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

Matthew Witman,[†,¶] Sanliang Ling,[‡,¶] David M. Grant,[‡] Gavin S. Walker,[‡] Sapan Agarwal,[†] Vitalie Stavila,[*,†] and Mark D. Allendorf[*,†]

†*Sandia National Laboratories, Livermore, California 94551, United States*

‡*Advanced Materials Research Group, Faculty of Engineering, University of Nottingham, University Park, Nottingham, UK*

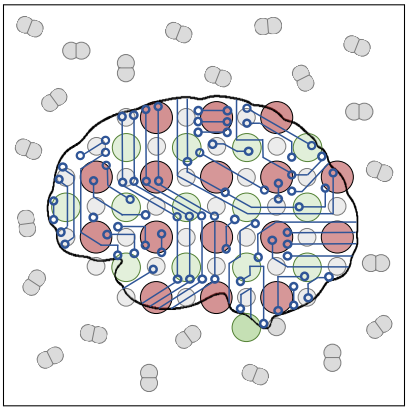
¶*Contributed equally to this work*

E-mail: vnstavi@sandia.gov; mdallen@sandia.gov

Abstract

An open question in the metal hydride community is whether there are simple, physics-based design rules that dictate the thermodynamic properties of these materials across the variety of structures and chemistry they can exhibit. While black box machine learning-based algorithms can predict these properties with some success, they do not directly provide the basis on which these predictions are made, therefore complicating the *a priori* design of novel materials exhibiting a desired property value. In this work we demonstrate how feature importance, as identified by a gradient boosting tree regressor, uncovers the strong dependence of the metal hydride equilibrium H_2 pressure on a volume-based descriptor that can be computed from just the elemental composition of the intermetallic alloy. Elucidation of this simple structure-property relationship is valid across a range of compositions, metal substitutions, and structural classes exhibited by intermetallic hydrides. This permits rational targeting of novel intermetallics for high-pressure hydrogen storage (low-stability hydrides) by their descriptor values, and we predict a known intermetallic to form a low-stability hydride (as confirmed by density functional theory calculations) that has not yet been experimentally investigated.

Graphical TOC Entry



Development of renewable energy technologies is more critical now than ever to avoid some of the catastrophic consequences of climate change.[1] Hydrogen is a clean energy carrier poised to make an impact throughout the renewable energy space, but storage and transportation of hydrogen gas (H_2) remains a significant challenge.[2,3] Decades of research have been devoted to storing hydrogen more economically/efficiently, and metal hydrides[4–10] are one of the most extensively studied materials for applications in H_2 storage for transportation, H_2 compressors, thermal energy storage, and H_2 getters.[11–14] Their practical applicability varies widely as a function of their thermodynamic properties which, when combined with other factors such as sustainability, cost, kinetics, capacity, has lead to thousands of metal hydrides being investigated experimentally. Thus, an open question is whether there exist simple materials design rules that dictate their thermodynamic properties across their varying chemical and structural space. Correlations have been elucidated from various experimental results[15–20] and empirical design rules derived,[21–25] such as the pressure dependence on interstitial volumes for a given intermetallic series. Computational screenings have also been performed,[26–28] but this problem has received comparatively little attention from a "big-data" perspective. In other energy related applications such as hydrogen storage or xenon/krypton separations in porous materials, big data approaches have been able to identify relatively simple materials descriptors and models that predict thermodynamic performance across a wide swathe of material space.[29,30] Can similar results be achieved for metal hydrides?

Statistical or machine learning (ML) techniques have the potential to answer this question and, despite sometimes lingering skepticism over their utility, are now continually employed to solve problems in the physical sciences.[31] Some prominent examples include generative models for drug design,[32] prediction of conductive metal organic frameworks,[33] classification of stable perovskites,[34] among others.[35–40] One natural way to segregate these studies is by those that use "black-box" vs. "explainable" ML techniques.[41] Black-box ML techniques are well-suited to make accurate predictions of materials properties but provide little visibility

into how the algorithm utilizes feature space to make a prediction, a potential limitation that explainable ML techniques seek to address. For example, "explainable" insight can be derived by simply extracting a given feature's importance when training a model[42] or by using ML methods whose mapping from features to prediction is directly interpretable by design.[43,44]

A few studies have applied ML techniques to make black-box predictions on the thermodynamic properties of metal hydrides. Rahnama et al. trained a model to use measured properties of metal hydrides to predict other measured properties.[45,46] However, this is not particularly predictive since one would have to perform an experiment or simulation to measure the materials' properties to use the model, and at that point the properties would be known, negating the need for a model. One of their main conclusions from this approach is that "composition formula was found to be an insignificant variable",[45] which is surprising given the large body of literature on doping and destabilization of metal hydrides. Hattrick-Simpers et al. trained a model on the Department of Energy's experimental metal hydride (HydPARK) database to predict hydriding enthalpies solely from the composition of the intermetallic phase, which was then used as a surrogate model to quickly evaluate the performance of novel intermetallic compositions for use in hydrogen compressors.[47] These studies did not exploit insights from explainable ML to determine what properties of intermetallic compounds dictate their thermodynamic performance.

In this work we also train an ML model on the HydPARK database using features derived *solely* from the intermetallic composition (no structural or hydride information); however, our major contribution is to use feature importance from gradient boosting trees to gain "explainable" insight into simple structure-property relationships that govern the thermodynamics of hydride formation. While our ML model can accurately predict the room temperature equilibrium H_2 pressure of intermetallic hydrides, its interpretability allows us to generalize the pressure dependence on the lattice volume in the $LaNi_5$ substitution series (a historically known design correlation[18–20]) over a surprisingly wide range of intermetallic chemistries and

structures. This unifies disparate experimental results onto a single structure-property relationship. Its elucidation provides thermodynamic insight into the underpinnings of the ML model predictions, which we further corroborate with density functional theory (DFT). We then utilize this to predict a known intermetallic for high-pressure H_2 storage applications whose hydriding properties have not yet been experimentally tested.

Metal hydride database selection.— Computational databases such as the Materials Project (MP),[48] OQMD,[49] and AFLOWLIB[50] contain large numbers of crystal structures and various DFT computed properties. However, only certain thermodynamic properties of metal hydrides can be readily calculated from electronic structure simulations, such as the enthalpy of dissociation of the hydride phase, ΔH. The entropy of dissociation, ΔS, on the other hand, requires a very computationally intensive estimate of the vibrational density of states and is impractical to compute for hundreds or thousands of structures. In contrast, the HydPARK database contains thermodynamic data that is not easily calculated but readily measurable, such as the equilibrium pressure of H_2, P_{eq}, at a given temperature, T. Therefore, ΔS can be computed using experimental data and the van't Hoff relationship,

$$\ln P_{eq} = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}. \tag{1}$$

Often we are interested in predicting the equilibrium H_2 pressure, $\ln P^o_{eq}$ (@25°C), as it indicates how much H_2 a material can deliver at room temperature and provides a standardized metric for comparing metal hydrides that accounts for *both* entropic and enthalpic hydriding effects.[51,52] Figure 1 shows the strong enthalpy-entropy correlation (which is not unique to this application[53,54]) for the hydriding reaction with a Spearman rank-order correlation coefficient (SC) of 0.68 with p-value < 0.01. Even if we specify some desired ΔH, e.g. 27 kJ (mol H_2)$^{-1}$, variations in ΔS yield $\ln P^o_{eq} \pm 10$. Thus assuming ΔS is just equal to the molar entropy of gaseous hydrogen (130.4 J (mol H_2)$^{-1}$ K^{-1}) is overly simplistic and ignores experimentally known secondary contributions.[17] Additionally, the HydPARK database con-

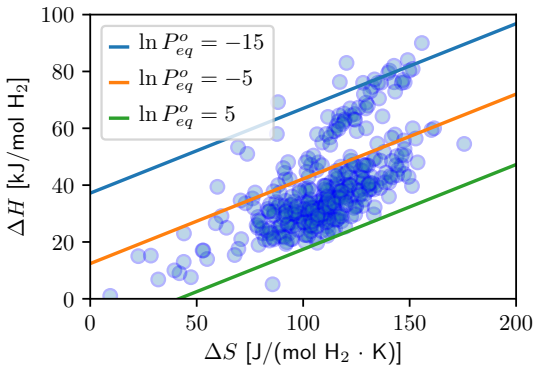


Figure 1: ΔH vs. ΔS for all complete entries in the HydPARK database and contour values for constant $\ln P^o_{eq}$.

tains many complex stoichiometries for which the exact crystal structure is not known and therefore cannot be included in a computational database of specific crystal structures. We therefore proceed with the HydPARK database for developing our ML model.

Next we clean and prepare the HydPARK database before training an ML model with additional details included in the Supplementary Information section S1. Briefly, we remove compositions with missing or unusable data (for which $\ln P^o_{eq}$ cannot be calculated), thereby reducing the size of the dataset from 2732 entries to 570 entries. Further investigation reveals significant spread in the reported experimental data for duplicate compositions (e.g. 6 different $CaNi_5$ entries) as well as incorrectly collected data in HydPARK, something we expect will challenge the development of a highly accurate ML model. We therefore remove duplicates, yielding 409 *unique* compositions, while minimizing the bias introduced by this process (more details in S1); however, these literature references need to be revisited individually and experiments repeated when necessary. Additionally, there is a large imbalance in both the distribution of thermodynamic properties (Table 1) and the population sizes of different metal hydride classes. For example, less than 3% of complete and unique entries are complex hydrides[55,56] which we discard since these ∼10 sample points are insufficient for an ML model to learn from. S12 contains our final version of the "ML ready" HydPARK database.

Table 1: The distribution of $\ln P^o_{eq}$ for complete and unique compositions in HydPARK. Even within the $\ln P^o_{eq} > -10$ subset, the data is non-uniformly distributed as shown in Figure 2d.

$\ln P^o_{eq}$ values	% of database
$\ln P^o_{eq} < -20$	2.9 %
$-20 < \ln P^o_{eq} < -10$	12.7 %
$-10 < \ln P^o_{eq}$	84.4 %

Machine learning with feature importance.— The Magpie code[57] was used to generate a set of 145 features derived solely from the intermetallic chemical composition for each HydPARK material. Therefore, no structurally specific features were included for training the ML model (other than what is implicitly encoded by the properties of the material's constituent elements), an approach which has shown great success in a variety of materials science applications.[57,58] In other words, we try to discover whether the thermodynamic properties of intermetallic hydrides can be *a priori* predicted from the intermetallic composition without any information about the hydride composition or structure. Next a gradient boosting tree regressor (the best performer in comparison to other regression techniques as shown in S2) was trained using scikit-learn[59] to predict $\ln P^o_{eq}$. A 10-fold validation was performed, and the combined test and train sets for each of the 10 models is shown in Figure 2a-b. As quantified by the mean absolute error (MAE), the model generalizes especially well to predict materials with $\ln P^o_{eq}$ values most commonly occurring in the dataset (Figure 2d). For materials in the wings of the $\ln P^o_{eq}$ distribution, the model can fit these samples well during train time but does not generalize as well during validation.

The mathematical foundation of gradient boosting trees is covered extensively in the literature.[60] This technique is particularly useful because it permits the calculation of feature importance, which generally scores how valuable each feature was in the construction of the boosting trees. Figure 2c shows the averaged Relative Importance (note each importance value is scaled by the constant factor that sets the most important feature to 100) across all 10-fold validations. While most Magpie names are intuitive, S3 explains the naming

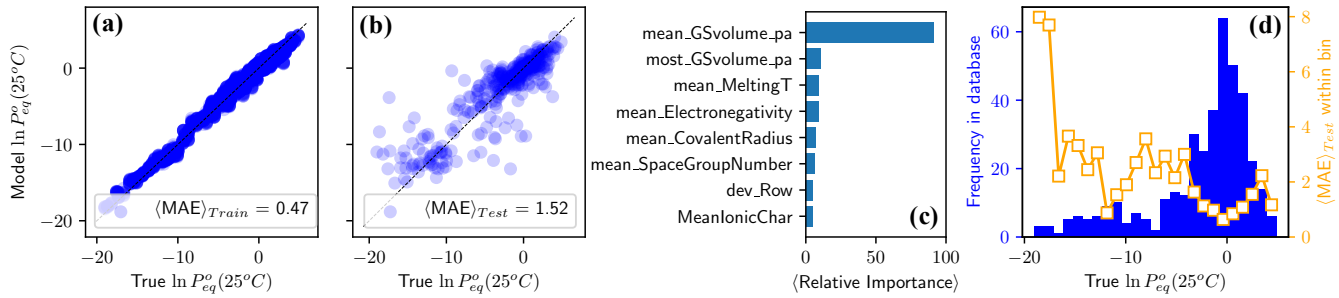


Figure 2: (a) Train and (b) test performance of the ML model on predicting $\ln P^o_{eq}$. Each plot contains the overlaid data from each 10-fold validation experiment, with the ⟨MAE⟩ corresponding to the average over all 10 models. (c) Among the 145 Magpie generated features, the mean relative feature importance of the 8 most important features over all 10 models is shown. (d) The histogram of the $\ln P^o_{eq}$ values for all HydPARK materials is overlaid with the test ⟨MAE⟩ computed within each individual bin (with bin width = 1).

of individual features in more detail. The most important descriptor, `mean_GSvolume_pa` (ν^{Magpie}_{pa}), is computed as

$$\nu^{Magpie}_{pa} = \sum f_i \nu_i, \tag{2}$$

where f_i is the composition fraction of element i, and ν_i is the volume occupied per atom in the ground state elemental solid of species i. In other words, it is Magpie's estimation of the specific volume per atom for a given composition. Notably, its average relative importance across all 10-fold validation models is close to 100 (not exactly 100 since one model yields `most_GSvolume_pa` as the most important), indicating that it is the most important descriptor regardless of the test/train split; it even remains so when removing data from the training set (see S4). We discuss secondary features in more detail later as they constitute important features when training individual models to predict $\ln P^o_{eq}$'s constituent components, ΔS and ΔH. Ultimately, the interpretable model suggests that a simple volume-based descriptor may be the *single* most useful feature, a surprising result given the wide ranging chemistries and structure types displayed by these materials.

An intermetallic hydride design principle.— The high importance of the ν^{Magpie}_{pa} descriptor warrants further investigation into a structurally specific volume descriptor. We cross-reference the compositions in the training set with the MP database to identify ∼80

overlapping structures for which we extract the DFT relaxed crystal structure corresponding to the lowest formation energy per atom. We then derive an analogous descriptor to ν_{pa}^{Magpie} based on the cell volume, V_{cell}, and the number of atoms in the cell, n_{atoms}, from the MP crystal structure

$$\nu_{pa}^{MP} = \frac{V_{cell}}{n_{atoms}}. \tag{3}$$

Hence ν_{pa} refers to the volume per atom in a crystal which can be either estimated (ν_{pa}^{Magpie}) or computed by DFT (ν_{pa}^{MP}). Figure 3a first compares $\ln P_{eq}^{o}$ vs. V_{cell}, from which DBSCAN,[59,61] a density-based clustering (unsupervised learning) algorithm, highlights distinct bands of materials obeying the same log-linear relationship. The common feature in these different colored "classes" is the number of atoms in the lattice cell. Thus, converting V_{cell} to ν_{pa}^{MP} in Figure 3b results in a majority of the data collapsing onto a single log-linear trend. And while ν_{pa}^{Magpie} does not explicitly encode any structural information, it still represents a highly correlated structure-property relationship (Figure 3c) because it reasonably predicts the true ν_{pa}^{MP} (S5), with small discrepancies arising when the volume of mixing is non-negligible. Note that the data associated with each material can be found in the ML ready HydPARK database (`https://github.com/mwitman1/MetalHydrideML`).

Figure 3 also shows how this result builds upon previous investigations of the volume dependence of thermodynamic properties in intermetallic hydrides. Cuevas, Zhang, and Reilly all demonstrated the log-linear dependence of H_2 plateau pressure on V_{cell} for $LaNi_5$ substituted materials (i.e. the AB5-type intermetallic).[18–20,62] Lundin et al. took these efforts a step further and, by considering local structure, correlated $\ln P_{eq}^{o}$ with the volume of interstices in AB5 substituted materials as well as cubic AB systems.[15,16] The advantage of using a data-driven approach with explainable ML is that we are able to recognize how this structure-property relationship encompasses different chemistries, different intermetallic classes (AB, AB2, AB5, solid solution), and different substitutions at A and B sites. For example, the data of Smith et al.[17] in Figure 3 corresponds to the "miscellaneous" hydride class R_6Fe_{23} [R=Ho,Er,Lu], whereby rare earth substitution expands ν_{pa} and leads to an

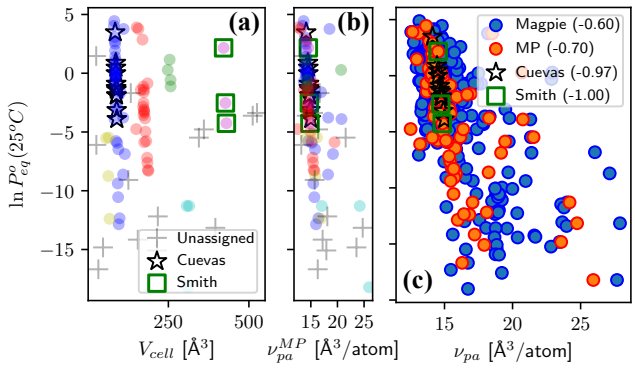


Figure 3: (a) The DBSCAN clustering of $\ln P^o_{eq}$ vs. V_{cell} for the ∼80 materials common to both the MP and HydPARK databases. Colored points correspond to materials identified with membership to the cluster, gray crosses represent unassigned materials, and open symbols correspond to various experimental results.[17,20] (b) Converting V_{cell} to ν_{pa}^{MP} collapses the subclasses onto a single log-linear relationship. (c) Compares the ν_{pa} structure-property descriptor for both Magpie and MP computed values with the SC in parentheses (all p-values < 0.01.

$\ln P^o_{eq}$ that collapses to the same correlation. Furthermore, we arrive at this relationship without any *a priori* knowledge of the intermetallic or hydride structures and instead only require the intermetallic composition (contrast this with the conclusions of Ref. 45). We stress that this simple structure-property relationship is less predictive than, and therefore not a substitute for, the ML model (see S7); rather, it provides an avenue for exploring *why* the ML model can predict the thermodynamic properties of metal hydrides over a range of chemistry and structural space.

A thermodynamic basis for ν_{pa}.— In order to understand *why* this structure property relationship exists, our discussion first returns to the individual contributions of $\Delta H/(RT^\circ)$ and $\Delta S/R$ to $\ln P^o_{eq}$. Figure 4a contrasts the strong negative correlation of $\Delta H/(RT^\circ)$ (SC $= -0.82$) and the very weak negative correlation $\Delta S/R$ (SC $= -0.23$) with $\ln P^o_{eq}$. In other words, the $\Delta H/(RT^\circ)$ contribution systematically decreases over a wider range of values than the $\Delta S/R$ contribution as evidenced by their ratio (Figure 4b). Consequently, there exists a strong structure-property relationship between ν_{pa}^{Magpie} and ΔH (Figure 4c), and an ML model of comparable accuracy can also be trained to predict ΔH (see S6) with ν_{pa}^{Magpie} as the dominantly important feature. A reasonably accurate model for ΔS can also be trained

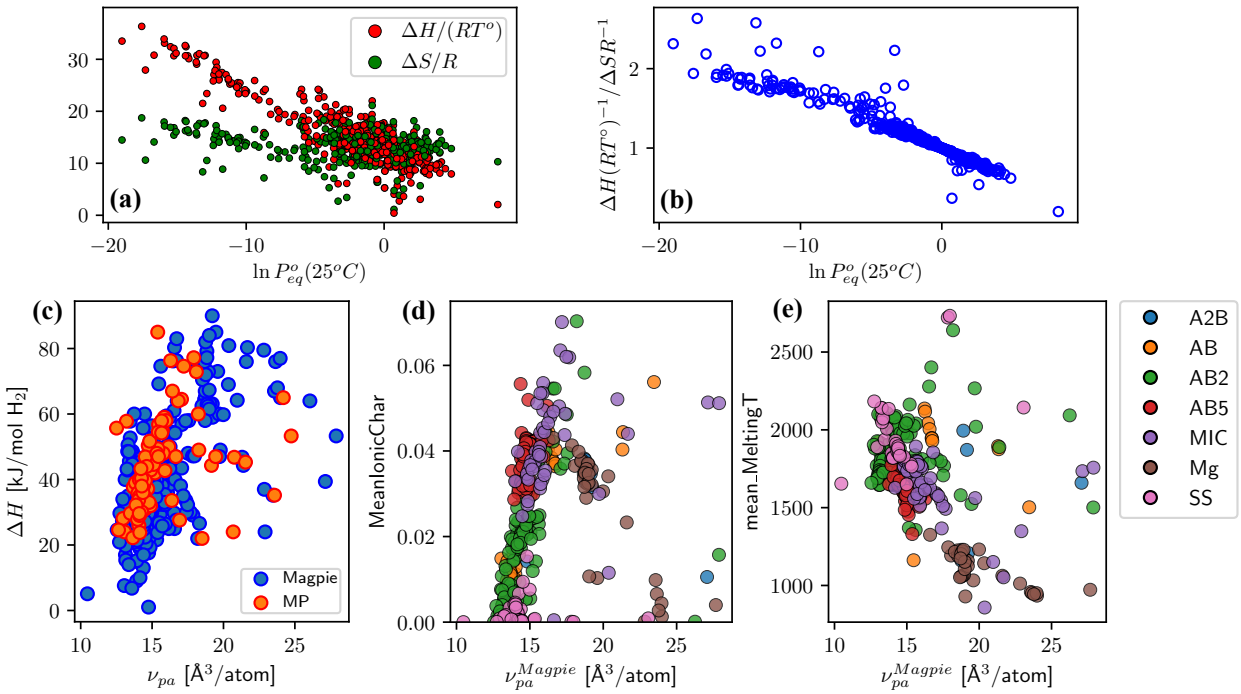


Figure 4: (a) $\Delta H/RT^o$ ($T^o = 25°$C) and $\Delta S/R$ plotted vs. $\ln P^o_{eq}$, with SC $= -0.82$ and SC $= -0.23$, respectively. (b) The ratio of these contributions plotted vs. $\ln P^o_{eq}$. (c) The correlation between ΔH and ν_{pa}. (d) Magpie's pairwise electronegativity differences (MeanIonicChar) and (e) mean melting temperature vs. ν^{Magpie}_{pa}, color-coded by intermetallic class.

(see S6) but with significantly different feature importance.

Empirical modeling of binary alloy formation enthalpies[21,63,64] has utilized cellular models incorporating properties like electronegativity differences and the difference in electron density at the boundary between dissimilar atoms. Extension to ternary hydrogen-containing alloys often relied on knowledge of structurally specific features.[16,22,25] ΔH has also been rationalized in terms of a qualitative rule of reversed stability: substituting La or Ni to stabilize the binary intermetallic results in a less stable hydride phase.[22] Interestingly, the same insights from these different modeling efforts can be qualitatively rationalized across the HydPARK materials via our data-driven approach. Specifically, ν_{pa} is a "synthetic feature" that encodes many of the other chemically specific features that affect ΔH; therefore, this feature can even be removed when training the ML model without significant loss in accuracy (S7). As a simple illustration in Figure 4d-e, materials with larger ν_{pa} tend to have

larger average pairwise electronegativity differences between elements (Magpie's MeanIonicChar feature), as well as lower mean melting temperatures. We can rationalize such trends as indicators for increasing hydrogen absorption strength or decreased energy penalty for lattice deformation, which we investigate further with DFT.

These structure-property relationships are also invaluable for outlier identification. For example, note that the thermodynamic properties of materials with $\nu_{pa} > \sim 17$ Å^3/atom deviate significantly from the trend exhibited by materials with $\nu_{pa} < 17$ Å^3/atom. These exactly correspond to the materials for which the model generalizes poorly (Figure 2b) due to a lack of materials in this descriptor regime, an insight only derived because the simple structure property relationships elucidated by the ML model's feature importance. This breakdown of the log-linear ($\ln P^o_{eq}$) or linear (ΔH) correlation at this critical threshold suggests that these materials require a different physical understanding than the ν_{pa} structure-property relationship, but the lack of data in this regime must be addressed before ML models have the chance to provide further data-driven insight. Other secondary benefits of the model's interpretability are discussed in S7, S8, and S10.

Thermodynamic insights from DFT.— We can further corroborate our insights into the ν_{pa} structure-property relationship by examining A site substitutions in the $LaNi_5$ series with DFT. First we define E_f as the formation energy of the intermetallic alloy with respect to the elemental crystals, ΔE_{def} as the energy penalty required to deform the intermetallic lattice to accommodate H absorption, and ΔE_H as the binding energy between hydrogen and metal atoms in the hydride lattice (see S11 for the definition). V is the volume of the hydrided lattice, and V_0 is the volume of the intermetallic lattice. Specific details on these calculations are provided in S11,[65–71] and we summarize DFT computed ΔH, E_f, ΔE_{def} and ΔE_H in Table 2. Note that in these calculations the final state is the hydride and the initial state is the intermetallic, i.e. the Δ's correspond to the hydriding reaction, not the dehydriding reaction. We also note that we considered a hydride composition of ANi_5H_7 (A = U, Ce or La), which corresponds to a hydrogen/metal ratio of 1.17 and is close to the

maximum hydrogen uptake reported for $LaNi_5$.[71]

Table 2: DFT computed properties for A site substitutions in the $LaNi_5$ series, including UNi_5 which does not exist in HydPARK. ΔH, ΔE_{def}, ΔE_H [=] kJ/(mol·H_2) and E_f[=]meV/atom.

	ν_{pa}	ΔH	E_f	ΔE_{def}	ΔE_H	V/V_0
UNi_5	13.17	-0.60	-285	65.2	-65.8	1.278
$CeNi_5$	13.76	-20.5	-353	49.3	-69.8	1.266
$LaNi_5$	14.38	-36.1	-224	44.3	-80.5	1.256

This demonstrates the experimental trend of increased hydride stability (ΔH) with ν_{pa} while the individual energy terms yield additional insight. First, there is no apparent correlation with E_f. Rather, increasing ν_{pa} more importantly indicates a propensity for a lower energy penalty of deformation ΔE_{def}, which can be rationalized by the reduced volume expansion required to form the hydride phase. It should be pointed out that in addition to the magnitude of volume expansion V/V_0, the elastic modulus of the intermetallic will also affect the energy penalty of deformation ΔE_{def}: assuming the same volume expansion, a stiffer intermetallic would require a bigger energy penalty to deform the lattice in comparison with a less stiff intermetallic. Second, the binding energy between hydrogen and metal atoms in the hydride lattice increases with ν_{pa}, which can be rationalized by the lower electronegativity of La in comparison with U (1.1/La vs. 1.38/U on Pauling scale), i.e. the binding between hydrogen and $LaNi_5$ is expected to be more ionic than that between hydrogen and UNi_5. Both of these effects promote greater hydride stability, and it is these trends which underpin the ν_{pa} structure-property relationship. The general trend of ΔH as a function of ν_{pa} may also be inferred from simple "chemical intuition", i.e. intermetallic alloys with larger ν_{pa} usually consist of elements with larger atomic radii, and elements with larger atomic radii tend to have lower electronegativities (see S9), because of reduced attractions to electrons in the outer valence shell. The result is that an intermetallic alloy with a larger ν_{pa} value tends to form more ionic bonds with hydrogen in the hydride phase, and therefore the enthalpy of formation of the hydride, ΔH, is usually bigger than that of an intermetallic alloy with

a smaller ν_{pa} value. We note that for intermetallic alloys with similar ν_{pa} values, the energy penalties of deforming the alloy lattice to accommodate the chemically absorbed hydrogen can be very different, which may result in a large scattering of the ΔH values.

ML informed targeting of a novel hydride phase.— We propose UNi_5 in Table 2 for two reasons. First, based on our ML informed results, we predict this A site substitution to $LaNi_5$ to reduce the stability of the metal hydride phase since it significantly reduces ν_{pa} (U has a smaller atomic radius than La). Second, UNi_5 is an experimentally synthesized intermetallic[72] in the Inorganic Crystal Structure Database (ICSD), but its hydrided form has not yet been reported in the ICSD nor is it contained in the HydPARK database (potentially due to the large H_2 pressures that may be necessary to synthesize it near room temperature). As confirmed by our DFT calculations, UNi_5H_7 has a very small reaction enthalpy of -0.6 kJ/(mol·H_2) and therefore should be a low stability hydride. As seen from Figure 1, UNi_5H_7, should it be synthesized in the future, would be one of the least stable hydrides in the entire HydPARK database. We note the hydriding reaction enthalpy of a metal alloy may differ depending on the amount of hydrogen that is absorbed. Nevertheless, the qualitative knowledge generated by our interpretable ML provides a path to rationally target novel hydride phases with a desired thermodynamic property, i.e. very low stability for high-pressure H_2 or hydrogen isotope storage applications.[73,74]

In conclusion, utilizing the HydPARK experimental metal hydride database, we have trained an ML model to predict the equilibrium plateau pressure, one of the most relevant thermodynamic quantities for practical applications which is also unique to this database (i.e. not contained in any computationally derived databases due to its dependence on ΔS). Exploiting the explainability of gradient boosting trees with our data-driven approach enables several key understandings. First, basic thermodynamic insight into intermetallic metal hydride formation can be derived from features generated only from the elemental composition of the intermetallic phase (a particularly useful capability if the exact crystal structure of an experimental compound is not known). Past experimental studies have elucidated

the dependence of equilibrium H_2 pressure on cell volume or structurally specific interstitial volumes, and the identification of our ν_{pa} structure-property relationship encompasses these observations across a range of intermetallic chemistries and structures. The thermodynamic basis for this correlation is attributed to the underlying structure-property relationship between ΔH and ν_{pa}; furthermore, materials not described by this simple structure-property relationship can now be investigated to determine the chemistry behind their outlying behavior. All of these insights are predicated on the physical interpretability of an ML model, which, when corroborated with DFT calculations, is ultimately used to propose a novel hydride of a known intermetallic with significant potential as a high-pressure hydrogen storage material.

Furthermore, we utilized a noisy, imbalanced database which required multiple heuristic steps to process and clean. This simply highlights that statistical learning techniques still have the power to help extract useful information in materials science applications, even when approximations must be made to prepare the training data. Future efforts in this space will benefit *greatly* from a concerted effort of the metal hydride community to centralize the reporting of experimental measurements such as ΔH, ΔS, P_{eq}, T, V_{cell} (if possible), etc.[75] There should also be a standardized method for reporting more complex phenomena such as hysteresis, existence of multiple hydride phases, sloping plateaus, etc. This could be incorporated into the framework of one of the many existing materials databases (MP, OQMD, AFLOWLIB) which would better position data-driven/ML based approaches to impact the discovery and understanding of metal hydrides. Less than 20% of the HydPARK database was used due to missing and/or duplicate information. Several errors were found in the dataset from a cursory manual investigation of the literature references therein, and standardized reporting in a central repository could help avoid such inconsistencies. Moreover, if more *complete* material entries existed in the HydPARK database with larger volumes per atom, our explainable ML approach might be able to elucidate the structure-property relationship(s) that differentiate them from the ν_{pa} discussed in this work. However, this will

be an unlikely accomplishment from an ML perspective until more/better data is acquired in this regime. Having gained explainable insights into the thermodynamics of hydride formation, future efforts can now be directed towards explainable ML models that discern whether a given composition (out of the essentially infinite number that may exist) will form a hydride and, if so, what its hydrogen content may be. We propose that combining all of these efforts will result in the data-driven discovery of novel, high-performing hydrides.

Acknowledgements.— 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 Energys National Nuclear Security Administration under contract DE-NA-0003525. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, nor usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. We acknowledge the use of the Athena supercomputer through the HPC Midlands+ Consortium, and the ARCHER supercomputer through membership of the UK's HPC Materials Chemistry Consortium, which are funded by EPSRC Grants No. EP/P020232/1 and No. EP/R029431/1, respectively.

References

(1) Masson-Delmotte, V.; Zhai, P.; Pörtner, H. O.; Roberts, D.; Skea, J.; Shukla, P.; Pirani, A.; Moufouma-Okia, W.; C.Péan,; Pidcock, R. et al. *Global warming of* 1.5°*C. An IPCC Special Report on the impacts of global warming of* 1.5°*C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change*; 2018.

(2) Schlapbach, L.; Züttel, A. Hydrogen-storage materials for mobile applications. *Nature* **2001**, *414*, 353–358.

(3) Yang, J.; Sudik, A.; Wolverton, C.; Siegel, D. J. High capacity hydrogen storage materials: Attributes for automotive applications and techniques for materials discovery. *Chem. Soc. Rev.* **2010**, *39*, 656–675.

(4) Dilts, J. A.; Ashby, E. C. Thermal decomposition of complex metal hydrides. *Inorg. Chem.* **1972**, *11*, 1230–1236.

(5) Stetson, N. T.; Yvon, K.; Fischer, P. Structure of the complex metal hydride $BaReH_9$. *Inorg. Chem.* **1994**, *33*, 4598–4599.

(6) Chen, Y.; Williams, J. Formation of metal hydrides by mechanical alloying. *J. Alloys Compd.* **1995**, *217*, 181–184.

(7) Zaluski, L.; Zaluska, A.; Ström-Olsen, J. Nanocrystalline metal hydrides. *J. Alloys Compd.* **1997**, *253-254*, 70–79.

(8) Mohtadi, R.; Orimo, S. I. The renaissance of hydrides as energy materials. *Nat. Rev. Mater.* **2016**, *2*, 1–16.

(9) Amsler, M.; Hegde, V. I.; Jacobsen, S. D.; Wolverton, C. Exploring the high-pressure materials genome. *Phys. Rev. X* **2018**, *8*, 041021.

(10) Schneemann, A.; White, J. L.; Kang, S.; Jeong, S.; Wan, L. F.; Cho, E. S.; Heo, T. W.; Prendergast, D.; Urban, J. J.; Wood, B. C. et al. Nanostructured metal hydrides for hydrogen storage. *Chem. Rev.* **2018**, *118*, 10775–10839.

(11) Schlapbach, L.; Züttel, A. Hydrogen-storage materials for mobile applications. *Nature* **2001**, *414*, 353–358.

(12) Lototskyy, M.; Yartys, V.; Pollet, B.; Bowman, R. Metal hydride hydrogen compressors: A review. *Int. J. Hydrogen Energy* **2014**, *39*, 5818–5851.

(13) Sartori, S.; Cuevas, F.; Latroche, M. Metal hydrides used as negative electrode materials for Li-ion batteries. *Appl. Phys. A* **2016**, *122*, 135.

(14) Bellosta von Colbe, J.; Ares, J.-R.; Barale, J.; Baricco, M.; Buckley, C.; Capurso, G.; Gallandat, N.; Grant, D. M.; Guzik, M. N.; Jacob, I. et al. Application of hydrides in hydrogen storage and compression: Achievements, outlook and perspectives. *Int. J. Hydrogen Energy* **2019**, *44*, 7780–7808.

(15) Lundin, C.; Lynch, F.; Magee, C. A correlation between the interstitial hole sizes in intermetallic compounds and the thermodynamic properties of the hydrides formed from those compounds. *J. Less Common Met.* **1977**, *56*, 19–37.

(16) Magee, C. B.; Liu, J.; Lundin, C. E. Relationships between intermetallic compound structure and hydride formation. *J. Less Common Met.* **1981**, *78*, 119–138.

(17) Smith, H.; Wallace, W.; Craig, R. An investigation of $R_6Fe_{23}H_x$ thermodynamics. *J. Less Common Met.* **1983**, *94*, 89–93.

(18) Reilly, J. Chemistry of intermetallic hydrides. 180th Meet. Electrochem. Soc. 1991.

(19) Zhang, W.; Kumar, M. P.; Visintin, A.; Srinivasan, S.; Ploehn, H. J. A microcalorimetric investigation of the thermodynamics and kinetics of hydriding-dehydriding reactions. *J. Alloys Compd.* **1996**, *242*, 143–152.

(20) Cuevas, F.; Joubert, J.-M.; Latroche, M.; Percheron-Guégan, A. Intermetallic compounds as negative electrodes of Ni/MH batteries. *Appl. Phys. A Mater. Sci. Process.* **2001**, *72*, 225–238.

(21) Miedema, A. The electronegativity parameter for transition metals: Heat of formation and charge transfer in alloys. *J. Less Common Met.* **1973**, *32*, 117–136.

(22) Van Mal, H.; Buschow, K.; Miedema, A. Hydrogen absorption in $LaNi_5$ and related compounds: Experimental observations and their explanation. *J. Less Common Met.* **1974**, *35*, 65–76.

(23) Buschow, K.; Van Mal, H.; Miedema, A. Hydrogen absorption in intermetallic compounds of thorium. *J. Less Common Met.* **1975**, *42*, 163–178.

(24) Westlake, D. Hydrides of intermetallic compounds: A review of stabilities, stoichiometries and preferred hydrogen sites. *J. Less Common Met.* **1983**, *91*, 1–20.

(25) Yan-Bin, W.; Northwood, D. O. Calculation of the enthalpy of metal hydride formation. *J. Less Common Met.* **1987**, *135*, 239–245.

(26) Alapati, S. V.; Johnson, J. K.; Sholl, D. S. Identification of destabilized metal hydrides for hydrogen storage using first principles calculations. *J. Phys. Chem. B* **2006**, *110*, 8769–8776.

(27) Alapati, S. V.; Karl Johnson, J.; Sholl, D. S. Using first principles calculations to identify new destabilized metal hydride reactions for reversible hydrogen storage. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1438.

(28) Wolverton, C.; Siegel, D. J.; Akbarzadeh, A. R.; Ozoliš, V. Discovery of novel hydrogen storage materials: an atomic scale computational approach. *J. Phys.: Condens. Matter* **2008**, *20*, 064228.

(29) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J.; Siegel, D. J. Theoretical limits of hydrogen storage in metalorganic frameworks: Opportunities and trade-offs. *Chem. Mater.* **2013**, *25*, 3373–3382.

(30) Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. What are the best materials to separate a xenon/krypton mixture? *Chem. Mater.* **2015**, *27*, 4459–4475.

(31) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.

(32) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276.

(33) He, Y.; Cubuk, E. D.; Allendorf, M. D.; Reed, E. J. Metallic metal-organic frameworks predicted by the combination of machine learning methods and ab initio calculations. *J. Phys. Chem. Lett.* **2018**, *9*, 4562–4569.

(34) Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; Ouyang, R.; Musgrave, C. B.; Ghiringhelli, L. M.; Scheffler, M. New tolerance factor to predict the stability of perovskite oxides and halides. *Sci. Adv.* **2019**, *5*, eaav0693.

(35) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 1–4.

(36) Deml, A. M.; O'Hayre, R.; Wolverton, C.; Stevanović, V. Predicting density functional theory total energies and enthalpies of formation of metal-nonmetal compounds by linear regression. *Phys. Rev. B* **2016**, *93*, 085142.

(37) Ziletti, A.; Kumar, D.; Scheffler, M.; Ghiringhelli, L. M. Insightful classification of crystal structures using deep learning. *Nat. Commun.* **2018**, *9*, 2775.

(38) Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.; Persson, K. A.; Ceder, G.; Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **2019**, *571*, 95–98.

(39) Davies, D. W.; Butler, K. T.; Walsh, A. Data-driven discovery of photoactive quaternary oxides using first-principles machine learning. *Chem. Mater.* **2019**, *31*, 7221–7230.

(40) AlQuraishi, M. End-to-end differentiable learning of protein structure. *Cell Syst.* **2019**, *8*, 292–301.e3.

(41) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.

(42) Muraoka, K.; Sada, Y.; Miyazaki, D.; Chaikittisilp, W.; Okubo, T. Linking synthesis and structure descriptors from a large collection of synthetic records of zeolite materials. *Nat. Commun.* **2019**, *10*, 4459.

(43) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, *2*, 083802.

(44) Agarwal, S.; Hudson, C. M. Probability series expansion classifier that is interpretable by design. *arXiv:1710.10301* **2017**, 1–6.

(45) Rahnama, A.; Zepon, G.; Sridhar, S. Machine learning based prediction of metal hydrides for hydrogen storage, Part I: Prediction of hydrogen weight percent. *Int. J. Hydrogen Energy* **2019**, *44*, 7337–7344.

(46) Rahnama, A.; Zepon, G.; Sridhar, S. Machine learning based prediction of metal hydrides for hydrogen storage, part II: Prediction of material class. *Int. J. Hydrogen Energy* **2019**, *44*, 7345–7353.

(47) Hattrick-Simpers, J. R.; Choudhary, K.; Corgnale, C. A simple constrained machine learning model for predicting high-pressure-hydrogen-compressor materials. *Mol. Syst. Des. Eng.* **2018**, *3*, 509–517.

(48) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G. et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.

(49) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501–1509.

(50) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L.; Sanvito, S.; Buongiorno-Nardelli, M. et al. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **2012**, *58*, 227–235.

(51) Baldi, A.; Gonzalez-Silveira, M.; Palmisano, V.; Dam, B.; Griessen, R. Destabilization of the Mg-H system through elastic constraints. *Phys. Rev. Lett.* **2009**, *102*, 226102.

(52) Ngene, P.; Longo, A.; Mooij, L.; Bras, W.; Dam, B. Metal-hydrogen systems with an exceptionally large and tunable thermodynamic destabilization. *Nat. Commun.* **2017**, *8*, 1846.

(53) Simon, C. M.; Kim, J.; Lin, L.-C.; Martin, R. L.; Haranczyk, M.; Smit, B. Optimizing nanoporous materials for gas storage. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5499.

(54) Rathee, V. S.; Sidky, H.; Sikora, B. J.; Whitmer, J. K. Role of associative charging in the entropyenergy balance of polyelectrolyte complexes. *J. Am. Chem. Soc.* **2018**, *140*, 15319–15328.

(55) Nakamori, Y.; Miwa, K.; Ninomiya, A.; Li, H.; Ohba, N.; Towata, S.-i.; Züttel, A.; Orimo, S.-i. Correlation between thermodynamical stabilities of metal borohydrides and cation electronegativites: First-principles calculations and experiments. *Phys. Rev. B* **2006**, *74*, 045126.

(56) Milanese, C.; Jensen, T.; Hauback, B.; Pistidda, C.; Dornheim, M.; Yang, H.; Lombardo, L.; Zuettel, A.; Filinchuk, Y.; Ngene, P. et al. Complex hydrides for energy storage. *Int. J. Hydrogen Energy* **2019**, *44*, 7860–7874.

(57) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2016**, *2*, 16028.

(58) Oliynyk, A. O.; Mar, A. Discovery of intermetallic compounds from traditional to machine-learning approaches. *Acc. Chem. Res.* **2018**, *51*, 59–68.

(59) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(60) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer New York: New York, NY, 2009.

(61) others,, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996; pp 226–231.

(62) Chao, B. S.; Klebanoff, L. E. In *Hydrogen Storage Technology: Materials and Applications*; Klebanoff, L., Ed.; CRC Press: Boca Raton, FL, 2013; pp 109–132.

(63) Miedema, A.; Boom, R.; De Boer, F. On the heat of formation of solid alloys. *J. Less Common Met.* **1975**, *41*, 283–298.

(64) Miedema, A. On the heat of formation of solid alloys. II. *J. Less Common Met.* **1976**, *46*, 67–83.

(65) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(66) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(67) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953–17979.

(68) Hobbs, D.; Kresse, G.; Hafner, J. Fully unconstrained noncollinear magnetism within the projector augmented-wave method. *Phys. Rev. B* **2000**, *62*, 11556–11570.

(69) Methfessel, M.; Paxton, A. T. High-precision sampling for Brillouin-zone integration in metals. *Phys. Rev. B* **1989**, *40*, 3616–3621.

(70) Kisi, E.; Buckley, C.; Gray, E. The hydrogen activation of $LaNi_5$. *J. Alloys and Compd.* **1992**, *185*, 369–384.

(71) Lartigue, C.; Bail, A. L.; Percheron-Guegan, A. A new study of the structure of $LaNi_5D_{6.7}$ using a modified Rietveld method for the refinement of neutron powder diffraction data. *J. Less Common Met.* **1987**, *129*, 65–76.

(72) Baenziger, N. C.; Rundle, R. E.; Snow, A. I.; Wilson, A. S. Compounds of uranium with the transition metals of the first long period. *Acta Crystallogr.* **1950**, *3*, 34–40.

(73) Filipek, S. M.; Paul-Boncour, V.; Gu gan, A. P.; Jacob, I.; Marchuk, I.; Dorogova, M.; Hirata, T.; Kaszkur, Z. Synthesis of novel deuterides in several Laves phases by using gaseous deuterium under high pressure. *J. Phys.: Condens. Matter* **2002**, *14*, 11261–11264.

(74) Wang, X.; Bei, Y.; Song, X.; Fang, G.; Li, S.; Chen, C.; Wang, Q. Investigation on high-pressure metal hydride hydrogen compressors. *Int. J. Hydrogen Energy* **2007**, *32*, 4011–4015.

(75) Coudert, F. Materials databases: the need for open, interoperable databases with standardized data and rich Metadata. *Adv. Theory Simul.* **2019**, 1900131.