

Lawrence Berkeley National Laboratory

LBL Publications

Title

Machine-learning and first-principles investigation of lightweight medium-entropy alloys for hydrogen-storage applications

Permalink

<https://escholarship.org/uc/item/2158w9bj>

Journal

International Journal of Hydrogen Energy, 154

ISSN

0360-3199

Authors

Banerjee, Tanumoy

Ji, Kevin

Xia, Weyi

et al.

Publication Date

2025-08-01

DOI

10.1016/j.ijhydene.2025.06.106

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Machine-learning and first-principles investigation of lightweight medium-entropy alloys for hydrogen-storage applications

Tanumoy Banerjee,^{1,2} Kevin Ji,¹ Weyi Xia,¹ Gaoyuan Ouyang,¹ Tyler Del Rose,¹ Ihor Z. Hlova,¹ Benjamin Ueland,¹ Duane D. Johnson,^{1,3} Cai-Zhuang Wang,¹ Ganesh Balasubramanian,⁴ Prashant Singh^{1,*}

¹Ames National Laboratory, US Department of Energy, Iowa State University, Ames, IA 50011, USA

²Lawrence Berkeley National Laboratory, US Department of Energy, Berkeley, CA 94720, USA

³Department of Materials Science & Engineering, Iowa State University, Ames, IA 50011, USA

⁴Department of Mechanical and Industrial Engineering, University of New Haven, West Haven, CT, 06516, USA

Abstract

The transition to a low-carbon economy demands efficient and sustainable energy-storage solutions, with hydrogen emerging as a promising clean-energy carrier and with metal hydrides recognized for their hydrogen-storage capacity. Here, we leverage machine learning (ML) to predict hydrogen-to-metal (H/M) ratios and solution energy by incorporating thermodynamic parameters and local lattice distortion (LLD) as key features. Our best-performing ML model provides improvements to H/M ratios and solution energies over a broad class of ternary alloys (easily extendable to multi-principal-element alloys), such as Ti-Nb-X (X = Mo, Cr, Hf, Ta, V, Zr) and Co-Ni-X (X = Al, Mg, V). Ti-Nb-Mo alloys reveal compositional effects in H-storage behavior, in particular Ti, Nb, and V enhance H-storage capacity, while Mo reduces H/M and hydrogen weight percent by 40-50%. We attributed results in molybdenum-rich alloys to slow hydrogen kinetics, as validated by our pressure-composition isotherm (PCT) experiments on pure Ti and Ti₅Mo₉₅ alloys. Density functional theory (DFT) and molecular simulations also confirm that Ti and Nb promote H diffusion, whereas Mo hinders it, highlighting the interplay between electronic structure, lattice distortions, and hydrogen uptake. Notably, our Gradient Boosting Regression model identifies LLD as a critical factor in H/M predictions. To aid material selection, we present two periodic tables illustrating elemental effects on (a) H₂ wt% and (b) solution energy, derived from ML, and provide a reference for identifying alloying elements that enhance hydrogen solubility and storage.

Keywords: Alloys, ML, DFT, MD, Hydrogen storage, Solution energy

* Corresponding author: psingh84@ameslab.gov/prashant40179@gmail.com

1. Introduction

Transitioning to a low-carbon economy, hydrogen (H) increasingly garners attention as a key clean-energy carrier in the shift toward renewable, efficient, and sustainable energy source. It can be produced through electrolysis, using excess electricity from renewable sources and this stored hydrogen can then be utilized in various ways to generate electricity during periods of high demand or, when renewable generation is low, providing a means to balance these intermittent energy sources [1–3]. Hydrogen has high energy density by weight making it an attractive option for storing and transporting energy efficiently

[4]. However, maximizing hydrogen's potential as an energy carrier faces continued hurdles, especially in storage and transportation. Metal hydrides are attractive H-storage materials because of their high storage capacities and favorable properties [5,6].

Complex solid-solution alloys (CSA) or multi-principal-element alloys (MPEA) are one such class of materials with vast numbers of unexplored compositions for adjusting desired properties, like H-storage behavior. However, large CSA design space limits efficient exploration using Edisonian “trial-and-error” methods in experiments and computations [7–12]. Artificial-intelligence (AI)/machine-learning (ML) methods, aided by extensive databases of material structures and properties, has extended design space by enabling robust high-throughput predictions of material properties. At the same time, it reveals correlations and patterns between physical factors that conventional methods might overlook.

AI/ML applications to predict improved metal hydrides for hydrogen storage is a complex and growing area of research. Several studies have focused on developing AI/ML platforms to enhance the property prediction for increasing hydrogen storage and its efficiency [13–16]. Consecutive regression and multi-class classification algorithms were used to predict H-weight percentage (wt%) and classify the hydride class, the key parameters for assessing H-storage capacity of metal hydrides [13,17]. By leveraging ML, our aim is to categorize metal hydrides via their characteristics and properties, yielding insights into identifying more efficient H-storage materials [17]. Furthermore, utilizing hybrid-ML framework using deep multi-layer perceptron (MLP) in identifying best-in-class H-storage metal hydrides demonstrates effectiveness of combining different ML approaches to enhance predictive accuracy [14] by overcoming associated high dimensionality of data and challenges imposed by the scarcity of experimental data. Literatures show that alloy structure and compositions along with several thermodynamic parameters are key to achieving higher H-storage capacity [9,15,16,18,19]. Leveraging explainable ML, Witman, *et al.* [15] showed that volume-based descriptor (calculated per atom) has a significantly strong correlation with equilibrium pressure that determine the tendency to form metal hydrides and their stability, whereas Lu *et al.* [19] studied the structure-property relationship of hydrides using ML and reported that the lattice constant, valency-electron concentration (VEC), and Z/r^3 (where Z and r are the valence electron number and atom radius) play key roles in the H-storage capacity of quaternary Ti-Cr-Fe-V alloy.

Molecular simulations are widely used to study H-storage properties to provide insights into H-adsorption capacities and binding energies, kinetics of H diffusion/adsorption, reaction pathways for H adsorption/dissociation, formation in H-evolution reactions [20–24], and on other material properties. Additionally, density functional theory (DFT) informed interpretable ML using Bayesian optimization shows that body-centered cubic (BCC) multi-principal elemental alloys (MPEA) theoretically exhibit high hydrogen

storage (2.83 wt%) and associated mechanical strength and electronic configurations [25]. Furthermore, BCC alloys exhibit a $H/M=2+$ by tuning alloy composition, and H-storage properties can be modified [26]. Suwarno, *et al.* [16] applied ML models to show the effects of alloy elements on H-storage capacity and solution energy, giving insight into alloy selection of elements for experiments. Other than metallic alloys, ML-based studies have been efficient in characterizing H-storage capacities of non-metallic materials, including metal-organic frameworks (MOFs) [27,28], $LiBH_4$ -based [29], and porous carbon-based materials [30]. Results show that several functional groups for MOFs and other thermodynamic parameters like the temperature, pressure and surface properties (like pore volume and surface pore fractions) play a major role in hydrogen uptake [11]. Overall, these studies reflect the significant role that ML can play by harnessing the combined power of computational methods of MD, DFT, and data-driven ML algorithms, to accelerate the discovery and optimization of metal hydrides with enhanced H-storage.

Here, we present a ML-guided approach that integrates fundamental ML principles, such as feature selection, model training, and validation, with atomistic simulations for the predictive design of solid-state H-storage materials. We utilize the US Department of Energy (DOE) H-storage dataset, which includes features like H wt.%, solution energy, entropy of formation, pressure, temperature, and other thermodynamic parameters across various chemical compositions. A key addition to our model is the inclusion of local lattice distortion (LLD), resulting from atomic-size mismatch between elements, shown to correlate with H-storage performance in MPEAs [25]. As a distinct difference between our model and others, we emphasize that H/M is more informative than wt.%, as it is a more effective descriptor for alloy structure and H-storage. We show that AI/ML framework can facilitate high-throughput search of H-storage materials over multi-dimensional alloy design space, exemplified for ternary metal hydrides with BCC and FCC crystal-structure, i.e., Ti-Nb-X ($X=Mo, Cr, Hf, Ta, V, Zr$) and Co-Ni-X ($X=Al, Mg, V$).

We also performed DFT and MD simulations to reveal the role of hydrogen-metal interaction on electronic-structure, local-lattice distortion (LLD), and H diffusion in affecting H-storage capacity on selected Ti-Nb-Mo and Co-Ni-X alloys with best H/M and H solubility. The correlation between H-storage capacity and solution energy with compositional wt.% has been shown, primarily focusing on refractory elements like Ti, Nb, and Mo. Overall, our study underscores the significant role that combining DFT, MD, and data-driven AI/ML algorithms can play in accelerating discovery and optimization of metal-hydrides with enhanced H-storage properties.

2. Methods

2.1 Data selection, preparation and feature engineering: In **Supplementary Figure S1 &S2**, we have shown the hydrogen storage (HyStr) experimental database consisting of H_2 wt% and hydrogen solution

energy of nearly 1700 metal alloys as a function of absorption temperature as well as the Pearson correlation matrix for all the features. The experimentally recoded database contains diverse classes of metal alloys including A_2B , AB , AB_2 , AB_5 , MIC, SS, Mg, Low/Medium/High-entropy Alloys. Like Pearson correlation matrix, the Spearman correlation coefficient (**Fig. S1b**) was used to assess the strength and direction of a monotonic relationship, making it more robust to non-linear associations and outliers, regardless of the linearity of the relationship [32]. Notably, the Pearson correlation in **Supplementary Figure S2** between H/M to solution energy is 0.3 and H/M to H_2 wt% is 0.7. **Supplementary Figure S2** shows that covalence (R_{cov}), average atomic radius (R_{avg}), average atomic volume (V_{avg}), and atomic packing fractions (APF) show strong positive correlations with both the H/M and solution energy of metal hydrides, indicating the importance of these atomic features to improve the accuracy of the ML model.

2.2 Feature engineering: Initially, for the complete dataset selected for the study, only the thermodynamic features like pressure, temperature, solution energy are there alongside each composition. Because of this scarcity of features, we have added different statistical-based atomic and electronic features like, Pauling electronegativity ($\chi_{pauling}$), Mulliken electronegativity ($\chi_{mulliken}$), covalent, atomic radii, atomic volume, weight, density, LLD, etc. and we approached our model development which could lead to attaining relatively higher accuracy without bias and overdependence on specific features. Mostly, we added the mean, variance, minimum and maximum values of each of those atomic, electronic and material features and trained the model using only these feature sets. The mean and variance were both weighted according to the proportion of the element within each alloy structures. Mathematically, the mean (μ), variance (σ^2), minimum (min), and maximum (max) values of a feature $X = \{x_1, x_2, \dots, x_n\}$ are calculated as, $\mu = \frac{1}{n} \sum_{i=1}^n x_i$; $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$; $\min(X) = \min\{x_1, x_2, x_3, \dots, x_n\}$; and $\max(X) = \max\{x_1, x_2, x_3, \dots, x_n\}$. Additionally, atomic packing fraction (APF), a measure of how densely the atoms is closely packed in each structure, is added as a feature which is calculated by the fraction of volume in a crystal structure that is occupied by the constituent atoms. APF can be a critical feature in hydrogen storage for alloys, as it is strongly correlated with the availability of interstitial sites where hydrogen atoms can reside within lattice structures. The APF is calculated by, $APF = \frac{\frac{4}{3}\pi\rho N_a \sum (r_i^3 c_i)}{\sum M_i c_i}$, where ρ is the density of the material, r_i , c_i , and M_i are the atomic radii, compositional fraction, and molar mass of i -th element respectively, and N_a is Avogadro's number. By addition of all these atomic, electronic, and structural features; calculated from elemental properties using the alloy compositions of the dataset; we finally reached a total of 38 features for each composition. This statistical approach helps in reducing bias and prevents over-reliance on all

thermodynamic features, thus improving the model's accuracy without overfitting, which is the primary objective of this work.

All compositions that could not be directly parsed as compositions, such as “Mg₂Ni_{1-y}Be_y” were removed. In addition, any values for temperature, pressure and formation enthalpy that were outliers by more than 5 standard deviations were removed, as well as values that were non-numerical or incorrectly formatted. After pre-processing, our sample had 1775 compositions and of these, only 1100 compositions had temperature and/or pressure values, and only 558 had values for formation enthalpy/solution energy. Furthermore, we used the *k* nearest neighbor (KNN) type unsupervised-learning technique that has been effective in predicting missing values of a dataset to make predictions. KNN was applied using *k*=5 to impute missing values based on the mean of the 5 nearest neighbors in the feature space. For each missing values, the model trains the KNN algorithm using datapoints where that feature is not missing, based on the remaining feature sets. It then predicts the missing values using 5 nearest neighbors in the feature space. This approach ensures that imputation is performed in an unsupervised manner without using any target labels. KNN-type supervised learning method computes the average or weighted average of their target values that is used as the predicted value for the new data points [33,34]. The dataset is divided into training, validation, and test sets, comprising 70%, 15%, and 15% of the total data points, respectively, to facilitate further analysis.

Evaluating the performance of machine learning models is a critical step in assessing their effectiveness. Here we have considered 4 parameters- R-squared (R²) score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), for evaluating our models' performance [35, 36] (see **Supplementary Information**). Although it is almost impossible to predict the performance of a machine-learning regression task just by one single evaluation parameter, studies show that the R² score (i.e., the coefficient of determination) becomes the most informative and truthful than any other performance matrices like MAE, MAPE or RMSE [36]. Additionally, the atomic size mismatch adopted to measure the LLD effect here is calculated as [35],

$$\delta = 100 \sqrt{\sum_{i=1}^n c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2} \quad (1)$$

Where c_i and r_i are the atomic percentage and radius of individual constituent elements of the alloy respectively.

2.3 Training of ML Models: Due to limited data points, we have applied several classical machine learning algorithms, mainly linear, decision tree, and ensemble-based models, and performed necessary model

tunings as per hyperparameter optimization to better the accuracy. First, we started by applying the classical ML models like Linear regression, Lasso, and Bayesian Ridge regressions and later upgraded to some ensemble and tree-based models like Gradient Boosting Regressor (GBR), Extreme Gradient Boosting Regressor (XGB), and Random Forests (RF).

The performance of machine learning models greatly depends on the hyperparameters before training the models and the optimization of these parameters becomes important as it directly impacts the generalization, efficiency, and stability of the model [37,38,39]. For this purpose, we employed the "GridSearchCV" from Scikit-learn [40] which facilitates an exhaustive search through a predefined randomized hyperparameter grid across a specified range. The best-performing hyperparameters for each of the best 3 predictive models are identified from the optimization and the selection criteria were based on the model's performance metrics, such as accuracy, precision, recall, and mean squared error averaged over 5-fold cross-validation (see Supplementary **Table S1**). By fine-tuning these parameters, we were able to enhance the generalization capability of the models, reduce training times, and mitigate the risks of overfitting, thereby achieving a more robust performance.

Uncertainty analysis: Uncertainty analysis in the Random Forest (RF) model is rooted in its ensemble structure, where predictions are generated by aggregating outputs from N decision trees. RF model is an ensemble of N decision trees, and for a given feature vector x , each tree T_i produces an estimate $y_i(x)$. The Random Forest prediction is taken to be the mean of these individual tree outputs, i.e., $\hat{y}(x) = \frac{1}{N} \sum_i^N y_i(x)$. To capture the RF model's uncertainty, we computed the variance of tree predictions: $\widehat{Var}(\hat{y}(x)) = \frac{1}{1-N} \sum_i^N [y_i(x) - \hat{y}(x)]^2$, while standard deviation of uncertainty in RF model was computed using formula $\sigma(\hat{y}(x)) = \sqrt{\widehat{Var}(\hat{y}(x))}$. This standard deviation is visualized as the vertical error bars in the prediction plots, allowing a direct assessment of confidence in the model outputs across training and test datasets.

2.4 MD simulation of H diffusion: The interatomic potentials for the molecular simulations in this study can be represented as, $U = U_{M-M} + U_{H-H} + U_{M-H}$, where U_{M-M} denotes the metal-metal interaction, and U_{H-H} and U_{M-H} represent the hydrogen-hydrogen and metal atom-hydrogen interactions respectively. Lennard Jones (LJ) 12-6 potential, presented in equation (2) [41,42] was applied with a cut-off distance of 10 Å to model the van der Waals interactions between each of these metal elements with gaseous hydrogen atoms (U_{M-H} and U_{H-H}).

$$U_{LJ}(r_{ij}) = 4 \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], r_c < r_{ij} \quad (2)$$

Where, σ_{ij} , ϵ_{ij} were the distance where potential energy becomes zero and the potential wall depth respectively, and r_{ij} is the distance between one LJ site to another. Lorentz-Berthelot mixing rule was applied to model the interaction parameters (σ and ϵ) for each of the U_{H-H} and U_{M-H} cross-interactions [41]. Inter and intra-atomic LJ interaction parameters for TiNbMo combinations with atomic hydrogen were taken from reported literature [43,44]. While the Lennard-Jones (LJ) 12-6 potential lacks the ability to capture the many-body effects and electronic characteristic of metal-hydrogen bonding, it is selected in this study as a computationally efficient qualitative approximation to model hydrogen-metal and hydrogen-hydrogen interactions, under conditions dominated by weak physisorption and van der Waals forces rather than strong chemisorption. Additionally, LJ potential has performed quite accurate to experimental data in modeling diffusion in gaseous phase over a large range of temperature (-50°C to $+90^{\circ}\text{C}$) [45]. This approach enables qualitative hydrogen diffusion trends in alloys, giving a preliminary screening before employing more accurate, but computationally intensive, many-body potentials such as embedded atoms (EAM) or modified embedded atoms (MEAM).

Each simulation supercell initially consisted of 60 atoms depending on the different elemental compositions selected, and we have added 6 hydrogen atoms at the octahedral positions within the lattice structure to assess the hydrogen diffusion inside bulk metallic structures. First, the energy minimization of each structure was performed using the conjugate gradient method [44], followed by isothermal-isobaric (NPT) equilibration at 300 K, and constant pressure heating to 1000 K at 4.7 K/ps and a subsequent NPT equilibration at 1000 K to relax the structures. Finally, the hydrogen diffusion simulations were carried out on the relaxed structures at 1000 K. We chose elevated temperature based on established practices to address the limitations of classical MD simulations in sampling diffusion kinetics applied to the MD simulation timescales. To address the slow timescales of hydrogen diffusion in metals, and metal hydrides, MD simulations are run at elevated temperatures to accelerate atomic movements and enable meaningful sampling of diffusion mechanisms within limited simulation and computation times [46-47]. Additionally, our purpose of applying high-temperature simulations is not intended to accurately predict the absolute diffusion coefficients at different operational temperatures, but to perform comparative analysis qualitatively across different alloy systems, where the relative trends in diffusion is retained. The average mean squared displacements (MSD) and subsequent radial distribution functions (RDF) were generated to qualitatively analyze hydrogen diffusion in different binary and ternary TiNbMo-based alloy systems. All the simulations were carried out using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software [48] and a 1 femtosecond (fs) timestep size was used with velocity verlet algorithm

[49,50] for performing the time integration of the equation of motions and the system temperature was controlled using Nosé-hoover thermostat [44,51,52].

2.5 DFT methods: We employed first-principles DFT as implemented in the Vienna Ab-initio Simulation Package (VASP) [53,54] for geometrical optimization (e.g., lattice constants, volume, bond length, and bond angles) and charge self-consistency. The generalized gradient approximation of Perdew, Burke, and Ernzerhof (PBE) was employed in all calculations [55] with a plane-wave cut-off energy of 520 eV. The choice of PBE over LDA or meta-GGA [56,57] functionals is based on the work of Söderling et al. [58] and Giese et al. [59] that establishes the effectiveness of GGA functionals. Large Supercell Random Approximates (SCRAPs), i.e., 60 atoms per cell, with the optimized disorder (zero-correlation) were generated (a single, optimized configurational representation) for DFT calculations. The energy and force convergence criterion of 10^{-8} eV and 10^{-6} eV/Å, respectively, were used for full (volume and atomic) relaxation of SCRAPs. The Monkhorst-Pack k -mesh was used for Brillouin zone integration during structural optimization and charge self-consistency calculations [60].

3. Results and discussions

By incorporating H/M, solution energy, and H₂ weight percent as predictive metrics, we discuss below that how ML helps filter out less promising materials early, reducing the need for extensive and computationally expensive DFT validation. This targeted approach was not only used to save time and resources but also increases the likelihood of discovering MPEAs with optimized hydrogen storage performance, driving faster progress in the field.

3.1 H/M, Solution Energy and H₂ wt.% predictive models

As described before, H-storage capacity in metals and alloys can be defined generally in 2 distinctive ways though they are closely related, i.e., (i) H₂ wt.% and (ii) H/M ratio. Typically, the gravimetric H-storage capacity is calculated from H wt.% which is the ratio of the mass of hydrogen absorbed within the compound to the mass of the compound including the stored hydrogen [61], which is given as,

$$H_2 \text{ wt}\% = \left(\frac{(H/M)M_H}{M_C + (H/M)M_H} \times 100 \right) \%$$

where M_H and M_C are the molar masses of hydrogen and the material/compound storing hydrogen, and **H/M** refers to the ratio of the number of hydrogen atoms to the total number of metal atoms in the unit cell, representing hydrogen occupancy per formula unit.

Here, we have developed a data-driven model to predict the H/M ratio. **Table 1** shows that RF performs the best in predicting H/M with a test set R^2 score of 74% with only the atomic, electronic and crystal properties of the alloy compositions. Additionally, the model was subjected to 5-fold cross-

validation that predicts the robustness and generalizability, resulting in a mean R^2 score of 58%. Despite the relatively high R^2 score, the MAPE was >40% suggesting a substantial deviation of some instances from the model's prediction. This discrepancy highlights that while the model captures the overall trend in the data, it may not be accurate in predicting all individual instances. The GBR model demonstrated strong predictive capabilities for H_2 wt%, achieving a test set R^2 of 0.85 and a cross-validation R^2 score of 82% which is substantially higher than the H/M model showcasing the features used for predicting H_2 wt% were able to capture the underlying patterns and able to effectively generalize the target variable whereas H/M prediction becomes inherently more complex or noisy. Supplementary Information **Table S2(A-B)** shows the performance of other machine learning models for H/M, solution energy prediction.

Table 1. Different evaluation parameter values (R^2 , Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE)) for the H/M, H_2 wt%, and solution energy models.

Models	Best performing ML model	Training data					Test data				
		R^2	MAE	RMSE	MAPE	CV	R^2	MAE	RMSE	MAPE	CV
H/M	RandomForest	0.89	0.11	0.16	22.51	0.70	0.74	0.17	0.24	41.60	0.58
Solution Energy	Extreme GradientBoost	0.99	1.05	1.62	3.75	0.66	0.74	9.82	17.35	22.18	0.57
H_2 wt. %	GradientBoost	0.98	0.06	0.18	7.26	0.77	0.85	0.27	0.45	33.6	0.82

Figure 1a-b shows the training and test data parity plot for the H/M model and **Figure 1c** shows the top 10 important features dominating the H/M prediction using the RF model. Average radius, covalent radius, and average electronegativity are the top 3 critical parameters, and interestingly, atomic packing fraction (APF), another important lattice parameter denoting the volumetric occupancy of atoms in the unit cell becomes a physically significant feature because hydrogen's susceptibility to adsorb in the interstitial gaps of lattice. The predictive model for H_2 wt% also shows similar important feature lists, however, average molecular weight (Z_{avg}) becomes the most important feature based on the GBR model (**Figure 1i**).

Additionally, the H_2 wt% model fits better (>85% R^2 on test data) than the H/M model (~74% R^2 on test data), indicating the strong effect of Z_{avg} and the model is able to capture the complexity and variance of H_2 wt% with the selected feature set in the dataset. As mentioned before, among the 1775 data points only 558 compositions of all hydride types (AB, AB₂, AB₅, MIC, and SS) have solution energy, and therefore for the solution energy (ΔH) predictive model only these 558 data points are used and **Figure 1d-f** shows the training, test data parity plots, and feature importance of the ΔH model. The relatively small number

of data points for solution energy (~558) limit the model's ability to make accurate and reliable predictions for alloy compositions (i.e. generalizability) on a wider scale of different types of hydrides (AB, AB₂, AB₅, MIC, or SS), particularly when predicting ΔH for unknown compositions. However, due to the lack of experimental solution energy data of other 1217 compositions, we were restricted to the available 558 entries. This could be improved by augmenting the dataset with high-throughput DFT calculations or curated experimental data for a wider range of compositions. It is found that other than R_{cov} , ρ_{min} , temperature, and LLD are among the top 5 feature sets. Our ML model for solution energy attains an R^2 score of 74% and a 5-fold cross-validation R^2 of 57% over the unseen dataset based on the XGB model which shows the best fitting and accuracy. As ΔH shows strong linear dependencies with equilibrium pressure and temperature (T), given by Van't Hoff's relation [61–63] shown in equation (3), therefore the temperature-dependency is relevant.

$$\ln\left(\frac{P_{eq}^0}{P}\right) = -(\Delta H)\left(\frac{1}{RT}\right) + \frac{\Delta S}{R} \quad (3)$$

This comes from replacing the expression of ΔG in $-RT \ln K = \Delta G$ by $\Delta G = \Delta H - T\Delta S$, where $K = \left(\frac{P_{eq}^0}{P}\right)$, is the equilibrium constant from the Pressure-Composition-Isotherm (PCI) of metal alloys for hydrogen storage, ΔG , ΔH , and ΔS are Gibbs free energy upon hydrogenation, change in standard enthalpy or formation enthalpy during hydrogenation and change in net entropy respectively, P_{eq}^0 is the equilibrium pressure at 25°C, R is the universal gas constant.

The uncertainty analysis of RF model in predicting H/M (**Fig. 1a, b**), XGB model for Solution energy (**Fig. 1d, e**), and GBR model for H₂ wt%. (**Fig. 1g, h**) reveals how well the model performs across training and test datasets, focusing on key regression metrics and the spread of prediction uncertainty (error bars). For GBR and XGB methods, uncertainties are estimated by tracking the model's predictions at each boosting stage and the standard deviation of these stage-wise predictions is calculated for each sample, representing the model's predictive uncertainty, whereas for the RF method, uncertainties are estimated by collecting predictions from each individual decision tree in the ensemble. The standard deviation of these tree-level predictions for each sample is computed to quantify the predictive uncertainty. A higher standard deviation indicates greater disagreement among the trees and thus lower confidence in the prediction. Each data point in **Fig. 1** (excluding feature importance plots) is accompanied by an error bar representing the uncertainty of predictions across individual predictions based on the best-fitted model. The HtoM model shows good predictive performance on the training data with an R^2 of 0.89, MAE of 0.11, and RMSE of 0.16, accompanied by relatively higher prediction error bars indicating high uncertainty. On the test dataset, the R^2 drops to 0.74, and both MAE (0.17) and RMSE (0.24) increase, with noticeably

larger error bars, especially at higher HtoM values. The sharp increase in uncertainty and error magnitude indicates strong overfitting and poor generalization to unseen compositions. For solution energy, the model performs extremely well on the training set ($R^2 = 0.99$), but test performance degrades significantly ($R^2 = 0.74$). In contrast, the hydrogen weight percent (H_2 wt%) model achieves high accuracy on the training data ($R^2 = 0.98$) with minimal uncertainty and maintains robust performance on the test set ($R^2 = 0.85$). Although prediction uncertainty slightly increases for higher H_2 values, it remains well controlled overall. Among the three targets, the H_2 wt% model demonstrates the best balance between accuracy and generalization, while the HtoM model exhibits moderate predictive uncertainty on new data, compared to the H_2 wt% model.

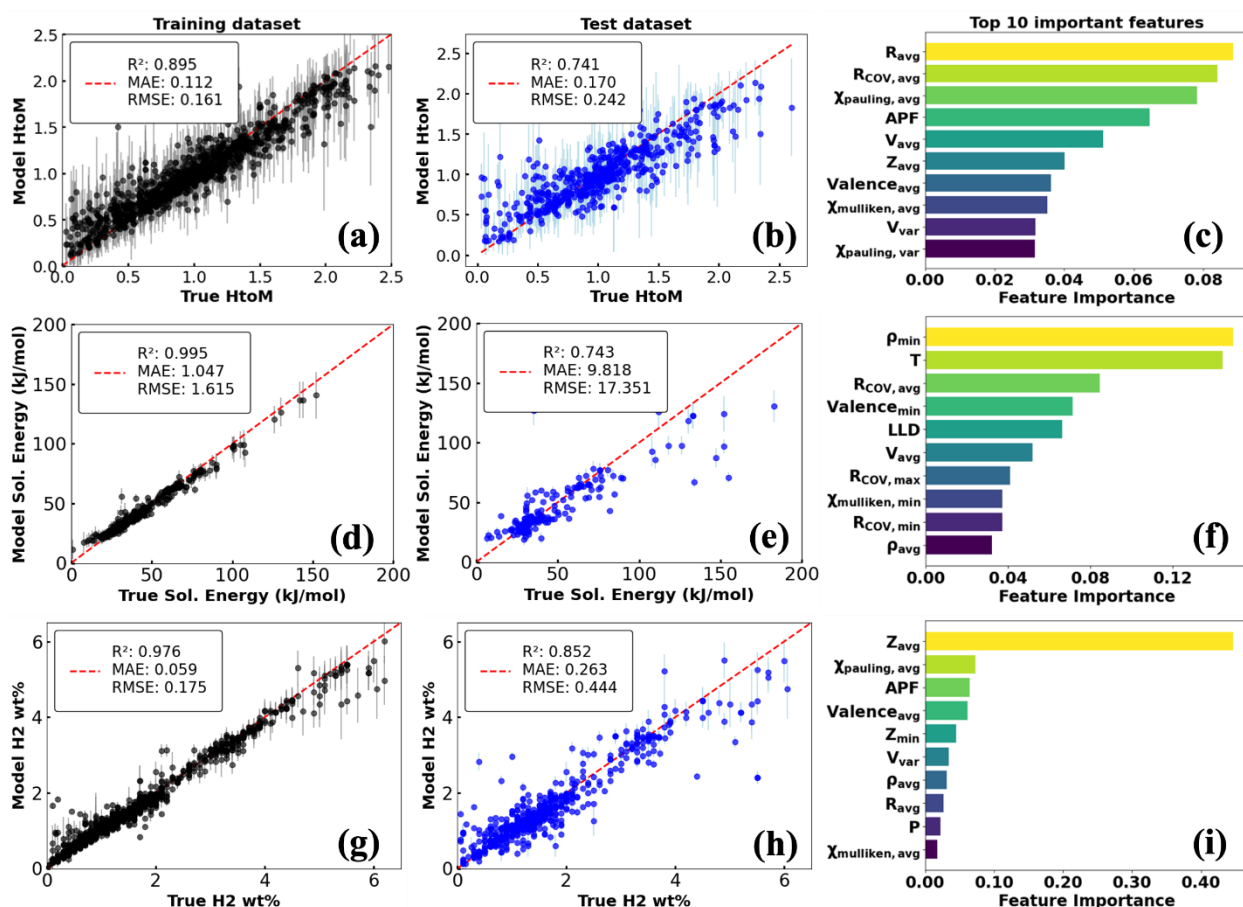


Figure 1. Parity plots and feature importance results for the 3 models: (a-c)-H/M, (d-f)-Solution energy, (g-i)- H_2 wt%, where (a, d, g) training (b, e, h) test data parity plots, (c, f, i) top 10 important features for the best performing models in each of those three cases. The parity plots show the uncertainty in each prediction points in the form of error bars.

Notably, the HtoM model exhibits good predictive reliability with moderate uncertainty on test data, however, the average uncertainty remains within 15-25% of the actual values. The solution energy model, although accurate on the training set, shows poor generalization, reflected by high prediction

uncertainty (~10-15%) in test data. The hydrogen weight percent model offers the best overall balance, maintaining accuracy and relatively low average uncertainty across both training (8%) and test sets (12%). This reflects a moderate increase in prediction uncertainty and suggests fair generalization with limited overfitting. As we indicated before, the ML predictive model on H₂ wt% performs better in prediction compared to the HtoM, as also indicated by relatively smaller average uncertainty values in both the training and test datasets.

3.2 Feature importance on H/M and solution energy predictions

Our predictions from the ML model are closely related to existing literature, which shows that hydrogen adsorption and desorption increase significantly with the addition of Ti in MPEAs [64,65]. It has also been reported that Ti and Zr exhibit some of the highest hydrogen uptake capacities among metals [66,67]. Our ML results, including models like RF for H/M, XGB for solution energy, and GBR for H₂ wt%, indicate that covalent radius (RCOV), average electronegativity (χ_{Pauling}), and average atomic radius (R_{avg}) are the most critical features, as shown in **Figure 1c-i**. These findings are consistent with literature, where covalent radius, LLD, and electronegativity are known to be key determinants in hydrogen adsorption capacity in MPEAs due to atomic and electronic structure considerations [68,69].

Interestingly, we found that LLD appears as one of the top 5 important features in the solution energy predictive model (**Figure 1f**), but adding LLD did not significantly affect the H/M and H₂ wt% predictions. This may suggest that while lattice distortion (LLD) plays an important role in the solution energy model, it does not directly influence the hydrogen uptake metrics in our ML predictions, as expected based on previous findings that high lattice strains from distortion are favorable for hydrogen adsorption in high-entropy alloys and alloys with low VEC [18,70]. On the other hand, when looking at hydrogen weight percentage prediction in **Figure 1i**, the top 5 features in the GBR model—atomic weight (Z), atomic-packing factor (APF), χ_{Pauling} , and average valence—contribute significantly (60-70%) to the prediction. Interestingly, atomic weight (Z) was not among the top 5 features in the H/M predictive model. Both covalent radius (R_{COV}) and electronegativity (χ_{Pauling}) consistently emerged as significant contributors across all tree-based models, highlighting their central importance in H/M prediction. We also noted that Vanadium and Titanium are two crucial elements that may improve H/M on alloying [64,68]. In **Supplementary Figure S3**, we have shown the actual vs predicted data points of both H/M and Solution Energy models with the number of data entries in continuous plots using 4 different ML models: GBR, RF, XGB, and Ridge regressions.

3.3 Effect of Different Elemental Compositions on ML developed Predictive Models

The BCC materials have been studied extensively for H-storage application as it is often identified to store $H/M = 1$ (monohydride) in a BCC or body-centered tetragonal (BCT) and $H/M = 2$ (dihydride) in face-centered cubic (FCC) phase [72–75]. However, hydrogen incorporation can either raise or lower the formation enthalpy of the hydride phase, potentially hindering reversible desorption by shifting the thermodynamic window [76–79]. Thus, in this section, we analyzed the H/M behavior of single-phase binary bcc refractory alloys using our GBR-predicted ML model to understand the role of refractory metal elements

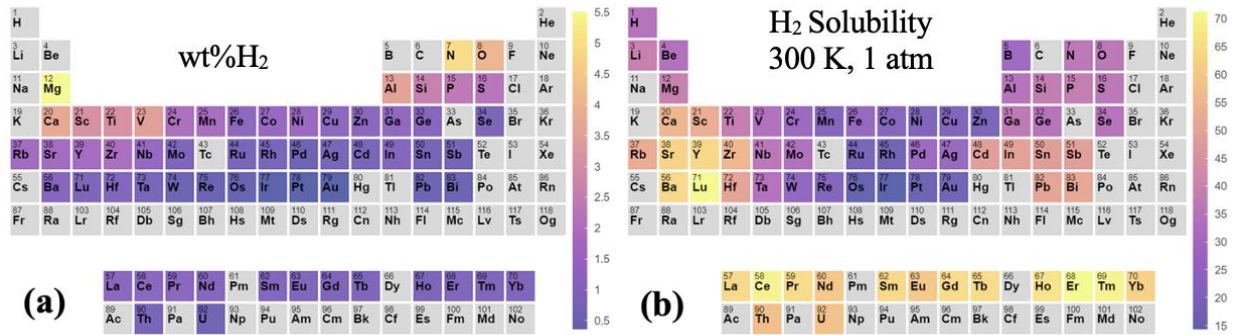


Figure 2. Periodic tables illustrating elemental effects on (a) H_2 wt% and (b) solution energy, derived from ML model predictions, which provides a reference for identifying alloying elements that enhance H solubility and storage.

Figure 2a-b shows the effect of binary elements on their ability to store hydrogen in terms of hydrogen weight percent and solution energy results based on our ML predictive model. Among the transition metal elements, V, Nb, Ti, and Zr show high H-storage ability with binary vanadium showing a 3.0-3.5 wt% H-storage ability, whereas the presence of other transition metals like Mo, Ta, and W shows hydrogen storage of only about 0.5-1.0 wt%. This is consistent with literature showing that V, Ti, Nb, and Zr have high affinity towards hydrogen and form stable solid solutions of hydrides but Mo, Cr exhibit low hydrogen affinity, leading to lower hydrogen storage [26,80]. Additionally, experimental results show that magnesium-based alloys exhibit 5.0-5.5 wt% hydrogen storage. However, magnesium can store hydrogen of about 6.0-6.4 wt % at ambient pressure and $>350^\circ\text{C}$.

3.4 Ternary Ti-Nb-X (X=Mo, Cr, Hf, Ta, V, Zr) MPEAs

Based on the results shown in **Figure 2**, in this section, we investigated the H-storage ability of transition and refractory alloy compositions having different compositions. We chose different ternary compositions of Ti-Nb alloys with Mo, Cr, V, Zr, Hf, and Ta to denote the effect of different alloying compositions on their H-storage ability, as shown in **Figure 3**.

From the H/M ternary plots, it is clearly seen that the Ti-rich and Ti-V-rich regions of all the compositions show a hydrogen-adsorption capacity of more than 2.4-3.5 wt%. In most of the

compositions, higher hydrogen weight percent regions also exhibit higher H/M ratio, however, for TiNbHf and TiNbZr 80-100% Ti-rich regions predict the highest H₂ storage of 2.5-3.0 wt% whereas the Hf- and Zr-rich regions show the highest H/M ratio based on our ML model's prediction. Heavier alloying elements such as Hf, Ta, and Zr, despite their high affinity for hydrogen (higher H/M ratios in Hf- Ta-, and Zr-rich regions, shown in **Figure 3g, j, p**), inherently reduce the overall hydrogen weight percent (only about 1-2 wt%) due to their large atomic masses (>90 g/mol), whereas high H/M regions in lighter NbTi-X alloys (NbTi-Mo, NbTi-V, and NbTi-Cr) also predicts high H₂ wt%. The model predictions also reveal that lighter, Ti-rich regions consistently exhibit both high H/M and high wt% H₂, emphasizing the favorable balance of light atomic weight and strong hydrogen affinity in such compositions. It agrees well with our earlier feature-interaction results showing average molecular weight is the most dominant feature (>40% effect) in predicting H₂ weight percent (**Figure 1i**), and atomic radius, electronegativity modulate shapes the H/M predictions (**Figure 1c**). This result also reflects the fact that the H₂ weight percent predictive model gives a more relevant understanding of overall H-storage capacity than the estimation of H/M. Additionally, these ternary plots allow for a comprehensive and systematic exploration of the compositional space. By visualizing the effects of varying concentrations of different transition elements, the predictions of H-storage capacity and stability can be further analyzed. This approach minimizes the need for exhaustive experimental testing by focusing efforts on the most promising compositions.

The model predictions in **Figure 3** show that Nb-Ti-based MPEAs are expected to have higher H/M and, therefore, a better fit for hydrogen storage applications. This finding corroborates with earlier reports [81] where early transition metals show improved stability for hydride formation while it rapidly decreases moving across the period for mid-to-late transition metals. Notably, Ti-based alloys also present several benefits including fast H-storage, long cycle life, better release rate, and low price, thus, it is considered one of the best materials choices for hydrogen storage. The choice of the third element, i.e., molybdenum, was based on prior knowledge that Mo-substitution can extend the BCC stability range. The superior ability of Nb-Ti alloys doped with Mo in stabilizing the BCC structure comes from reduced unit-cell volume or lattice parameter in the BCC phase (due to smaller atomic radii of Mo-atom compared to Ti-atom) that inhibits the transformation of Ti into HCP phase. The coarse packing and high density of interstitial sites (tetrahedral and octahedral) in the BCC structure also make these alloys attractive for superior hydrogen absorption.

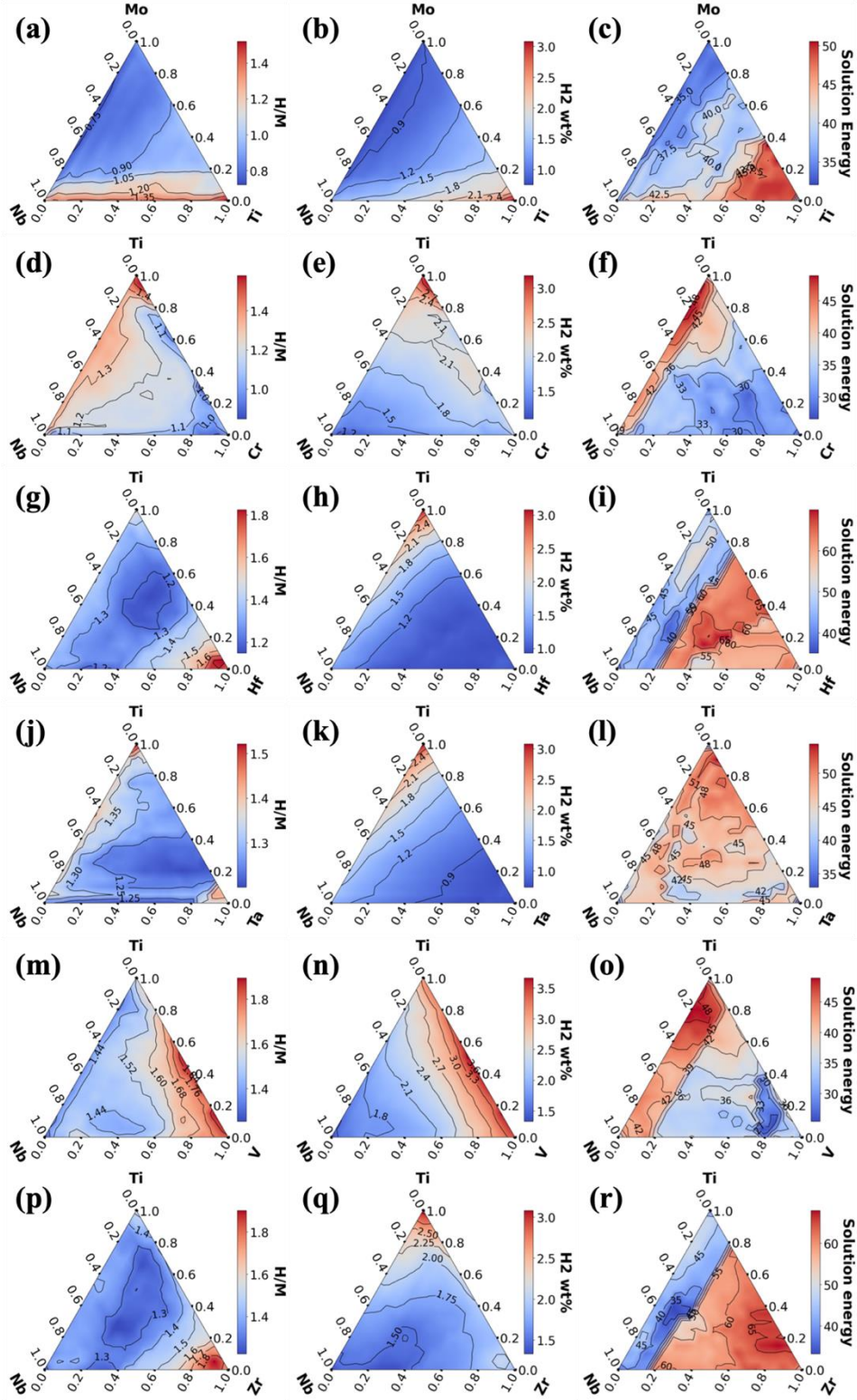


Figure 3. Ternary plots of different transitional metal compositions showing the distribution of H/M, hydrogen weight percent and solution energy. (a-c) TiNbMo, (d-f) TiNbCr, (g-i) TiNbHf, (j-l) TiNbTa, (m-o) TiNbV, (p-r) TiNbZr. (a,d,g,j,m,p) are for the H/M, (b,e,h,k,n,q) are for hydrogen weight percent and (c,f,i,l,o, and r) are for the solution energy (kJ/mole) based on our ML models predictions.

The ternary plots indicate that the combination of Ti, Nb, and Mo yields a favorable H/M ratio with a stable Nb-Ti rich solution energy region. The solution energy plots showed that the Ti-Nb-Mo combination has a relatively low solution energy, indicating stable hydride formation. Stability is essential for the reversibility of hydrogen absorption and desorption, making this composition a reliable candidate for hydrogen storage. Additionally, the combination of Mo, Nb, and Ti offers a good balance between thermodynamics (solution energy) and kinetics (hydrogen absorption) which is critical for H storage.

3.5 Phase stability and electronic-structure analysis of Ti-Nb-Mo Alloys:

To further understand the phase stability and hydrogen storage behavior, in **Fig. 4a**, we show the ternary diagram of formation enthalpy for the bcc Ti-Nb-Mo alloy. The phase stability diagram reveals two distinct thermodynamic regions: (1) the high-entropy region, located at the center of the ternary diagram, which shows higher stability, and (2) the vortex regions or edges, corresponding to binary or near-unary regions, which exhibit weaker stability.

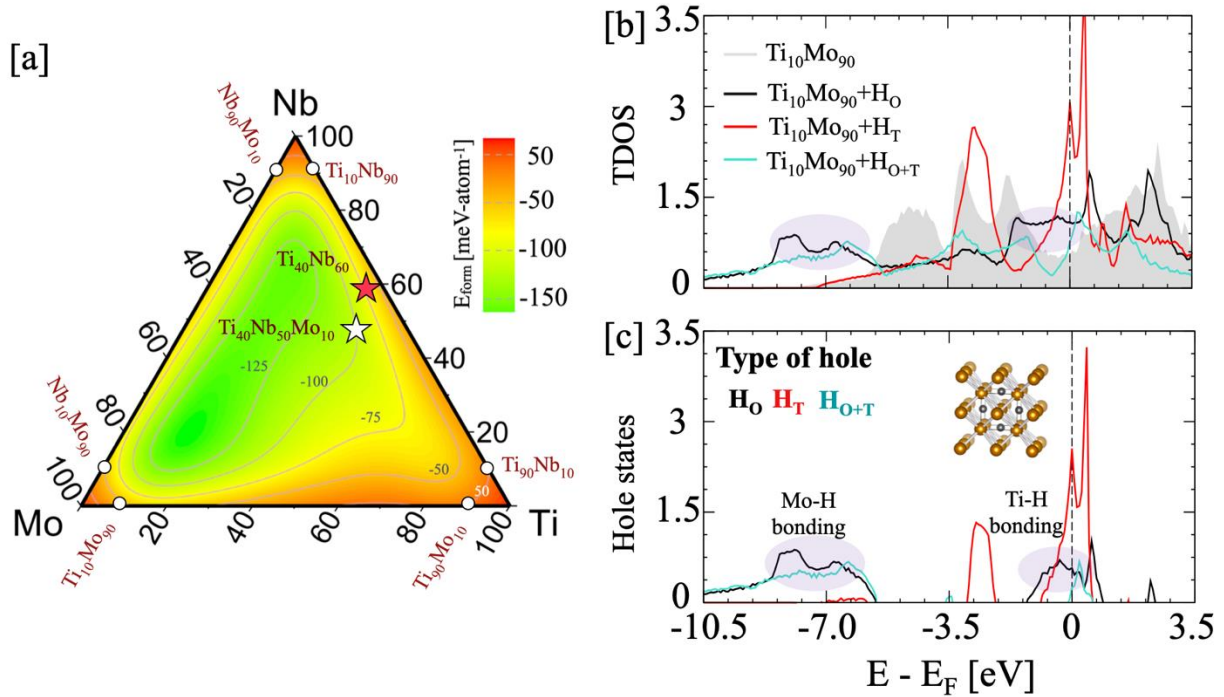


Figure 4. (a) Thermodynamic stability (formation enthalpy) analysis of ternary BCC Ti-Nb-Mo MPEAs. (b) The electronic-structure analysis of Mo-rich $\text{Mo}_{90}\text{Ti}_{10}$ shows strong Mo-*d* and H-*s* bonding at very low energy below Fermi energy, and (c) total hole states where H-*s* goes are near very deep energy states, i.e., -7.0 eV, not very favorable for hydrogen storage or absorption/desorption cycle. See related discussion on relaxation effect on DOS of Mo-Ti alloy in **See Fig. 6 and Fig. 7 and A3**. While the experimental validation of detrimental role of Mo on hydrogen absorption is discussion in **section 3.5.3**.

In the near-equiatomic region, the formation enthalpy of the bcc phase sharply increases. Interestingly, the Mo-Nb-rich region is thermodynamically more stable, while the Ti-rich region shows moderate to low phase stability. From a hydrogen absorption perspective, the high stability of hydride formation in the Mo-Nb region makes the desorption process slower and requires more energy to break the metal-hydrogen bonds. This is consistent with the observation that very high VEC metal atoms are less favorable, as the contraction of metal-*d* bands across the period reduces the extent and overlap of the *d*-states. The phase stability data from DFT further confirms the computational insights gained from the ML models, offering a detailed understanding of the thermodynamic and structural factors influencing hydrogen storage performance in these alloys.

The ternary diagram in **Fig. 4a** shows DFT-calculated phase stability (formation enthalpy) for bcc Ti-Nb-Mo alloy. The phase stability diagram can be divided into thermodynamic regions: (1) high-entropy region, i.e., the central part of the ternary diagram, with higher stability, and (2) vortex regions or edges, i.e., binary or near unary regions, with weaker stability. The formation enthalpy of the bcc phase increases sharply in the near equiatomic region. Interestingly, Mo-Nb rich region is thermodynamically more stable while the Ti-rich region shows moderate to low phase stability, which is interestingly from a hydrogen absorption point of view because the very high high-stability of hydride formation makes the desorption process slow as it will require higher energy to break metal-hydrogen bonds. The reason for not choosing very high VEC metal atoms lies in the contraction of metal-*d* bands going across the period due to increased nuclear charge that reduces the extent and overlap of the *d*-states [78].

In **Fig. 4b**, we show the electronic structure (electronic total density-of-states, TDOS) of binary bcc $\text{Ti}_{10}\text{Mo}_{90}$ and its hydride where hydrogen is added at octahedral (O; 3 sites), tetrahedral (T; 6 sites), and mixed octahedral-tetrahedral (O+T; 3+6 sites) sites as shown in the inset. The grey-shaded region is the total DOS of $\text{Mo}_{90}\text{Ti}_{10}$ without hydrogen addition while black, red, and cyan represent TDOS with hydrogen at three different interstitials. The H-*s* states for octahedral and tetrahedral hydrogen occupation are mainly concentrated in low-energy {-6 to -10 eV} and mid-energy {-3 eV} to high at/around Fermi energy, respectively, where the Fermi energy is defined at zero on the x-axis. The **Fig. 4c** shows the hole-states added due to hydrogen addition to alloy, which was calculated by subtracting TDOS of pristine $\text{Ti}_{10}\text{Mo}_{90}$ with TDOS of hydrogenated alloys or by employing the analytical relationship, i.e., $D_h(\varepsilon) = \frac{dN}{d\varepsilon} = \frac{V}{2\pi^2} \left[\frac{2m_h}{(\hbar/2\pi)^2} \right]^{3/2} [\varepsilon + E_F]^{1/2}$, where $N(k) = \frac{k^3 V}{3\pi^2}$, and $k^2 = 2m_h(\varepsilon + E_F)/(\hbar/2\pi)^2$ where *N* is the number of electronic states, *V* is the unit-cell volume, *m_h* is hole mass, and *E_F* is the Fermi-energy.

The electronic TDOS of $\text{Ti}_{10}\text{Mo}_{90}+\text{H}$ was scaled by the difference of percent lattice change for one-to-one comparison. We observed in **Fig. 4b,c** that the low-energy bands in $\text{Ti}_{10}\text{Mo}_{90}+\text{H}$ show pronounced hydrogen character. Additionally, metal character arising from formation of bonding and antibonding states between the metal e_g and H-s states was also observed. The observation of higher density of H-s states from tetrahedral or mixed (O+T) hydrogen near/at the Fermi energy (**Fig. 4b**) suggests that these states will require lower energy to break metal-hydrogen bonds for hydrogen desorption. Despite higher density of hydrogen hole states in the tetrahedral case over octahedral case, a peak in the TDOS at the Fermi energy shows a key drawback, i.e., energy instability, which has been reported in other refractory metal alloys [81]. Interestingly, the H addition at the mixed O+T sites in {Ti, Nb} shows no such instability, suggestive of a higher concentration of hole-states, and perhaps advantageous for improved H absorption.

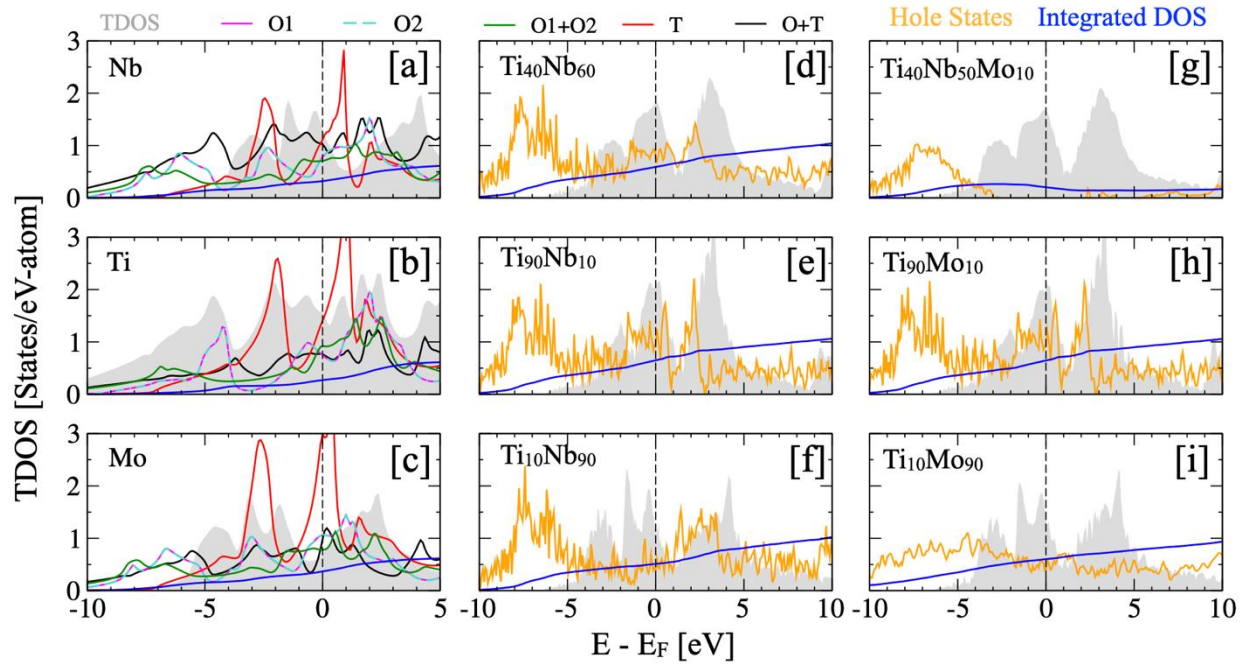


Figure 5. (a-c) Total electronic density of states (TDOS) of simple metal {Nb, Ti, Mo}-hydrides in the body-centered cubic structure. (d-i) TDOS and hole-states (considering octahedral hydrogen) in six ML (GB) predicted Ti-Nb-Mo MEAs. The solid blue line shows the trendline for the integrated density of states.

In **Fig. 5a-i**, we present an analysis of electronic DOS of all alloying compositions predicted from GB-based ML model as well with constituent elements {Nb, Ti, Mo} to understand the electronic-structure origin of improved hydrogen-storage behavior. If we carefully look at the electronic DOS of {Nb, Ti, Mo} metals, in **Fig. 5a-c**, we can clearly see that low VEC refractory metals, such as Ti, show higher spread in electronic states of metal- d bands while high VEC shows stronger localization of d -states, e.g., Nb and Mo. Notably, titanium is known to be a better hydride-former compared to other transition metals, and

remarkably, the electronic DOS of the Ti-H system (**Fig. 5b**) shows that the population of H-s states is much closer to Fermi energy (-3 to -7 eV) compared to Mo-H (-5 to -9 eV) and Nb-H (-4 to -8 eV).

The ML models in **Fig. 1** show a somewhat positive correlation of LLD with hydrogen solubility in hydrogen-storage materials. To extract LLD, we employed a first-principles DFT method to optimize the disorder supercell of chosen compositions (atomic and lattice) using a revised definition of atomic distortion as $\Delta u_i^{rlx-ideal} \sim \sqrt{|u_{rlx}^i(x, y, z) - u_{ideal}^i(x_0, y_0, z_0)|^2}$ or $\sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}$, correct for crystalline, partially-order or disorder solids [82,83]. Here, $u_{rlx}^i(x, y, z)$ and $u_{ideal}^i(x_0, y_0, z_0)$ are positions of i^{th} atom at relaxed and special lattice symmetry points. The atomic displacement of each atom in the crystalline lattice was used as LLD ($\sim \Delta u_i^{rlx-ideal}$), which is tabulated in **Table 2**.

Table 2: DFT-calculated local-lattice distortion, formation-enthalpy (or H solubility ΔE_{form}), formation enthalpy difference, supercell volume, and ML-predicted formation enthalpy for selected set of BCC Ti-Nb-Mo alloys (see related results in **Table 3** and discussion in **Section 3.5.1-3.5.3**).

System	LLD [RLX-Ideal]	E_{form} [meV/atom]		ΔE_{form} [Ideal - RLX]	V [Å ³ /atom]	$\Delta H_{hydride}$ (ML) [kJ/mol]	$\Delta H_{hydride}$ (ML) [meV/atom]
		Ideal	Relaxed				
Ti₉₀Mo₁₀	0.026	-127.4	-191.1	63.8	16.6	48.96	0.51
Ti₁₀Mo₉₀	0.171	-60.8	-62.4	1.7	15.7	33.89	0.35
Nb₉₀Mo₁₀	0.193	+10.7	+10.2	0.6	17.9	35.19	0.36
Nb₁₀Mo₉₀	0.038	+19.9	+16.7	3.1	15.9	33.49	0.35
Ti₉₀Nb₁₀	0.216	-101.9	-132.9	31.0	17.1	48.65	0.50
Ti₁₀Nb₉₀	0.034	+101.8	+99.9	1.8	18.03	41.22	0.43
Ti₄₀Nb₆₀	0.121	-11.9	-31.4	19.5	17.49	42.31	0.44
Ti₄₀Nb₅₀Mo₁₀	0.124	+1.5	-30.6	32.0	17.48	40.95	0.42

The comparison between ideal and relaxed supercell shows that LLD has a strong effect on formation enthalpy ($H_{hydride}$), which is further highlighted in $\Delta H_{hydride}^{ideal-rlxd}$. **Table 2** also shows the ML-predicted solution energy of each of the selected BCC Ti-Nb-Mo alloy sets (in kJ/mol and meV/atom). Our calculations show that a strong local lattice distortion (LLD), quantified via LLD, significantly influences phase stability and hydrogen uptake in Ti-Nb-Mo alloys. Specifically, among the predicted compositions, only Ti₄₀Nb₆₀, and Nb₄₀Ti₅₀Mo₁₀ exhibit both high LLD, i.e., 0.121 and 0.124 respectively. Two alloys also lie within the optimal formation enthalpy window of about -25 meV-per-atom, with calculated values of -31.4 and -30.6 meV/atom, respectively. This range is comparable to thermal energy at room temperature

($k_B \cdot T \sim 25 \text{ meV}$), indicating that these compositions strike a favorable balance between phase stability and hydrogen reversibility under ambient conditions. These compositions also correspond to the top 5% in our ML-predicted hydrogen-to-metal ratios ($H/M > 1.6$), suggesting a direct correlation between enhanced LLD, favorable thermodynamics, and improved storage capacity. Both phase stability and LLD suggest improved hydrogen storage, however, looking back into electronic-structure $\text{Ti}_{40}\text{Nb}_{60}$ (**Fig. 5d**) and $\text{Ti}_{40}\text{Nb}_{50}\text{Mo}_{10}$ (**Fig. 5g**), we found an absence of hole states in $\text{Ti}_{40}\text{Nb}_{50}\text{Mo}_{10}$ compared to $\text{Ti}_{40}\text{Nb}_{60}$, which is indicative of lower hydrogen absorption.

Notably, our analysis suggests that the addition of LLD as a feature gives us a better understanding and ability to predict the hydrogen storage abilities of these chemically complex materials. The stability of MPEAs is also dominated by the atomic and electronic features of their constituent elements in the solid solution including electronegativity, electron affinity, atomic weight, volume, radius, and covalent bond at different temperature and pressure conditions. We have shown in **Fig. 1** the importance of including these features in predicting the hydrogen storage performance of metal hydrides [80]. Finally, we also want to emphasize the significance of LLD in the refractory alloys [84] as the atomic size mismatch, which is systematically used as LLD qualifier, however, it is not accurate enough to predict LLD. In general, a higher LLD in alloys means additional lattice strain, which is seen as a useful feature for quantifying hydrogen adsorption as well as hydride formation [31,70].

3.5.1 Local lattice distortion and structural stability: **Table 3** provides a detailed analysis of LLD, phase stability, and volume change in body-centered cubic (BCC) multi-principal element alloys (MPEAs), particularly focusing on the effects of hydrogen incorporation and oxygen impurities. The formation energy (E_{form}), percentage change upon relaxation, and equilibrium volume per unit cell and per atom (V_0) are examined across different alloy compositions. For pure Mo, the formation energy is relatively low at 25.9 meV/atom, indicating stability, while the atomic volume per atom is $15.78 \text{ \AA}^3$. Upon hydrogen incorporation ($\text{Mo}+\text{H}$), the formation energy increases drastically to 624.5 meV/atom, reflecting a significant destabilizing effect due to H interstitials, while the atomic volume expands to $19.40 \text{ \AA}^3$. This volume expansion suggests that hydrogen introduces substantial lattice strain, altering interatomic distances and promoting local lattice distortions.

For the $\text{Ti}_5\text{Mo}_{95}$ alloy, the negative formation energy of -33.8 meV/atom indicates a thermodynamically stable phase, with minimal volume change after relaxation (-2.7% - 2.7%), suggesting limited structural distortion. However, when hydrogen is introduced ($\text{Ti}_5\text{Mo}_{95}+\text{H}$), the formation energy sharply increases to 660.6 meV/atom, and the atomic volume per atom expands to $19.75 \text{ \AA}^3$, indicating strong lattice distortion effects. The inclusion of oxygen impurities further exacerbates these

effects, as seen in $\text{Ti}_5\text{Mo}_{95}+\text{H}$ [O1+O2], where the formation energy rises to 715.4 meV/atom, and the atomic volume expands significantly to 25.35 Å³ per atom. This highlights the combined impact of H interstitials and oxygen-induced lattice strain, which leads to substantial alterations in phase stability and mechanical behavior.

Table 3. Analysis of formation energy (ΔE_{form}), local lattice distortion (LLD), and volume change in BCC MPEAs with hydrogen incorporation at energetically favorable octahedral sites. ALA6 (Ames Lab Alloy #6) is the alloy composition developed at Ames, $\text{Mo}_{72.3}\text{W}_{12.8}\text{Ta}_{10}\text{Ti}_{2.5}\text{Zr}_{2.5}$ [88,89].

MPEAs	ΔE_{form} [meV/atom]		%Change	V [Å ³ per atom]
	Ideal	Relaxed [LLD]	[Ideal-RLX]/Ideal	
Mo	0	0	0	15.8
Mo + H [O]	624.5	115.3	81.5	19.4
Ti ₅ Mo ₉₅	-33.8	-34.7	-2.7	15.8
Ti ₅ Mo ₉₅ + H [O]	660.6	81.4	87.7	19.8
Ti ₅ Mo ₉₅ + H [O1+O2]	715.4	194.9	72.8	25.4
ALA6	-14.9	-23.4	-57.0	16.1
ALA6 + H	623.5	82.0	86.8	19.6

We also analyzed a Mo-rich high-entropy composition, i.e., ALA6 ($\text{Mo}_{72.3}\text{W}_{12.8}\text{Ta}_{10}\text{Ti}_{2.5}\text{Zr}_{2.5}$ [88,89], which consists of multiple elements, the formation energy remains negative at -14.9 meV/atom, signifying stability. The equilibrium atomic volume is 16.12 Å³, larger than that of pure Mo. Upon hydrogen incorporation (ALA6 + H), however, the formation energy rises sharply to 623.5 meV/atom, demonstrating a destabilizing effect similar to Mo-based systems. The volume expansion is also pronounced, with the per-cell volume increasing from 1934.4 Å³ to 2351.1 Å³ and the per-atom volume reaching 19.59 Å³, reinforcing the role of hydrogen in modifying the local atomic structure and increasing lattice strain. Overall, the results demonstrate that hydrogen significantly destabilizes these alloys by increasing formation energy and inducing local lattice distortions, with oxygen further amplifying these effects in Mo-based systems. The observed volumetric expansion in hydrogenated alloys indicates that interstitial hydrogen alters atomic spacing, affecting phase stability and mechanical properties. These insights are particularly relevant for understanding hydrogen embrittlement, phase transformations, and structural integrity in refractory MPEAs, which are critical for applications in hydrogen storage and high-temperature environments.

3.5.2 Relaxation and hydrogen effect on electronic structure: In this section, we discussed the effect of relaxation on the electronic-structure of Mo, Mo-H, Ti-Mo, and Ti-Mo-H alloys. The presence of a sharp peak in the Mo-H TDOS at the E_{Fermi} in **Fig. 6** shows electronic instability, making it difficult to maintain the crystal integrity upon hydrogen absorption. A high density of states at E_{Fermi} indicates that the system is prone to electronic excitations, which can lead to structural distortions and phase instability. This

instability results in weak metal-hydrogen interactions, reducing the material's ability to effectively store hydrogen. In the case of $\text{Ti}_5\text{Mo}_{95}$, the unrelaxed state shows a pronounced peak at E_{Fermi} , implying poor stability when hydrogen is introduced. Upon relaxation, the redistribution of states reduces the intensity of this peak, slightly improving stability. However, the fundamental electronic structure still lacks the necessary characteristics for effective hydrogen retention. The lack of hydrogen states near the Fermi energy, weak bonding environment, and the presence of unstable electronic states make it challenging for the material to maintain its structural integrity under hydrogenation, further confirming that $\text{Ti}_5\text{Mo}_{95}$ is a poor hydrogen storage candidate.

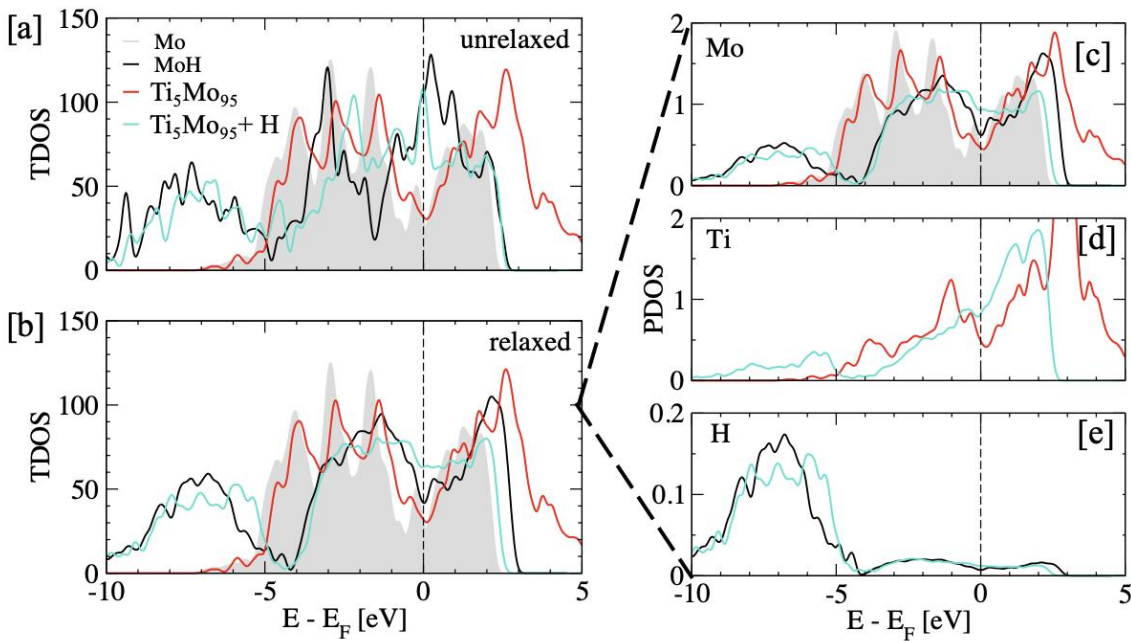


Figure 6. Total and projected density of states (DOS) for Mo, MoH, $\text{Ti}_5\text{Mo}_{95}$, and $\text{Ti}_5\text{Mo}_{95} + \text{H}$ in (a) unrelaxed, and (b) relaxed structures, indicating significant electronic redistribution upon relaxation. (c-e) The PDOS for Mo, Ti, and H, respectively, highlighting the weak metal-hydrogen interactions in $\text{Ti}_5\text{Mo}_{95} + \text{H}$. The absence of H-states near the E_{Fermi} confirms unfavourability of $\text{Ti}_5\text{Mo}_{95}$ for storage. The relaxation removed the sharp peak at the Fermi level in Mo-H that suggests electronic instability [82], which still indicates poor hydrogen storage capability.

Finally, we want to summarize some points related to hydrogen storage behavior in BCC refractory alloys, where strongly bonded H-s states to metal d -states—typically indicated by deep, well-localized hydrogen-derived states far below the Fermi level—are generally unfavorable for efficient hydrogen storage. These deep-lying states reflect strong metal-hydrogen bonding, which, while stabilizing hydride phases, leads to high desorption enthalpies and sluggish hydrogen release kinetics. Such bonding characteristics hinder the reversibility and practicality of hydrogen storage, as significant energy input is

required to liberate hydrogen from the lattice. This suggests that optimal hydrogen storage requires a balance: hydrogen must be sufficiently stabilized to enter the lattice but not so strongly bound that desorption becomes energetically prohibitive. Therefore, overly stable, low-lying H-s states correlate with poor hydrogen release properties, limiting the material's usefulness for practical storage applications.

3.5.3 Validation of weak hydrogen absorption in Mo-rich alloys: The hydrogen absorption behavior of Ti and $\text{Ti}_5\text{Mo}_{95}$ alloys, showing temperature (T) and hydrogen content (wt%) as functions of time (t). Ti and $\text{Ti}_5\text{Mo}_{95}$ compositions were prepared by arc melting of pure elemental ingots on a water-cooled copper hearth, with five cycles of flipping and remelting the alloy ingot to achieve higher compositional uniformity. The hydrogen absorption and desorption tests in Ti and Mo-Ti were done by placing the samples into a pressure-composition isotherm (PCT) experiments (PCT) using Pro2000 system that allows for the control of pressure and temperature for the addition/removal of hydrogen to/from material systems. With this setup, after dynamic vacuum activation at 375°C, we expose samples to a high-pressure (~25 bar) hydrogen atmosphere and monitor the pressure changes during absorption/desorption at temperatures ranging from room temperature to 375°C. The amount of hydrogen absorbed was determined by monitoring the change in hydrogen pressure during measurement along with calibrated volumes and the ideal gas law.

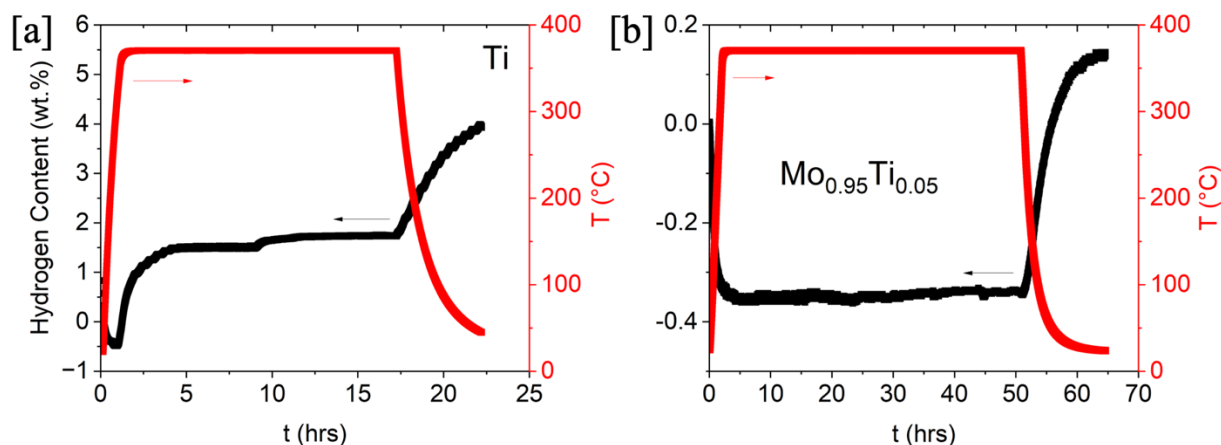


Figure 7. Hydrogen absorption behavior of (a) Ti and (b) $\text{Ti}_5\text{Mo}_{95}$ alloys. Ti shows rapid hydrogen uptake (~3.0 wt.%) while $\text{Ti}_5\text{Mo}_{95}$ absorbs less hydrogen (~0.2 wt.%) due to slower kinetics.

Our experiments on hydrogen absorption process for Ti (**Fig. 7a**) and $\text{Ti}_5\text{Mo}_{95}$ (**Fig. 7b**) also confirms our predictions that Mo-based alloys have slower absorption kinetics and lower capacity, likely due to the presence of Mo, which weakens hydrogen solubility (see **Fig. 10**). The desorption process also differs, with Ti releasing hydrogen more readily, whereas $\text{Ti}_5\text{Mo}_{95}$ shows a more extended desorption

period. These results highlight the influence of alloying on hydrogen storage performance, with Ti being more suitable for high-capacity applications and Mo-containing alloys exhibiting lower absorption.

3.6 Comparative analysis with molecular simulations

Figure 8 shows molecular simulation results for different binary and ternary systems of TiNbMo-based alloys to qualitatively assess hydrogen diffusion at a temperature of 1000 K. Simulations were carried out under two conditions for each composition, one with ideal unrelaxed crystal structure and the other where the atomic arrangement of the structures are relaxed to find its minimum energy configuration.

Figure 8 (a-b) shows the average mean-squared displacements (MSD) of elemental compositions and molecular hydrogen for both unrelaxed and relaxed structures of $\text{Ti}_{90}\text{Nb}_{10}$, $\text{Ti}_{90}\text{Mo}_{10}$, $\text{Nb}_{90}\text{Mo}_{10}$, $\text{Ti}_{40}\text{Nb}_{60}$ and $\text{Ti}_{40}\text{Nb}_{50}\text{Mo}_{10}$ systems. This clearly shows that ‘Ti’ shows a positive effect in hydrogen diffusion for relaxed binary solutions of different weight percent of Ti-Nb systems compared to the unrelaxed ones suggesting that LLD induced during relaxation plays quite an important role in hydrogen diffusion in these structures whereas all the other 3 combinations involving ‘Mo’ ($\text{Ti}_{90}\text{Mo}_{10}$, $\text{Nb}_{90}\text{Mo}_{10}$ and $\text{Ti}_{40}\text{Nb}_{50}\text{Mo}_{10}$) shows a 20-30% reduction in hydrogen diffusivity i.e., the involvement of ‘Mo’ downplays the LLD effect. We have also provided the discussion on role of relaxation on electronic-structure, which corroborates with findings of MD, where relaxation changes diffusion kinetics.

Figure 8 (c-e) depicts the time evolution of average elemental MSDs and radial distribution functions for the $\text{Ti}_{40}\text{Nb}_{60}$ structure. Additionally, when 10% of Nb in $\text{Ti}_{40}\text{Nb}_{60}$ is replaced with ‘Mo’ to form $\text{Ti}_{40}\text{Nb}_{50}\text{Mo}_{10}$, average hydrogen MSD for the relaxed structure (as in **Figure 8 (b)** and **Supplementary Figure S4(a-b)**) shows a six-fold increase than for $\text{Ti}_{40}\text{Nb}_{60}$, although, the MSDs for the unrelaxed structures are similar for these 2 compositions. These findings are consistent with our results obtained from the ML model (**Figure 1**), suggesting that increasing “Mo” shows a negative impact on H/M (see **Fig. 7**). **Supplementary Figure S5** also shows that for both the $\text{Mo}_{10}\text{Nb}_{90}$ and $\text{Mo}_{90}\text{Nb}_{10}$, the relaxed structures have lower average hydrogen MSDs which is contrary to the $\text{Ti}_{40}\text{Nb}_{60}$ (**Figure 8 (c-d)**). This suggests that the lattice distortions induced by the atomic positional reorientation restrict the interstitial diffusion of hydrogen inside the lattice structure, despite maintaining structural integrity, as observed in the radial distribution functions (RDF) in **Supplementary Figure S5 (c)** and **(f)**. **Supporting figures S6** and **S7** show the corresponding MSD and RDF of other structures ($\text{Ti}_{10}\text{Nb}_{90}$, $\text{Ti}_{90}\text{Nb}_{10}$, $\text{Ti}_{10}\text{Mo}_{90}$, and $\text{Ti}_{90}\text{Mo}_{10}$) analyzed in the study and **Supplementary Table S3** shows all the average MSD values of each of the elemental compositions averaged over 10 picoseconds (ps) for different Ti-Nb-Mo systems in the presence of hydrogen.

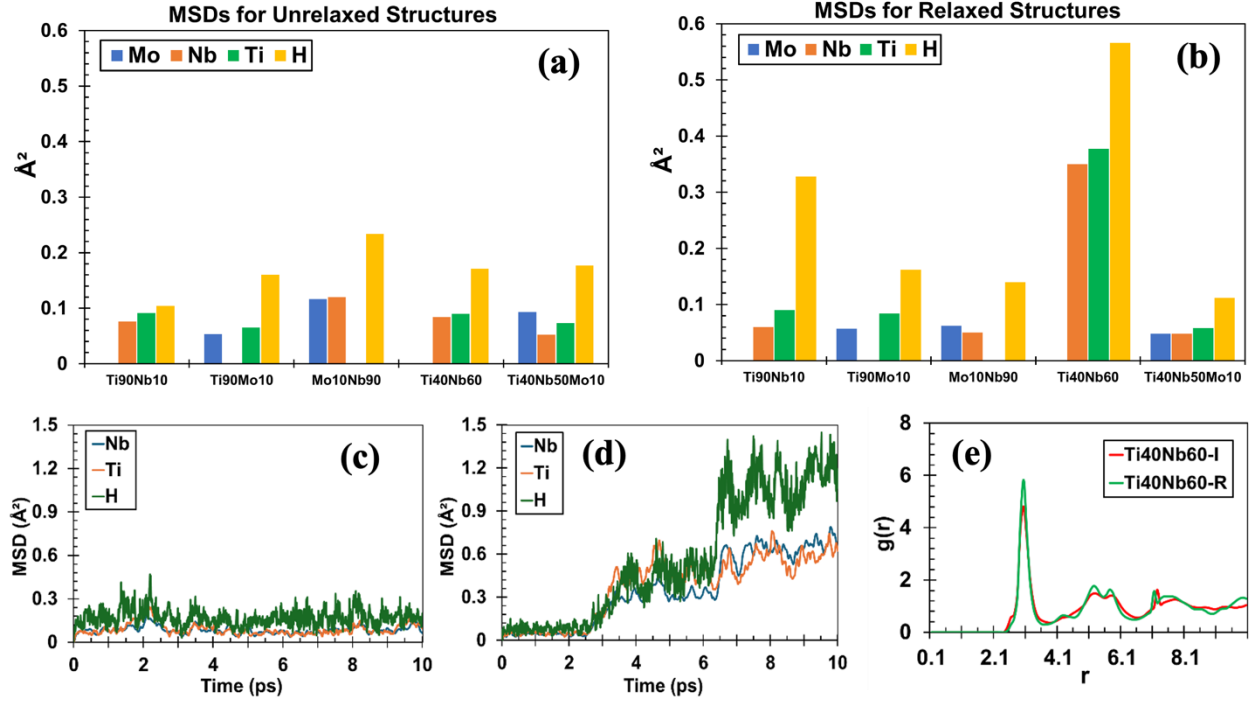


Figure 8. (a-b) Mean square displacements (MSD) of hydrogen and different Ti-Nb-Mo systems in the presence of hydrogen. (c-d) MSDs of elemental compositions in hydrogen environment for (c) unrelaxed, (d) relaxed structures of $\text{Ti}_{40}\text{Nb}_{60}$, and (e) radial distribution functions of both cases ($\text{Ti}_{40}\text{Nb}_{60}\text{-I}$ refers to the unrelaxed and $\text{Ti}_{40}\text{Nb}_{60}\text{-R}$ refers to the relaxed structure respectively).

3.7 CoNi-X (X=Al, Mg, V) systems

Here, we discuss different FCC ternary compositions of Co-Ni-X alloys (**Figure 9**) with X=Al, Mg, and V to denote the effect of different alloying compositions on their hydrogen-storage ability.

From the H/M ternary plots, it is found that the V-rich ($V > 80\%$) zone in CoNiV (**Fig. 9g-i**) and Co-Ni-rich regions of Co-Ni-Al (**Fig. 9a-c**) compositions shows a hydrogen adsorption capacity of more than 1.8-3.0 wt%. However, for CoNi-Mg (**Fig. 9d-f**) with Mg content of $>70\%$ in the solid solution, our model predicts an H/M of 1.4-1.8 and correspondingly an H_2 weight percent of 4%. With increasing Mg content, this causes the alloy to become increasingly more stable as the HCP phase has larger interstitial gaps for hydrogens to occupy. Additionally, by increasing the Mg and Al content beyond 60% in CoNiMg and CoNiAl systems, the solution energy formation becomes more favorable, and the hydrogen weight percent is 1.5-4.0%. However, excessive Al content can also lead to grain coarsening and alter the solidification path, potentially affecting hydrogen storage properties and the presence of high-Mg contributes to high hydrogen weight percentages [85]. This also suggests that while Al and Mg play a crucial role in the thermodynamical stability of the alloy after hydrogenation, the presence of Co and Ni contributes to the formation of intermetallic compounds (like Mg_2Ni) having better hydrogen adsorption and desorption

kinetics. The results suggest a trade-off between hydrogen storage capacity and thermodynamic stability while V-rich compositions show high H/M ratios, Mg-rich compositions offer more favorable solution energy formation [86,87]. Therefore, based on these results, 3 promising compositions are identified to perform an atomistic study on the hydrogen diffusion in bulk alloy phases starting with the initial FCC phases of $\text{Ni}_{70}\text{Co}_{20}\text{Al}_{10}$, $\text{Ni}_{10}\text{Co}_{10}\text{V}_{80}$, and HCP phase of $\text{Ni}_{15}\text{Co}_{10}\text{Mg}_{75}$.

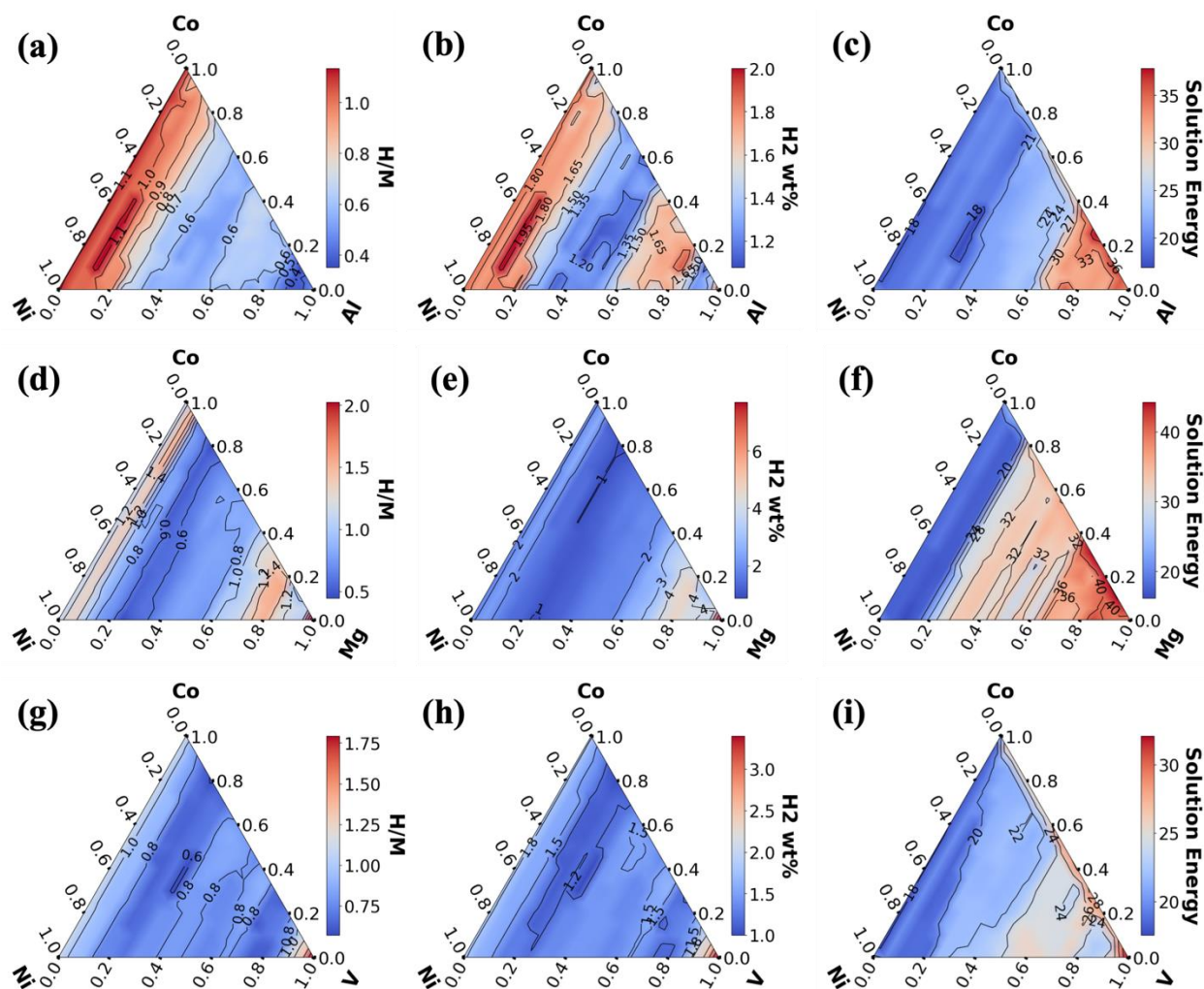


Figure 9. Ternary plots of different transitional metal compositions show the distribution of H/M, H₂ wt %, and solution energy. (a-c) CoNi-Al, (d-f) CoNi-Mg, (g-i) CoNi-V. (a, d, g) are for the H/M, (b, e, h) are for hydrogen weight percent, and (c, f, i) are for the solution energy based on our ML models predictions.

Based on ML predictions of different CoNi-X (X=Al, Mg, V) ternaries, here we present an atomistic insight of hydrogen diffusion characteristics in 3 chosen alloys. **Figure 10** shows the average logarithmic MSDs over 1 ns of hydrogen diffusion simulation at 2 different temperatures on the 3 selected CoNi-based ternary compositions. The actual average MSDs are shown in **Table S4** in **Supplementary information**. We observe a striking 2 orders of magnitude higher average hydrogen mean squared displacements in Mg-

based alloy than the V-rich or Al-rich CoNi-X alloy establishing our previous observation from the ML study (Figure 9(d-f)) that Mg has a strong positive effect on the hydrogen storage capacity and helps in thermodynamic stability of ternary CoNi-Mg alloy. Finally, the thermodynamic stability of CoN-X systems enables multi-principal element alloying without phase separation, ensuring a robust host lattice for hydrogen absorption. Local lattice distortion or average mean square displacements in Co-Ni alloys is a key factor that allows controlled distortions, which enhance hydrogen binding without excessive trapping.

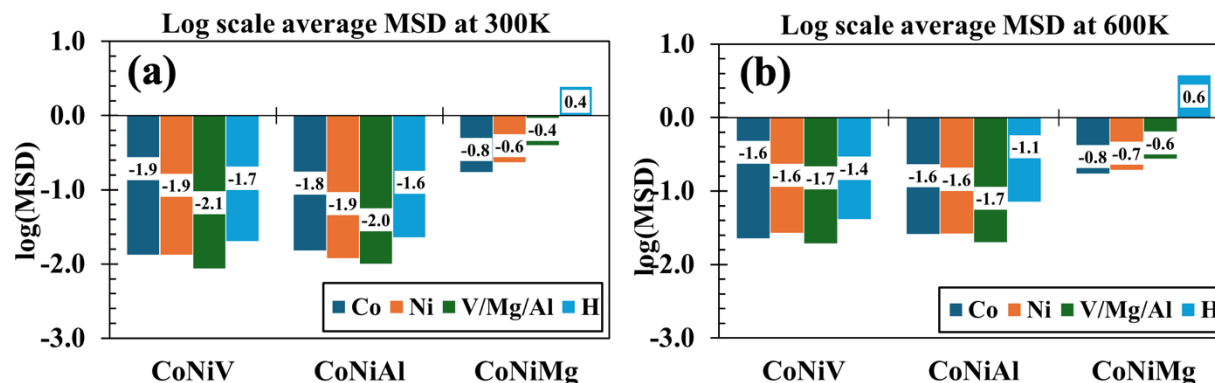


Figure 10. (a-b) Average mean-square displacements (MSD) of elemental compositions in the presence of hydrogen for different CoNi-X ternary systems (in log-scale). The HCP structure of Mg-rich CoNi-Mg exhibit almost 2 orders of magnitude higher hydrogen MSD at both the different temperatures than the V- and Al- ternary systems.

4. Summary

We performed comprehensive feature engineering to identify key properties governing the hydrogen-to-metal (H/M) ratio and solution energy in chemically complex solid-solution alloys. Local lattice distortion (LLD) is a critical factor in refractory-based alloys, influencing both solute miscibility (e.g., H) and mechanical properties. To capture its effect, we incorporated atomic, electronic, and lattice features in our machine-learning (ML) models for predicting H/M and H solubility. Our results highlight that atomic radius, covalent radius, and electronegativity, along with lattice features such as atomic packing fraction (APF) and LLD, dominate predictive outcomes across ensemble-based ML models. Notably, LLD significantly impacts solution energy predictions, with our best-performing model, XGBoost (XGB), achieving an R^2 of 74% and a mean absolute percentage error (MAPE) below 10% on test data, alongside a 57% average R^2 in cross-validation. Additionally, atomic weight, packing fraction, electronegativity, and valence collectively contribute over 60% to H wt.% predictions in Gradient Boosting Regression (GBR) and XGB models. Temperature and average density emerge as key parameters for solution energy predictions, aligning with Van't Hoff equation for thermodynamic equilibrium.

To validate, we tested multiple alloy compositions, revealing that Ti, Nb, and V enhance H-storage capacity across diverse compositions. However, adding just 10-20% Mo reduces H/M and H-wt.% by 40-50%. Specifically, predictions for Ti-Nb-Mo system indicate that increasing Ti and Nb content can elevate H/M to ~ 1.4 and H-storage capacity to ~ 3.0 wt%, whereas Mo negatively impacts storage performance. Molecular simulations confirm these trends, showing that Ti and Nb promote H diffusion, while Mo hinders it. Furthermore, DFT reveals that hydrogen preferentially hybridizes with Mo *d*-orbitals at low energies below the Fermi energy, making Mo unfavorable for hydrogen storage. Phase stability and lattice distortions in bcc Ti-Nb-Mo alloys significantly influence their H-absorption behavior. Despite moderate to low phase stability in Ti-rich regions, favorable Ti-H bonding dynamics enhance H uptake. Incorporating LLD in ML models improves prediction accuracy, underscoring the interplay between electronic structure and atomic displacements in these complex alloys. Finally, the ML-derived periodic table on (a) hydrogen wt% and (b) solution energy will further aid material selection by identifying alloying elements that enhance hydrogen solubility and storage. These findings provide a framework for selecting optimal H-storage materials and guiding future experimental efforts.

5. Acknowledgments

This work was supported by the Laboratory-Directed Research and Development (LDRD) program at Ames National Laboratory. Kevin Ji is grateful for the research opportunity at Ames National Laboratory supported by the U.S. Department of Energy (DOE), Office of Science, Science Undergraduate Laboratory Internships (SULI) program. The Ames National Laboratory is supported by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences, Materials Science and Engineering Division. The Ames National Laboratory is operated for the U.S. DOE by Iowa State University under contract DE-AC02-07CH11358. TB and GB acknowledge the support from the National Science Foundation awards CMMI-1944040 and CMMI-2436601, respectively. PS would like to thank Drs. Matthew D. Wittman, Mark D. Allendorf, and Vitalie Stavila of Sandia National Laboratory for useful discussion and access to updated HyMark database.

7. References

- [1] LeRoy RL, Bowen CT, LeRoy DJ. The Thermodynamics of Aqueous Water Electrolysis. *J Electrochem Soc* 1980;127:1954–62. <https://doi.org/10.1149/1.2130044>.
- [2] Elberry AM, Thakur J, Santasalo-Aarnio A, Larmi M. Large-scale compressed hydrogen storage as part of renewable electricity storage systems. *Int J Hydrogen Energy* 2021;46:15671–90. <https://doi.org/10.1016/J.IJHYDENE.2021.02.080>.
- [3] Banerjee T, Bravo J, Sarunac N, D’Agostini M, Romero C. Sustainable energy storage solutions for coal-fired power plants: A comparative study on the integration of liquid air energy storage and hydrogen energy storage systems. *Energy Convers Manag* 2024;310:118473. <https://doi.org/10.1016/J.ENCONMAN.2024.118473>.

- [4] Rosen MA, Koohi-Fayegh S. The prospects for hydrogen as an energy carrier: an overview of hydrogen energy and hydrogen energy systems. *Energy Ecol Environ* 2016;1:10–29. <https://doi.org/10.1007/S40974-016-0005-Z>.
- [5] Davoodabadi A, Mahmoudi A, Ghasemi H. The potential of hydrogen hydrate as a future hydrogen storage medium. *IScience* 2021;24:101907. <https://doi.org/10.1016/J.ISCI.2020.101907>.
- [6] Bellosta von Colbe J, Ares JR, Barale J, Baricco M, Buckley C, Capurso G, et al. Application of hydrides in hydrogen storage and compression: Achievements, outlook and perspectives. *Int J Hydrogen Energy* 2019;44:7780–808. <https://doi.org/10.1016/J.IJHYDENE.2019.01.104>.
- [7] Witman M, Ek G, Ling S, Chames J, Agarwal S, Wong J, et al. Data-Driven Discovery and Synthesis of High Entropy Alloy Hydrides with Targeted Thermodynamic Stability. *Chemistry of Materials* 2021;33:4067–76. <https://doi.org/10.1021/ACS.CHEMMATER.1C00647>
- [8] Kronberg R, Lappalainen H, Laasonen K. Hydrogen Adsorption on Defective Nitrogen-Doped Carbon Nanotubes Explained via Machine Learning Augmented DFT Calculations and Game-Theoretic Feature Attributions. *Journal of Physical Chemistry C* 2021;125:15918–33. <https://doi.org/10.1021/ACS.JPCC.1C03858>
- [9] Dong S, Wang Y, Li J, Li Y, Wang L, Zhang J. Exploration and design of Mg alloys for hydrogen storage with supervised machine learning. *Int J Hydrogen Energy* 2023;48:38412–24. <https://doi.org/10.1016/J.IJHYDENE.2023.06.108>.
- [10] Verma A, Wilson N, Joshi K. Solid state hydrogen storage: Decoding the path through machine learning. *Int J Hydrogen Energy* 2024;50:1518–28. <https://doi.org/10.1016/J.IJHYDENE.2023.10.056>.
- [11] Borja NK, Fabros CJE, Doma BT. Prediction of Hydrogen Adsorption and Moduli of Metal–Organic Frameworks (MOFs) Using Machine Learning Strategies. *Energies* 2024, Vol 17, Page 927 2024;17:927. <https://doi.org/10.3390/EN17040927>.
- [12] Gao J, Guo X, Wu Y, Xiao W, Hao L. The hydrogen absorption process prediction of AB₂ hydrogen storage device based on data-driven approach. *Int J Hydrogen Energy* 2024;58:657–67. <https://doi.org/10.1016/J.IJHYDENE.2024.01.174>.
- [13] 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;44:7337–44. <https://doi.org/10.1016/J.IJHYDENE.2019.01.261>.
- [14] Bhattacharjee S, Das P, Ram S, Lee S-C. A Hybrid Machine Learning Framework for Predicting Hydrogen Storage Capacities: Unsupervised Feature Learning with Deep Neural Networks 2024.
- [15] Witman M, Ling S, Grant DM, Walker GS, Agarwal S, Stavila V, et al. Extracting an Empirical Intermetallic Hydride Design Principle from Limited Data via Interpretable Machine Learning. *Journal of Physical Chemistry Letters* 2020;11:40–7. <https://doi.org/10.1021/ACS.JPCLETT.9B02971>.
- [16] Suwarno S, Dicky G, Suyuthi A, Effendi M, Witantyo W, Noerochim L, et al. Machine learning analysis of alloying element effects on hydrogen storage properties of AB₂ metal hydrides. *Int J Hydrogen Energy* 2022;47:11938–47. <https://doi.org/10.1016/J.IJHYDENE.2022.01.210>.
- [17] 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;44:7345–53. <https://doi.org/10.1016/J.IJHYDENE.2019.01.264>.
- [18] Zlotea C, Bouzidi A, Montero J, Ek G, Sahlberg M. Compositional effects on the hydrogen storage properties in a series of refractory high entropy alloys. *Front Energy Res* 2022;10:991447. <https://doi.org/10.3389/FENRG.2022.991447>.

- [19] Lu Z, Wang J, Wu Y, Guo X, Xiao W. Predicting hydrogen storage capacity of V–Ti–Cr–Fe alloy via ensemble machine learning. *Int J Hydrogen Energy* 2022;47:34583–93. <https://doi.org/10.1016/J.IJHYDENE.2022.08.050>.
- [20] Wenfeng H, Xiaoqiang T, Chuanxiao C, Shiquan Z, Tian Q, Xueling Z, et al. Molecular dynamics simulation of hydrogen adsorption and diffusion characteristics in graphene pores. *Int J Hydrogen Energy* 2024;69:883–94. <https://doi.org/10.1016/J.IJHYDENE.2024.05.040>.
- [21] Dubouis N, Grimaud A. The hydrogen evolution reaction: from material to interfacial descriptors. *Chem Sci* 2019;10:9165–81. <https://doi.org/10.1039/C9SC03831K>.
- [22] Banerjee T, Balasubramanian G. Predictive Modeling of Molecular Mechanisms in Hydrogen Production and Storage Materials. *Materials* 2023, Vol 16, Page 6050 2023;16:6050. <https://doi.org/10.3390/MA16176050>.
- [23] Oyinbo ST, Jen TC. Reactive molecular dynamics simulations of nickel-based heterometallic catalysts for hydrogen evolution in an alkaline KOH solution. *Comput Mater Sci* 2022;201:110860. <https://doi.org/10.1016/J.COMMATSCI.2021.110860>.
- [24] Herrero CP, Ramírez R. Diffusion of hydrogen in graphite: a molecular dynamics simulation. *J Phys D Appl Phys* 2010;43:255402. <https://doi.org/10.1088/0022-3727/43/25/255402>.
- [25] Halpren E, Yao X, Chen ZW, Singh CV. Machine learning assisted design of BCC high entropy alloys for room temperature hydrogen storage. *Acta Mater* 2024;270:119841. <https://doi.org/10.1016/J.ACTAMAT.2024.119841>.
- [26] Marques F, Balcerzak M, Winkelmann F, Zepon G, Felderhoff M. Review and outlook on high-entropy alloys for hydrogen storage. *Energy Environ Sci* 2021;14:5191–227. <https://doi.org/10.1039/D1EE01543E>.
- [27] Giappa RM, Tylanakis E, Di Gennaro M, Gkagkas K, Froudakis GE. A combination of multi-scale calculations with machine learning for investigating hydrogen storage in metal organic frameworks. *Int J Hydrogen Energy* 2021;46:27612–21. <https://doi.org/10.1016/J.IJHYDENE.2021.06.021>.
- [28] Yang C, Qi J, Wang A, Zha J, Liu C, Yao S. Application of machine learning in MOFs for gas adsorption and separation. *Mater Res Express* 2023;10:122001. <https://doi.org/10.1088/2053-1591/AD0C07>.
- [29] Ding Z, Chen Z, Ma T, Lu CT, Ma W, Shaw L. Predicting the hydrogen release ability of LiBH₄-based mixtures by ensemble machine learning. *Energy Storage Mater* 2020;27:466–77. <https://doi.org/10.1016/J.ENSME.2019.12.010>.
- [30] Maulana Kusdhany MI, Lyth SM. New insights into hydrogen uptake on porous carbon materials via explainable machine learning. *Carbon N Y* 2021;179:190–201. <https://doi.org/10.1016/J.CARBON.2021.04.036>.
- [31] Modi P, Aguey-Zinsou KF. Room Temperature Metal Hydrides for Stationary and Heat Storage Applications: A Review. *Front Energy Res* 2021;9:616115. <https://doi.org/10.3389/FENRG.2021.616115>.
- [32] Hauke J, Kossowski T. Comparison of values of pearson’s and spearman’s correlation coefficients on the same sets of data. *Quaestiones Geographicae* 2011;30:87–93. <https://doi.org/10.2478/V10117-011-0021-1>.
- [33] Zhang S. Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software* 2012;85:2541–52. <https://doi.org/10.1016/J.JSS.2012.05.073>.
- [34] García-Laencina PJ, Sancho-Gómez JL, Figueiras-Vidal AR, Verleysen M. K nearest neighbours with mutual information for simultaneous classification and missing data imputation. *Neurocomputing* 2009;72:1483–93. <https://doi.org/10.1016/J.NEUCOM.2008.11.026>.
- [35] Chiba T, Hino H, Akaho S, Murata N. Correlation and causation. *J Agric Res* 1921;20:557-. <https://doi.org/10.1371/JOURNAL.PONE.0169981>.

- [36] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci* 2021;7:1–24. <https://doi.org/10.7717/PEERJ-CS.623>.
- [37] Kuhn M, Johnson K. Applied predictive modeling. *Applied Predictive Modeling* 2013:1–600. <https://doi.org/10.1007/978-1-4614-6849-3>.
- [38] Elshaw R, Maher M, Sakr S. Automated Machine Learning: State-of-The-Art and Open Challenges. *arXiv:1906.02287* 2019.
- [39] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 2020;415:295–316. <https://doi.org/10.1016/J.NEUCOM.2020.07.061>.
- [40] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 2012;12:2825–30.
- [41] Delhommelle J, Millié P, Delhommelle' J, Millié' P. Inadequacy of the Lorentz-Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation. *Mol Phys* 2001;99:619–25. <https://doi.org/10.1080/00268970010020041>.
- [42] Filippova VP, Kunavin SA, Pugachev MS. Calculation of the parameters of the Lennard-Jones potential for pairs of identical atoms based on the properties of solid substances. *Inorganic Materials: Applied Research* 2015;6:1–4. <https://doi.org/10.1134/S2075113315010062>.
- [43] Jacobson DW, Thompson GB. Revisiting Lennard Jones, Morse, and N-M potentials for metals. *Comput Mater Sci* 2022;205:111206. <https://doi.org/10.1016/J.COMMATSCI.2022.111206>.
- [44] Varshney V, Patnaik SS, Muratore C, Roy AK, Voevodin AA, Farmer BL. MD simulations of molybdenum disulphide (MoS₂): Force-field parameterization and thermal transport behavior. *Comput Mater Sci* 2010;48:101–8. <https://doi.org/10.1016/J.COMMATSCI.2009.12.009>.
- [45] Barraco, M, Neyertz, S, Benes, NE, & Brown, D. Comparison of eight classical Lennard-Jones-based H₂ molecular models in the gas phase at temperatures and pressures relevant to hydrogen on-board storage tanks. *The Journal of Physical Chemistry A* 2023; 127(30): 6335–6346.
- [46] Zhou, XW, El Gabaly, F, Stavila, V, Allendorf, MD. Molecular dynamics simulations of hydrogen diffusion in aluminum. *The Journal of Physical Chemistry C* (2016); 120(14): 7500–7509.
- [47] Konorev, SI, Kozubski, R, Albrecht, M, Vladymyrskyi, IA. Self-diffusion of Fe and Pt in L10-Ordered FePt: Molecular dynamics simulation. *Computational Materials Science* (2021); 192: 110337.
- [48] Plimpton S. Fast parallel algorithms for short-range molecular dynamics 1993. <https://doi.org/10.2172/10176421>.
- [49] Paul WB. Molecular dynamics simulation, elementary methods. By J. M. Haile, Wiley, Chichester 1992, 489 pp. ISBN 0-471-81966-2. *Advanced Materials* 1993;5:223–4. <https://doi.org/10.1002/ADMA.19930050319>.
- [50] Swope WC, Andersen HC, Berens PH, Wilson KR. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys* 1982;76:637–49. <https://doi.org/10.1063/1.442716>.
- [51] Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys* 1984;81:511–9. <https://doi.org/10.1063/1.447334>.
- [52] Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A (Coll Park)* 1985;31:1695. <https://doi.org/10.1103/PhysRevA.31.1695>.
- [53] Kresse G, Hafner J. Ab initio molecular dynamics for liquid metals. *Phys Rev B* 1993;47:558. <https://doi.org/10.1103/PhysRevB.47.558>.
- [54] Kresse G, Furthmüller J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput Mater Sci* 1996;6:15–50. [https://doi.org/10.1016/0927-0256\(96\)00008-0](https://doi.org/10.1016/0927-0256(96)00008-0).

- [55] Perdew JP, Wang Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys Rev B* 1992;45:13244. <https://doi.org/10.1103/PhysRevB.45.13244>.
- [56] Singh P, Harbola MK, Hemanadhan M, Mookerjee A, Johnson DD. Better band gaps with asymptotically corrected local exchange potentials. *Phys Rev B* 2016;93:085204. <https://doi.org/10.1103/PHYSREVB.93.085204>.
- [57] Singh P, Harbola MK, Sanyal B, Mookerjee A. Accurate determination of band gaps within density functional formalism. *Phys Rev B Condens Matter Mater Phys* 2013;87:235110. <https://doi.org/10.1103/PHYSREVB.87.235110>.
- [58] Söderlind P, Turchi PEA, Landa A, Lordi V. Ground-state properties of rare-earth metals: an evaluation of density-functional theory. *Journal of Physics: Condensed Matter* 2014;26:416001. <https://doi.org/10.1088/0953-8984/26/41/416001>.
- [59] Giese TJ, York DM. A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method. *J Chem Theory Comput* 2018;14:1564–82. <https://doi.org/10.1021/ACS.JCTC.7B01175>.
- [60] Singh R, Sharma A, Singh P, Balasubramanian G, Johnson DD. Accelerating computational modeling and design of high-entropy alloys. *Nature Computational Science* 2021 1:1 2021;1:54–61. <https://doi.org/10.1038/s43588-020-00006-7>.
- [61] Sandrock G. A panoramic overview of hydrogen storage alloys from a gas reaction point of view. *J Alloys Compd* 1999;293–295:877–88. [https://doi.org/10.1016/S0925-8388\(99\)00384-9](https://doi.org/10.1016/S0925-8388(99)00384-9).
- [62] Fukai Y. *The Metal-Hydrogen System* 2005;21. <https://doi.org/10.1007/3-540-28883-X>.
- [63] Andreasen A, Risø National Lab. R (DK). MRDepartment. Predicting formation enthalpies of metal hydrides 2004.
- [64] Pálsson GK, Wang YT, Azofeifa D, Raanaei H, Sahlberg M, Hjörvarsson B. Influence of titanium and vanadium on the hydrogen transport through amorphous alumina films. *J Alloys Compd* 2010;494:239–44. <https://doi.org/10.1016/J.JALLCOM.2009.12.178>.
- [65] Shimohata Y, Hamamoto Y, Nishi K, Tanabe K. Improved kinetic model of hydrogen absorption and desorption in titanium with subsurface transport. *Fusion Engineering and Design* 2021;173:112833. <https://doi.org/10.1016/J.FUSENGDES.2021.112833>.
- [66] Chernyayeva T, Ostapov A. HYDROGEN IN ZIRCONIUM Part 1 2013.
- [67] Hamamoto Y, Uchikoshi T, Tanabe K. Comprehensive modeling of hydrogen transport and accumulation in titanium and zirconium. *Nuclear Materials and Energy* 2020;23:100751. <https://doi.org/10.1016/J.NME.2020.100751>.
- [68] Pei P, Song X, Zhao M, Zhang P, Chen G. Influences of V Content on Hydrogen Storage Properties in Low Vanadium Ti-V-Cr Alloy. *Rare Metal Materials and Engineering* 2008;37:1419–23. [https://doi.org/10.1016/S1875-5372\(09\)60039-4](https://doi.org/10.1016/S1875-5372(09)60039-4).
- [69] Gong J, Li Y, Song X, Wang Y, Chen Z. Hydrogen storage of high entropy alloy NbTiVZr and its effect on mechanical properties: A first-principles study. *Vacuum* 2024;219:112754. <https://doi.org/10.1016/J.VACUUM.2023.112754>.
- [70] Sahlberg M, Karlsson D, Zlotea C, Jansson U. Superior hydrogen storage in high entropy alloys. *Scientific Reports* 2016 6:1 2016;6:1–6. <https://doi.org/10.1038/srep36770>.
- [71] Sengar SK, Mehta BR, Kulriya PK, Khan SA. Enhanced hydrogenation and reduced lattice distortion in size selected Pd-Ag and Pd-Cu alloy nanoparticles. *Appl Phys Lett* 2013;103. <https://doi.org/10.1063/1.4826580/26430>.
- [72] Wu C, Zheng X, Chen Y, Tao M, Tong G, Zhou J. Hydrogen storage and cyclic properties of V₆₀Ti_(21.4+x)Cr_(6.6-x)Fe₁₂ (0 ≤ x ≤ 3) alloys. *Int J Hydrogen Energy* 2010;35:8130–5. <https://doi.org/10.1016/J.IJHYDENE.2010.01.017>.
- [73] Huang Z, Cuevas F, Liu X, Jiang L, Wang S, Latroche M, et al. Effects of Si addition on the microstructure and the hydrogen storage properties of Ti_{26.5}V₄₅Fe_{8.5}Cr₂₀Ce_{0.5} BCC solid

- solution alloys. *Int J Hydrogen Energy* 2009;34:9385–92. <https://doi.org/10.1016/J.IJHYDENE.2009.09.055>.
- [74] Aoki M, Noritake T, Ito A, Ishikiriya M, Towata SI. Improvement of cyclic durability of Ti–Cr–V alloy by Fe substitution. *Int J Hydrogen Energy* 2011;36:12329–32. <https://doi.org/10.1016/J.IJHYDENE.2011.07.019>.
- [75] Balcerzak M. Hydrogenation properties of nanocrystalline TiVMn body-centered-cubic alloys. *Int J Hydrogen Energy* 2020;45:15521–9. <https://doi.org/10.1016/J.IJHYDENE.2019.02.092>.
- [76] Biswas A, Morozovska AN, Ziatdinov M, Eliseev EA, Kalinin S V. Multi-objective Bayesian optimization of ferroelectric materials with interfacial control for memory and energy storage applications. *J Appl Phys* 2021;130:204102. <https://doi.org/10.1063/5.0068903/15272032>.
- [77] Miedema AR, Buschow KHJ, Van Mal HH. Which intermetallic compounds of transition metals form stable hydrides ? *Journal of the Less Common Metals* 1976;49:463–72. [https://doi.org/10.1016/0022-5088\(76\)90057-6](https://doi.org/10.1016/0022-5088(76)90057-6).
- [78] Boeri L, Hennig R, Hirschfeld P, -Superconductivity in CeBeH al, J Buschow KH, P Bouten PC, et al. Hydrides formed from intermetallic compounds of two transition metals: a special class of ternary alloys. *Reports on Progress in Physics* 1982;45:937. <https://doi.org/10.1088/0034-4885/45/9/001>.
- [79] de Boer FR, Boom R, Mattens WCM, Miedema AR, Niessen AK. *Cohesion in Metals - Volume 1 of Cohesion and Structure* 1988.
- [80] Somo TR, Lototsky M V., Yartys VA, Davids MW, Nyamsi SN. Hydrogen storage behaviours of high entropy alloys: A Review. *J Energy Storage* 2023;73:108969. <https://doi.org/10.1016/J.EST.2023.108969>.
- [81] Smithson H, Marianetti CA, Morgan D, Van Der Ven A, Predith A, Ceder G. First-principles study of the stability and electronic structure of metal hydrides. *Phys Rev B* 2002;66:144107. <https://doi.org/10.1103/PhysRevB.66.144107>.
- [82] Singh P, Smirnov A V., Johnson DD. Ta-Nb-Mo-W refractory high-entropy alloys: Anomalous ordering behavior and its intriguing electronic origin. *Phys Rev Mater* 2018;2:055004. <https://doi.org/10.1103/PHYSREVMATERIALS.2.055004>.
- [83] Singh P, Vela B, Ouyang G, Argibay N, Cui J, Arroyave R, et al. A ductility metric for refractory-based multi-principal-element alloys. *Acta Mater* 2023;257:119104. <https://doi.org/10.1016/J.ACTAMAT.2023.119104>.
- [84] Song H, Tian F, Hu QM, Vitos L, Wang Y, Shen J, et al. Local lattice distortion in high-entropy alloys. *Phys Rev Mater* 2017;1:023404. <https://doi.org/10.1103/PHYSREVMATERIALS.1.023404>.
- [85] Dong X, Li Y, Zhai Y, Liu Z, Zhang G, Yang F. Study on Microstructure and Hydrogen Storage Properties of Mg₈₀Ni₁₆–xAlxY₄ (x = 2, 4, 8) Alloys. *Metals (Basel)* 2024;14:126. <https://doi.org/10.3390/MET14010126>.
- [86] Yartys VA, Lototsky M V., Akiba E, Albert R, Antonov VE, Ares JR, et al. Magnesium based materials for hydrogen based energy storage: Past, present and future. *Int J Hydrogen Energy* 2019;44:7809–59. <https://doi.org/10.1016/J.IJHYDENE.2018.12.212>.
- [87] Baran A, Polański M. Magnesium-Based Materials for Hydrogen Storage—A Scope Review. *Materials* 2020;13. <https://doi.org/10.3390/MA13183993>.
- [88] Ouyang G, Singh P, Su R. et al. Design of refractory multi-principal-element alloys for high-temperature applications. *npj Comput Mater* 2023;9:141. <https://doi.org/10.1038/s41524-023-01095-4>.
- [89] Singh, P, Sharma, A, Smirnov, AV et al. Design of high-strength refractory complex solid-solution alloys. *npj Comput Mater* 2018;4:16. <https://doi.org/10.1038/s41524-018-0072-0>.