Materials Today Physics 43 (2024) 101384

Contents lists available at ScienceDirect

Materials Today Physics

journal homepage: www.journals.elsevier.com/materials-today-physics

Unveiling future superconductors through machine learning

Zihao Bai [a,b,c,d], Mangladeep Bhullar [a], Akinwumi Akinpelu [a], Yansun Yao [a,*]

[a] Department of Physics and Engineering Physics, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 5E2, Canada
[b] Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin–Madison, Madison, WI, 53706, United States
[c] Center for Theoretical Physics and School of Physics, Jilin University, Changchun, 130012, China
[d] Department of Materials Science and Engineering, University of Wisconsin–Madison, Madison, WI, 53706, United States

ARTICLE INFO

Keywords:
Superconductivity
Machine learning
Superhydrides
High-throughput computing
First principles calculations

ABSTRACT

The recent discovery of superconductivity above 200 K in hydrides of sulfur and lanthanum under high pressure marked a significant advance toward the realization of room-temperature superconductivity. While binary hydrides have almost been completely studied theoretically, experimental evidence suggests that the next breakthrough in finding high-temperature and low-pressure limits is likely connected with ternary and higher hydrides. Unlike the traditional synthesis-test-repeat approach, experimental discovery of superhydrides under high pressure often follows *prior* theoretical predictions. In this Minireview, we describe how various artificial intelligence schemes enable and enrich each stage of the discovery cycle of superhydrides and new developments made toward predicting ternary and higher hydrides. As a new enabling tool, machine learning-informed material simulation is still making its way into this field but is already playing an essential role in augmenting the prediction of new superhydrides through automated and iterative machine-learning processes. The review concludes with a perspective on outstanding challenges and possible future developments in the field.

1. Introduction

Superconductivity under ambient conditions has been one of the most sought-after discoveries in physical sciences [1]. Materials exhibiting this quantum phenomenon have zero *DC* electrical resistivity below a critical temperature T_c, *i.e.*, in the superconducting state. Superconducting materials are also perfect diamagnets ($X\nu = -1$) in the Meissner state, where they expulse magnetic fields when subjected to a temperature below T_c. These exceptional properties find superconductors practical applications in industrial and scientific sectors, *i.e.*, powerful electromagnets, lossless electrical transmission, and sensitive magnetometers. However, the requirement for low temperatures is a significant challenge for the widespread application of superconductors. Currently, the highest T_c discovered in superconductors at atmospheric pressure is only about 138 K, held by cuprate ceramics of mercury (Hg), barium (Ba) and calcium (Ca) [2]. On the other hand, achieving superconductivity with even higher T_c is possible under extreme pressures. One such group of candidates is hydrogen-rich superhydrides, especially lanthanum hydride (LaH_{10}), which has a measured T_c of about 250 K under 170 GPa pressure [3,4]. Hydrogen sulfide (H_3S) also exhibits superconductivity at 203 K under 150 GPa pressure [5,6], and clathrate hydride CaH_6 superconducts at a T_c of 215 K under 172 GPa pressure [7]. The superconducting hydrides were not accidently discovered; the synthesis followed theoretical predictions [8–10]. This signifies a paradigm shift in the history of superconductors as theoretical prediction now guides experiments toward new superconductors. With extensive R&D, it is possible to realize hydrogen-rich superconductors operating under temperature and pressure close to ambient conditions.

To synthesize high-T_c superconductors under pressure, miniature-sized samples must be packed in diamond anvil cells (DAC). After the synthesis, characterizing electrical resistivity and the Meissner effect in samples under DAC conditions requires highly specialized techniques. This technically challenging procedure dictates that the standard synthesis-test-repeat cycles for material discovery do not apply to high-pressure superconductors. Thus, experimental investigations of high-T_c superhydrides have been only focused on a handful of candidates. To accelerate the discovery, theoretical predictions are actively employed to perform computational materials searches under extreme pressure for new superconducting materials. Using efficient global optimization algorithms [11–15], crystal structure search can be carried forward for superhydrides corresponding to any given stoichiometry from scratch (without the need for experimental input). Once candidate materials are

* Corresponding author.
E-mail address: yansun.yao@usask.ca (Y. Yao).

https://doi.org/10.1016/j.mtphys.2024.101384
Received 22 December 2023; Received in revised form 26 February 2024; Accepted 28 February 2024
Available online 1 March 2024

identified, density functional theory for superconductors [16] or Eliashberg theory of superconductivity [17] are often employed to estimate the strength of electron-phonon coupling (*epc*), which is the essential characteristic of the conventional superconductivity per the BCS theory [18,19].

Theoretical and experimental evidence suggests that finding superconductivity at higher temperatures and lower pressures is likely connected with ternary or higher hydrides [20]. However, predicting crystal structures of ternary hydrides could be lengthy if they are entirely based on *ab initio* calculations. The complexity of the ternary system means a much larger atom count in the unit cell and a significant increase in the number of landscape valleys in the phase space. Meanwhile, the stoichiometry number for ternary systems is significantly higher than for the binary systems. An estimate suggests there are less than 300 stable binary hydrides at any pressure, but the number of ternary hydrides will exceed 1700 [21]. Consequently, an exhaustive *ab initio* exploration of ternary hydrides would be much more challenging. Thus, there is a demand for developing material prediction schemes with comparable accuracy to *ab initio* methods while simultaneously being computationally efficient. Theoretical approaches that can balance accuracy and efficiency are desirable for the computational discovery of new superconductors.

Over the past few years, artificial intelligence has made its way into material science and has proven to be a successful tool for exploring hydrogen-rich superconductors. Machine learning (ML), a universal fitting tool without priori physics formulae, has significant roles to play in superconductor research. The three focused utilities of ML in superconductor discovery are: (i) reparameterizing the semi-empirical superconductivity equation for a better estimate of T_c in conventional superconductors, (ii) accelerating crystal structure search in potential superconductor space using ML interatomic potentials, and (iii) screening candidate superconductors from existing material database or during high-throughput calculation using efficient T_c and *epc* descriptors. The existing ML efforts in the workflow of new superconductor discovery are depicted in Fig. 1. This brief review summarizes a selection of studies that used ML to investigate hydride superconductors. We do not attempt to present an exhaustive review but only hope to give the reader a survey of studies that represent the recent efforts in this area and what can be achieved with this new method. The milestone activities reviewed in this paper are listed in Table 1.

2. Machine learning accelerated superconductor prediction

2.1. Reparameterizing the McMillan equations

High-pressure experiments [5] suggest that superconducting superhydrides are conventional or BCS superconductors described by the *epc* mechanism. The estimate of T_c is derived from the BCS theory [18,19], which incorporates Debye temperature Θ_D, electronic density of states (DOS) at the Fermi level N (E_f) and electron-phonon attractive interaction V,

$$T_c = 0.85\Theta_D \exp\left[-\frac{1}{N(E_f)V}\right]. \tag{1}$$

Hydrogen-rich systems possess high Debye temperature Θ_D and phonon frequencies primarily due to the low mass of the proton. The Migdal-Eliashberg theory [17] extends the BCS theory to the strong coupling regime, relating the microscopic superconducting theory to the observed T_c in real materials. The Eliashberg spectral function $\alpha^2F(\omega)$ describes the superconducting state in this formalism. The strength of attractive interaction is characterized by the *epc* parameter λ, a weighted average of $\alpha^2F(\omega)$ over the entire vibrational region. The $\alpha^2F(\omega)$ can be obtained for specific materials by tunnelling experiments or density functional perturbation theory [33] calculations. The superconducting gap Δ_0 and T_c are obtained in a complete solution of the Eliashberg equation [34]. The Eliashberg equation involves a nonlinear system and must be solved self-consistently [35,36], which is computationally demanding. It is, therefore, desirable to derive an explicit formula for T_c from approximately fitting the experimental results. A commonly used formula is the McMillian equation [37],

$$T_c = \frac{\omega_D}{1.45}\exp\left(-\frac{1.04(1+\lambda)}{\lambda-\mu^*(1+0.62\lambda)}\right), \tag{2}$$

where ω_D represents the Debye frequency, and μ^* is the Coulomb pseudopotential. The average *epc* in the system is represented by λ, which is a weighted average of $\alpha^2F(\omega)$ over the entire vibrational range,

$$\lambda = 2\int_0^\infty \frac{\alpha^2F(\omega)}{\omega}d\omega. \tag{3}$$

The McMillan equation is inaccurate for strong coupling superconductors ($\lambda > 1$). A modification of the McMillan equation to the large λ regime was made by Allen and Dynes [38],

$$T_c = \frac{f_1 f_2 \omega_{\log}}{1.20}\exp\left(-\frac{1.04(1+\lambda)}{\lambda-\mu^*(1+0.62\lambda)}\right), \tag{4}$$

where $f_1 = \left(1+\left(\frac{\lambda}{2.46(1+3.8\mu^*)}\right)^{\frac{3}{2}}\right)^{\frac{1}{3}}$ and $f_2 = \left(1+\frac{\lambda^2(\overline{\omega}_2/\omega_{\log}-1)}{\lambda^2+[1.82(1+6.3\mu^*)(\overline{\omega}_2/\omega_{\log})]^2}\right)$.

In these formulas, $\overline{\omega}_2$ and $\omega_{\log}$ are the weighted root mean square and logarithmic average frequencies. The Allen and Dynes modified McMillian equation was fitted based on 217 numerical solutions of the Eliashberg equations to three types of $\alpha^2F(\omega)$ shapes, those of Pb and Hg, and that of a single Einstein model. The limited function set dictates that this equation would not work well for superconductors whose λ is

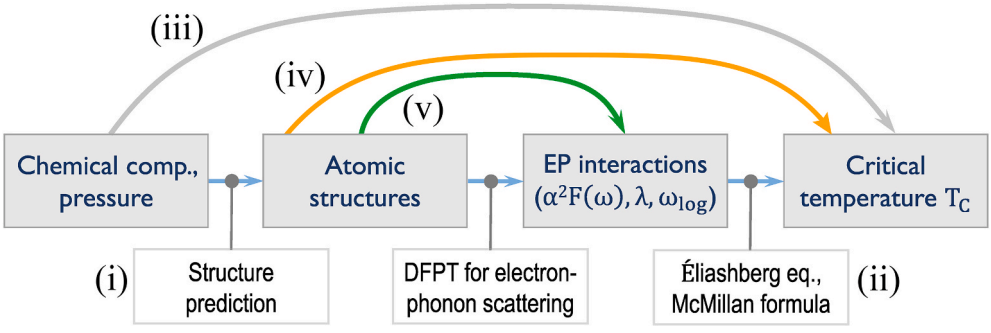

Fig. 1. A typical workflow to calculate the T_c of a material [22]. Existing ML efforts assist in four lines to accelerate the workflow: (i) using ML potentials to accelerate CSP, (ii) generating formulas between T_c and *epc* parameters, (iii) predicting T_c from descriptors calculated before CSP, (iv) predicting T_c from descriptors calculated after CSP, and (v) predicting *epc* parameters from descriptors calculated after CSP. Reprinted (figure) with permission from Ref. [22] and modified, Copyright (2023) by the American Physical Society.

Table 1
Milestone activities from 2018 to 2023 reviewed in this paper. The types of activities correspond to the five types (i to v) in Fig. 1.

Year	Work	Type	Ref	New superhydrides
2018	Neural network model assisting in researching T_c distribution of superhydrides	iii	[23]	–
2019	Evolutionary algorithm (EA) based material informatics approach to discover new superhydrides	iii	[24]	$KScH_{12}$ (122 K 300 GPa), $GaAsH_6$ (98 K 180 GPa)
2020	Neural network model that simultaneously searches for superhydrides with high T_c and low operational pressure	iii	[25]	RbH_{12} (115 K, 50 GPa)
2021	New electron structure descriptor developed: networking value	iv	[26]	–
2021	High throughput superhydride calculation with a Gaussian process regression model combined with Gaspari-Gyorffy *epc* estimate	iv	[27]	NaH_6 (248–279 K 100 GPa)
2022	Predicting *epc* parameters and T_c from structural information by the atomistic line graph neural network (ALIGNN) and classical force-field descriptor (CFID)	iv, v	[28]	–
2022	Correcting the Allen-Dynes formula for large λ and multimodal $\alpha^2F(\omega)$	ii	[29]	–
2022	Employing structural descriptors based on the smooth overlap of atomic positions (SOAP) to train a T_c model using a dataset comprising 5713 superconductors (larger scope than superhydrides)	iv	[30]	–
2023	Combining highly accurate MILPs with SSCHA techniques to achieve careful convergence and acceleration in SSCHA calculations	i	[31]	–
2023	Using ephemeral data-derived potentials (EDDP) accelerated CSP to search for ambient superconductivity of Lu-N-H	i	[32]	Confirmed that no room-temperature conventional superconductivity in Lu-N-H
2023	Gaussian process regression model to predict *epc* parameters from atomic structures under pressure	v	[22]	CrH (15.7 K, 0 GPa), CrH_2 (10.7 K, 0 GPa)

significantly deviating from 1.6 (the λ for Hg and Pb) and those whose $\alpha^2F(\omega)$ shapes differ considerably from the unimodal Einstein model.

A very recent development of correcting the Allen-Dynes formula for strong coupling was formulated by Xie et al. [29]. The authors suggest that the Allen-Dynes formula primarily applies to situations where λ is approximately 1.6, and the $\overline{\omega}_2/\omega_{\log}$ ratio is close to 1. The latter is taken as a shape indicator, *i.e.*, '1' corresponding to $\alpha^2F(\omega)$ of the unimodal Einstein model. Superconducting superhydrides have significantly larger λ and multimodal $\alpha^2F(\omega)$ – they contribute to systematic errors when applying the Allen-Dynes formula to estimate their T_c's. Xie et al. constructed the training dataset consisting of $\alpha^2F(\omega)$ of real materials and randomly generated $\alpha^2F(\omega)$ with three random Gaussian peaks. The latter is inexpensive to generate, and the arbitrarily trained multimodal data are more realistic than the unimodal model. They also included the data on superhydrides in the testing set. The authors trained a symbolic regression (SISSO framework) ML model using these datasets. This model exhibits an enhanced accuracy for high-T_c superhydrides while maintaining the same level of performance for the low-T_c superconductors as the Allen-Dynes formula. The ML model is used to reparameterize the McMillian equation, yielding a formula that closely resembles the Allen-Dynes formula,

$$T_c^{ML} = \frac{f_\omega f_\mu \omega_{\log}}{1.20} \exp\left(-\frac{1.04(1+\lambda)}{\lambda - \mu^*(1+0.62\lambda)} \right), \tag{5}$$

with $f_\omega = 1.92\left(\frac{\lambda+\omega_{\log}/\overline{\omega}_2-\sqrt[3]{\mu^*}}{\sqrt{\lambda}\exp(\omega_{\log}/\overline{\omega}_2)}\right) - 0.08$, and $f_\mu = \frac{6.86\ \exp(-\lambda/\mu^*)}{1/\lambda-\mu^*-\omega_{\log}/\overline{\omega}_2} + 1$.

Moreover, Xie et al. trained two additional models: a random forest (RF) model and an artificial neural network (ANN) model. Compared to Symbolic Regression and ANN, the RF model exhibits the lowest root-mean-square error (RMSE) in the training and testing datasets. Yet, the RF model experiences two notable problems. Firstly, its output encounters discontinuities as the value of λ increases due to its piecewise-constant form. Secondly, it cannot extrapolate points beyond those in the training dataset. On the other hand, ANN performs quite similarly to the Symbolic Regression, yet it has a greater model complexity. Symbolic Regression stands out for its simplicity and invertibility, which allows extracting λ from experimentally measured T_c and $\alpha^2F(\omega)$. However, it should be noted that $\alpha^2F(\omega)$ is very expensive to calculate since the calculation of *epc* matrices is based on DFPT [33], which led to the development of descriptors for *ep* superconductivity based on electronic quantities such as DOS, electron configuration and bonding relevant quantities. Similarly, quantities related to the atoms and structures, such as hydrogen concentration, space group, mass per atom, etc., can also be used. The *ep* descriptors play a crucial role in the high-throughput workflow of screening materials, facilitating the discovery of new superconductors. Several studies have achieved notable success in devising workable descriptors of T_c, which are used to screen existing material databases for new superconductors [22,39,40].

2.2. Models aid in predicting new superhydrides

In 2019, Ishikawa et al. developed an evolutionary algorithm (EA) based material informatics approach to discover new superhydrides [24]. After the database collection, the authors employed genetic programming (GP) to train a predictor for T_c and predicted potential candidates for high T_c superhydrides by regression analysis. The crystal structures of the candidate materials are predicted using a genetic algorithm (GA). After the first principles-based functionality validation, the predicted structures are selected to augment the datasets. By repeating these procedures, both the database and the predictor are improved iteratively. The authors employed five quantities for the T_c predictors, including hydrogen concentration, space group, mass per atom, and effective screened Coulomb repulsion constant (the same as the one used in the Allen-Dynes formula).

An interesting concept in this work is employing the predictors trained on binary superhydrides to predict ternary superhydrides. A dataset for hypothetical ternary hydrides was generated by replacing some heavy elements in binary hydrides with their neighboring elements in the periodic table, *i.e.*, elements with $Z-1$ or $Z+1$. For instance, H_3S is used to generate H_6PCl, and LaH_{10} is used to create $BaCeH_{20}$. All other parameters apart from the atomic masses match the original binary hydrides. A GA structure search is then performed to identify the targeted ternary hydrides' crystal structures. Using this approach, the authors predicted two ternary superhydrides, a $KScH_{12}$ with an estimated T_c of 122 K at 300 GPa and $GaAsH_6$ with a T_c of 98 K at 180 GPa. The outcomes of this study show that, qualitatively, regression analysis combined with GP predictor is practical in predicting ternary superhydrides. Even though some regression analyses based on the GP predictor exhibit some errors, this work successfully transforms GP techniques into developing T_c predictors. It renders predictors based on simple quantities rather than expensive $\alpha^2F(\omega)$. As noted by the authors, GP models perform well as predictors for a small number of datasets and parameters, as shown in this work. Although GP models may have a lower accuracy than ANN models, they can generate/output symbolic formulas and offer good interpretability if the formula is not overly complicated. It isn't easy to interpret ANN models meaningfully for their black box features. The lower accuracy of GP models may be compensated by the cycle of the materials informatics workflow.

In 2020, Hutcheon et al. trained a neural network model that simultaneously searches for superhydrides with high T_c and low operational pressure [25]. The authors trained the model using atomic and electronic descriptors for binary hydrides: hydrogen content (n), van der

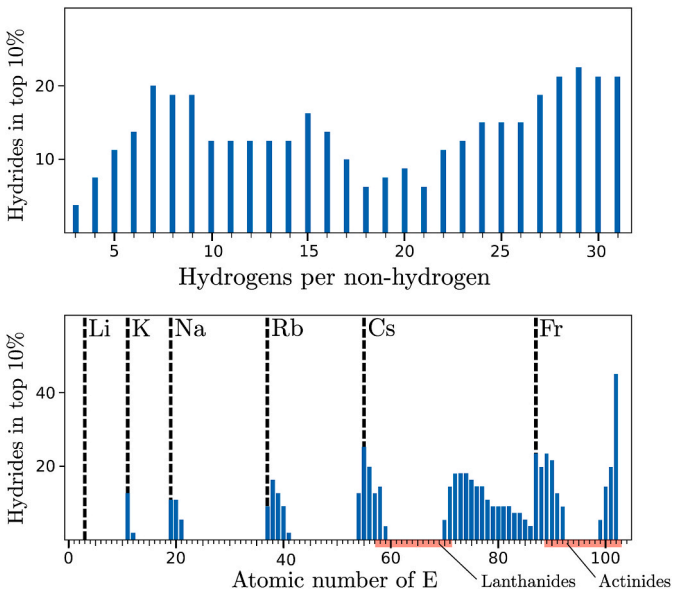

Fig. 2. The top 10% of binary hydrides (EH_n) predicted to be superconducting closest to ambient conditions (atmospheric pressure and room temperature). Upper and Bottom: The distribution of hydrogen number n and atomic number of nonhydrogen element E in this group of hydrides. Dashed lines indicate the positions of alkali metals. Reprinted (figure) with permission from Ref. [25], Copyright (2020) by the American Physical Society.

Waals radius, atomic and mass numbers, and electronic configuration of the non-hydrogen element. The training dataset includes computational and experimental data of various binary hydrides in the literature. This model was applied to predict the T_c and operational pressure of the binary hydrides (n = 1–32) of all possible elements in the Periodic Table. The results are interesting: hydrides of heavy alkali and alkaline earth metals are the most promising candidates for attaining superconductivity at the ambient-condition limit (Fig. 2, lower panel). The distribution of hydrogen content within the close-to-ambient-condition superhydrides is relatively uniform, suggesting that different stoichiometries should be considered for each binary hydride (Fig. 2, upper panel). The authors then focused on Cs and Rb binaries, which have not been studied as extensively as other hydrides in the first two groups. They successfully predicted a superhydride operating near ambient conditions, specifically RbH_{12} (115 K, 50 GPa), identified with structure search and characterized by first principles calculations.

2.3. Models unveiling T_c distribution of superhydrides

Based on the T_c data of several metal hydrides, Semenok et al. suggest [41,23] that most high-T_c superconducting hydrides should be located in the 'lability belt' region, which lies at the border between *s/p* and *s/d* blocks in the periodic table, specifically in the proximity of Groups IIA and IIIB. This is possibly due to the anomalously strong *epc* for hydrides in which the metal atoms have minor energy differences between *s*-, *p*-, and *d*-orbitals under high pressure. The authors collected data on superhydrides from the literature. They employed a neural network model based on the maximum T_c values of the known hydrides to predict the maximum T_c for the hydrides of elements that lack data. Using the distribution of the maximum T_c of elements in the periodic table, they predicted several new superhydrides and rectified several previous results. An in-depth analysis of the data yields the following trends.

(1) Most high-T_c metal superhydrides are found within the 'lability belt': Sc-T-La-Ac for the d^1 belt, Mg–Ca–Sr–Ba-Ra for the d^0 belt, and Th for the s^2d^2 belt. The physical phenomenon involved is that the orbital populations in these elements are sensitive to the atomic environment, causing strong *epc*.
(2) Hydrides attain maximum T_c when the average electron transfer between metal and hydrogen is approximately 0.3 e^-/atom. The superhydrides with the highest T_c feature a DOS at the Fermi level of about 1 state/Ry/H-atom.
(3) The superconducting properties notably diminish when *d*- and *f*-electrons in the hydrides increase. Notably, the T_c of lanthanide and actinide hydrides almost decrease monotonically as *f*-electrons increase.
(4) $XH_{10\pm2}$ should be the optimal hydrogen content, which could be achieved at about 150–250 GPa. The superconducting properties decrease when the stoichiometry and the electron doping level deviate from this optimal value.
(5) The stabilization pressure of hydrides decreases when moving down the periodic table. A pressure of 50 GPa is sufficient to stabilize superconducting actinide polyhydrides, such as UH_7 and PaH_8.

It is worth noting that these T_c and pressure trends are based on many calculated or experimentally measured data by the authors or in the literature, in addition to the data obtained in neural network models. Therefore, the conclusions regarding the distribution of superhydrides hold greater reliability than the results solely predicted by the ML models. Nonetheless, the work of Hutcheon et al. [25], in which the neural network models exclusively made predictions on the distribution of superhydrides, reveals similar trends in some respects. For example, in the work of Hutcheon et al. the T_c of hydrides for elements in the second and third groups are also relatively high. However, some notable discrepancies exist in their ML predictions, such as the underestimated T_c of about 100 K for Y, La, and Th hydrides, even when the data are included in the training set. These errors are considerably high relative to the RMSE of 33.7 K for its entire dataset, which might be due to the inefficient descriptors chosen to train the models.

2.4. New descriptors

Two subsequent research works introduced advanced descriptors to train ML models for superhydrides. First, in 2021, Belli et al. developed an electron structure descriptor, networking value [26], based on the electron localization function (ELF) [42]. Networking value φ is defined as the highest value of the ELF that creates an isosurface in the whole crystal, which is used to design a descriptor,

$$\Phi_{DOS} = \varphi H_f \sqrt[3]{H_{DOS}}. \quad (6)$$

Here H_f is the hydrogen fraction of the compound, and H_{DOS} is the hydrogen fraction of the total DOS at the Fermi level. By analyzing more than a hundred predicted hydrogen-based superconductors, the authors determine a strong correlation between the T_c and Φ_{DOS} (Fig. 3, left). Therefore, creating an electronic bonding network between localized units is critical to enhancing the T_c in superhydrides. Networking value combined with the hydrogen fraction in a superhydride and hydrogen contribution to the DOS at the Fermi level can predict the T_c within an accuracy of about 60 K. To calculate Φ_{DOS}, ELF is first calculated, a real-space function based on electron density. ELF is 1 when electrons are perfectly localized and 0.5 for homogeneous electron gas. Hence, bonds, cores, and lone pairs of electrons would feature ELF maxima. The ELF value between the regions of ELF maxima describes the degree of delocalization, indicating how easily electrons can transfer between localized units. Following the calculation of ELF, the maximum ELF value (φ) that enables ELF saddle points and the atoms to form a 3D network spanning through the crystal is determined (Fig. 3, right). It should be noted that this descriptor can be directly obtained through electronic structure calculation and, therefore, is considerably cost-effective compared to the $\alpha^2F(\omega)$. It could be used for the high-

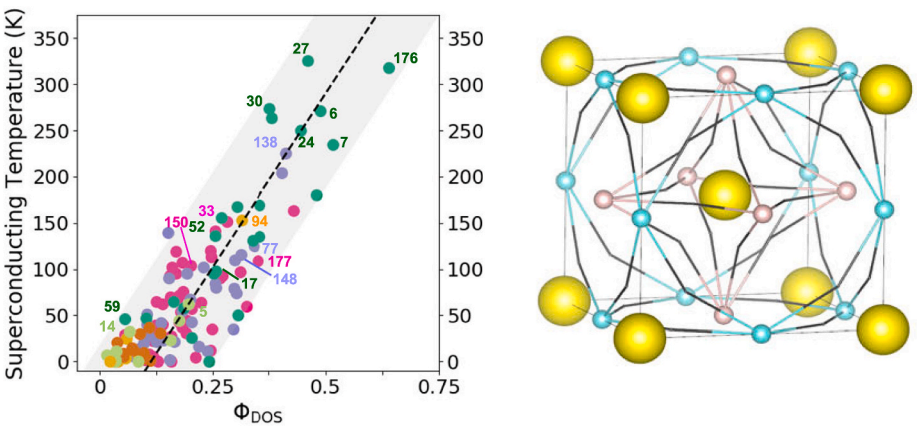

Fig. 3. Networking values. Left: the correlation between T_c and descriptor Φ_{DOS} based on the networking value φ for more than a hundred predicted binary hydrides. The dotted line represents a fit ($T_c = 750\Phi_{DOS} - 85$) K. Right: Three-dimensional network formed by the ELF saddle points (black turns) and the atoms (large and small spheres) for H_3S with a networking value φ of 138. Reprinted (figure) with permission from Ref. [26] under the terms of the Creative Commons Attribution 4.0 International License Copyright (2021) The Author(s). Published by Springer Nature.

throughput search of superhydrides. On the other hand, these quantities are electron-structure specific, making them more relevant to T_c than simpler material fingerprints like space groups and hydrogen fractions. Effective input descriptors play a crucial role in enhancing the performance of numerical models.

Initially used for transition metal superconductors, the multiple scattering-based Gaspari-Gyorffy (GG) theory [43] has recently been extended to study metal hydrides under high pressure [27,44,45]. In 2021, Shipley et al. realized the high-throughput discovery of binary superhydrides by designing one T_c model based on Gaspari-Gyorffy estimates of *epc* [27]. As shown by McMillan [37], the *epc* constant λ could be written in a simplified form,

$$\lambda = \frac{N(E_F)\langle I^2\rangle}{M\langle\omega^2\rangle}, \tag{7}$$

where $\langle I^2\rangle$ is the Fermi surface averaged *ep* matrix element and $\langle\omega^2\rangle$ is the mean squared phonon frequency. $N(E_F)$ is the total DOS at the Fermi level. Considering *epc* for element-specific contributions, Eq. (7) is written as,

$$\lambda = \sum_j \frac{\eta_j}{M_j\langle\omega_j^2\rangle}. \tag{8}$$

Here, the vibrational modes are divided by different atomic characteristics, where j specifies atom type and η_j is the Gaspari-Gyorffy *epc* estimate or Hopfield parameter. Gaspari-Gyorffy theory [43] estimates η_j using quantities available from self-consistent DOS calculations, avoiding expensive $\langle I^2\rangle$ calculations. With the calculated $\langle\omega_j^2\rangle$ and η_j, λ is approximated, and T_c can be determined using the McMillan equation and its variants. However, the calculation of $\langle\omega^2\rangle$ still requires computationally demanding phonon calculations. Consequently, Shipley et al. trained a model to directly utilize η_j calculated using the Gaspari-Gyorffy theory and other DFT descriptors to predict T_c [27]. The model developed in this work is a Gaussian process regression model for binary superhydride X_nH_m, with input descriptors including η_H, η_X, M_X, $N(E_F)$, and $N_H(E_F)/N(E_F)$ ratio.

The methodology for conducting high throughput superhydride discovery is as follows: First, the authors calculated the properties of 160 literature structures and trained the initial ML model. Then, structure search calculations are performed to identify binary hydrides across the periodic table under a certain pressure (ranging from 10 GPa to 500 GPa). Structures on and just above the convex hulls are selected, and their descriptor properties are computed to predict the T_c using the model. Next, *epc* calculations are performed on the best-selected structures to obtain accurate T_c. Occasionally, calculations are extended to middle- or low-performance structures to augment the training set with 'low-performance data' to improve the model behavior. These results are fed back to the training set to retrain the model. The same structure search process is performed at all pressures of interest, through which the model is improved iteratively. Finally, the training set is augmented by 119 new data points, on which the final machine-learning model is trained. It should be noted that in these training iterations, only dynamically stable structures are selected to augment the training set. All search structures considered previously are recalculated with this final model again to avoid missing anything of interest. Based on the T_c results obtained during the training iterations, more focused structure searches are performed on promising systems. Predicted structures are screened by convex hull analysis and T_c model screening. The systems with the highest predicted T_c values (including those found in training iterations) are subject to fully converged *epc* calculations to obtain accurate T_c. It must be noted that, in this work, dynamically unstable structures are also subjected to analysis to a certain extent.

This work successfully uncovered several new superhydrides with good performance characteristics. In total, 36 dynamically stable superhydrides with T_c above 100 K were identified, and 27 were new. Also, 18 of the 36 superhydrides have T_c values above 200 K. In particular, NaH_6 is calculated to have a T_c of 248–279 K under the pressure of 100 GPa. The superconductivity of this system has been overlooked before, although sodium polyhydrides have been synthesized experimentally [46]. The prediction of NaH_6 (Fig. 4) pushed the high T_c – low synthesis pressure limit and provided new insight into the superconductivity in superhydrides. Shipley et al. [27] achieve a high-throughput discovery of superhydrides, employing both ML screening and high-throughput *epc* calculations to yield an efficient superconductor search. The success is based on several aspects: First, the authors implemented optimizations to accelerate the calculation of *epc* by nearly an order of magnitude. This enables the training set to expand with more first principles data iteratively and increases the model's performance. Secondly, the double screening procedure, convex hull and T_c-model screening significantly reduce the calculation cost and increase the possibility of identifying superhydrides with better performance. In addition, dynamical stability is also used to help select the candidate structures. Third, a pivotal reason is that most of the descriptors are directly based on electronic structure, which helps increase the relevance between the descriptors and T_c. In particular, Gaspari-Gyorffy *epc* estimates and atomic masses directly relate to T_c through the McMillan equation. Lastly, low-performance data are not entirely discarded, further improving the model's performance.

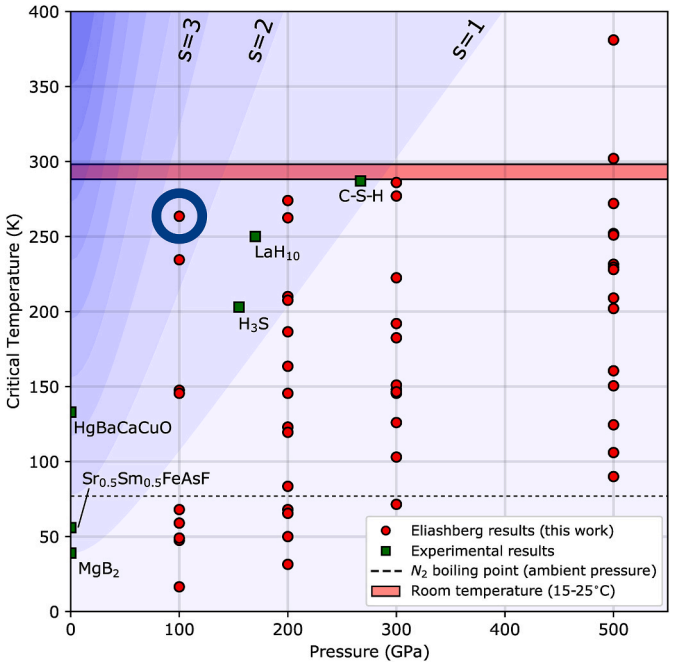

Fig. 4. Pressure distribution of T_c calculated by Eliashberg theory for the most promising superhydrides according to the ML screening (red dots), alongside experimental results for specific superconductors (green squares). The background is shaded according to the figure of merit S. The red dot in the blue circle is NaH_6. Note that the experimental data for C–S–H hydride has been retracted due to the questions raised regarding how the data have been processed and analyzed. Reprinted (figure) with permission from Ref. [27] and modified, Copyright (2021) by the American Physical Society. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2.5. Structural features

Compared to the limited dataset, i.e., a few hundred entries utilized in ML models for *superhydride* predictions, the dataset for modelling T_c of *other types of superconductors* (where the databases are available) could encompass thousands of entries, yielding better-performed models. In 2022, Zhang et al. employed structural descriptors based on the smooth overlap of atomic positions (SOAP) [47] to train a T_c model using a dataset comprising 5713 superconductors [30]. To construct the dataset, the authors extracted the structure information from the ICSD database [48] and superconductivity-specific information from the SuperCon database [49]. In the dataset, 2089 entries are cuprate ceramics, 762 are iron-based superconductors, and 2862 mainly consist of conventional superconductors. Three distinct types of ML models were trained [50], including random forest regression (RFR), xgboost regression (XGBR), and support vector regression (SVR), among which the RFR model shows the best coefficient of determination (R^2) of 0.929, while the XGBR model exhibits the best RMSE of 7.54 K. The authors also developed ML models employing only 'element descriptors' derived from elemental property statistics without crystal structure information. Consequently, the models without crystal structure information yielded the best R^2 of 0.895 and the best RMSE of 8.94 K. This result demonstrates that utilizing descriptors based on crystal structure information, such as SOAP, weighted atom-centered symmetry functions (wACSFs) [51,52], and bispectrum coefficient descriptors [53,54], might serve to increase the performance of the ML models. On the other hand, considering the large dataset used (5713 structures), it is unsurprising that this model achieved a lower RMSE of 7.54 K compared to the RMSE of ~30 K in superhydride models built upon the datasets of several hundred entries. This serves as a reminder that further increasing the performance of superhydride models requires a large dataset by implementing more efficient high-throughput superconductivity calculations.

Other than exploiting the structural descriptors, another important consideration is about the target predicted quantities that most correlate with structure features. In 2022, Choudhary et al. developed a workflow to find conventional superconductors through pre-screening of 1736 materials with high Debye temperature Θ_D and electronic density of states at Fermi-level N (E_f), and first principles *ep* calculations for 1058 entries of them [28]. Remarkably, they achieved a cost-effective, high-throughput *ep* calculation approach by extensive benchmark testing and convergence checks. With this method, they established a large and systematic database of BCS superconducting properties and identified 105 dynamically stable materials with $T_c \geq 5$ K. To accelerate these screening processes, they developed ML models to predict N (E_f), Θ_D, *epc* properties ($\alpha^2F(\omega)$, $\omega_{\log}$, and λ), and T_c. For *epc* and T_c models, they used two methods: the atomistic line graph neural network (ALIGNN) [55] and classical force-field descriptor (CFID) [56]. These two methods, respectively, consider crystal structural features in the form of graphs and (numeric) descriptors. They found that while $\omega_{\log}$ could be well predicted using the crystal structure, λ and T_c are more challenging to predict. However, the performance could be improved by predicting the Eliashberg function $\alpha^2F(\omega)$ as intermediate quantities to calculate T_c through McMillan-Allen-Dynes type equations. Specifically, the mean absolute error (MAE) of predicting a T_c directly from structures would be more significant than that in calculating the T_c from predicted $\omega_{\log}$ and λ. The MAEs of these two models would be larger than that in calculating T_c from predicted $\alpha^2F(\omega)$. This implies that learning more basic and informative *epc* quantities such as $\alpha^2F(\omega)$ could perform better than direct predictions of the integrated quantity T_c. It should be noted that although the workflow could be applied to high-pressure superconductivity research, the authors only analyzed systems under atmospheric pressure in this work.

In 2023, Tran et al. utilized structural descriptors with the invariance of transformation to train a Gaussian process regression model on a dataset of 584 atomic structures [22]. Unlike the previous work [28], which lacks studies of high-pressure superconductors, the dataset contains structures under various pressures, in which 505 entries contain hydrogen. With the same concept as the previous work, this model aims to predict λ and $\omega_{\log}$ directly from atomic structures, which speeds up the *ep* superconductor screening by skipping the *ep* scattering calculation through DFPT. With predicted λ and $\omega_{\log}$, T_c could then be obtained using the McMillan formula. Screened by this model and validated by first principles calculation, the authors predicted two possible superhydrides at ambient pressure: CrH (15.7 K) and CrH_2 (10.7 K). This model benefits from the direct and physics-inspired correlation between atomic structures and *ep* properties, thus being accurate and robust for structures under different pressures. This enables the model to be applicable in screening large atomic-structural databases and on-the-fly structure searches.

2.6. Machine learning in crystal structure prediction (CSP)

Other than helping develop quantitative structure-property relationships (QSPR) for T_c, ML schemes also significantly contribute to CSP, thereby accelerating the exploration of superhydrides. Modern CSP plays a critical role in determining the unknown structures for the stoichiometry of materials with little to no experimental inputs [57–59]. Commonly used CSP algorithms locate ground-state structures using various strategies: particle swarm optimization (PSO) explores the energy landscape with flock intelligence [60,61], genetic algorithms [62–66] improve structure population by propagating beneficial traits, random search [67,68] samples the structural space in an unbiased but sensible fashion, and metadynamics [69] explores the energy landscape by accelerating the occurrence of rare events. As the exploration for superhydrides is extended from binary to ternary and higher superhydrides, the challenges to perform CSP significantly escalate due to the

explosive increase of the combinational possibilities [70]. There can be the same number of binary hydrides for the 83 non-radioactive elements in the periodic table, but the ternary and quaternary hydride combinations are 6723 and 5.5×10^5, respectively. Moreover, each hydride combination can have various stoichiometries and metastable states. While the ground state is typically required for superhydrides, there are instances where metastable states exhibit very high T_c and, therefore, must be considered [71–73]. With this in mind, efficient CSP searches need to be developed.

ML addresses this problem with two strategies: building reliable yet computationally inexpensive machine learning interatomic potentials (MLIP) and constructing effective ML generative models for crystal structures. MLIPs have been combined with mainstream CSP algorithms to accelerate the search for stable structures on the potential energy surface (PES) that are beyond the reach of DFT-based calculations [32,74–79]. The structural descriptors are central to the construction of MLIPs. In 2007, Behler and Parrinello proposed the first successful descriptor based on atom-centered symmetry functions invariant to rotation, translation, and atom swap [80]. Commonly used descriptors in MLIP include bispectrum [53,81], SOAP [47], weighted symmetry functions [51], and many others, providing multiple options for balancing accuracy and speed.

Recently, MLIPs were employed to study a ternary hydride system, Lu-H-N. In 2023, an experimental team claimed the achievement of near-ambient-pressure room temperature superconductivity in nitrogen-doped Lu–H hydrides. This claim was immediately met with skepticism regarding the reliability of data [82]. Interestingly, the discovery of 'ambient superconductivity' contradicted most theoretical studies [83–85]. Particularly, Ferreira et al. conducted a comprehensive analysis of the phase diagram of the Lu-N-H superhydrides, looking for superconducting phases using ephemeral data-derived potentials (EDDP) accelerated CSP [32]. The authors sampled over 200,000 different hydride combinations of Lu-N-H and identified 52 possible candidates for conventional superconductors. However, whether doped with N or not, the computed T_c's for Lu hydrides are not remotely compatible with room temperature. At the time of writing, the experimental report of 'ambient superconductivity' in Lu-H-N has been retracted. The MLIPs accelerated CSP had played an essential role in the community effort, confirming that no room-temperature conventional superconductivity is expected in this compound.

2.7. Machine learning in the calculation of quantum nuclear and anharmonic effects

Quantum nuclear and anharmonic effects are significant in superhydrides because of the small mass of hydrogen atoms. The energy landscape can undergo substantial changes under these effects. For instance, the high-symmetry structures of H_3S and LaH_{10} are stable once the quantum nuclear and anharmonic effects are accounted for [86,87]. Otherwise, they would distort to low symmetry structures in standard DFT calculations and yield much lower T_c. Generally, the stochastic self-consistent harmonic approximation (SSCHA) framework performs calculations considering both quantum nuclear and anharmonic effects [88]. ML also plays a role in assisting with the SSCHA calculation. In 2023, Lucrezi et al. combined highly accurate MILPs with SSCHA techniques to achieve careful convergence and acceleration in SSCHA calculations [31]. Based on compressive sensing (an ML algorithm), compressive sensing lattice dynamics (CSLD) methods can also be used in anharmonic phonon renormalization [89,90]. Recently, Kruglov et al. developed a solution for finite temperature evolutionary CSP through quick anharmonic free energy calculations using canonical ensemble (NVT) molecular dynamics with ML potentials [91]. The ML potential is constructed by on-the-fly training Moment Tensor Potentials [92] and corrected using thermodynamic perturbation theory to the *ab initio* level.

3. Existing problems and future perspectives

First and foremost, the CSP methods based on structural optimizations usually conduct the search under Born Oppenheimer approximation at $T = 0$ K to avoid expensive entropy and quantum nuclear calculations. This might not have sufficient accuracy for the hydrogen-rich materials where quantum-nuclear and anharmonic effects must be accounted for. For example, structural stabilities or superconducting properties are reported to be significantly affected by quantum-nuclear and anharmonic effects in the superhydrides under pressure [3,87,93,94]. However, it is computationally highly challenging to calculate energy corrections of quantum-nuclear and anharmonic effects in CSP on the fly. Addressing the PES's metastable and dynamically unstable structures becomes essential considering this situation. The saddle points on the PES within the Born-Oppenheimer approximation are dynamically unstable structures. In many cases, dynamically unstable structures could be stabilized by anharmonic or thermal effects, as observed in several superhydrides [3,71–73,87,94]. These structures can be located using saddle-point *ab initio* random structure searching methods (sp-AIRSS) [95]. Although filtering out and properly treating these dynamically unstable structures is expensive, solving this problem in future high-throughput schemes remains crucial.

Furthermore, finding relatively cost-effective descriptors more relevant to superconducting critical temperature T_c is important. Appropriate descriptors can directly influence the performance of machine learning QSPR models that predict superconducting properties. The *epc* matrices directly correspond to the T_c of conventional superconductors, yet they are computationally challenging to calculate. Newly developed networking values [26] are relevant to T_c and are solely based on electronic calculation, significantly more cost-effective than *epc* calculations based on DFPT. It could be possible to derive alternate descriptors with better performance directly from simple DFT results. Meanwhile, testing the performance of different descriptors could also help to understand the physics pictures behind the superconductivity in various types of superconductors [39].

In addition, we may consider designing ML schemes for superconductor discovery in a modularized way. As shown in Fig. 1, ML could be used in different processes across four nodes, including chemical composition and pressure, atomic structures, *epc* parameters, and T_c. Building type iii models, i.e., from chemical composition directly to T_c, may not be accurate because it skips essential information in the intermediate steps and, therefore, has a low correlation. A few practical schemes with intermediate steps are listed in Table 1 and focus on type ii, type iv, and type v models, which exhibit improved performance. Furthermore, models of type iv are reported to perform worse than type v connected with type ii using the same training dataset [28]. This implies that we may achieve higher performance and accuracy by training models of different sub-processes and connecting them.

The scale and quality of the superhydride database need to be further increased. Currently, most machine-learning QSPR models for superhydrides are based on datasets limited to hundreds of samples [22,27]. While it must be acknowledged that the computational and experimental costs of finding new superhydrides to supplement datasets are high, there are still some valuable thoughts. First, it is essential to construct databases specific to superhydrides, including experimental and computational superconductor properties. There are good databases for general superconductors like SuperCon [49], but to the authors' knowledge, few databases specific to superhydrides or other high-pressure superconductors exist. An example of such a database is the CALYPSO Database [96], which includes nearly one hundred entries of critical temperature and structures of superhydrides and other high-pressure superconductors. Second, to further supplement databases for superhydrides, it may be beneficial to use ML techniques to extract information from the literature [20]. Third, it is crucial to encourage the scientific community to share additional data about their work, including 'negative results' with low T_c or unstable structures. Not

only will this increase the scale of databases, but it also helps to increase the sampling diversity of the training datasets. The samples with low superconducting performance can also improve the accuracy of ML models, and certain unstable results might be stabilized after considering quantum-nuclear and anharmonic effects. Lastly, due to the difficulty in conducting high-throughput experiments for high-pressure systems, optimizing further and designing more efficient high-throughput superconducting calculation methods to generate abundant data becomes essential. With more extensive and diverse datasets, the performance of QSPR models can be further improved.

Finally, it is helpful to implement different ML algorithms for specific problems. Models based on various algorithms exhibit different performances when applied to different problems. Prioritizing and optimizing the algorithms that show high accuracy are desirable. The development and testing of the new algorithms are anticipated in the future.

CRediT authorship contribution statement

Zihao Bai: Writing – original draft, Investigation, Conceptualization, Writing – review & editing, Visualization. **Mangladeep Bhullar:** Writing – review & editing, Investigation. **Akinwumi Akinpelu:** Writing – review & editing, Investigation. **Yansun Yao:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yansun Yao reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

This project was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] V.L. Ginzburg, The Physics of a Lifetime: Reflections on the Problems and Personalities of 20th Century Physics, Springer, Berlin, 2001.
[2] G.F. Sun, K.W. Wong, B.R. Xu, Y. Xin, D.F. Lu, T_c enhancement of $HgBa_2Ca_2Cu_3O_{8+\delta}$ by Tl substitution, Phys. Lett. 192 (1994) 122–124.
[3] A. Drozdov, et al., Superconductivity at 250 K in lanthanum hydride under high pressures, Nature 569 (2019) 528–531.
[4] M. Somayazulu, M. Ahart, A.K. Mishra, Z.M. Geballe, M. Baldini, Y. Meng, V. V. Struzhkin, R.J. Hemley, Evidence for superconductivity above 260 K in lanthanum superhydride at megabar pressures, Phys. Rev. Lett. 122 (2019): 027001.
[5] A.P. Drozdov, M.L. Eremets, I.A. Troyan, V. Ksenofontov, S.I. Shylin, Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system, Nature 525 (2015) 73–76.
[6] I. Troyan, A. Gavriliuk, R. Rüffer, A. Chumakov, A. Mironovich, I. Lyubutin, et al., Observation of superconductivity in hydrogen sulfide from nuclear resonant scattering, Science 351 (2016) 1303–1306.
[7] L. Ma, et al., High-temperature superconducting phase in clathrate calcium hydride CaH_6 up to 215 K at a pressure of 172 GPa, Phys. Rev. Lett. 128 (2022): 167001.
[8] D. Duan, Y. Liu, F. Tian, D. Li, X. Huang, Z. Zhao, H. Yu, B. Liu, W. Tian, T. Cui, Pressure-induced metallization of dense $(H_2S)_2H_2$ with high-T_c superconductivity, Sci. Rep. 4 (2014) 6968.
[9] H. Liu, I.I. Naumov, R. Hoffmann, N.W. Ashcroft, R.J. Hemley, Potential high-T_c superconducting lanthanum and yttrium hydrides at high pressure, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) 6990–6995.
[10] H. Wang, J.S. Tse, K. Tanaka, T. Iitaka, Y. Ma, Superconductive sodalite-like clathrate calcium hydride at high pressures, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 6463–6466.
[11] C.J. Pickard, R.J. Needs, High-pressure phases of silane, Phys. Rev. Lett. 97 (2006): 045504.
[12] A.R. Oganov, A.O. Lyakhov, M. Valle, How evolutionary crystal structure prediction works-and why, Acc. Chem. Res. 44 (2011) 227–237.
[13] D.C. Lonie, E. Zurek, XTALOPT: an open-source evolutionary algorithm for crystal structure prediction, Comput. Phys. Commun. 182 (2011) 372–387.
[14] A.N. Kolmogorov, S. Shah, E.R. Margine, A.F. Bialon, T. Hammerschmidt, R. Drautz, New superconducting and semiconducting Fe-B compounds predicted with an ab initio evolutionary search, Phys. Rev. Lett. 105 (2010): 217003.
[15] Y. Wang, J. Lv, L. Zhu, Y. Ma, CALYPSO: a method for crystal structure prediction, Comput. Phys. Commun. 183 (2012) 2063–2070.
[16] L.N. Oliveira, E.K.U. Gross, W. Kohn, Density-functional theory for superconductors, Phys. Rev. Lett. 60 (1988) 2430–2433.
[17] G.M. Eliashberg, Interaction between electrons and lattice vibrations in a superconductor, J. Exp. Theor. Phys. Lett. 11 (1960) 696–702.
[18] J. Bardeen, L.N. Cooper, J.R. Schrieffer, Microscopic theory of superconductivity, Phys. Rev. 106 (1957) 162–164.
[19] J. Bardeen, L.N. Cooper, J.R. Schrieffer, Theory of superconductivity, Phys. Rev. 108 (1957) 1175–1204.
[20] B. Lilia, et al., The 2021 room-temperature superconductivity roadmap, J. Phys. Condens. Matter 34 (2022): 183002.
[21] J.A. Flores-Livas, L. Boeri, A. Sanna, G. Profeta, R. Arita, M. Eremets, A perspective on conventional high-temperature superconductors at high pressure: methods and materials, Phys. Rep. 856 (2020) 1–78.
[22] H. Tran, T.N. Vu, Machine-learning approach for discovery of conventional superconductors, Phys. Rev. Mater. 7 (2023): 054805.
[23] D.V. Semenok, I.A. Kruglov, I.A. Savkin, A.G. Kvashnin, A.R. Oganov, On distribution of superconductivity in metal hydrides, Curr. Opin. Solid State Mater. Sci. 24 (2018): 100808.
[24] T. Ishikawa, T. Miyake, K. Shimizu, Materials informatics based on evolutionary algorithms: application to search for superconducting hydrogen compounds, Phys. Rev. B 100 (2019): 174506.
[25] M.J. Hutcheon, A.M. Shipley, R.J. Needs, Predicting novel superconducting hydrides using machine learning approaches, Phys. Rev. B 101 (2020): 144505.
[26] F. Belli, T. Novoa, J. Contreras-García, I. Errea, Strong correlation between electronic bonding network and critical temperature in hydrogen-based superconductors, Nat. Commun. 12 (2021) 5381.
[27] A.M. Shipley, M.J. Hutcheon, R.J. Needs, C.J. Pickard, High-throughput discovery of high-temperature conventional superconductors, Phys. Rev. B 104 (2021): 054501.
[28] K. Choudhary, K. Garrity, Designing high-T_C superconductors with BCS-inspired screening, density functional theory, and deep-learning, npj Comput. Mater. 8 (2022) 244.
[29] S.R. Xie, et al., Machine learning of superconducting critical temperature from Eliashberg theory, npj Comput. Mater. 8 (2022) 14.
[30] J. Zhang, Z. Zhu, X.D. Xiang, K. Zhang, S. Huang, C. Zhong, H.-J. Qiu, K. Hu, X. Lin, Machine learning prediction of superconducting critical temperature through the structural descriptor, J. Phys. Chem. C 126 (2022) 8922–8927.
[31] R. Lucrezi, E. Kogler, S. Di Cataldo, M. Aichhorn, L. Boeri, C. Heil, Quantum lattice dynamics and their importance in ternary superhydride clathrates, Commun. Phys. 6 (2023) 298.
[32] P.P. Ferreira, L.J. Conway, A. Cucciari, S. Di Cataldo, F. Giannessi, E. Kogler, L.T. F. Eleno, C.J. Pickard, C. Heil, L. Boeri, Search for ambient superconductivity in the Lu-N-H system, Nat. Commun. 14 (2023) 5367.
[33] S. Baroni, S. de Gironcoli, A. Dal Corso, Phonons and related crystal properties from density-functional perturbation theory, Rev. Mod. Phys. 73 (2001) 515–562.
[34] Y. Yao, J.S. Tse, K. Tanaka, F. Marsiglio, Y. Ma, Superconductivity in lithium under high pressure investigated with density functional and Eliashberg theory, Phys. Rev. B 79 (2009): 054524.
[35] E.R. Margine, F. Giustino, Anisotropic Migdal-Eliashberg theory using Wannier functions, Phys. Rev. B 87 (2013): 024505.
[36] P.B. Allen, B. Mitrović, Theory of superconducting T_c, Solid State Phys. 37 (1983) 1–92.
[37] W.L. McMillan, Transition temperature of strong-coupled superconductors, Phys. Rev. 167 (1968) 331–344.
[38] P.B. Allen, R. Dynes, Transition temperature of strong-coupled superconductors reanalyzed, Phys. Rev. B 12 (1975) 905–922.
[39] V. Stanev, C. Oses, A. Gilad Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, Machine learning modeling of superconducting critical temperature, npj Comput. Mater. 4 (2018) 29.
[40] K. Matsumoto, T. Horide, An acceleration search method of higher T_c superconductors by a machine learning algorithm, Appl. Phys. Express 12 (2019): 073003.
[41] D.V. Semenok, A.G. Kvashnin, I.A. Kruglov, A.R. Oganov, Actinium hydrides AcH_{10}, AcH_{12}, and AcH_{16} as high-temperature conventional superconductors, J. Phys. Chem. Lett. 9 (2018) 1920–1926.
[42] A.D. Becke, K.E. Edgecombe, A simple measure of electron localization in atomic and molecular systems, J. Chem. Phys. 92 (1990) 5397.
[43] G.D. Gaspari, B.L. Gyorffy, Electron-phonon interactions, *d* resonances, and superconductivity in transition metals, Phys. Rev. Lett. 28 (1972) 801–805.
[44] D.A. Papaconstantopoulos, B.M. Klein, M.J. Mehl, W.E. Pickett, Cubic H3S around 200 GPa: an atomic hydrogen superconductor stabilized by sulfur, Phys. Rev. B 91 (2015): 184511.

[45] P.-H. Chang, S. Silayi, D.A. Papaconstantopoulos, M.J. Mehl, Pressure-induced high-temperature superconductivity in hypothetical H3X (X=As, Se, Br, Sb, Te and I) in the H3S structure with *Im*-3*m* symmetry, J. Phys. Chem. Solid. 139 (2020): 109315.
[46] V.V. Struzhkin, D.Y. Kim, E. Stavrou, T. Muramatsu, H. -k. Mao, C.J. Pickard, R. J. Needs, V.B. Prakapenka, A.F. Goncharov, Synthesis of sodium polyhydrides at high pressures, Nat. Commun. 7 (2016): 12267.
[47] A.P. Bartók, R. Kondor, G. Csányi, On representing chemical environments, Phys. Rev. B 87 (2013): 184115.
[48] D. Zagorac, H. Müller, R. Ruehl, J. Zagorac, S. Rehme, Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features, J. Appl. Crystallogr. 52 (2019) 918–925.
[49] National Institute of Materials Science, Materials Information Station, SuperCon http://supercon.nims.go.jp/index_en.html..
[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[51] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, P. Marquetand, wACSF-Weighted atom-centered symmetry functions as descriptors in machine learning potentials, J. Chem. Phys. 148 (2018): 241709.
[52] S. Rostami, M. Amsler, S.A. Ghasemi, Optimized symmetry functions for machine-learning interatomic potentials of multicomponent systems, J. Chem. Phys. 149 (2018): 124106.
[53] A.P. Bartók, M.C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons, Phys. Rev. Lett. 104 (2010): 136403.
[54] H. Yanxon, D. Zagaceta, B.C. Wood, Q. Zhu, Neural network potential from bispectrum components: a case study on crystalline silicon, J. Chem. Phys. 153 (2020): 054118.
[55] K. Choudhary, B. DeCost, Atomistic line graph neural network for improved materials property predictions, npj Comput. Mater. 7 (2021) 185.
[56] K. Choudhary, B. DeCost, F. Tavazza, Machine learning with force-field-inspired descriptors for materials: fast screening and mapping energy landscape, Phys. Rev. Mater. 2 (2018): 083801.
[57] L. Zhang, Y. Wang, J. Lv, Y. Ma, Materials discovery at high pressures, Nat. Rev. Mater. 2 (2017): 17005.
[58] A.R. Oganov, C.J. Pickard, Q. Zhu, R.J. Needs, Structure prediction drives materials discovery, Nat. Rev. Mater. 4 (2019) 331–348.
[59] Y. Yao, Theoretical methods for structural phase transitions in elemental solids at extreme conditions: statics and dynamics, J. Phys. Condens. Matter 34 (2022): 363001.
[60] Y. Wang, J. Lv, L. Zhu, Y. Ma, Crystal structure prediction via particle-swarm optimization, Phys. Rev. B 82 (2010): 094116.
[61] J. Lv, Y. Wang, L. Zhu, Y. Ma, Particle-swarm structure prediction on clusters, J. Chem. Phys. 137 (2012): 084104.
[62] C.W. Glass, A.R. Oganov, N. Hansen, USPEX-Evolutionary crystal structure prediction, Comput. Phys. Commun. 175 (2006) 713–720.
[63] Z. Falls, P. Avery, X. Wang, K.P. Hilleke, E. Zurek, The XtalOpt evolutionary algorithm for crystal structure prediction, J. Phys. Chem. C 125 (2020) 1601–1620.
[64] W.W. Tipton, R.G. Hennig, A grand canonical genetic algorithm for the prediction of multi-component phase diagrams and testing of empirical potentials, J. Phys. Condens. Matter 25 (2013): 495401.
[65] G. Trimarchi, A. Zunger, Global space-group optimization problem: finding the stablest crystal structure without constraints, Phys. Rev. B 75 (2007): 104113.
[66] Y. Yao, J.S. Tse, K. Tanaka, Metastable high-pressure single-bonded phases of nitrogen predicted via genetic algorithm, Phys. Rev. B 77 (2008): 052103.
[67] C.J. Pickard, R.J. Needs, Structures at high pressure from random searching, Phys. Status Solidi B 246 (2009) 536–540.
[68] C.J. Pickard, R.J. Needs, Ab initio random structure searching, J. Phys. Condens. Matter 23 (2011): 053201.
[69] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12562–12566.
[70] F.H. Stillinger, Exponential multiplicity of inherent structures, Phys. Rev. E 59 (1999) 48–59.
[71] Z. Zhang, T. Cui, M.J. Hutcheon, A.M. Shipley, H. Song, M. Du, V.Z. Kresin, D. Duan, C.J. Pickard, Y. Yao, Design principles for high-temperature superconductors with a hydrogen-based alloy backbone at moderate pressure, Phys. Rev. Lett. 128 (2022): 047001.
[72] H. Xie, Y. Yao, X. Feng, D. Duan, H. Song, Z. Zhang, S. Jiang, S.A.T. Redfern, V. Z. Kresin, C.J. Pickard, T. Cui, Hydrogen pentagraphenelike structure stabilized by hafnium: a high-temperature conventional superconductor, Phys. Rev. Lett. 125 (2020): 217001.
[73] H. Xie, W. Zhang, D. Duan, X. Huang, Y. Huang, H. Song, X. Feng, Y. Yao, C. J. Pickard, T. Cui, Superconducting zirconium polyhydrides at moderate pressures, J. Phys. Chem. Lett. 11 (2020) 646–651.
[74] W. Ibarra-Hernández, S. Hajinazar, G. Avendaño-Franco, A. Bautista-Hernández, A. N. Kolmogorov, A.H. Romero, Structural search for stable Mg-Ca alloys accelerated with a neural network interatomic model, Phys. Chem. Chem. Phys. 20 (2018) 27545–27557.
[75] E.V. Podryabinkin, E.V. Tikhonov, A.V. Shapeev, A.R. Oganov, Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning, Phys. Rev. B 99 (2019): 064114.
[76] Q. Tong, X. Luo, A.A. Adeleke, P. Gao, Y. Xie, H. Liu, Q. Li, Y. Wang, J. Lv, Y. Yao, Y. Ma, Machine learning metadynamics simulation of reconstructive phase transition, Phys. Rev. B 103 (2021): 054107.
[77] P.A. Santos-Florez, H. Yanxon, B. Kang, Y. Yao, Q. Zhu, Size-dependent nucleation in crystal phase transition from machine learning metadynamics, Phys. Rev. Lett. 129 (2022): 185701.
[78] P. Avery, X. Wang, C. Oses, E. Gossett, D.M. Proserpio, C. Toher, S. Curtarolo, E. Zurek, Predicting superhard materials via a machine learning informed evolutionary structure search, npj Comput, Mater 5 (2019) 1–11.
[79] C.J. Pickard, Ephemeral data derived potentials for random structure search, Phys. Rev. B 106 (2022): 014102.
[80] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. 98 (2007): 146401.
[81] A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, G.J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, J. Comput. Phys. 285 (2015) 316–330.
[82] X. Ming, Y.-J. Zhang, X. Zhu, Q. Li, C. He, Y. Liu, T. Huang, G. Liu, B. Zheng, H. Yang, J. Sun, X. Xi, H.-H. Wen, Absence of near-ambient superconductivity in $LuH_{2\pm x}N_y$, Nature 620 (2023) 72–77.
[83] S. Cai, J. Guo, H. Shu, L. Yang, P. Wang, Y. Zhou, J. Zhao, J. Han, Q. Wu, W. Yang, T. Xiang, H.-k. Mao, L. Sun, No evidence of superconductivity in the compressed sample prepared from the lutetium foil and H_2/N_2 gas mixture, Matter Radiat. Extremes 8 (2023): 048001.
[84] Đ. Dangić, P. Garcia-Goiricelaya, Y.-W. Fang, J. Ibañez-Azpiroz, I. Errea, *Ab initio* study of the structural, vibrational, and optical properties of potential parent structures of nitrogen-doped lutetium hydride, Phys. Rev. B 108 (2023): 064517.
[85] Z. Huo, D. Duan, T. Ma, Z. Zhang, Q. Jiang, D. An, H. Song, F. Tian, T. Cui, First-principles study on the conventional superconductivity of N-doped fcc-LuH_3, Matter Radiat. Extremes 8 (2023): 038402.
[86] I. Errea, M. Calandra, C.J. Pickard, J.R. Nelson, R.J. Needs, Y. Li, H. Liu, Y. Zhang, Y. Ma, F. Mauri, Quantum hydrogen-bond symmetrization in the superconducting hydrogen sulfide system, Nature 532 (2016) 81–84.
[87] I. Errea, F. Belli, L. Monacelli, A. Sanna, T. Koretsune, T. Tadano, R. Bianco, M. Calandra, R. Arita, F. Mauri, J.A. Flores-Livas, Quantum crystal structure in the 250-kelvin superconducting lanthanum hydride, Nature 578 (2019) 66–69.
[88] L. Monacelli, R. Bianco, M. Cherubini, M. Calandra, I. Errea, F. Mauri, The stochastic self-consistent harmonic approximation: calculating vibrational properties of materials with full quantum and anharmonic effects, J. Phys. Condens. Matter 33 (2021): 363001.
[89] Y. Xia, V.I. Hegde, K. Pal, X. Hua, D. Gaines, S. Patel, J. He, M. Aykol, C. Wolverton, High-throughput study of lattice thermal conductivity in binary rocksalt and zinc blende compounds including higher-order anharmonicity, Phys. Rev. X 10 (2020): 041029.
[90] F. Zhou, W. Nielson, Y. Xia, V. Ozoliņš, Lattice anharmonicity and thermal conductivity from compressive sensing of first-principles calculations, Phys. Rev. Lett. 113 (2014): 185501.
[91] I.A. Kruglov, A.V. Yanilkin, Y. Propad, A.B. Mazitov, P. Rachitskii, A.R. Oganov, Crystal structure prediction at finite temperatures, npj Comput. Mater. 9 (2023) 197.
[92] A.V. Shapeev, Multiscale Model. Simul. 14 (2016) 1153–1173.
[93] I. Errea, M. Calandra, C.J. Pickard, J. Nelson, R.J. Needs, Y. Li, H. Liu, Y. Zhang, Y. Ma, F. Mauri, High-pressure hydrogen sulfide from first principles: a strongly anharmonic phonon-mediated superconductor, Phys. Rev. Lett. 114 (2015): 157004.
[94] A.M. Shipley, M.J. Hutcheon, M.S. Johnson, R.J. Needs, C.J. Pickard, Stability and superconductivity of lanthanum and yttrium decahydrides, Phys. Rev. B 101 (2020): 224511.
[95] B. Monserrat, N.D. Drummond, P. Dalladay-Simpson, R.T. Howie, P.L. Ríos, E. Gregoryanz, C.J. Pickard, R.J. Needs, Structure and metallicity of phase V of hydrogen, Phys. Rev. Lett. 120 (2018): 255701.
[96] CALYPSO Database http://db.calypso.cn/.