OPEN ACCESS
IOP Publishing | Songshan Lake Materials Laboratory
Materials Futures
Mater. Futures 3 (2024) 025602 (14pp)
https://doi.org/10.1088/2752-5724/ad4a94

Data-driven design of high pressure hydride superconductors using DFT and deep learning

Daniel Wines* and **Kamal Choudhary**

Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, United States of America

E-mail: daniel.wines@nist.gov

Received 23 February 2024, revised 1 May 2024
Accepted for publication 12 May 2024
Published 31 May 2024

Abstract

The observation of superconductivity in hydride-based materials under ultrahigh pressures (for example, H_3S and LaH_{10}) has fueled the interest in a more data-driven approach to discovering new high-pressure hydride superconductors. In this work, we performed density functional theory (DFT) calculations to predict the critical temperature (T_c) of over 900 hydride materials under a pressure range of (0–500) GPa, where we found 122 dynamically stable structures with a T_c above MgB_2 (39 K). To accelerate screening, we trained a graph neural network (GNN) model to predict T_c and demonstrated that a universal machine learned force-field can be used to relax hydride structures under arbitrary pressures, with significantly reduced cost. By combining DFT and GNNs, we can establish a more complete map of hydrides under pressure.

Supplementary material for this article is available online

Keywords: superconductivity, density functional theory, high-throughput, materials discovery, hydrides, deep learning

1. Introduction

The discovery of superconductors with a high transition temperature (T_c) has been a long-standing goal in the condensed matter physics community. The highest measured T_c for a Bardeen–Cooper–Schrieffer (BCS) [1, 2] conventional superconductor (mediated by electron–phonon interactions) in ambient conditions has been measured to be 39 K for MgB_2 [3]. These moderately low T_c values (significantly less than liquid nitrogen) of traditional BCS superconductors can be overcome by applying ultra-high pressures to hydrogen-rich compounds [4, 5]. Metallic hydrogen and hydrogen-rich compounds are ideal for high temperature superconductivity because hydrogen atoms provide strong electron–phonon coupling (EPC) and high frequency phonon modes [6, 7]. Recent experimental observation (within the past decade) of conventional superconductivity in hydride-based materials under high pressure has reinvigorated the interest in searching for new hydride superconductors and minimizing the required pressure to sustain superconductivity at a high temperature [4, 5].

This renowned interest in high pressure hydrides was reignited in 2015 with the experimental measurement of superconductivity at 203 K for H_3S under pressure [8]. In addition to H_3S, a T_c above 250 K has been experimentally observed for LaH_{10} above 170 GPa [9–11], with several experimental and computational studies confirming these results [9–14]. A T_c of 243 K has been observed for YH_9 at 201 GPa [15] and superconductivity has also been measured for ThH_{10} [16], YH_4 [17], LuH_3 [18], zirconium polyhydrides [19], and tin hydrides [20] under pressure. More recently, calcium superhydrides such as

* Author to whom any correspondence should be addressed.

CaH_6 have been measured to have a T_c of up to 210 K at (160–190) GPa [21]. It has also been demonstrated that doping can play a role in enhancing T_c for hydride materials under pressure [22–25]. Recent claims of room temperature superconductivity in Ca-H-S (under 190 GPa) [26] and Lu-H (under 10 GPa) [27] systems have been met with scrutiny from the scientific community, with [26] and [27] being retracted. Despite the lack of reproducibility in these recent claims, the search for new hydride superconductors with a high T_c continues on the experimental and computational front.

In contrast to unconventional superconductors, where there is not a well-defined theoretical framework to describe superconductivity, BCS theory can be utilized as a tool to screen potential superconducting materials in conjunction with electronic structure methods such as density functional theory (DFT). With the advancements in computing power, DFT databases, and robust and efficient computational workflows, high-throughput calculations (on the order of hundreds to thousands) of superconducting properties are now in the realm of possibility [28–35]. In addition, various machine learning and data-driven approaches have been used to aid in the discovery of new and novel superconductors [36–48]. In our previous works, we expanded the existing JARVIS (Joint Automated Repository for Various Integrated Simulations) [33, 35] DFT database to include calculations for over 1000 bulk [36] and 150 two-dimensional (2D) superconductors [37] (all at zero pressure). We also developed a deep-learning tool (trained on our DFT dataset) to predict the EPC parameters and T_c of any material [36] at a significantly reduced cost. We went on to utilize this deep-learning tool for forward and inverse design of new superconductors [36, 45].

In addition to our own work, there has been a recent high-throughput DFT and machine learning study which performed EPC calculations for over 7000 superconducting candidates [48] (all at zero pressure), which makes it one of the largest high-throughput DFT study of conventional superconductors to date. In contrast to the DFT and machine learning works that span several material classes, there have been few high-throughput computational efforts that have focused on high pressure hydrides [46]. Recently, the Superhydra database was created by Saha *et al* [49] to map superconductivity in high pressure hydrides, but the DFT calculations have so far been limited to binary compounds at a single pressure value of 200 GPa. In addition, Belli *et al* [50] analyzed the electronic and structural properties of more than 100 hydride superconductors calculates with DFT (at several pressures ranging from 0 GPa to 500 GPa) and found a strong correlation between T_c and electronic bonding network, establishing a clear path to engineer the T_c [50]. In a seminal work in 2021, Shipley *et al* [46] surveyed the landscape of binary hydrides up until 500 GPa for the entire periodic table and performed high-throughput DFT calculations for the best candidates at pressure values of 10, 100, 200, 300 and 500 GPa. From these results, they found 36 dynamically stable high pressure hydride superconductors with a T_c above 100 K, with 18 of these materials having a T_c over 200 K [46].

Building on our previous efforts within JARVIS and inspired by the recent efforts of other researchers focused on hydrogen-based superconductors, we established a high-throughput workflow to screen potential hydride superconductors. Specifically, we extended our search beyond binary compounds. We ran DFT calculations for over 900 hydrogen-based materials. These calculations involved relaxing the hydride structures under various amounts of pressure and then performing EPC calculations to determine superconducting properties. We went on to further analyze dynamically stable structures (containing all positive phonon frequencies in the phonon density of states) with the highest T_c and the structures that were not previously known to be superconducting. After running these DFT calculations, we trained a deep learning model (using the relaxed high-pressure structures and computed T_c values) to predict T_c given a crystal structure, specifically tailored to high pressure hydrides. Such deep learning models have shown remarkable potential for superconductor design [36, 45, 48, 51–53]. After testing and benchmarking our newly trained deep learning property prediction model, we used a unified machine learning-based force-field to relax the hydride structures under pressure, further reducing the number of DFT calculations needed to obtain a complete map of superconducting properties under pressure.

2. Methodology

In order to compute the EPC and T_c of hydride materials, we performed non-spin polarized density functional perturbation theory (DFPT) [54, 55] with the Gaussian broadening (interpolation) method [56]. Similarly to our previous workflows in [36], [37], and [45], we used the Quantum Espresso [57] software package, Garrity–Bennett–Rabe–Vanderbilt (GBRV) pseudopotentials [58] and the Perdew–Burke–Ernzerhof functional revised for solids PBEsol [59]. For Lu, a Topsakal–Wentzcovitch [60] pseudopotential was used. Due to the high-throughput nature of this study, which is primarily focused on screening potential high-pressure hydride candidates, we used preconverged k-points from the JARVIS database [61], a $2 \times 2 \times 2$ q-point grid and a kinetic energy cutoff of 610 eV (45 Ry). It was demonstrated in [36] and [37] that these lower q-point grids can still be effective for material screening purposes. For structures not already in the JARVIS dataset (materials added from the literature), we used the automated k-point convergence scheme in JARVIS [61].

The Eliashberg spectral function ($\alpha^2F(\omega)$) [62] that determines the EPC can be obtained from:

$$\alpha^2F(\omega) = \frac{1}{2\pi N(\epsilon_F)} \sum_{qj} \frac{\gamma_{qj}}{\omega_{qj}} \delta(\omega - \omega_{qj}) w(q) \tag{1}$$

where $N(\epsilon_F)$ is the density of states (DOS) at the Fermi level ϵ_F, ω_{qj} is the mode frequency, δ is the Dirac-delta function, $w(q)$ is the weight of the q point, and γ_{qj} is the phonon mode j linewidth at wave vector q:

$$\gamma_{qj} = 2\pi\omega_{qj} \sum_{nm} \int \frac{d^3k}{\Omega_{BZ}} |g^j_{kn,k+qm}|^2 \delta(\epsilon_{kn} - \epsilon_F) \delta(\epsilon_{k+qm} - \epsilon_F). \tag{2}$$

The integral is over the first Brillouin zone, $g^{j}_{kn,k+qm}$ is the electron–phonon matrix element, and ϵ_{kn} and ϵ_{k+qm} are the DFT eigenvalues with wavevector k and $k+q$ within the nth and mth bands. The relation between γ_{qj} and the mode EPC parameter is:

$$\lambda_{qj} = \frac{\gamma_{qj}}{\pi h N(\epsilon_{\mathrm{F}})\omega^{2}_{qj}}. \tag{3}$$

The EPC parameter can now be written as:

$$\lambda = 2\int \frac{\alpha^{2}F(\omega)}{\omega}\mathrm{d}\omega = \sum_{qj}\lambda_{qj}w(q) \tag{4}$$

where $w(q)$ is the weight of a q point. The McMillan–Allen–Dynes [63] formula can be used to approximate the superconducting transition temperature (T_{c}) and is written as:

$$T_{\mathrm{c}} = \frac{\omega_{\log}}{1.2}\exp\left[-\frac{1.04(1+\lambda)}{\lambda-\mu^{*}(1+0.62\lambda)}\right] \tag{5}$$

where

$$\omega_{\log} = \exp\left[\frac{\int \mathrm{d}\omega \frac{\alpha^{2}F(\omega)}{\omega}\ln\omega}{\int \mathrm{d}\omega \frac{\alpha^{2}F(\omega)}{\omega}}\right] \tag{6}$$

In equation (5), the constant μ^{*} is the effective Coulomb potential parameter, which we take to be 0.1. To generate our DFT dataset, it took roughly 11 500 core hours. Approximately 7% of the computational time was devoted to relaxing the structures under pressure while the remaining 93% of the computational time was devoted to DFPT calculations of the EPC.

To predict the superconducting properties of each high pressure hydride material, we used the atomistic line graph neural network (ALIGNN) [64]. Specifically, we trained an ALIGNN model on our new high pressure hydride DFT dataset, and trained an ALIGNN model on a combination of our high pressure hydride DFT dataset and our bulk superconducting DFT dataset from [36]. It is important to note that both of these ALIGNN models were trained on DFT relaxed structures under pressure. Within ALIGNN, the crystal structure is represented as a graph (elements are nodes, bonds are edges). Each node in the graph has nine input features, which include block, electronegativity, valence electrons, group number, electron affinity, ionization energy, atomic volume, and covalent radius. The bond distances are the edge features and the radial basis function cutoff is 8 Å. In addition, a periodic graph construction with 12-nearest neighbors is used. The line graph is constructed from the atomistic graph, using bond distances as nodes and bond angles as edges. The node and edge features are updated with edge-gated graph convolution using a propagation function. One layer is composed of an edge-gated graph convolution on the bond graph and an edge-gated convolution on the line graph. The line graph convolution produces bond messages that propagate to the atomistic graph, where the atom and bond features are updated further. For superconducting transition temperature, we used 300 epochs for training, a 80:10:10 split and a batch size of 16 (the test set was not used at all during training). The hyperparameters were kept the same as the original ALIGNN paper by Choudhary and DeCost [64]. ALIGNN was designed to be a flexible framework that could be applied to any material class for an incredibly diverse range of material properties. In fact, this formalism has been tested on over 52 properties for materials and molecules that range the entire periodic table, with great success [64]. Extensive benchmarking and testing of model parameters such as node features, optimized hyperparameters, number of epochs and batch size are presented in [64]. For consistency and due to the fact that we have a similar dataset size [36], we used the same number of epochs and the same batch size as our previous superconducting work [36], in addition to the same set of hyperparameters. In addition to the several follow-up studies which use ALIGNN [35] (including our previous superconducting work [36]) which used the same nine node features, we used the same nine node features in the current work. It is entirely possible that adding an additional descriptor to ALIGNN can improve the model to cater to the characteristics of BCS superconductors under pressure, but that is something beyond the scope of this work. ALIGNN is implemented in PyTorch [65] and the deep graph library [66]. To train our ALIGNN models for our newly generated hydride dataset and a combination of our hydride dataset and previous dataset, it roughly took 4 and 7 GPU hours, respectively.

3. Results and discussion

3.1. High-throughput DFT results

Figure 1 depicts our full high-throughput workflow used to investigate high pressure hydride superconductors. The first step involves screening the existing JARVIS-DFT database, which contains over 80 000 materials and millions of computed properties. The first task was to screen materials that contain hydrogen and have a unit cell of 8 or less atoms (structures where the EPC can feasibly be computed, with the exception of certain binary superhydrides such as LaH_{10}). After this first round of screening, we screened certain DFT computed properties from JARVIS-DFT such as the formation energy per atom (E_{form}), the energy above the convex hull (E_{Hull}), the electronic band gap (E_{gap}), and the magnetic moment per atom (M). To ensure the superconducting candidate materials were the most thermodynamically stable, metallic and nonmagnetic, the criteria we established (similar to our previous work [36, 37, 45]) was $E_{\mathrm{form}} < 0$ eV/atom, $E_{\mathrm{Hull}} < 0.5$ eV/atom, $E_{\mathrm{gap}} < 0.05$ eV, and $M < 0.05$ μ_{B}. These properties in the JARVIS-DFT database are calculated with the Opt-b88-vdW [67] functional, which tends to underestimate the band gap. This band gap underestimation is not of concern for screening purposes since we only wish to identify metallic structures. We chose such a generously high E_{Hull} cutoff due to the possibility of applied pressure stabilizing structures (it is possible that a structure at 0 GPa could lie very far above the hull but at higher pressure could lie on the hull). In addition to pulling candidate structures from JARVIS-DFT, we added candidate structures from literature. These included H-based structures from the 3DSC [47] database. In addition, several binary compounds from the work of Shipley

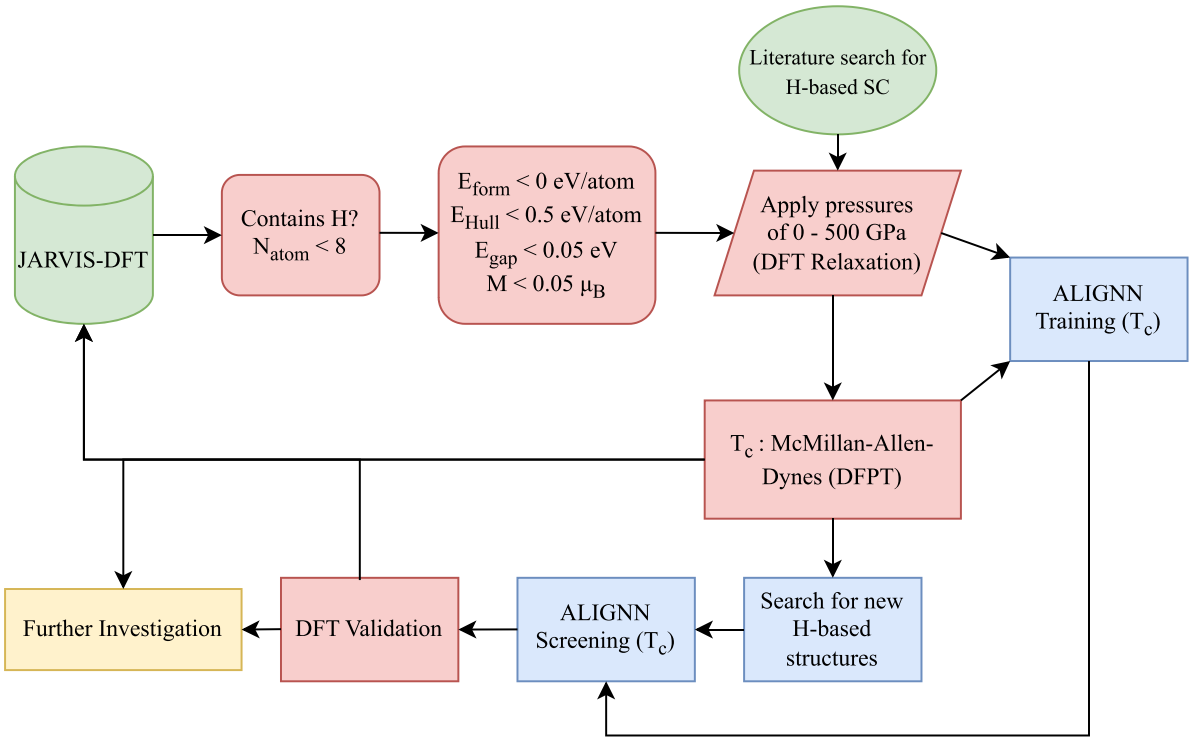

Figure 1. The full workflow for new high pressure hydride superconductors using JARVIS, DFT, and ALIGNN. This workflow involves screening structures from the JARVIS-DFT database (filtering structures that contain H, are thermodynamically stable, nonmagnetic and metallic) and adding structures from literature, then performing DFT calculations at applied pressures of 0 GPa to 500 GPa and computing T_c. The ALIGNN model is then trained on DFT data to predict T_c with lower cost, enabling enhanced screening of candidate superconductors.

et al [46] were included in our calculations. After identifying the potential hydride superconductors, we applied pressures of (0, 100, 200, 300 and 500) GPa to each material and relaxed the structure. To be consistent with our previous work [36], we performed our new DFT calculations (taking H-based structures from JARVIS, 3DSC [47] and Shipley *et al* [46]) with the same workflow and settings as in [36] (same PBEsol functional, pseudopotential, etc). After obtaining the final structure under pressure, we performed DFPT calculations to obtain the EPC and then used the McMillan–Allen–Dynes equation to estimate T_c. In addition to obtaining results for over 900 EPC calculations for T_c, we used our DFT data to train an ALIGNN model from scratch to predict T_c specifically for hydride structures under pressure. By obtaining a well-trained deep learning model for T_c, we can eliminate the need for expensive DFT calculations. Specifically, we can use ALIGNN as a screening tool to identify potential candidates outside of the training set, which we can verify with DFT (see figure 1).

Figure 2(a) depicts a full summary of the over 900 DFT calculations performed in this study. These calculations were performed for the same set of materials at (0, 100, 200, 300 and 500) GPa (500 GPa was selected as an example of extreme pressure). Figure 2(b) depicts a subset of figure 2(a), which includes the dynamically stable structures with a T_c above MgB_2 (39 K). Figure 2 also depicts reference experimental values for MgB_2 (39 K at ambient pressure), H_3S (203 K at 150 GPa) and LaH_{10} (250 K at 170 GPa) and the dotted line represents the boiling point of liquid N_2 at ambient pressure (77 K). Table 1 and table S1 (supplementary) display the contents of figure 2(b) in tabulated form, with additional details of where the initial structure originated before external pressure was applied in our DFT calculations. It is important to note that a few simulations were performed at intermediate pressures of 50 GPa and 250 GPa. These are included in the tables but omitted from figure 2 for visual purposes.

At this stage, it is important to discuss the successes and limitations of our theoretical calculations within the context of this work. The goal of this work is to use high-throughput DFT to (1) identify new high pressure hydride superconductors and (2) generate a substantial amount of data to train a reliable deep learning model to predict $T_{c]}$. Due to the high-throughput nature of these calculations and limitations to the theoretical framework, accuracy of the results may be impacted. There are two main sources of variation that can arise when calculating T_c. The first type of potential variation stems from the underlying DFT calculations, where calculation parameters such as the size of the q-point grid, choice of exchange–correlation functional, and choice of pseudopotential can impact the accuracy of the result. The second type of variation can occur from the theoretical framework used to estimate the T_c. Two types of approaches are normally used to estimate T_c from the underlying DFT results for EPC, either solving the Migdal–Eliashberg equations directly [68] or plugging the DFPT computed λ and ω_{log} into the McMillan–Allen–Dynes equation (as we did in this work, equation (5)). Both of

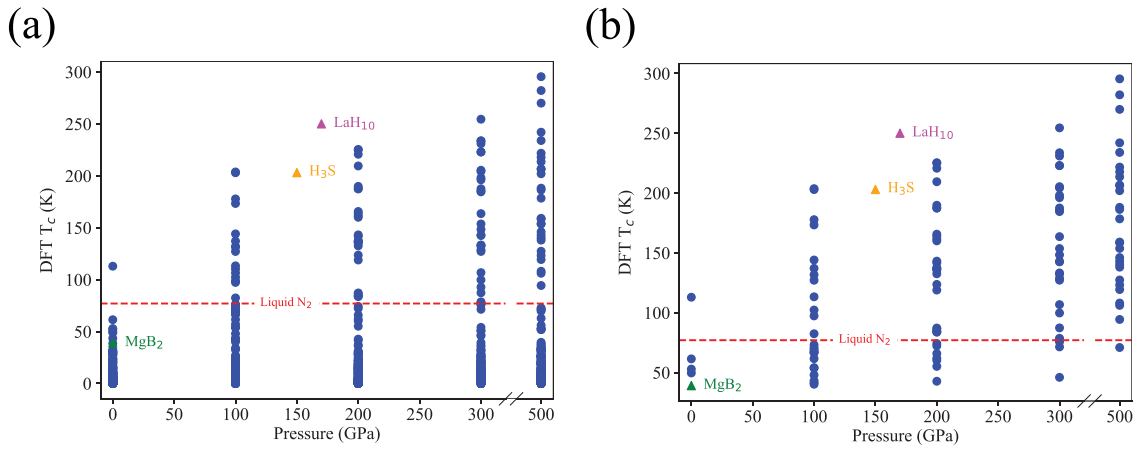

Figure 2. The DFT calculated T_c for the entire dataset at applied pressures of 0–500 GPa. For reference, experimental values for MgB_2, H_3S and LaH_{10} are given. The dotted line represents the boiling point of liquid N_2 at ambient pressure. (a) Depicts all of the DFT results while (b) depicts the dynamically stable structures with a T_c above MgB_2 (39 K).

these approaches utilize the empirical constant μ^*, which can introduce variations in the T_c results. In our previous work [36, 37], we found that varying μ^* from 0.03 to 0.18 can change the T_c by up to 40%, depending on the material, with the variation being more apparent for structures with a higher T_c. At larger EPC strengths, the variation between results obtained from the Migdal–Eliashberg equation and results obtained with the McMillan–Allen–Dynes equation begin to diverge, with the McMillan–Allen–Dynes approach underestimating the T_c [12, 46]. In fact, it was demonstrated that the variations arising from μ^* combined with the variations arising from the Migdal–Eliashberg vs. McMillan–Allen–Dynes approach can be up to 100 K for materials with large λ [46]. An example of this includes the underestimation of our DFT results vs. experimental values for the T_c of H_3S and LaH_{10} (we computed a max T_c of (160–187) K at (200–250) GPa for LaH_{10} and a T_c of 154 K at 200 GPa for H_3S). Nonetheless, important qualitative trends can be identified from these calculations and previously unidentified superconductors can be revealed. This can motivate future, more in-depth studies of selected superconducting materials, with the goal of achieving quantitative experimental accuracy for a smaller, focused subset of materials (rather than screening hundreds, or thousands of candidates with a less computationally costly approach). In order to achieve even higher accuracy and eliminate empirical biases, methods such as Superconducting-DFT and explicit calculations of μ^* with the random phase approximation can be performed [69–71].

It is clear from table 1 and table S1 that the majority of the structures that have a high T_c come from the dataset of Shipley *et al* [46], which were recalculated with our workflow. The main motivation for adding the previously discovered structures from Shipley *et al* [46] was to provide high-T_c data for our deep learning model. All of the structures shown in table 1 and table S1 are dynamically stable (possess positive phonon frequencies in the phonon density of states). In addition to dynamical stability of these hydride compounds, we assessed the thermodynamic stability of these structures by computing the formation energy of each material (added to table 1 and table S1). Each formation energy value is computed by subtracting the respective energies of the most energetically favorable elemental solids from the total energy of each hydride compound. For an accurate reference point, we calculated the total energy of each elemental solid under pressure (i.e. to calculate the formation energy of CaH_6 at 100 GPa, we subtracted off the energies of elemental Ca and H at 100 GPa). We observe that all of structures in table 1 and most structures in table S1 possess negative formation energy. By obtaining these formation energy results, we can easily compute the convex hull [72, 73] for selected compounds to observe how thermodynamically stable a given phase is with respect to other reported phases (see figure 4, discussed in the next paragraph). Due to a lack of formation energy data for hydride-based structures at pressures above 0 GPa in most DFT databases [28–30, 33, 35], constructing a convex hull for a given hydride superconductor has been a challenging task. We hope that by providing a large set of formation energy data for a wide variety of hydride structures under various amounts of pressure can allow others to create their own phase diagrams for compounds of interest (similar to those reported in figure 4). We acknowledge that it is possible for a material to exhibit strong EPC and also have a band gap, which would destroy the chances of the candidate material to be superconducting. To investigate this, we calculated the band gap of all of our materials using the PBEsol functional. We found that out of all of the nearly 900 computations we performed, only 41 structures possessed a band gap. Out of these 41 structures, 30 have a band gap and a finite T_c (almost all of these structures have a T_c less than MgB_2). These structures are listed in table S2.

In addition to adding previously discovered high-T_c to our dataset, our workflow revealed some hydride-based structures in JARVIS to be superconducting. These included MgH_2 (82 K at 100 GPa), AsHO (62 to 78 K at 200 to 300 GPa), ScH_3 (69 K at 100 GPa), and $KAlH_3$ (52 K at 0 GPa). Figure 3 depicts some of the selected materials from JARVIS that were revealed to be superconductors, in addition to some of the corresponding Eliashberg spectral functions. Along with this manuscript,

Table 1. A summary of DFT results, including structure (chemical formula), applied pressure, superconducting critical temperature (T_c), formation energy per atom (E_{form}) at each respective pressure, and the source of the original structure (JARVIS-DFT, 3DSC [47], or Shipley *et al* [46] dataset) for dynamically stable structures with the highest T_c. A continuation of this data is in table S1.

Structure	Pressure (GPa)	T_c (K)	E_{form} (eV/atom)	Source of structures
MgH_{12}	500	296	−10.9	Shipley *et al*
NaH_6	500	282	−14.4	Shipley *et al*
SrH_{10}	500	270	−13.2	Shipley *et al*
MgH_{12}	300	255	−6.7	Shipley *et al*
NaH_6	500	242	−12.0	Shipley *et al*
NaH_6	300	234	−4.0	Shipley *et al*
NaH_9	500	234	−11.3	Shipley *et al*
MgH_6	300	231	−7.5	Shipley *et al*
NaH_6	200	225	−5.4	Shipley *et al*
LiH_6	300	223	−6.4	Shipley *et al*
LiH_3	500	222	−10.9	Shipley *et al*
MgH_6	200	221	−3.4	Shipley *et al*
MgH_{10}	500	218	−11.1	Shipley *et al*
MgH_6	500	214	−11.8	Shipley *et al*
CaH_6	200	210	−5.9	Shipley *et al*
LiH_6	500	207	−10.5	Shipley *et al*
LiH_2	300	205	−7.2	Shipley *et al*
NaH_6	100	204	−1.9	Shipley *et al*
MgH_{10}	300	204	−6.9	Shipley *et al*
LiH_2	500	202	−11.2	Shipley *et al*
CaH_6	300	198	−8.3	Shipley *et al*
YH_9	300	196	−5.7	Shipley *et al*
LiH_2	200	190	−3.1	Shipley *et al*
LiH_6	200	188	−3.6	Shipley *et al*
LiH_3	300	188	−6.9	Shipley *et al*
YH_9	500	188	−9.5	Shipley *et al*
LaH_{10}	250	187	−4.7	JVASP-149370
KH_{10}	500	186	−15.6	Shipley *et al*
LaH_{10}	300	185	−5.7	JVASP-149370
NaH_9	300	185	−7.2	Shipley *et al*
CaH_{15}	500	179	−11.3	Shipley *et al*
CaH_6	100	178	−3.5	Shipley *et al*
MgH_{13}	100	173	−1.8	Shipley *et al*
H_3S	200	166	−5.7	JVASP-79487
MgH_{13}	200	164	−3.6	Shipley *et al*
CaH_{15}	300	164	−6.8	Shipley *et al*
LiH_3	200	162	−3.4	Shipley *et al*
LaH_{10}	200	160	−3.8	JVASP-149370
ScH_{12}	500	159	−11.2	Shipley *et al*
Na_2H_{11}	500	159	−12.1	Shipley *et al*
H_3S	250	154	−3.9	JVASP-79487
KH_{10}	300	154	−9.3	Shipley *et al*
Na_2H_{11}	500	154	−12.1	Shipley *et al*
CaH_{10}	300	148	−7.3	Shipley *et al*
ScH_{14}	500	146	−11.0	Shipley *et al*
ScH_6	100	144	−2.0	Shipley *et al*
YH_9	200	143	−3.9	Shipley *et al*
MgH_{13}	300	143	−6.7	Shipley *et al*
H_3S	300	143	−8.1	JVASP-79487
MgH_{14}	500	143	−10.8	Shipley *et al*
MgH_{12}	200	142	−3.6	Shipley *et al*
SrH_{15}	500	140	−12.1	Shipley *et al*
H_3S	500	138	−12.6	JVASP-79487
Na_2H_{11}	100	137	−1.9	Shipley *et al*
ScH_6	200	137	−3.9	Shipley *et al*
KH_{10}	200	137	−6.2	Shipley *et al*
CaH_{15}	200	136	−4.7	Shipley *et al*

all relaxed structures and full DFT calculations are available for further analysis and future utilization. Figure 4 depicts the computed phase diagrams for a few of the selected materials previously mentioned (using pymatgen [72, 73]), and we see that all of these structures lie on the respective convex hulls. Specifically in figures 4(a) and (b)), we demonstrate how the phase diagram of Mg-H changes with pressure. Although figure 4 depicts phase diagrams of binary systems, it is possible to construct these phase diagrams for ternary systems using pymatgen. Interestingly, when pressure is added to $KAlH_3$, a metal-to-semiconductor transition is induced from (100–500) GPa, quenching the superconducting properties (see table S2). To our knowledge, there have been no experimental reports of superconductivity in $KAlH_3$. In addition, we found that the SiH_4 (identified from 3DSC [47]) structure possesses a T_c of 71 and 72 for 200 GPa and 300 GPa respectively. It is important to note a different phase of ScH_3 has experimentally and theoretically been reported to be superconducting [18, 74] (with a $Fm\bar{3}m$ space group), but to our knowledge, the phase of ScH_3 (with a Cm space group) reported in our manuscript is identified as a superconductor with DFT for the first time. With regards to MgH_2, the phase reported in our work has a space group of $Fm-3m$, which differs from the more stable reported α-MgH_2 with a space group of Pnma [75]. α-MgH_2 (Pmna) undergoes a semiconductor to metal transition with higher pressures above 170 GPa, reaching a T_c up to 23 K [75]. Our reported MgH_2 structure ($Fm-3m$), which has been found to be metastable [76, 77], retains its metallicity from (0–500) GPa. To our knowledge, previous studies have not explored the superconducting capabilities of the metastable MgH_2 structure (Fm-3m) in detail (besides our previous work at 0 GPa [36]). SiH_4 (Silane) is a material that has been previously studied in great detail experimentally and through first-principles, with a variety of phases and pressure-induced phase transitions being reported [78–84]. In this work, we obtained a structure for SiH_4 (space group $P2_1/c$) from the 3DSC [47] database (which maps entries of the experimental Supercon database to Materials Project entries). This structure is found in both JARVIS and Materials Project (JVASP-5281, mp-23739). Although this $P2_1/c$ structure of Silane is insulating at zero pressure, our calculations indicate that the application of large pressure results in an insulator-to-metal transition, making superconductivity possible. Although it has been demonstrated theoretically that the thermodynamic stability of $P2_1/c$ can vary with pressure (i.e. the $C2/c$ phase is more stable than the $P2_1/c$ phase for pressure ranges below 400 GPa [78–81]), we find that $P2_1/c$ Silane possesses strong superconducting properties at (200–300) GPa. To our knowledge, DFT calculations for T_c at 200–300 GPa have not been carried out for $P2_1/c$ Silane prior to this study and the superconducting properties of $P2_1/c$ Silane have not been experimentally measured at these pressure values. Although our DFT results show dozens of potential high pressure hydride superconductor candidates, very few of our predicted structures have been experimentally realized. In fact, only a handful of materials from our dataset have been synthesized under pressure including

H_3S [8], LaH_{10} [9–11], CaH_6 [21], YH_9 [15], and polyhydra of Li-H (LiH_2 and LiH_6) [85, 86]. In addition, polyhydra of Sr-H [87] and Na-H [88] with different stoichiometry than our DFT calculations have been synthesized (SrH_2, NaH_3, NaH_7).

Our plethora of DFT data on hydride materials under extreme pressures allows us to analyze certain trends in the results. Figure 5 depicts a thorough analysis of DFT results for selected materials at various pressures. Figures 5(a) and (b) depicts how the phonon density of states changes when pressure is applied. From figures 5(a) and (b), we observe that the phonon DOS is blueshifted, and the application of high pressure can cause the first peak in the phonon DOS to change from negative to positive (triggering an unstable-to-stable transition), as is the case for LaH_{10}, H_3S, MgH_{12}, MgH_6, SrH_{10}. Figure 5(c) depicts how T_c changes with respect to applied pressure. It is clear from the figure that there is not necessarily a linear relationship between applied pressure and an increase in T_c. In fact, for certain materials such as LaH_{10} and H_3S, T_c increases up to some critical value of pressure and then decreases. This enforces the point that the relationship between superconductivity and applied pressure is entirely material dependent and cannot be described by simple models or equations (i.e. linear or polynomial fitting). Figure 5(d) depicts the Eliashberg function of LaH_{10} under different amounts of pressure. As expected, the area under the curve is directly proportional to the change in T_c at each pressure value (see equation (4)).

3.2. Machine learning results

The main outputs of this effort are (1) a high quality and diverse set of DFT data for hydride materials under pressure and (2) a well-developed deep learning model to predict T_c at a lower computational cost. The motivation to train a deep learning model to predict T_c for hydrides under various amounts of pressure stems from the fact that (1) each DFT calculation (for the structural relaxation under pressure and the DFPT calculation for the EPC) can be incredibly computationally expensive and (2) (as previously mentioned) there is no clear relationship between superconductivity and applied pressure, where it is entirely material dependent and cannot be described by simple models or equations. To address these two main concerns, we trained an ALIGNN model on our DFT data to predict T_c. A similar procedure was carried out in [36], where an ALIGNN model was trained on 1000 bulk materials. The T_c model of [36] achieved an MAE of 1.84 K, but the maximum T_c in this dataset was 33 K. In this work, we trained two separate ALIGNN models on different datasets (the datasets consisted of relaxed structures under pressure and T_c value). One model was trained on purely the hydride dataset (over 900 structures) and the other model was trained on the hydride dataset plus the dataset from [36] (900+1000 structures). Figures 6(a) and (b) depict the performance of both of these ALIGNN models on a random 10% test set for each dataset (the random test set is different for each model to minimize bias that arises from partitioning the data into train:test:validation). The MAE for the ALIGNN model trained on the hydride dataset is 12 K and the MAE for the ALIGNN model trained on the hydride plus the JARVIS dataset from [36] is 8 K. We find the mean absolute deviation (MAD) of the ALIGNN model trained on the hydride dataset to be 32 K while the MAD of the ALIGNN model trained on the hydride plus the JARVIS dataset from [36] is 19 K. The MAE values of these models are much larger due in part to the fact that the maximum T_c in the datasets are much larger (over 280 K). Figures 6(a) and (b) also demonstrates that the addition of data in these deep learning models can reduce the overall MAE. Ideally, if an additional, larger DFT dataset of high-pressure hydride materials was developed (that was consistent with our calculation method, in terms of DFT functional and pseudopotential), it could be used to augment the training set and could potentially improve performance. Figure 6 also depict some of the outlying ALIGNN predictions on the test set, which can be attributed to random model error. Figures S1–S3 depict additional details regarding both ALIGNN models. Figure S1 shows the training and validation MAE for T_c (K) as a function of the number of epochs. Figure S2 shows the performance of both ALIGNN models on the training and validation sets. Due to the fact that the MAE values could be somewhat inflated because a large portion of the data lies under 50 K, we plotted the data separately for T_c above 50 K and below 50 K and computed each MAE separately for both ALIGNN models (see figure S3). In our previous work in [36], we computed T_c by predicting λ and ω_{log} separately with ALIGNN and plugged both quantities into equation (5). We found that this only resulted in a 4% increase in accuracy. For this reason, we decided to directly predict T_c with ALIGNN. We hope that other researchers can utilize this ALIGNN model specifically fine tuned for high-pressure hydrides to screen the broader materials space for high-T_c hydride candidates prior to verifying with DFT (as shown in the workflow in figure 1). It is apparent from figure 6 and figure S3 that the errors in T_c prediction from ALIGNN are relatively high. Despite this limitation of the model, we believe that it can be useful for the classification of potential high-T_c superconductors. The goal is not to outright replace DFT predictions, but to aid in the screening of potential candidates prior to verification by DFT.

Motivated by this idea, we went on to quantify the classification of superconductors for both ALIGNN models. Our first classification metric is whether or not ALIGNN can predict if a structure is superconducting or not ($T_c > 0.1$ K). We define a false positive as when ALIGNN predicts a $T_c > 0.1$ K while DFT predicts a $T_c < 0.1$ K, and a false negative as when ALIGNN predicts a $T_c < 0.1$ K while DFT predicts a $T_c > 0.1$ K. In the few cases where ALIGNN erroneously predicts a negative T_c value with magnitude larger than 0.1 K (unphysical result), these entries are counted as both false positive and false negative. For the ALIGNN model trained only on the hydride DFT dataset, the test set contained 95 materials. Out of these 95 structures, there were 8 false positives and 4 false

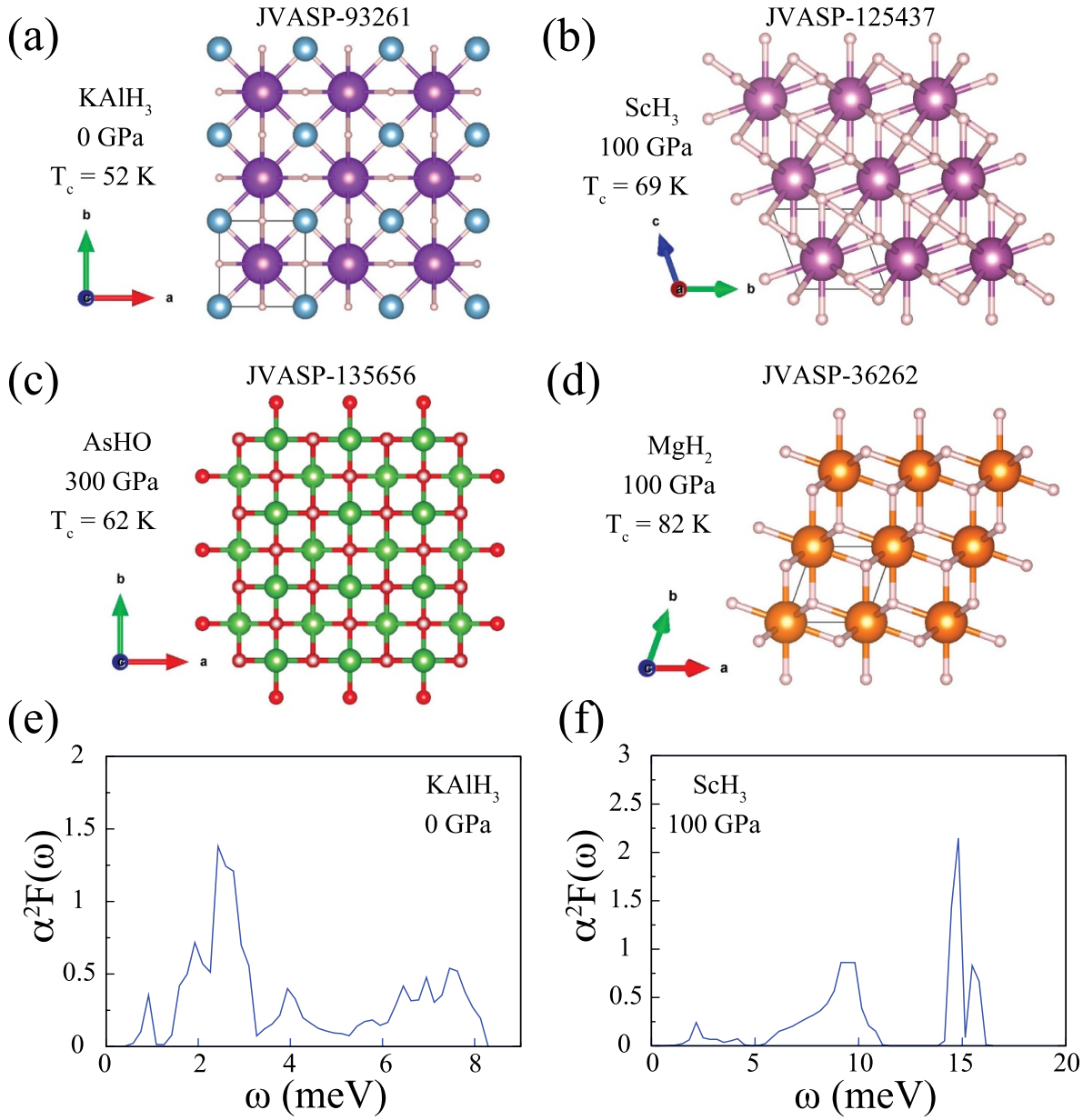

Figure 3. (a)–(d) Selected structures from JARVIS that were revealed to be superconducting. (e) and (f) depict selected Eliashberg spectral functions.

negatives (with 3 entries being counted for both categories as unphysical results). For the ALIGNN model trained on the hydride+JARVIS-SC DFT dataset, the test set contained 199 materials. Out of these 199 structures, there were 31 false positives and 9 false negatives (with 8 entries being counted for both categories as unphysical results). Our second classification metric is whether or not ALIGNN can predict if a structure is a high temperature superconducting or not ($T_c >$ 50 K). We define a false positive as when ALIGNN predicts a $T_c > 50$ K while DFT predicts a $T_c < 50$ K, and a false negative as when ALIGNN predicts a $T_c < 50$ K while DFT predicts a $T_c > 50$ K. For the ALIGNN model trained only on the hydride DFT dataset, we identified 0 false positives and 5 false negatives (16 materials in the test set have a DFT $T_c > 50$ K). For the ALIGNN model trained on the hydride DFT+JARVIS-SC dataset, we identified 2 false positives and 3 false negatives (17 materials in the test set have a DFT $T_c >$ 50 K). These metrics demonstrate that although the prediction errors are relatively high in figure 6, these ALIGNN models can be useful in classification tasks for T_c, especially for high-T_c materials. By adding the ALIGNN model to the screening workflow for high pressure hydrides, the number of DFT calculations for superconducting properties can be significantly reduced.

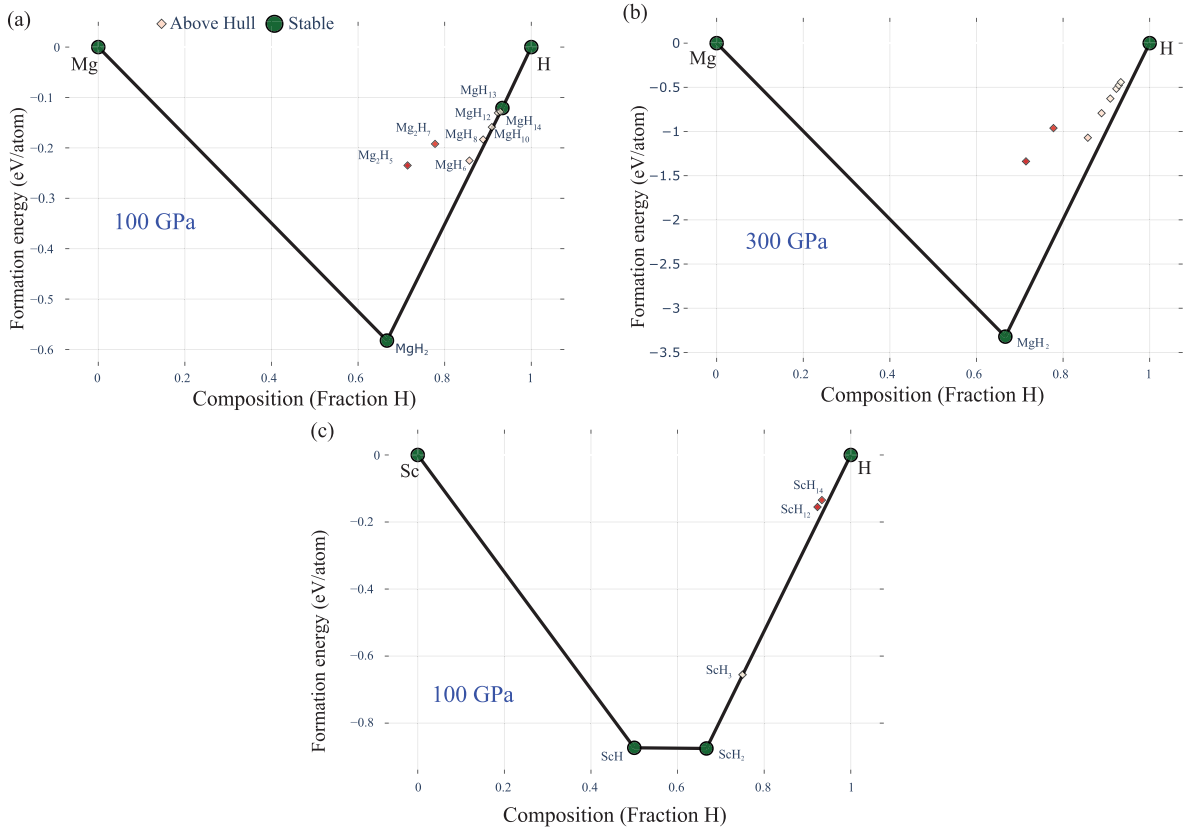

Figure 4. Phase diagrams for some of the selected compounds reported in this work: Mg-H at (a) 100 GPa, (b) 300 GPa and (c) Sc-H at 100 GPa. Formation energy data and code to produce these phase diagrams are available with this manuscript for users to construct their own phase diagrams and add their own DFT data.

In the previous paragraph, we discussed how ALIGNN trained on DFPT calculations for T_c can accelerate the prediction of T_c. Although this is an extremely useful tool, the ALIGNN model for T_c assumes that the relaxed structure under pressure is provided. In addition, if only the relaxed structure with zero applied pressure is provided, it will usually yield a low value for T_c, which may result in the structure being disregarded as a candidate for a potential superconductor. Although the DFPT computation of the EPC is much more computationally expensive than the DFT structural relaxation under pressure, this relaxation can still be a time consuming endeavor. One way to circumvent this and eliminate the cost of DFT structural relaxation is to use a universal machine learned force field (such as ALIGNN-Force-Field [89], M3GNET/MatGL [90] and GemNet-OC [91]) to relax the structure prior to ALIGNN prediction of T_c. In this work, we used the pretrained ALIGNN-Force-Field (FF) [89] (universal machine learning force field) with the Atomic Simulation Environment (ASE) [92] FIRE [93] optimizer to relax the hydride-based structures under various amounts of pressure. Pressure was induced in each structure by compressing the volume and allowing the atoms and bond angles to relax under that constant volume value. ALIGNN-FF was developed to handle chemically and structurally diverse crystalline systems, where the entirety of the previous JARVIS-DFT dataset was used as training, (which contains 4 million energy-force entries for 89 elements of the periodic table, 307113 which are used for training) [89]. It is important to note that none of the high pressure calculations in this work were used to train this ALIGNN-FF.

The ALIGNN-FF+ALIGNN-Superconducting (ALIGNN-SC) results along with DFT are depicted in figure 7. Specifically, we tested both ALIGNN-SC models (one model trained on the hydride dataset, the other trained on the hydride dataset plus the dataset from [36]) on the relaxed structures from ALIGNN-FF under pressure. As seen in figure 7 there is good qualitative agreement between DFT and ALIGNN and meaningful trends can be extracted from the purely machine learning results. Regardless if the structures under pressure were in the training, validation or test sets during ALIGNN training, there were no DFT calculations performed at intermediate pressure values such as (50, 150, 250, and 400) GPa (with the exception of DFT calculations for LaH_{10} at 50 GPa and 250 GPa). From figure 7, we observe that ALIGNN-FF+ALIGNN-SC can interpolate reasonably well to intermediate values of pressure that were not part

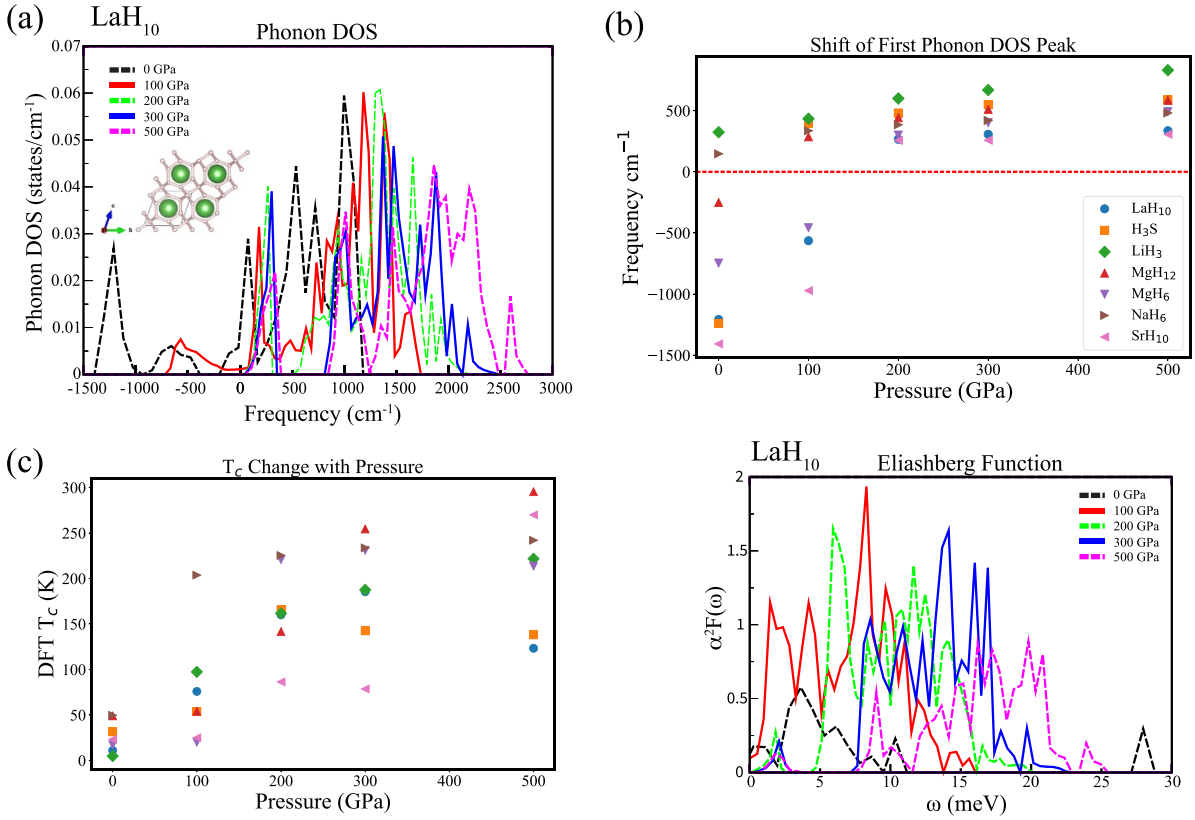

Figure 5. An analysis of DFT results for select materials at various pressures: (a) the overlapping phonon density of states for LaH_{10}, (b) the shift of the first peak in the phonon density of states with respect to applied pressure (points below the red dotted line indicate dynamical instability), (c) T_c as a function of pressure, and (d) the overlapping Eliashberg function for LaH_{10}.

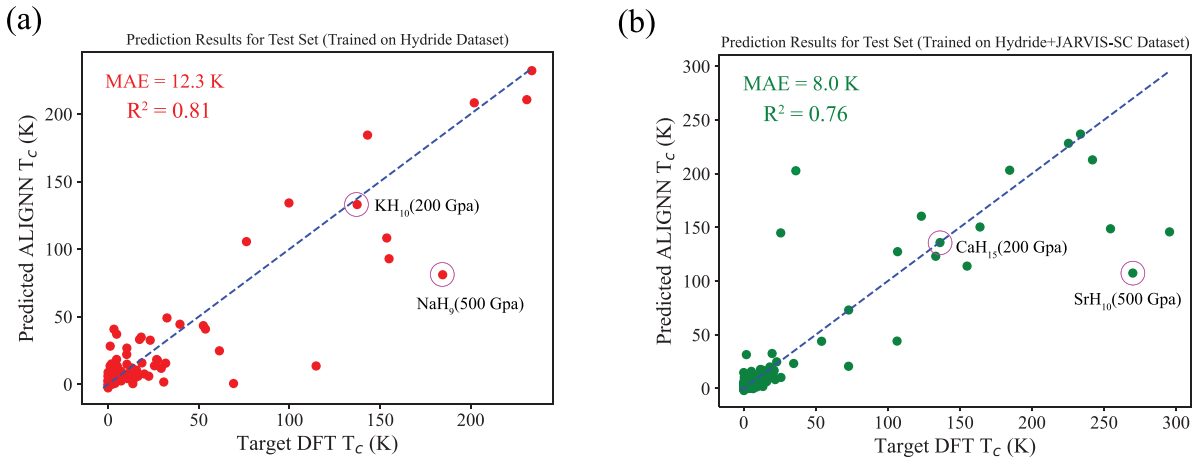

Figure 6. The performance of two separate ALIGNN models for T_c on the test set (random 10% of each dataset). (a) Is the ALIGNN model trained on the hydride dataset and (b) is the ALIGNN model trained on dataset which contains the hydride dataset and the JARVIS superconductor dataset from [36]. The mean absolute error (MAE) and R^2 value are depicted in the inset for each model and the dotted line is shown for reference.

of the original dataset, which allows us to reliably compute T_c for a larger number of pressure values and obtain a finer map of how superconducting properties change with pressure, avoiding unnecessary DFT calculations. This can also give us an indication of where the critical pressure value lies (pressure value which results in the maximum value of T_c) and is a cheaper alternative to brute-force DFT calculations at several values of pressure.

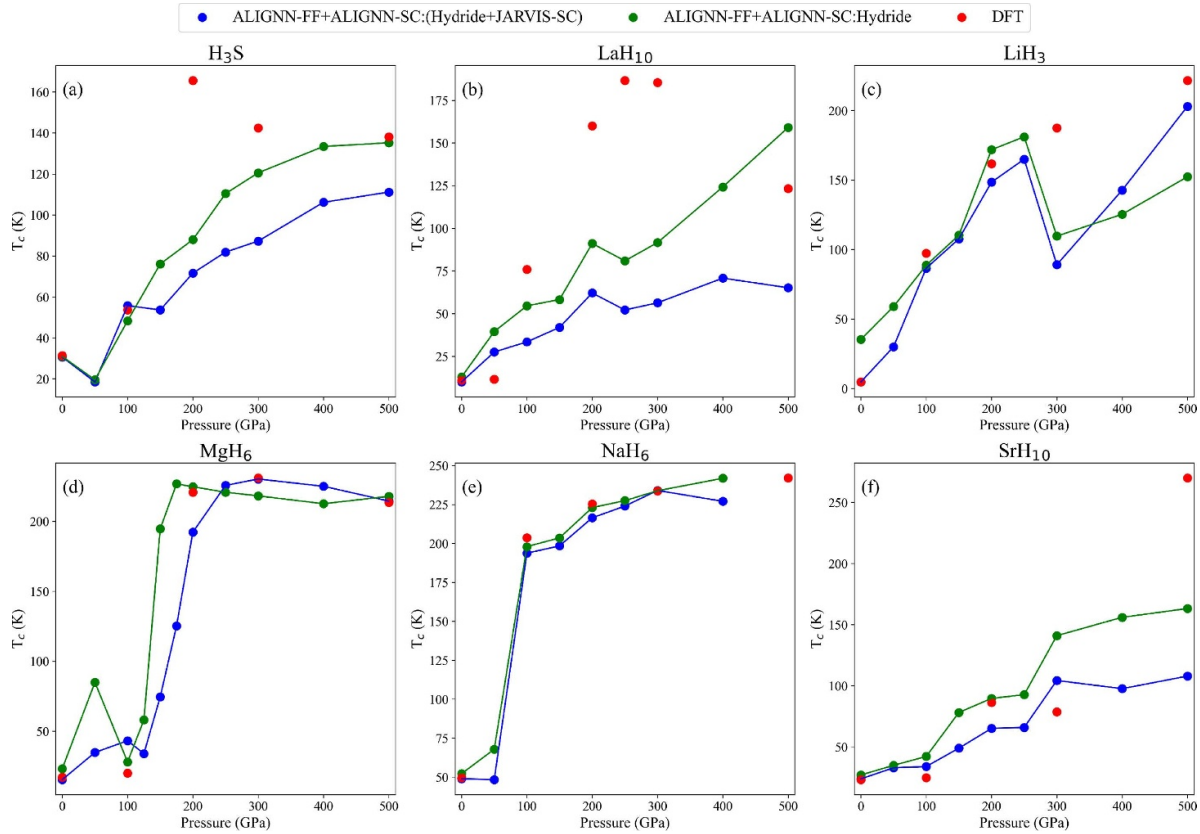

Figure 7. T_c results from DFT (red) along with T_c results from ALIGNN-SC (green for the ALIGNN-SC model from figure 6(a), blue for the ALIGNN-SC model from figure 6(b)) that were obtained using ALIGNN-FF relaxed structures at various pressures.

4. Conclusion

In this work, we have computed the superconducting properties of over 900 hydride-based materials with DFT under ultrahigh pressures (0 GPa–500 GPa). In addition to adding previously discovered high-T_c materials to our dataset and revealing previously undiscovered superconductors from the JARVIS database, we trained an ALIGNN model to predict T_c with reasonable accuracy. We took this one step further and coupled our ALIGNN model for T_c to the ALIGNN-FF universal force-field to relax the structures with minimal computational cost prior to ALIGNN T_c prediction. By utilizing these deep learning tools for structural relaxation under pressure and property prediction, enhanced materials screening can be enabled and the number of DFT calculations (and overall computational cost) can be significantly reduced, allowing for a broader search of material space for new high pressure hydride superconductors.

5. Future perspectives

Data-driven discovery and design of novel high pressure hydride superconductors is a rapidly changing and exciting field of physics and materials science. Coupling high-throughput first-principles (DFT) simulations with machine learning techniques can allow for a more broad search of the high pressure superconductor landscape. Most importantly, these machine learning models can be used as effective classification tools to screen candidate systems prior to more in-depth theoretical calculations and eventual experimental synthesis. In order for this field to progress, the amount of publicly available and unique high-throughput DFT data must increase, which can be one of the most effective routes to improving the accuracy of these machine learning models. Collaborative efforts such as the JARVIS-Leaderboard [94], which allows users to view and contribute materials science data and machine learning models and assess accuracy/fidelity (while enhancing transparency and reproducibility) and OPTIMADE (Open Databases Integration for Materials Design) [95], which allows for the easy and interoperable access to information across different materials databases in a uniform format can play an important role in granting access to relevant data and metrics with the goal of advancing the field of materials design.

Data availability statement

Software packages mentioned in the article can be found at https://github.com/usnistgov/jarvis. All code and data specific to this work can be found at https://doi.org/10.6084/m9.figshare.25270012 and model benchmarks can be found at https://pages.nist.gov/jarvis_leaderboard/.

Notes

Please note that the use of commercial software (VASP) does not imply recommendation by the National Institute of Standards and Technology.

Conflicts of interest

The authors declare no competing interests.

Acknowledgments

All authors thank the National Institute of Standards and Technology for funding, computational, and data-management resources.

ORCID iDs

Daniel Wines https://orcid.org/0000-0003-3855-3754
Kamal Choudhary https://orcid.org/0000-0001-9737-8074

References

[1] Bardeen J, Cooper L N and Schrieffer J R 1957 Microscopic theory of superconductivity *Phys. Rev.* **106** 162–4
[2] Bardeen J, Cooper L N and Schrieffer J R 1957 Theory of superconductivity *Phys. Rev.* **108** 1175–204
[3] Nagamatsu J, Nakagawa N, Muranaka T, Zenitani Y and Akimitsu J 2001 Superconductivity at 39 K in magnesium diboride *Nature* **410** 63–64
[4] Eremets M I *et al* 2022 High-temperature superconductivity in hydrides: experimental evidence and details *J. Supercond. Nov. Magn.* **35** 965–77
[5] Zhang S, Zhang M and Liu H 2021 Superconductive hydrogen-rich compounds under high pressure *Appl. Phys.* A **127** 684
[6] Ashcroft N W 1968 Metallic hydrogen: a high-temperature superconductor? *Phys. Rev. Lett.* **21** 1748–9
[7] Ashcroft N W 2004 Hydrogen dominant metallic alloys: high temperature superconductors? *Phys. Rev. Lett.* **92** 187002
[8] Drozdov A P, Eremets M I, Troyan I A, Ksenofontov V and Shylin S I 2015 Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system *Nature* **525** 73–76
[9] Drozdov A P *et al* 2019 Superconductivity at 250 K in lanthanum hydride under high pressures *Nature* **569** 528–31
[10] Geballe Z M, Liu H, Mishra A K, Ahart M, Somayazulu M, Meng Y, Baldini M and Hemley R J 2018 Synthesis and stability of lanthanum superhydrides *Angew. Chem., Int. Ed.* **57** 688–92
[11] Somayazulu M, Ahart M, Mishra A K, Geballe Z M, Baldini M, Meng Y, Struzhkin V V and Hemley R J 2019 Evidence for superconductivity above 260 K in lanthanum superhydride at megabar pressures *Phys. Rev. Lett.* **122** 027001
[12] Liu H, Naumov I I, Hoffmann R, Ashcroft N W and Hemley R J 2017 Potential high-T_C superconducting lanthanum and yttrium hydrides at high pressure *Proc. Natl Acad. Sci.* **114** 6990–5
[13] Liu H, Naumov I I, Geballe Z M, Somayazulu M, Tse J S and Hemley R J 2018 Dynamics and superconductivity in compressed lanthanum superhydride *Phys. Rev.* B **98** 100102
[14] Kruglov I A *et al* 2020 Superconductivity of LaH_{10} and LaH_{16} polyhydrides *Phys. Rev.* B **101** 024508
[15] Kong P *et al* 2021 Superconductivity up to 243 K in the yttrium-hydrogen system under high pressure *Nat. Commun.* **12** 5075
[16] Semenok D V, Kvashnin A G, Ivanova A G, Svitlyk V, Fominski V Y, Sadakov A V, Sobolevskiy O A, Pudalov V M, Troyan I A and Oganov A R 2020 Superconductivity at 161 K in thorium hydride ThH_10: synthesis and properties *Mater. Today* **33** 36–44
[17] Shao M, Chen W, Zhang K, Huang X and Cui T 2021 High-pressure synthesis of superconducting clathratelike YH_4 *Phys. Rev.* B **104** 174509
[18] Shao M, Chen S, Chen W, Zhang K, Huang X and Cui T 2021 Superconducting ScH_3 and LuH_3 at megabar pressures *Inorg. Chem.* **60** 15330–5
[19] Zhang C, He X, Li Z, Zhang S, Feng S, Wang X, Yu R and Jin C 2022 Superconductivity in zirconium polyhydrides with Tc above 70 K *Sci. Bull.* **67** 907–9
[20] Hong F *et al* 2022 Possible superconductivity at 70 K in tin hydride SnH_x under high pressure *Mater. Today Phys.* **22** 100596
[21] Li Z *et al* 2022 Superconductivity above 200 K discovered in superhydrides of calcium *Nat. Commun.* **13** 2863
[22] Denchfield A, Park H and Hemley R J 2024 Electronic structure of nitrogen-doped lutetium hydrides *Phys. Rev. Materials* **8** L021801
[23] Ge Y, Zhang F and Hemley R J 2021 Room-temperature superconductivity in boron-and nitrogen-doped lanthanum superhydride *Phys. Rev.* B **104** 214505
[24] Ge Y, Zhang F, Dias R P, Hemley R J and Yao Y 2020 Hole-doped room-temperature superconductivity in $H_3S_{1-x}Z_x$ (Z = C, Si) *Mater. Today Phys.* **15** 100330
[25] Fan F, Papaconstantopoulos D, Mehl M and Klein B 2016 High-temperature superconductivity at high pressures for $H_3Si_xP_{1-x}$, $H_3P_xS_{1-x}$ and $H_3C_{1-x}S_{1-x}$ *J. Phys. Chem. Solids* **99** 105–10
[26] Snider E, Dasenbrock-Gammon N, McBride R, Debessai M, Vindana H, Vencatasamy K, Lawler K V, Salamat A and Dias R P 2020 RETRACTED ARTICLE: room-temperature superconductivity in a carbonaceous sulfur hydride *Nature* **586** 373–7
[27] Dasenbrock-Gammon N *et al* 2023 RETRACTED ARTICLE: evidence of near-ambient superconductivity in a N-doped lutetium hydride *Nature* **615** 244–50
[28] Jain A *et al* 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation *APL Mater.* **1** 011002
[29] Saal J E, Kirklin S, Aykol M, Meredig B and Wolverton C 2013 Materials design and discovery with high-throughput density functional theory: the Open Quantum Materials Database (OQMD) *JOM* **65** 1501–9
[30] Kirklin S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Rühl S and Wolverton C 2015 The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies *npj Comput. Mater.* **1** 15010
[31] Haastrup S *et al* 2018 The computational 2D materials database: high-throughput modeling and discovery of atomically thin crystals *2D Mater.* **5** 042002
[32] Gjerding M N *et al* 2021 Recent progress of the computational 2D materials database (C2DB) *2D Mater.* **8** 044002
[33] Choudhary K *et al* 2020 The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design *npj Comput. Mater.* **6** 1–13
[34] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 Python materials genomics (pymatgen): a

robust, open- source python library for materials analysis *Comput. Mater. Sci.* **68** 314–9
[35] Wines D, Gurunathan R, Garrity K F, DeCost B, Biacchi A J, Tavazza F and Choudhary K 2023 Recent progress in the JARVIS infrastructure for next-generation data-driven materials design *Appl. Phys. Rev.* **10** 041302
[36] Choudhary K and Garrity K 2022 Designing high-TC superconductors with BCS-inspired screening, density functional theory and deep-learning *npj Comput. Mater.* **8** 244
[37] Wines D, Choudhary K, Biacchi A J, Garrity K F and Tavazza F 2023 High-throughput DFT-based discovery of next generation two-dimensional (2D) superconductors *Nano Lett.* **23** 969–78
[38] García-Nieto P J, García-Gonzalo E and Paredes-Sánchez J 2021 Prediction of the critical temperature of a superconductor by using the WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques *Neural Comput. Appl.* **33** 17131–45
[39] Stanev V, Oses C, Kusne A G, Rodriguez E, Paglione J, Curtarolo S and Takeuchi I 2018 Machine learning modeling of superconducting critical temperature *npj Comput. Mater.* **4** 29
[40] Zhang J, Zhu Z, Xiang X D, Zhang K, Huang S, Zhong C, Qiu H-J, Hu K and Lin X 2022 Machine learning prediction of superconducting critical temperature through the structural descriptor *J. Phys. Chem.* C **126** 8922–7
[41] Meredig B *et al* 2018 Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery *Mol. Syst. Des. Eng.* **3** 819–25
[42] Roter B and Dordevic S 2020 Predicting new superconductors and their critical temperatures using machine learning *Physica* C **575** 1353689
[43] Menon D and Ranganathan R 2022 A generative approach to materials discovery, design and optimization *ACS Omega* **7** 25958–73
[44] Seegmiller C C, Baird S G, Sayeed H M and Sparks T D 2023 Discovering chemically novel, high-temperature superconductors *Comput. Mater. Sci.* **228** 112358
[45] Wines D, Xie T and Choudhary K 2023 Inverse design of next-generation superconductors using data-driven deep generative models *J. Phys. Chem. Lett.* **14** 6630–8
[46] Shipley A M, Hutcheon M J, Needs R J and Pickard C J 2021 High-throughput discovery of high-temperature conventional superconductors *Phys. Rev.* B **104** 054501
[47] Sommer T, Willa R, Schmalian J and Friederich P 2023 3DSC - a dataset of superconductors including crystal structures *Sci. Data* **10** 816
[48] Cerqueira T F, Sanna A and Marques M A 2024 Sampling the materials space for conventional superconducting compounds *Adv. Mater.* **36** 2307085
[49] Saha S, Di Cataldo S, Giannessi F, Cucciari A, von der Linden W and Boeri L 2023 Mapping superconductivity in high-pressure hydrides: the Superhydra project *Phys. Rev. Mater.* **7** 054806
[50] Belli F, Novoa T, Contreras-García J and Errea I 2021 Strong correlation between electronic bonding network and critical temperature in hydrogen-based superconductors *Nat. Commun.* **12** 5381
[51] Choudhary K *et al* 2022 Recent advances and applications of deep learning methods in materials science *npj Comput. Mater.* **8** 59
[52] Choudhary K 2024 AtomGPT: atomistic generative pre-trained transformer for forward and inverse materials design (arXiv:2405.03680)
[53] Burdine C and Blair E 2023 Discovery of novel superconducting materials with deep learning *2023 IEEE Int. Conf. on Quantum Computing and Engineering* (*QCE*) pp 1335–41
[54] Baroni S, Giannozzi P and Testa A 1987 Green's-function approach to linear response in solids *Phys. Rev. Lett.* **58** 1861
[55] Gonze X 1995 Perturbation expansion of variational principles at arbitrary order *Phys. Rev.* A **52** 1086
[56] Wierzbowska M, de Gironcoli S and Giannozzi P 2005 Origins of low-and high-pressure discontinuities of T_c in niobium (arXiv:cond-mat/0504077)
[57] Giannozzi P *et al* 2020 Quantum ESPRESSO toward the exascale *J. Chem. Phys.* **152** 154105
[58] Garrity K F, Bennett J W, Rabe K M and Vanderbilt D 2014 Pseudopotentials for high-throughput DFT calculations *Comput. Mater. Sci.* **81** 446–52
[59] Perdew J P, Ruzsinszky A, Csonka G I, Vydrov O A, Scuseria G E, Constantin L A, Zhou X and Burke K 2008 Restoring the density-gradient expansion for exchange in solids and surfaces *Phys. Rev. Lett.* **100** 136406
[60] Topsakal M and Wentzcovitch R 2014 Accurate projected augmented wave (PAW) datasets for rare-earth elements (RE = La–Lu) *Comput. Mater. Sci.* **95** 263–70
[61] Choudhary K and Tavazza F 2019 Convergence and machine learning predictions of Monkhorst-Pack k-points and plane-wave cut-off in high-throughput DFT calculations *Comput. Mater. Sci.* **161** 300–8
[62] Marsiglio F 2020 Eliashberg theory: a short review *Ann. Phys., NY* **417** 168102
[63] McMillan W 1968 Transition temperature of strong-coupled superconductors *Phys. Rev.* **167** 331
[64] Choudhary K and DeCost B 2021 Atomistic line graph neural network for improved materials property predictions *npj Comput. Mater.* **7** 1–8
[65] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library *Advances in Neural Information Processing Systems* vol 32 pp 8026–37
[66] Wang M *et al* 2019 Deep graph library: towards efficient and scalable deep learning on graphs (arXiv:1909.01315)
[67] Klimeš J, Bowler D R and Michaelides A 2009 Chemical accuracy for the van der Waals density functional *J. Phys.: Condens. Matter* **22** 022201
[68] Eliashberg G M 1960 Interactions between electrons and lattice vibrations in a superconductor *Sov. Phys. -JETP* **11** 3 (Engl. transl.)
[69] Sanna A, Pellegrini C, Liebhaber E, Rossnagel K, Franke K J and Gross E K U 2022 Real-space anisotropy of the superconducting gap in the charge-density wave material 2H-$NbSe_2$ *npj Quantum Mater.* **7** 6
[70] Marques M A L, Lüders M, Lathiotakis N N, Profeta G, Floris A, Fast L, Continenza A, Gross E K U and Massidda S 2005 *Ab initio* theory of superconductivity. II. Application to elemental metals *Phys. Rev.* B **72** 024546
[71] Lüders M, Marques M A L, Lathiotakis N N, Floris A, Profeta G, Fast L, Continenza A, Massidda S and Gross E K U 2005 *Ab initio* theory of superconductivity. I. Density functional formalism and approximate functionals *Phys. Rev.* B **72** 024545
[72] Ong S P, Jain A, Hautier G, Kang B and Ceder G 2010 Thermal stabilities of delithiated olivine MPO4 (M = Fe, Mn) cathodes investigated using first principles calculations *Electrochem. Commun.* **12** 427–30
[73] Ong S P, Wang L, Kang B and Ceder G 2008 Li–Fe–P–O_2 phase diagram from first principles calculations *Chem. Mater.* **20** 1798–807
[74] Kim D Y, Scheicher R H, kwang Mao H, Kang T W and Ahuja R 2010 General trend for pressurized

superconducting hydrogen-dense materials *Proc. Natl Acad. Sci.* **107** 2793–6
[75] Lonie D C, Hooper J, Altintas B and Zurek E 2013 Metallization of magnesium polyhydrides under pressure *Phys. Rev.* B **87** 054107
[76] El-Eskandarany M S, Banyan M and Al-Ajmi F 2018 Discovering a new MgH_2 metastable phase *RSC Adv.* **8** 32003–8
[77] Boonchot C, Tsuppayakorn-Aek P, Pinsook U and Bovornratanaraks T 2021 Stability and electronic structure of magnesium hydride and magnesium deuteride under high pressure *J. Phys.: Conf. Ser.* **2145** 012026
[78] Pickard C J and Needs R J 2006 High-pressure phases of silane *Phys. Rev. Lett.* **97** 045504
[79] Strobel T A, Goncharov A F, Seagle C T, Liu Z, Somayazulu M, Struzhkin V V and Hemley R J 2011 High-pressure study of silane to 150 GPa *Phys. Rev.* B **83** 144102
[80] Chen X-J, Wang J-L, Struzhkin V V, Mao H-K, Hemley R J and Lin H-Q 2008 Superconducting behavior in compressed solid SiH_4 with a layered structure *Phys. Rev. Lett.* **101** 077002
[81] Zhang H, Jin X, Lv Y, Zhuang Q, Liu Y, Lv Q, Bao K, Li D, Liu B and Cui T 2015 High-temperature superconductivity in compressed solid silane *Sci. Rep.* **5** 8845
[82] Kim D Y, Scheicher R H, Lebègue S, Prasongkit J, Arnaud B, Alouani M and Ahuja R 2008 Crystal structure of the pressure-induced metallic phase of SiH_4 from *ab initio* theory *Proc. Natl Acad. Sci.* **105** 16454–9
[83] Strobel T A, Somayazulu M and Hemley R J 2009 Novel pressure-induced interactions in silane-hydrogen *Phys. Rev. Lett.* **103** 065701
[84] Degtyareva O, Canales M M, Bergara A, Chen X-J, Song Y, Struzhkin V V, Mao H-K and Hemley R J 2007 Crystal structure of SiH_4 at high pressure *Phys. Rev.* B **76** 064123
[85] Howie R T, Narygina O, Guillaume C L, Evans S and Gregoryanz E 2012 High-pressure synthesis of lithium hydride *Phys. Rev.* B **86** 064108
[86] Pépin C, Loubeyre P, Occelli F and Dumas P 2015 Synthesis of lithium polyhydrides above 130 GPa at 300 K *Proc. Natl Acad. Sci.* **112** 7673–6
[87] Smith J S, Desgreniers S, Klug D D and Tse J S 2009 High-density strontium hydride: an experimental and theoretical study *Solid State Commun.* **149** 830–4
[88] Struzhkin V V, Kim D Y, Stavrou E, Muramatsu T, Mao H-K, Pickard C J, Needs R J, Prakapenka V B and Goncharov A F 2016 Synthesis of sodium polyhydrides at high pressures *Nat. Commun.* **7** 12267
[89] Choudhary K, DeCost B, Major L, Butler K, Thiyagalingam J and Tavazza F 2023 Unified graph neural network force-field for the periodic table: solid state applications *Digit. Discovery* **2** 346–55
[90] Chen C and Ong S P 2022 A universal graph deep learning interatomic potential for the periodic table *Nat. Comput. Sci.* **2** 718–28
[91] Gasteiger J, Shuaibi M, Sriram A, Günnemann S, Ulissi Z, Zitnick C L and Das A 2022 GemNet-OC: developing graph neural networks for large and diverse molecular simulation datasets (arXiv:2204.02782)
[92] Larsen A H *et al* 2017 The atomic simulation environment—a Python library for working with atoms *J. Phys.: Condens. Matter* **29** 273002
[93] Bitzek E, Koskinen P, Gähler F, Moseler M and Gumbsch P 2006 Structural relaxation made simple *Phys. Rev. Lett.* **97** 170201
[94] Choudhary K *et al* 2024 JARVIS-Leaderboard: a large scale benchmark of materials design methods *npj Comput. Mater.* **10** 93
[95] Evans M *et al* 2024 Developments and applications of the OPTIMADE API for materials discovery, design and data exchange *Digit. Discovery* (https://doi.org/10.1039/D4DD00039K)