

Biased screening for multi-component materials with Structures of Alloy Generation And Recognition (SAGAR)

Chang-Chun He^{a,1}, Ji-Hai Liao^{a,1}, Shao-Bin Qiu^a, Yu-Jun Zhao^a, Xiao-Bao Yang^{a,b,*}

^a Department of Physics, South China University of Technology, Guangzhou 510640, China

^b Guangdong Provincial Key-lab for Computational Science, Shenzhen 518000, China

ARTICLE INFO

Keywords:

Material exploration
Structural symmetry
Biased screening

2020 MSC:

20-04

20-08

ABSTRACT

Exploring new materials has attracted more and more attention, due to the improvement of technology and industry. Theoretically, the first-principles calculations combined with global optimization have been applied to predicting the properties of materials accurately and accelerate the discovery of new materials. For certain multi-component materials, the atoms are nearly at the lattice sites and the material exploration can be enhanced with biased screening. Based on the space symmetry, the calculations of duplicate structures can be avoided, and the specified constraints are applied to filter the energetically unstable candidates out. To provide the reasonable candidates for the first-principles calculations, we have released a program of Structures of Alloy Generation And Recognition (SAGAR), which can be conveniently used through the website. Herein, we will present several examples to show that the biased screening is practical and efficient, determining structures with high stabilities and novel properties. We have studied the distribution of up to 50% Cl vacancy concentration in NaCl and revealed semiconductor–metal transition in the boron/nitrogen co-doped diamond. The ground state anti-ferromagnetic configuration of VCl₂ is determined, as well as the low-lying hydrogenated-C₆₀ structures with high symmetry. Our results indicate that the biased screening with proper constraint will effectively enhance the material exploration.

1. Introduction

The physical, chemical and biological properties of materials rigorously depend on the atomic structures and chemical compositions of materials [1,2]. Combined with global optimization algorithm, the first-principles calculations based on density functional theory have become more and more crucial in material design, because various structures with novel properties have been predicted before the rapid development of experimental observation [3]. Group intelligence algorithms, as implemented in CALYPSO [4] and USPEX [5], have been successfully applied in the global optimization of stable structures, especially at high pressure. For example, H₂S was predicted to be a strongly anharmonic phonon-mediated superconductor [6], which has been confirmed experimentally under high pressure (around 150 GPa) with the T_c of 203 K [7]. Surprisingly, xenon is predicted to react with iron and nickel with the existence of XeFe₃/XeNi₃ in the Earth's inner core [8], providing a reasonable explanation why more than 90% of the expected amount of Xe is missing. Besides, the chemically inert helium and

sodium are predicted to form Na₂He at high pressure (>113 GPa) [9], where the eight-center-two-electron bonds within Na₈ cubes are stabilized by the electron pairs restricted in interstices sites. In addition, *ab initio* random structure searching algorithm [10] and random superlattices with atomic random distribution over different planes algorithm [11] predict a series of novel structures in recent decades [12–14].

The stable structures can be predicted based on the first-principles calculations combined with unbiased global optimization, however, the prediction of alloy can be enhanced with biased screening, because the lattice can be pre-determined and the atomic distribution is crucial. Cluster expansion (CE) method combined with first-principles calculations is employed to efficiently determine the ground state structures [15], which is applicable to simple lattices and complex lattices [16]. For alloy system, evolution algorithm has also been applied to select efficient cluster interactions when CE method is employed [17]. For example, vacancy (denoted by “□”) ordering has been found in Sc_{1-x}□_xS even at low temperature after numerous structures were evaluated by CE method [18], where proper metal vacancies will lower

* Corresponding author at: Department of Physics, South China University of Technology, Guangzhou 510640, China.

E-mail address: scxbyang@scut.edu.cn (X.-B. Yang).

¹ These authors contributed equally to this work.

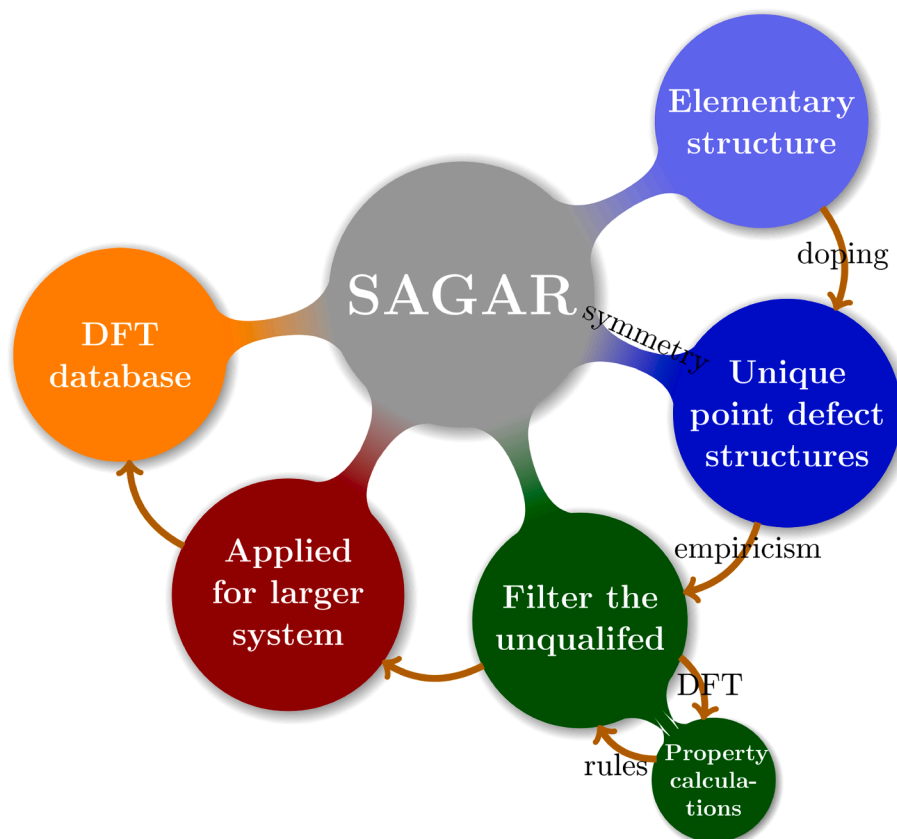


Fig. 1. The workflow of SAGAR, providing the biased screening method for the exploration of multi-component materials, carrying out structure generation, property calculation and database construction.

Fermi level and leave antibonding d band empty to enhance the stability. Moreover, degenerate gapped transparent conductors can cause a self-regulating instability, where metal vacancies will be spontaneously exothermically formed to promote conduction band electrons to decay into hole states [19]. The potential energy surface of boron is particularly complex due to the electron deficiency [20], therefore, it is a challenge to determine the local minimum structures because of the complex energy landscapes of elemental boron [21]. In our previous studies, we have determined the stable structures of boron clusters based on the triangular lattice fragments [22] and filtered the monolayers [23] with various vacancies distribution, indicating high efficiency of biased screening.

For the structures with certain concentration of vacancies [24], the vacancies are located at a pre-determined lattice, the initial distribution for local optimization is the dominant to the structural stabilities and chemical/physical properties. Similarly, enumerating the possible structures is an common method to explore novel properties for the complex multi-component system [28,29]. In order to confirm the ground state of ferromagnetism, various anti-ferromagnetic configurations should be considered, in which the opposite orientation of spins can be regarded as two elements and the problem is same to the determination of ground state of binary alloy [30,31]. For the system with hydrogen atoms absorption, the distribution of surface atoms with/without H atom is crucial, which can also be enumerated when the number of H atoms is small [25]. All these problems will be converted into the determination of multi-component materials, in which the duplicate configurations should be removed before the first-principles calculations, and the biased screening should be adopted to accelerate the searching procedure. In this paper, we propose a biased screening strategy to explore novel materials based on the domain knowledge, focusing on the systems with pre-determined lattices. We demonstrate

an effective avenue to determine the atomic distribution in the systems with high concentration vacancies and multi-components, which can be extended for the distribution of magnetic moments and adsorption with localized bonds. The calculations of duplicate structures are avoided based on the structural recognition. The specific constraints are applied to filter the energetically unstable candidates out to enhance the efficiency, which will also improve our understanding of given materials.

2. Method

Here, we use four examples to demonstrate that appropriate constraints will be derived according to the enumerated calculation, which will reduce the searching space and the computational cost. We firstly consider the 2D Na_2Cl of unconventional stoichiometries, corresponding to the NaCl crystals with high concentrations of Cl vacancies. It is not clear that whether there is an ordered distribution of Cl vacancies, similar to the one of binary alloys. For boron/nitrogen (B/N) doped bulk carbon (C) diamond structures, there are various local bonding of C/B/N atoms for this ternary alloy. The dependence of stabilities and properties on the local atomic environment can be derived from the results of small supercells, and the proper constraints will provide the biased search for novel structures in larger supercells. When there is a magnetic elements, we should take account of the possible magnetic configurations, where the symmetry of magnetic space group should be considered. To determine the ground state of magnetic configurations, all possible AFM configurations should be involved to confirm that whether the FM configuration is the ground state. For hydrogenated C_{60} , the C–H bonds are totally localized, where there is not more than one H atom on each C atom. Thus, the distribution of C atoms with/without H atom is crucial, which can be enumerated when the number of H atoms is small. Though the appropriate constraints will accelerate the searching, the rules are

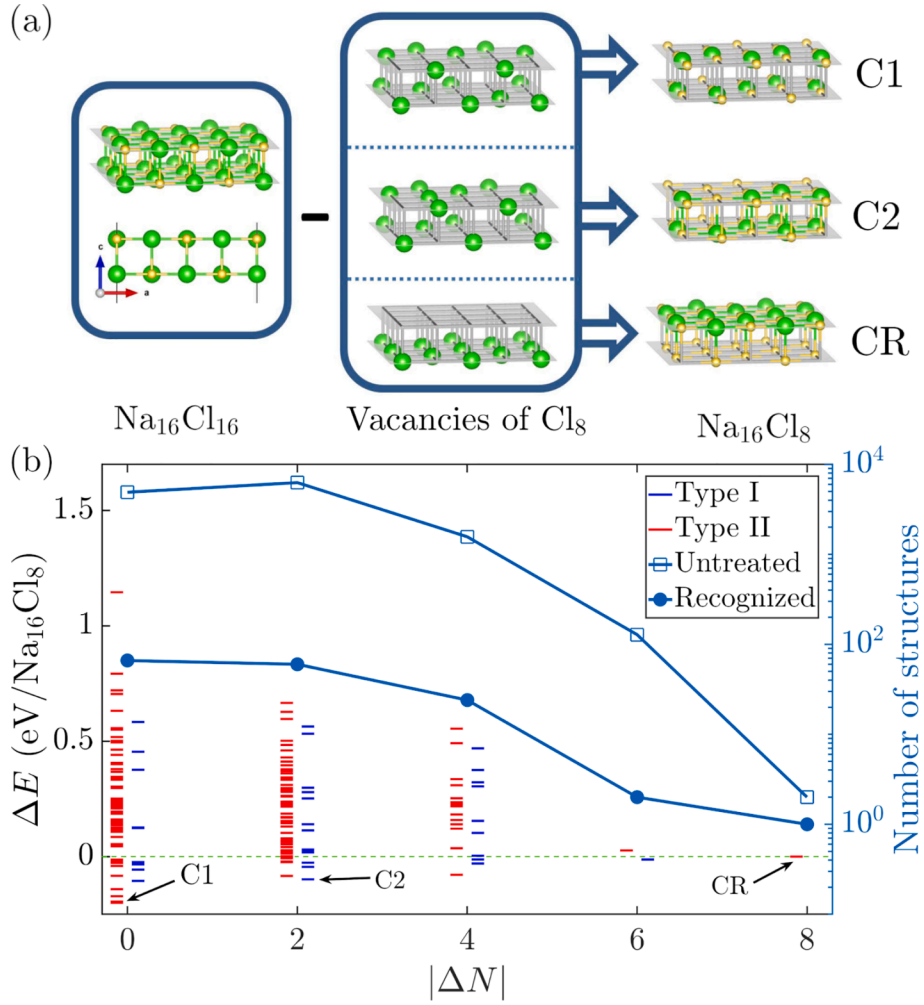


Fig. 2. (a) Schematic diagram of the process in generating a series of allotropes of Na₁₆Cl₈. (b) The energy of the allotropes of Na₁₆Cl₈ used the energy of CR configuration as a reference, as well as the number of configurations with/without recognized, as a function of $|\Delta N|$, the absolute value of the difference between the numbers of Cl atoms in the two layers. Type I: each Na atom has at least one Na–Cl bond; Type II: at least one Na atom has none Na–Cl bond.

different for these four systems and the high-throughput first-principles calculations are required.

In these four systems, the common tasks are to determine the atomic distribution and the structural recognition is necessary before the first-principle calculations. Among the possible candidates of C₆₀H_n, about 99% duplicate structures can be removed by the unique sequence number algorithm [32]. However, the number of unique structures is still growing exponentially as the number of H atom increases, which is still far away from the computation ability that we can afford. Therefore, it is impossible to generate or calculate all unique isomers, then it is genius to choose some special or representative structures for further calculations, namely, a biased screening with the specific constraints should be adopted to filter the potential structures, which can help us to economically explore the unknown fields. In fact, this is the application of Bayesian theory in material design fields [33], helping us to intelligently search novel structures.

Fig. 1 shows the workflow of Structures of Alloy Generation And Recognition (SAGAR) [34], providing the biased screening method for the exploration of multi-component materials. Firstly, all the possible point defect structures of small systems can be enumerated, where duplicate structures can be removed according to the structural symmetry. Combined the results from the first-principles calculations with the domain empirical knowledge, appropriate rules can be proposed to filter the larger systems with numerous configurations. Simultaneously, the database will be constructed with these structures and properties

from the first-principles calculations [35]. Note that there are various rules due to the variety of materials and the filtering the unqualified materials requires the analysis based on the domain knowledge.

All the first-principles calculations in this paper are based on the structural recognition by SAGAR [34], where the convenient usage can be realized through the website of online version for SAGAR with the only input of position file (including the lattice parameters and atomic positions). For a certain structure, we can obtain the primitive cell and all the possible super-cells with various number of atoms [36,37]. For a host crystal structure or molecule, the unique point defect structures at any concentration can be easily generated through our website, which can efficiently accelerate the searching of multi-components materials. Similarly, the unique magnetic configurations can also be generated to find possible anti-ferromagnetic configurations. Furthermore, the tetrahedral interstitial point defect structures in metal alloys can be generated for the design of hydrogen storage media [38].

For the given structures, the calculations were performed using the Vienna *Ab initio* Simulation Package (VASP) [39]. The projector-augmented plane wave approach [40] was adopted with the Perdew–Burke–Ernzerhof (PBE) scheme of generalized gradient approximation functional [41]. The energy cutoff of the plane wave was carefully tested and the criteria of the energies convergence were set to be 0.001 eV/Å. For 2D material, the distance of vacancy layer was also carefully tested. For total energy calculation, the Brillouin zone was sampled with allowed spacing between k-points in 0.1 Å⁻¹, with

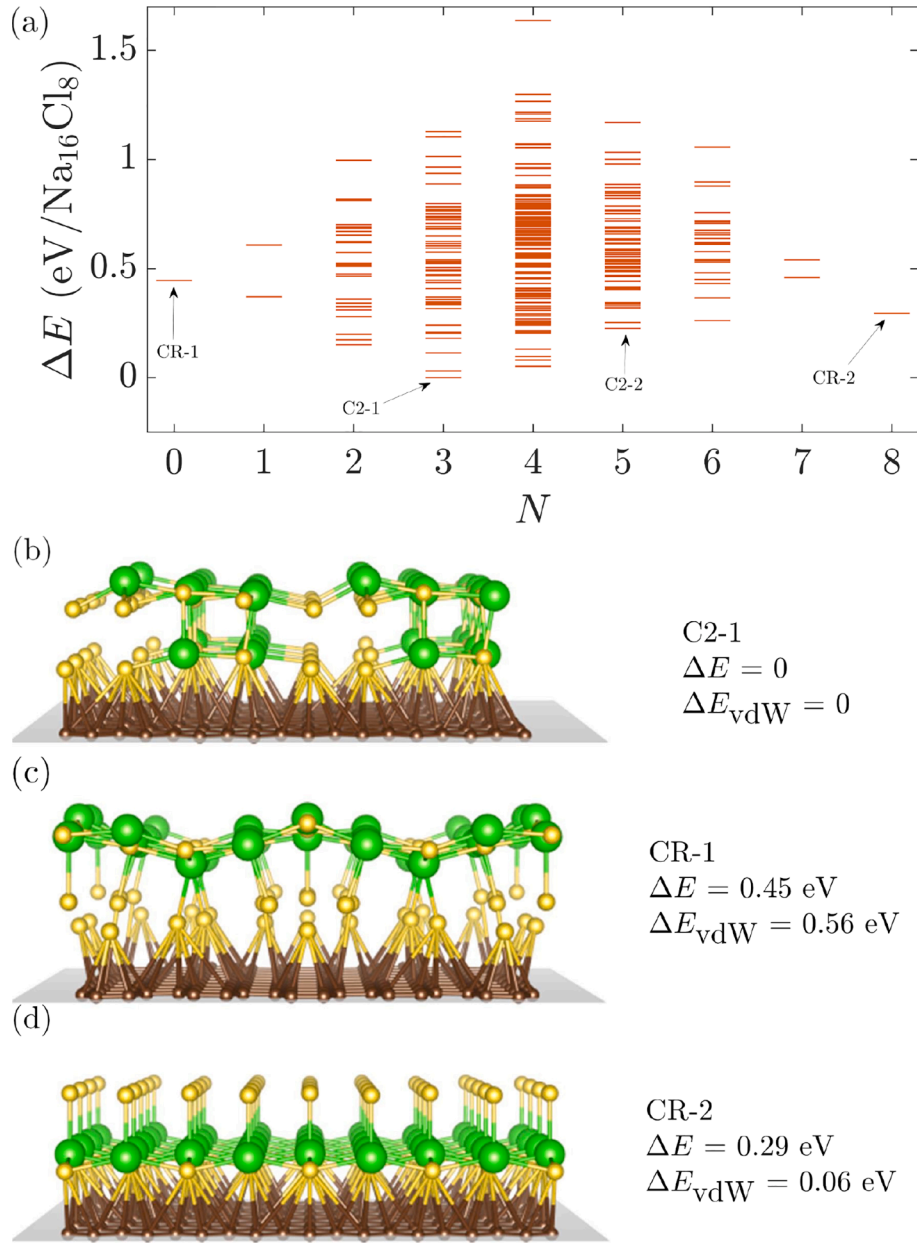


Fig. 3. (a) The relative energy of the allotropes of Na₂Cl on graphene as a function of N , which is the number of Cl atoms in graphene's adjacent layer. Representative configurations of Na₂Cl on graphene: (b) C2-1, with the lowest energy, (c) CR-1, a structure from Ref. [24] after full relaxation, (d) CR-2, the CR configuration adsorbed on graphene in which the NaCl layer is adjacent to graphene.

Γ -centered Monkhorst-Pack k -point grid for high-throughput calculations.

3. Results and discussion

In the following, we will take Na₂Cl, B/N doped diamond, VCl₂, and hydrogenated C₆₀ as four examples to demonstrate how to find out the appropriate constraints. We focus on the distribution of high Cl vacancies in the NaCl crystals and how the local structures of C/B/N atoms affect the structural stabilities. Considering the spin polarization, we determine the magnetic configuration of VCl₂. For hydrogenated C₆₀, we obtain the stable structures based on the distribution of C atoms with/without H atom.

3.1. Na₂Cl crystals on graphene

Bulk NaCl is the prototype of ionic crystals with a wide band gap of ~ 9 eV [42], in which the crystallographic plane (100) is the most energetically favorable surface. Thin insulating NaCl films have recently received an increasing interest as inert barriers to decouple the electronic states of adsorbed nanostructures from those of the metallic substrate [43]. For example, an exotic hexagonal NaCl thin films on the (110) diamond surface were crystallized in the experiment following a theoretical prediction based on USPEX code [44]. In bulk NaCl and the corresponding thin films, the intrinsic Cl vacancy defects are well defined and stable, which have been studied commonly as color centers both experimentally and theoretically [45,46]. It is particularly intriguing that the unconventional stoichiometric Na₂Cl and Na₃Cl thin films were also experimentally prepared on graphene surface under ambient conditions [24], where the highly ordered nonstoichiometric

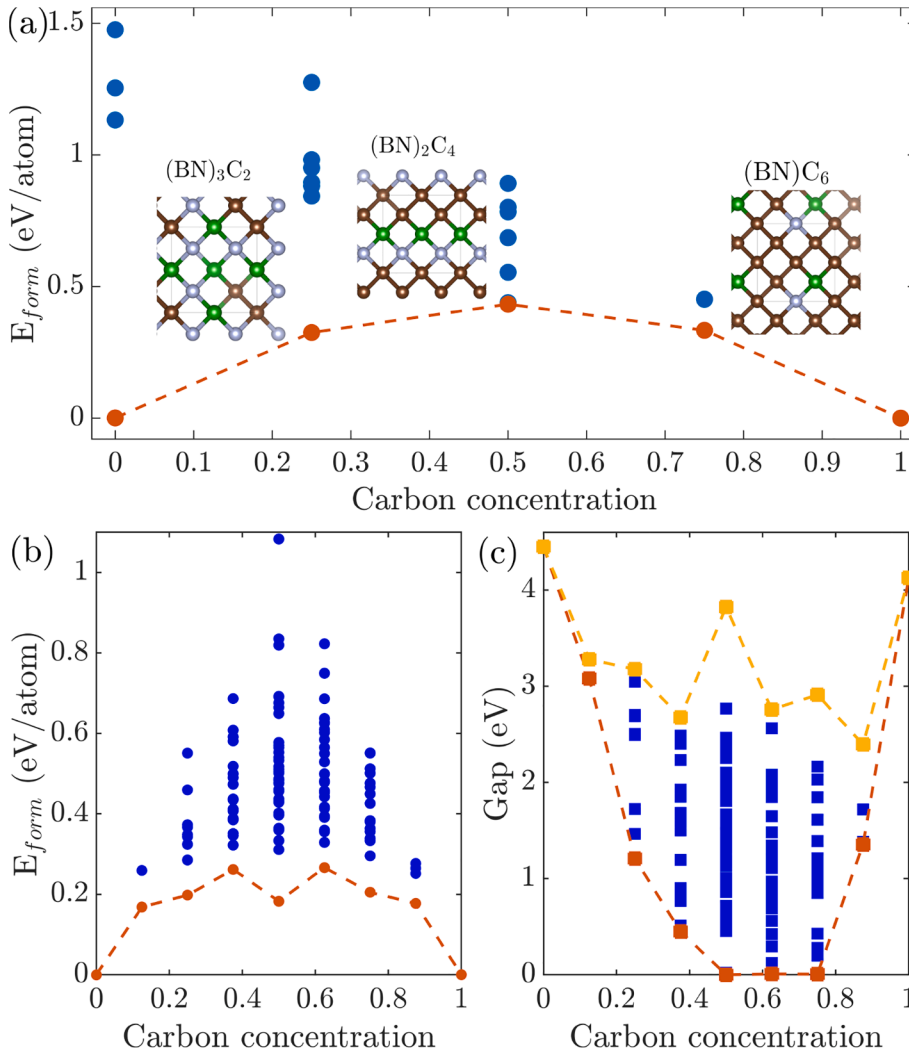


Fig. 4. (a) The formation energy of 21 structures (denoted by triangles) of $C_{1-2x}(BN)_x$ in the conventional unit cell, the brown line show the structures at the convexhull, the brown dash line indicates the ground state structures at each carbon concentration, and the crystal structures are plotted next to the triangles. (b) The formation energy of 127 structures (denoted by blue circles) of $C_{1-2x}(BN)_x$. (c) Band gap of these structures (denoted by blue squares). The structures at the convexhull are denoted by red points.

Na_2Cl film was reported to contain a $NaCl(100)$ layer and a pure Na layer. However, the uncertainties seem to be too large for one to be fully confident in the exact crystal structure.

With a $2 \times 2 \times 1$ supercell of $NaCl(100)$ bilayer ($Na_{16}Cl_{16}$), we have considered the unique configurations with the 50% Cl vacancy concentration. The schematic diagram of the process in generating a series of allotropes of $Na_{16}Cl_{18}$ is shown in Fig. 2a, in which the positions of Cl vacancies are shown in the middle. We have determined the unique 153 $Na_{16}Cl_{18}$ isomers by SAGAR [34] with the structural recognition, among the $C_{16}^8 = 12870$ configurations from combination. The difference between the numbers of Cl atoms in the two layers is defined as $|\Delta N|$, and the one with $|\Delta N| = 8$ is reported to be most stable [24]. In this structure, all Cl vacancies are in the same layer, labeled as CR (see Fig. 2b). The number of unique isomers increases as $|\Delta N|$ decreases, and most of isomers of $Na_{16}Cl_{18}$ correspond to $|\Delta N| = 0$. Note that a few configurations with $|\Delta N| = 0, 2, 4, 6$ contain lower energies compared to CR, where the one with lowest energy corresponds to $|\Delta N| = 0$ (labeled as C1). Thus, our enumerated calculation shows that the recently reported CR structure is not the ground state of $Na_{16}Cl_{18}$, though it is more energetically stable than $\sim 80\%$ of the possible isomers. For the Na_2Cl configurations from NaCl, most of the configurations belong to Type I, in which each Na atom has at least one Na-Cl bond. There might be Na atoms without Cl neighboring (type II) and the Na atoms will be largely relaxed. The C1 structure did not contain the Na atoms without Cl neighboring, maintaining high structural stabilities.

To investigate the adsorption of Na_2Cl on graphene, we choose a

rectangle supercell containing 180/64/32 C/Na/Cl atoms with $a = 22.14 \text{ \AA}$ and $b = 21.30 \text{ \AA}$, to achieve a small lattice mismatch of $\sim 3\%$. Note that two different surfaces of a configuration of Na_2Cl adsorb on graphene, resulting in nonequivalent structures. The CR structure with $|\Delta N| = 8$ corresponds to $N = 0$ (CR-1) and $N = 8$ (CR-2), where N is the number of Cl atoms adjoining graphene. Previous studies have concluded that the unconventional crystallization of Na_2Cl on graphene were attributed to cation- π interactions between the Na ions and the aromatic rings in the graphenes. Thus, we compared the results with and without ignored van der Waals interaction. According to the relative total energies shown in Fig. 3a, the most stable adsorption configuration (C2-1 shown in Fig. 3b) is the one with $|\Delta N| = 2$. For the recently reported CR structure with $|\Delta N| = 8$, the stable adsorption configuration is the one with NaCl layer adjoining graphene, regardless of whether or not considering van der Waals interaction.

3.2. B/N co-doped diamond

Traditional semiconductor materials include II-VI, III-V group and their alloys, whose band gap can be effectively engineered by the compositions [47]. With similar atomic radius, the band gap control in graphene through boron and nitrogen co-doping has been intensively investigated in recent years [48,49]. Here we take the B/N substituting C in diamond as an example to investigate how the local motifs modulate the structural stabilities and electronic properties in this ternary semiconductor.

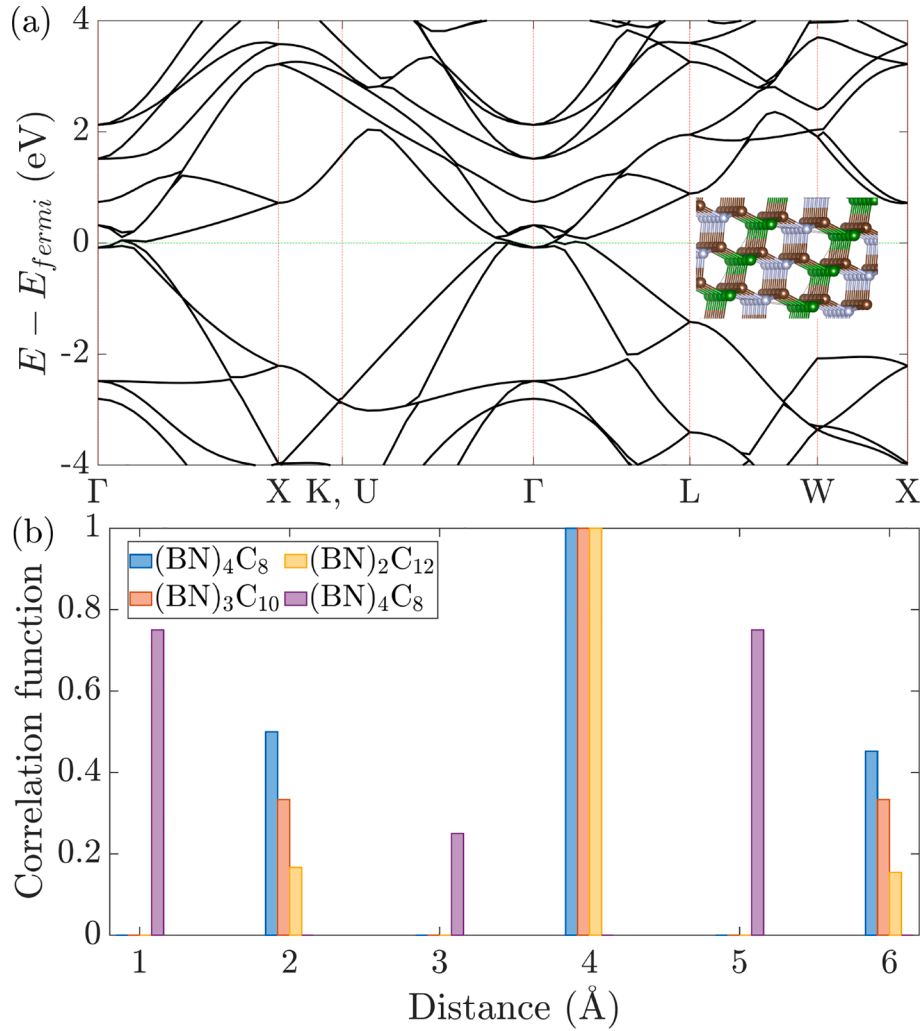


Fig. 5. (a) Structure of the metallic $(\text{BN})_4\text{C}_8$, the band structures are plotted in the subfigure. (b) Correlation function of $(\text{BN})_x\text{C}_{1-x}$ structures, the correlation function of metallic structures are plotted by bar graph, while the purple bar graph represents a $(\text{BN})_4\text{C}_8$ structure with band gap of 3.8 eV.

With the diamond lattice, we firstly selected a conventional unit cell of 8 atoms to enumerate all the possible configurations of $\text{C}_{1-2x}(\text{BN})_x$ without any constraint conditions. The formation energy is defined by $E_f = E_{\text{C}_{1-2x}(\text{BN})_x} - (1-2x)E_{\text{C}} - xE_{\text{BN}}$, where E_{C} and E_{BN} are the energies of diamond and bulk BN structures. The formation energies are found to be positive and the alloy should be formed at high temperature. For each given C concentration except the bulk diamond, our calculations show that the B and N atoms will form bonds in the structures with lowest formation energy in Fig. 4a. It indicates that the B-N bonds will extraordinarily enhance the structural stability rather than B-B bonds or N-N bonds. Therefore, a constraint condition that the configurations with B-B or N-N bonds should be filtered out before we proceed the first-principles calculations.

Herein, we select a diamond supercell with 16 atoms to demonstrate the biased screening of $\text{C}_{1-2x}(\text{BN})_x$. According to the formation energy showed in the Fig. 4b, we can conclude that the $\text{C}_{1-2x}(\text{BN})_x$ is a phase separation system at low temperature. We employ a linear regression (LR) bond energy model to analyze the dependence of energies on the bonds among B/C/N atoms, where the Lasso regression [50] is applied to remove unnecessary parameters in this LR model. As a result, the fitting error of energy is found to be ~ 25 meV/atom with just one pair (B-N) and two three-body interactions (C-C-C, C-N-C). Particularly, the interactions of B-N and C-C-C clusters are negative, indicating that C-C bonds and B-N bonds will remarkably enhance the stability in the $\text{C}_{1-2x}(\text{BN})_x$ system. The interaction of C-N-C cluster is positive, because

the N atoms should be preferably bonded with B atoms. To search the possible stable structures in larger supercells, we can exclude the configurations with C-N-C clusters to reduce the computational cost.

Interestingly, the band gap will vanish at some certain carbon concentrations as shown in Fig. 4c, which induce the semiconductor-metal transition. Fig. 5a depicts the band structures of a metallic $(\text{BN})_4\text{C}_8$ structure with the semimetal, where the valance band maximum and conduction band minimum cross at the Fermi surface. This configuration is found to be a layered stacked structure, where each plane is totally occupied by B/C/N atoms. To understand the origin of metallic properties, we consider the B-N correlation function, which is defined by $f(k) = \frac{1}{\text{Num}} \sum_{i=1}^{\text{Num}} n_{\text{B},r_k}^i / n_{r_k}$, where Num represents the total number of boron atoms in the unit cell, n_{r_k} represents the total coordinate number at the distance of r_k , where k is the order of near neighbor, and n_{B,r_k}^i represents the number of B atoms at the distance of r_k away from the i -th N atom. As shown in Fig. 5b, there are all C atoms at the nearest neighbor of B/N atoms for the metallic structures, while all BN bonds form at the nearest neighbor for the $(\text{BN})_4\text{C}_8$ structure with a large gap (shown in purple bar). The values of B-N correlation function are large at 1/3/5 th nearest neighbor for semiconducting $\text{C}_{1-2x}(\text{BN})_x$, while they are large 2/4/6 th nearest neighbor for metallic ones. Thus, we can divide the diamond structure into two sub-lattices and the $\text{C}_{1-2x}(\text{BN})_x$ tend to be metal when B/N atoms occupy the same sub-lattice.

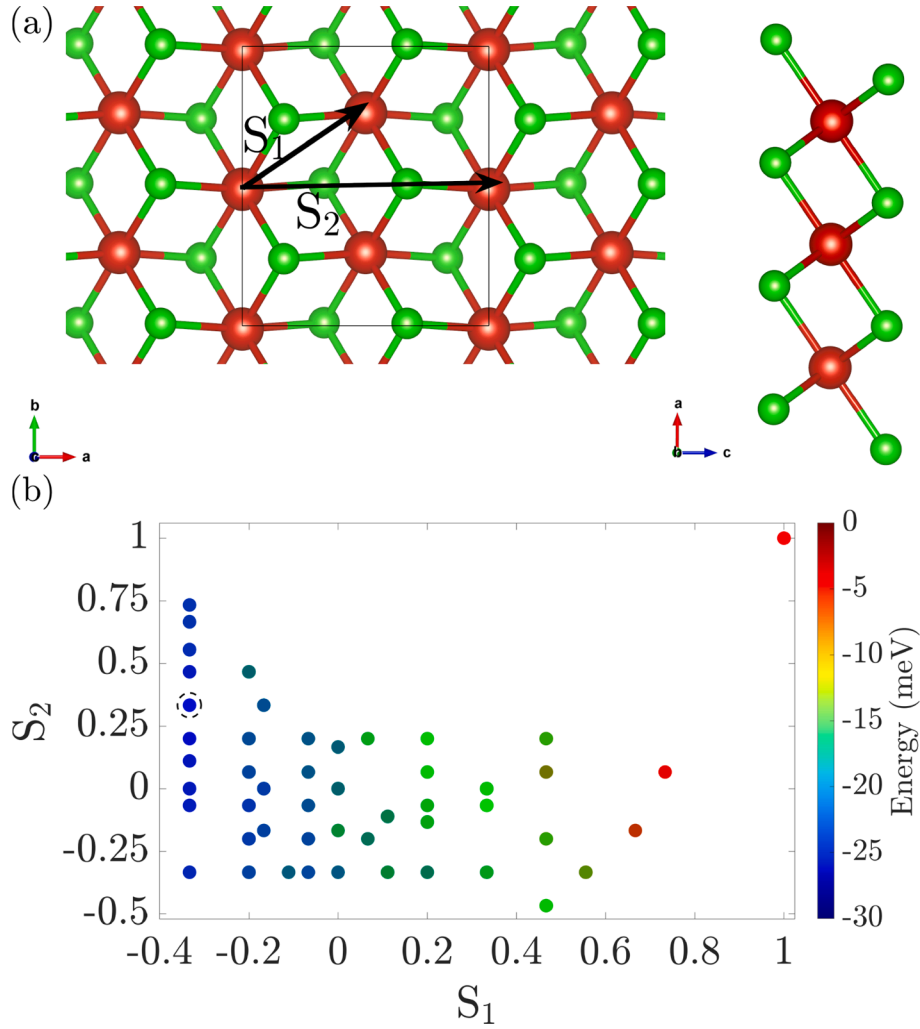


Fig. 6. (a) views of VCl_2 as seen from the z axis (left) and the y axis (right), the red/green atoms represent V/Cl atoms, respectively. The unit cell is also shown. (b) The relation of correlation coefficients, where the red/blue color is correspondent to high/low energy.

3.3. Magnetic properties in VCl_2

2D materials, such as graphene and its analogues, have displayed novel physical chemical properties, whose potential applications in electronics devices have also attracted extensive interests [51]. After the intrinsic magnetism at finite temperatures in 2D material CrI_3 was confirmed experimentally [52], the metal-functionalized monolayer magnetic materials (MoN_2 [53], Fe_2Si [54], CrB_4 [55]) are promising to have potential application in spintronics fields due to high Curie temperature above room-temperature. However, it is a nontrivial problem to confirm whether the FM configuration is the ground state or not, because all the AFM configurations should be taken into account. On the other hand, it is significant to calculate enough magnetic configuration to determine a proper exchange-coupling parameters in Ising model or Heisenberg model, which are used for the Monte Carlo simulations to obtain correct Curie temperature. Hence, as many AFM configurations as possible should be generated and calculated.

As shown in Fig. 6a, we take 2D VCl_2 with the space-group of C2/m as an example to demonstrate how to determine the stable magnetic configuration. Hermite normal form matrices are employed to generate all possible supercells with given sizes [36] and all the AFM configurations are generated by the combination, which is further filtered to obtain the unique configurations by SAGAR [34]. Due to the expensive computational cost, we only screen all the supercells with the sizes less than 6. However, still numerous magnetic configurations need to be

evaluated, which is far away beyond our computational ability if there is not any constraint. The clustering of spins with the same direction will reveal FM characteristics, therefore, the AFM configurations can be constructed with the dispersed distribution of parallel spins. Similar to the correlation coefficients in CE method [15], we define $S_1 = \frac{1}{6 \times \text{Num}} \sum_{\langle i,j \rangle} \sigma_i \sigma_j$ and S_2 is defined by $S_2 = \frac{1}{6 \times \text{Num}} \sum_{\langle i,j \rangle} \sigma_i \sigma_j$, where i, j are nearest/next-nearest neighboring for S_1 / S_2 , respectively. Num is the number of magnetic atoms and such a definition will maintain the absolute values of S_1, S_2 not more than 1. According to the dependence of magnetic configuration stabilities on S_1 and S_2 (shown in Fig. 6b), we find the AFM configurations will be lower in energy than FM ($S_1 = S_2 = 1$), confirming that AFM coupling will be more energetically favorable in this 2D VCl_2 system. For the ground state magnetic configuration (marked with the dark dash circle), the AFM coupling is energetically preferable at the nearest neighboring, while it is not consistent for the next-nearest neighboring as shown in Fig. 7a.

Fig. 7 shows the magnetic configurations with low energies, where the energy difference of conventional AFM configuration (d) is 27.78 meV compared to the FM one. In the conventional AFM configuration, the spins with same direction will distribute in a line, while those AFM configurations with lower energies prefer a zigzag pattern. It indicates that spins at the next-nearest neighboring should be FM coupling, though the ones at the nearest neighboring is AFM coupling. Unexpectedly, according to the Goodenough–Kanamori rules [56,57], the magnetic ordering is usually be FM when the metal–nonmetal–metal

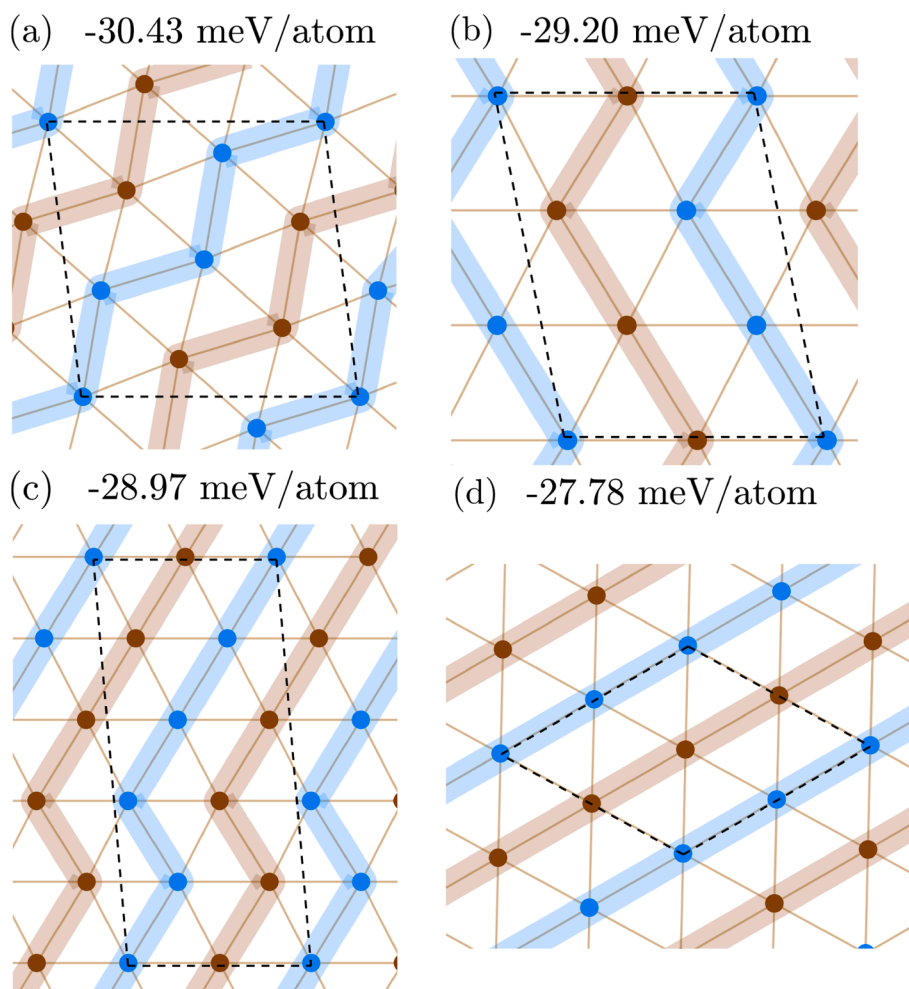


Fig. 7. (a)–(c) shows the top three structures with the lowest energy, the brown/cyan atoms represent spin up/down, respectively, (d) depicts a conventional AFM configuration in planar triangular lattice, where the brown points represent spin up vanadium atoms and cyan points represent spin down vanadium atoms, while the chloride atoms are ignored. The energy difference compared with the FM configuration is shown above each subfigure.

bond angle is near 90° , while AFM occurs if the angle is greater than 90° . In this VCl_2 system, the magnetic ordering is found to be AFM and the angle of V-Cl-V is about 90° .

3.4. Hydrogenated C_{60}

Organic semiconductors have rich applications in fields of organic lighting systems due to the flexibility and low cost, which have been extensively used in organic light-emitting devices (OLEDs) and organic field-effect transistors (OFETs) [58]. Since the discovery of C_{60} in 1985 [26], scientists have found that the hydrogenation of C_{60} would result in a considerable chemical modulation of C_{60} [59], while the number of C_{60} derivatives is enormous for theoretical prediction. Previous experiment showed that the $\text{C}_{60}\text{H}_{18}$ with C_{3v} point group is the ground state structure [27]. To reduce the computational cost, we focus on the candidates with high symmetry.

Taking C_{60}H_n as an example, we should determine the distribution of C atoms with/without H atom. Attributed to I_h group of C_{60}H_n with 120 unique permutation symmetry operations, the point group of high symmetric structures C_{60}H_n must be the subgroup, which has many nontrivial subgroups with 2, 3, 4, 5, 6, 8, 10, 12, 20, 24 elements. To generate $\text{C}_{60}\text{H}_{18}$ configurations with six fold symmetry, we can use all the unique C_{60}H_3 and operate them with the six-elements subgroup, where the 303 three-H atoms structures can be obtained from the unique sequence algorithm [32]. Note that the configurations with duplicate positions of H atoms should be removed.

Based on the structural recognition, we can generate the highly symmetric $\text{C}_{60}\text{H}_{18}$ structures with 3, 6, 12-fold symmetry by this method. Additionally, the high symmetric structures of $\text{C}_{60}\text{H}_{36}$ can also be generated similarly and several isomers without any symmetry are also calculated for comparison. As shown in Fig. 8a, we confirm that the $\text{C}_{60}\text{H}_{18}$ with C_{3v} point group (which can be viewed at the $\text{C}_{60}\text{H}_{18}$ website) is the ground state structure which is in agreement with the experimental study [27]. Though the configuration with highest symmetry is not most stable, the configurations with H disordered distributions are less stable, which can be filtered out firstly to reduce the computational cost. Note that the $\text{C}_{60}\text{H}_{18}$ structures whose carbon atoms linked with hydrogen atoms can connect with each other through nearest neighborhood will be more stable, indicating that H atoms tend to be aggregated rather dispersive when the number of H atoms are small. Similarly, the $\text{C}_{60}\text{H}_{36}$ structure with T point group (which can be viewed at the $\text{C}_{60}\text{H}_{36}$ website) is found to be ground state in our calculations, as shown in the Fig. 8b. However, we find that the hydrogen atoms tend to be dispersive in this system, and those structures with aggregated hydrogen atoms are less stable.

4. Conclusion

In summary, we have effectively reduced the cost of the first-principles calculations with the usage of SAGAR, where the biased screening method will further enhance the efficiency of exploring multi-component materials. As a result, we have determined the distribution

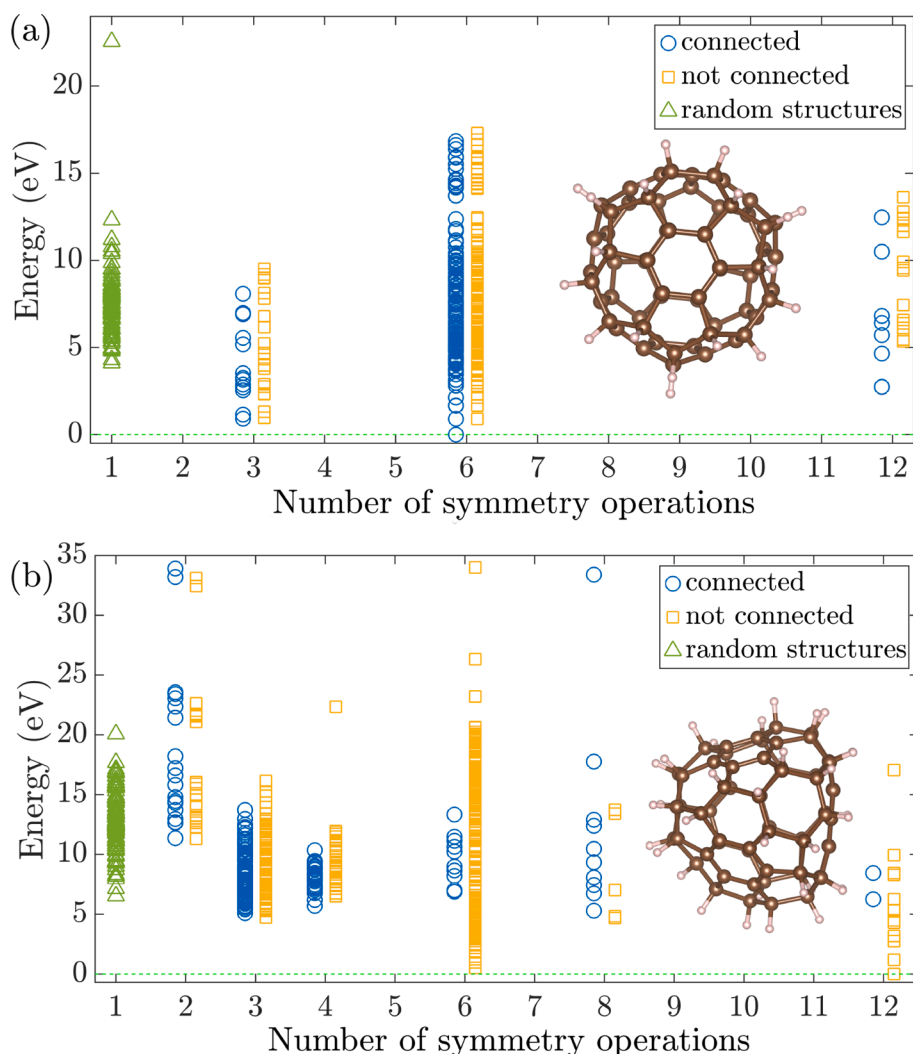


Fig. 8. (a)/(b) show that the energy of $C_{60}H_{18}/C_{60}H_{36}$ isomers varies with number of symmetry operation, the zero energy level is set to be the energy of the ground state structure we have found. The circle symbols represent those structures with nearest neighborhood (NN) connected carbon atoms, while the square symbols represent those structures without connected NN, it is clear that the circle symbols possess the lower energy than square symbols.

of up to 50% Cl vacancy concentration in NaCl, which is energetically favorable than those reported up to now. The semiconductor–metal transition in BCN systems is attributed to B–N atomic distribution, where the electronic properties are dominated by the B–N correlation function. The AFM configurations of VCl_2 with lower energies prefer a zigzag pattern, indicating that spins at the next-nearest neighboring should be FM coupling, though the ones at the nearest neighboring is AFM coupling. For hydrogenated- C_{60} , the stable structures with high symmetry are more stable than those with the P1 symmetry, where H atoms tend to be dispersed when the number of H atoms are small. Thus, the biased screening with SAGAR will provide an effective avenue for the material exploration, in which the usage of website will facilitate the generation of possible configurations.

CRediT authorship contribution statement

Chang-Chun He: Methodology, Software, Writing - original draft. **Ji-Hai Liao:** Methodology, Software, Writing - original draft. **Shao-Bin Qiu:** Visualization, Investigation. **Yu-Jun Zhao:** Supervision, Writing - review & editing, Validation. **Xiao-Bao Yang:** Supervision, Writing - review & editing, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Guangdong Provincial Key-lab for Computational Science and Materials Design program (No.2019B030301001) and National Natural Science Foundation of China (No. U1601212).

References

- [1] T. Le, V.C. Epa, F.R. Burden, D.A. Winkler, Quantitative structure–property relationship modeling of diverse materials properties, *Chem. Rev.* 112 (5) (2012) 2889–2919, <https://doi.org/10.1021/cr200066h>, url:https://doi.org/10.1021/cr200066h.
- [2] A.R. Oganov, C.J. Pickard, Q. Zhu, R.J. Needs, Structure prediction drives materials discovery, *Nat. Rev. Mater.* 4 (5) (2019) 331–348, <https://doi.org/10.1038/s41578-019-0101-8>, url:https://doi.org/10.1038/s41578-019-0101-8.
- [3] S.M. Woodley, R. Catlow, Crystal structure prediction from first principles, *Nat. Mater.* 7 (12) (2008) 937–946, <https://doi.org/10.1038/nmat2321>, url:https://doi.org/10.1038/nmat2321.

- [4] Y. Wang, J. Lv, L. Zhu, Y. Ma, Crystal structure prediction via particle-swarm optimization, *Phys. Rev. B* 82 (2010), <https://doi.org/10.1103/PhysRevB.82.094116>, 094116 url:<https://link.aps.org/doi/10.1103/PhysRevB.82.094116>.
- [5] A.O. Lyakhov, A.R. Oganov, H.T. Stokes, Q. Zhu, New developments in evolutionary structure prediction algorithm uspx, *Comput. Phys. Commun.* 184 (4) (2013) 1172–1182, <https://doi.org/10.1016/j.cpc.2012.12.009>, url:<http://www.sciencedirect.com/science/article/pii/S0010465512004055>.
- [6] I. Errea, M. Calandra, C.J. Pickard, J. Nelson, R.J. Needs, Y. Li, H. Liu, Y. Zhang, Y. Ma, F. Mauri, High-pressure hydrogen sulfide from first principles: A strongly anharmonic phonon-mediated superconductor, *Phys. Rev. Lett.* 114 (2015), <https://doi.org/10.1103/PhysRevLett.114.157004>, 157004 url:<https://link.aps.org/doi/10.1103/PhysRevLett.114.157004>.
- [7] A.P. Drozdov, M.I. Erements, I.A. Troyan, V. Ksenofontov, S.I. Shylin, Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system, *Nature* 525 (7567) (2015) 73–76, <https://doi.org/10.1038/nature14964>, url:<https://doi.org/10.1038/nature14964>.
- [8] L. Zhu, H. Liu, C.J. Pickard, G. Zou, Y. Ma, Reactions of xenon with iron and nickel are predicted in the earth's inner core, *Nat. Chem.* 6 (7) (2014) 644–648, <https://doi.org/10.1038/nchem.1925>, url:<https://doi.org/10.1038/nchem.1925>.
- [9] X. Dong, A.R. Oganov, A.F. Goncharov, E. Stavrou, S. Lobanov, G. Saleh, G.-R. Qian, Q. Zhu, C. Gatti, V.L. Deringer, A stable compound of helium and sodium at high pressure, *Nat. Chem.* 9 (5) (2017) 440–445, <https://doi.org/10.1038/nchem.2716>, url:<https://doi.org/10.1038/nchem.2716>.
- [10] C.J. Pickard, R.J. Needs, Ab initio random structure searching, *J. Phys.: Condens. Matter* 23 (5) (2011), <https://doi.org/10.1088/0953-8984/23/5/053201>, 053201.
- [11] V. Stevanović, Sampling polymorphs of ionic solids using random superlattices, *Phys. Rev. Lett.* 116 (2016), <https://doi.org/10.1103/PhysRevLett.116.075503>, 075503 url:<https://link.aps.org/doi/10.1103/PhysRevLett.116.075503>.
- [12] C.J. Pickard, R.J. Needs, High-pressure phases of silane, *Phys. Rev. Lett.* 97 (2006), <https://doi.org/10.1103/PhysRevLett.97.045504>, 045504 url:<https://link.aps.org/doi/10.1103/PhysRevLett.97.045504>.
- [13] A. Reinhardt, C.J. Pickard, B. Cheng, Predicting the phase diagram of titanium dioxide with random search and pattern recognition, *Phys. Chem. Chem. Phys.* 22 (2020) 12697–12705, <https://doi.org/10.1039/D0CP02513E>, url:<https://doi.org/10.1039/D0CP02513E>.
- [14] X. Yang, J. Wang, J. Zheng, M. Guo, R.-Z. Zhang, Screening for planar carbon allotropes using structure space sampling, *J. Phys. Chem. C* 124 (11) (2020) 6379–6384, <https://doi.org/10.1021/acs.jpcc.9b10778>, url:<https://doi.org/10.1021/acs.jpcc.9b10778>.
- [15] J. Sanchez, F. Ducastelle, D. Gratias, Generalized cluster description of multicomponent systems, *Physica A* 128 (1) (1984) 334–350, [https://doi.org/10.1016/0378-4371\(84\)90096-7](https://doi.org/10.1016/0378-4371(84)90096-7), url:<http://www.sciencedirect.com/science/article/pii/0378437184900967>.
- [16] A. Seko, K. Shitara, I. Tanaka, Efficient determination of alloy ground-state structures, *Phys. Rev. B* 90 (2014), <https://doi.org/10.1103/PhysRevB.90.174104>, 174104 url:<https://link.aps.org/doi/10.1103/PhysRevB.90.174104>.
- [17] G.L.W. Hart, V. Blum, M.J. Walorski, A. Zunger, Evolutionary approach for determining first-principles hamiltonians, *Nat. Mater.* 4 (5) (2005) 391–394, <https://doi.org/10.1038/nmat1374>, url:<https://doi.org/10.1038/nmat1374>.
- [18] G.L.W. Hart, A. Zunger, Origins of nonstoichiometry and vacancy ordering in $\text{sc-1-x}\square\text{-xs}$, *Phys. Rev. Lett.* 87 (2001), <https://doi.org/10.1103/PhysRevLett.87.275508>, 275508 url:<https://link.aps.org/doi/10.1103/PhysRevLett.87.275508>.
- [19] O.I. Malyi, M.T. Yeung, K.R. Poeppelmeier, C. Persson, A. Zunger, Spontaneous non-stoichiometry and ordering in degenerate but gapped transparent conductors, *Mater. 1* (1) (2019) 280–294, <https://doi.org/10.1016/j.matt.2019.05.014>, url:<http://www.sciencedirect.com/science/article/pii/S2590238519300396>.
- [20] T. Ogitsu, E. Schwegler, G. Galli, β -rhombohedral boron: At the crossroads of the chemistry of boron and the physics of frustration, *Chem. Rev.* 113 (5) (2013) 3425–3449.
- [21] V.L. Deringer, C.J. Pickard, G. Csányi, Data-driven learning of total and local energies in elemental boron, *Phys. Rev. Lett.* 120 (2018), 156001.
- [22] S.G. Xu, Y.J. Zhao, J.H. Liao, X.B. Yang, Understanding the stable boron clusters: A bond model and first-principles calculations based on high-throughput screening, *J. Chem. Phys.* 142 (21) (2015), <https://doi.org/10.1063/1.4922059>, 214307 url:<https://doi.org/10.1063/1.4922059>.
- [23] S.G. Xu, X.T. Li, Y.J. Zhao, J.H. Liao, H. Xu, X.B. Yang, An electron compensation mechanism for the polymorphism of boron monolayers, *Nanoscale* 10 (2018) 13410–13416, <https://doi.org/10.1039/C8NR01230J>, url:<https://doi.org/10.1039/C8NR01230J>.
- [24] G.S. Shi, L. Chen, Y.Z. Yang, D.Y. Li, Z. Qian, S.S. Liang, L. Yan, L.H. Li, M.H. Wu, H.P. Fang, Two-dimensional na-cl crystals of unconventional stoichiometries on graphene surface from dilute solution at ambient conditions, *Nat. Chem.* 10 (7) (2018) 776–779, <https://doi.org/10.1038/s41557-018-0061-4>, url:<https://doi.org/10.1038/s41557-018-0061-4>.
- [25] J. Greeley, M. Mavrikakis, Surface and subsurface hydrogen: adsorption properties on transition metals and near-surface alloys, *J. Phys. Chem. B* 109 (8) (2005) 3460–3471, <https://doi.org/10.1021/jp046540q>, url:<https://doi.org/10.1021/jp046540q>.
- [26] H.W. Kroto, J.R. Heath, S.C. O'Brien, R.F. Curl, R.E. Smalley, C_{60} : Buckminsterfullerene, *Nature* 318 (6042) (1985) 162–163, <https://doi.org/10.1038/318162a0>, url:<https://doi.org/10.1038/318162a0>.
- [27] C.-S. Chen, T.-H. Chuang, Y.-H. Liu, W.-Y. Yeh, The solitary isomer of $\text{C}_{60}\text{H}_{18}$ is proven to have a C_{3v} crown shape: Crystal structure determination and synthesis of its triruthenium cluster complex, *Chem. – Eur. J.* 21 (48) (2015) 17229–17233, <https://doi.org/10.1002/chem.201503410>, url:<https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/chem.201503410>.
- [28] X.-P. Zhao, H.-Y. Huang, C. Wen, Y.-J. Su, P. Qian, Accelerating the development of multi-component cu-al-based shape memory alloys with high elastocaloric property by machine learning, *Comput. Mater. Sci.* 176 (2020), <https://doi.org/10.1016/j.commatsci.2020.109521>, 109521 url:<http://www.sciencedirect.com/science/article/pii/S0927025620300124>.
- [29] R. Otis, M. Emelianenko, Z.-K. Liu, An improved sampling strategy for global energy minimization of multi-component systems, *Comput. Mater. Sci.* 130 (2017) 282–291, <https://doi.org/10.1016/j.commatsci.2017.01.019>, url:<http://www.sciencedirect.com/science/article/pii/S0927025617300332>.
- [30] A. Ruban, On segregation in multicomponent alloys: Surface segregation in austenite and ferronimn alloys, *Comput. Mater. Sci.* 187 (2021), <https://doi.org/10.1016/j.commatsci.2020.110080>, 110080 url:<http://www.sciencedirect.com/science/article/pii/S0927025620305711>.
- [31] D. Chatain, P. Wynblatt, Surface segregation in multicomponent high entropy alloys: Atomistic simulations versus a multilayer analytical model, *Comput. Mater. Sci.* 187 (2021), <https://doi.org/10.1016/j.commatsci.2020.110101>, 110101 url:<http://www.sciencedirect.com/science/article/pii/S0927025620305929>.
- [32] Y.-H. Cheng, J.H. Liao, Y.J. Zhao, X.B. Yang, An extended cluster expansion for ground states of heterofullerenes, *Sci. Rep.* 7 (1) (2017) 16211, <https://doi.org/10.1038/s41598-017-16469-0>, url:<https://doi.org/10.1038/s41598-017-16469-0>.
- [33] P.M. Tagade, S.P. Adiga, S. Pandian, M.S. Park, K.S. Hariharan, S.M. Kolake, Attribute driven inverse materials design using deep learning bayesian framework, *npj Comput. Mater.* 5 (1) (2019) 127, <https://doi.org/10.1038/s41524-019-0263-3>, url:<https://doi.org/10.1038/s41524-019-0263-3>.
- [34] Structures of Alloy Generation And Recognition, url:<https://github.com/scut-cmp/sagar> (2020 (accessed 2020-05-14)).
- [35] A.H. Larsen, J.J. Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dulak, J. Friis, M.N. Groves, B. Hammer, C. Hargus, E.D. Hermes, P.C. Jennings, P. B. Jensen, J. Kermode, J.R. Kitchin, E.L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J.B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K.S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K.W. Jacobsen, The atomic simulation environment—a python library for working with atoms, *J. Phys.: Condens. Matter* 29 (27) (2017), <https://doi.org/10.1088/1361-648x/aa680e>, 273002 url:<https://doi.org/10.1088/1361-648x/aa680e>.
- [36] G.L.W. Hart, R.W. Forcade, Algorithm for generating derivative structures, *Phys. Rev. B* 77 (2008), <https://doi.org/10.1103/PhysRevB.77.224115>, 224115 url:<https://link.aps.org/doi/10.1103/PhysRevB.77.224115>.
- [37] X.-T. Li, S.-G. Xu, X.-B. Yang, Y.-J. Zhao, An intrinsic representation of atomic structure: From clusters to periodic systems, *J. Chem. Phys.* 147 (14) (2017), <https://doi.org/10.1063/1.4997292>, 144106 url:<https://doi.org/10.1063/1.4997292>.
- [38] Y.-L. Jiang, Y.-Z. Chen, H. Wang, X.-B. Yang, Theoretical investigation of v2xf2 (1-x)zr and scxy1-xfe2 (0<x<1) quasi-binary alloy: Stable structures, mechanical and electronic properties, *Phys. Lett. A* 384 (26) (2020), <https://doi.org/10.1016/j.physleta.2020.126658>, 126658 url:<http://www.sciencedirect.com/science/article/pii/S0375960120305259>.
- [39] G. Kresse, J. Hafner, Ab initio molecular dynamics for open-shell transition metals, *Phys. Rev. B* 48 (1993) 13115–13118, <https://doi.org/10.1103/PhysRevB.48.13115>, url:<https://link.aps.org/doi/10.1103/PhysRevB.48.13115>.
- [40] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B* 59 (1999) 1758–1775, <https://doi.org/10.1103/PhysRevB.59.1758>, url:<https://link.aps.org/doi/10.1103/PhysRevB.59.1758>.
- [41] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* 77 (1996) 3865–3868, <https://doi.org/10.1103/PhysRevLett.77.3865>, url:<https://link.aps.org/doi/10.1103/PhysRevLett.77.3865>.
- [42] N.O. Lipari, A.B. Kunz, Energy bands and optical properties of nacl, *Phys. Rev. B* 3 (1971) 491–497, <https://doi.org/10.1103/PhysRevB.3.491>, url:<https://link.aps.org/doi/10.1103/PhysRevB.3.491>.
- [43] H.-Y.T. Chen, G. Pacchioni, Properties of two-dimensional insulators: a dft study of co adsorption on nacl and mgo ultrathin films, *Phys. Chem. Chem. Phys.* 16 (2014) 21838–21845, <https://doi.org/10.1039/C4CP03470H>, url:<https://doi.org/10.1039/C4CP03470H>.
- [44] K.A. Tikhomirova, C. Tantardini, E.V. Sukhanova, Z.I. Popov, S.A. Evlashin, M. A. Tarkhov, V.L. Zhdanov, A.A. Dudin, A.R. Oganov, D.G. Kvashnin, Exotic two-dimensional structure: The first case of hexagonal nacl, *J. Phys. Chem. Lett.* 11 (10) (2020) 3821–3827, <https://doi.org/10.1021/acs.jpclett.0c00874>, url:<https://doi.org/10.1021/acs.jpclett.0c00874>.
- [45] V. Zlasek, T. Hildebrandt, M. Henzler, Surface color centers on epitaxial nacl films, *Phys. Rev. B* 62 (2000) 2912–2919, <https://doi.org/10.1103/PhysRevB.62.2912>, url:<https://link.aps.org/doi/10.1103/PhysRevB.62.2912>.
- [46] J. Repp, G. Meyer, S. Paavilainen, F.E. Olsson, M. Persson, Scanning tunneling spectroscopy of cl vacancies in nacl films: Strong electron-phonon coupling in double-barrier tunneling junctions, *Phys. Rev. Lett.* 95 (2005), <https://doi.org/10.1103/PhysRevLett.95.225503>, 225503 url:<https://link.aps.org/doi/10.1103/PhysRevLett.95.225503>.
- [47] C.-Z. Ning, L. Dou, P. Yang, Bandgap engineering in semiconductor alloy nanomaterials with widely tunable compositions, *Nat. Rev. Mater.* 2 (12) (2017) 17070, <https://doi.org/10.1038/natrevmats.2017.70>, url:<https://doi.org/10.1038/natrevmats.2017.70>.
- [48] L.S. Panchakarla, K.S. Subrahmanyam, S.K. Saha, A. Govindaraj, H. R. Krishnamurthy, U.V. Waghmare, C.N.R. Rao, Synthesis, structure, and properties of boron- and nitrogen-doped graphene, *Adv. Mater.* 21 (46) (2009)

- 4726–4730, <https://doi.org/10.1002/adma.200901285>, url:<https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.200901285>.
- [49] C.C. He, S.-B. Qiu, J.S. Yu, J.H. Liao, Y.J. Zhao, X.B. Yang, Atom classification model for total energy evaluation of two-dimensional multicomponent materials, *J. Phys. Chem. A* 124 (22) (2020) 4506–4511, <https://doi.org/10.1021/acs.jpca.0c02431>, url:<https://doi.org/10.1021/acs.jpca.0c02431>.
- [50] H. Zou, The adaptive lasso and its oracle properties, *J. Am. Statist. Assoc.* 101 (476) (2006) 1418–1429, <https://doi.org/10.1198/016214506000000735>, url:<https://doi.org/10.1198/016214506000000735>.
- [51] A.H. Castro Neto, F. Guinea, N.M.R. Peres, K.S. Novoselov, A.K. Geim, The electronic properties of graphene, *Rev. Mod. Phys.* 81 (2009) 109–162, <https://doi.org/10.1103/RevModPhys.81.109>, url:<https://link.aps.org/doi/10.1103/RevModPhys.81.109>.
- [52] B. Huang, G. Clark, E. Navarro-Moratalla, D.R. Klein, R. Cheng, K.L. Seyler, D. Zhong, E. Schmidgall, D.H. Cobden, X. Xu, Layer-dependent ferromagnetism in a van der waals crystal down to the monolayer limit, *Nature* 546 (7657) (2017) 270–273, <https://doi.org/10.1038/nature22391>, url:<https://doi.org/10.1038/nature22391>.
- [53] F. Wu, C. Huang, H. Wu, C. Lee, K. Deng, E. Kan, P. Jena, Atomically thin transition-metal dinitrides: High-temperature ferromagnetism and half-metallicity, *Nano Letters* 15 (12) (2015) 8277–8281, PMID: 26575002. doi:10.1021/acs.nanolett.5b03835. url:<https://doi.org/10.1021/acs.nanolett.5b03835>.
- [54] Y. Sun, Z. Zhuo, X. Wu, J. Yang, Room-temperature ferromagnetism in two-dimensional fe₂si nanosheet with enhanced spin-polarization ratio, *Nano Lett.* 17 (5) (2017) 2771–2777, <https://doi.org/10.1021/acs.nanolett.6b04884>, url:<https://doi.org/10.1021/acs.nanolett.6b04884>.
- [55] X.Y. Li, X.X. Li, J.L. Yang, Room-temperature ferromagnetism in transition metal embedded borophene nanosheets, *J. Phys. Chem. Lett.* 10 (15) (2019) 4417–4421, <https://doi.org/10.1021/acs.jpclett.9b01667>, url:<https://doi.org/10.1021/acs.jpclett.9b01667>.
- [56] J.B. Goodenough, Theory of the role of covalence in the perovskite-type manganites [La.*m*(II)]MnO₃, *Phys. Rev.* 100 (1955) 564–573, <https://doi.org/10.1103/PhysRev.100.564>, url:<https://link.aps.org/doi/10.1103/PhysRev.100.564>.
- [57] J. Kanamori, Superexchange interaction and symmetry properties of electron orbitals, *J. Phys. Chem. Solids* 10 (2) (1959) 87–98, [https://doi.org/10.1016/0022-3697\(59\)90061-7](https://doi.org/10.1016/0022-3697(59)90061-7), url:<http://www.sciencedirect.com/science/article/pii/0022369759900617>.
- [58] A.R. Murphy, J.M.J. Fréchet, Organic semiconducting oligomers for use in thin film transistors, *Chem. Rev.* 107 (4) (2007) 1066–1096, <https://doi.org/10.1021/cr0501386>, url:<https://doi.org/10.1021/cr0501386>.
- [59] K. Tokunaga, S. Ohmori, H. Kawabata, K. Matsushige, A density functional theory study on the hole transfer in fullerene hydride C₆₀H₂, *Jpn. J. Appl. Phys.* 47 (2) (2008) 1089–1093, <https://doi.org/10.1143/jjap.47.1089>, url:<https://doi.org/10.1143%2Fjjap.47.1089>.