Article

High-Throughput Prediction of the Band Gaps of van der Waals Heterostructures via Machine Learning

Rui Hu, Wen Lei, Hongmei Yuan, Shihao Han and Huijun Liu *

Key Laboratory of Artificial Micro- and Nano-Structures of Ministry of Education and School of Physics and Technology, Wuhan University, Wuhan 430072, China; ruihu16@whu.edu.cn (R.H.); leiwen64@whu.edu.cn (W.L.); hmyuan@whu.edu.cn (H.Y.); hansh123@whu.edu.cn (S.H.)
* Correspondence: phlhj@whu.edu.cn

Abstract: Van der Waals heterostructures offer an additional degree of freedom to tailor the electronic structure of two-dimensional materials, especially for the band-gap tuning that leads to various applications such as thermoelectric and optoelectronic conversions. In general, the electronic gap of a given system can be accurately predicted by using first-principles calculations, which is, however, restricted to a small unit cell. Here, we adopt a machine-learning algorithm to propose a physically intuitive descriptor by which the band gap of any heterostructures can be readily obtained, using group III, IV, and V elements as examples of the constituent atoms. The strong predictive power of our approach is demonstrated by high Pearson correlation coefficient for both the training (292 entries) and testing data (33 entries). By utilizing such a descriptor, which contains only four fundamental properties of the constituent atoms, we have rapidly predicted the gaps of 7140 possible heterostructures that agree well with first-principles results for randomly selected candidates.

Keywords: machine learning; van der Waals heterostructures; band gap; high-throughput screening

1. Introduction

With the rapid and successful development in the study of two-dimensional (2D) materials [1–5], there has been a growing interest in the van der Waals heterostructures (vdWHs) due to their unique structures and rich physical properties [6–8]. Analogous to building with Legos, a vdWH can be viewed as stacking different 2D systems on top of each other. In fact, the weak interlayer interaction permits the superposition of virtually any given pair of 2D materials, such as graphene/phosphorene [9], silicene/graphene [10], phosphorene/SnX_2 (X = S, Se) [11], graphene/hexagonal boron nitride [12], WSe_2/Bi_2Te_3 [13], and so on. On the experimental side, the vdWHs can be prepared by using either bottom-up or top-down approaches [14–16]. The former makes use of successive deposition techniques, while the latter requires the fabrication of individual layers first and subsequent peeling off and assembly [7].

The emergence of vdWHs offers additional degree of freedom to tailor the electronic structure of 2D materials [17,18], especially for the band-gap tuning that leads to various applications such as thermoelectric [19,20] and optoelectronic conversions [21,22]. In general, the electronic band gap of a given vdWH can be predicted by using first-principles calculations [23–25]. Although the conventional local density approximation (LDA) [26] or generalized gradient approximation (GGA) [27] are relatively computationally efficient, they suffer from obvious band-gap underestimation. Accurate calculations require state-of-the-art techniques, such as the Heyd–Scuseria–Ernzerhof (HSE) hybrid functional [28] or the *GW* approximation of many-body effects [29]. However, both of them are computationally expensive and, thus, are restricted to small systems. As an alternative, the machine learning (ML) method has recently attracted considerable attention for band-gap prediction, which can efficiently manage a huge search space at an extremely low cost [30–37]. For example, by choosing 28 primary atomic properties as input features, Pilania et al. [30] obtained

a kernel ridge regression (KRR) model to predict the band gaps of 1378 unique double perovskites, where the Pearson correlation coefficient can be as high as 97%. On top of the 136 compositional properties of inorganic solids, Zhuo et al. [31] proposed a support vector regression (SVR) model for gap prediction and found a small root mean square error (RMSE) of 0.45 eV. In addition, by leveraging 42 initial elemental features of chalcogenides, Wang et al. [33] developed a stacked ensemble learning (SEL)-gap model with a coefficient of determination (R^2) value of 90%. It should be noted that the above-mentioned ML models usually contain a large number of input features, which is actually not beneficial for the high-throughput discovery of desired systems. Moreover, most of these models appear as black boxes, so a direct understanding of the underlying physics is quite necessary.

In this work, using group III, IV, and V elements as prototypical examples of constituent atoms, we adopt the SISSO (Sure Independence Screening and Sparsifying Operator) method [38,39] to propose a physically intuitive three-dimensional (3D) descriptor, by which the band gap of any vdWHs can be readily obtained. The strong predictive power of our descriptor is demonstrated by the good agreement between the SISSO-predicted gaps ($E_{g,pre}$) and those calculated using accurate HSE functional ($E_{g,cal}$), either inside or beyond the training data. As the input features only contain four fundamental properties of the constituent atoms, the 3D descriptor is very beneficial for the accelerated discovery of vdWHs with the desired band gaps.

2. Methodology

To obtain reliable training data for ML, we have calculated the band gaps of 325 vdWHs by using first-principles pseudopotential method, as implemented in the Vienna ab-initio simulation package (VASP) [40]. The hybrid functional within the HSE scheme [28] is adopted to overcome the gap underestimation in standard density functional theory (DFT), and the vdW interaction is considered by using the DFT-D3 exchange functional [41]. The plane-wave cutoff energy is set as 450 eV, and a $19 \times 19 \times 1$ Monkhorst-Pack $\boldsymbol{k}$-mesh is sampled in the Brillouin zone. The energy convergence threshold is 10^{-6} eV, and the relaxed structure is determined until the residual force on each atom is less than 0.01 eV Å^{-1}.

Based on the 325 samples in the training data, we adopt the SISSO approach [38,39] to obtain an optimized descriptor for predicting the band gap. Here, the input features include only four fundamental properties of constituent atoms, i.e., the atomic number Z, the Pauling electronegativity χ, the number of valence electrons VE, and the atomic radius r. A combination of algebraic operations is then recursively performed to extend the feature space, as defined by $H^{(m)} \equiv \left\{ I, +, -, \times, /, exp, log, |-|, \sqrt{}, ^{-1}, ^2, ^3 \right\}$. Here, m means dimensional analysis, so that only meaningful combinations are allowed. By equipping the feature space with nonlinear operators in $H^{(m)}$, the intrinsically linear relation between observables and descriptor in the compressed sensing formalism is made nonlinear. At each iteration, $H^{(m)}$ operates on all possible combinations, and over 10^{10} features are constructed up to a complexity cutoff of 3. Such huge size can be effectively reduced by combining the sure independence screening (SIS) with the sparsifying operators (SO). Here, the SIS scores each feature with a metric and keeps only the top ranked. The subset extracted by the SIS is set to 80,000. After dimensionality reduction, the SO is used to pinpoint the optimal descriptor, which turns out to be physically interpretable. It should be mentioned that the SISSO algorithm demonstrates obvious advantages compared with other established ML approaches that suffer from huge and highly correlated feature spaces.

3. Results and Discussion

In the present work, we focus on the vdWH, which is composed of two graphene-like monolayers via vdW interactions along the out-of-plane direction. As illustrated in Figure 1, the system can be labeled by a nominal formula of AB/CD, where the A and B represent the two nonequivalent atoms in the upper layer, and the C and D represent those in the lower layer. Among these four atoms, the Pauling electronegativity of the A (C) atom is

smaller than that of the B (D) atom by default. Considering the fact that vdWH permits the superposition of virtually any given pair of 2D materials [6], here we randomly select 26 graphene-like structures [42–45] as the constituent monolayers, including group V (N, P, As, Sb), group IV-IV (SiC, SiGe, SnSi, GeC, SnGe, SnC), and group III-V (BN, BP, BAs, BSb, AlN, AlP, AlAs, AlSb, GaN, GaP, GaAs, GaSb, InN, InP, InAs, InSb). For simplicity, only the conventional AA stacking pattern is considered, and the twist angle is $0°$, which can in principle create $C_{26}^2 = 325$ different vdWHs (including 22 binary, 132 ternary, and 171 quaternary systems). Note that the vdW thickness is determined by minimizing the total energies of these heterostructures, where the vdW functional is explicitly considered in the DFT calculations. Besides, the stability and possible existence of 325 vdW heterostructures has been demonstrated in previously published work [43]. As there are only four atoms in the primitive cell, it is computationally ready to obtain the electronic band gaps of all these vdWHs from first-principles calculations, even with the high-level HSE scheme. The results are summarized in Table S1 of the Supplementary Materials, which will be used as the original dataset for SISSO training.
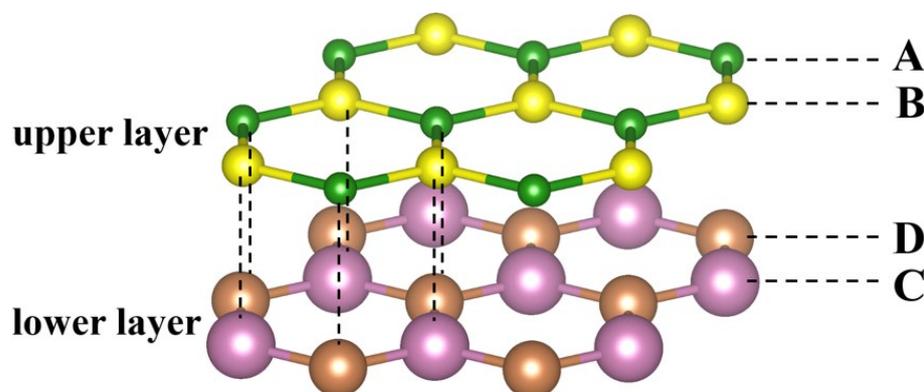
Figure 1. The crystal structure of AB/CD vdWH, where A, B, C, and D represent group III (B, Al, Ga, In, and Ta), group IV (C, Si, Ge, Sn, and Pb), and group V elements (N, P, As, Sb, and Bi).

For any ML approaches, the selection of input features plays a crucial role in deriving the optimal model. In principle, one can adopt those related to the investigated system or the constituent atoms. Here, we consider the latter, since it can be readily obtained and is very beneficial for the high-throughput screening of the desired candidates. Among the 58 fundamental properties of the constituent atoms [46], we find that choosing the atomic number Z, the Pauling electronegativity χ, the number of valence electrons VE, and the atomic radius r could pinpoint an optimal descriptor that enables the accurate prediction of the band gap of the vdWHs. Table 1 lists these four kinds of input features for several constituent atoms from groups III, IV, and V. It should be mentioned that the 325 entries in the original dataset are randomly divided into 292 for effective training and 33 for real-time testing. Consequently, the SISSO-identified descriptor for gap prediction is given in a 3D form:

$$E_{g,\ pre} = 0.85 \times D_1 - 25.32 \times D_2 - 0.0003 \times D_3 - 1.45 \tag{1}$$

where D_1, D_2, and D_3 are defined as:

$$D_1 = \frac{(VE_A + VE_C)(\chi_B \times \chi_D)}{(r_A + r_C)(\chi_A \times \chi_C)} \tag{2}$$

$$D_2 = \left| \frac{1}{VE_B(r_A + r_D)} - \frac{1}{VE_D(r_B + r_C)} \right| \tag{3}$$

$$D_3 = \frac{|Z_A - Z_D| + |Z_B - Z_C|}{(\chi_B{}^3 + \chi_D{}^3)} \tag{4}$$

Table 1. The input features used for SISSO training, which includes the atomic number (Z), the Pauling electronegativity (χ, in units of eV), the number of valence electrons (VE), and the atomic radius (r, in units of Å) for group III (B, Al, Ga, In, and Ta), group IV (C, Si, Ge, Sn, and Pb), and group V elements (N, P, As, Sb, and Bi).

Elements	Z	χ	VE	r
B	5	2.04	3	0.95
Al	13	1.61	3	1.43
Ga	31	1.81	3	1.4
In	49	1.78	3	1.66
Tl	81	1.62	3	1.73
C	6	2.55	4	0.86
Si	14	1.98	4	1.34
Ge	32	2.01	4	1.4
Sn	50	1.96	4	1.58
Pb	82	2.33	4	1.75
N	7	3.04	5	0.8
P	15	2.19	5	1.3
As	33	2.18	5	1.5
Sb	51	2.05	5	1.6
Bi	83	2.02	5	1.7

Figure 2 shows the intuitive linear correlation between the SISSO-predicted gaps ($E_{g,\,pre}$) and those calculated by using HSE scheme ($E_{g,\,cal}$). It is obvious that both the training (Figure 2a) and testing data (Figure 2b) are evenly distributed around the dashed line with slope one, suggesting the higher prediction accuracy of our SISSO descriptor. Indeed, the Pearson correlation coefficient is found to be 94% and 92% for the training and testing sets, respectively.
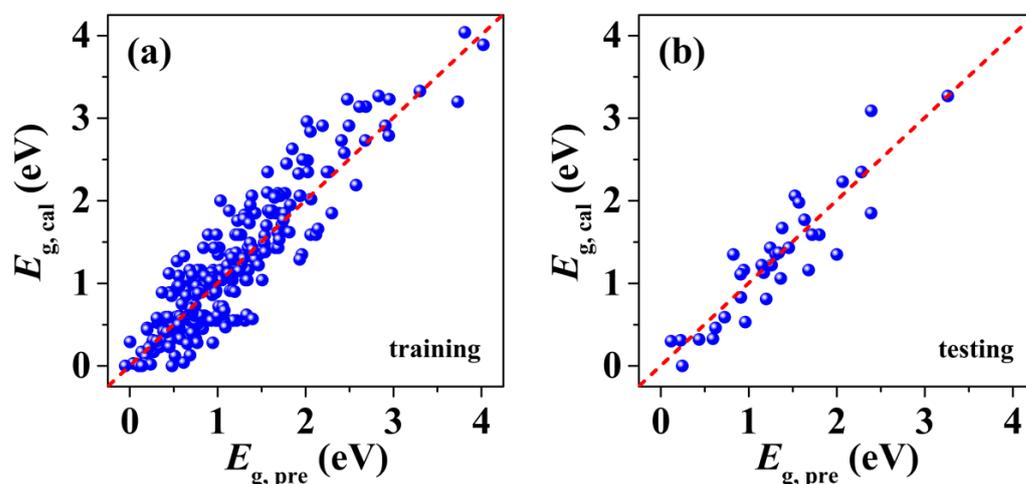

Figure 2. The intuitive linear correlation between the SISSO-predicted band gaps and those calculated by first-principles for vdWHs in the (**a**) training data and (**b**) testing data. The red dashed line represents equality.

In addition to the strong predictive power discussed above, we should emphasize that the SISSO-derived descriptor is physically interpretable. Here, we focus on the first term ($0.85 \times D_1$) of Equation (1), since it plays a major role [39] in the gap prediction. For the convenience of discussion, we rewrite Equation (2) as:

$$D_1 = \frac{(VE_A + VE_C)}{(r_A + r_C)} \times \frac{(\chi_B \times \chi_D)}{(\chi_A \times \chi_C)} \tag{5}$$

That is, the SISSO-predicted gap is proportional to $\frac{(VE_A+VE_C)}{(r_A+r_C)}$ and $\frac{(\chi_B \times \chi_D)}{(\chi_A \times \chi_C)}$. For the $\frac{(VE_A+VE_C)}{(r_A+r_C)}$ term, if the A and C atoms are located in the same row of the periodic table, the number of valence electrons ($VE_A + VE_C$) increases, while the atomic radius ($r_A + r_C$) decreases with the increasing atomic number. As a consequence, the $\frac{(VE_A+VE_C)}{(r_A+r_C)}$ term increases, which leads to a larger band gap. This is reasonable since a decreased atomic radius ($r_A + r_C$) usually corresponds to a smaller bond length (or stronger bond strength), which would enhance band splitting, so a larger band gap can be expected. On the other hand, if the A and C atoms are in the same column of the periodic table, the number of valence electrons ($VE_A + VE_C$) keeps constant, while the atomic radius ($r_A + r_C$) increases with increasing atomic number. That is, we find a decreased $\frac{(VE_A+VE_C)}{(r_A+r_C)}$ and, thus, a smaller band gap, which can be understood by the same reason discussed above. For the $\frac{(\chi_B \times \chi_D)}{(\chi_A \times \chi_C)}$ term, since the Pauling electronegativities of the B and D atoms (anion-like) are relatively larger compared with those of the A and C atoms (cation-like), the significant difference between them would give rise to a larger band gap. Such a finding is consistent with the general belief that the band gap of an inorganic compound is approximately proportional to the electronegativity difference between the anions and cations [47,48].

Beyond the original dataset of 325 vdWHs, we have employed the SISSO-derived descriptor to predict the band gaps of a vast number of possible AB/CD heterostructures. Using group III, IV, and V elements as examples, it is assumed that A, B, C, and D can be chosen from 15 atoms: B, Al, Ga, In, Tl, C, Si, Ge, Sn, Pb, N, P, As, Sb, and Bi. As illustrated in Figure 3a, we can obtain a total of 7140 possible vdWHs, which includes 315 binary (calculated by $3 \times C_{15}^2$), 2730 ternary ($6 \times C_{15}^3 = 2730$), and 4095 quaternary systems ($3 \times C_{15}^4 = 4095$). By leveraging the 3D descriptor given in Equation (1), we can quickly predict the band gaps ($E_{g, pre}$) of all these 7140 vdWHs (including 325 entries in the original dataset). Figure 3b illustrates the distribution of $E_{g, pre}$, where we see that most investigated heterostructures have smaller band gaps in the range of 0–1 eV. Besides, there are 786 systems exhibiting intermediate gaps from 1 to 3 eV, and 23 systems with gaps exceeding 3 eV. To further verify the predictive power of our 3D descriptor, we have calculated the band gaps of six vdWHs (BP/BiN, BSb/AsP, GeP/PP, SiAs/NN, AlC/AlAs, GeN/NN) by using an accurate first-principles approach (HSE scheme). Note that these vdWHs are randomly selected from Figure 3 and do not belong to the original dataset. As can be seen from Table 2, the band gaps obtained from the first-principles calculation ($E_{g, cal}$) almost coincide with those predicted from the 3D descriptor ($E_{g, pre}$). Once again, such an observation confirms the strong reliability of our SISSO approach, although only four fundamental properties of the constituent atoms are involved in the derived descriptor.
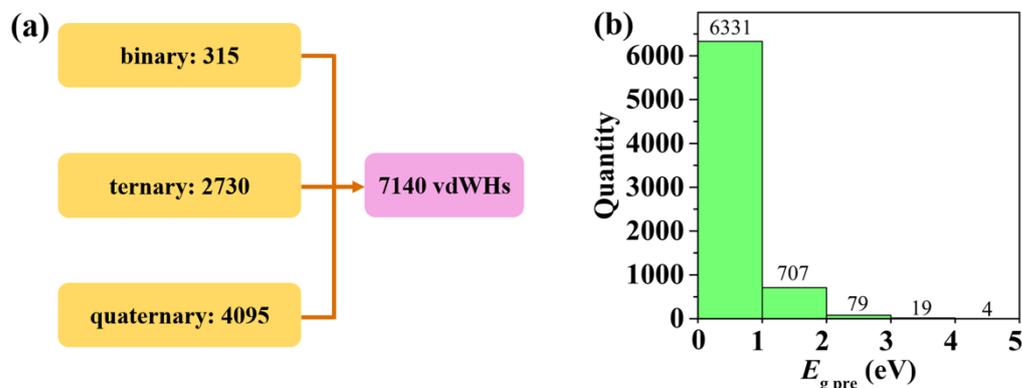

Figure 3. (**a**) Schematic illustration of 7140 AB/CD vdWHs, which includes 315 binary, 2730 ternary, and 4095 quaternary systems. (**b**) Distribution of 7140 vdWHs according to their SISSO-predicted band gaps.

Table 2. Comparisons of the SISSO-predicted band gaps and those calculated by first-principles (HSE scheme), for six randomly selected vdWHs beyond the original dataset.

vdWHs	$E_{g,\,pre}$ (eV)	$E_{g,\,cal}$ (eV)
BP/BiN	0.35	0.38
BSb/AsP	0.89	0.84
GeP/PP	1.45	1.43
SiAs/NN	1.94	1.88
AlC/AlAs	2.10	2.09
GeN/NN	2.53	2.45

Before concluding our work, we should mention that the high-throughput results shown in Figure 3 also provide a very useful diagram to screen 2D functional materials with the desired band gap for potential applications, such as thermoelectric and optoelectronic conversions. Here, we focus on the thermoelectric materials that usually require a smaller gap (0–1 eV) to realize a relatively larger power factor. Among the 6331 possible candidates in Figure 3b, we choose the BSb/AsP vdWH as a checking example, where the SISSO-predicted gap is found to be 0.89 eV. Figure 4a plots the band structure of the BSb/AsP vdWH, as obtained from the first-principles calculations. We see that the conduction-band minimum (CBM) and the valence-band maximum (VBM) are both located at the **K** point, which gives a direct gap of 0.84 eV (HSE value) and is very close to the SISSO prediction. Based on the band structure, we can readily evaluate the Seebeck coefficient (*S*) by using the Boltzmann transport theory [49]. Note that *S* describes the ratio of voltage difference to the temperature gradient imposed on a given system and plays an important role in determining the thermoelectric-conversion efficiency that is usually evaluated by the dimensionless figure-of-merit $ZT \propto S^2$. Figure 4b shows the Seebeck coefficient of the BSb/AsP vdWH as a function of carrier concentration, where the result is the same along the armchair and zigzag directions. Remarkably, we observe that the *S* value at a typical-carrier concentration of 10^{19} cm^{-3} can reach 486 μV/K and 501 μV/K for the *n*- and *p*-type systems, respectively. Such values are significantly larger than those of state-of-the-art thermoelectric materials, suggesting a very promising thermoelectric performance by our vdWHs. If the carrier concentration is increased to 5×10^{19} and 1×10^{20} cm^{-3}, we can, respectively, obtain an *S* value of ~300 μV/K and ~200 μV/K for both *n*- and *p*-type vdWHs, which are still comparable to those of many good thermoelectric materials.
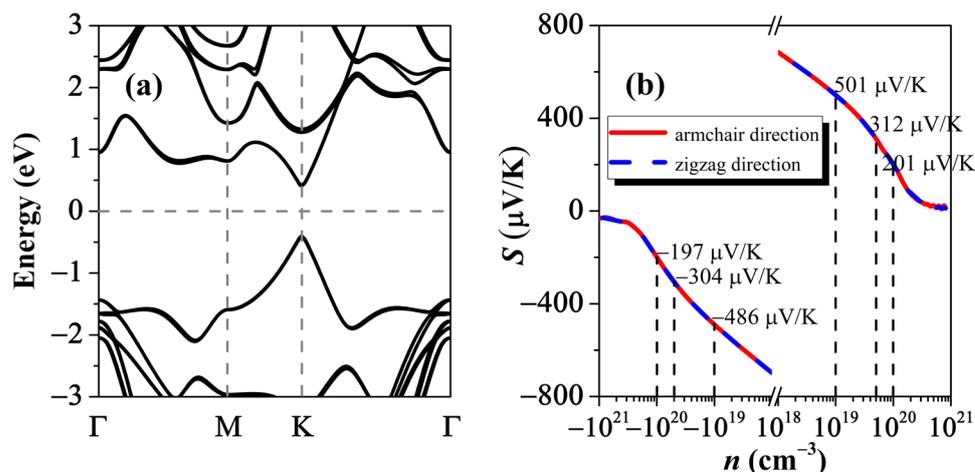
Figure 4. (**a**) The energy band structure, and (**b**) the Seebeck coefficient of the screened BSb/AsP vdWH.

4. Summary

In conclusion, our work offers a high-throughput solution to the fundamentally important issue of band-gap prediction, using the vdWHs as a prototypical class of examples.

Unlike many ML algorithms, where the derived model usually appears as a black box, the SISSO approach employed in the present work can pinpoint a physically intuitive 3D descriptor that requires only four fundamental properties of the constituent atoms as input features. By leveraging such a simple-yet-efficient data-driven descriptor, one can deal with a huge search space (even for those systems with constituent atoms beyond Table 1) to screen promising candidates with the desired electronic band gap for potential applications, such as thermoelectric and optoelectronic conversions. It should be also noted that the SISSO approach established here can be adopted to predict the optical gap and entire band structure of vdWHs, as long as the corresponding training set is available.

References

1. Novoselov, K.S.; Geim, A.K.; Morozov, S.V.; Jiang, D.; Zhang, Y.; Dubonos, S.V.; Grigorieva, I.V.; Firsov, A.A. Electric field effect in atomically thin carbon films. *Science* **2004**, *306*, 666. [CrossRef] [PubMed]
2. Zhang, H. Ultrathin two-dimensional nanomaterials. *ACS Nano* **2015**, *9*, 9451. [CrossRef] [PubMed]
3. Gupta, A.; Sakthivel, T.; Seal, S. Recent development in 2D materials beyond graphene. *Prog. Mater. Sci.* **2015**, *73*, 44. [CrossRef]
4. Tan, C.; Cao, X.; Wu, X.-J.; He, Q.; Yang, J.; Zhang, X.; Chen, J.; Zhao, W.; Han, S.; Nam, G.-H.; et al. Recent advances in ultrathin two-dimensional nanomaterials. *Chem. Rev.* **2017**, *117*, 6225. [CrossRef]
5. Akinwande, D.; Huyghebaert, C.; Wang, C.-H.; Serna, M.I.; Goossens, S.; Li, L.-J.; Wong, H.-S.P.; Koppens, F.H.L. Graphene and two-dimensional materials for silicon technology. *Nature* **2019**, *573*, 507. [CrossRef]
6. Geim, A.K.; Grigorieva, I.V. Van der Waals heterostructures. *Nature* **2013**, *499*, 419–425. [CrossRef]
7. Novoselov, K.S.; Mishchenko, A.; Carvalho, A.; Neto, A.H.C. 2D materials and van der Waals heterostructures. *Science* **2016**, *353*, aac9439. [CrossRef]
8. Liu, Y.; Weiss, N.O.; Duan, X.D.; Cheng, H.C.; Huang, Y.; Duan, X.F. Van der Waals heterostructures and devices. *Nat. Rev. Mater.* **2016**, *1*, 16042. [CrossRef]
9. Padilha, J.E.; Fazzio, A.; da Silva, A.J.R. Van der Waals heterostructure of phosphorene and graphene: Tuning the Schottky barrier and doping by electrostatic gating. *Phys. Rev. Lett.* **2015**, *114*, 066803. [CrossRef]
10. Shi, L.; Zhao, T.S.; Xu, A.; Xu, J.B. Ab initio prediction of a silicene and graphene heterostructure as an anode material for Li-and Na-ion batteries. *J. Mater. Chem. A* **2016**, *4*, 16377–16382. [CrossRef]
11. Xia, C.; Du, J.; Li, M.; Li, X.; Zhao, X.; Wang, T.; Li, J. Effects of Electric Field on the Electronic Structures of Broken-gap Phosphorene/Sn X_2 (X = S, Se) van der Waals Heterojunctions. *Phys. Rev. Appl.* **2018**, *10*, 54064. [CrossRef]
12. Li, T.; Tang, Z.; Huang, Z.; Yu, J. Substrate effects on the thermal performance of in-plane graphene/hexagonal boron nitride heterostructures. *Carbon* **2018**, *130*, 396. [CrossRef]
13. Liu, H.; Zhu, X.; Sun, X.; Zhu, C.; Huang, W.; Zhang, X.; Zheng, B.; Zou, Z.; Luo, Z.; Wang, X.; et al. Self-powered broad-band photodetectors based on vertically stacked WSe_2/Bi_2Te_3 *p–n* heterojunction. *ACS Nano* **2019**, *13*, 13573. [CrossRef]

14. Ponomarenko, L.A.; Geim, A.K.; Zhukov, A.A.; Jalil, R.; Morozov, S.V.; Novoselov, K.S.; Grigorieva, I.V.; Hill, E.H.; Cheianov, V.V.; Fal'ko, V.I.; et al. Tunable metal–insulator transition in double-layer graphene heterostructures. *Nat. Phys.* **2011**, *7*, 958. [CrossRef]

15. Haigh, S.J.; Gholinia, A.; Jalil, R.; Romani, S.; Britnell, L.; Elias, D.C.; Novoselov, K.S.; Ponomarenko, L.A.; Geim, A.K.; Gorbachev, R. Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices. *Nat. Mater.* **2012**, *11*, 764. [CrossRef]

16. Liu, Y.; Huang, Y.; Duan, X. Van der Waals integration before and beyond two-dimensional materials. *Nature* **2019**, *567*, 323. [CrossRef]

17. Koda, D.S.; Bechstedt, F.; Marques, M.; Teles, L.K. Trends on band alignments: Validity of Anderson's rule in SnS_2-and $SnSe_2$-based van der Waals heterostructures. *Phys. Rev. B* **2018**, *97*, 165402. [CrossRef]

18. Deng, X.Q.; Jing, Q. Tunable electronic and optical properties of InSe/arsenene heterostructure by vertical strain and electric field. *Phys. Lett. A* **2021**, *405*, 127427. [CrossRef]

19. Jia, P.-Z.; Zeng, Y.-J.; Wu, D.; Pan, H.; Cao, X.-H.; Zhou, W.-X.; Xie, Z.-X.; Zhang, J.-X.; Chen, K.-Q. Excellent thermoelectric performance induced by interface effect in $MoS_2/MoSe_2$ van der Waals heterostructure. *J. Phys. Condens. Matter* **2020**, *32*, 055302. [CrossRef]

20. Hu, R.; Zhou, Z.Z.; Sheng, C.Y.; Wang, L.; Liu, J.H.; Han, S.H.; Liu, H.J. Surprisingly good thermoelectric performance of a black phosphorus/blue phosphorus van der Waals heterostructure. *Phys. Chem. Chem. Phys.* **2020**, *22*, 22390. [CrossRef]

21. Lin, S.; Lu, Y.; Xu, J.; Feng, S.; Li, J. High performance graphene/semiconductor van der Waals heterostructure optoelectronic devices. *Nano Energy* **2017**, *40*, 122. [CrossRef]

22. Lin, X.; Wang, F.; Shan, X.; Miao, Y.; Chen, X.; Yan, M.; Zhang, L.; Liu, K.; Luo, J.; Zhang, K. High-performance photodetector and its optoelectronic mechanism of MoS_2/WS_2 vertical heterostructure. *Appl. Surf. Sci.* **2021**, *546*, 149074. [CrossRef]

23. Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **1993**, *47*, 558. [CrossRef]

24. Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **1994**, *49*, 14251. [CrossRef]

25. Hu, R.; Li, Y.H.; Zhang, Z.H.; Fan, Z.Q.; Sun, L. O-Vacancy-line defective $Ti_2 CO_2$ nanoribbons: Novel magnetism, tunable carrier mobility, and magnetic device behaviors. *J. Mater. Chem. C* **2019**, *7*, 7745. [CrossRef]

26. Kohn, W.; Sham, L.J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133. [CrossRef]

27. Perdew, J.P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865. [CrossRef]

28. Heyd, J.; Scuseria, G.E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2006**, *124*, 219906. [CrossRef]

29. Shishkin, M.; Kresse, G. Self-consistent G W calculations for semiconductors and insulators. *Phys. Rev. B* **2007**, *75*, 235102. [CrossRef]

30. Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B.P.; Ramprasad, R.; Gubernatis, E.J.; Lookman, T. Machine learning bandgaps of double perovskites. *Sci. Rep.* **2016**, *6*, 19375. [CrossRef] [PubMed]

31. Zhuo, Y.; Tehrani, A.M.; Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *J. Phys. Chem. Lett.* **2018**, *9*, 1668. [CrossRef]

32. Rajan, A.C.; Mishra, A.; Satsangi, S.; Vaish, R.; Mizuseki, H.; Lee, K.-R.; Singh, A.K. Machine-learning-assisted accurate band gap predictions of functionalized MXene. *Chem. Mater.* **2018**, *30*, 4031. [CrossRef]

33. Wang, X.; Xu, Y.; Yang, J.; Ni, J.; Zhang, W.; Zhu, W. ThermoEPred-EL: Robust bandgap predictions of chalcogenides with diamond-like structure via feature cross-based stacked ensemble learning. *Comput. Mater. Sci.* **2019**, *169*, 109117. [CrossRef]

34. Lee, J.; Seko, A.; Shitara, K.; Nakayama, K.; Tanaka, I. Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques. *Phys. Rev. B* **2016**, *93*, 115104. [CrossRef]

35. Weston, L.; Stampfl, C. Machine learning the band gap properties of kesterite I2−II−IV−V4 quaternary compounds for photovoltaics applications. *Phys. Rev. Mater.* **2018**, *2*, 085407. [CrossRef]

36. Huang, Y.; Yu, C.; Chen, W.; Liu, Y.; Li, C.; Niu, C.; Wang, F.; Jia, Y. Band gap and band alignment prediction of nitride-based semiconductors using machine learning. *J. Mater. Chem. C* **2019**, *7*, 3238. [CrossRef]

37. Liang, J.; Zhu, X. Phillips-inspired machine learning for band gap and exciton binding energy prediction. *Phys. Chem. Lett.* **2019**, *10*, 5640. [CrossRef]

38. Ghiringhelli, L.M.; Vybiral, J.; Levchenko, S.V.; Draxl, C.; Scheffler, M. Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503. [CrossRef]

39. Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L.M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, *2*, 083802. [CrossRef]

40. 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. [CrossRef]

41. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104. [CrossRef]

42. Şahin, H.; Cahangirov, S.; Topsakal, M.; Bekaroglu, E.; Akturk, E.; Senger, R.T.; Ciraci, S. Monolayer honeycomb structures of group-IV elements and III-V binary compounds: First-principles calculations. *Phys. Rev. B* **2009**, *80*, 155453. [CrossRef]

43. Özçelik, V.O.; Azadani, J.G.; Yang, C.; Koester, S.J.; Low, T. Band alignment of two-dimensional semiconductors for designing heterostructures with momentum space matching. *Phys. Rev. B* **2016**, *94*, 035125. [CrossRef]

44. Peng, B.; Zhang, D.; Zhang, H.; Shao, H.; Ni, G.; Zhu, Y.; Zhu, H. The conflicting role of buckled structure in phonon transport of 2D group-IV and group-V materials. *Nanoscale* **2017**, *9*, 7397. [CrossRef]

45. Zhou, Z.Z.; Liu, H.J.; Fan, D.D.; Cao, G.H. A comparative study of the thermoelectric performance of graphene-like BX (X = P, As, Sb) monolayers. *J. Phys. Condens. Matter* **2019**, *31*, 385701. [CrossRef]

46. Features. Available online: https://xenonpy.readthedocs.io/en/latest/features.html (accessed on 30 May 2022).

47. Di Quarto, F.; Sunseri, C.; Piazza, S.; Romano, M.C. Semiempirical correlation between optical band gap values of oxides and the difference of electronegativity of the elements. Its importance for a quantitative use of photocurrent spectroscopy in corrosion studies. *J. Phys. Chem. B* **1997**, *101*, 2519. [CrossRef]

48. Kandpal, H.C.; Felser, C.; Seshadri, R. Covalent bonding and the nature of band gaps in some half-Heusler compounds. *J. Phys. D Appl. Phys.* **2006**, *39*, 776. [CrossRef]

49. Madsen, G.K.H.; Singh, D.J. BoltzTraP. A code for calculating band-structure dependent quantities. *Comput. Phys. Commun.* **2006**, *175*, 67. [CrossRef]