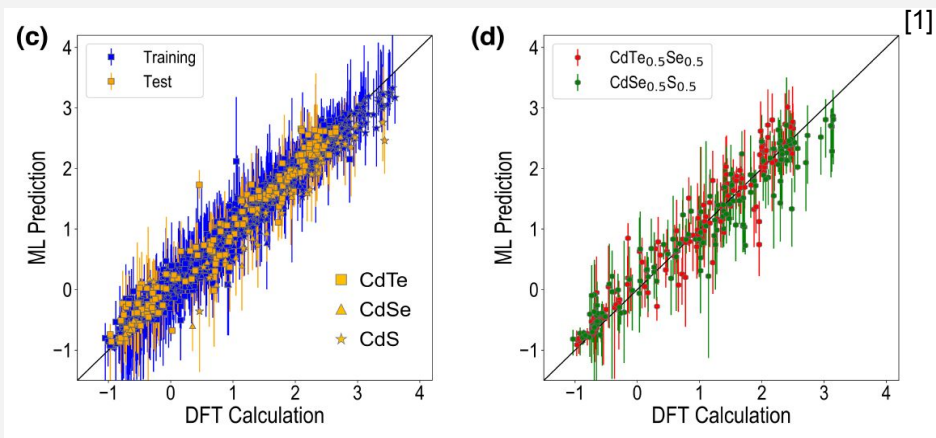We don't have DFT values for this semiconductor/impurity compound yet
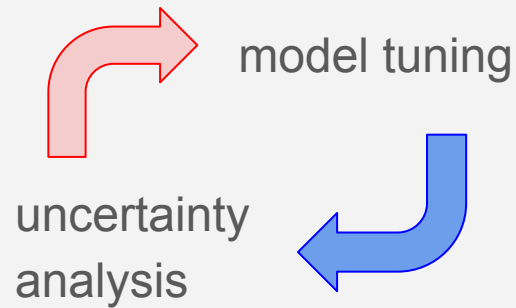

Transition level Prediction

Test how models predict on alloys

Prepare work for publication!

(c) — Training / Test; CdTe, CdSe, CdS; ML Prediction vs DFT Calculation
(d) — $CdTe_{0.5}Se_{0.5}$, $CdSe_{0.5}S_{0.5}$; ML Prediction vs DFT Calculation [1]

model tuning

uncertainty analysis

[1] Kanakkithodi, Toriyama et al. *arXiv:1906.02244v1* (2020)
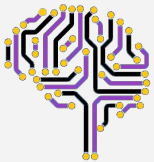
Special thanks to...

Argonne National Lab:
- Dr. Maria Chan
- Dr. Arun Mannodi Kanakkithodi

UW DIRECT:
- Dr. Dave Beck
- Dr. Dan Schwartz
- Kelly Thornton

https://github.com/lmjacoby/ai_semiconductors