Journal of Materials Chemistry A

High-Throughput Computation of Novel Ternary B-C-N Structures and Carbon Allotropes with Electronic-Level Insight into Superhard Materials from Machine Learning

Journal:	*Journal of Materials Chemistry A*
Manuscript ID	TA-ART-09-2021-007553.R1
Article Type:	Paper
Date Submitted by the Author:	29-Oct-2021
Complete List of Authors:	Al-Fahdi, Mohammed; University of South Carolina System, Mechanical Engineering Ouyang, Tao; Xiangtan University, Laboratory for Quantum Engineering and Micro-Nano Energy Technology & Department of Physics Hu, Ming; University of South Carolina, Mechanical Engineering

High-Throughput Computation of Novel Ternary B-C-N Structures and Carbon Allotropes with Electronic-Level Insights into Superhard Materials from Machine Learning

Mohammed Al-Fahdi,[1] Tao Ouyang,[2,*] and Ming Hu[1,*]

[1]Department of Mechanical Engineering, University of South Carolina, Columbia, SC 29208, US

[2]Hunan Key Laboratory for Micro-Nano Energy Materials and Device and School of Physics and Optoelectronics, Xiangtan University, Xiangtan 411105, Hunan, China

Abstract

Discovering new materials with desired properties have been a dominant and crucial topic of interest in the field of materials science in the past few decades. In this work, novel carbon allotropes and ternary B-C-N structures were generated using the state-of-art RG^2 code. All structures were fully optimized using density function theory with first-principles calculations. Several hundred carbon allotropes and ternary B-C-N structures were identified to be superhard materials. The thermodynamic stability of some randomly selected superhard materials was confirmed by evaluating the full phonon dispersions in the Brillouin zone. The new carbon allotropes and ternary B-C-N structures possess a wide range of mechanical properties generally and Vickers hardness specifically. Through 2D Pearson's correlation map, we first reproduced the well-accepted explanation and relationship of Vickers hardness of the generated structures with other mechanical properties such as shear modulus, bulk modulus, Pugh's ratio, universal anisotropy, and Poisson's ratio. We then propose two fundamentally new descriptors from electronic level, namely local potential and electron localization function averaged over a unit cell, both of which exhibit strong correlation with Vickers hardness. More importantly, these descriptors are easy to access from first-principles calculations (at least two orders of magnitude faster than traditional calculation of elastic constants), and thus can serve as a fast and accurate approach for screening superhard materials. We also combined these new descriptors with known composition and structure descriptors in the machine learning training process. The new descriptors significantly enhance the performance of the trained machine learning model in predicting Vickers hardness of unknown materials, which provides strong evidence for local potential and electron localization function to be considered in future high-throughput computation. This work unravels more fundamental but previously unexplored knowledge about superhard materials and the newly proposed electronic level descriptors are expected to accelerate the discovery of new superhard materials.

[*] Authors to whom all correspondence should be addressed. E-Mail: ouyangtao@xtu.edu.cn (T.O.); hu@sc.edu (M.H.)

I. Introduction

Materials with various mechanical properties are desired in multiple applications, which range from soft materials being used as biological tissues and hard electronic materials such as thermoelectric materials or solar cells to superhard materials like concrete and steel materials. Superhard materials are needed and desired in a wide variety of real-world applications such as cutting tools, drilling, abrasives for lapping, grinding, polishing, automated and aerospace applications, medical plants, and armor plating [1-5]. Traditionally, the hardness of materials is measured by a machine named Vickers hardness test developed in 1921 by Robert Smith and George Sandland [6]. The machine has a tip that indents through the materials and evaluates the how much the materials is capable to resisting the indentation through assessing the plastic deformation of those materials [7-8]. The load applied by the tip has some effect of the values of Vickers hardness which might make the value of Vickers hardness different with different loads, and this phenomenon is named the indentation size effect [8]. In the indentation size effect, the hardness decreases as the load keeps increasing which can be attributed to change in the microstructure or the sizable elastic recovery [9-10]. Materials with hardness of 40 GPa or higher are considered to be superhard materials [1, 11-13]. However, the experiments are too expensive to perform. High-throughput computation has been widely performed to accelerate new materials discovery. This process can further speed up by combining with the state-of-the-art machine learning approach. Therefore, machine learning is highly expected to predict Vickers hardness in a faster and cheaper way [1, 14]. In fact, a machine learning model named support vector machine has already been used to screen thousands of materials to predict superhard materials [14]. The machine learning model was able to identify promising ternary superhard materials: $ReWC_{0.8}$ and $Re_{0.5}W_{0.5}C$ which were later synthesized and proved to be superhard materials experimentally at low indentation loads [14]. Extreme Gradient Boosting (XGBoost) has been developed and improved in 2016 [15] from gradient boosting by Friedman in 2001 [16]. Gradient Boosting can also be named Gradient Boosting Machine (GBM) or Gradient Boosted Regression Tree (GBRT) [15]. Extreme Gradient Boosting is considered to be an ensemble learning algorithm in the machine learning world. Extreme Gradient Boosting algorithm builds a tree ensemble model in which each tree makes predictions with regularized terms to avoid overfitting. Then, the sum of each prediction from each tree is the final prediction of the algorithm from the provided data. Gradient Tree Boosting has demonstrated to yield the state-of-the-art results in classification problems [17] and was used in Netflix prize [18]. Extreme Gradient Boosting algorithm is available as an open source package and free for everyone to use [15]. The algorithm was used in winning competitions in Kaggle, a website for data scientists to participate and show their coding skills in real-world data science competitions [15]. Among the 29 challenges at Kaggle, Extreme Gradient Boosting was used in 17 codes of the winning and best performing machine learning models. The highly efficient performance of Extreme Gradient Boosting algorithm in capturing hidden trends that prevail in datasets is the reason why we utilize the algorithm in our work.

Several Carbon allotropes are known for being superhard materials as reported in previous studies [19-22]. Carbon can exhibit a wide range of allotropic forms due to its ability to form various hybridizations with different bond angles [23-24]. Carbon can have different form of hybridization states such as sp^3, sp^2, and sp-hybridized states because it has four valence electrons, so it can make single, double, triple, and aromatic Carbon-Carbon bonds. It is known scientifically that three carbon allotropes exist naturally. Those natural carbon allotropes are amorphous carbon, diamond, and graphite with hybrid sp^2/sp^3, sp^3, and sp^2 hybridized carbon atoms, respectively [25]. A substantial amount of research has been conducted to discover and explore new carbon allotropes in which some were synthesized and fabricated experimentally. The newly discovered carbon allotropes can have different dimensionalities and hybridizations. Some examples of the newly discovered carbon allotropes with different dimensionalities are fullerenes (zero-dimensional carbon allotrope (0-D)), carbon nanotubes (one-dimensional material (1-D)), and finally graphene (two-dimensional material (2-D)) [26-33]. Carbon allotropes with their hybridizations and dimensionalities that were synthesized experimentally or predicted theoretically are

one-dimensional (1-D) sp-carbyne, two-dimensional (2-D) $sp-sp^2$-graphyne, and three-dimensional (3-D) $sp-sp^3$-yne-diamond [34-35]. Several methods were utilized to synthesize those carbon allotropes. One of the methods that was recently used in 2011 is cold-compressing graphite with pressure equivalent to 17 GPa which also transformed the hybridization state of graphite from sp^2 to sp^2/sp^3 [34]. After this discovery, more carbon allotropes are synthesized experimentally through high-pressure techniques, examples on those carbon allotropes are bct-C_4 carbon [30], M-carbon [36], O-carbon [26], W-carbon [37], and Z-carbon [38]. In all these previous studies, carbon was the sole element that was used in the process of superhard materials discovery. However, light elements such as B, C, N, and O are known for having higher Vickers hardness because they make strong covalent bonds which makes them difficult to break [39]. Carbon allotropes [40-41] and ternary compositions were formed from these elements such as BC_4N [42], BC_2N [43], C-N-O [44], B-N-O [45-46], and B-C-O [41, 47]. However, in the previous studies, experiments were utilized to implement their ideas and synthesize the previously mentioned structures which makes it extremely expensive to synthesize those materials in terms of time, resources, and cost. Recently, computational tools were developed to help the materials discovery process become easier and more accessible. These methods hold great promise to promote the new material discovery. Some of those crystal structure prediction (CSP) tools are CALYPSO [48] and USPEX [49]. In this work, RG^2 package [50-54] was used to generate the new and novel carbon allotropes with new ternary B-C-N structures.

II. Procedure of Materials Generation, Screening, and Computational Methods

A) Structures Generation

The RG^2 package [20, 50-54] was used to generate the novel carbon and B-C-N structures. The main input parameters of the RG^2 package are the target constitute elements, a range for the number of bonded atoms for all species, number of inequivalent atoms in unit or primitive cell, target space group/groups, and bond feature information such as bond length, bond angle and the tolerance allowed for both derivations. With these input parameters, RG^2 package builds the correct labeled quotient graph. In this work regarding the generated carbon allotropes, various carbon allotrope structures with different number of carbon atoms were arbitrarily distributed in a stochastic unit cell with a random lattice constant and arbitrarily selected space group, and the same applies for B-C-N system structures. This will allow for global search of all possible space groups. The number of symmetrically independent atoms usually ranges between 2 and 10, having the majority between 3 and 7. Then, an initial structure with the same number of atoms is generated which is also predicated on symmetry. The package computed the matrix that consists of the distances of the carbon (or B-C-N) atoms in a specific generated carbon allotrope (or B-C-N) structure and build the labeled quotient graph (LQG). The code can also relax the structure if the structures have physically reasonable bond lengths and angles. The new structures generated by RG^2 package were named according to the rules which are essentially the input parameters used to generate the structure. Those parameters are separated by a hyphen to distinguish the parameters and add another parameter after each hyphen. The naming procedure applied to the generated structures start from left to right. The naming rules are implemented in the following order (from left to right): number of space group, number of nonequivalent atoms in the structures, the names of elements (C in carbon allotropes and $B_\# C_\# N_\#$ in B-C-N structures), loop or ring (predominantly ring), structure local ID, and possible additional hyphen to differentiate two IDs.

B) Structure Optimization using Density Functional Theory (DFT)

First-principles calculations by density functional theory (DFT) were carried out with the computational chemistry software: Vienna ab initio simulation package (VASP) [55-57]. The structures were optimized

at first before performing any following calculations. The structure optimization convergence criteria of force and energy are 10^{-4} eV/A and 10^{-7} eV, respectively. VASP performed structure relaxation with full degree of freedom in terms of allowing the atomic coordinates, lattice size, lattice constant, and lattice shape to change to reach the convergence criteria of force and energy in the structure optimization process. Perdew-Burke Ernzerhof (PBE) of the generalized gradient approximation (GGA) was used for exchange-correlation functional [58]. The kinetic energy cutoff was set to be 520 eV for the electronic wavefunction having a plane wave basis set which was obtained using the Projector Augmented-Wave method [57, 59]. The chosen kinetic energy cutoff (520 eV) was selected as recommended by VASP manual which should be at least 1.3 times the largest ENMAX of an atomic specie in a composition. In both B-C-N structures and carbon allotropes, the largest ENMAX was 400 eV which means that the kinetic energy cutoff should be at least 520 eV which was implemented in the calculations. The Monkhorst-pack k-mesh grids [60] selected to sample the Brillouin zone in the calculations were determined depending on the lattice constants. The product of number of k-mesh in one direction and the lattice constant in the same direction is roughly set as 60, which is dense enough and equivalent to a regular diamond lattice (primitive cell containing 2 carbon atoms) at k-mesh size of 16×16×16. The local potential is a measure of repulsive and attractive interactions in the unit cell. Local potential was calculated throughout the unit cell using VASP which is outputted in a file named "LOCPOT". Electron localization function is a measure of the probability of finding an electron at a certain point in the unit cell which is calculated by VASP and outputted in a file named "ELFCAR"

C) Mechanical Properties Calculations

The finite difference method implemented in VASP was used to calculate the elastic stiffness tensor matrix of the optimized conventional cell. The results can be read from the OUTCAR file outputted from VASP calculations. The elastic stiffness matrix is computed through performing six different distortions on the lattice or unit cell and calculating the elastic constants using the relationship between stress and strain [61]. The elastic stiffness matrix is different for each crystal system in terms of the number of independent elastic constants and zero elastic constant in some of the terms in the matrix depending on the type of crystal system [62]. The general form of the elastic stiffness matrix in which all the elastic constants are independent (21 independent elastic constants) is applied to the least symmetrical crystal system which is the triclinic crystal system. The elastic stiffness tensor for the triclinic crystal system is shown below:

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56} \\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66} \end{bmatrix} \quad (1)$$

where c_{ij} is the elastic constant element in direction ij in the elastic stiffness matrix given that $c_{ij} = c_{ji}$ in all the elements in the elastic stiffness matrix. The number of independent elastic constants in the previous matrix shown in Eq. (1) is 21 which is the maximum number of independent elastic constants for a crystal system. The number of independent elastic constants decreases if the crystal system becomes more symmetrical. In fact, some elements in the elastic stiffness matrix become zeros when the crystal system is more symmetrical. The previous statements about the elastic constants in the elastic stiffness matrix is demonstrated in the equation (2) which shows the elastic stiffness matrix of a cubic crystal system which is the most symmetrical crystal system.

$$c_{cubic} = \begin{bmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{bmatrix} \quad (2)$$

In the cubic crystal system, the following elastic constants are equal: $c_{21} = c_{31}$, $c_{33} = c_{22} = c_{11}$, and $c_{66} = c_{55} = c_{44}$. The number of independent elastic constants for the cubic crystal system is therefore 3 which is the minimum number of independent elastic constants. That confirms the statement mentioned earlier that the number of independent elastic constant decreases as the crystal system becomes more symmetrical. It is also shown that several elastic constants in the elastic stiffness tensor have the value of zero. The elastic stiffness matrices for all other crystal systems can be found in Ref. [62].

The compliance matrix is the inverse of the elastic matrix. The compliance matrix shown below will be useful in the coming equations which will be used in calculating the mechanical properties.

$$[c_{ij}] = [s_{ij}]^{-1} \quad (3)$$

Two common definitions can be used to calculate bulk (B) and shear (G) moduli using the elastic constants in elastic stiffness matrix or the compliance matrix. The two definitions are Voigt's [62] and Reuss' [64] approximations. Reuss' approximation assumes uniform stress throughout the lattice. On the other hand, Voigt's approximation assumes uniform strain throughout the lattice. Voigt's approximation equations of bulk (B_V) and shear (G_V) moduli are as follows:

$$\begin{aligned} 9B_V &= (c_{11} + c_{22} + c_{33}) + 2(c_{12} + c_{23} + c_{31}) \\ 15G_V &= (c_{11} + c_{22} + c_{33}) - (c_{12} + c_{23} + c_{31}) + 4(c_{44} + c_{55} + c_{66}) \end{aligned} \quad (4)$$

Reuss approximation equations for bulk (B_R) and shear (G_R) moduli are as follows:

$$\begin{aligned} \frac{1}{B_R} &= (s_{11} + s_{22} + s_{33}) + 2(s_{12} + s_{23} + s_{31}) \\ 15/G_R &= 4(s_{11} + s_{22} + s_{33}) - 4(s_{12} + s_{23} + s_{31}) + 3(s_{44} + s_{55} + s_{66}) \end{aligned} \quad (5)$$

In Hill's [65] approximation, Voigt's and Reuss' approximations are lower and upper bounds, respectively. Hill found that the average of these two approximations yields better results that are close to the experimental results, i.e.

$$\begin{aligned} B_{VRH} &= \frac{(B_V + B_R)}{2} \\ G_{VRH} &= \frac{(G_V + G_R)}{2} \end{aligned} \quad (6)$$

In this work, the results from Hill's approximations will be considered for the calculations of bulk and shear moduli.

Young's modulus and Poisson's ratio are calculated using the following formulae:

$$E = \frac{9BG}{3B+G} \quad (7)$$

$$\nu = \frac{3B-2G}{2(3B+G)} \quad (8)$$

Researchers have carried out a great deal of effort to quantify and calculate anisotropy of a material. In 1948, a definition to quantify anisotropy was introduced by Zener. He used some of elastic constants in his definition. The formula that he came up with was $A = \frac{2c_{44}}{c_{11}-c_{12}}$ [66]. Chung in 1967 presented another formula to calculate anisotropy which includes both Voigt's and Reuss' approximations in the formula [67]. Chung's formula is $A^c = \frac{G_V-G_R}{G_V+G_R}$. The definitions of anisotropy developed by Zener and Chung yielded sufficiently accurate results for the cubic crystal systems since cubic crystal systems possess an isotropic bulk resistance. The previous two definitions did not give acceptable results regarding the other crystal systems besides the cubic crystal systems because the other crystal systems -not cubic- manifest anisotropic bulk resistance [62]. Therefore, quantifying anisotropy accurately required that all the contributions of anisotropy must be represented in an equation for all the crystal systems. To this date, Ranganathan [68] was able to mathematically formulate anisotropy which was able to overcome and solve the obstacles in the previous two definitions of anisotropy. Ranganatha's definition is written below in equation (9):

$$A^U = 5\frac{G_V}{G_R} + \frac{B_V}{B_R} - 6 \quad (9)$$

In this work, Ranganatha's definition of anisotropy will be implemented in the calculations.

A significant amount of effort and research was achieved to understand and build a model for Vickers hardness. In fact, it was explained with various approaches and theories. The following is a brief summary of some of the theories that explained Vickers hardness: (1) strength of the chemical bond between the atoms [69]; (2) a thermodynamic concept that explains how the chemical bonding pertains to the density of energy [70]; (3) bond length, charge density, and ionicity [71]; (4) how the electron holding energy of a bond and electronegativity contribute to hardness [72]. Progress also occurred to mathematically calculate Vickers hardness. In 1998, Teter observed the correlation between shear modulus and Vickers hardness [73] which was successful to a certain extent. The formula that Teter developed is shown below:

$$\mathrm{H_{V,Teter}} = 0.151 \times \mathrm{G} \quad (10)$$

Chen observed that the existence of a discrepancy between the calculations performed by Teter's model and the experiment results [74]. Chen hypothesized that plastic deformation was not considered in Teter's model. Chen suggested that the new mathematical model should include Pugh's ratio to add the plastic deformation contribution to hardness. Pugh's ratio (k) is calculated by the ratio of shear modulus and bulk modulus as shown below in equation 11.

$$k = \frac{G}{B} \quad (11)$$

Pugh's ratio is a measure of how brittle a material is. As the value of Pugh's ratio increases, the more brittle the material becomes [74]. Chen and his coworkers proposed another mathematical formula which was more accurate than Teter's model:

$$\mathrm{H_{V,Chen}} = 2(\mathrm{k^2G})^{0.585} - 3 \quad (12)$$

Chen's model demonstrated high accuracy in calculating Vickers hardness when compared with the experimental results. However, Chen's mathematical model did not yield physically accurate results when

calculating Vickers hardness of ductile materials. The reason for that is the presence of an intercept term which does not hold any physical meaning. Tian observed this phenomenon in 2012 when calculating Vickers hardness of some ductile materials which turned out to have negative Vickers hardness according to Chen's model [19]. Tian then proposed another model for Vickers hardness with superior performance in calculating Vickers hardness for both ductile and brittle materials, as shown below:

$$H_{V,Tian} = 0.92k^{1.137}G^{0.708} \quad (13)$$

Tian's model considered the plastic deformation represented in Pugh's ratio in his mathematical model which was not considered in Teter's model [19, 73]. Moreover, results from Tian's model on ductile materials were in good agreement with the experimental results unlike Chen's model which gave unrealistic and negative Vickers hardness when calculating hardness of ductile materials [19, 74]. Tian's model surpassed the other models in a wide range of ductile and brittle materials with various values of Vickers hardness and had extraordinarily accurate results when compared with the experimental results [19]. Therefore, in this work Tian's model is used in calculating Vickers hardness of all carbon allotropes and ternary B-C-N structures.

D) Phonon Dispersion Calculations

The phonon dispersion of selected carbon allotropes and B-C-N structures were calculated. The supercell used to calculate the phonon dispersion of those structures is 2 × 2 × 2 for all the structures. PHONOPY [75] package was utilized in calculating the second order harmonic constant. The phonon dispersion for all the selected structures were calculated by PHONOPY through the finite displacement method except for 217-5-64-B15C5N12-r6-np-id0-20210507-093520. The phonon dispersion of 217-5-64-B15C5N12-r6-np-id0-20210507-093520 was calculated by PHONOPY using ALM method [76] since ALM method gave no negative frequency by considering the finite temperature effect. The dynamical matrix in the reciprocal space was then derived from the derivative of energy to plot the phonon dispersion for the selected structures.

III. Results and Discussion

A) Superhard structures screening and thermodynamical stability

682 carbon allotropes were initially generated from the RG^2 package [20, 50-54]. 578 carbon allotropes were not found in the SACADA database [77] (pure carbon) after cross-checking. However, 669 carbon allotropes were selected and screened from the 682 original carbon allotropes because anisotropy (AU Ranganathan's definition) [68] set in this work is between 0 and 3. Out of the screened 669 carbon allotropes, 635 structures fulfill the condition of being superhard materials with Vickers hardness higher than 40 GPa [1, 11-13, 20]. Regarding B-C-N structures, the RG^2 package initially generated 977 B-C-N structures, and 833 structures remained after screening the materials with AU outside of 0-3 range. 404 B-C-N structures are found to be superhard out of the remaining 833 B-C-N structures after the elastic constants are calculated. Figure 1 shows some selected structures from B-C-N systems and carbon allotropes. All structures were obtained from VESTA software [78]. 4×4×1 and 3×3×3 supercells are plotted in Figs. 1(a) and 1(b), respectively, while 2×2×2 supercells are plotted in both Figs. 1(c) and 1(d).

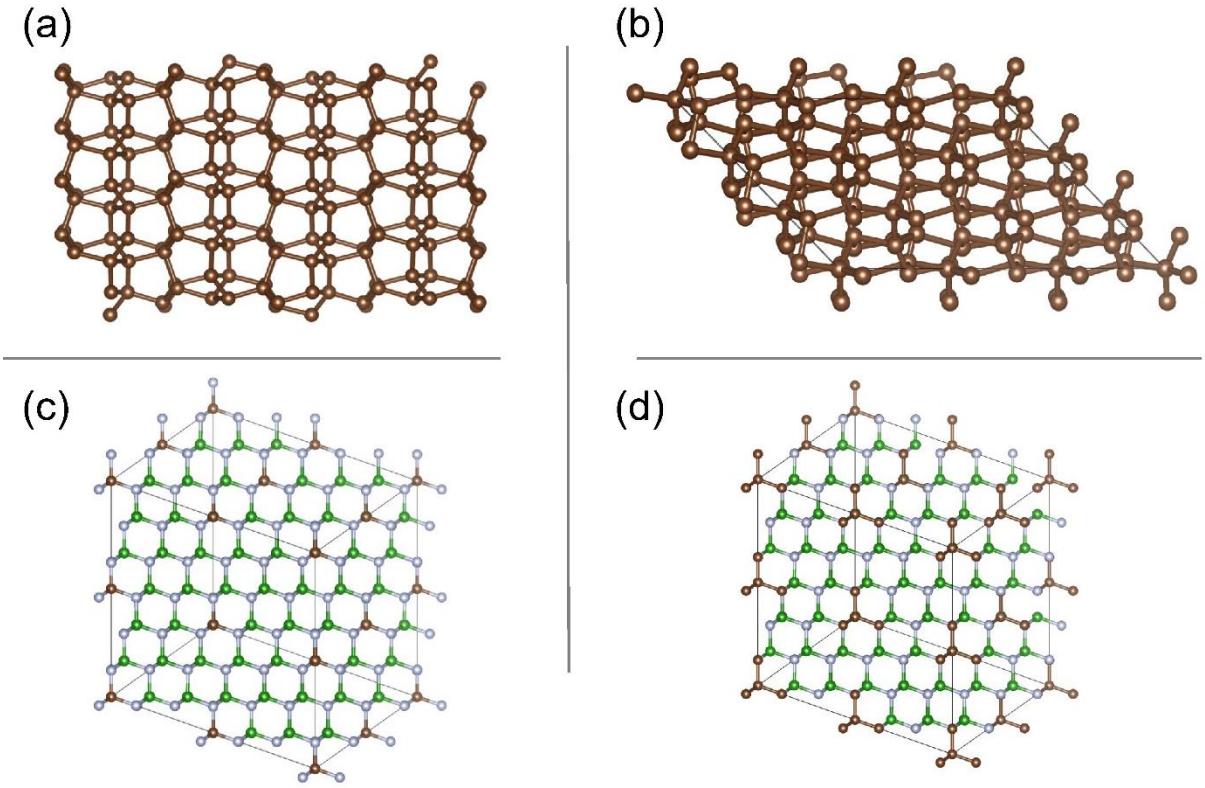

Figure 1. Supercell structures of carbon allotropes (a) 56-4-32-C-r567-np-id0-20210426-175823, (b) 122-2-20-C-r567-np-id0-20210426-180030, and ternary B-C-N structures (c) 217-5-64-B15CN16-r6-np-id0-20210507-094012, (d) 217-5-64-B15C5N12-r6-np-id0-20210507-093520. Color code: dark brown, carbon; green, boron; light silver, nitrogen. Naming procedure of the structures can be found in section **II. Procedure of Materials Generation, Screening, and Computational Methods** in subsection *A) Structures Generation.*

Figure 2 shows the phonon dispersion plots of all the previously mentioned selected structures in the same previous order as well. No negative or imaginary frequency was found in any of the phonon dispersions of the selected structures. The absence of the negative frequencies in the phonon dispersion plots indicates the thermodynamical stability of those structures [20, 79-81], which means that these structures could be synthesized experimentally [82] provided that the formation energy is also negative for all the reported materials. Table 1 shows information about the material IDs of some selected materials and information about their composition, Vickers hardness, and their classification in terms of whether they are superhard materials or not. We show the structures of some reported materials (two carbon allotropes and two ternary B-C-N structures) in Figure 1 and the corresponding phonon dispersions of those materials in Figure 2.

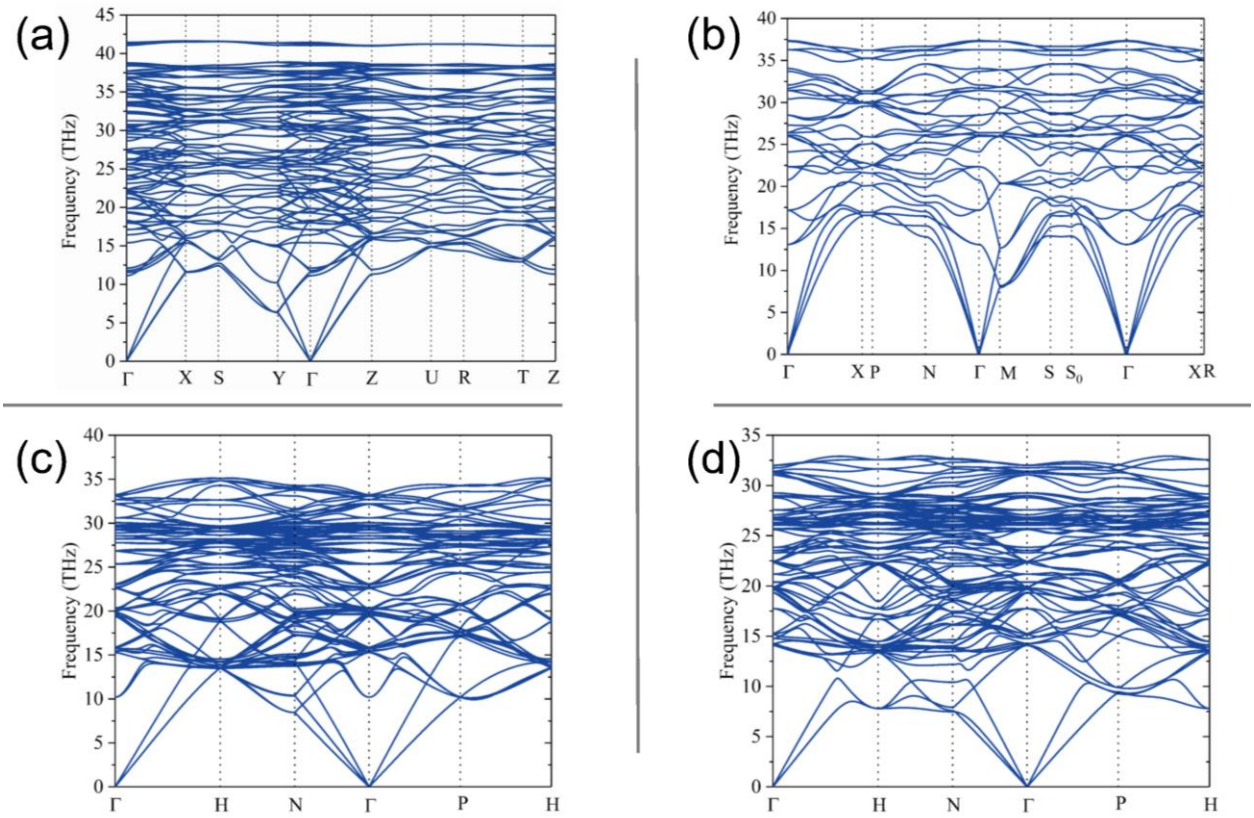

Figure 2. Phonon dispersion of carbon allotropes (a) 56-4-32-C-r567-np-id0-20210426-175823, (b) 122-2-20-C-r567-np-id0-20210426-180030, and B-C-N ternaries (c) 217-5-64-B15CN16-r6-np-id0-20210507-094012, (d) 217-5-64-B15C5N12-r6-np-id0-20210507-093520. Naming procedure of the structures can be found in section **II. Procedure of Materials Generation, Screening, and Computational Methods** in subsection *A) Structures Generation*

Table 1: Vickers hardness of the selected carbon allotropes and ternary B-C-N compounds with crystal system information.

Materials	Composition	Crystal System (Space Group Number)	Vickers Hardness (GPa)	Superhard Materials Classification
122-2-20-C-r567-np-id0-20210426-180030	C_{20}	$I\bar{4}2d$ (122)	93.5658	Superhard
57-4-24-C-r56-np-id0-20210426-175825	C_{24}	Pbcm (57)	89.2500	Superhard
63-4-40-C-r567-np-id0-20210426-175904	C_{40}	Cmcm (63)	89.1263	Superhard
56-4-32-C-r567-np-id0-20210426-175823	C_{32}	Pccn (56)	88.3379	Superhard
217-5-64-B15CN16-r6-np-id0-20210507-094012	$B_{30}C_2N_{32}$	$I\bar{4}3m$ (217)	43.2490	Superhard
217-5-64-B15C5N12-r6-np-id0-20210507-093520	$B_{30}C_{10}N_{24}$	$I\bar{4}3m$ (217)	41.5995	Superhard

217-5-64-B4CN3-r6-np-id0-20210507-093640	$B_{32}C_8N_{24}$	$I\bar{4}3m$ (217)	32.7131	Not Superhard
215-3-8-B4CN3-r6-np-id0-20210506-082304	$B_4C_1N_3$	$P\bar{4}3m$ (215)	24.0686	Not Superhard

B) Correlation between Vickers hardness and other mechanical properties

Descriptors from composition of the materials were used to predict mechanical properties in previous studies [1, 14, 83]. Composition along with structural descriptors were also used to predict mechanical properties [84-85]. Combining the structural descriptors to the composition descriptors for machine learning training to predict mechanical properties proved to enhance the results of the model, yielding lower mean absolute error when compared to using the composition descriptors alone which yielded higher mean absolute error [84]. Here, we investigate the electron localization function (ELF) and local potentials (LOCPOT) as potential promising descriptors to predict Vickers hardness.

Local potential reflects the interatomic interactions between positive (protons in the nucleus) and negative (electrons surrounding the nucleus) charges in the unit cell [86-87]. The local potential varies in different areas of the unit cell depending on the strength of the surrounding charges and the distance from those charges. Positive local potentials convey the repulsive interactions taking place at a particular spatial point. Negative local potential values signify the attraction interactions at a specific point of space. The local potentials used in the unit cells [88] can be expressed as

$$V_{LOCPOT}(r) = V(r) + \int \frac{n(r')}{|r-r'|} dr' + V_{XC}(r) \quad (14)$$

where $V(r)$ is the ionic potential, $V_{XC}(r)$ is the exchange-correlation potential, and the middle (second) term is the Hartree potential. Local potentials for each material can be obtained from the file named "LOCPOT" outputted by VASP. ELF measures the probability of an electron localized at a specific point in a lattice or unit cell [89] and can be expressed as

$$ELF = \frac{1}{1+\left(\frac{D}{D_h}\right)^2} \quad (15)$$

The ratio $\left(\frac{D}{D_h}\right)$ is a dimensionless localization index which indicates the electron localization for the uniform electron gas. ELF value has a range between 0 and 1. The value 0 in ELF means it is unlikely that there will be an electron at that spatial point in space (lattice or unit cell in our work) [90-91]. The value 1 in ELF denotes that the electrons are localized at that point of space. When ELF is equal to 0.5, that suggests that the electron cloud density is delocalized like in metallic bonds. The ELF information for each material can be obtained from the file "ELFCAR" which is generated by VASP.

Pearson correlation matrix was created by open-source Python to give insights of how much each property correlates with the other properties [92]. The Pearson correlation matrix is shown in Figure 3. Figure 3 shows how much each property correlates with the other properties starting with Vickers hardness from Tian's model and ending with average ELF. Atomic properties shown in Figure 3 such as packing fraction (PF), volume per atom (vpa), and density were produced by Matminer [93]. Although the Pearson correlation matrix shows the full information about the correlation for each property with the other properties, Vickers hardness will be the dominant topic of the discussion to expound compared to other properties.

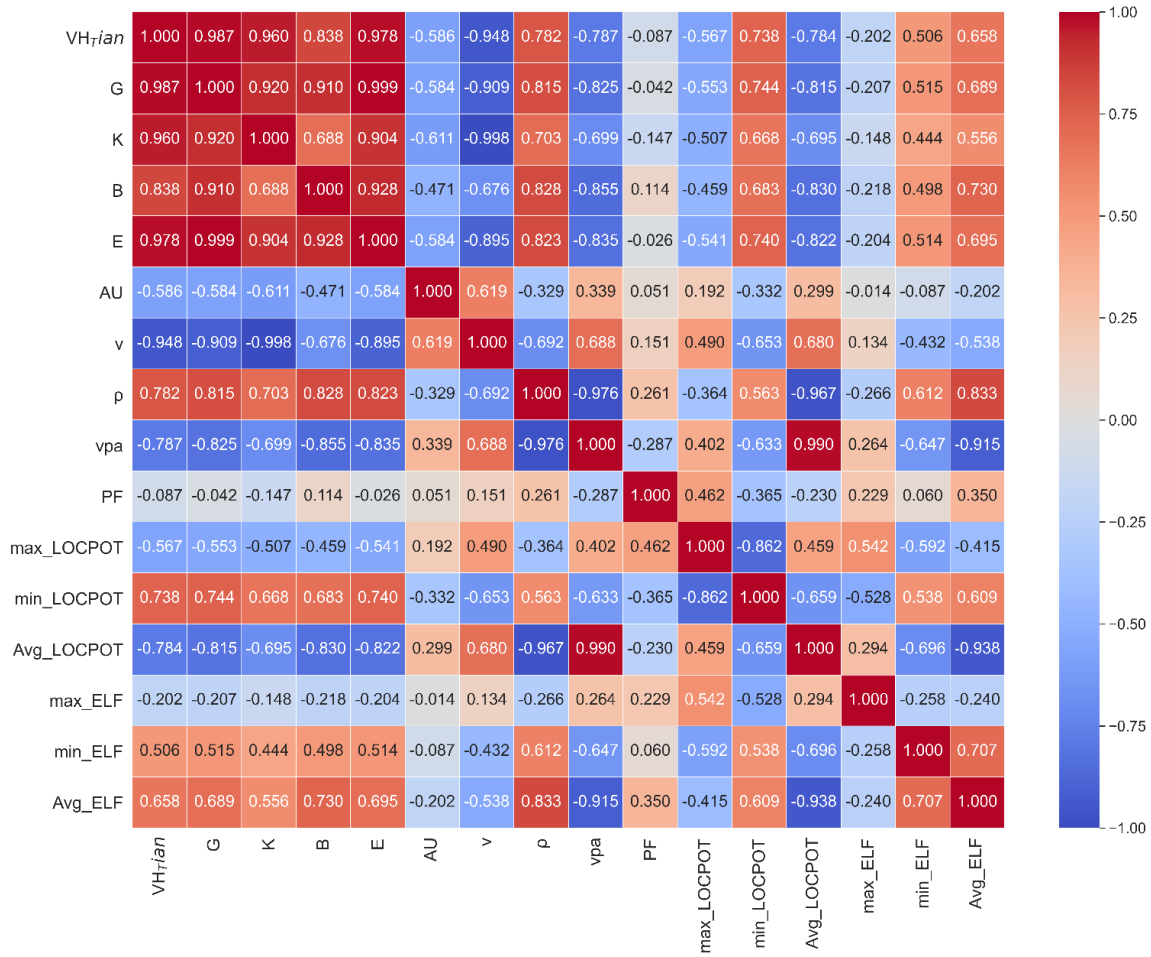

Figure 3. Pearson correlation matrix for various properties. The properties with their symbols are: Vickers hardness from Tian's model (VH_{Tian}), Shear Modulus (G), Pugh's ratio (k), Bulk modulus (B), Young's modulus (E), Universal Anisotropy (AU), Poisson's ratio (ν), density (ρ), volume per atom (vpa), packing fraction (PF), maximum local potential (max_LOCPOT), minimum local potential (min_LOCPOT), average local potential (Avg_LOCPOT), maximum local ELF (max_ELF), minimum local ELF (min_ELF), and average ELF (Avg_ELF).

A Pearson correlation matrix correlates two parameters with each other having the values of the correlation between −1 to 1 as seen in the color bar of Figure 3. The order of the parameters in columns is the same order of the parameters in the rows, so the diagonal elements in the correlation matrix represent the correlation results of the same identical parameters, that is why the value +1 always appears in the diagonal. −1 value denotes the complete opposite correlation between two parameters. The value +1 indicates that there is absolute commensurate correlation between two parameters. In the correlation matrix shown in Figure 3, no perfect inverse correlation (value of −1) exists, but perfect direct correlation exists between the same parameters in columns and rows (diagonal parameters). For example, shear modulus (G) in the second row is in perfect direct correlation with shear modulus in the second column (G) which is represented by the value 1, and the same applies to all the diagonal elements (correlation results between two identical parameters) in the correlation matrix shown in Figure 3. The value 0 denotes a non-existent correlation between two parameters. Although, no correlation has the value of zero, some correlation results are extremely close to 0 which conveys insignificant inverse or direct correlation between those two parameters. Also as seen in Figure 3, most values are not −1, 0, or +1, but somewhere between two of those values. The values between 0 to +1 connotes the direct correlation, but

the degree of the direct correlation depends on whether that value is closer to 0 or +1. If the value is closer to 0, then the direct correlation is weak. However, if the value is closer to +1, then the direct correlation is strong. The same applies to the values between 0 to −1 which represent the reverse correlation results between two parameters. If the value is close to −1, the inverse relation is strong. On the other hand, the inverse correlation is weak if the value is close to 0.

The mechanical and atomic features (descriptors) that are strongly and directly commensurate to Vickers hardness are shear modulus, Pugh's ratio, bulk modulus, Young's modulus, density, minimum local potential, minimum ELF, average ELF. The mechanical and atomic features (descriptors) that conversely affect Vickers hardness are universal anisotropy, volume per atom, maximum local potential, and average local potential. Packing fraction and maximum ELF descriptors were not included in either inverse or direct correlations since their correlations to Vickers hardness are extremely low which is -0.087 for packing fraction and -0.202 for maximum ELF. Pearson correlation matrix confirmed some obvious and well-known correlations between Vickers hardness with some other material properties. Shear modulus and Pugh's ratio are vastly proportional to Vickers hardness with correlation coefficients of 0.987 and 0.960, respectively. Shear modulus and Pugh's ratio are part of the equation of Vickers hardness shown in Eq. (13) which is why the correlation coefficients are tremendously high. Although elastic modulus is not part of the definition of Vickers hardness, the correlation between elastic modulus with Vickers hardness was extraordinarily high with a correlation coefficient of 0.978 which is higher than the correlation coefficient of Pugh's ratio (part of the definition of Vickers hardness). Having higher elastic modulus was reported to be an indicator of having higher Vickers hardness as seen from Refs. [19, 74, 94-96]. Furthermore, machine learning model for elastic modulus prediction was used to predict new superhard materials [14]. Besides, bulk modulus correlation coefficient of 0.838 is also noticeably high from Figure 3. One of the definitions of Vickers hardness was a linear correlation with bulk modulus which manifested its success in accurately calculating Vickers hardness in materials such as diamond, Si, Ge, GaSb, and InSb [97-98]. However, the linear model of bulk modulus failed to calculate Vickers hardness in a diverse range of other materials [97-98]. It was also reported that a non-linear correlation between Vickers hardness and bulk modulus failed to calculate Vickers hardness in a wide range of materials [99]. The linear and non-linear correlations of bulk modulus to calculate Vickers hardness failed, because the variation of shear modulus with bulk modulus must also be considered in calculating Vickers hardness which was shown in Refs. [19, 74].

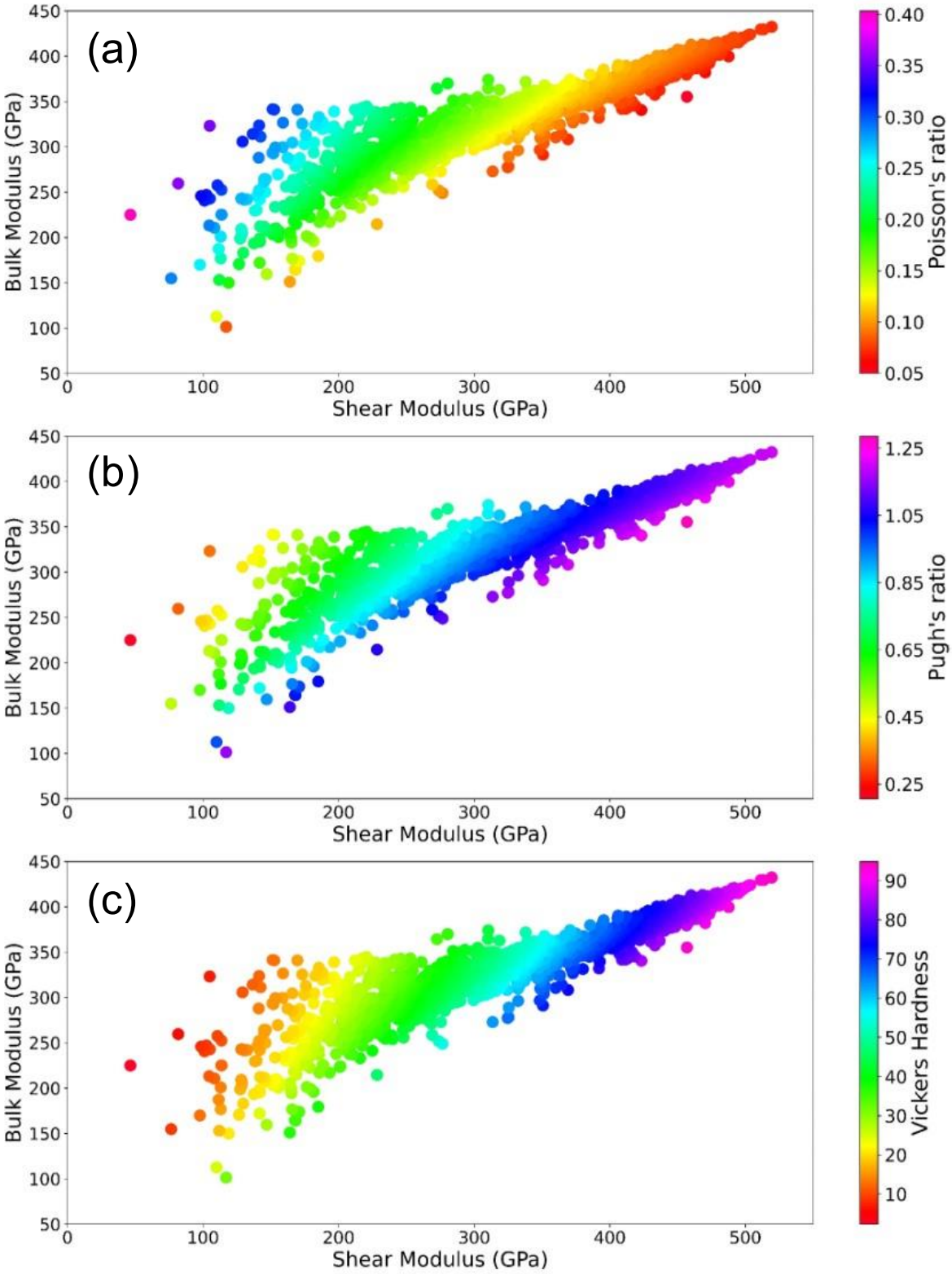

Figure 4. A 2D map of bulk modulus vs. shear modulus with (a) Poisson's ratio, (b) Pugh's ratio, and (c) Vickers hardness as color bar to show the wide range of Vickers hardness in terms of new brittle and ductile carbon allotropes, how both bulk and shear moduli affect Vickers hardness, and where superhard carbon allotropes are located with respect to both shear and bulk moduli.

Figure 4 shows the 2D map of bulk modulus vs. shear modulus with (a) Poisson's ratio, (b) Pugh's ratio, and (c) Vickers hardness as color bars. Figure 4 (c) directly shows how to alter the results of Vickers hardness even with the same values of bulk modulus. As can be seen from Figure 4(c), bulk modulus, shear modulus, and Vickers hardness have a relationship among each other. When shear modulus is equal to 200 GPa the Vickers hardness decreases as bulk modulus increases which conveys some inverse relation between Vickers hardness and bulk modulus having constant shear modulus which explains why the correlation between Vickers hardness and bulk modulus is not closer to 1 compared to shear modulus as an example. However, the general trend as can be seen from Figure 4 (c) is that Vickers hardness increases as bulk modulus rises as well. Therefore, the direct correlation is confirmed between bulk modulus and Vickers hardness which is in good agreement with Pearson's correlation matrix shown in Figure 3. Furthermore, when bulk modulus is constant with a value of 300 GPa as an example, the value of Vickers hardness grows as shear modulus rises as well, which implies that Vickers hardness rises as shear modulus increases with constant bulk modulus. The previous few statements illustrated from Figure 4(c) elucidate why the model of linear [97-98] or non-linear [99] bulk modulus that calculates Vickers hardness failed. Although the overall trend from Figure 3 is that Vickers hardness of the materials grows as bulk modulus rises, Figure 4(c) reveals some visual insight on Vickers hardness which declines as bulk modulus increases with constant shear modulus. Generally speaking, materials with high Vickers hardness are likely to be found in the upper-right corner of the 2D map of bulk modulus vs. shear modulus.

Poisson's ratio has an extremely high opposite correlation with Vickers hardness with a correlation coefficient of −0.948. This result of Pearson correlation matrix confirms that brittle materials have high Vickers hardness. Poisson's ratio and Pugh's ratio can be used as criteria to distinguish brittle and ductile materials. According to Pugh's ratio, a material is brittle when $k \geq k_{cr}$ and ductile when $k < k_{cr}$ where $k_{cr} \approx 0.571$ [100-103]. According to Poisson's ratio, a material is brittle when $\upsilon \leq 0.33$ and ductile or malleable when $\upsilon > 0.33$ [104]. Figures 4 (a) and 4 (b) show how Pugh's ratio and Poisson's ratios differ with bulk modulus and shear modulus to further explain the relationship between Vickers hardness with brittle and ductile materials.

As can be observed from Poisson's and Pugh's ratios' criteria for brittle and ductile materials, they are opposite each other in the sense that brittle (ductile) materials exist when the Poisson's ratio is low (high) and the Pugh's ratio is high (low) [105]. That was also visually seen in Figures 4 (a) and 4 (b). The areas where Pugh's ratio is high are the same areas where low Poisson's ratio values occur, and vice versa. Figures 4 (a) and 4 (b) confirm that superhard materials usually have low Poisson's ratio and high Pugh's ratio. That also explains why the correlation between Pugh's ratio and Poisson's ratio is −0.998 in Pearson correlation matrix (Figure 3). In fact, if Eqs. (8) and (11) that represent Poisson's ratio and Pugh's ratio equations, respectively, are combined, they will yield equation 16 in Ref. [106]:

$$\upsilon = \frac{3-2k}{2(3+k)} = \frac{3-2k}{6+2k} \tag{16}$$

Eq. (16) also proves that Poisson's ratio and Pugh's ratio are inversely proportional to each other. Note that Eq. (16) restricts Poisson's ratio value in the range $-1 < \nu \leq \frac{1}{2}$, while for Pugh's ratio $\infty \geq k > 0$ [105].

In Figure 5, we also plot the composition triangle of contour distribution of bulk modulus, shear modulus, and Vickers hardness for the new ternary B-C-N compounds reported in this work. Figure 5(a) illustrates that high bulk modulus compounds are concentrated in the region where B atoms are about 50% while C and N atoms have diverse percentages. Similar trend is found for shear modulus of ternary B-C-N structures in Figure 5(b). As clearly seen in Figure 5(c), the contour distribution of Vickers hardness vastly similar to that for shear modulus, while it has large differences with the contour of bulk modulus.

We also see that high Vickers hardness compounds are likely to occur in the region where B to N ratio is around 1:1 or less with higher concentration of C atoms (more than 50% to both B and N).

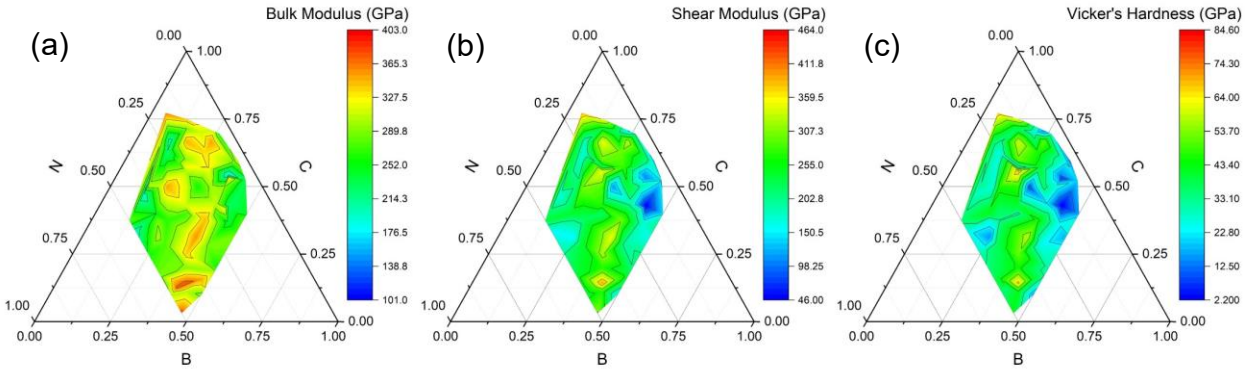

Figure 5. Composition triangle of contour distribution of (a) bulk modulus, (b) shear modulus, and (c) Vickers hardness for the new ternary B-C-N materials reported in this work.

C) Insights on Vickers hardness from atomic and electronic level features

In the previous discussions, mechanical properties such as bulk, shear, and elastic moduli, Poisson's ratio, and Pugh's ratio were used to explain Vickers hardness. Although they are useful to explain Vickers hardness, atomic descriptors are still needed to explain Vickers hardness at the atomic level. Atomic descriptors can be consistent and proportional to Vickers hardness as was shown in the Pearson correlation matrix in Figure 3. Therefore, Table 2 and Table 3 will show some atomic descriptors that will help in expounding Vickers hardness in more details. Table 2 shows carbon allotropes and B-C-N structures with high Vickers hardness and their respective material properties, while Table 3 shows carbon allotropes and B-C-N structures with low Vickers hardness and their material properties. The two tables are presented to differentiate the materials with high Vickers hardness from the materials with low Vickers hardness in terms of their material properties. Six materials will be shown in each table: the first three are ternary B-C-N structures, and the second three materials are carbon allotropes.

Table 2: Selected superhard B-C-N structures and carbon allotropes with corresponding material properties.

Materials ID	Composition	Vickers Hardness, (GPa)	Universal Anisotropy	Elastic Modulus, (GPa)	Density, (kg/m^3)	Volume per atom, $(m^3/atom)$	Average local potential	Average ELF
165-4-24-BC4N-r46x-np-id0-20210507-094212	$B_4C_{16}N_4$	84.589	0.0393	983.533	3.4271	5.8839	−12.9668	0.42604
160-4-24-BC6N-r6-p-id0-20210506-082421	$B_3C_{18}N_3$	83.800	0.0448	990.202	3.4414	5.8433	−13.0187	0.42762
164-4-16-BC6N-r46-np-id0-20210507-093548	$B_2C_{12}N_2$	83.359	0.0804	982.632	3.4208	5.8787	−12.9803	0.42629
166-8-48-C-r6-np-id0-20210426-180510	C_{48}	92.816	0.0216	1098.295	3.4904	5.7141	−13.1787	0.43122
194-8-100-C-r567-np-id0-20210426-180626	C_{100}	92.329	0.0932	1039.989	3.4153	5.8396	−13.0537	0.42916
194-6-64-C-r567-np-id0-20210426-180618	C_{64}	89.191	0.0953	1001.570	3.3795	5.9015	−12.9823	0.42762

Table 3: Selected non-superhard B-C-N ternaries and carbon allotropes with corresponding material properties.

Materials ID	Composition	Vickers Hardness, (GPa)	Universal Anisotropy	Elastic Modulus, (GPa)	Density, (kg/m^3)	Volume per atom, $(m^3/atom)$	Average local potential	Average ELF
103-4-32-B2CN-r468-p-id0-20210506-085316	$B_{16}C_8N_8$	5.584	2.3552	221.851	2.7062	7.3079	−11.3952	0.41134
230-3-96-B2CN-r46-np-id0-20210506-084739	$B_{48}C_{24}N_{24}$	8.403	2.9596	260.923	2.5922	7.6294	−11.0451	0.38016
169-4-24-BCN2-r46-p-id1-20210507-093500	$B_6C_6N_{12}$	8.902	1.7577	197.058	2.3904	8.8285	−10.0149	0.32822
14-5-20-C-r3489x-np-id0-20210426-175559	C_{20}	17.008	2.6245	366.645	2.7341	7.2947	−11.6681	0.39784
166-3-108-C-r456x-np-id0-20210426-180505	C_{108}	22.759	1.0010	444.769	2.6250	7.5977	−11.1035	0.37629
191-3-36-C-r4568-np-id0-20210426-180605	C_{36}	32.523	1.5164	554.205	2.9314	6.8037	−11.8796	0.39783

Table 2 shows some instances of superhard materials and their material properties, and Table 3 shows other instances in which the materials are not superhard. It can be noticed that universal anisotropy tends to be lower in superhard materials shown in Table 2, whereas anisotropy is higher in materials with low Vickers hardness shown in Table 3. Therefore, anisotropic materials tend to have lower Vickers hardness. The results shown in Table 2 and 3 confirm the inverse correlation between Vickers hardness and universal anisotropy shown in Figure 3 in Pearson correlation matrix which has a correlation coefficient of −0.586. Regarding density, superhard materials tend to be denser than non-superhard materials which conveys that superhard materials are more packed. The correlation coefficient between density and Vickers hardness is 0.782 as shown in Figure 3 in Pearson correlation matrix. Volume per atom is essentially the volume divided by the number of atoms. If the volume per atom is low, then the material has a higher density and vice versa. Therefore, volume per atom and Vickers hardness are inversely commensurate with each other having a correlation coefficient of −0.787 as shown in Pearson correlation matrix in Figure 3. Furthermore, B-C-N materials in Table 3 have less carbon fraction than those in Table 2, which confirms the results in Figure 5 that the B-C-N structures with high Vickers hardness have higher carbon atomic fractions or concentrations compared to B-C-N structures with low Vickers hardness.

At the electronic level, local potential reflects the type of interactions between the nucleus with positive charge and the electrons with the negative charge [86-87]. Negative local potential indicates attraction and positive local potential signifies repulsive interactions. Interesting enough, the average local potential for superhard materials is more negative which implies that more attractive interactions exist in superhard materials which might also explain why superhard materials are denser and more packed. Pearson's correlation map (Figure 3) clearly shows that, the correlation coefficient between Vickers hardness and average local potential is -0.784, which essentially indicates that materials with higher Vickers hardness tend to have lower (more negative) local potentials, and vice versa. Moreover, average ELF indicates the probability of finding electrons at a certain point in the unit cell. Superhard material are more inclined to have higher average ELF which suggests that superhard materials have higher electron density and are more packed since the electrons are distributed throughout the unit cell whereas non-superhard materials have less average ELF which implies that some empty spaces exist in the unit cell of non-superhard materials which also explains why they have less density. It is also known from the literature that electron concentration is directly proportional to hardness [19]. The correlation coefficient between average ELF and Vickers hardness is 0.658, which indicates that Vickers hardness positively correlates with the

average ELF distribution throughout the unit cell. It is also worth pointing out that, both average local potential and ELF are easy to access in first-principles calculations, e.g. they can be immediately outputted after structure optimization. Our first-principles calculations of several hundred carbon allotropes and B-C-N ternaries have shown that, the average CPU time for obtaining local potential and ELF takes only a few minutes, while traditional calculations of elastic constants by the Hessian matrix (matrix of the second derivatives of the energy with respect to the atomic positions, i.e. IBRION = 6 DFPT calculations in VASP runs) take several hours for each structure. Then, obtaining local potential and ELF is at least two orders of magnitude faster than direct calculations of elastic constants. Therefore, local potential and ELF are tremendously promising and helpful for fast and accurate screening superhard materials once the structures are optimized.

D) Machine learning insights from local potential and electron localization function

From the previous results and discussion through Pearson correlation matrix, we know that the generated descriptors of local potential and ELF are (directly or inversely) proportional to Vickers hardness depending on the descriptor. A machine learning algorithm needs to be implemented to confirm that the generated descriptors from local potential and ELF affect the results of trained machine learning models for predicting Vickers hardness by reducing the error in the machine learning models, which will also verify the novelty of the generated descriptors to enhance the predictions performed by the machine learning models. Composition descriptors have been widely utilized in machine learning to predict mechanical properties with acceptable successful results [1, 14, 83]. Combining structural descriptors with composition descriptors has also proved to generally enhance the machine learning models to predict material properties, in particular for mechanical properties [84]. The composition descriptors that were used in Refs. [1, 14, 83] utilized the atomic and chemical properties of constituent elements as follows: (1) atomic number, (2) atomic weight, (3) period, (4) group, (5) families, (6) Mendeleev number, (7) atomic radius, (8) covalent radius, (9) Zunger radii sum, (10) ionic radius, (11) crystal radius, (12) Pauling electronegativity, (13) Martynov and Batsanov (MB) electronegativity, (14) Gordy electronegativity, (15) Mullinke electronegativity, (16) Allred and Rockow (AR) electronegativity, (17) metallic valence, (18) number of valence electrons, (19) Gilmor number of valence electron, (20) valence *s*, (21) valence *p*, (22) valence *d*, (23) outer shell electrons, (24) 1st ionization potential (kJ/mol), (25) polarizability (A^2), (26) melting point (K), (27) boiling point (K), (28) density (g/mL), (29) specific heat (J/g K), (30) heat of fusion (kJ/mol), (31) heat of vaporization (kJ/mol), (32) thermal conductivity (W/(mK)), (33) heat atomization (kJ/mol), (34) cohesive energy, and (35) electron affinity (kJ/mol). The composition descriptors that are produced based on the atomic features from the elements are the maximum, minimum, difference (range) between maximum and minimum, and average of the elemental properties of the constituent elements in a composition, which in total produces $35 \times 4 = 140$ descriptors for each material from the composition. The descriptors from the structures are density, packing fraction, volume per atom, and space group (total of 4 descriptors for each material). The new descriptors from local potential and ELF that will be added to the machine learning model are the maximum, average, and minimum values of local potential (total of 3 descriptors for each material) and ELF (total of 3 descriptors for each material).

The machine learning models were trained based on three levels of descriptors. The first machine learning model is trained with composition descriptors only as in Refs. [1, 14, 83]. The second machine learning model is trained with both composition and structural descriptors, while the third machine learning model is trained with the combination of composition, structural, local potential, and ELF descriptors. To gauge the novelty of the new local potential and ELF descriptors, the results from the machine learning model that uses local potential and ELF descriptors must be compared to the machine learning model that uses previously existing descriptors namely composition descriptors used in previous work [14, 83] and see whether the new descriptors enhance the new machine learning model. Carbon allotropes have the same

composition descriptors which makes it impossible to train a machine learning model based on carbon allotropes using composition descriptors solely. Machine learning models learn from the variance that exists in the features, and carbon allotropes have no variance due to having the same values (zero variance) in the features generated from composition. For that reason, B-C-N materials are selected for training the machine learning model since B-C-N structures can be distinguished from composition descriptors due to the various B-C-N elemental fractions from one B-C-N material to another. The machine learning algorithm used in the training process is the ensemble learning algorithm named Extreme Gradient Boosting [15] implemented in the open-source software, Python. Some hyperparameters in the machine learning algorithm, Extreme Gradient Boosting, is 0.03 for the learning rate and 600 for the number of trees in the forest. The data was split into 80% for training and 20% for testing. The quality of the machine learning models is judged based on two parameters: (1) Mean Absolute Error (MAE), and (2) R^2 score between the predictions and test data. MAE represents the average error quantity between the "true" values and predicted values. In this work, MAE measures the error between test DFT data for Vickers hardness and the predicted Vickers hardness from machine learning. R^2 score is a statistical measure which measures the variance proportion between an independent variable with a dependent variable. In this work, R^2 is a manifestation of how consistent the predicted results are with the test data. Comparing the testing data, which is usually 20% of the total data, with the machine learning model predicted results is a standard procedure in machine learning, which can confirm that the machine learning model was able to capture the trends in the training dataset and can predict the test data whether with/without underfitting or overfitting. Underfitting occurs when the machine learning model cannot accurately predict the testing data due to the lack of descriptors that can capture the trends which can be shown by large MAE and low R^2 score while comparing the predicted results with the test data. Overfitting occurs when the machine learning model learns from the noise in the training set which will show low MAE for the training set. However, when the machine learning model has the issue of overfitting, the machine learning model will have high MAE and low R^2 score in the testing dataset. The goal from this scheme is to confirm that the generated descriptors from local potential and ELF affect the results of machine learning models trained on Vickers hardness by reducing MAE and increasing R^2 score in the testing dataset to verify the novelty of the generated descriptors to enhance Vickers hardness predictions. In Figure 6, the results of MAE and R^2 score are compared among the three machine learning models between the test data and the machine learning model predicted results.

As seen in Figure 6 (a), MAE decreases as the number of descriptors increases from 1- composition descriptors only with MAE = 11.7, then 2- composition and structure descriptors with an MAE score of 5.7 to 3- combined composition, structure, local potential, and ELF descriptors with an MAE of 4.3. This proves that adding local potential and ELF descriptors improves the performance of machine learning models trained on Vickers hardness. With combined descriptors, the model predicts Vickers hardness closer to the "true" value calculated from DFT with much lower error (i.e., MAE in this work). R^2 scores show the consistency levels of the machine learning predicted Vickers hardness to the calculated values by DFT. The machine learning model trained on solely composition descriptors have the worst consistency levels between the test data and the predicted results with R^2 score of only 27.8%, but R^2 score has significantly improved to 72.8% by adding structural descriptors to the original composition descriptors. The novelty of our work manifests itself by improving R^2 score further to 88.9% through adding new descriptors of local potential and electron localization function. The results prove that adding local potential and ELF descriptors improved MAE score from 11.7 using composition descriptors only used in previous work [14, 83] to an MAE score of 4.3 combining descriptors from previous work to the novel local potential and ELF descriptors which is an impactful novelty in this work for future studies on superhard materials.

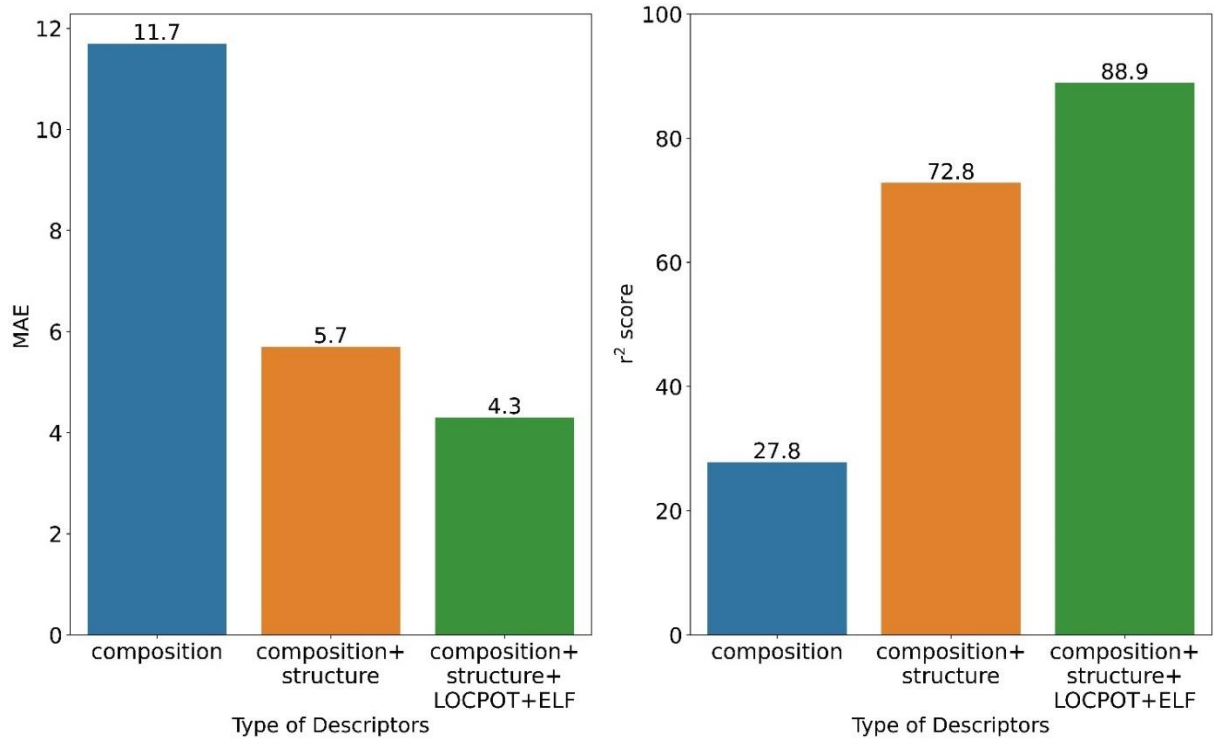

Figure 6. Comparison of the prediction results with the test data using the two gauging machine learning parameters: (a) MAE and (b) R^2 score among three machine learning models with the three different types of descriptors.

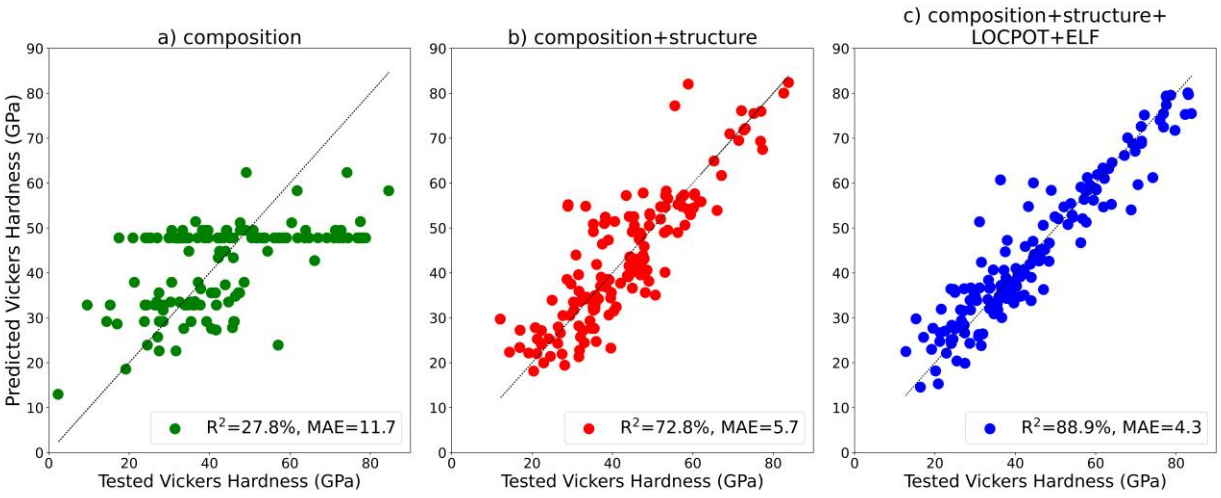

Figure 7. Predicted Vickers hardness vs. tested Vickers hardness data from three machine learning models trained on Vickers hardness using the following descriptors in each model: (a) composition descriptors only, (b) composition and structure descriptors, (c) combined composition, structural, local potential (LOCPOT), and electron localization function (ELF) descriptors. The dashed lines define the perfect prediction slope to be the consistency indicator between the predicted results and test data.

Figure 7 shows the predicted Vickers hardness results against the test data calculated by DFT for all three machine learning models which should give more insights into the results of MAE and R^2 score. It shows how the predicted results of Vickers hardness from the three machine learning models vary with the test data of Vickers hardness calculated by DFT. In the first machine learning model shown in Figure 7 (a) that only used composition descriptors, R^2 score was substantially low which shows enormous inconsistency of the predicted Vickers hardness from machine learning with the calculated test results from DFT, and MAE was substantially high consequently. The reason for the colossal inconsistency in the predicted Vickers hardness is that the machine learning model predicted that many materials have the same Vickers hardness close to 50 GPa which can be shown as a straight horizontal-like line of points when in fact those materials have different values of Vickers hardness as shown by their respective Vickers hardness values from the test results in the x-axis. Those materials located on the horizontal-like line might have the same composition, and that is why the machine learning model predicted their Vickers hardness to be equal. What caused the machine learning model to have the same prediction of Vickers hardness for the materials in the horizontal-like line is the weights from training the machine learning model in which the machine learning model was not able to identify the difference in all those materials based on composition only which begs the necessity for another descriptor or descriptors to be used by the machine learning model as an input to distinguish those materials from each other. The phenomenon that happened in Figure 7 (a) shows what is known in data science as underfitting which occurs when a machine learning model is unable to capture the relationship between the features and the predicted output. In Figure 7 (b), the machine learning model was trained using the same composition descriptors plus structure descriptors, and the model shows better consistency between the predicted results with the tested Vickers hardness data compared to the first one which used composition descriptors only. The second machine learning model has noticeably lower MAE and higher R^2 score compared to the first one which in consequence yielded more consistency between the predicted results and test data. Adding the structure descriptors solved the underfitting issue in the results from the machine learning model that used the composition descriptors shown Figure 7 (a). The results shown in Figure 7 (c) were from the third machine learning model that used the same composition and structure descriptors from the machine learning model shown in Figure 7 (b) along with a total of six other descriptors from ELF and local potential. In the third machine learning model, the predicted and test data look even less dispersed and more consistent than the second machine learning model in Figure 7 (b) which explains why the machine learning model has lower MAE and higher R^2 score. The results from the machine learning model show that the issue of underfitting shown in Figure 7 (a) was enhanced. Moreover, overfitting does not occur since the machine learning model was able to accurately predict the results in the test data which proves that the machine learning model did not learn from the noise data during the training process. The results shown in Figure 7 (c) validate the novelty of the descriptors from local potential and ELF as was already shown in Figure 6 (c) as well. It is also worth noting that, we have run the machine learning code multiple times in which the machine learning code gave different results in each run. However, the R^2 score and MAE did not change much in all of the runs. For example, when all the composition, structural, ELF and local potential descriptors were combined, we had R^2=90% as the highest percentage and R^2=87.8% as the lowest percentage. This is normal since the shuffle function is activated in the train and test split part of the code. However, the results were always similar and consistent in each run among all the three machine learning models. The train and test data might differ between different runs, which is why the R^2 score and MAE results differ in each run. However, the testing and training sets have a wide range of Vickers hardness values, which was ensured in each run. The tested Vickers hardness in Figure 7 shows that the testing data in all the models have a wide range of data points from low Vickers hardness to high Vickers hardness.

Before closing, we would like to explain Vickers hardness from more fundamental electronic level. In Figure 8 we compare the spatial distribution of local potentials between two representative B-C-N ternaries, i.e. a non-superhard material ($B_6C_6N_{12}$) and superhard material ($B_3C_{18}N_3$). The plots were produced through VESTA code [78]. Both plots were produced from the same plane which was [001].

The plots were made at different distances from the origin to find the image in which the comparison between the local potential in both materials can be explained more clearly. The blue color in the VESTA code [78] represents extremely low values of local potentials which signifies attractive interactions [86-88]. The opposite is implemented to the red color which represents positive values of local potentials (repulsive interactions) [86-88].

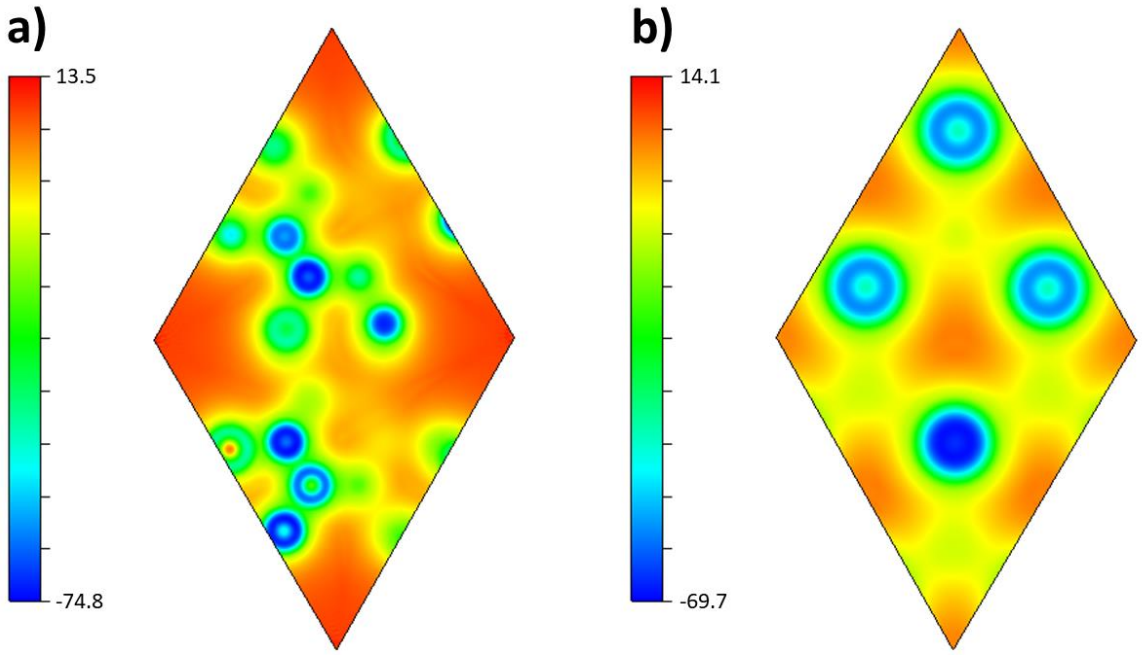

Figure 8. Spatial distribution of local potentials in (a) 169-4-24-BCN2-r46-p-id1-20210507-093500 ($B_6C_6N_{12}$, a non-superhard material) with distance of 0.3 Å from origin and (b) 160-4-24-BC6N-r6-p-id0-20210506-082421 ($B_3C_{18}N_3$, a superhard material) with distance of 1.8 Å from origin. The plots were created by VESTA code [78] and viewed along [001] plane.

For $B_6C_6N_{12}$ shown in Figure 8 (a) the Vickers hardness is 8.9 GPa, which belongs to non-superhard material, while for $B_3C_{18}N_3$ shown in Figure 8 (b) its Vickers hardness is 83.8 GPa and the material is superhard. The blue color in Figure 8 denotes attractive interactions between positive and negative charges (i.e., nucleus and surrounding electrons). That is why the blue color in both plots in Figure 8 is almost perfectly circular. As the distance increases from the nucleus, the color changes from blue, to light blue, to green, to yellow, and finally red, which is where the electrons have increasingly more repulsive interaction between each other. Figure 8 (a) was taken at a distance where the blue color was more common and red color was less dominant compared to the other distances from the origin. Figure 8 (b) was produced where the blue color could be viewed, and the red color could also be viewed clearly compared to the other distances. It is shown that in Figure 8 (a) the red color is more dominant even though at that distance red color is less prevalent compared to the other distances from the origin, which means the red color is even more pervasive in other distances compared to the selected distance. Therefore, the interactions are extremely repulsive at the selected distance and more repulsive at other distances. In Figure 8 (b) the local potential was less red in other distances, but that distance was selected for comparison purposes to show that even when red is more common in $B_3C_{18}N_3$ at some distance, the red color in $B_3C_{18}N_3$ is still less prevalent than $B_6C_6N_{12}$. The fact, that the local potential in $B_6C_6N_{12}$ shown in Figure 8 (a) has a higher red color concentration than $B_3C_{18}N_3$ shown in Figure 8 (b), essentially demonstrates that more repulsive interactions prevail in $B_6C_6N_{12}$ as compared to $B_3C_{18}N_3$. The higher concentration of negative and attractive local potential also explains why superhard materials have higher

density and are generally more packed compared to the low dense and less packed non-superhard materials which have higher repulsive, positive local potentials as was seen in $B_3C_{18}N_3$ and $B_6C_6N_{12}$. This observation made from Figure 8 confirms the difference of local potentials between superhard materials and non-superhard materials, which was initiated by the correlation coefficient in Pearson correlation matrix in Figure 3, and then was perceived by the machine learning model through reducing MAE as presented in Figure 7.

IV. Conclusions

Materials discovery has been a crucial topic in the recent years with the goal to generate and discover materials with specific demands and desirable properties. In this work, we report hundreds of new ternary B-C-N compounds and carbon allotropes structures with super high hardness generated by the state-of-the-art RG^2 package. The atomic configuration of each material was optimized using high precision first-principles calculations. Vickers hardness was calculated using Tian's model. Several descriptors (atomic and mechanical descriptors) were used to explain Vickers hardness from Pearson correlation in this work. To the best of our knowledge, we introduce, for the first time, new descriptors namely local potential and electron localization function to further explain and enhance machine learning based models to boost the accuracy for Vickers hardness predictions with lower mean absolute error and higher R^2 score. The spatial distribution of local potentials has successfully, for the first time, provided deep insights into interpreting Vickers hardness at the atomic level, and the average local potentials can be used as a measure to give some initial idea on the magnitude of Vickers hardness of the material. The machine learning model was utilized in deducing physical insights from atomic level features, i.e., local potential and electron localization function on Vickers hardness. We believe this work pushes our knowledge and understanding of superhard materials into a promising direction through introducing new but easily accessible electronic level descriptors (at least two orders of magnitude faster than traditional direct calculation of elastic constants from first-principles), which will help in accelerating the materials discovery procedure that uses machine learning algorithms to discover advanced materials with fascinating and captivating properties.

Acknowledgements

Research reported in this publication was supported in part by the NSF (award number 2030128) and SC EPSCoR/IDeA Program under NSF OIA-1655740 via SC EPSCoR/IDeA SAN program (20-SA05). T.O. acknowledges Scientific Research Fund of Hunan Provincial Education Department (No. 20K127).

References

[1] Tehrani, A.M.; Brgoch, J. Hard and superhard materials: A computational perspective. J. Solid State Chem. 2019, 271, 47–58.

[2] Kanyanta, V. Hard, superhard and ultrahard materials: An overview. In Microstructure-Property Correlations for Hard, Superhard, and Ultrahard Materials; Springer: Berlin/Heidelberg, Germany, 2016; pp. 1–23.

[3] Park B. Current and Future Applications of Nanotechnology. Issues Environ. Sci. Technol. 2007, 24, 1–19.

[4] Presting H., Ko U. Future nanotechnology developments for automotive applications. Mater. Sci. Eng. C. 2003, 23, 737–741.

[5] Andre V.P., Han D., Shih W.M., Yan H. Challenges and Opportunities for structural DNA nanotechnology. Nat. Nanotechnol. 2011, 6, 763–772.

[6] Smith, R. L.; Sandly, G. E. An Accurate Method of Determining the Hardness of Metals, with Particular Reference to Those of a High Degree of Hardness. Proceedings of the Institution of Mechanical Engineers 1922, 102, 623–641.

[7] Dao, M. Multi-fidelity deep learning for extracting mechanical properties from instrumented indentation. Proc. Natl. Acad. Sci. 2020.

[8] Pharr, G. M.; Herbert, E. G.; Gao, Y. The Indentation Size Effect: A Critical Examination of Experimental Observations and Mechanistic Interpretations. Annu. Rev. Mater. Res. 2010, 40, 271–292.

[9] Iost, A.; Bigot, R. Indentation size effect: reality or artefact? J. Mater. Sci. 1996, 31, 3573–3577.

[10] E. Broitman. Indentation hardness measurements at macro-, micro-, and nanoscale: a critical overview. Tribo. Lett. 2017, 65, 23.

[11] Guo, X.; Li, L.; Liu, Z.; Yu, D.; He, J.; Liu, R.; Xu, B.; Tian, Y.; Wang, H.-T. Hardness of covalent compounds: Roles of metallic component and d valence electrons. Journal of Applied Physics 2008, 104, 023503.

[12] Haines, J.; Leger, J.M.; Bocquillon, G. Synthesis and design of superhard materials. Annu. Rev. Mater. Res. 2001, 31, 1–23.

[13] Vepˇrek, S. Nanostructured superhard materials. In Handbook of Ceramic Hard Materials; Riedel, R., Ed.; Wiley: Weinheim, Germany, 2000; p. 109

[14] Tehrani, A. M.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. J. Am. Chem. Soc. 2018, 140, 9844–9853.

[15] Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **2016**, 785-794.

[16] Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189-1232.

[17] Li, P. Robust LogitBoost and Adaptive Base Class (ABC) LogitBoost. *Proceedings of the Twenty-Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI'10),* **2012**, 302-311.

[18] J. Bennett and S. Lanning. The netflix prize. In *Proceedings of the KDD Cup Workshop*, **2007**, 3-6.

[19] Tian, Y.; Xu, B.; Zhao, Z. Microscopic theory of hardness and design of novel superhard crystals. Int. J. Refract. Met. Hard Mater. 2012, 33, 93–106

[20] Al-Fahdi, M.; Rodriguez, A.; Ouyang, T.; Hu, M. High-Throughput Computation of New Carbon Allotropes with Diverse Hybridization and Ultrahigh Hardness. Crystals 2021, 11, 783.

[21] He, C.; Sun, L.; Zhang, C.; Peng, X.; Zhang, K.; Zhong, J. Four superhard carbon allotropes: a first-principles study. Phys. Chem. Chem. Phys. 2012, 14, 8410.

[22] He, C.; Sun, L. Z.; Zhong, J. Prediction of superhard carbon allotropes from the segment combination method. J. Superhard Mater. 2013, 34, 386–399.

[23] Zhao, Z.; Xu, B.; Wang, L.-M.; Zhou, X.-F.; He, J.; Liu, Z.; Wang, H.-T.; Tian, Y. Three Dimensional Carbon-Nanotube Polymers. ACS Nano 2011, 5, 7226–7234.
[24] Miller, E. D.; Nesting, D. C.; Badding, J. V. Quenchable Transparent Phase of Carbon. Chem. Mater. 1997, 9, 18–22.
[25] Balaban, A. Carbon and Its Nets. Symmetry 2 1989, 397–416.
[26] Wang, Z.; Zhao, Y.; Tait, K.; Liao, X.; Schiferl, D.; Zha, C.; Downs, R. T.; Qian, J.; Zhu, Y.; Shen, T. A quenchable superhard carbon phase synthesized by cold compression of carbon nanotubes. Proceedings of the National Academy of Sciences 2004, 101, 13699–13702.
[27] Ivanovskaya VV, Ivanovskii AL. Simulation of novel superhard carbon materials based on fullerenes and nanotubes. J Superhard Mater 2010;32:67–87.
[28] Li Q, Ma Y, Oganov A, Wang H, Wang H, Xu Y, et al. Superhard monoclinic polymorph of carbon. Phys Rev Lett 2009, 102, 175506.
[29] Sheng X-L, Yan Q-B, Ye F, Zheng Q-R, Su G. T-carbon: a novel carbon allotrope. Phys. Rev. Lett. 2011, 106, 155703.
[30] Umemoto K, Wentzcovitch RM, Saito S, Miyake T. Body-centered tetragonal C4: a viable sp3 carbon allotrope. Phys. Rev. Lett. 2010, 104, 125504.
[31] Zhou X-F, Qian G-R, Dong X, Zhang L, Tian Y, Wang H-T. Ab initio study of the formation of transparent carbon under pressure. Phys. Rev. B 2010, 82, 134126.
[32] Zhu Q, Oganov AR, Salvadó AM, Pertierra P, Lyakhov AO. Denser than diamond: Ab initio search for superdense carbon allotropes. Phys. Rev. B 2011, 83, 193410.
[33] Mao WL, Mao H-K, Eng PJ, Trainor TP, Newville M, Kao C-C, Heinz DL, Shun J, Meng Y, Hemley RJ. Bonding changes in compressed superhard graphite. Science 2003, 302, 425–7.
[34] Diederich, F. & Kivala, M. All-carbon scaffolds by rational design. Adv. Mater. 2010; 22; 803–812.
[35] Hirsch, A. The era of carbon allotropes. Nat. Mater. 2010; 9; 868–871.
[36] Li, Q.; Ma, Y.; Oganov, A. R.; Wang, H.; Wang, H.; Xu, Y.; Cui, T.; Mao, H.-K.; Zou, G. Superhard Monoclinic Polymorph of Carbon. Phys. Rev. Lett. 2009, 102.
[37] Wang, J. T., Chen, C. F. & Kawazoe, Y. Low-temperature phase transformation from graphite to sp3 orthorhombic carbon. Phys. Rev. Lett. 2011, 106, 075501.
[38] Zhao, Z.; Xu, B.; Zhou, X.-F.; Wang, L.-M.; Wen, B.; He, J.; Liu, Z.; Wang, H.-T.; Tian, Y. Novel Superhard Carbon: C-Centered OrthorhombicC8. Phys. Rev. Lett. 2011, 107.
[39] Kurakevych, O. O. Superhard phases of simple substances and binary compounds of the B-C-N-O system: from diamond to the latest results (a Review). J. Superhard Mater. 2009, 31, 139–157.
[40] Li, J.; Zhang, R.-Q. New superhard carbon allotropes based on C20 fullerene. Carbon 2013, 63, 571–573.
[41] Liu, C.; Chen, M.; He, J.; Yu, S.; Liang, T. Superhard B2CO phases derived from carbon allotropes. RSC Adv. 2017, 7, 52192–52199.
[42] Zhao Y, He DW, Daemen LL, Shen TD, Schwarz RB, Zhu Y, et al. Superhard B-C-N materials synthesized in nanostructured bulks. J. Mater. Res. 2002, 17, 3139–45
[43] Solozhenko VL, Andrault D, Fiquet G, Mezouar M, Rubie DC. Synthesis of superhard cubic BC2N. Appl. Phys. Lett. 2001, 78, 1385–7
[44] Steele, B. A.; Oleynik, I. I. Ternary Inorganic Compounds Containing Carbon, Nitrogen, and Oxygen at High Pressures. Inorg. Chem. 2017, 56, 13321–13328.
[45] Li, Q.; Wang, J.; Zhang, M.; Li, Q.; Ma, Y. Superhard-driven search of the covalent network in the B3NO system. RSC Adv. 2015, 5, 35882–35887.
[46] Bhat, S.; Wiehl, L.; Molina-Luna, L.; Mugnaioli, E.; Lauterbach, S.; Sicolo, S.; Kroll, P.; Duerrschnabel, M.; Nishiyama, N.; Kolb, U.; Albe, K.; Kleebe, H.-J.; Riedel, R. High-Pressure Synthesis of Novel Boron Oxynitride B6N4O3 with Sphalerite Type Structure. Chem. Mater. 2015, 27, 5907–5914.

[47] Wang, S.; Oganov, A. R.; Qian, G.; Zhu, Q.; Dong, H.; Dong, X.; Esfahani, M. M. D. Novel superhard B–C–O phases predicted from first principles. Phys. Chem. Chem. Phys. 2016, 18, 1859–1863.
[48] Zhang, X.; Qin, J.; Sun, X.; Xue, Y.; Ma, M.; Liu, R. First-principles structural design of superhard material of ZrB4. *Physical Chemistry Chemical Physics* **2013**, *15*, 20894.
[49] Lyakhov, A. O.; Oganov, A. R. Evolutionary search for superhard materials: Methodology and applications to forms of carbon and TiO2. *Physical Review B* **2011**, *84*.
[50] Shi, X.; He, C.; Pickard, C.J.; Tang, C.; Zhong, J. Stochastic generation of complex crystal structures combining group and graph theory with application to carbon. Phys. Rev. B 2018, 97, 014104.
[51] He, C.; Shi, X.; Clark, S.J.; Li, J.; Pickard, C.J.; Ouyang, T.; Zhang, C.; Tang, C.; Zhong, J. Complex low energy tetrahedral polymorphs of group IV elements from first principles. Phys. Rev. Lett. 2018, 121, 175701.
[52] Yin, H.; Shi, X.; He, C.; Martinez-Canales, M.; Li, J.; Pickard, C.J.; Tang, C.; Ouyang, T.; Zhang, C.; Zhong, J. Stone-Wales graphene: A two-dimensional carbon semimetal with magic stability. Phys. Rev. B 2019, 99, 041405.
[53] Zhou, N.; Zhou, P.; Li, J.; He, C.; Zhong, J. Si-Cmma: A silicon thin film with excellent stability and Dirac nodal loop. Phys. Rev. B 2019, 100, 115425.
[54] Ouyang, T.; Cui, C.; Shi, X.; He, C.; Li, J.; Zhang, C.; Tang, C.; Zhong, J. Systematic enumeration of low-energy graphyne allotropes based on a coordination-constrained searching strategy. Phys. Status Solidi Rapid Res. Lett. 2020, 14.
[55] Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. Comput. Mater. Sci. 1996, 6, 15–50.
[56] Kresse, G.; Furthmüller, J. Efficient Iterative Schemes Forab Initiototal-Energy Calculations Using a Plane-Wave Basis Set. Phys. Rev. B 1996, 54, 11169–11186.
[57] Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. Phys. Rev. B 1999, 59, 1758–1775.
[58] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. Phys. Rev. Lett. 1996, 77, 3865–3868.
[59] Blöchl, P. E. Projector Augmented-Wave Method. Phys. Rev. B 1994, 50, 17953–17979.
[60] Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integrations. Phys. Rev. B 1976, 13, 5188–5192.
[61] Page, Y.L.; Saxe, P. Symmetry-general least-squares extraction of elastic data for strained materials from ab initio calculations of stress. Phys. Rev. B 2002, 65.
[62] Zhang, S.; Zhang, R. AELAS: Automatic ELAStic property derivations via high-throughput first-principles computation. Comput. Phys. Commun. 2017, 220, 403–416.
[63] Voigt, W. Lehrbuch der Kristallphysik; Springer Science and Business Media: Berlin/Heidelberg, Germany, 1966
[64] Reuss, A.; Angew, Z. Berechnung der Fließgrenze von Mischkristallen auf Grund der Plastizitätsbedingung für Einkristalle. J. Math. Mech. 1929, 9, 49–58
[65] Hill, R. The elastic behaviour of a crystalline aggregate. Proc. Phys. Soc. Sect. A 1952, 65, 349–354.
[66] Zener, C.M.; Siegel, S. Elasticity and anelasticity of metals. J. Phys. Chem. 1949, 53, 1468.
[67] Chung, D.H.; Buessem, W.R. The elastic anisotropy of crystals. J. Appl. Phys. 1967, 38, 2010–2012.
[68] Ranganathan, S.; Ostoja-Starzewski, M. Universal elastic anisotropy index. Phys. Rev. Lett. 2008, 101, 055504.
[69] Šim ˚unek, A.; Vacká˘r, J. Hardness of covalent and ionic crystals: First-principle calculations. Phys. Rev. Lett. 2006, 96, 085501.
[70] Mukhanov, V.A.; Kurakevych, O.O.; Solozhenko, V.L. Thermodynamic model of hardness: Particular case of boron-rich solids. J. Superhard Mater. 2010, 32, 167–176.

[71] Gao, F.; He, J.; Wu, E.; Liu, S.; Yu, D.; Li, D.; Zhang, S.; Tian, Y. Hardness of covalent crystals. Phys. Rev. Lett. 2003, 91, 015502.
[72] Li, K.; Wang, X.; Zhang, F.; Xue, D. Electronegativity identification of novel superhard materials. Phys. Rev. Lett. 2008, 100, 235504.
[73] Teter, D.M. Computational alchemy: The search for new superhard materials. MRS Bull. 1998, 23, 22–27.
[74] Chen, X.-Q.; Niu, H.; Li, D.; Li, Y. Modeling hardness of polycrystalline materials and bulk metallic glasses. Intermetallics 2011, 19, 1275–1281.
[75] Togo, A.; Tanaka, I. First principles phonon calculations in materials science. Scr. Mater. 2015, 108, 1–5.
[76] Tadano, T.; Tsuneyuki, S. First-Principles Lattice Dynamics Method for Strongly Anharmonic Crystals. *J. Phys. Soc. Japan* **2018**, *87*, 041015.
[77] Hoffmann, R.; Kabanov, A.A.; Golov, A.A.; Proserpio, D.M. Homo citans and carbon allotropes: For an ethics of citation. Angew. Chem. Int. Ed. 2016, 55, 10962–10976.
[78] Momma, K.; Izumi, F. VESTA 3for three-dimensional visualization of crystal, volumetric and morphology data. J. Appl. Crystallogr. 2011, 44, 1272–1276.
[79] Al-Fahdi, M.; Zhang, X.; Hu, M. Phonon transport anomaly in metavalent bonded materials: contradictory to the conventional theory. *Journal of Materials Science* **2021**, *56*, 18534–18549.
[80] Zhao, Y.; Al-Fahdi, M.; Hu, M.; Siriwardane, E. M. D.; Song, Y.; Nasiri, A.; Hu, J. High-Throughput Discovery of Novel Cubic Crystal Materials Using Deep Generative Neural Networks. *Adv. Sci.* **2021**, 2100566.
[81] Yue, S.-Y.; Qin, G.; Zhang, X.; Sheng, X.; Su, G.; Hu, M. Thermal transport in novel carbon allotropes with sp2 or sp3 hybridization: An ab initio study. Phys. Rev. B 2017, 95, 085207.
[82] Emery, A.A.; Wolverton, C. High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO3 perovskites. Sci. Data 2017, 4, 170153.
[83] Zhang, Z.; Tehrani, A. M.; Oliynyk, A.; Day, B.; Brgoch, J. Finding the Next Superhard Material through Ensemble Learning. *Adv. Mater.* 2020.
[84] Choudhary, K.; Decost, B.; Tavazza, F. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape. *Phys. Rev. Mater.* 2018, *2*.
[85] Choudhary, K.; Cheon, G.; Reed, E.; Tavazza, F. Elastic properties of bulk and low-dimensional materials using van der Waals density functional. *Phys. Rev. B* 2018, *98*.
[86] Doll, K.; Saunders, V.R.; Harrison, N.M. Analytical Hartree-Fock gradients for periodic systems. Int. J. Quantum Chem. 2000, 82, 1–13.
[87] Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. Int. J. Quantum Chem. 1976, 10, 325–340.
[88] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* 1965, *140*.
[89] Silvi, B.; Savin, A. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* 1994, *371*, 683–686.
[90] Fuentealba, P.; Chamorro, E.; Santos, J. C. Chapter 5 Understanding and using the electron localization function. *Theoretical and Computational Chemistry Theoretical Aspects of Chemical Reactivity* **2007**, 57–85.
[91] Lai, W.; Wang, Y.; Morelli, D. T.; Lu, X. From Bonding Asymmetry to Anharmonic Rattling in Cu12Sb4S13Tetrahedrites: When Lone-Pair Electrons Are Not So Lonely. *Adv. Funct. Mater.* **2015**, *25*, 3648–3657.
[92] Adler, J.; Parmryd, I. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytom. Part A 2010, 77, 733–742.
[93] Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N.E.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; et al. Matminer: An open source toolkit for materials data mining. Comput. Mater. Sci. 2018, 152, 60–69

[94] Jiang, X.; Zhao, J.; Jiang, X. Correlation between hardness and elastic moduli of the covalent crystals. *Comput. Mater. Sci.* **2011**, *50*, 2287–2290.
[95] Jiang, X.; Zhao, J.; Wu, A.; Bai, Y.; Jiang, X. Mechanical and electronic properties of B12-based ternary crystals of orthorhombic phase. *J. Phys. Condens. Matter.* **2010**, *22*, 315503.
[96] Miao, N.; Sa, B.; Zhou, J.; Sun, Z. Theoretical investigation on the transition-metal borides with Ta3B4-type structure: A class of hard and refractory materials. *Comput. Mater. Sci.* **2011**, *50*, 1559–1566.
[97] Gilman, J.J. Hardness e a strength microprobe. In The Science of Hardness Testing and its Research Applications; Westbrook, J.H., Conrad, H., Eds.; American Society of Metal: Metal Park, OH, USA, 1973; Chapter 4.
[98] Liu, A.Y.; Cohen, M.L. Prediction of new low compressibility solids. Science 1989, 245, 841–842.
[99] Gao, F.M.; Gao, L.H. Microscopic models of hardness. J. Superhard Mater. 2010, 32, 148–166.
[100] Pugh, S. XCII. Relations between the elastic moduli and the plastic properties of polycrystalline pure metals. London, Edinburgh Dublin Philos. Mag. J. Sci. **1954**, *45*, 823–843.
[101] Kanchana, V.; Vaitheeswaran, G.; Zhang, X.; Ma, Y.; Svane, A.; Eriksson, O. Lattice dynamics and elastic properties of the4felectron system: CeN. *Phys. Rev. B* **2011**, *84*.
[102] Varshney, D.; Sapkale, R.; Dagaonkar, G. J.; Varshney, M. Pressure dependent mechanical and thermodynamical properties of Hg0.91Mn0.09Te semiconductor. *Eur. Phys. J. B* 2011, *79*, 495–502.
[103] Niu, H.; Chen, X.-Q.; Liu, P.; Xing, W.; Cheng, X.; Li, D.; Li, Y. Extra-electron induced covalent strengthening and generalization of intrinsic ductile-to-brittle criterion. *Sci. Rep.* 2012, *2*.
[104] N. Frantsevich, F. F. Voronov, and S. A. Bokuta, Elastic Constants and Elastic Moduli of Metals and Insulators Handbook (Naukova Dumka, Kiev, 1983).
[105] Wu, S.-C.; Fecher, G.H.; Naghavi, S.S.; Felser, C. Elastic properties and stability of Heusler compounds: Cubic Co2YZ compounds with L21 structure. J. Appl. Phys. 2019, 125, 082523.
[106] J. F. Nye, Physical Properties of Crystals: Their Representation by Tensors and Matrices (Oxford Science Publications, Oxford, 1985).