Extracting an Empirical Intermetallic Hydride Design Principle from Limited Data via Interpretable Machine Learning

Matthew Witman¹, Sanliang Ling², David M. Grant², Gavin S. Walker², Sapan Agarwal¹, Vitalie Stavila¹, Mark Allendorf¹

¹ Sandia National Laboratories, Livermore, CA USA
² University of Nottingham, Nottingham, UK
Acknowledgements

Collaboration Group:
Sandia – Sapan Agarwal, Vitalie Stavila, and Mark Allendorf
Nottingham – Sanliang Ling, David Grant, and Gavin Walker

Funding:
The authors gratefully acknowledge research support from the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, 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.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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-NA-0003525.
Economical storage of hydrogen is critical for enabling a variety of zero emission technologies. E.g., the Toyota Mirai is a commercially available fuel cell vehicle whose hydrogen is compressed and stored onboard. A material that meets all DOE technical targets for onboard hydrogen storage could send the technology mainstream:

https://www.businessinsider.com/this-toyota-fuel-cell-car-can-power-your-house-2014-11

https://www.energy.gov/eere/fuelcells/hydrogen-storage

How is hydrogen stored?

| Physical-based | Material-based |
|----------------|----------------|
| Compressed Gas | Liquid H₂       |
| Cold/Cryo Compressed | Interstitial hydride |
| Liquid organic | Complex hydride |
| Adsorbent | Chemical hydrogen |

Ex. MOF-5 | Ex. BN-methyl cyclopentane | Ex. LaNi₅H₆ | Ex. NaAlH₄ | Ex. NH₃BH₃
(1) Thermodynamics of hydriding, (2) hydrogen capacity, and (3) absorption kinetics decide applicability of the hydride

Table 2 Intermetallic compounds and their hydrogen-storage properties

| Type       | Metal | Hydride  | Structure | mass% | $p_{eq}$ | $T$     |
|------------|-------|----------|-----------|-------|----------|---------|
| Elemental  | Pd    | PdH$_{0.6}$ | Fm3m     | 0.56  | 0.020 bar, 298 K |
| AB$_5$     | LaNi$_5$ | LaNi$_5$H$_6$ | P6/mmm  | 1.37  | 2 bar, 298 K       |
| AB$_2$    | ZrV$_2$ | ZrV$_2$H$_{6.5}$ | Fd3m   | 3.01  | 10$^{-8}$ bar, 323 K |
| AB       | FeTi  | FeTiH$_2$ | Pm3m     | 1.89  | 5 bar, 303 K       |
| A$_2$B    | Mg$_2$Ni | Mg$_2$NiH$_4$ | P6222  | 3.59  | 1 bar, 555 K       |
| Body-centred cubic | TiV$_2$ | TiV$_2$H$_4$ | b.c.c.  | 2.6   | 10 bar, 313 K       |

Near atmospheric equilibrium pressure of H$_2$ occurs at room temperature
Near atmospheric equilibrium pressure of H$_2$ occurs at > 280 C

Schlapbach, I and Züttel, A. Hydrogen-storage materials for mobile applications. Nature, 2001, 414, 6861.
Research Question: Can machine learning (ML) yield physics-based insight to facilitate the design of novel metal hydrides exhibiting targeted thermodynamic properties?

1. Train an ML model to predict the equilibrium plateau pressure, $P_{eq}$, of a metal hydride from only the alloy composition.

2. Utilize the ML model's interpretability to understand the underlying structure-property relationships from which $P_{eq}$ can be predicted.

3. Use these structure-property relationships to a priori identify known intermetallic compositions whose hydrides have not been reported and are predicted to exhibit a desired $P_{eq}$.
DOE's experimental HYDPARK database contains alloy compositions and their hydriding thermodynamics

**HYDPARK database**

- ~2500 compositions

**Data cleaning**

1. Compute \( \Delta S = R \ln P_{eq} + \frac{\Delta H}{T} \)
2. Compute \( \ln P_{eq}^o = -\frac{\Delta H}{R(25 \, ^\circ C)} + \frac{\Delta S}{R} \)

**ML ready database**

- ~400 compositions

- **Comp.** | \( \Delta H \) | \( P_{eq} \) | \( T \) | ...
  - LaNi\(_5\)
  - ...
  - Er\(_6\)Fe\(_{23}\)
  - ...

- **Comp.** | \( \Delta H \) | \( P_{eq} \) | \( T \) | ...
  - LaNi\(_5\)
  - ...
  - Er\(_6\)Fe\(_{23}\)

* Must remove incomplete and handle duplicate compositions
* Complex hydrides excluded b/c only ~10 entries are complete
Aside: Why train an ML model to predict $\ln P_{eq}^o$ and not $\Delta H$?

1. $\ln P_{eq}^o$ accounts for both the enthalpic and entropic contributions to the free energy of hydriding.
2. Indicates the practical applicability of a given hydride for a given application (vehicular storage, high-pressure storage, H$_2$ getters, etc.)
3. A clear enthalpy-entropy trade-off exists, i.e. it is a “stretch” to argue an optimal $\Delta H$ exists for targeting a given $\ln P_{eq}^o$. 

![Graph showing the relationship between $\Delta H$ and $\Delta S$.]
Structurally agnostic featurization is required and gradient boosting trees yield insights from these features

**Magpie:**
each of the \( n = 400 \) compositions is mapped to a 145 dimensional vector computed from elemental properties

\[
X \in \mathbb{R}^{n \times 145} \\
x_{\text{LaNi}_5} = \{v_{pa}^{\text{Magpie}}, \ldots, 145\}
\]

**Gradient Boosting Trees:**
train a model, \( F \), by minimizing the mean squared error of its \( \ln P_{eq} \) predictions

\[
\hat{y} = F(X) \in \mathbb{R}^{n \times 1}\]

\[
\text{loss} = \frac{1}{n} \sum_i (\hat{y}_i - y_i)^2
\]

**Feature importance:**
Several ways to calculate, e.g. average number of times a feature is used to split data across all trees

An example Magpie descriptor:

\[
v_{pa}^{\text{Magpie}} = \sum_i f_i v_i
\]

- \( f_i \equiv \) composition fraction of element \( i \)
- \( v_i \equiv \) ground state volume per atom of elemental solid \( i \)
ML model can predict $\ln P_{eq}^o$ with decent accuracy using input features derived only from the intermetallic composition.
ML model can predict $\ln P_{eq}^0$ with decent accuracy using input features derived only from the intermetallic composition.

$\Delta H$ model:

$\Delta S$ model:
The $\nu_{pa}: \ln P^0_{eq}$ structure-property relationship extends over a wide range of metal substitutions and intermetallic classes.

1. Compute the structurally specific volume per atom for ~70 available structures in the Materials Project (MP) via:

$$V_{cell} \equiv \text{Volume of the intermetallic lattice computed in MP}$$

$$\nu_{pa}^{MP} = \frac{V_{cell}}{n_{atoms}}$$

2. Investigate equilibrium pressure as a function of $\nu_{pa}^{MP}$ and $\nu_{pa}^{MP}$:

Cuevas et al. noted the dependence of $\ln P^0_{eq}$ on $V_{cell}$ in LaNi$_5$ substitutions.

Smith et al. noted the same trend for R$_x$Fe$_{23}$ [R=Ho,Er,Lu] substitutions.
On the importance of $\nu_{pa}$

$\nu_{pa}$ encodes information about other features:

An equally accurate ML model can still be trained after removing $\nu_{pa}$ from the feature list:
Novel hydride phase of a known intermetallic for high-pressure H₂ storage predicted based on νₚₐ (and validated with DFT)

DFT computed properties for AB₅ + 3.5H₂ → AB₅H₇:

1. ΔH [kJ/molH₂] ≡ hydriding enthalpy
2. ΔE_{def} [kJ/molH₂] ≡ energy penalty to deform lattice to accommodate H absorption
3. ΔEₖ [kJ/molH₂] ≡ binding energy of H

Δ ≡ forward hydriding reaction

|        | νₚₐ | ΔH | Eₚ | ΔE_{def} | ΔEₖ | V/V₀ |
|--------|-----|----|----|----------|------|------|
| UNi₅   | 13.17 | -0.60 | -285 | 65.2 | -65.8 | 1.278 |
| CeNi₅  | 13.76 | -20.5 | -353 | 49.3 | -69.8 | 1.266 |
| LaNi₅  | 14.38 | -36.1 | -224 | 44.3 | -80.5 | 1.256 |

*Several U containing compounds in HydPARK, but no UNi₅ (even though it exists in ICSD)
Key Takeaways

1. ML models with experimental data provide a powerful tool to explore phenomena too expensive to simulate directly with computational approaches.

2. Equilibrium pressure in intermetallic hydrides can be predicted just from the alloy composition, despite noisy/incomplete data.

3. Explainable insights from the ML model permit the rational design of novel materials with targeted thermodynamic properties.
Thank you for your attention.

Questions?
Backup: DFT computed properties

Notation: $E(X)_{Y}$ denotes the energy of chemical system X in the geometry of system Y

Enthalpy of hydriding:
$$\Delta H = \left[ E(AB_5H_7)_{AB_5H_7} - E(AB_5)_{AB_5} - 3.5 \times E(H_2)_{H_2} \right] / 3.5 \text{ [kJ/molH}_2\text{]}$$

Lattice deformation energy of hydriding:
$$\Delta E_{def} = \left[ E(AB_5)_{AB_5H_7} - E(AB_5)_{AB_5} \right] / 3.5$$

Binding energy of hydriding:
$$\Delta E_H = \left[ E(AB_5H_7)_{AB_5H_7} - E(AB_5)_{AB_5H_7} - 3.5 \times E(H_2)_{H_2} \right] / 3.5$$