A Simple Constrained Machine Learning Model for Predicting Materials for High Pressure Hydrogen Compression

Using machine learning techniques to predict chemistries of materials with novel properties has been of great interest to the materials community over the past few years. Unfortunately, many of these approaches will predict the figure of merit for any combination of materials with no insight as to the potential stability of the material or its engineering feasibility. While evaluation of materials stability can be challenging, the application of techno-economic constraints to materials design is relatively straightforward and can be implemented early in the prediction process. In this manuscript we used a free open source materials machine learning platform on a free open source experimental database to generate thousands of new alloy combinations with favorable enthalpies of formation for high pressure compressors. We then whittled down our list of potential alloys using a series of engineering constraints such as enthalpy values, cost, and simple stochiometric rules. This enabled us to focus our stability check to the Fe-Mn-Ti-X system. Comparison of the Fe-Mn-Ti to CALPHAD, previous experimental studies, and multiple DFT studies resulted in contradictory predictions of stability indicating that it is a system with the potential to provide insights to materials scientists and engineers.

A Simple Constrained Machine Learning Model for Predicting Materials for High Pressure Hydrogen

Compression

Jason R. Hattrick-Simpers[1], Kamal Choudhary[1], Claudio Corgnale[2]

[1] Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD

20899

[2] Greenway Energy LLC, Aiken, SC 29803

Abstract

Here we present the results of using techno-economic analysis as constraints for machine learning

guided studies of new metal hydride materials. Using existing databases for hydrogen storage alloys a

regression model to predict the enthalpy of hydrogenation was generated with a mean absolute error of

8.56 kJ/mol and a mean relative error of 28%. Model predictions for new hydride materials were

constrained by techno-economic analysis and used to identify 6110 potential alloys matching the criteria

required for hydrogen compressors. Additional constraints such as alloy cost, composition, and likely

structure were used to reduce the number of possible alloys for experimental verification to less than

400. Finally, expert heuristics and a novel machine learning approach to approximating alloy stability

were employed to select the Fe-Mn-Ti-X alloy system for future experimental studies.

Introduction

Recently there has been an explosion in the use of machine or advanced statistical methods

materials science. For about a decade, groups have been building automated data analysis tools to

extract knowledge from large datasets[1–4]. Efforts in this area have included automated phase diagram

generation from combinatorial samples and the use of unsupervised techniques to extract information

from spectral scanning probe microscope datasets[5–7]. In the past few years, new materials discovery via

machine learning has become increasingly of interest.[8–10] Large data mining efforts have demonstrated

that machine learning models, trained on computational and experimental datasets, can create

predictions of materials that satisfy key scientific or technological criteria, the so-called "inverse design"

problem[11–13]. For instance, there has been great progress in using large datasets built from finite

element analysis simulations to design structural alloys with favorable microstructures and elastic

properties.[14] Recently, Ward et. al. demonstrated a generalized machine learning platform, called

Magpie, that ingests experimental or theoretical datasets, maps material composition into a multi-

dimensional attribute space, and creates models that can predict material performance.[15]

In its current iteration, Magpie enables the predictions of tens of thousands of potential

materials with promising properties via algorithms which are largely physics agnostic. In a recent paper

from Ward et. al. the band gaps of more than 4500 compounds from the OQMD database were

predicted via regression analysis, an exciting development since band gaps are computationally difficult

to predict via DFT.[15] From their work, they identified 223 materials that were likely to have favorable

band gaps. This presents a new challenge for the materials science field; in that it is not possible to

experimentally validate all these new materials in a reasonable amount of time. Even high-throughput

(combinatorial) methods are not capable of efficiently screening 223 compounds distributed among tens

of three component systems at the rate of hypothesis generation. An additional problem is that of

experimental material stability. The previous example creates a model that can calculate the band gap

for any potential compound but does not determine if it is (meta) stable or its likely crystal structure. In this case, using the OQMD dataset partially mitigated this concern, but many studies focus on exploring materials outside of OQMD or other existing theoretical databases. Material properties are intimately linked to their crystal structure, so thaving an idea of what structures a new compound is likely to form in would provide an expert insight into whether the material will exhibit the predicted properties. Interestingly, predictions are also made in the absence of other concerns such as material cost or compatibility with existing technologies or health and safety concerns associated with production, which could be pivotal in deciding which material leads to pursue.

Here, we demonstrate a layered approach to constraining the theoretical-experimental search space by applying a series of techno-economic criteria prior to and after materials prediction. Our proof of principle case is in the prediction of new high-pressure hydrogen storage alloys to be used in vehicle fueling stations. Although the DOE currently sees multiple avenues for successfully delivering hydrogen to fueling stations, each requires the compression of hydrogen to pressures exceeding 500 bar. This presents a techno-economic challenge as the compression technology must meet specific technical criteria to be viable (e.g. H_2 delivery pressure, compressor specific energy, etc), but it also must meet economic criteria (e.g. uninstalled capital cost, annual maintenance cost, and lifetime). In principle, compressors making use of hydrogen storage alloys would have lower cost, higher flow rates, and better reliability than mechanical compressors. However, there is currently no readymade materials technology that matches all the DOE techno-economic criteria. There is, however, a wealth of materials data within open hydride databases that could be used to train machine learning models to guide researchers.

In this study, the existing DOE metal hydride database was consumed, filtered, and then used as training data for a regression model that predicts the enthalpy and entropy of formation for metal hydrides. The regression model was found to accurately predict values of enthalpy, Pearson's

correlation coefficient ~ 0.8, but was substantially worse at modeling entropy. Overfitting of the training

set was tested by using the model to predict the values of enthalpy for 47 held out samples with a mean

average error (MAE) of 3.37 kJ/mol and a mean relative error (MRE) of 20%. The model was then used

to predict a set of new potential hydrogen storage alloys constrained by TEA considerations including

elements to be used and the enthalpies required to meet technical targets, resulting in 6110 alloy

predictions. An additional set of TEA constraints including overall alloy cost and proximity of the

composition to Laves-type phases was used to further reduce the potential alloys by more than an order

of magnitude. Although this substantially contracts the potential materials search space it does not

address the issue of material stability or structure. Subsequently, a combination of expert heuristics and

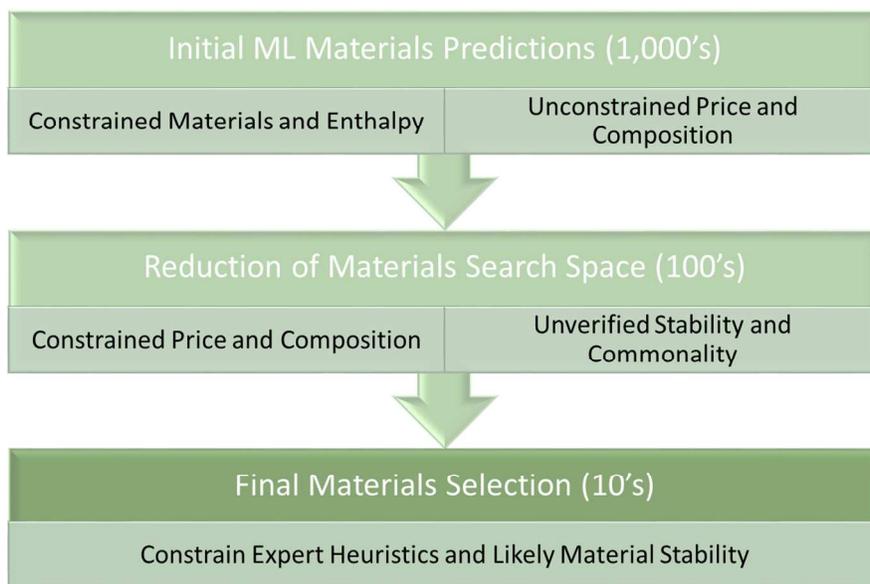cluster expansion was used to narrow down to quaternaries of the form Fe-Mn-Ti-X and to determine if

Figure 1. Schematic representation of the process for generating initial materials predictions and reducing the overall materials search space to those likely to have the apporpriate properties, cost, and were likely to be stable.

the materials are likely to be stable near the predicted region in the Fe-Mn-Ti ternary.

Techno-Economic Analysis

Metal hydride-based hydrogen compressors

The US DOE envisions two main options for large scale hydrogen delivery. One employs a hydrogen compression unit (pressures up to about 100 bar) with hydrogen storage in geologic formations. The delivery scenario is realized by a separate high pressure compression system (with pressures up to approximately 875 bar) that delivers the hydrogen to distributed pipelines[16]. TEA was performed targeting high pressure compression systems to reach pressures on the order of 875 bar as required by the DOE targets [16], shown in Table 1.

Table 1 DOE Techno-economic targets for refueling station hydrogen compressors at 100 kg$_{H2}$/h

	FY2020	Ultimate target
Availability	85%	≥90%
Compressor specific energy (kWh/kg)	1.6 (100 bar pipeline delivery)	1.4 (120 bar pipeline delivery)
Uninstalled capital cost ($) for 100 kg/h	275,000 (100 bar pipeline delivery)	170,000 (120 bar pipeline delivery)
Annual maintenance	4%	2%
Outlet operating pressure (bar)	875	950
Lifetime (years)	10	>10

Presently, mechanical compressors cannot achieve the DOE targets. Mechanical compressors account for over half of the station's cost, have lower reliability than that required by the DOE targets and have low flow rates for a mature fuel cell market. Failure in conventional compressors is often due

to fatigue associated with moving parts, and is exacerbated by the repeated starts and stops expected at fueling stations.[17]

Hydrogen compressors based on metal hydride materials have the potential to overcome all the limitations associated with traditional mechanical compressors and to achieve all the targets shown in Table 1. MH materials absorb hydrogen through an exothermic chemical reaction and release the absorbed hydrogen reversibly, through an endothermic chemical reaction. The equilibrium pressures for the chemical reactions are a direct function of their operating temperatures. Therefore, hydrogen can be absorbed at low temperatures and pressures and, by providing higher temperature thermal power during the desorption process, the hydrogen pressure can be increased without the use of external electric power.

Typical MH equilibrium isotherm profiles are shown in Figure 2, for two isotherm pressure profiles at different temperatures (T_1 and T_2 with $T_2 > T_1$) for a nominal MH material. For the current application, the relatively flat absorption/desorption regions, where the MH phase change occurs and most of the hydrogen is absorbed/desorbed, are of interest.
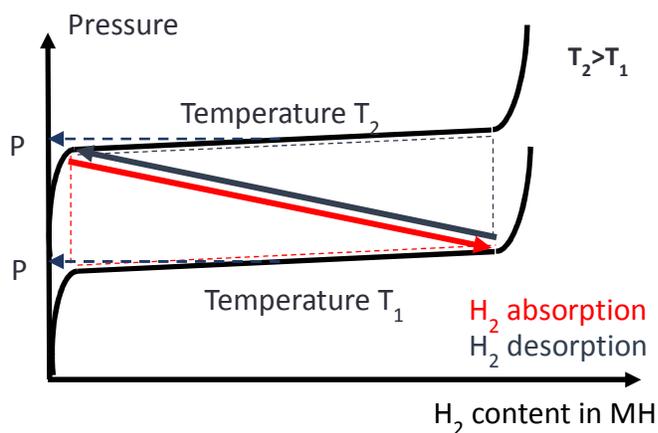

Figure 2. Typical equilibrium temperature and pressure relationship for common MH materials during absorption at T1 and desorption at T2

The reactions in these regions can be approximated as occurring at a relatively constant pressure, for a fixed temperature, thus identifying the equilibrium operating conditions (e.g. P_1 - T_1 and P_2 - T_2 in Figure 2). The objective is to find a material operating at P_1 on the order of 100 bar and corresponding T_1 on the order of 30-40 °C and P_2 on the order of 875 bar and corresponding T_2 of about 120-140 °C. In addition, the isotherm two-phase reaction should occur at near constant pressure with minimal hysteresis (i.e. variation of the equilibrium pressures) during charging and discharging.

A comprehensive review of the available MH materials and some of the already developed heat transfer and pressure vessel concepts, operating at maximum pressures on the order of 600-700 bar can be found in Reference 18. Currently available MH for hydrogen compression are based on intermetallic materials. They can be classified into two main groups: (1) the Rare Earth MH materials, such as $LaNi_5$ or $MmNi_5$ hydrides and (2) the Ti-based MH materials. Given the current operating pressure range, the Ti-based MH materials (laves phase, or AB_2 phase) represent the only feasible group. These materials are generally based on Ti (A element) with different possible combinations of other metal elements, such as Cr, Mn, V, Ni, as the B elements. Sometime small quantities of Zr are included in the 'A' term, to replace Ti. In general, the A and B site can incorporate different alloying elements and exhibit improved material performance (e.g. reduced hysteresis and flatter plateaus) for the required operating temperatures and pressures. Depending on the formulation of the AB_2 materials, the operating pressures can range between 10 bar and over 1000 bar. A comprehensive list of existing AB_2 materials for hydrogen compression applications, both for low and high pressures, can be found in Reference 18.

Acceptable MH material property envelope for hydrogen compressor

An inverse system analysis has been carried with the objective of assessing the technical and economic properties of the ideal MH compression system meeting the DOE targets. A comparison with

the currently AB_2 available materials has also been carried out, to evaluate the gaps between the ideal and the current MHs and to guide the discovery of new materials.

Techno-economic model and assumptions

A high level steady state lumped parameter techno-economic model was developed to analyze the performance of the proposed compression system. The system mass balance is expressed as follows. The mass of each MH material (the compression system is comprised of two parallel MH) can be estimated as:

$$m_{MH} = \frac{\dot{m}_{H2} \cdot \Delta t}{wf} \qquad (1)$$

Equation 1 assumes that the absorption time is equal to the desorption time (Δt).

The equilibrium pressure and temperature values during absorption and desorption are estimated using the Van't Hoff equation (Equation 2), which is derived from the Gibbs energy expression:

$$p = \exp\left(\frac{\Delta H}{RT} - \frac{\Delta S}{R}\right) \qquad (2)$$

Equation 2 expresses the Van't Hoff relationship assuming the material hysteresis is negligible, thus implying the reaction enthalpy and entropy values are the same for charging and discharging processes.

The system energy balance was assessed as follows. During the hydrogen charging/discharging process, the cooling/heating power was estimated including: (1) the chemical reaction latent thermal power, (2) the MH material sensible cooling/heating power to be provided to reach the required operating pressure, (3) the wall and tubing material sensible cooling/heating power to achieve the

required operating temperatures for the MH material. More details about the technical model will be provided in a future publication.

The technical assumptions relative to the ideal MH material and system properties as well as the initial constraints, data, and assumed degrees of freedom are summarized in Table 2. The unknown quantity of the technical analysis is the material reaction enthalpy required to achieve the DOE efficiency target for the 2020 scenario

Table 2 Ideal metal hydride material properties assumptions and constrains

MH material properties	
Bulk density (kg m^{-3})	3000
Overall material thermal conductivity (W m^{-1} K^{-1})	8
wf (%)	1.1
Δt (min)	8
Porosity (%)	35
System constraints	
Pressure (bar)	100-875
Temperatures (°C)	40-120
Hydrogen flow rate (kg h^{-1})	100

The ideal material properties are assumed based on typical values for Ti-based laves materials. The overall thermal conductivity value takes into account inclusion of expanded natural graphite at 10 wt% or metal foam structures, to increase the thermal conductivity.[19] The weight capacity of the material (1.1%) was assumed based on the typical Ti-based high pressures AB_2 values, between 1% and 1.7%.[18] The conservative assumption is justified by the reduced Ti content in the ideal metal hydride

material (as explained in the next section), which results in a decrease of the material weight capacity. A reasonable charging/discharging time of 8 minutes (i.e. total cycling time of 16 minutes) was assumed, based on the fast kinetics typical of these materials. The system was analyzed for hydrogen flow rates of 100 kg/h, to be compressed between 100 bar and 875 bar at temperatures between 40 °C (MH equilibrium temperature corresponding to 100 bar) and 120 °C (MH equilibrium temperature corresponding to 875 bar).

The economic-financial model was developed as described in the following. The system installed cost was calculated based on the MH material cost and the cost of the additional equipment required to transfer the required heating/cooling power. A baseline initial configuration based on traditional shell and tube heat exchangers was assumed. The required heating/cooling thermal power was assessed based on the DOE efficiency target (required electricity per kg of hydrogen), evaluating the primary thermal input to produce the required electricity. An overall thermal-electric efficiency of 34% was assumed, including the electric transmission inefficiencies. The lifetime cost of the MH compressor system was calculated and compared with the corresponding values obtained for the mechanical compressor that achieves the DOE targets.

Results and considerations

The techno-economic model was applied to find the thermodynamic properties (i.e. reaction enthalpy) and economic properties (i.e. cost of the MH material) required to fully meet the DOE targets, with the assumptions and constrains shown in Table 2 and described above. The techno-economic analysis results are shown in Table 3.

Table 3 Ideal MH material properties required to meet the DOE 2020 techno-economic targets

Reaction enthalpy (kJ/mol$_{H2}$)	≤22

Operating Pressure (bar)/ Temperature (°C)	100-875 / 40-120
FOB material cost ($/kg)	≤250
Raw MH material cost ($/kg)	≤1.6

A MH material with a reaction enthalpy on the order of 22 kJ/molH$_2$, reaching pressures of 875 bar at temperatures on the order of 120 °C, can meet the DOE 2020 efficiency target. The enthalpy target represents a feasible value for AB$_2$ type high pressure metal hydrides, however, very few known Ti based MHs can meet this target.[18] From an economic perspective, a treated material having a cost of 250 $/kg (or lower) can achieve the lifetime cost required to meet the DOE economic targets. The material cost is relative to the Free On Board (FOB) cost of the MH powder, acquired in large quantities and already processed (i.e. annealing process included in the cost). The most common Ti-based MH materials, acquired in minimum quantities of about 10 kg, have costs on the order of $800 /kg, including heat treatments, with values on the order of 1,800 $/kg for hydrides with high Cr content, such as TiCr$_{1.9}$.[20] Based on the data available from JMC (USA)[20,21] the ratio between the current Ti MH material FOB cost and the raw MH material cost is approximately 155, ranging between a minimum of 130 and a maximum of 180. In addition, the data available from JMC (USA) showed that only a reduction of the Ti content to values of approximately 20-25%, with inclusion of other inexpensive elements (e.g. Fe, Mn), along with massive quantity production have the potential to bring the cost down to the required values. Thus, the required FOB cost target can be achieved only if the following conditions are met: (1) raw MH material cost equal to (250 $/kg /155) = 1.6 $/kg, (2) expensive elements are eliminated from the material formulation, including V, Zr, Mo, Cr; and (3) the Ti molar content is reduced to less than 20-25% in the material formulation. The results of the TEA in Table 3 were used as constraints for the ML model described below.

DOE Database Ingestion

To provide MagPie a training data set, the entire "Hydrogen Storage Materials Database" was downloaded from the Department of Energy's Fuel Cell Technologies Office.[22] Although the website offers several search and filtering tools, none of those were employed and all dataset cleaning was done locally. The hydrogen storage materials database contains the composition and hydrogen gravimetric capacity of 2722 different hydrogen storage materials spread out over traditional interstitial Laves phase metal hydrides, complex hydrides, magnesium hydrides, solid-solution interstitial hydrides, and miscellaneous hydrides. More fundamental thermodynamic data, such as the enthalpy and entropy of
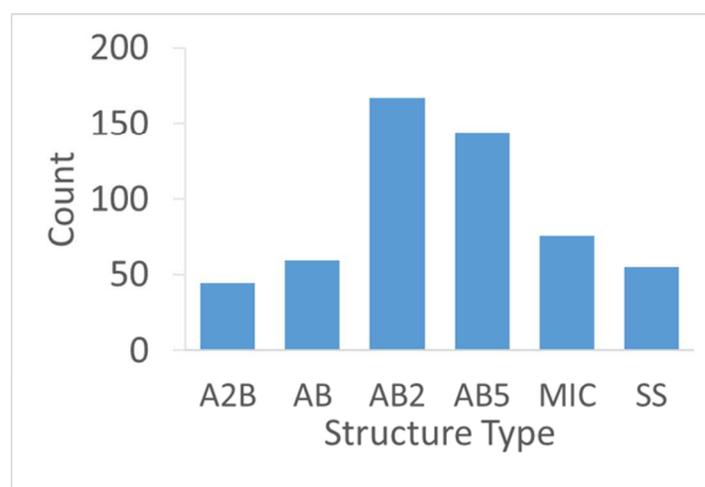

Figure 3. Histogram of structural types contained in the training data set.

formation, are reported only for a subset of the overall alloys.

For this study, only reversible metal alloys were used (e.g. no complex hydrides), leaving 1815 compounds in the training set. A second filter was applied to eliminate compounds for which the enthalpy of formation was not explicitly reported in the table, further reducing the initial training set to 545 compounds. Laves phases are heavily represented in the resulting dataset, particularly AB_2 and AB_5 structures like $TiCr_2$ or $LaNi_5$ which have been heavily studied in the technical literature, see the

histogram in Figure 3. The solid solution and miscellaneous labeled data was kept because Lave phase

materials are often heavily A or B site substituted. We reasoned that information from the solid solution

data could empirically inform the model about the role of substitution in determining the value of

enthalpy and entropy. Also, including miscellaneous dataset provided additional stoichiometries of

metal hydrides that didn't fall into the traditional Laves phase regions potentially reducing bias in the

models. Entropy values not contained within the original dataset were calculated using the Van't Hoff

relationship if the equilibrium pressure was given for any temperature. This yielded a total of 503 total

compounds with associated entropies for training.

Machine Learning Model Development

The machine learning model was created using the MagPie code developed by Ward *et. al.*[15]

Magpie first transforms each compound in the training set into a set of 145 attributes that are built from

properties like stoichiometries, elemental property statistics, electronic structure, and ionic compound

attributes. The models and attributes, as built, did not explicitly include the structure of each

compound. This provides the developed machine learning model a robust set of machine interpretable

descriptors for subsequent model training.

A few simple models from the Weka software platform for instance RepTree, Random Forest

Regression, and Neural Networks[23] were used to predict hydride enthalpies. We found that 10-fold

validated Random Forest Regression provided the highest Spearman (0.8031) and Pearson (0.7558)

correlation coefficients of the three methods tested. RepTree provided Pearson and Spearman

coefficients of 0.6829 and 0.6747, respectively, while for the neural network both coefficients were

below 0.5. For Random Forest Regression, the mean absolute error (MAE) was 8.56 kJ/mol, the mean

relative error (MRE) is 28% and the ROC value of 0.79. In Magpie, the ROC value is calculated by

classifying the predictions as being above or below a moving threshold value from the minimum to

maximum predicted value. The calculated MAE is relatively large; however, a cursory investigation of the dataset gives a strong indication that the error in the predictions is at least partially caused by the spread of the enthalpy values reported for any given compound. For instance, Mg_2Ni has 13 different values for the enthalpy reported ranging from 31.3 kJ/mol to 71.3 kJ/mol with a mean of 62.9 kJ/mol and an experimental MAE of 5.2 kJ/mol. As an extreme example, TiCu has two enthalpy values reported, 126 kJ/mol and 75 kJ/mol.

To validate the model was not overfit, we used a holdout dataset. 46 alloys outside of the Hydrogen Storage Materials Database and their associated thermodynamic data were selected from the literature. The model was used to predict their enthalpy of formation and the predictions were then compared to the literature values. Comparison between the known and predicted values revealed a lower MAE of 3.4 kJ/mol and lower mean relative error (MRE) of 20% than was obtained during the 10-fold cross validation of the full training set. The 10 compounds with the highest relative error include several Ti – Cr based AB_2 structures that have B-site substitution with Mn, Fe, Mo, etc, and a few TiCrMn related alloys. These compounds have a relative error in excess of 37.5% with a maximum error of 64.4%. The next group of alloys all show similar structures/base materials but with the relative error decreasing to below 20%. Interestingly, one alloy that the model under-predicts is $TiCr_{1.78}$, the model predicts an enthalpy of 21 kJ/mol but the experimental value is reported as 28.04 kJ/mol. The experimental value here is somewhat surprising. In the hydride database, $TiCr_2$, $TiCr_{1.9}$, and $TiCr_{1.8}$ have averages reported enthalpies of 26.35 kJ/mol, 26.19 kJ/mol, and 19.6 kJ/mol, respectively, marking the predicted value as being intuitively correct.

Including these 46 materials into the training of the model decreases the Pearson's Correlation to 0.7690 and increases the Spearman's Correlation to 0.8095, with little impact on the overall MAE and ROC AUC. Although for the hold-out compounds the MAE reduces to 1.11 kJ/mol and the MRE reduces

to 4.6%. The model developed using both archival and hold-out data was used for all subsequent

studies.

Similar analysis was performed to develop models for predicting the entropy of hydrogenation.

The best performing model was still Random Forest Regression, unfortunately a convincing case for a

strong model couldn't be made. For instance, the Pearson's and Spearman's Correlations were found to

not exceed 0.46 and 0.42, respectively. Here the issue is likely the dispersion of data in the dataset.

Although it is expected that, on average, the entropy should fall somewhere around 130 J/mol*K, values
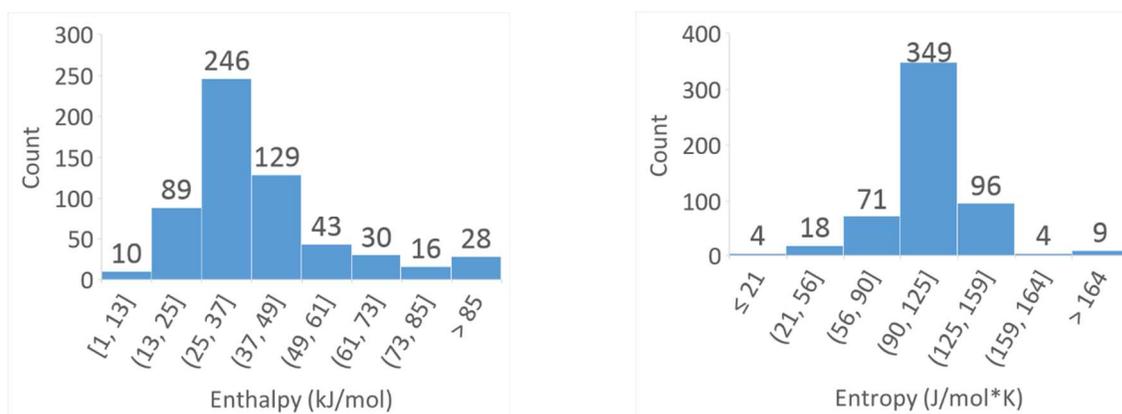

Figure 4. Histograms of enthalpy and entropy values for the training data set (original DOE dataset and the holdout set). Note that the entropy values have a large number of unphysical values above and below the value of 130 J/mol for H2 gas phase hydrogen.

in the dataset vary from 9 J/mol*K to 625 J/mol*K with an average of 110 J/mol*K and a standard

deviation of 35 J/mol*K, see Figure 4. An attempt was made to constrain the training data set to a

"realistic range" of 95 J/mol*K to 140 J/mol*K, this reduced the training set to 282 entries but didn't

substantially improve the predictions.

Predicting Materials Based on TEA Constraints

From the above model, predictions for the enthalpy of formation of metal hydrides can be made

with reasonable certainty, given the spread in experimental values from the archival dataset. To limit

the range of predictions that are made, a set of criteria from TEA were implemented as filters at

different stages during the prediction cycle. Filters for the elemental constituents of the alloy and overall enthalpy were implemented prior predicting new alloys. Elements considered in this study were identified chosen based on a balance of their known hydrogen storage properties and cost. The full list of elements considered was Ca, Al, Si, Fe, Mg, Na, Mn, Zn, Cr, Mo, and Ti. The maximum number of elements per compound was varied from 2 – 4. Based on the targets identified by the TEA and a MAE of 8 kJ/mol, predictions having enthalpies above 15 kJ/mol and below 40 kJ/mol were considered. Any compound within 0.3 at% of any compound contained in the dataset, as measured by the L1 norm (Manhattan distance), was discarded from the predicted dataset. Once the predictions were run, a total of 6110 different alloys were proposed with values of enthalpy ranging from 18 kJ/mol to 30 kJ/mol.

To reduce the number of possible alloys for consideration, a series of post-prediction filters were employed based on the TEA and expert heuristics. The first filter down sampled the predicted alloys based on a $1.6/kg threshold for the alloy cost. This dropped the number of possible alloys by nearly a factor of six down to 962 possibilities. The second filter was used to limit potential material compositions to those similar to AB, AB_2, A_2B, or AB_5 Laves phases. The rationale was that the training data was predominantly composed of Laves phases so that the predictions were likely to be most accurate for that class of materials. This criterion further reduced the number of alloys to be considered to the 533 contained in the table in the SI. As a note, the price of Ti depends very strongly on the purity. The previous discussion supposed a Ti cost of $3.8/kg. If higher purity Ti is required ($8/kg), then 336 alloys are possible.

Several trends are apparent in the remaining materials set, firstly ordered in terms of relative abundance of elements the predictions contain: Fe (396 predictions), Mn (376 predictions), Ti (177 predictions), Si (141 predictions), Al (131 predictions), Mg (111 predictions), Cr (59 predictions), and Mo (14 predictions). Indicating that almost all remaining predicted compounds were ternary or quaternary alloys that used Fe-Mn as a base alloy. No alloys containing Ca, Na, and Zr were present in the final list.

The predicted enthalpies varied from 22 kJ/mol up to 30 kJ/mol, the predictions are skewed heavily towards higher enthalpies with more than 90% of all predictions being above 27 kJ/mol. Not surprisingly Ti containing compounds dominate the predictions at lower enthalpies and thus drive up the cost. Compounds with enthalpies below 25 kJ/mol are all more than $1.2/kg as a consequence.

The final step in choosing alloys for subsequent study was verifying that they can form stable single-phase alloys either solid solution or laves phases and were likely to absorb hydrogen. Fe-Mn was not reported in the database as being a hydrogen storage alloy and indeed the binary phase diagram for Fe-Mn does not exhibit any Laves or intermetallic type structures. The database does, however, contain many AB_2 structures that incorporate Mn and Fe as co-B site substituents. Some of these compounds including ZrMnFe and $Zr_{0.7}Ti_{0.3}MnFe$ show enthalpies between 20.4 kJ/mol and 10 kJ/mol, therefore compounds containing mixtures of Mn and Fe would tend to be identified as favoring low enthalpies. Both $TiMn_2$ and FeTi are known intermetallics that can store hydrogen at low enthalpies, 26.7 kJ/mol and 24.6 kJ/mol, respectively. Since Mn and Fe are miscible within one another, compounds of the form $Ti(Mn_{1-x}Fe_x)_2$ could be stable for $0 < x < 1$. Likewise, Mn and Ti have some miscibility and thus $FeTi_2$-Mn alloys could exist with Mn substituting into both the A and B site. One predicted alloy, $Ti_{0.25}Mn_{0.25}Fe_{0.50}$ could be seen as an A and B site substituted $TiMn_2$ or as an extension of the Fe-Mn solid solution. There is also a known $TiFe_2$ C_{14} phase with an alloying window at high temperature, although this composition has not been reported to absorb hydrogen.[24] Phase diagrams from Murakami et. al. and Hughes suggested a large region of solubility between $TiFe_2$ and $TiMn_2$ which would include Ti compositions down to 20 at% at 1273 K.[25] Subsequent CALPHAD studies confirmed the $Ti(Mn_{1-x}Fe_x)_2$ tie-line but were inconclusive in terms of the range of A-site substitution.[26]

Table 4 Machine learning proposed compositions from the Mn-Fe-Ti-X system with promising enthalpies of formation

Alloy Composition	Enthalpy

$Ti_{0.20}Mn_{0.20}Fe_{0.50}Si_{0.10}$	25.8
$Ti_{0.25}Mn_{0.25}Fe_{0.50}$	21.8
$Ti_{0.25}Mn_{0.20}Al_{0.05}Fe_{0.50}$	24.4
$Ti_{0.25}Mn_{0.20}Fe_{0.50}Si_{0.05}$	25.5
$Mg_{0.05}Ti_{0.20}Mn_{0.25}Fe_{0.50}$	26.8
$Mg_{0.10}Ti_{0.15}Mn_{0.25}Fe_{0.50}$	26.6
$Ti_{0.25}Mn_{0.10}Al_{0.05}Fe_{0.60}$	26.4
$Ti_{0.25}Mn_{0.10}Fe_{0.60}Si_{0.05}$	26.7
$Ti_{0.30}Mn_{0.10}Al_{0.05}Fe_{0.55}$	27.0
$Mg_{0.10}Ti_{0.20}Mn_{0.20}Fe_{0.50}$	27.3

From this analysis, 10 alloys (Table 4) were selected for future study. Ti-Mn-Fe was chosen as the base alloy and quaternary alloys containing Mg, Si and Al were selected as potential additives. Mg and Al were chosen to improve gravimetric capacity and reduce alloy cost. While Si was chosen primarily to reduce cost. Since all of the alloys were based on the Ti-Mn-Fe alloy a genetic algorithm guided DFT approach was used to evaluate the likelihood of alloy stability off the $TiMn_2$ – $TiFe_2$ tieline.

We used the genetic algorithm for structure and phase prediction (GASP) to identify the low-energy Ti-Mn-Fe structures[27] using Vienna Ab-initio Software Package (VASP)[28] as energy calculator using density functional theory. The genetic algorithm started with an initial population of random structures that broadly sampled the phase space. The structures were then relaxed and low-energy structures were preferentially selected as parents to create child structures using genetic operators such as mutation and mating. When enough child structures had been created, they in turn were selected to make offspring of their own. In the TiMnFe structure searches, the number of atoms varied, and we

used an upper limit of 30 atoms per cell. The maximum lattice length was constrained to 50 Å and only

create primitive unit cells were created. We employed the phase-diagram searching mode of the

algorithm, which allows the stoichiometry to vary, and we stoped the searches after 500 structure

relaxations. We employed Opb88vdw functional, 600 eV plane wave cutoff and k-point mesh density of

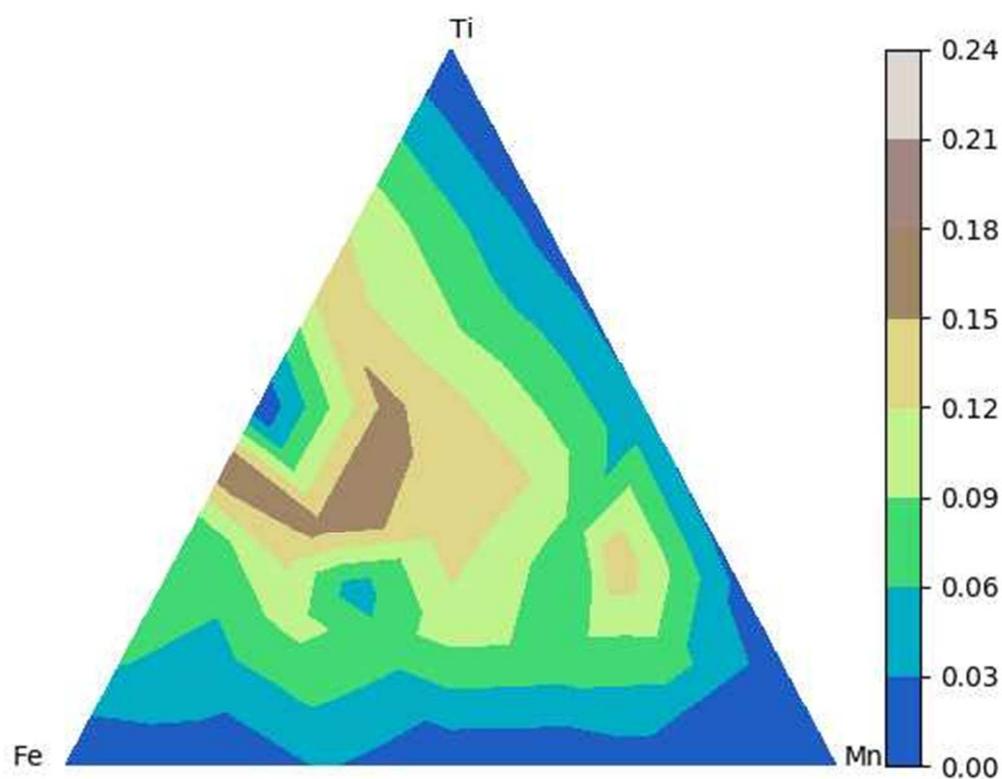only 20 k-points per Angstrom for spin-polarized DFT calculations using JARVIS-DFT workflow.[29]

Figure 5. Predicted stability of Ti-Mn-Fe based on crystal structure searches using genetic

algorithm and heat of formation calculations using density functional theory. Points are colored based

on relative likelihood of formability at 0 K. The color bar shows the energy above hull values. The higher

the energy above hull value the more unstable material should be.

In Figure 5 the energy above convex hull is shown in a colormap with linear interpolation for Ti-Mn-Fe system for structures identified during genetic algorithm based DFT calculations. The genetic algorithm attempted 5800 structure searches and narrowed down to predict 100 structures for which DFT calculations were performed. All the phases found during genetic algorithm search are used in calculating the convex hull and then the stable ones are used in generating the triangular surface plot in Figure 5. These constraints were based on bond-length and other criteria mentioned above. The predicted stabilities were based on heat of formation values obtained from DFT. All the TiMnFe based compounds had negative formation energies, indicating that these compounds should be energetically possible to form. The heat of formation data was then converted to a convex hull to realize relative stability of materials. Deep blue color in Figure 5 indicates that the systems are on the hull or are the stable material. Here, the $TiMn_2$-$TiFe_2$ region is not observed to be stable (as shown in Figure 5), along the $TiMn_2$ – $TiFe_2$ tieline and only the $TiMn_2$ phase was found to be stable. In contrast, the Materials Project, clearly shows stability of both the $TiMn_2$ and $TiFe_2$ phases and previous experimental work and theoretical work have illustrated the stability along the tieline.[30] Here, it is likely that the genetic algorithm was not able to identify the stable phase from the 100 structures calculated via DFT.

Conclusions

Here we presented the results of using techno-economic analysis as constraints for machine learning guided studies of new metal hydride materials. Using existing databases for hydrogen storage alloys and regression analysis we were able to identify 6110 potential alloys that met the thermodynamic criteria required for hydrogen compressors. Additional constraints such as alloy cost and composition were used to reduce the number of possible alloys for experimental verification to less than 400. Finally, expert heuristics and a novel machine learning approach to approximating alloy stability was employed to select the Fe-Mn-Ti-X alloy system for future experimental studies. Interestingly, we observed conflicting theoretical predictions of materials stability in the Fe-Mn-Ti range

with two separate DFT studies and an early CALPHAD study providing quite different predictions. In contrast, previous experimental studies of the Fe-Mn-Ti have indicated a large range of alloy stability in the composition regions predicted to be interesting. This suggests that Fe-Mn-Ti would be a potentially interesting system for exploration both experimentally (to reconfirm the ranges of solubility) and computationally via modern CALPHAD techniques.

Conflict of Interest:

There are no conflicts of interest to declare

Acknowledgements:

References:

1   C. Suh, A. Rajagopalan, X. Li and K. Rajan, *Data Sci. J.*, 2002, **1**, 19–26.

2   Y. Saad, D. Gao, T. Ngo, S. Bobbitt, J. R. Chelikowsky and W. Andreoni, *Phys. Rev. B*, 2012, **85**, 104104.

3   I. Takeuchi, M. Lippmaa and Y. Matsumoto, *MRS Bull.*, 2006, **31**, 999–1003.

4   A. Corma, M. Diazcabanas, M. Moliner and C. Martinez, *J. Catal.*, 2006, **241**, 312–318.

5   R. Le Bras, R. Bernstein, J. M. Gregoire, S. K. Suram, C. P. Gomes, B. Selman and R. B. van Dover,
        Proc. 28th Int. Conf. Artif. Intell. AAAI, 2014.

6   A. G. Kusne, T. Gao, A. Mehta, L. Ke, M. C. Nguyen, K.-M. Ho, V. Antropov, C.-Z. Wang, M. J.
        Kramer, C. Long and I. Takeuchi, *Sci. Rep.*, 2014, **4**, 6367.

7   L. Vlcek, A. Maksov, M. Pan, R. K. Vasudevan and S. V. Kalinin, *ACS Nano*, 2017, **11**, 10313–10320.

8   J. K. Bunn, J. Hu and J. R. Hattrick-Simpers, *JOM*, 2016, **68**, 2116–2125.

9   A. G. Kusne, D. Keller, A. Anderson, A. Zaban and I. Takeuchi, *Nanotechnology*, 2015, **26**, 444002.

10   S. K. Suram, J. A. Haber, J. Jin and J. M. Gregoire, *ACS Comb. Sci.*, 2015, **17**, 224–233.

11   A. R. Oganov, A. O. Lyakhov and M. Valle, *Acc. Chem. Res.*, 2011, **44**, 227–237.

12   G. Hautier, C. C. Fischer, A. Jain, T. Mueller and G. Ceder, *Chem. Mater.*, 2010, **22**, 3762–3767.

13   J. Ling, M. Hutchinson, E. Antono, S. Paradiso and B. Meredig, *Integr. Mater. Manuf. Innov.*, 2017,
        6, 207–217.

14   D. B. Brough, D. Wheeler, J. Warren and S. R. Kalidindi, *Curr. Opin. Solid State Mater. Sci.*, 2017,
        21, 129–140.

15   L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, *npj Nat. Comput. Mater.*, 2016, **2**, 16028.

16   *DOE MYRDD Plan, Section 3.2, Hydrogen Delivery*, 2015.

17   *DOE Funding Opportunity Announcement (FOA) Number: DE-FOA-0001412, AOI 1, Topic 2*, 2016.

18   M. Lototskyy, V. A. Yartys, B. G. Pollet and R. C. Bowman, *Int. J. Hydrogen Energy*, 2014, **39**,

5818–5851.

19   C. Corgnale, B. Hardy, D. Tamburell, S. Garrison and D. Anton, *Int. J. Hydrogen Energy*, 2012, **37**,

        2812–2824.

20   *JMC (USA), Inc. Personal Communication*, 2017.

21   *Certain commercial equipment, instruments, or materials are identified in this paper to foster*

        understanding. Such identification does not imply recommendation or endorsement by the

        National Institute of Standards and Technology, nor does it imply that the products or services

        identified are necessarily the best available for the purpose.

22   *Hydrogen Storage Materials Database http://hydrogenmaterialssearch.govtools.us/search.aspx*, .

23   E. Frank, M. Hall, L. Trigg, G. Holmes and I. H. Witten, *Bioinformatics*, 2004, **20**, 2479–81.

24   G. D. Sandrock, J. J. Reilly and J. R. Johson, *Metallurgical Considerations in the Production and Use*

        of FeTi Alloys for Hydrogen Storage, 1976.

25   D. Dew-Hughes, *Metall. Trans. A*, 1980, **11**, 1219–1225.

26   D. Dew-Hughes and L. Kaufman, *Calphad*, 1979, **3**, 175–203.

27   B. C. Revard, W. W. Tipton, A. Yesypenko and R. G. Hennig, *Phys. Rev. B*, 2016, **93**, 54117.

28   G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15–50.

29   K. Choudhary, I. Kalish, R. Beams and F. Tavazza, *Sci. Rep.*, 2017, **7**, 5179.

30   A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner,

        G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 11002.

Nomenclature

FOB = Free on Board

MH = Metal hydride

m = mass (kg)

$\overset{\bullet}{m}$ = mass flow rate (kg s^{-1})

P = pressure (bar)

T = temperature (K)

ΔH = metal hydride reaction enthalpy (kJ mol$_{H2}^{-1}$)

ΔS = metal hydride reaction entropy (kJ K^{-1} mol$_{H2}^{-1}$)

R = Universal gas constant (8.314 J mol$_{H2}^{-1}$ K^{-1})

Δt = charging or discharging time (s)

wf = weight capacity of the metal hydride (kg$_{H2}$ kg$_{MH}^{-1}$)

Subscripts

MH = metal hydride

H2 = hydrogen