SAND2021-14831C

Sandia National Laboratories

A data-driven roadmap towards Pareto optimal hydrogen storage alloys

PRESENTED BY

Matthew Witman

Staff Scientist, Energy Nanomaterials Department
Sandia National Laboratories, Livermore, CA USA

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.

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.

Collaborators & Acknowledgements

Experiments

Mark Allendorf
Vitalie Stavila
Jeffery Chames
Robert Horton
David Grant (Nottingham)
Gavin Walker (Nottingham)
Gustav Ek (Uppsala)
Martin Sahlberg (Uppsala)
Claudia Zlotea (Paris Est)
Anis Bouzidi (Paris Est)

Machine Learning

Sapan Agarwal
Justin Wong
Emily Allendorf

Computational Chemistry

Sanliang Ling (Nottingham University)

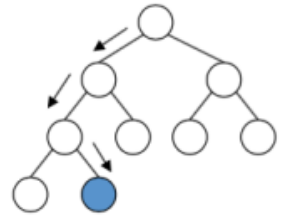
Data Infrastructure

Rachel Hurst (NREL)
Nick Wunder (NREL)
Max Gallant (NREL)

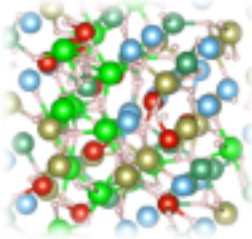
The authors gratefully acknowledge research support from the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Hydrogen and Fuel Cell Technologies Office through the Hydrogen Storage Materials Advanced Research Consortium (HyMARC). This work was supported by the Laboratory Directed Research and Development (LDRD) program at Sandia National Laboratories.

Seminar Organization

Approach: Using data-driven/machine learning techniques for materials discovery

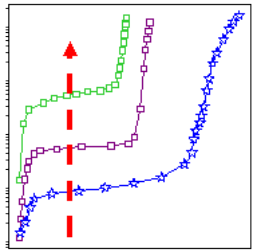


Application: Discovering more efficient materials-based hydrogen storage



Results: Success stories in data-driven discovery of hydrogen storage materials

- New technical capabilities/physical insights
- Towards Pareto optimal materials



Future Work: Overcoming current limitations in ML-driven discovery for hydrogen storage materials

$$v = \mathrm{avg}(v_1^{t0}, v_2^{t0})$$

$$v_1^{t1} = (v_1^{t0} \oplus v_2^{t0} \oplus \mathbf{e}_{12}) \cdot \mathbf{W} + \mathbf{b}$$

Approach:

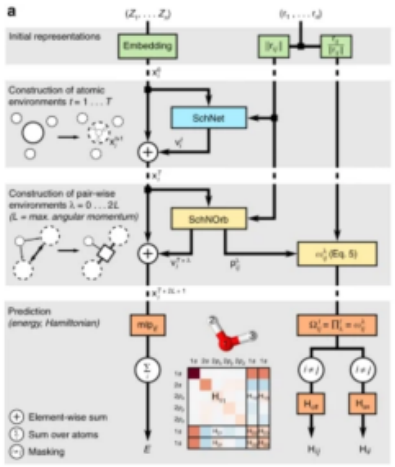
Why use data-driven/machine learning techniques for materials discovery?

Data science/ML models in materials/chemical sciences can be applied to problems of "all shapes and sizes"

Angstroms … to … **Acres**

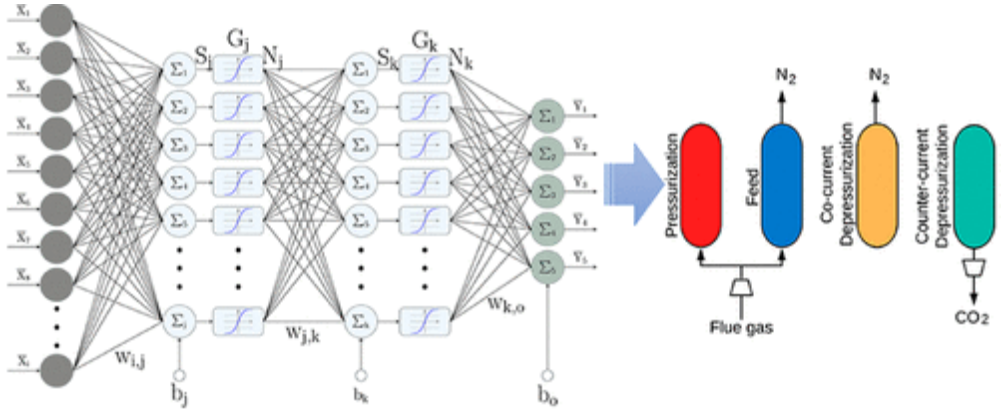
Small molecule property predictions[1]

Inputs
(***Z***, ***r***)



Outputs
(*E*)

Chemical plant performance prediction[2]



Inputs
(PSA unit parameters)

Outputs
(CO_2 purity of PSA units)

ML models **are not**: a magic box that solve all our problems, will always require "truth" data from experiments or first principles calculations (often difficult in materials science)

ML models **are**: *surrogates that execute **many orders of magnitude** faster than an experiment or first principles calculation to make task X tractable (high-throughput screening, optimization, active search, etc.)*

[1] Schutt, et al. *Nat. Comm.*10, **2019**
[2] Leperi, et al. *I&EC*. 58 (39), **2019**

Three main approaches to materials discovery

Experiments

Chemical intuition

↓

Synthesize material

↓

Characterize material

↓

Test Performance

Physics-based simulations

First-principles theory

↓

Run *simulation* & predict performance

Machine learning (ML)

Material property/training data

↓

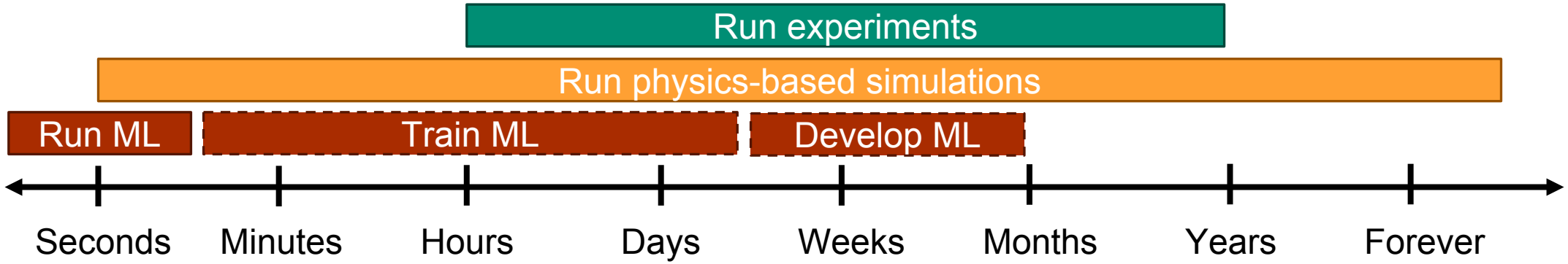
Convert materials to features

↓

Construct & train a model

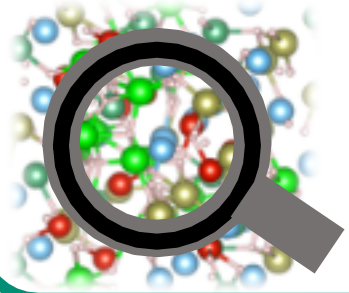
↓

Run *model* & predict performance



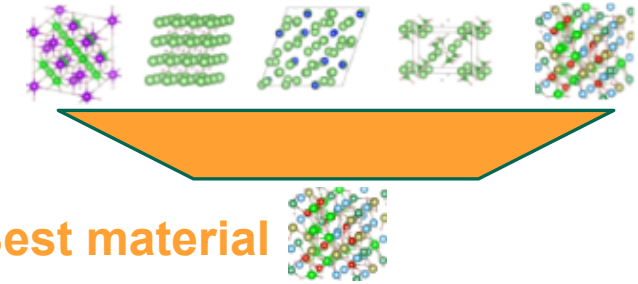
Generally speaking, data science and machine learning techniques can…

Predict properties and elucidate design rules for optimal materials

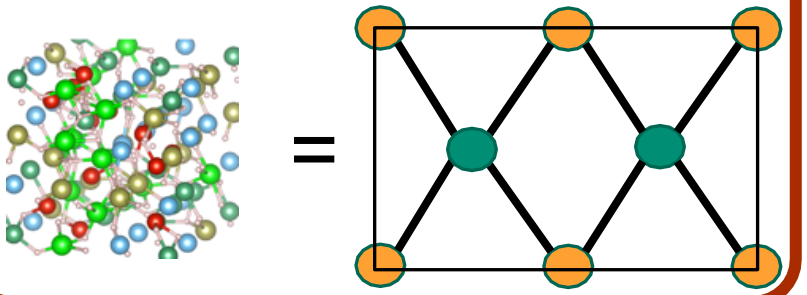


$$P_{eq} = m \quad + b$$

High-throughput screen materials orders of magnitude faster than experiments or simulations



Accelerate physics-based simulations when lacking experimental training data



Tools provide a roadmap for…

H_2 storage: Data-driven discovery of Pareto optimal hydrogen storage alloys

H_2 generation: Data-driven discovery of water-splitting materials (STCH, liquid metals, etc.)

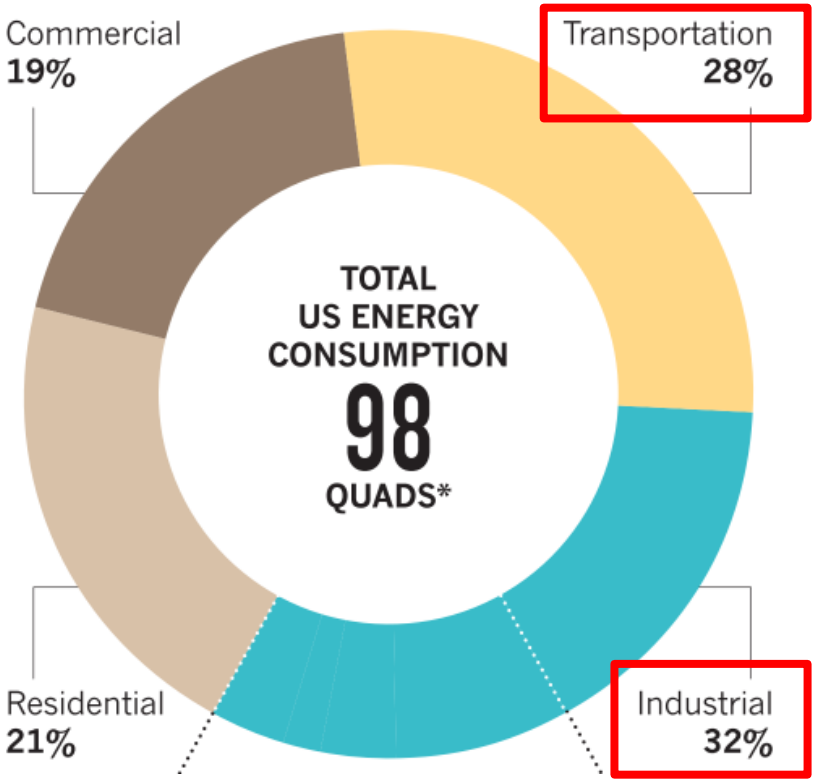
Application:

Discovering more efficient materials-based hydrogen storage

Discovering improved H_2 storage and generation techniques will facilitate progress towards a hydrogen economy

H_2 to play a critical role in…

Heavy duty (planes, trains, ships)



Sholl, Lively. Nature 532 (7600) **2016**.

High T applications

Personal transportation (Toyota/Hyundai)

Air Travel (Airbus)

Home energy storage (Lavo)

Utility-scale generation (Shell)

First Energy Earthshot Aims to Slash the Cost of Clean Hydrogen by 80% to $1 per Kilogram in One Decade

"Clean hydrogen is a game changer" -- Secretary Granholm

Are there undiscovered materials that could improve upon conventional H_2 storage and generation technologies?

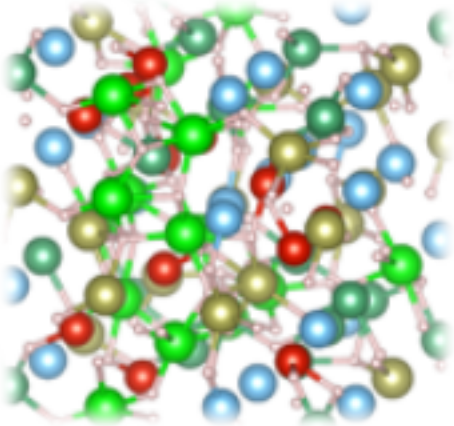
H_2 Storage objectives:

- High H_2 gravimetric and/or volumetric density
- Fast, reversible release near ambient T
- Practical/cost-effective
- Reduced infrastructure cost

700 bar

"Conventional" (Compressed gas)

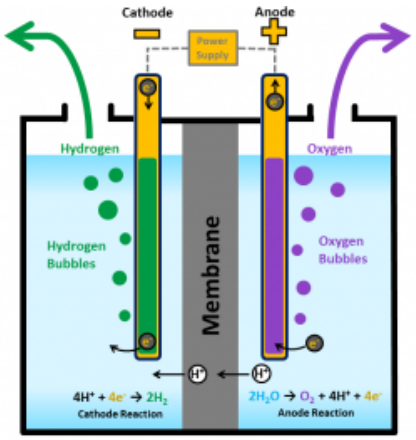
vs.



Material X ??

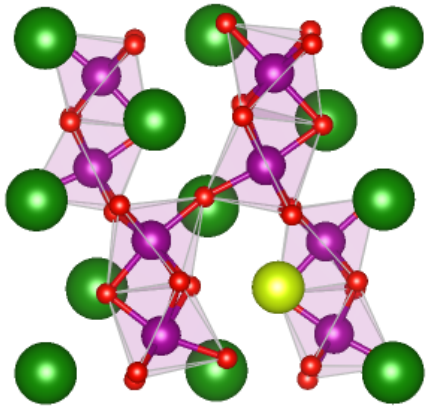
H_2 Generation objectives:

- Water-splitting using only renewable energy
- Practical/cost-effective



"Conventional"
1.2 V in theory
1.8 V in practice

vs.

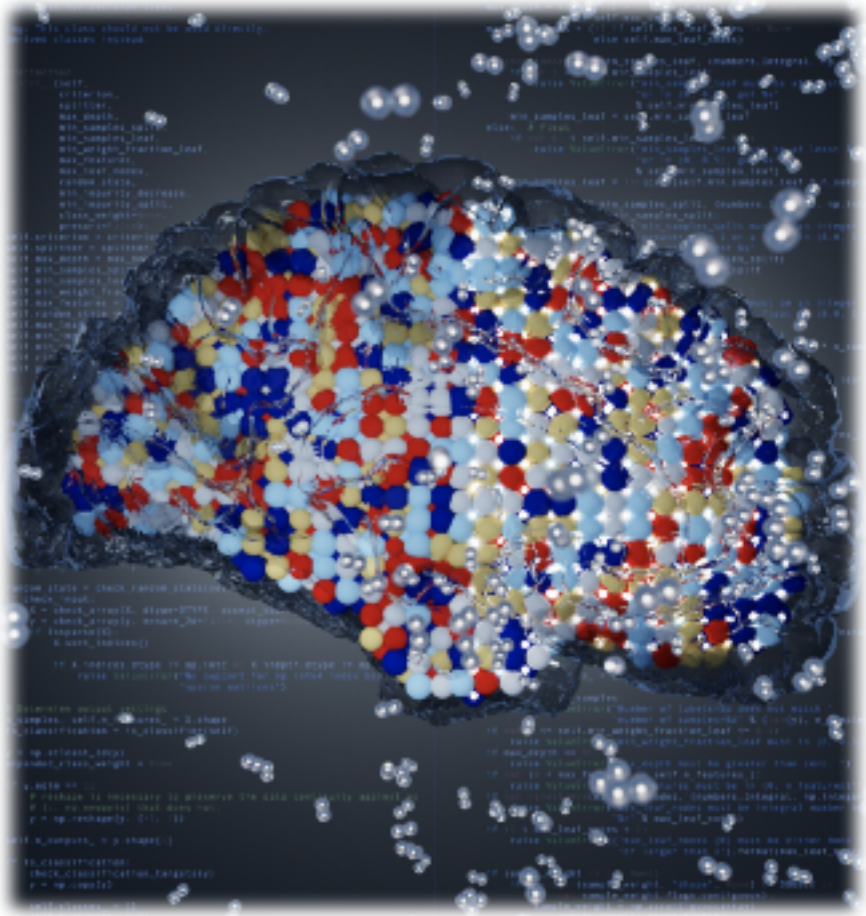


Material Y ??

Results: A data-driven roadmap towards Pareto optimal hydrogen storage alloys

- **Milestone #1** : Explainable ML models predict metal hydride thermodynamics
- **Milestone #2** : ML enables discovery of destabilized high entropy alloy (HEA) hydrides
- **Milestone #3** : ML screening & identification of *Pareto optimal* HEA hydrides

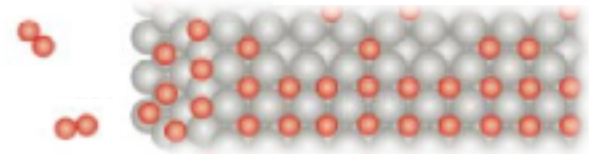
Milestone #1: Explainable ML models predict metal hydride thermodynamics



Milestone #1: Explainable ML models predict metal hydride thermodynamics

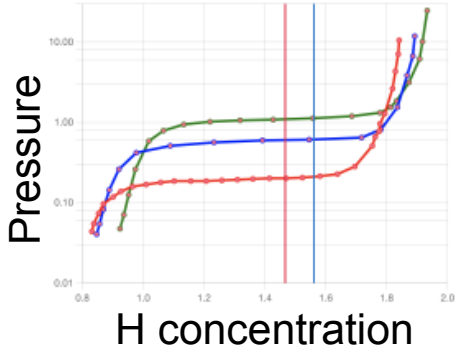
(1) $\ln(P^o_{eq}/P_o)$ target property

H_2 Metal Hydride



$$\ln(P^o_{eq}/P_o) = -\frac{\Delta H}{R(25^oC)} + \frac{\Delta S}{R}$$

Extracted from isotherms & van't Hoff



$\Delta H, \Delta S$ are tabulated in HydPARK.csv
- Missing data, errors, etc.
- ***Only*** 15% of 2500 entries usable

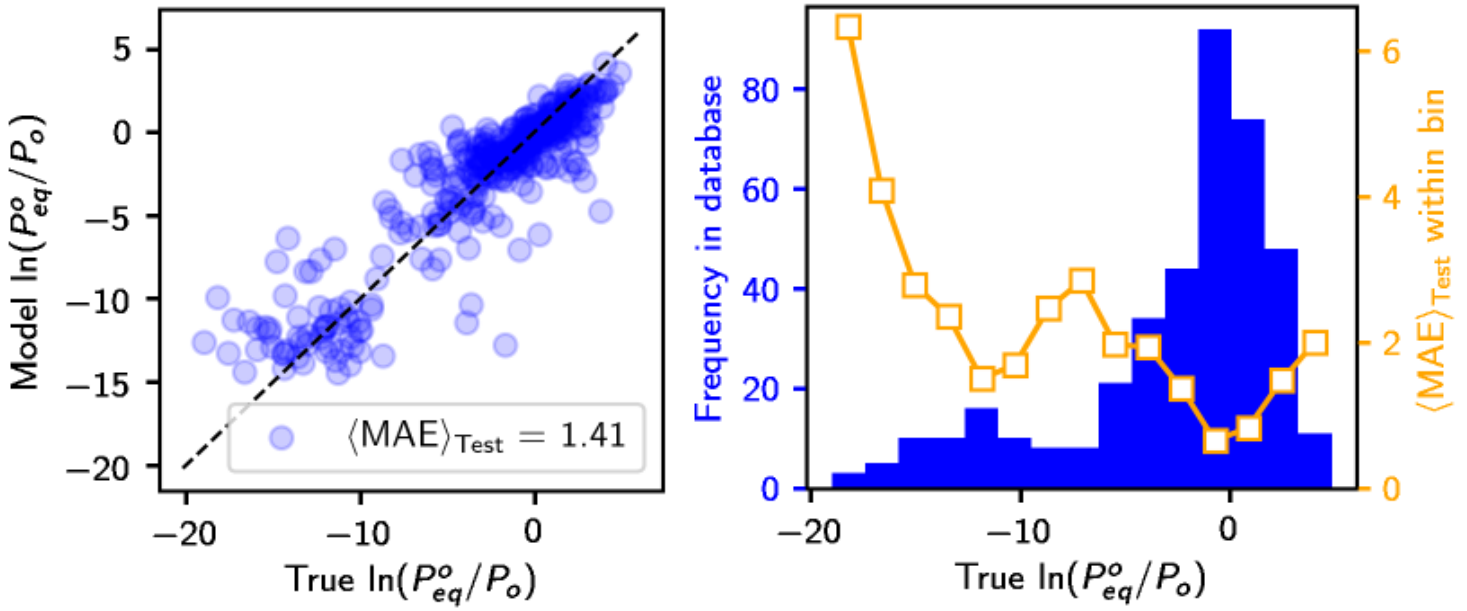
(2) Constructing a compositional ML model[1]

- Features derived from substituent elemental properties:
  $f_i \equiv$ fraction of element i, $p_i \equiv$ elemental property of i
- Simple operations map composition to scalar feature:
  $\bar{p} = \sum_i f_i p_i$, $\hat{p} = \sum_i f_i |p_i - \bar{p}|$, $\ddot{p} = \max(p_i) - \min(p_i)$, etc…
- Any composition mapped to same dimensional feature vector:
  $$\boldsymbol{x}_{\text{Magpie}} = \{\bar{\nu}_{pa}, \bar{r}_{cov}, \bar{\chi}, \ldots\} \in \mathbb{R}^{145}$$
- **Necessary if exact structure unknown** (e.g. $TiFe_{0.92}Nb_{0.08}$)
- Gradient boosting trees (ensemble of *m* trees):
  $y = F_m(\boldsymbol{x}) + h_m(\boldsymbol{x})$
  $h_m(\boldsymbol{x})$, a new estimator, corrects the previous model, $F_m(\boldsymbol{x})$
  i.e. is fit to minimize residual error, $y - F_m(\boldsymbol{x})$
- **Expressive, interpretable model if sufficient data**

[1] Ward, et al. *npj Comp. Mater.* 2016
Witman, Ling, Grant, Walker, Agarwal, Stavila, Allendorf. *J. Phys. Chem. Lett.*, 11 (1), **2020**

Milestone #1: Explainable ML models predict metal hydride thermodynamics

Thermodynamics (equilibrium plateau pressure) model validation

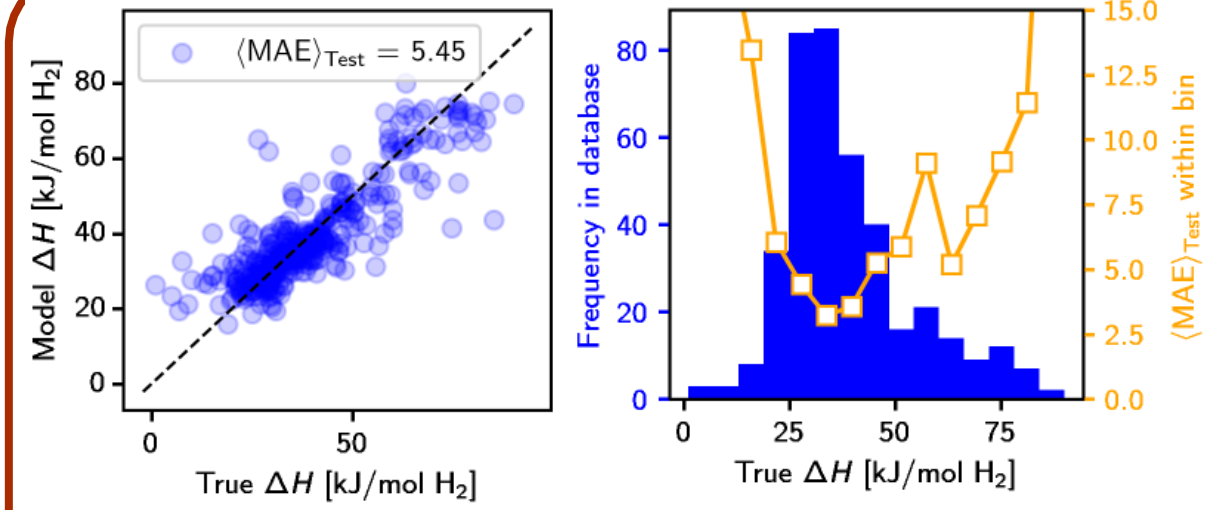


*ML model **can predict** $\hat{y} \equiv \ln(P^o_{eq}/P_o)$ with cross-validation test <MAE> = 1.4

Witman, Ling, Grant, Walker, Agarwal, Stavila, Allendorf. *J. Phys. Chem. Lett.*, 11 (1), **2020**

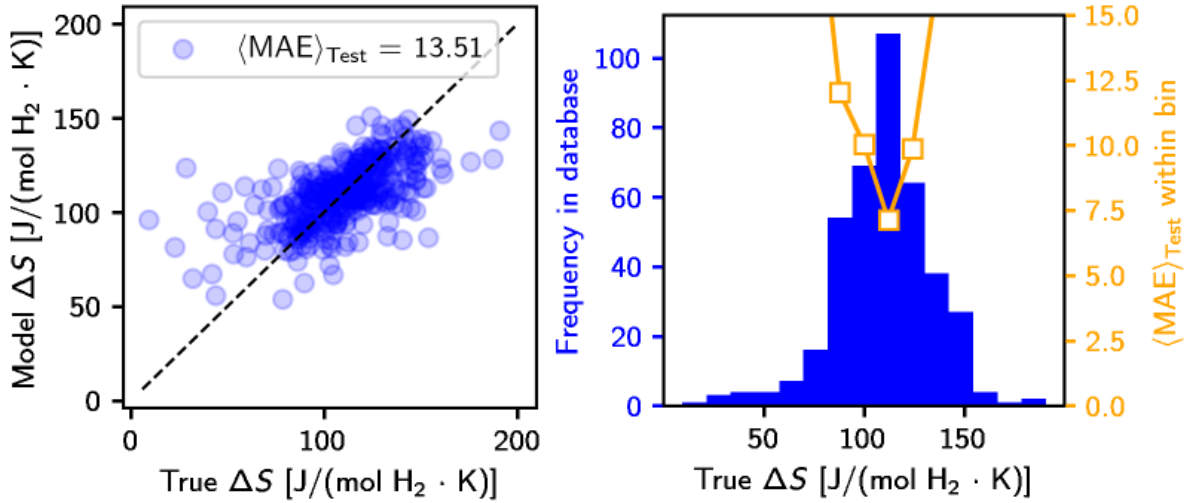
Milestone #1: Explainable ML models predict metal hydride thermodynamics

(3) Thermodynamics (individual ΔH and ΔS) model validation



Can predict $\hat{y} \equiv \Delta H$ with cross-validation test
<MAE> = 5.5 kJ/molH2

Can predict $\hat{y} \equiv \Delta S$ with cross-validation test
<MAE> = 14 J/(mol K)



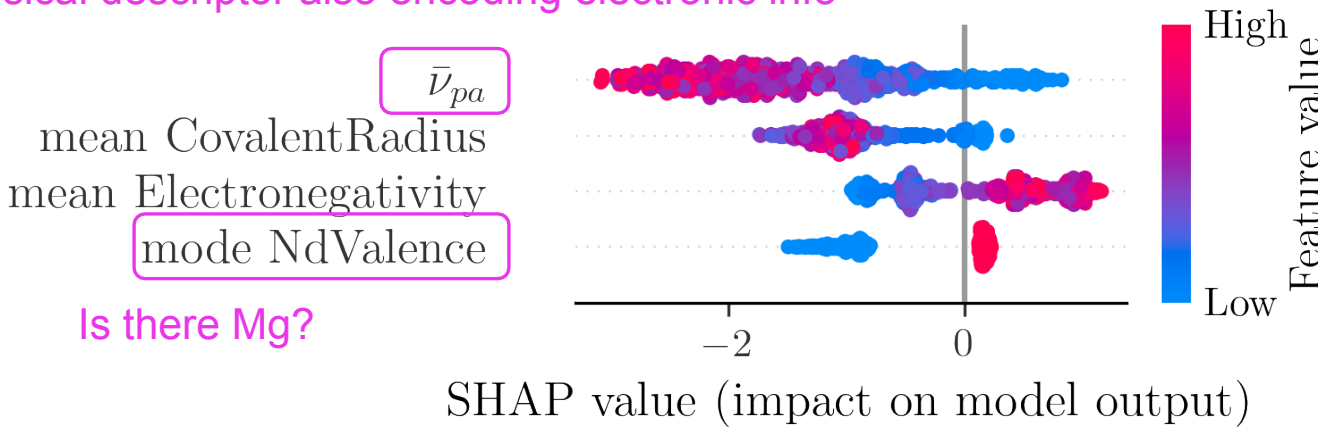
Witman, Ling, Grant, Walker, Agarwal, Stavila, Allendorf. *J. Phys. Chem. Lett.*, 11 (1), **2020**

Milestone #1: Explainable ML models predict metal hydride thermodynamics

Model interpretability with SHapely Additive Predictions (SHAP)[1]

How does a model's output depend on the value of a given feature?

"Elemental solid ground state volume/atom":
A physical descriptor also encoding electronic info



Is there Mg?

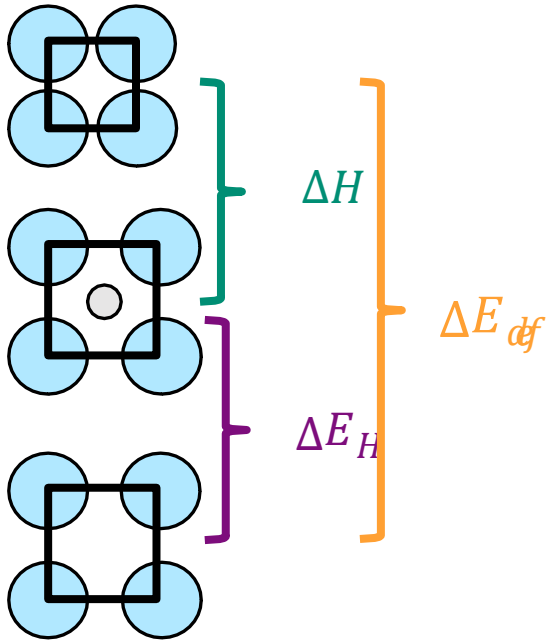
***Linear correlation** with $\bar{\nu}_{pa}$: $\ln\left(\frac{P^o_{eq}}{P_o}\right) \approx -m\,\bar{\nu}_{pa} + b$

[1] Lundberg, et al. *NIPS*, **2017**.
Witman, Ling, Grant, Walker, Agarwal, Stavila, Allendorf. *J. Phys. Chem. Lett.*, 11 (1), **2020**

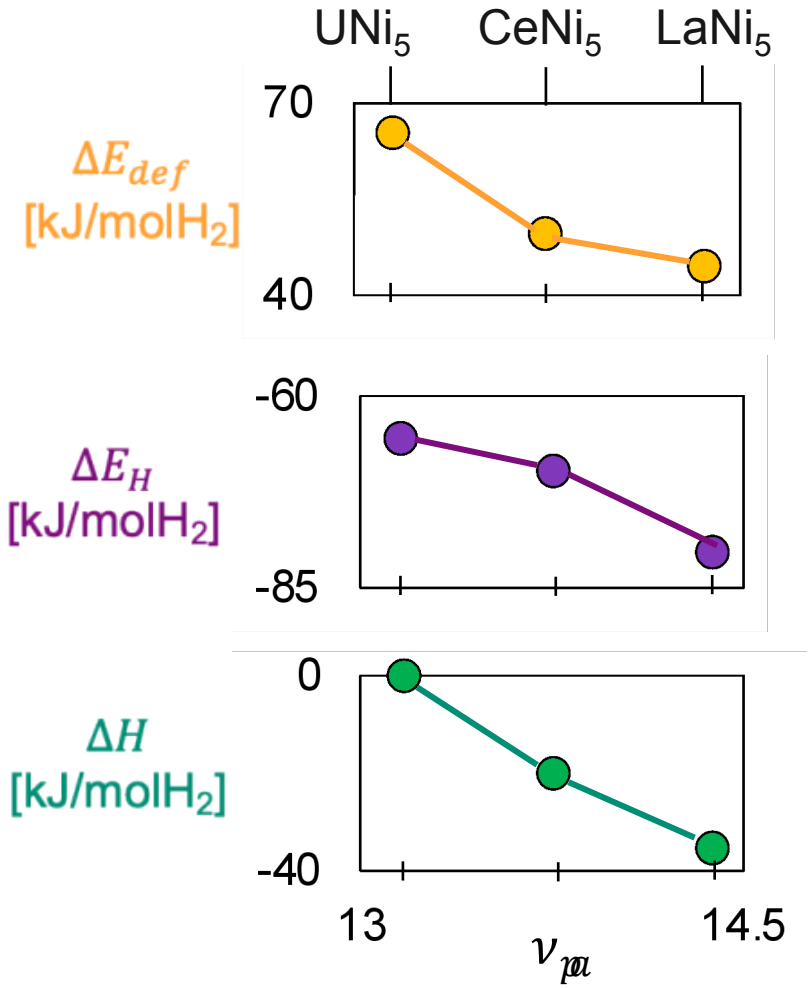
Milestone #1: Explainable ML models predict metal hydride thermodynamics

A DFT case study on correlation of hydride thermodynamics with $\bar{\nu}_{pa}$

Probe the reaction:
$AB_5 + 3.5H_2 \rightarrow AB_5H_7$



Lattice deformation penalty
H stabilization energy
Hydride enthalpy

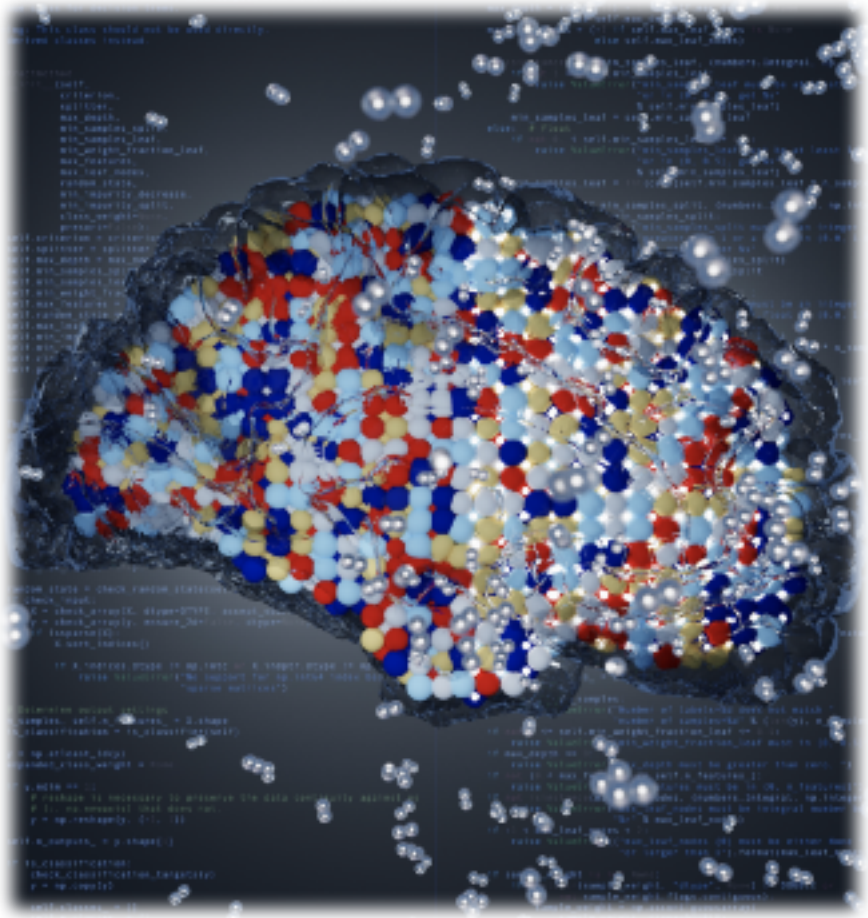


Decreases with $\bar{\nu}_{pa}$

Decreases with $\bar{\nu}_{pa}$

Net decreases with $\bar{\nu}_{pa}$

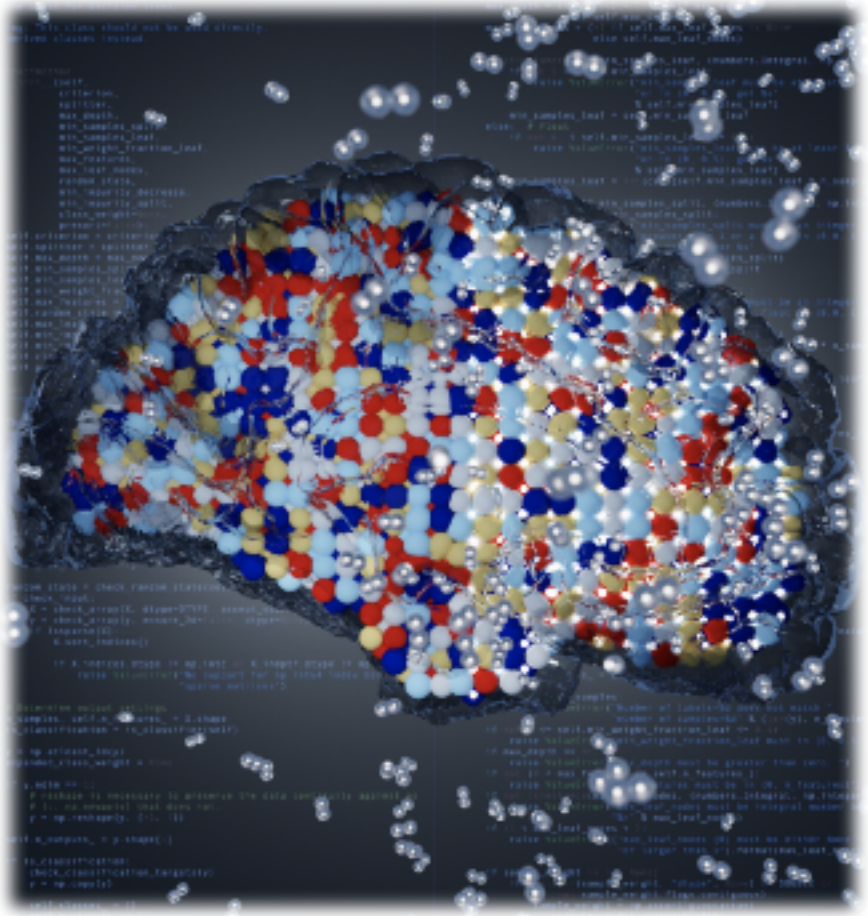
Witman, Ling, Grant, Walker, Agarwal, Stavila, Allendorf. *J. Phys. Chem. Lett.*, 11 (1), **2020**



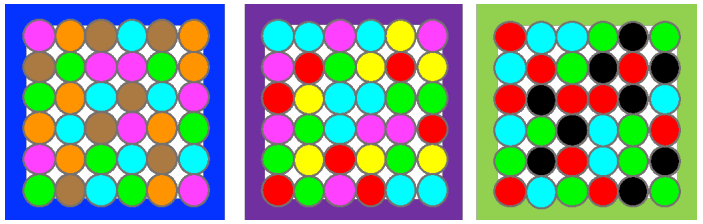
Conclusions:

- Compositional ML models can predict metal hydride thermodynamics
- Interpretability of ML models reveals design rules of increasing complexity
- DFT can validate ML-established design rules for specific hydride classes with limited throughput

Milestone #2: ML enables discovery of destabilized high entropy alloy (HEA) hydrides



Milestone #2: ML enables discovery of destabilized high entropy alloy (HEA) hydrides



1. Define chemical/materials exploration space

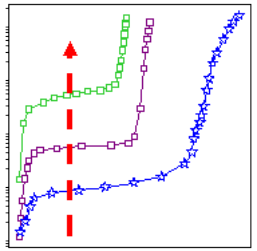
$\ln\left(\frac{P_{eq}}{P_o}\right)$ Model

2. High-throughput screening/property prediction

Phase Model

Manual Select

3. Apply necessary additional down-selection criteria

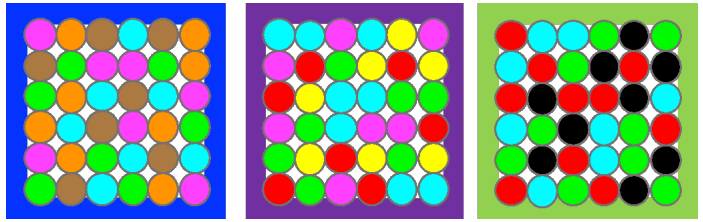


4. Experimental validation

Witman, Ek, Ling, Chames, Agarwal, Wong, Allendorf, Sahlberg, Stavila. *Chem. Mater.* 30 (11), **2021**

Milestone #2: ML-based discovery of destabilized high entropy alloy (HEA) hydrides

(1) HEA overview:



- > 4 elements, ~ equimolar
- Defined lattice type
- Solid solution character necessitates a compositional ML model

(2) Enumerating refractory HEA space

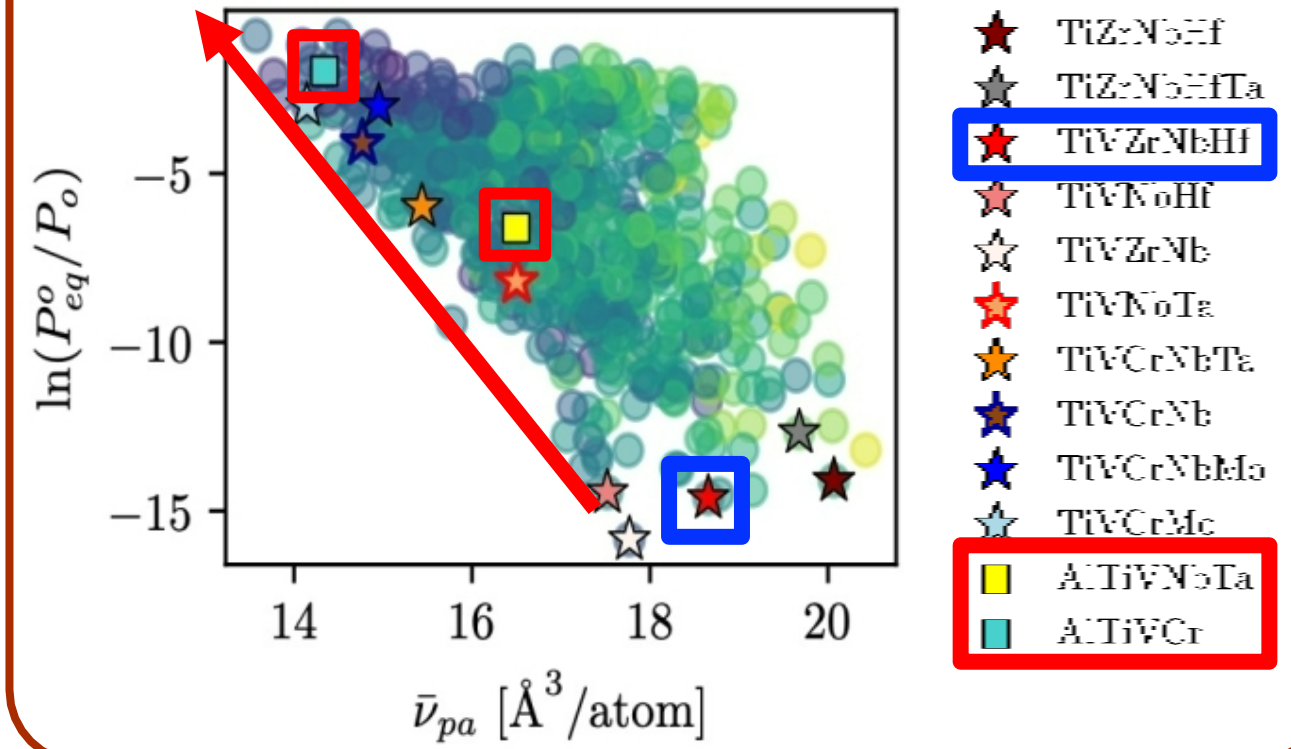
E = {Al, Ti, V, Cr, Zr, Nb, Mo, Pd, Hf, Ta}

$$\binom{E}{4} + \binom{E}{5} + \binom{E}{6} \rightarrow 672 \text{ compositions}$$

Far too many for experiments…

(3) Screening refractory HEA space

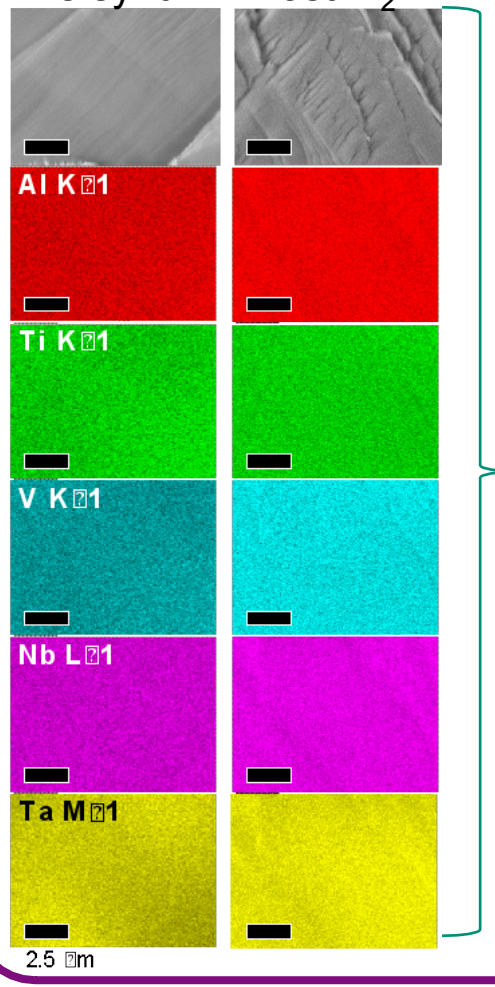
Destabilized hydrides experimentally confirmed!



Witman, Ek, Ling, Chames, Agarwal, Wong, Allendorf, Sahlberg, Stavila. *Chem. Mater.* 30 (11), **2021**

Milestone #2: ML enables discovery of destabilized high entropy alloy (HEA) hydrides

(1) AlTiVNbTa & AlTiVCr synthesis



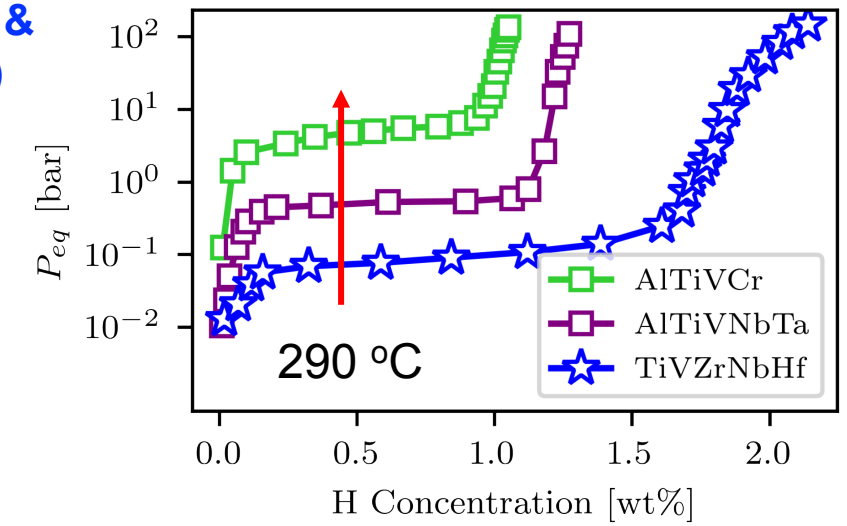
(Gustav, Martin, & Jeffrey)

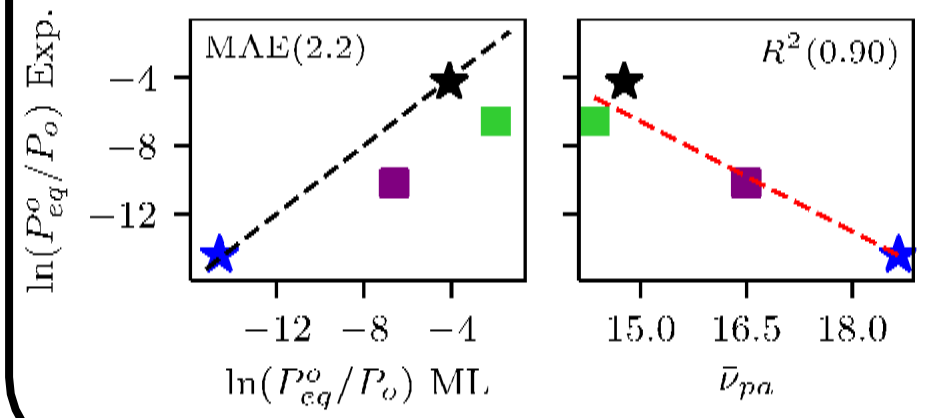
No elemental segregation

(2) ML model & design rule confirmed by PCT experiments

(Vitalie & Robert)

Successfully targeted destabilized hydrides (increase in P_{eq})





Validated ML model & design rule

Witman, Ek, Ling, Chames, Agarwal, Wong, Allendorf, Sahlberg, Stavila. *Chem. Mater.* 30 (11), **2021**

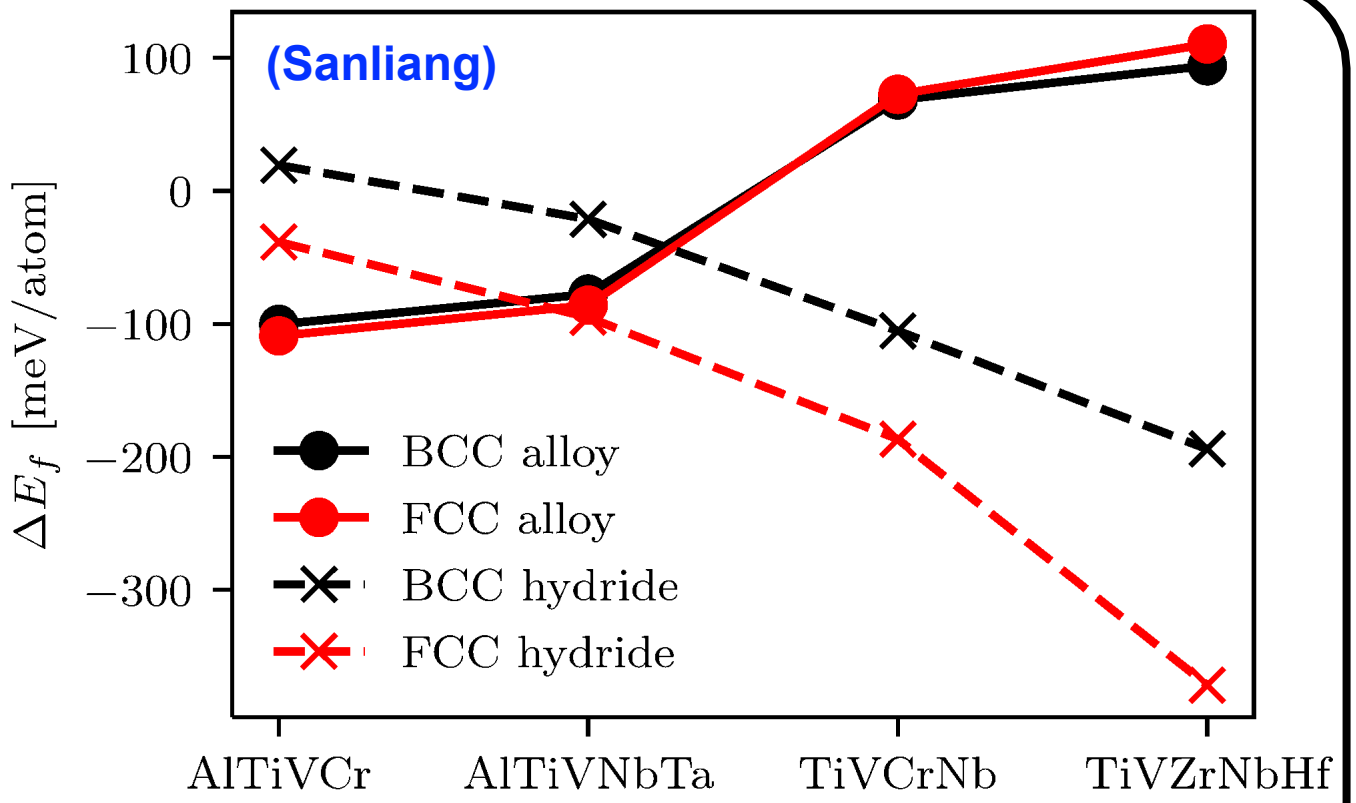
Milestone #2: ML enables discovery of destabilized high entropy alloy (HEA) hydrides

Investigating thermodynamic trends with Density Functional Theory

Compute 0K formation enthalpy of alloy and hydride for single HEA config.:

$$\Delta E_{f,\text{alloy}} = \frac{1}{N}\left(E_{\text{alloy}} - \sum_{i\in\{\text{alloy}\}} E_i\right)$$

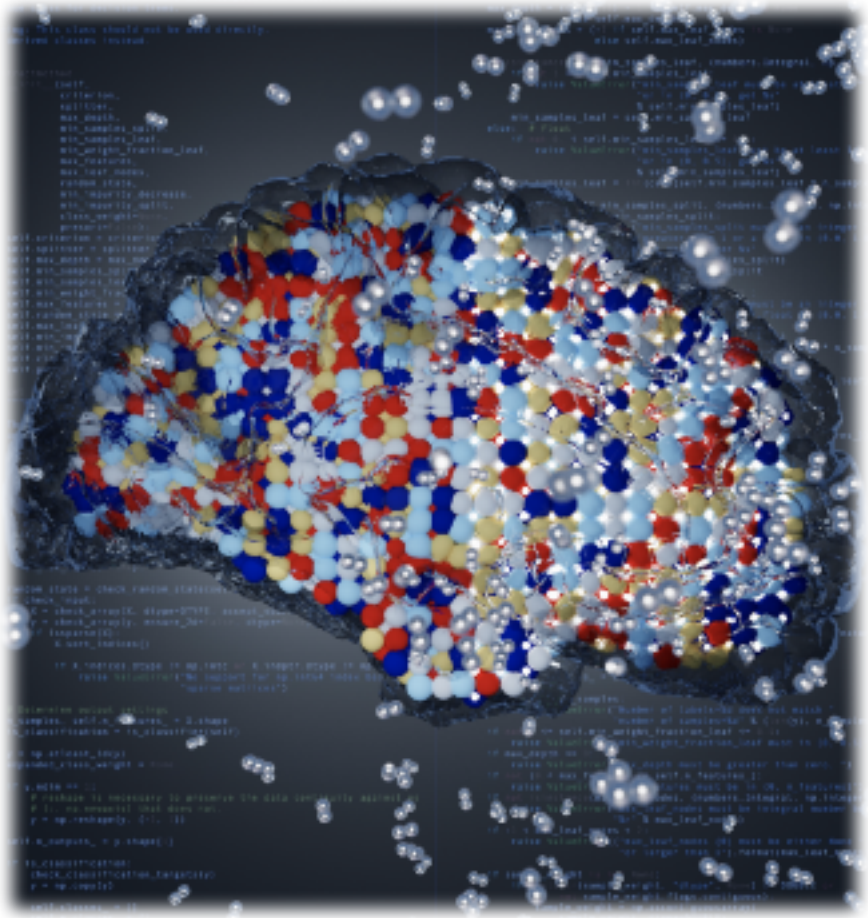
$$\Delta E_{f,\text{hydride}} = \frac{1}{N}\left(E_{\text{hydride}} - \left(\frac{N}{2}\right)E_{H_2} - \sum_{i\in\{\text{alloy}\}} E_i\right)$$



Analysis Reveals:

- Enthalpic driving force for BCC alloy -> FCC transition @ high H/M ratios
- The general correlation of less stable lattice = more stable hydride (similar to $\bar{\nu}_{pa}$ design rule)

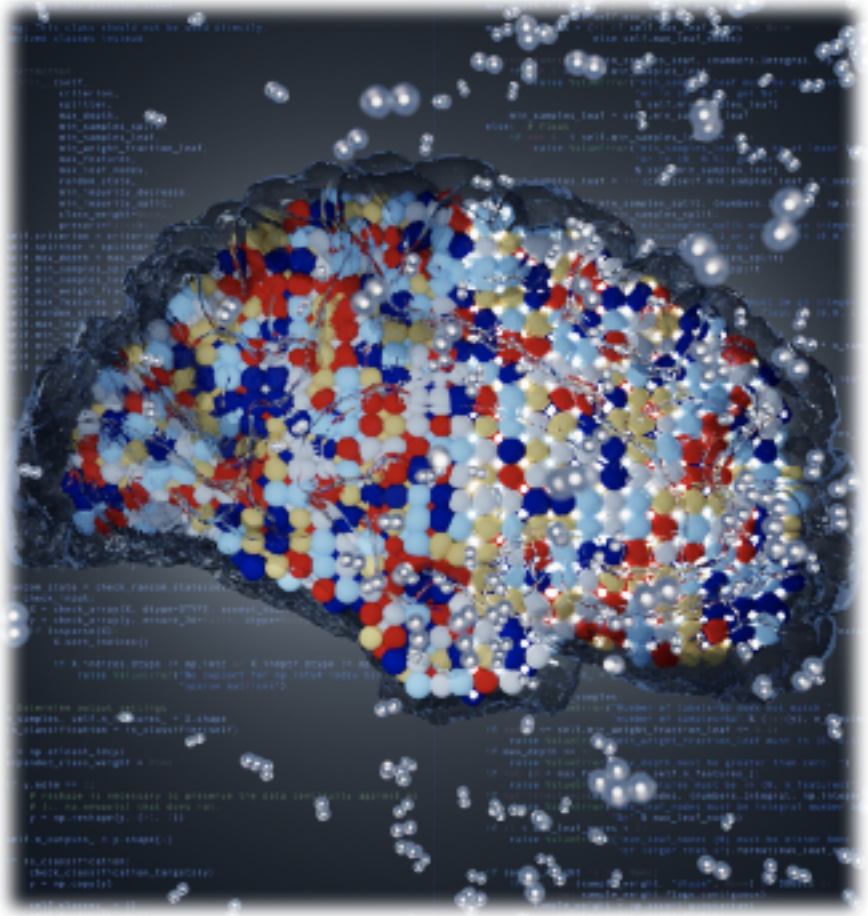
Witman, Ek, Ling, Chames, Agarwal, Wong, Allendorf, Sahlberg, Stavila. *Chem. Mater.* 30 (11), **2021**



Conclusions:

- Introduced a powerful ML capability for high-throughput screening
- ML-directed and experimentally validated synthesis of HEA hydrides
- DFT-enabled insights into thermodynamic trends

Milestone #3: ML screening & identification of *Pareto* optimal HEA hydrides



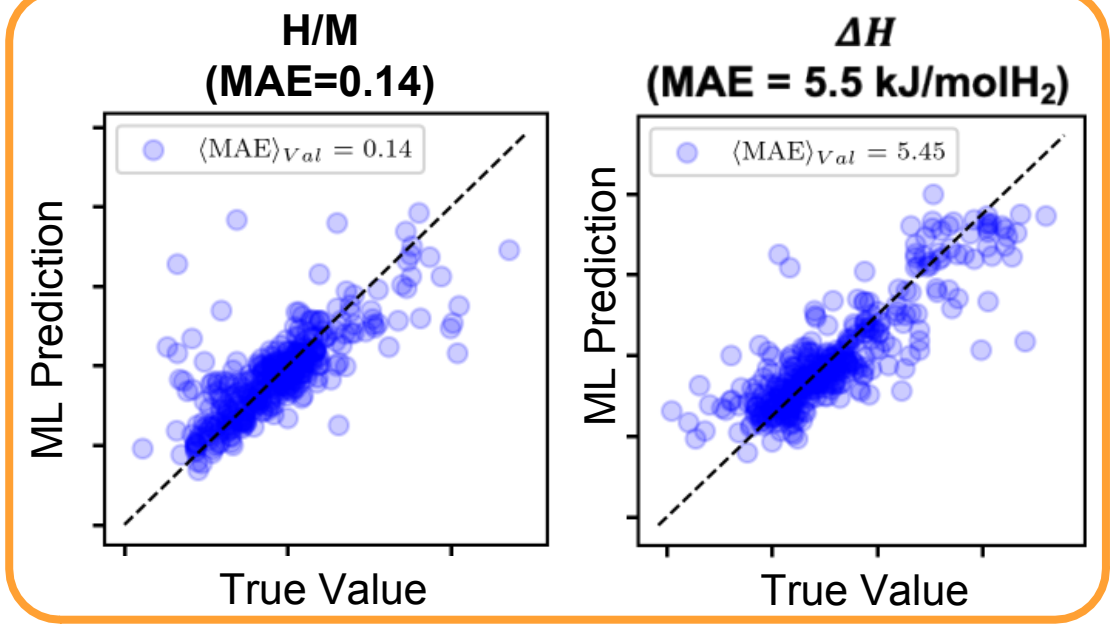
Milestone #3: ML screening & identification of Pareto optimal HEA hydrides

(1) Screening an expansive HEA space

E = {Mg, Al, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Zr, Nb, Mo, Pd, Hf, Ta}

$$\binom{E}{4} + \binom{E}{5} + \binom{E}{6} \rightarrow 20{,}944 \text{ compositions}$$

(2) Multiple ML property predictions

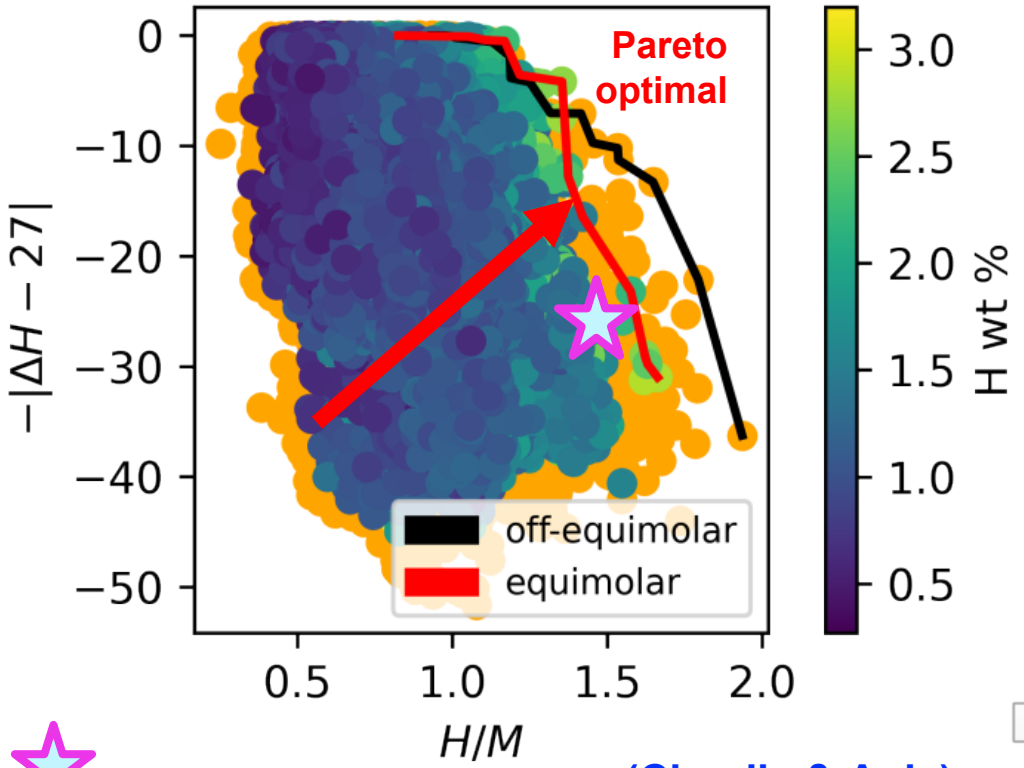


In preparation

(3) Identification of ~100 Pareto optimal materials for stationary storage

Objectives / Quantity to maximize:

- Optimal thermodynamics -> $-|\Delta H - 27|$
- High volumetric capacity -> H/M
- High gravimetric capacity -> Hwt%
- Raw material cost -> -cost



= Experimental validation (Claudia & Anis)

Milestone #3: ML screening & identification of Pareto optimal HEA hydrides

(1) Screening an expansive HEA space

E = {Mg, Al, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Zr, Nb, Mo, Pd, Hf, Ta}

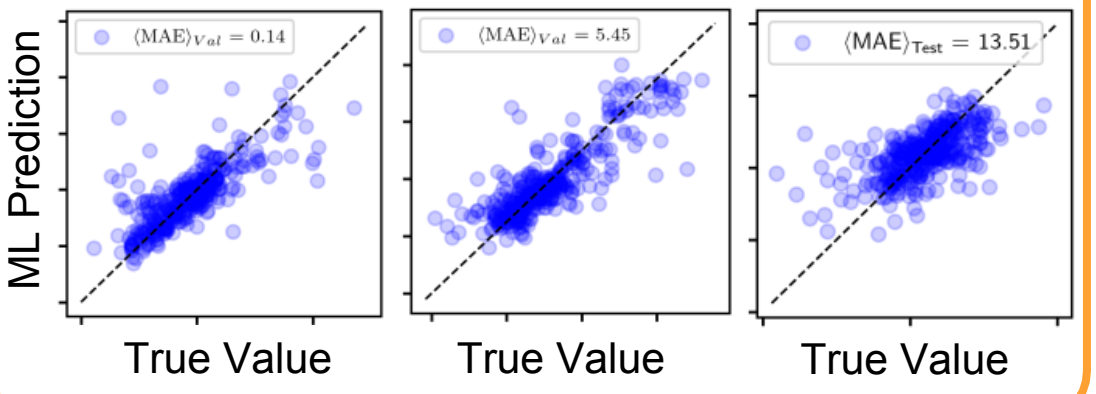
$$\binom{E}{4} + \binom{E}{5} + \binom{E}{6} \rightarrow 20{,}944 \text{ compositions}$$

(2) Multiple ML property predictions

H/M (MAE = 0.14)

ΔH (MAE = 5.5 kJ/molH$_2$)

ΔS (MAE = 14 J/molK)

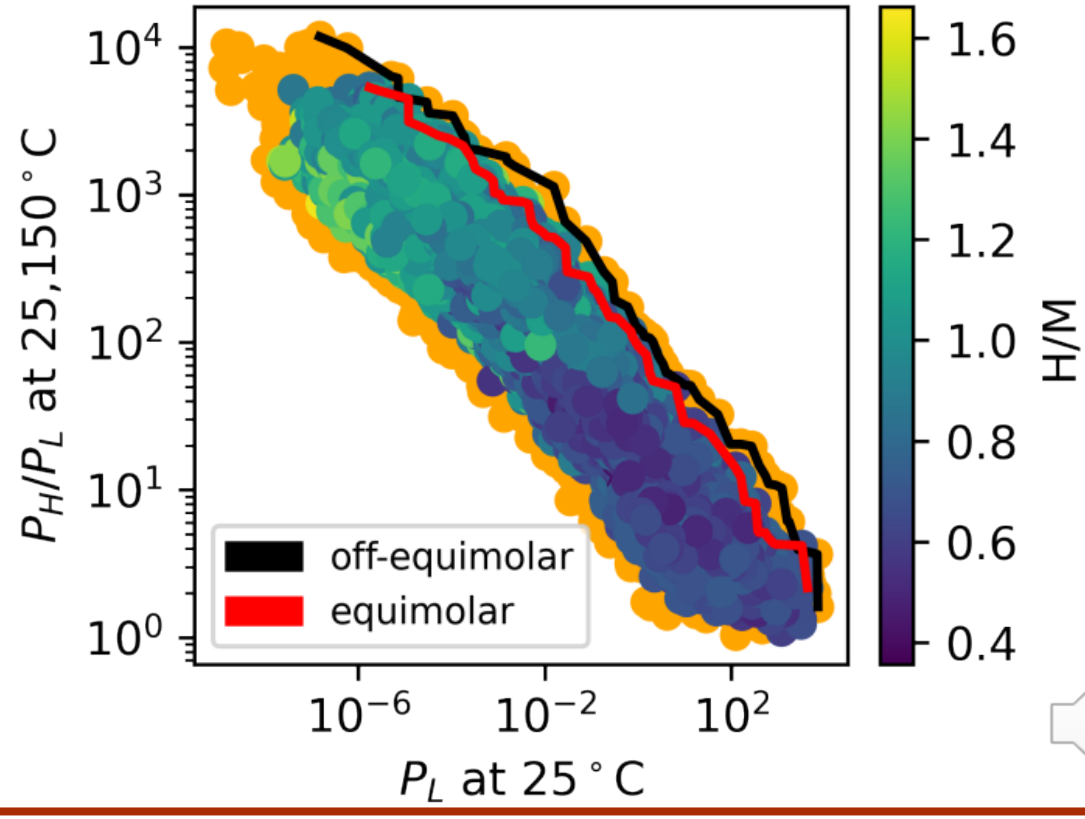


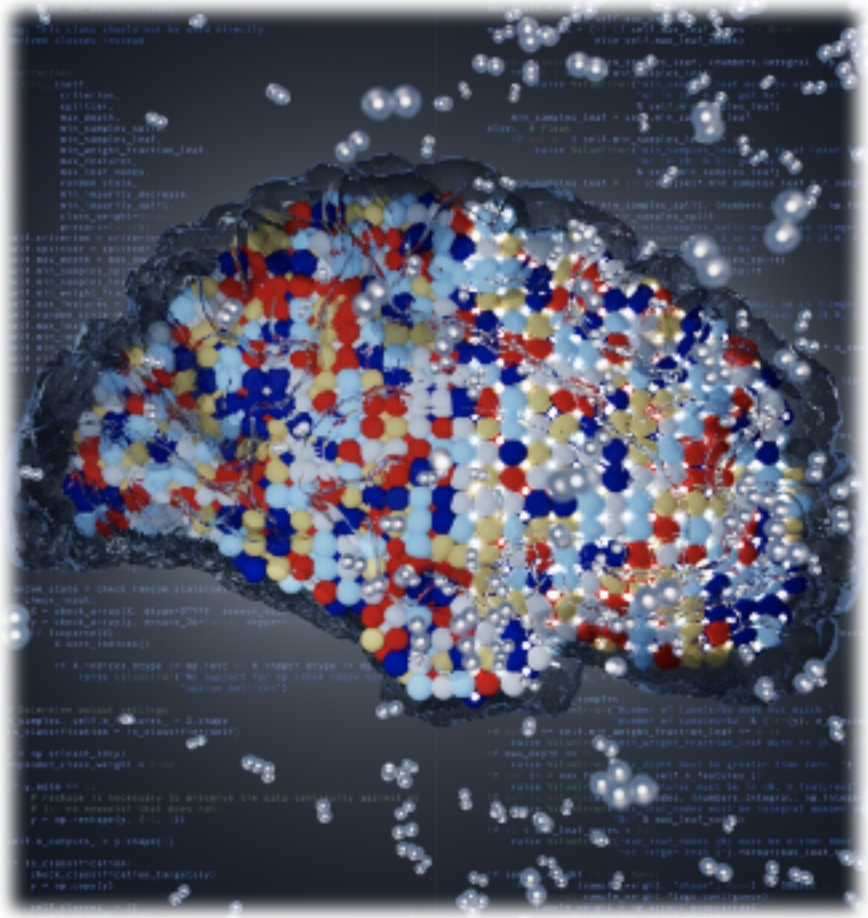
In preparation

(3) Identification of ~100 Pareto optimal materials for hydrogen compressors

Objectives / Quantity to maximize:

- P_{eq} @ T_L= 25 C -> P_L
- Compression ratio @ T_H= 150 C -> P_H/P_L
- High volumetric capacity -> H/M
- Raw material cost -> -cost





Conclusions:

- Given a set of ML models for various hydride properties, Pareto frontiers can be determined
- Random off-equimolar perturbations substantially advance the Pareto frontier, indicating potential for future improvement
- We have highlighted (with experimental validation pending) Pareto optimal compressor and stationary storage materials

Future Work:

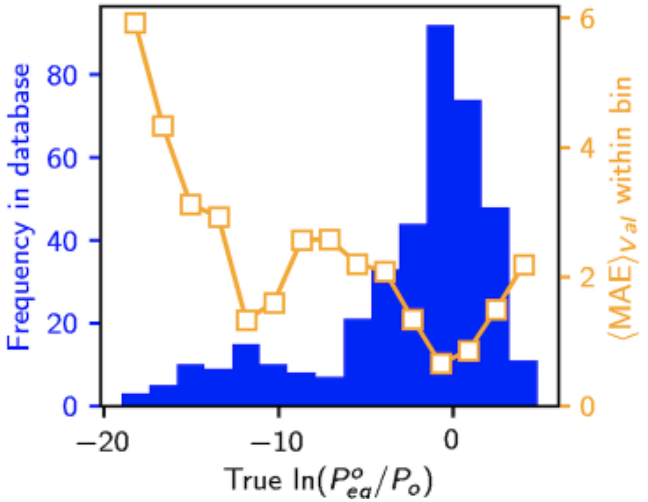
Overcoming barriers to ML-driven hydride discovery
(e.g. how to deal with limited data)

More/Better Experimental Data: Promoting standardization, reproducibility, and community-based "high-throughput" data acquisition

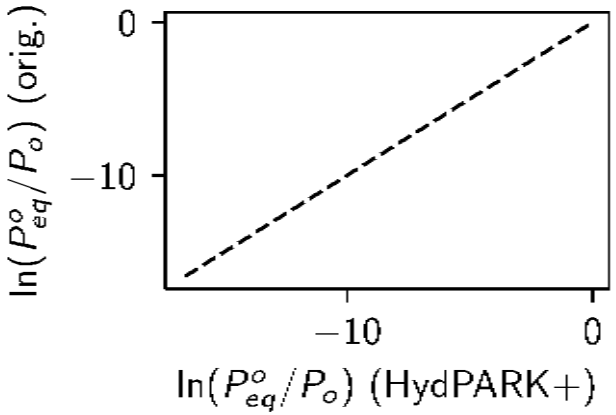
Challenge: Accuracy and accessibility of HydPARK data

Lack of updatability: data stored in "offline" CSV file
Lack of completeness: **15%** of HydPARK entries are usable
Lack of consistency: different values for identical materials

Very non-uniform distribution of target



Addition of new data systematically shifts predictions



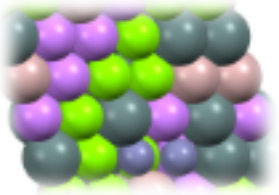
Solution: NREL DataHub app and backend for standardized PCT upload and storage

Much faster data acquisition than can be achieved by one group alone

More/Better Computational Data: If experimental measurements and physics-based simulation are too expensive, accelerate the latter with ML

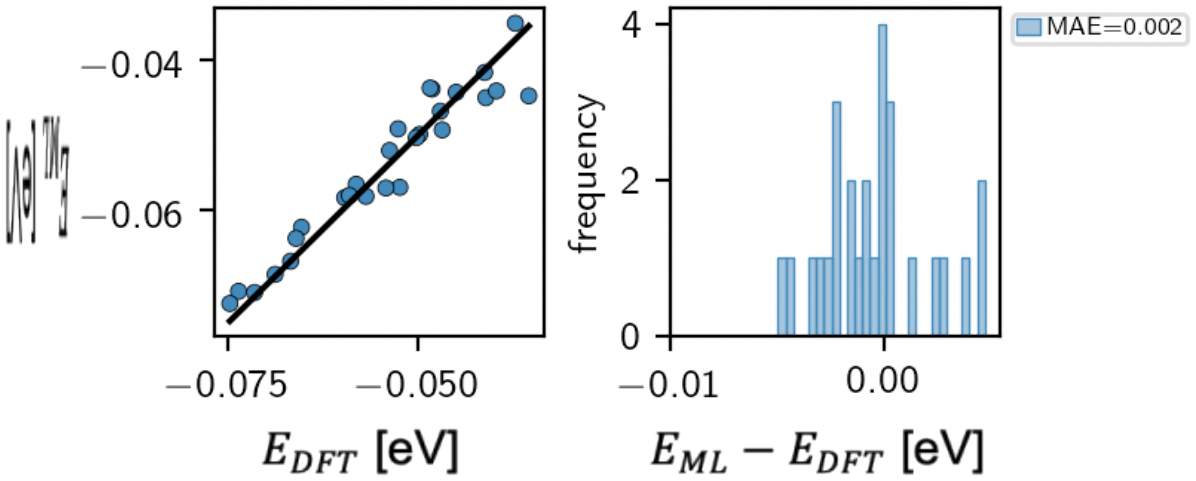
Challenge: Need an accurate, many-element potential energy surface (PES) for simulations (MD or MC)

PES choices: Force Field , ML PES, DFT



Not necessarily transferable/ generalizable

Too expensive for extensive sampling in complex systems



Solution: Direct atomistic simulation & screening of metal hydride thermodynamics?

Thank you for your attention.

Questions?

Contact: mwitman@sandia.gov