

Application of machine learning methods for predicting new superhard materials

Cite as: J. Appl. Phys. **128**, 075102 (2020); <https://doi.org/10.1063/5.0012055>

Submitted: 27 April 2020 . Accepted: 24 July 2020 . Published Online: 17 August 2020

Efim Mazhnik , and Artem R. Oganov 



View Online



Export Citation



CrossMark

Lock-in Amplifiers
up to 600 MHz



Application of machine learning methods for predicting new superhard materials

Cite as: J. Appl. Phys. 128, 075102 (2020); doi: 10.1063/5.0012055

Submitted: 27 April 2020 · Accepted: 24 July 2020 ·

Published Online: 17 August 2020



Efim Mazhnik^{1,a)}  and Artem R. Oganov^{1,2,3,b)} 

AFFILIATIONS

¹Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, 3 Nobel Street, Moscow 121205, Russia

²Moscow Institute of Physics and Technology, 9 Institutsky Lane, Dolgoprudny 141700, Russia

³International Center for Materials Discovery, Northwestern Polytechnical University, Xi'an 710072, China

Note: This paper is part of the special collection on Machine Learning for Materials Design and Discovery

a) Author to whom correspondence should be addressed: efim.mazhnik@skoltech.ru

b) a.oganov@skoltech.ru

ABSTRACT

Superhard materials are of great interest in various practical applications, and an increasing number of research efforts are focused on their development. In this article, we demonstrate that machine learning can be successfully applied to searching for such materials. We construct a machine learning model using neural networks on graphs together with a recently developed physical model of hardness and fracture toughness. The model is trained using available elastic data from the Materials Project database and has good accuracy for predictions. We use this model to screen all crystal structures in the database and systematize all the promising hard or superhard materials, and find that diamond (and its polytypes) are the hardest materials in the database. Our results can be further used for the investigation of interesting materials using more accurate *ab initio* calculations and/or experiments.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0012055>

I. INTRODUCTION

Hardness is the ability of a material to resist localized plastic deformations. Such deformations can include indentation, scratching, densification, and fracture. Superhard materials (those that have hardness above 40 GPa) are in need in many industrial applications including cutting, drilling, and polishing technologies. Their practical importance prompted significant efforts toward the development of such new materials over the past several decades.¹

With recent developments in computational materials science,² it is now possible to predict new materials without performing many trial-and-error experiments. This is usually done either by searching among existing materials using crystal structure databases or by means of computer algorithms capable of creating completely new structures and evaluating their suitability.^{3–5} These methods significantly facilitated the development of the materials with the desired properties, sometimes very exotic. The prominent examples include H₃S^{6,7} and ThH₁₀^{8,9}—superconductors with some of the highest critical temperatures, the highest-T_c superconductor LaH₁₀,^{10–13} the high-pressure transparent form of sodium,¹⁴ the unusual helium compound

Na₂He,¹⁵ Sr₅P₃—the novel electride material,¹⁶ and three novel high-*k* dielectric polymers.¹⁷

In order to use any of the prediction methods, we need to specify exactly what are the desired properties that we need to optimize and have a method for evaluating them from the crystal structure. In the case of superhard materials, these properties are usually the hardness and the fracture toughness.

There are various scales to determine the hardness. In this article, we will focus on the Vickers hardness, which is one of the most common and easy to use experimentally. The Vickers hardness is defined as the resistance to the deformation caused by pyramidal diamond indenter in which two edges of pyramidal face intersect at 136° (Fig. 1).

The value is determined by the ratio

$$H = \frac{F}{A} = \frac{2F \sin(17\pi/45)}{d^2}, \quad (1)$$

where F is the applied force, A is the indentation area, and d is the average length of the diagonal of indentation.

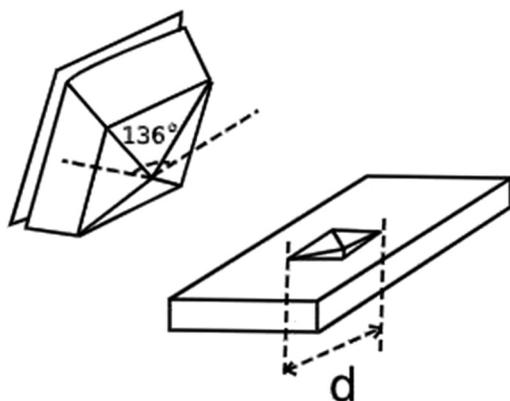


FIG. 1. A schematic view of the Vickers hardness indentation test.

Fracture toughness is defined as the resistance of the material to the propagation of a crack. It can be measured directly by intentionally introducing a crack on a sample and recording the tensile stress at which the crack propagates or indirectly, by noticing the moment during the Vickers hardness test at which a crack begins to form.

Knowing both of these parameters, we can search for the materials in the upper-right corner of the hardness vs fracture toughness plot (Fig. 2).

There are not many materials that can be regarded as superhard. Diamond stands out as the hardest known material with the Vickers

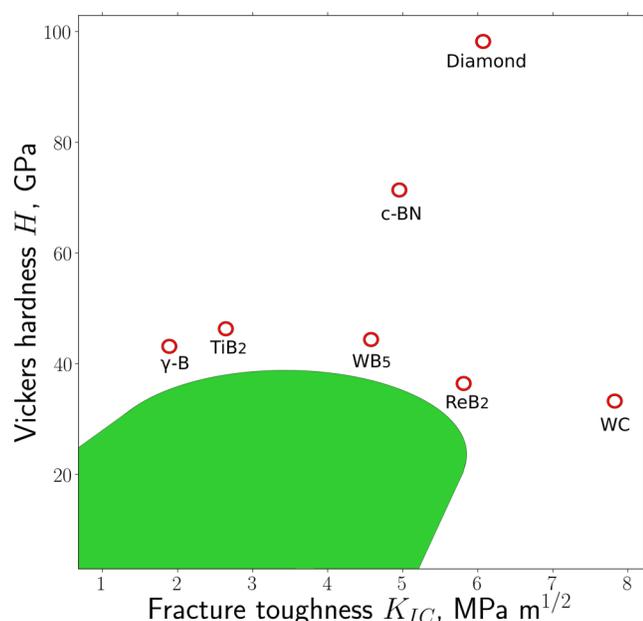


FIG. 2. Plot of the hardness vs fracture toughness. Most of the materials are in the bottom-left corner.

hardness around 90 GPa and also with a high fracture toughness. There are other carbon allotropes with very high, but, nevertheless, lower hardness (e.g., Ref. 18). Tungsten carbide (WC), apart from the high value of hardness, has an outstanding fracture toughness, and due to high hardness and fracture toughness, this material has many applications, especially as the main component of hard alloy composites. Although it has been argued¹⁹ and even computationally demonstrated^{20,21} that no material can be harder than diamond (with lonsdaleite having the same hardness as diamond), some materials that have lower hardness but other attractive characteristics (cost, fracture toughness, thermal or chemical stability, electrical conductivity) can find their applications in many industries.

Unfortunately, there are no accurate theoretical models that could be used to calculate the hardness and fracture toughness from the crystal structure. The main reason is that the underlying physical principles complex and include both elastic and plastic effects. Another reason is that the properties itself are ill-defined and can depend on the different experimental factors including the applied load, surface roughness, concentration of defects in the sample, loading time, shape of the indenter, degree of elastic recovery, etc.

However, there are several approximate models^{21–25} which allow estimation of hardness and fracture toughness from the other properties of the crystal. We have recently developed a model of such type,²⁶ which turned out to be particularly useful. Using this model, which is based on the physical principles, we can estimate the Vickers hardness

$$H = \frac{0.096E(1 - 8.5\nu + 19.5\nu^2)}{1 - 7.5\nu + 12.2\nu^2 + 19.6\nu^3} \quad (2)$$

and fracture toughness

$$K_{IC} = \alpha_0^{-1/2} V_0^{1/6} \left(\frac{E(1 - 13.7\nu + 48.6\nu^2)}{1 - 15.2\nu + 70.2\nu^2 - 81.5\nu^3} \right)^{3/2}, \quad (3)$$

where E is Young's modulus, ν is Poisson's ratio, V_0 is the volume per atom, and α_0 is a constant that equals $\alpha_0 = 8840$ GPa for ionic and covalent materials, but is smaller for typical free-electron metals.

Essentially, we reduced the calculation of the hardness-related properties H and K_{IC} to the elastic properties E and ν . The latter can be calculated with the quantum-mechanical equations; however, these calculations are very time-consuming and not suitable for screening large databases.

Machine learning methods are good candidates for solving such problems. Over the last decade, these methods were extensively applied in different areas, such as computer vision, recommendation systems, speech recognition, and data extraction. As a result, a big part of our daily lives is governed by different machine learning algorithms. They have also begun to be actively used in materials science. Recent examples include the calculation of different properties of materials^{27–29} and accelerating quantum calculations.³⁰

Such methods had also been already used to identify superhard materials. For example, in Ref. 31, the famous Teter relationship³² was used to determine the value of hardness. However, such a relationship is known to be inaccurate, especially for materials with a pronounced plastic effect.²⁶ Moreover, the authors made no

attempt to screen a large number of structures. In Ref. 33, the authors used elastic moduli as a proxy to identify superhard compounds. However, as in the previous case, the connection between hardness and elasticity is more complicated. Both works made no estimation of fracture toughness, which is also a very important parameter for superhard compounds.

The main idea of machine learning models is to use a dataset with the already known values (in our case, E and ν obtained by quantum-mechanical calculations) and to expand it to the unknown data. By passing the dataset through the model, we can optimize its internal parameters to reduce the error and to obtain reasonable results. The whole process is usually divided into epochs in each of which the error is evaluated for the particular random subset of data and the corresponding corrections of internal parameters are made. In some sense, these parameters are related to the patterns found by the model in the training data, which influence the target properties. After the model is trained, the calculation of the property is almost instant compared to the rigorous quantum approach. It allows us to use the big databases of crystal structures for screening for the compounds with the optimal properties.

It is important to mention that it seems reasonable to use the dataset for hardness and fracture toughness values directly bypassing the additional step of converting the E and values into H and K_{IC} . However, to the best of the authors' knowledge, no such dataset with sufficient amount of structures exists at the current moment.

II. REPRESENTATION OF THE CRYSTAL STRUCTURE

In order to use any of the methods of machine learning, we need to decide how to represent the crystal structure as a set of numbers to use it further in the model. This set is usually called a descriptor, or a feature vector, or a fingerprint. The naive approach would be to represent the structure in terms of Cartesian coordinates of each atom. However, this approach will not produce any reasonable results, because in this case, structures that are in fact equivalent may have very different descriptors. The ideal descriptor

will stay the same under the operations of rotations, reflection, translation, and permutation of the chemically equivalent atoms because they are not changing the crystal structure and thus not changing the properties. In addition, the length of the descriptor should be the same and independent from the size of a structure.

Many efforts were put into the development of such descriptions. In particular, they are widely used in the construction of machine learning interatomic potentials.^{34–36} However, such descriptors are usually difficult to apply to problems where the structures in the dataset have many atom types. Furthermore, these descriptors are particularly optimized for predicting the energy, which has additional useful properties. For example, for short-range interactions, it can be decomposed as the sum of energies from the local environments of each atom.

In the case of hardness and fracture toughness, it is harder to create specific descriptors because no underlying physical phenomena are fully understood. The problem can be solved by using graph convolutional networks. These networks can treat descriptors themselves as an internal parameter and thus can optimize them for particular properties, without any external supervision from humans.

In this method, atom and bond properties are encoded in node feature vectors v_i and edge feature vectors u_{ij} . The node feature vector is the unique characteristic of the type of atom (not related to its spatial position!). For example, in the simplest case, it can encode the atomic number of the element in the periodic table ([1, 0, ...] for hydrogen, [0, 1, ...] for helium, etc.). It can also contain other properties of the elements such as number of valence electrons, atomic mass, electronegativity, heat capacity, thermal conductivity, and so on. This additional information can become very useful with the small number of structures in the dataset; for large datasets, it becomes irrelevant. The edge feature vector characterizes the distance between nodes. We will use the Gaussian approximation with constant step s and cut-off distance R_{cut} . In this case, if the distance between atoms i and j is equal to d , the edge feature vector is

$$u_{ij} = \left[\exp\left(-\frac{(d-0 \cdot s)^2}{s^2}\right), \exp\left(-\frac{(d-1 \cdot s)^2}{s^2}\right), \dots, \exp\left(-\frac{(d-R_{\text{cut}} \cdot s)^2}{s^2}\right) \right]. \quad (4)$$

The crystal structure is periodic in space generated from the collection of unique sites. We can always consider only unique sites of the structure and will give the same results. Also, we can take a fixed number of nearest neighbors for each atom for which edge features will be considered.

Let us analyze a specific example. Tungsten carbide (WC) (Fig. 3) has a hexagonal crystal structure with two unique sites, corresponding to tungsten and carbon atoms.

The feature vectors can be taken to be [1, 0] and [0, 1]. Let the number of neighbors be 8. For any tungsten atom, we have six neighboring carbon atoms and two other tungsten atoms. Distances are [2.212, 2.212, 2.212, 2.212, 2.212, 2.212, 2.853, 2.853] Å. If we let s to be 1 Å and R_{cut} be 4 Å, the corresponding Gaussian distances

will be six times [0.0075, 0.2301, 0.9560, 0.5375, 0.0409] Å and 2 times [0.0003, 0.0323, 0.4831, 0.9786, 0.2683] Å. We can notice that the first type of distance concentrates around the third index (meaning 2 Å) and the second type around the fourth (meaning 3 Å). For any carbon atom, we have six neighboring tungsten atoms and two other carbon atoms. Distances and corresponding Gaussian distances are the same. In the current work, we chose R_{cut} to be 8 Å. We did not observe any noticeable influence on the accuracy of predictions if we increase this value further.

Similar to the ordinary neural networks, graph convolutional networks have several types of layers including linear, activation, convolution, and pooling layers.

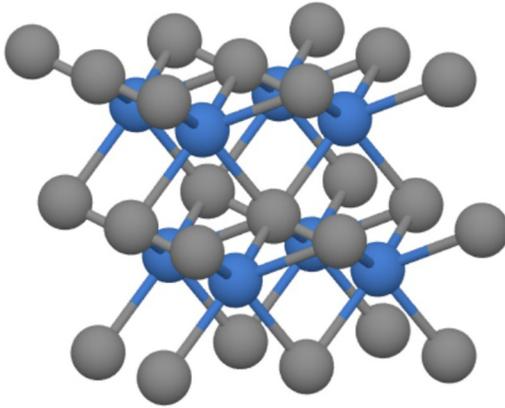


FIG. 3. Crystal structure of tungsten carbide (WC). Tungsten atoms are shown in blue color and carbon atoms are shown in gray.

The linear and activation layers perform simple operations on the separate feature vectors, while the convolutional layer is more complicated and can be constructed in many ways. We adopted the definition from Ref. 37, which emphasizes the differences of the interaction strength between the neighbors

$$v_i^{t+1} = v_i^t + \sum_j \sigma(z_{ij}^t W_j^t + b_j^t) \odot \sigma(z_{ij}^t W_j^t + b_j^t), \quad (5)$$

where σ is a sigmoid activation function, \odot is an element-wise multiplication, and z_{ij}^t a vector, which is formed by the concatenation of v_j^t , v_j^t , and u_{ij}^t .

The pooling layer is where the descriptor is formed. It is the normalized summation of all the feature vectors. We can stack several layers of such types, which will define the transformation of the crystal structure into a one-dimensional vector (descriptor) governed by the unknown coefficients. After the pooling layer, we can use the ordinary neural network layers to obtain the final property. By optimizing the coefficients, we can “teach” the network to predict the correct values of the target properties. In this way, the descriptor becomes an internal parameter of the neural network.

III. RESULTS AND DISCUSSION

In order to train our network, we used the database of crystal structures by Materials Project³⁸ obtained via Python Materials Genomics (pymatgen)³⁹ package. This database contains 124 515 inorganic crystal structures (as of December 2019) with properties calculated using density functional theory (DFT).^{40,41} The central property for our work is the stiffness (elastic) tensor C , which fully characterizes the elastic behavior of a material via relationship

$$\sigma_{ij} = C_{ijkl} \epsilon_{kl}, \quad (6)$$

where σ_{ij} and ϵ_{kl} are the components of the stress and strain tensors, respectively. Due to symmetry constraints

$$C_{ijkl} = C_{jikl} = C_{ijlk} = C_{klij}, \quad (7)$$

elastic tensor has only 21 independent components. Therefore, it is reasonable to represent this tensor as a 6×6 symmetric matrix. By applying perturbations to the lattice vectors and measuring the resulting stresses, Materials Project has already calculated elastic tensors for 8033 structures (ignoring those which have warnings). Given the compliance matrix S , which is inverse to the stiffness matrix, we can calculate the aggregate elastic moduli using Voigt–Reuss–Hill approximation.⁴² First, we obtain the bulk modulus B which is determined as an average between Voigt average B_V and Reuss average B_R ,

$$9B_V = (C_{11} + C_{22} + C_{33}) + 2(C_{12} + C_{23} + C_{31}), \quad (8)$$

$$1/B_R = (S_{11} + S_{22} + S_{33}) + 2(S_{12} + S_{23} + S_{31}), \quad (9)$$

and shear modulus G , which is also an average of G_V and G_R ,

$$15G_V = (C_{11} + C_{22} + C_{33}) - (C_{12} + C_{23} + C_{31}) + 3(C_{44} + C_{55} + C_{66}), \quad (10)$$

$$15/G_R = 4(S_{11} + S_{22} + S_{33}) - 4(S_{12} + S_{23} + S_{31}) + 3(S_{44} + S_{55} + S_{66}). \quad (11)$$

After that, we can calculate Young’s modulus

$$E = \frac{9BG}{3B + G} \quad (12)$$

and Poisson’s ratio

$$\nu = \frac{3B - 2G}{6B + 2G}. \quad (13)$$

In principle, we could predict the whole elastic tensor by using the neural network. However, it is too large (given the size of the available training dataset) to maintain the desired accuracy for the target properties. Also, Poisson’s ratio is too noisy and thus it is undesirable to predict it directly. Therefore, we decided to predict the pair (B, G) and associate it with an output of our network. The pair (E, ν) can be easily calculated after that.

The overall number of atom types in the dataset is 63. Except for oxygen, they are distributed pretty evenly (Fig. 4), which reduces the bias of the network toward specific types.

The whole network consists of nine layers (Fig. 5). Each atom is described by a vector of size 63. After the first linear (embedding) layer, the number of values describing each atom reduces to the same number—40. It allows the network to optimize the representation of chemical elements for specific properties. At the next step, three convolutions are performed and after pooling, the whole structure is described by one vector of size 40. Essentially, at this step, we obtained a descriptor for the structure. After that, two linear transformations with softmax activation layers are performed to obtain the final vector of size 2.

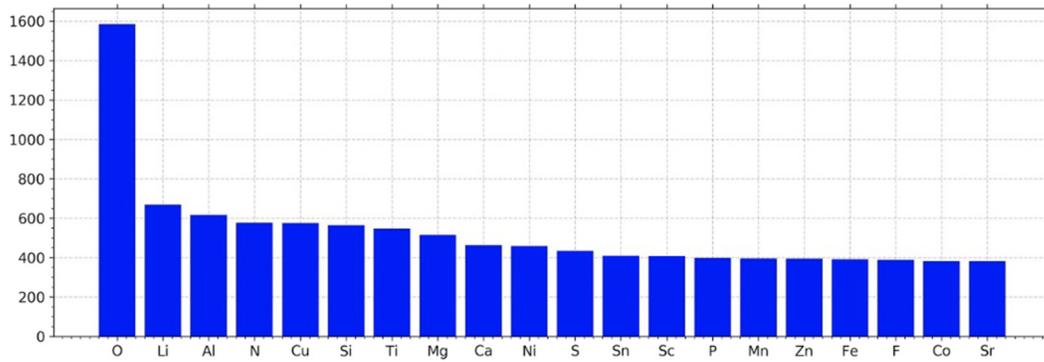


FIG. 4. Distribution of the most frequent chemical elements in the dataset.

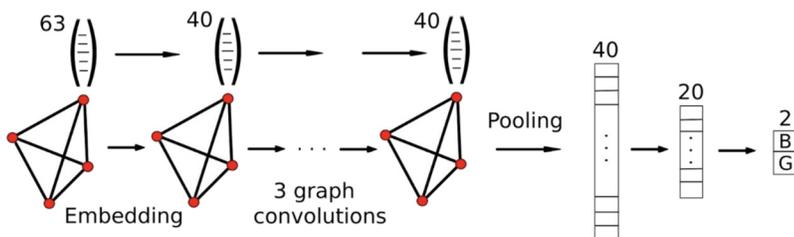


FIG. 5. The scheme of the graph neural network used in the current article.

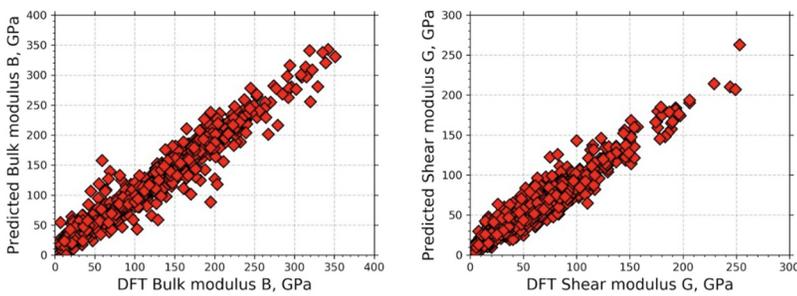


FIG. 6. Results for the bulk modulus B and shear modulus G on the validation dataset.

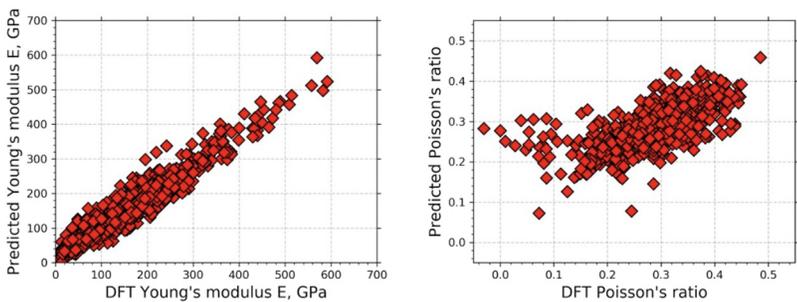


FIG. 7. Results for Young's modulus E and Poisson's ratio ν on the validation dataset.

TABLE I. The average values of target properties together with the errors on the validation dataset. MAE stands for mean absolute error and RMSE stands for root mean square error.

	Average, GPa	MAE, GPa	RMSE, GPa
B	111.83	11.11	19.54
G	54.81	8.24	11.43
E	138.95	19.15	26.23
ν	0.286	0.041	0.105

This vector contains the predicted bulk modulus B and shear modulus G . This architecture was specifically optimized for these target parameters.

The whole dataset of 8033 structures was divided into a training dataset and a validation dataset containing 7229 and 804 crystal structures, respectively. This was done in order to reduce overfitting, which can happen when training deep neural networks.

The network was constructed on the graphical processing unit (GPU) using PyTorch package. Initially, all layers in the network were filled with random numbers. The process of training the network is the process of optimizing these numbers. The optimization is performed using back-propagation algorithm, which minimizes the difference between predicted and DFT-calculated values with respect to the loss (error) function. The loss function \mathcal{L} is defined as

$$\mathcal{L} = \sum (B_{\text{pred}} - B_{\text{DFT}})^2 + \sum (G_{\text{pred}} - G_{\text{DFT}})^2, \quad (14)$$

where “pred” and “DFT” denote the predicted and DFT-calculated values, respectively. The whole training process is divided into epochs. In each epoch, a small subset of the dataset containing 256 structures is used to calculate the loss function and to update the coefficients in layers of the network by using Adam optimizer⁴³ with a learning rate of 0.001 and a zero weight decay. The total number of epochs is 5000. After that, we took the coefficients of the layers that gave a minimal value of the loss function on the validation dataset.

TABLE II. Results of network predictions for several known compounds and their corresponding experimental values. id denotes id of the material in Materials Project database and the energy above convex hull per atom d_{hull} is provided for understanding of stability. H denotes the Vickers hardness and K_{IC} denotes the fracture toughness. Experimental values are obtained from Ref. 26.

	id	Space group	d_{hull} , eV	E_{pred} , GPa	ν_{pred}	H_{pred} , GPa	H_{exp} , GPa	$K_{IC\text{pred}}$, MPa m ^{1/2}	$K_{IC\text{exp}}$, MPa m ^{1/2}
Diamond	66	227	0.135	1 124	0.063	101	96	6.2	5.3–6.7
WC	1 894	187	0.000	640	0.208	32	27	6.1	7.5
c-BN	1 639	216	0.076	801	0.136	62	66	5.1	5.0
Y ₂ O ₃	2 652	206	0.000	161	0.301	8	6	1.2	0.7
B ₄ C	69 6746	166	0.038	341	0.213	17	30	2.4	3.1–3.7
Si	149	227	0.000	113	0.279	6	12	0.7	0.8–1.0
TiN	492	225	0.000	291	0.306	15	18	2.7	3.4–5.0
TiC	631	225	0.000	388	0.218	19	29	2.9	2.0–3.8
AlN	661	186	0.000	300	0.241	14	18	2.2	2.8
Ge	32	227	0.000	92	0.275	5	8	0.5	0.6
α -SiO ₂	6 945	92	0.006	82	0.191	5	11	0.3	—
GaAs	2 534	216	0.000	83	0.281	4	7	0.4	0.4

The comparison between predicted and DFT-calculated values on the validation dataset is presented in Fig. 6 for bulk modulus and shear modulus and in Fig. 7 for Young’s modulus and Poisson’s ratio. Also, we provide the errors for each case (Table I). These results are comparable with those obtained by other machine learning methods.^{28,44} We can notice that predictions for B , G , and E work quite well while for Poisson’s ratio the network overestimates predictions for low values on the validation dataset. This can be connected with the fact that materials of this type are under-represented in the database.

After we developed the model for prediction of Young’s modulus and Poisson’s ratio, we were able to estimate the Vickers hardness and the fracture toughness using the expressions (2) and (3). The results for several known crystal structures are presented in Table II. As we can see, predictions are reasonable and models are definitely capable of distinguishing superhard materials. However, the amount of experimental data is not sufficient for direct comparison with the predictions; instead, we compare our machine learning model with hardnesses and fracture toughnesses calculated from the known elastic properties. We calculated the hardness and the fracture toughness using elasticity data from the Materials Project. In Table III, we list the compounds, which have high values of both properties. We also provide the corresponding predictions of the network in order to understand the accuracy. The comparison between the predicted values with those calculated by our model from known elastic data is also depicted in Fig. 8.

Table III contains all entries from Materials Project with sufficiently high predicted hardness and fracture toughness, and for some compounds (e.g., C, SiC, etc.), many entries are present. These correspond to different phases (some of which are hypothetical). Entries corresponding to different polytypes of the same compound have very similar properties (e.g., diamond, lonsdaleite, and other polytypes) and sometimes even the same space group—such structures differ only in the stacking sequence of layers. For the sake of completeness, we did not discard any of these cases.

The next step is to use the model for phases, for which no elastic properties are known and try to find the materials which

TABLE III. Examples of materials that have high values of both hardness and fracture toughness according to the elastic data from Materials Project and our model. *id* denotes id of the material in Materials Project database and the energy above convex hull per atom d_{hull} is provided for understanding of stability. E , ν , H , and K_{IC} stand for Young's modulus in GPa, Poisson's ratio, the Vickers hardness in GPa, and the fracture toughness in $\text{MPa m}^{1/2}$, respectively with values calculated using the elastic tensor from Materials Project. H_{pred} and $K_{IC\text{pred}}$ correspond to the Vickers hardness and fracture toughness predicted by the network. By \star , we denote intermetallics, for which values of hardness are lower and values of fracture toughness are usually much higher than those calculated by our empirical model.

	id	Space group	d_{hull} , eV	E , GPa	ν	H , GPa	K_{IC} , $\text{MPa m}^{1/2}$	H_{pred} , GPa	$K_{IC\text{pred}}$, $\text{MPa m}^{1/2}$
C (lonsdaleite)	47	194	0.160	1123	0.070	100	6.2	102	6.3
C	616 440	194	0.141	1118	0.072	99	6.2	102	6.2
C (diamond)	66	227	0.135	1117	0.073	99	6.2	101	6.2
C	569 567	166	0.174	1118	0.072	99	6.2	101	6.2
C	611 426	194	0.145	1121	0.071	99	6.2	103	6.3
C	24	206	0.833	1055	0.046	96	5.4	93	5.6
BC ₂ N	30 148	17	0.542	912	0.079	80	4.7	77	4.7
C	1 008 395	139	0.332	935	0.113	77	5.3	77	4.9
BC ₂ N	629 458	25	0.537	891	0.089	77	4.6	79	4.9
BC ₂ N	1 008 523	115	0.995	862	0.086	75	4.4	63	4.8
BN (h-BN)	2 653	186	0.093	856	0.118	70	4.7	63	5.1
C ₃ N ₄	571 653	215	0.489	866	0.134	67	5.6	70	5.2
C ₃ N ₄	2 852	220	0.493	874	0.150	64	7.3	65	6.4
BN (c-BN)	1 639	216	0.076	860	0.149	63	7.0	62	5.1
BC ₅	1 018 649	156	0.269	797	0.148	58	6.2	76	4.7
BeCN ₂	15 703	122	0.004	669	0.147	49	4.7	54	3.9
TiB ₂	1 145	191	0.000	569	0.125	45	2.9	48	2.6
CrB ₄	27 710	71	0.004	590	0.145	44	4.0	44	2.9
HfB ₂	1 994	191	0.000	550	0.135	43	3.1	45	2.7
C ₃ N ₄	1 985	176	0.325	747	0.195	41	6.8	58	6.4
MnB ₄	1 010	71	0.019	551	0.148	40	3.7	39	3.9
N	999 498	199	0.000	561	0.161	39	4.2	36	4.3
ZrB ₂	1 472	191	0.000	515	0.139	39	3.1	41	2.7
ReB ₂	1 773	194	0.000	639	0.184	38	5.6	39	5.4
Si ₃ N ₄	2 075	227	0.147	582	0.170	38	4.7	35	3.9
TcB ₂	1 019 317	194	0.000	568	0.165	38	4.6	27	4.5
B ₂ CN	1 008 527	115	0.311	617	0.183	37	5.0	29	4.1
VB ₂	1 491	191	0.000	565	0.171	37	4.5	38	4.2
VN	1 002 105	221	0.677	536	0.162	37	4.1	29	5.2
Mg(B ₆ C) ₂	568 803	74	0.000	489	0.143	37	2.9	23	2.8
Ti ₃ B ₄	1 025 170	71	0.000	486	0.139	37	2.8	41	2.4
V ₂ B ₃	9 208	63	0.000	555	0.170	36	4.4	40	4.1
VB	9 973	63	0.000	543	0.168	36	4.3	42	4.0
V ₃ B ₄	569 270	71	0.000	545	0.174	35	4.3	40	4.0
B ₆ O	1 346	166	0.000	477	0.149	35	3.0	34	2.1
WC	1 894	187	0.000	674	0.208	34	6.6	32	6.1
CN ₂	1 009 818	119	0.720	697	0.212	34	6.4	49	5.3
CrB	260	63	0.010	522	0.173	34	4.0	34	4.2
AlN	1 330	225	0.172	508	0.168	34	3.8	29	3.2
HfNbB ₄	38 818	71	0.000	517	0.180	32	4.1	35	4.2
TiB	7 857	62	0.000	436	0.149	32	2.8	30	2.9
ZrB ₁₂	1 084	225	0.017	468	0.166	31	3.3	33	2.8
TaB	1 097	63	0.000	503	0.185	30	4.1	27	4.0
Si ₂ W	1 620	139	0.000	440	0.165	30	3.3	22	3.1
SiC	570 690	160	0.032	434	0.161	30	3.1	31	3.0
SiC	568 735	156	0.000	434	0.161	30	3.1	31	3.0
SiC	570 791	156	0.000	433	0.162	30	3.1	31	3.0
SiC	568 696	156	0.000	434	0.160	30	3.1	31	3.0
SiC	570 641	160	0.031	434	0.161	30	3.1	31	3.0

TABLE III. (Continued.)

	id	Space group	d_{hull} , eV	E , GPa	ν	H , GPa	K_{IC} , MPa m ^{1/2}	H_{pred} , GPa	K_{ICpred} , MPa m ^{1/2}
SiC	567 551	186	0.000	434	0.161	30	3.1	31	3.0
SiC	9 947	160	0.003	434	0.161	30	3.1	31	3.0
SiC	567 505	186	0.000	435	0.161	30	3.1	31	3.0
SiC	11 714	186	0.000	434	0.161	30	3.1	31	3.0
SiC	570 985	156	0.000	434	0.161	30	3.1	31	3.0
SiC	568 619	156	0.001	433	0.161	30	3.1	31	3.0
SiC	582 034	156	0.000	434	0.161	30	3.1	31	3.0
SiC	7 631	186	0.000	434	0.161	30	3.1	31	3.0
SiC	568 656	156	0.000	434	0.161	30	3.1	31	3.0
ScB ₁₂	8 772	225	0.000	438	0.165	30	3.0	24	2.6
SiC	8 062	216	0.001	433	0.158	30	3.0	31	3.0
Be ₃ N ₂	18 337	206	0.000	433	0.162	30	2.9	29	2.8
Os	49	194	0.000	639	0.235	29	6.9	29	6.9
ReN ₂	1 019 055	127	0.503	622	0.227	29	5.9	30	6.0
CrC	1 018 050	187	0.077	591	0.212	29	5.2	25	4.8
B ₂ Mo	2 331	166	0.000	537	0.196	29	4.4	29	4.4
SiC	7 140	186	0.004	430	0.163	29	3.1	31	3.0
B ₄ C	530 074	1	0.064	435	0.168	29	3.0	17	2.4
B ₈ O	758 933	12	0.026	418	0.160	29	2.7	31	2.3
Os	8 643	225	0.133	582	0.262	28	6.7	29	6.8
ReC	1 009 735	187	0.267	589	0.262	28	6.4	27	5.8
TcOs ₃ ★	867 212	194	0.000	601	0.235	28	6.3	32	6.6
MoC	2 305	187	0.002	586	0.221	28	5.5	24	4.9
TaN	1 459	187	0.027	578	0.214	28	5.3	27	5.1
Si ₂ Mo	2 592	139	0.000	413	0.167	28	3.0	21	3.0
CrSi ₂	8 937	139	0.000	399	0.159	28	2.7	21	2.6
B	161	166	0.026	401	0.160	28	2.6	35	2.1
GaB ₃ N ₄	1 019 740	215	0.398	558	0.217	27	4.8	24	4.2
V ₃ B ₂	2 091	127	0.000	475	0.191	27	3.7	26	3.7
MgSiN ₂	1 017 514	166	0.282	421	0.171	27	2.9	24	2.9
GaBN ₂	1 007 823	115	0.383	410	0.168	27	2.8	25	3.0
WC	1 008 630	221	0.918	571	0.235	26	5.5	28	6.5
WN ₂	999 549	187	0.029	559	0.236	26	5.3	19	3.5
B ₂ W	569 803	194	0.000	545	0.217	26	4.8	25	4.2
C	1 008 374	65	0.437	546	0.216	26	4.5	44	4.2
C	570 002	229	0.763	535	0.214	26	4.4	24	4.4
SiO ₂ (stishovite)	6 947	136	0.196	489	0.200	26	3.8	30	4.0
BeSiN ₂	7 913	33	0.000	406	0.174	26	2.8	22	2.9
Re ₃ Ir ★	974 430	139	0.017	521	0.266	25	5.8	24	5.5
OsN ₂	21 264	58	0.272	554	0.243	25	5.3	27	5.1
BN	644 751	62	0.271	521	0.213	25	4.2	38	5.1
NbB	2 580	63	0.000	474	0.201	25	3.9	25	3.8

have a high values of both Vickers hardness and fracture toughness. In Table IV, we present the structures that contain no elastic properties in Materials Project database but were identified as having a high value of hardness and fracture toughness. In this list, we see especially many borides, a bit fewer carbides and nitrides—confirming conclusions of Ref. 2 that very hard materials are more likely to be found among metal borides than metal carbides or nitrides. For convenience, we plot all these structures as well as structures with known elastic data, on one Ashby plot (Fig. 9).

Some of the presented structures are metastable and some are hypothetical. Moreover, some part of these structures can be a result of poor convergence during calculations and special cases should be investigated further. While having good accuracy in general, the network can also give worse results for some structures, belonging to classes that are underrepresented in the database. For example, it is hardly possible that Fe₂C₄Cl₂O₉ is a superhard compound. Still, many of the structures look promising and this suggests that this method can be used with other data or algorithms to

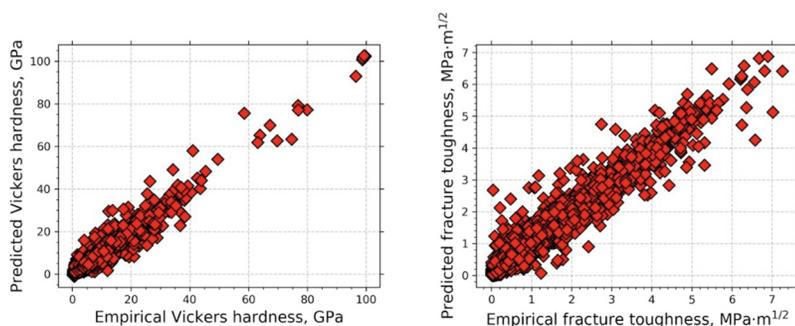


FIG. 8. Comparison between the predictions of the network and values calculated using the known elastic data for hardness and fracture toughness.

TABLE IV. Materials predicted by the neural network to have high values of both Vickers hardness (≥ 25 GPa) and fracture toughness (≥ 2.5 MPa m^{1/2}) and for which there are no elastic data in Materials project database. id denotes id of the material in Materials Project database and the energy above convex hull per atom d_{hull} is provided for understanding of stability. E , ν , H , and K_{IC} stand for Young's modulus in GPa, Poisson's ratio, the Vickers hardness in GPa, and the fracture toughness in MPa m^{1/2}, respectively. By \star , we denote intermetallics, for which values of hardness are lower and values of fracture toughness are usually higher than those calculated by our empirical model.

	id	Space group	d_{hull} , eV	E_{pred} , GPa	ν_{pred}	H_{pred} , GPa	$K_{IC\text{pred}}$, MPa m ^{1/2}
C	569 517	166	0.145	1 140	0.061	102	6.3
C	611 448	194	0.142	1 137	0.061	102	6.2
C	1 190 171	62	0.290	1 129	0.059	102	6.2
C	1 080 826	12	0.298	1 110	0.061	100	6.1
C	1 078 845	65	0.265	1 085	0.067	97	5.9
C ₁₁ N ₄	1 104 073	111	0.305	1 006	0.078	88	5.4
BC ₇	1 078 935	115	0.208	1 003	0.075	88	5.4
BC ₇	1 080 030	156	0.242	970	0.084	84	5.2
BC ₇	1 079 661	160	0.240	967	0.085	84	5.2
BC ₇	1 079 046	25	0.240	959	0.083	83	5.1
BC ₅	1 077 125	119	0.235	941	0.083	82	5.0
BC ₇	1 095 030	215	0.274	903	0.092	77	4.7
BC ₅	1 077 743	44	0.264	891	0.093	76	4.7
BC ₅	1 095 514	2	0.294	849	0.098	72	4.4
C ₁₁ N ₄	1 104 513	16	1.247	870	0.114	71	4.8
BC ₂ N	1 079 201	8	0.888	833	0.103	70	4.3
B ₂ CN ₂	1 228 638	160	0.461	790	0.131	62	4.7
C	1 188 817	221	0.763	747	0.119	61	4.0
C ₃ N ₄	9 410	159	0.288	798	0.146	59	6.1
BC ₃	1 239 206	217	0.269	735	0.122	59	3.9
N	1 056 857	225	0.000	814	0.165	55	7.6
B ₂ (CN ₂) ₃	989 468	2	0.439	705	0.136	54	4.3
H(C ₃ N ₂) ₄	976 247	221	0.454	643	0.116	53	3.1
C ₂ N ₃	1 078 791	36	0.546	716	0.149	52	5.5
BeCN ₂	1 189 451	33	0.000	663	0.135	51	3.8
SrC ₇ N ₁₀	1 245 820	7	0.337	683	0.149	50	5.2
PbC ₇ N ₁₀	1 246 015	7	0.348	685	0.149	50	5.2
BaC ₇ N ₁₀	1 245 412	7	0.378	680	0.148	50	5.0
VC ₃	1 067 129	221	2.503	743	0.168	49	6.7
MoB ₄	1 228 687	191	0.260	632	0.135	49	3.8
C ₃ N ₂	1 105 655	221	0.256	596	0.116	49	2.8
CaC ₇ N ₁₀	1 247 296	7	0.334	667	0.151	48	5.1
C ₃ N ₂	1 188 347	215	0.256	592	0.115	48	2.7
HfTiB ₄	1 224 263	191	0.013	565	0.111	47	2.7
CrB ₄	1 078 278	58	0.000	576	0.118	47	2.7

TABLE IV. (Continued.)

	id	Space group	d_{hull} , eV	E_{pred} , GPa	v_{pred}	H_{pred} , GPa	K_{ICpred} , MPa m ^{1/2}
CN ₂	1 018 655	164	0.588	600	0.140	46	3.8
TiVB ₄	1 216 667	191	0.004	574	0.123	46	2.8
ZrTiB ₄	1 215 178	191	0.018	549	0.114	45	2.6
MnB ₄	1 078 253	58	0.010	573	0.138	44	3.3
HfZrB ₄	1 224 184	47	0.004	544	0.121	44	2.7
TaTiB ₄	1 217 898	191	0.000	572	0.140	43	3.6
MnB ₄	1 106 184	14	0.000	563	0.145	42	3.6
BN	13 151	136	0.177	650	0.175	41	5.4
BN	601 223	194	2.543	605	0.164	41	4.7
HC ₂ N ₃	1 103 408	36	0.286	574	0.154	41	4.0
V ₅ B ₆	1 206 441	65	0.000	560	0.150	41	4.0
V ₃ ReB ₄	1 216 475	25	0.009	588	0.163	40	4.8
B ₂ (CN ₂) ₃	989 466	15	0.661	561	0.153	40	4.0
TiNbB ₄	1 216 692	191	0.000	551	0.151	40	3.9
TiCrB ₄	1 216 966	191	0.047	523	0.138	40	3.0
MoB ₄	1 106 346	194	0.534	565	0.161	39	4.4
V ₃ Cr ₃ B ₈	1 216 493	25	0.003	558	0.159	39	4.3
HfTaB ₄	1 224 283	191	0.000	550	0.157	39	4.2
Mo ₇ B ₂₄	1 228 730	187	0.154	575	0.170	38	4.7
VCrB ₂	1 216 398	38	0.003	554	0.162	38	4.3
Fe ₂ C ₄ Cl ₂ O ₉	864 958	43	3.358	561	0.164	38	4.1
ZrTaB ₄	1 215 209	191	0.000	538	0.157	38	4.1
CN ₂	1 102 681	122	0.560	538	0.159	38	3.8
BC ₂ N	1 080 483	5	0.902	505	0.140	38	2.8
VCrB ₄	1 216 375	191	0.037	538	0.160	37	4.0
C	568 410	65	0.506	539	0.161	37	4.0
V ₂ (B ₂₄ C) ₃	1 217 011	16	0.024	483	0.140	37	2.7
InN ₃	975 606	221	2.393	789	0.231	36	8.9
TiMoB ₄	1 217 026	191	0.061	543	0.166	36	4.3
BN	1 077 506	44	0.301	557	0.173	36	4.3
CN ₂	1 077 595	36	0.697	524	0.164	36	3.8
MoB ₃	1 080 111	166	0.012	571	0.182	35	4.8
Mo ₂ B ₅	7 229	166	0.448	579	0.184	35	4.8
B ₂ CN	1 079 333	51	0.224	549	0.175	35	4.2
HfNbB ₄	1 224 328	191	0.004	529	0.166	35	4.2
VB ₄ Os	1 216 395	187	0.177	534	0.170	35	4.2
ReB ₄	1 190 213	194	0.814	571	0.184	34	4.6
CrB	1 080 664	141	0.000	542	0.176	34	4.3
TaB ₄	1 189 303	194	0.461	536	0.174	34	4.2
ZrNbB ₄	1 215 211	191	0.000	514	0.169	34	4.0
ReN ₂	1 077 354	12	0.023	630	0.200	33	5.7
TaVB ₄	1 217 818	191	0.002	537	0.179	33	4.3
TiWB ₂	1 217 023	38	0.011	525	0.178	33	4.3
TaCrB ₄	1 217 958	191	0.039	527	0.176	33	4.1
WB ₄	29 651	194	0.599	522	0.173	33	4.0
OsN ₂	1 102 074	205	0.619	609	0.203	32	5.5
AlC ₃	1 065 540	221	2.748	533	0.185	32	4.2
ZrMoB ₄	1 215 250	191	0.088	518	0.178	32	4.1
Si ₂ N ₂ O	2 948	164	0.192	505	0.174	32	3.9
Si ₂ N ₂ O	4 644	141	0.189	504	0.175	32	3.9
CrWB ₂	1 226 242	38	0.016	553	0.193	31	4.6
Si ₂ N ₂ O	4 400	15	0.371	539	0.187	31	4.3

TABLE IV. (Continued.)

	id	Space group	d_{hull} , eV	E_{pred} , GPa	v_{pred}	H_{pred} , GPa	K_{ICpred} , MPa m ^{1/2}
NbVB ₄	1 220 351	191	0.014	522	0.186	31	4.1
TiMoB ₂	1 217 028	38	0.000	507	0.182	31	4.1
V ₉ Cr ₃ B ₈	1 216 445	10	0.000	509	0.182	31	4.0
TiBC	1 232 377	194	0.307	505	0.180	31	3.9
Cr ₂ B ₃	12 054	63	0.028	492	0.174	31	3.7
YVB ₄	1 191 641	55	0.000	450	0.161	31	3.2
SiC	1 200 692	160	0.000	439	0.155	31	3.0
SiC	1 200 848	160	0.032	438	0.155	31	3.0
SiC	1 204 356	160	0.000	438	0.155	31	3.0
SiC	570 804	160	0.031	438	0.155	31	3.0
SiC	571 286	156	0.000	438	0.155	31	3.0
SiC	571 298	156	0.000	439	0.155	31	3.0
SiC	1 200 168	160	0.031	439	0.155	31	3.0
SiC	624 397	160	0.031	439	0.155	31	3.0
SiC	11 713	160	0.003	438	0.155	31	3.0
SiC	1 197 730	160	0.000	439	0.156	31	3.0
LuAlB ₁₄	1 197 767	74	0.054	420	0.149	31	2.5
ReOs ₃ ★	867 141	194	0.000	634	0.225	30	6.6
N	1 061 298	194	0.000	616	0.213	30	5.4
ReN ₂	1 077 096	65	0.152	560	0.200	30	4.9
CrMoB ₂	1 226 228	38	0.040	543	0.193	30	4.5
OsWB ₄	1 228 631	187	0.000	527	0.189	30	4.3
YMo ₃ B ₇	504 874	62	0.000	504	0.185	30	4.0
Ti ₃ MoB ₄	1 217 095	6	0.002	464	0.170	30	3.5
Sc ₂ CrB ₆	510 306	55	0.000	424	0.155	30	2.8
ReC ₂	1 008 802	123	0.907	636	0.231	29	6.3
MnOs ₃ ★	1 186 060	194	0.000	613	0.217	29	6.1
Ir ₃ Os ★	1 184 761	221	0.035	605	0.219	29	6.0
AlW ₃ C ₄	1 228 839	25	0.294	569	0.207	29	5.1
ReC ₂	1 019 051	194	0.838	549	0.202	29	4.7
SiO ₂	10 948	60	0.254	510	0.191	29	4.0
MnCrB ₂	1 221 636	38	0.032	502	0.187	29	3.9
SiO ₂	32 667	14	0.244	500	0.190	29	3.8
HfB ₁₂	1 001 600	225	0.043	451	0.173	29	3.2
FeB ₄	1 079 437	58	0.000	450	0.173	29	3.1
Si(GeN ₂) ₂	1 020 663	227	0.205	436	0.169	29	3.1
Ce(CrB ₃) ₂	2 873	71	0.000	420	0.164	29	2.9
B	632 401	1	0.280	414	0.157	29	2.7
LiB ₁₁	1 103 613	216	0.343	411	0.154	29	2.5
B	22 046	134	0.094	409	0.153	29	2.5
IrOs ★	1 223 655	187	0.028	612	0.228	28	6.3
ReOs ★	1 219 509	187	0.000	597	0.231	28	6.2
MoWC ₂	1 221 393	25	0.000	588	0.220	28	5.5
IrN ₂	1 102 235	205	0.311	574	0.211	28	5.1
CrIr ₃ ★	1 183 749	194	0.007	552	0.206	28	5.1
ReB ₃	7 839	194	1.067	570	0.215	28	5.0
V ₁₁ FeB ₈	1 216 907	10	0.014	491	0.190	28	3.9
Mn ₂ MoB ₄	1 078 108	71	0.011	497	0.192	28	3.9
ZrBC	1 232 384	194	0.449	490	0.191	28	3.9
Er ₄ TaV ₇ B ₂₄	1 225 684	6	0.000	462	0.184	28	3.5
TiB ₁₂	1 245 924	225	0.150	429	0.173	28	3.0
Sm(CrB ₃) ₂	1 079 198	71	0.008	407	0.163	28	2.8

TABLE IV. (Continued.)

	id	Space group	d_{hull} , eV	E_{pred} , GPa	V_{pred}	H_{pred} , GPa	K_{ICpred} , MPa m ^{1/2}
B	570 602	134	0.223	411	0.166	28	2.7
ZrBeB	1 215 258	187	0.012	392	0.153	28	2.6
Be ₃ N ₂	1 070 456	160	0.049	399	0.160	28	2.5
Ir ₄ Os ★	1 223 677	166	0.005	580	0.226	27	5.8
Ir ₃ Ru ★	974 358	139	0.000	569	0.216	27	5.5
Nb ₁₁ S ₁₂	684 971	167	0.174	564	0.216	27	5.1
FeIr ₃ ★	10 596	221	0.000	550	0.210	27	5.1
FeIr ₃ ★	1 184 374	194	0.012	545	0.213	27	5.0
B ₄ MoIr	1 228 634	187	0.039	497	0.198	27	4.0
V ₁₁ ReB ₈	1 216 792	10	0.009	490	0.196	27	3.9
Hf(Nb ₂ B ₃) ₄	1 224 510	71	0.005	485	0.193	27	3.9
Ta(MnB ₂) ₂	1 077 930	71	0.000	488	0.196	27	3.8
TmCrB ₄	1 191 371	55	0.000	419	0.173	27	3.0
Be ₄ CrMo ★	1 227 375	164	0.021	397	0.164	27	2.7
B	570 316	134	0.153	400	0.166	27	2.6
IrOsRu ★	1 223 662	156	0.071	557	0.233	26	5.6
ReC	1 079 494	194	0.644	546	0.255	26	5.5
OsRu ★	1 220 023	187	0.000	554	0.233	26	5.5
OsN ₂	1 018 852	127	0.906	564	0.246	26	5.5
NbB ₂ W	1 220 349	38	0.000	513	0.207	26	4.4
W ₂ B ₅	8 079	166	0.547	514	0.208	26	4.2
W ₂ B ₅	570 938	194	0.551	514	0.208	26	4.2
MnMo ₃ B ₄	1 221 732	6	0.049	504	0.207	26	4.2
Nb ₃ V ₅ (B ₂ Ir) ₄	1 220 611	6	0.073	498	0.201	26	4.1
Ta ₅ B ₆	28 629	65	0.000	489	0.199	26	4.0
Ir ₄ W ₅ B ₂₀	1 228 750	187	0.145	485	0.199	26	3.9
Ta ₃ TiB ₄	1 217 975	25	0.000	472	0.194	26	3.8
VC ₃	1 178 822	1	0.886	487	0.197	26	3.8
HoMo ₃ B ₇	504 877	62	0.000	475	0.195	26	3.7
TiNbB ₂	1 216 709	38	0.000	451	0.189	26	3.5
Cr ₃ NiB ₆	1 226 349	38	0.042	464	0.192	26	3.4
Lu ₄ V ₅ B ₁₈	1 223 330	6	0.000	434	0.181	26	3.2
Ti ₇ Ir ₂ Rh ₄ B ₈	1 217 063	6	0.064	431	0.184	26	3.2
Fe ₂ B ₇	1 194 531	55	0.000	434	0.183	26	3.0
YCrB ₄	20 450	55	0.000	407	0.175	26	2.9
BN	1 599	8	0.136	395	0.171	26	2.7
Th(CrB ₃) ₂	9 357	71	0.000	395	0.169	26	2.7
Os ₃ W ★	1 186 374	194	0.000	538	0.260	25	5.9
Re ₃ Os ★	867 264	194	0.000	541	0.255	25	5.8
Os ₂ C ₃	1 189 780	116	1.220	472	0.311	25	5.8
Ta ₅ N ₆	34 761	193	0.853	547	0.245	25	5.4
Ta ₃ N ₅	1 205 002	62	0.166	553	0.237	25	5.2
OsN ₂	568 862	14	0.272	541	0.224	25	4.8
Ta ₂ N ₃	1 208 406	62	0.042	535	0.225	25	4.8
ReN ₂	1 102 441	62	0.089	531	0.218	25	4.6
MnW ₃ B ₄	1 221 699	25	0.049	528	0.216	25	4.6
ReN ₂	1 019 077	11	0.046	527	0.219	25	4.6
ReN ₂	1 019 078	13	0.143	519	0.213	25	4.4
Nb ₄ Mo ₁₁ B ₂₀	1 220 650	71	0.016	511	0.211	25	4.3
TaMoB ₂	1 217 965	38	0.000	496	0.207	25	4.1
DyVB ₄	1 191 899	55	0.000	420	0.184	25	3.0
YbSiO ₃	1 187 516	221	0.000	397	0.174	25	2.7

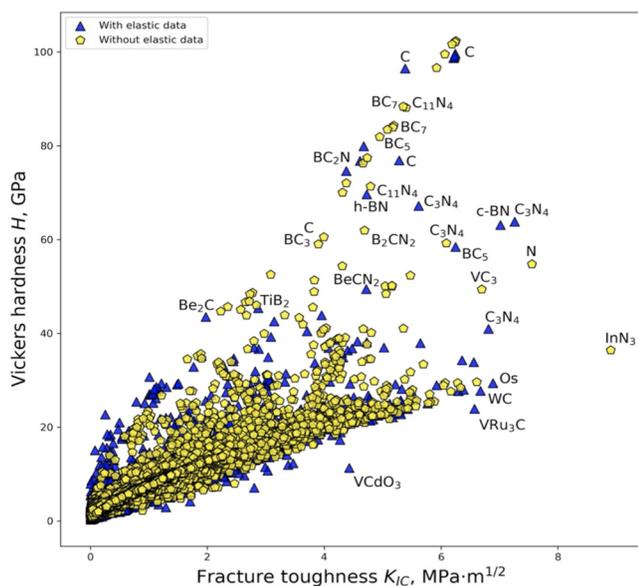


FIG. 9. Ashby plot showing all phases (around 120 000 structures) in Materials Project in the space hardness—fracture toughness. Note that fracture toughnesses of pure metals, metallic alloys, and intermetallics are strongly underestimated (this figure focuses on ceramics and carbides, borides, and nitrides for which results are accurate). The hardest materials are indeed diamond and its polytypes (lonsdaleite and others).

produce even better results. In particular, we see a great opportunity for such methods to be implemented as part of the algorithms, which can generate new crystal structures in a clever way. This could significantly facilitate the development of new materials and open the way to new discoveries.

ACKNOWLEDGMENTS

We thank the Russian Science Foundation (Grant No. 19-72-30043) for financial support.

DATA AVAILABILITY

The data that support the findings of this study are available within the article. All crystal structures, which are part of the calculations, can be accessed via Materials Project database using their ids. Additional information regarding the current article is available from the corresponding author upon reasonable request.

REFERENCES

- A. G. Kvashnin, Z. Allahyari, and A. R. Oganov, "Computational discovery of hard and superhard materials," *J. Appl. Phys.* **126**, 040901 (2019).
- A. R. Oganov, C. J. Pickard, Q. Zhu, and R. J. Needs, "Structure prediction drives materials discovery," *Nat. Rev. Mat.* **4**, 331–348 (2019).
- A. R. Oganov and C. W. Glass, "Crystal structure prediction using *ab initio* evolutionary techniques: Principles and applications," *J. Chem. Phys.* **124**, 244704 (2006).

- J. Wang, J. Lv, L. Zhu, and Y. Ma, "Crystal structure prediction via particle-swarm optimization," *Phys. Rev. B* **82**, 094116 (2010).
- C. J. Pickard and R. J. Needs, "High-pressure phases of silane," *Phys. Rev. Lett.* **97**, 045504 (2006).
- M. Einaga, M. Sakata, T. Ishikawa, K. Shimizu, M. I. Eremets, A. P. Drozdov, I. A. Troyan, N. Hirao, and Y. Ohishi, "Crystal structure of the superconducting phase of sulfur hydride," *Nat. Phys.* **12**, 835 (2016).
- A. F. Goncharov, S. S. Lobanov, I. Kruglov, X.-M. Zhao, X.-J. Chen, and A. R. Oganov, "Hydrogen sulfide at high pressure: Change in stoichiometry," *Phys. Rev. B* **93**, 174105 (2016).
- A. G. Kvashnin, D. V. Semenok, I. A. Kruglov, I. A. Wrona, and A. R. Oganov, "High-temperature superconductivity in a Th-H system under pressure conditions," *ACS Appl. Mater. Interfaces* **10**, 43809 (2018).
- D. V. Semenok, A. G. Kvashnin, A. G. Ivanova, V. Svitlyk, V. Y. Fomin, A. V. Sadakov, O. A. Obolevskiy, V. M. Udalov, I. A. Troyan, and A. R. Oganov, "Superconductivity at 161 K in thorium hydride ThH₁₀: Synthesis and properties," *Mater. Today* **33**, 36 (2020).
- F. Peng, Y. Sun, C. J. Pickard, R. J. Needs, Q. Wu, and Y. Ma, "Hydrogen clathrate hydrates in rare earth hydrides at high pressures: Possible route to room-temperature superconductivity," *Phys. Rev. Lett.* **119**, 107001 (2017).
- H. Liu, I. I. Naumov, R. Hoffmann, N. W. Ashcroft, and R. J. Hemley, "Potential high-Tc superconducting lanthanum and yttrium hydrides at high pressure," *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6990 (2017).
- M. Somayazulu, M. Ahart, A. K. Mishra, Z. M. Geballe, M. Baldini, Y. Meng, V. V. Struzhkin, and R. J. Hemley, "Evidence for superconductivity above 260 K in lanthanum superhydride at megabar pressures," *Phys. Rev. Lett.* **122**, 027001 (2019).
- A. P. Drozdov, P. P. Kong, V. S. Minkov, S. P. Besedin, M. A. Kuzovnikov, S. Mozaffari, L. Balicas, F. F. Balakirev, D. E. Graf, V. B. Prakapenka, E. Greenberg, D. A. Knyazev, M. Tkacz, and M. I. Eremets, "Superconductivity at 250 K in lanthanum hydride under high pressures," *Nature* **569**, 528 (2019).
- Y. Ma, M. Eremets, A. R. Oganov, Y. Xie, I. Troyan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, "Transparent dense sodium," *Nature* **458**, 182 (2009).
- X. Dong, A. R. Oganov, A. F. Goncharov, E. Stavrou, S. Lobanov, G. Saleh, G.-R. Qian, Q. Zhu, C. Gatti, V. L. Deringer, R. Dronskowski, X.-F. Zhou, V. B. Prakapenka, Z. Konôpková, I. A. Popov, A. I. Boldyrev, and H.-T. Wang, "A stable compound of helium and sodium at high pressure," *Nat. Chem.* **9**, 440 (2017).
- J. Wang, K. Hanzawa, H. Hiratsma, J. Kim, N. Umezawa, K. Iwanaka, T. Tada, and H. Hosono, "Exploration of stable strontium phosphide-based electrides: Theoretical structure prediction and experimental validation," *J. Am. Chem. Soc.* **139**(44), 15668 (2017).
- V. Sharma, C. Wang, R. G. Lorenzini, R. Ma, Q. Zhu, D. W. Sinkovits, G. Pilania, A. R. Oganov, S. Kumar, G. A. Sotzing, S. A. Boggs, and R. Ramprasad, "Rational design of all organic polymer dielectrics," *Nat. Commun.* **5**, 4845 (2014).
- W. L. Mao, H. Mao, P. J. Eng, T. P. Trainor, M. Newville, C. Kao, D. L. Heinz, J. Shu, Y. Meng, and R. J. Hemley, "Bonding changes in compressed superhard graphite," *Science* **302**, 425 (2003).
- V. V. Brazhkin and V. L. Solozhenko, "Myths about new ultrahard phases: Why materials that are significantly superior to diamond in elastic moduli and hardness are impossible," *J. Appl. Phys.* **125**, 130901 (2019).
- Z. Allahyari and A. R. Oganov, "Co-evolutionary search of materials with optimal properties in the space of all possible compounds," *NPJ Comp. Mat.* **6**, 55 (2020).
- A. O. Lyakhov and A. R. Oganov, "Evolutionary search for superhard materials: Methodology and applications to forms of carbon and TiO₂," *Phys. Rev. B* **84**, 092103 (2011).
- A. Šimůnek and J. Vackář, "Hardness of covalent and ionic crystals: First-principle calculations," *Phys. Rev. Lett.* **96**(8), 085501 (2006).
- F. Gao, J. He, E. Wu, S. Liu, D. Yu, D. Li, S. Zhang, and Y. Tian, "Hardness of covalent crystals," *Phys. Rev. Lett.* **91**, 015502 (2003).

- ²⁴K. Li, X. Wang, F. Zhang, and D. Xue, "Electronegativity identification of novel superhard materials," *Phys. Rev. Lett.* **100**, 235504 (2008).
- ²⁵X. Chen, H. Niu, D. Li, and Y. Li, "Modeling hardness of polycrystalline materials and bulk metallic glasses," *Intermetallics* **19**, 1275 (2011).
- ²⁶E. Mazhnik and A. R. Oganov, "A model of hardness and fracture toughness of solids," *J. Appl. Phys.* **126**, 125109 (2019).
- ²⁷W. Ye, C. Chen, I.-H. Chu, Z. Wang, and S. P. Ong, "Deep neural networks for accurate predictions of crystal stability," *Nat. Commun.* **9**, 3800 (2018).
- ²⁸O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, and A. Tropsh, "Universal fragment descriptors for predicting properties of inorganic crystals," *Nat. Commun.* **8**, 15679 (2017).
- ²⁹S. K. Kauwe, J. Graser, A. Vazquez, and T. D. Sparks, "Machine learning prediction of heat capacity for solid inorganics," *Integr. Mater. Manuf. Innov.* **7**, 43 (2018).
- ³⁰R. Jalem, K. Kanamori, I. Takeuchi, M. Nakayama, H. Yamasaki, and T. Saito, "Bayesian-driven first-principles calculations for accelerating exploration of fast ion conductors for rechargeable battery application," *Sci. Rep.* **8**, 5845 (2018).
- ³¹P. Avery, X. Wang, C. Oses, E. Gossett, D. M. Proserpio, C. Toher, S. Curtarolo, and E. Zurek, "Predicting superhard materials via a machine learning informed evolutionary structure search," *npj Comput. Mater.* **5**, 89 (2019).
- ³²D. M. Teter, "Computational alchemy: The search for new superhard materials," *MRS Bull.* **23**(01), 22 (1998).
- ³³A. M. Tehrani, A. O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T. D. Sparks, and J. Brgoch, "Machine learning directed search for ultraincompressible, superhard materials," *J. Am. Chem. Soc.* **140**, 9844 (2018).
- ³⁴I. A. Kruglov, P. E. Dolgirev, A. R. Oganov, A. B. Mazitov, S. N. Pozdnyakov, E. A. Mazhnik, and A. V. Yanilkin, "Machine learning interatomic potentials for global optimization and molecular dynamics simulation," in *Materials Informatics*, edited by O. Isayev, A. Tropsha, and S. Curtarolo (Wiley, 2019), pp. 253–288.
- ³⁵A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," *Phys. Rev. B* **87**, 184115 (2013).
- ³⁶A. V. Shapeev, "Moment tensor potentials: A class of systematically improvable interatomic potentials," *Multiscale Model. Simul.* **14**, 1153 (2016).
- ³⁷T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," *Phys. Rev. Lett.* **120**, 145301 (2018).
- ³⁸A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, "Commentary: The materials project: A materials genome approach to accelerating materials innovation," *APL Mater.* **1**, 011002 (2013).
- ³⁹S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, "Python materials genomics (pymatgen): A robust, open-source python library for materials analysis," *Comput. Mater. Sci.* **68**, 314 (2013).
- ⁴⁰P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.* **136**, B864 (1964).
- ⁴¹W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.* **140**, A1133 (1965).
- ⁴²R. Hill, "The elastic behaviour of a crystalline aggregate," *Proc. Phys. Soc.* **65**, 349 (1952).
- ⁴³D. P. Kingma and J. Ba, Adam: "A method for stochastic optimization," [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- ⁴⁴M. de Jong, W. Chen, R. Notestine, K. Persson, G. Ceder, A. Jain, M. Asta, and A. Gamst, "A statistical learning framework for materials science: Application to elastic moduli of k-nary inorganic polycrystalline compounds," *Sci. Rep.* **6**, 34256 (2016).