SAND2022-1512D

Sandia Academic Alliance Fall 2021 University of Illinois LDRD Virtual Poster Session

"Opening the "Black Box": An Experimentally-Validated Explainable Machine Learning Framework for Materials Design and Discovery"

Authors: Grace Lu (PhD Candidate in Materials Science and Engineering), Vitalie Stavila (Sandia, Energy Nanomaterials Department), Dallas Trinkle (U of IL, Dept of Materials Science and Engineering)

Introduction / Motivation

- Materials that diffuse hydrogen slowly are useful for storing hydrogen. Conversely, materials that diffuse hydrogen quickly are important for use in hydrogen **fuel cells**.
- Hydrogen's small mass results in intrinsically quantum mechanical effects, namely a large contribution to diffusion from zero-point vibrational energy.

- Previously, machine learning models have successfully been used to predict diffusion in metals [1, 2].
- Most machine learning models are a 'black box,' but we want to enable **expert-guided extrapolation** beyond our dataset by revealing relationships between metal properties and hydrogen diffusion.
- Our findings can therefore be expanded beyond the metals and alloys in our dataset.

Approach

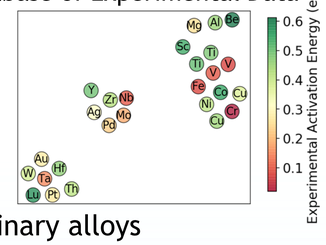
- **Feature selection: 196 total features**
  - Previous researchers [1] have used the Wolverton group **MAGPIE** features database [3] to predict vacancy-based diffusion in metals. This database includes a variety of characteristics such as atomic data, crystal structure, electrical properties, and elastic constants.
  - We further added features from theoretical models of diffusion, which were previously used successfully in a machine learning model for C, O, and B interstitial diffusion in metals [3].
  - **Electrical thermal conductivity** is correlated with the solubility activation energy of hydrogen [4].

- **Model selection: ElasticNET**
  - Linear model which combines properties of LASSO and ridge regression
  - Cuts down the number of features and keeps coefficients small
  - Very similar to normal least-squares linear regression, but is more stable against overfitting
  - **Explainable:** coefficients for the features give us an idea of their relative importance
  - Helps us generate a physical understanding for why properties may be important in hydrogen diffusivity

Current Status/ Results (If any)

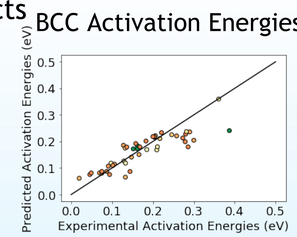
- Gathered a database of experimental hydrogen diffusion data, including data collected at Sandia:
  - *50* metals
  - *88* binary random alloys
  - Contains only high temperature (>300 K) results
  - All three isotopes of hydrogen included

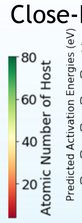
Database of Experimental Data



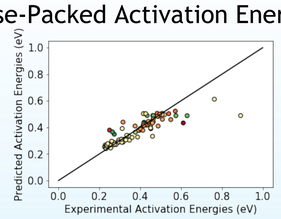
- Hydrogen diffusion activation energy predictions for metals and binary alloys
- Fit two separate models for BCC and close-packed metals as BCC metals have more prominent quantum mechanical effects

BCC Activation Energies





Close-Packed Activation Energies



Challenges

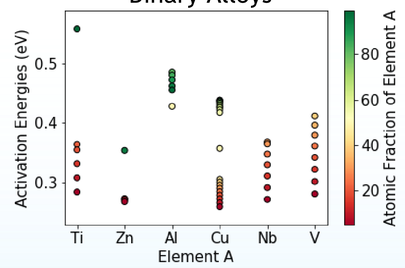
- Large number of features and relatively **small amount of data** (especially by machine learning standards)
- Limits the types of models that can be used since S more flexible and accurate models learn 'too well', so they can't generalize to new data

- Hydrogen-metal interactions are more difficult to predict due to their **highly quantum-mechanical** nature.
- Previous models, which focused on larger interstitials, cannot simply be applied here.

Next Steps/ Future Work

- Compare predictions from the model to new experimental/density functional theory results

Predicted Hydrogen Diffusion in Pd Binary Alloys



- Expand the model to work with more complicated random alloys such as high-entropy alloys

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology &Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. _____

[1] H. Wu *et al.*, *Computational Materials Science. 134*, 5 160-5 (2017).
[2] Y. Zeng *et al. Computational Materials Science.* 144 232-47 (2016)
[3] L. Ward *et al*, *npj Comput Mater.* 16028 (2016).
[4] D.S. Shupe and R.E. Stickney, *The Journal of Chemical Physics. 51,* 4 1620-5 (1969)

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.