Journal of Materials Chemistry C

PAPER

Cite this: J. Mater. Chem. C, 2019, 7, 3238

Band gap and band alignment prediction of nitride-based semiconductors using machine learning†

Yang Huang,[ab] Changyou Yu,[c] Weiguang Chen,[d] Yuhuai Liu,[ce] Chong Li,[a] Chunyao Niu,[a] Fei Wang*[a] and Yu Jia*[af]

Received 5th November 2018,
Accepted 5th February 2019

DOI: 10.1039/c8tc05554h

rsc.li/materials-c

Nitride has been drawing much attention owing to its wide range of applications in optoelectronics and there remains plenty of room for materials design and discovery. Here, a large set of nitrides has been designed, with their band gap and alignment being studied by first-principles calculations combined with machine learning. The band gap and band offset against wurtzite GaN accurately calculated by the combination of the screened hybrid functionals of HSE and DFT-PBE were used to train and test machine learning models. After comparison among different machine learning techniques, when elemental properties are taken as features, support vector regression (SVR) with radial kernel performs best for predicting both the band gap and band offset with a prediction root mean square error (RMSE) of 0.298 eV and 0.183 eV, respectively. The former is within the HSE calculation uncertainty and the latter is small enough to provide reliable predictions. Additionally, when the band gap calculated by DFT-PBE was added into the feature space, the band gap prediction RMSE decreased to 0.099 eV. Through a feature engineering algorithm, the elemental feature space-based band gap prediction RMSE further drops by around 0.005 eV and the relative importance of elemental properties for band gap prediction was revealed. Finally, the band gap and band offset of all designed nitrides were predicted and two trends were noticed: as the number of cation types increases, the band gap tends to narrow while the band offset tends to increase. The predicted results will provide useful guidance for precise investigation of nitride engineering.

Introduction

Machine learning, a popular data mining technology that has been widely used in computer vision, speech recognition and natural language processing, has also recently been effectively used for materials research,[1] specifically, in property prediction[2] and prescreening in high-throughput searches for materials.[3,4] On the other hand, nitride semiconductor materials have emerged as one of the most important classes of materials in the modern semiconductor industry over the past 40 years. This family of materials, which traditionally consists of wurtzite III–N binary compounds, such as AlN, GaN, and InN, and later involved II–IV compounds like $Zn(Sn,Ge)N_2$ with various forms of alloys, has shown multiple significant applications in light-emitting diodes, lasers, photodetectors and photovoltaics owing to a broad range of band gap values ranging from deep UV to terahertz.[5] Despite their great accomplishments, nitride semiconductors are still relatively unexplored compared to other families of materials such as oxides and there remains broad space for materials discovery and design.[6] Nitride design is highly motivated by the great number of possible but unexplored structures with promising optoelectronic properties. The structural diversity and property uniqueness of nitrides partly originates from the high valence (−3) of the nitrogen element, which requires either metal elements with a high valence or a large number of low valence metal elements in various combinations in a formula unit of a nitride compound. In nitride semiconductor

[a] International Laboratory for Quantum Functional Materials of Henan, School of Physics and Engineering, Zhengzhou University, Zhengzhou 450001, China. E-mail: wfei@zzu.edu.cn, jiayu@zzu.edu.cn
[b] Materials Science and Engineering Program, University of California San Diego, La Jolla, California 92093, USA
[c] National Center for International Joint Research of Electronic Materials and Systems, School of Information Engineering, Zhengzhou University, Zhengzhou, Zhengzhou 450001, China
[d] Quantum Materials Research Center, College of Physics and Electronic Engineering, Zhengzhou Normal University, Zhengzhou 450044, China
[e] Institute of Materials and Systems for Sustainability, Nagoya University, Nagoya 464-8603, Japan
[f] Key Laboratory for Special Functional Material of Ministry of Education, and School of Materials Science, Henan University, Kaifeng 475004, China

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8tc05554h

design, the property most worth-investigating is band gap because it is the determining factor that affects the performance of nitride semiconductors in optoelectronic devices and is thus also regarded as the most commercially significant property. It has been demonstrated that band gap exhibits satisfying tunability when alloying different metal elements or altering compositions in nitride compounds.[7,8] In addition, nitride materials have also been used in semiconductor heterojunctions.[9,10] In semiconductor heterojunction engineering, the band offset between two connecting materials acts as the key parameter that determines junction performance, such as potential barrier and mobility.[11] Therefore, designing new nitride materials through elemental and compositional modulation followed by band gap and band offset measurement or calculation should be an effective way of discovering new nitride semiconductors.

Owing to the aforementioned extremely large amount of possible nitride structures, experimentally, it is difficult to fabricate and characterize the overall possible new nitride semiconductors. From a theoretical perspective, although conventional density functional theory (DFT) is relatively computationally efficient, it suffers from obvious band gap underestimation.[12] Accurate band gap calculations require advanced methods, such as the screened hybrid functionals of Heyd–Scuseria–Ernzerhof (HSE)[13] or many body perturbation theories;[14] however, these are both far more computationally expensive than DFT and are not able to be applied to large materials sets. For band offset calculation, although using DFT for interfacial potential alignment is accurate enough,[15] typically, a superlattice that consists of hundreds of atoms needs to be built up even for a simple compound, which is also computationally expensive when applied to large materials sets. A successful machine learning model is typically trained by a small subset of a large dataset and is able to predict the whole dataset within an acceptable error. In the case of band gap and band offset calculation in new nitrides, if accurate first-principles calculations are performed on a small subset of nitrides or their junctions and the results are used to train a machine learning model, it is highly likely that all nitrides in the design space can be accurately predicted. Work on band gap prediction by using machine learning methods has been reported: Zhuo *et al.* used 136 engineered elemental features and an SVR model trained and tested on 3896 various forms of semiconductors for experimental band gap prediction, achieving a RMSE of 0.45 eV.[16] By using 18 features including both elemental properties and low-level DFT computational results of compounds, Lee *et al.* used an SVR model on 270 binary and ternary compounds and achieved a RMSE of 0.24 eV in experimental band gap prediction.[17] Weston *et al.* trained and tested an SVR model on 284 I_2–II–IV–VI_4 kesterite compounds with HSE calculated band gaps by using 12 elemental features, achieving a RMSE of 0.283 eV.[18] To the best of our knowledge, by using a machine learning approach, neither systematic work on nitride band gap prediction nor band offset prediction for bulk materials have ever been reported.

In the present paper, 16-atom constructed wurtzite nitrides in an orthorhombic cell were studied and 68 115 possible materials were considered based on all possible cation–nitrogen combinations in the design space. 300 out of the total 68 115 materials were randomly selected and their band gap was calculated by using the hybrid functionals of the HSE method and the band offset against wurtzite GaN was calculated using the combination of HSE and DFT-GGA (Generalized Gradient Approximation) based on interface models. The calculated results were used to train and test machine learning models. Various machine learning models were tested and their performances were compared with each other in terms of RMSE. By using 18 accessible elemental properties as features, radial kernel SVR with an RMSE of 0.183 eV performed best for band offset prediction. For band gap prediction, radial kernel SVR is again the best model and shows an RMSE of 0.298 eV with the same 18 elemental features. Through feature engineering, 26 elemental properties were taken as features and the RMSE decreased by around 0.005 eV compared to 0.298 eV and the relative importance of elemental properties for band gap prediction was found. Our results show that the designed nitrides exist in all valuable band gap ranges and, interestingly, as the number of types of cations increases from 1 to 8, the mean band gap decreases and mean band offset increases. Both the predicted values and discovered trends with cation type number will be useful as guidance for computational and experimental investigations on nitride engineering with higher precision.

Methodology

a. Materials design space

The materials studied in this paper have been derived from a 16-atom 2 × 2 × 2 supercell of wurtzite GaN by cation transmutations and combinations. In the design space, +2, +3 and +4 cations were considered to occupy the positions near nitrogen anions. +2 cations are from group II A: Be^{2+}, Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+} and group II B: Zn^{2+} and Cd^{2+}; +3 cations are from group III A: Al^{3+}, Ga^{3+} and In^{3+} and group III B: Sc^{3+} and Y^{3+}; +4 cations are from group IV A: Si^{4+}, Ge^{4+} and Sn^{4+} and group IV B: Ti^{4+}, Zr^{4+} and Hf^{4+}. With the consideration of the proper size of the total materials set, occupations are divided into three

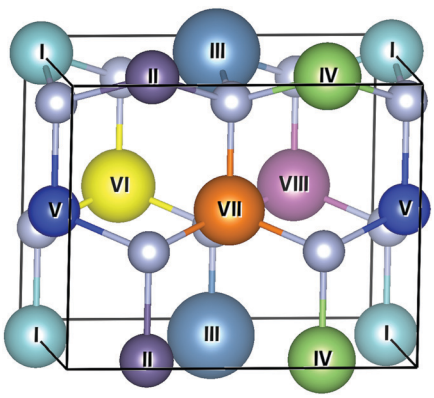

Fig. 1 The nitride structure in the design space and positions of 16 ions. The eight labelled balls are cations and the other eight unlabelled balls are nitrogen atoms.

types and rules are shown in Fig. 1. All eight cation positions are entirely symmetric. Type 1: I, III, V and VII are occupied by +2 cations while II, IV, VI and VIII are occupied by +4 cations; type 2: all eight cation positions are occupied by +3 cations; type 3: I and III are occupied by +2 cations, II and IV are occupied by +4 cations while all the rest of the cation positions are occupied by +3 cations. In total, 68 115 different nitrides were constructed and 300 materials were randomly selected for training and testing of the machine learning model. These 300 nitrides with feature space and all computational results from first-principles calculations are listed in Table S1 in the ESI.† For the nitride compounds designed in this work, owing to the high electronegativity of the nitrogen atom, the binding energy of the system is usually high and it can be expected that the nitrides in the design space should be thermally stable. It has been checked that the formation energies of the 300 randomly selected materials are all negative. These calculated formation energies are listed in Table S1 in the ESI.† Therefore, the designed 16-atom supercell based on wurtzite GaN is reasonable in this study.

b. First-principles calculation

The first-principles calculations were performed using the plane-wave pseudopotential as implemented in the VASP code.[19,20] The electron–core interactions are described with the frozen-core projected augmented wave pseudopotentials.[21] The generalized gradient approximation (GGA) formulated by Perdew, Burke, and Ernzerhof (PBE) as the exchange–correlation functional[22] with a cut-off energy of 500 eV for basic functions was chosen in all of our calculations. A reciprocal space sampling of $6 \times 5 \times 6$ Monkhorst–Pack mesh[23] of the Brillouin zone is used in the structural optimizations. All the structures are fully relaxed until the forces on each atom are smaller than 0.01 eV $Å^{-1}$ with a tetrahedron method with Blöchl corrections in broadening of 0.05 eV. The screened hybrid functional of HSE with $\alpha = 0.31$[24] rather than the typical value of 0.25 was performed on PBE optimized structures for band gap calculation. A comparison test between the HSE calculated bandgap based on an exchange parameter of 0.31, typical value 0.25 and experimental or GW calculated band gap values for seven nitrides in the design space indicated that 0.31 leads to more accurate bandgap values. The test results are tabulated in Table S5 in the ESI.† For band offset calculation, 300 superlattices were formed on the PBE optimized isolated nitrides along the (001) direction as $(XN)_n/(GaN)_n$, where $n = 5$ and XN represents each of the 300 nitrides we have randomly selected. All the energy levels in isolated materials have been calculated through the HSE method and energy levels in the constructed interface models have been calculated at the DFT-PBE level.

By using the DFT-PBE method, for each constructed nitride compound, the energy of the compound (E_C) and the energy of the most stable elementary substance of every component element ($E_{I–VIII}$) are calculated. Formation energies (E_f) were obtained through eqn (1):

$$E_f = E_C - \sum_{I}^{VIII} E_i. \quad (1)$$

Table 1 Computational band offset comparison between our method and Wei's results

	Band offset calculated (eV)	Band offset reported (eV)
AlN/GaN	−1.11	−1.28[25]
InN/AlN	1.16	1.11[25]

The band offset against wurtzite GaN was calculated using Wei's core level method[25,26] with the following eqn (2):

$$\Delta E_V(XN/GaN) = \Delta E_{V,C'}(XN) - \Delta E_{V,C}(GaN) + \Delta E_{C'/C}(XN/GaN) + A_V(XN) + A_V(GaN) \quad (2)$$

where $\Delta E_{V,C'}(XN) = E_V(XN) - E_{C'}(XN)$, $\Delta E_{V,C}(GaN) = E_V(GaN) - E_C(GaN)$, E_V is the energy level of the valence band maximum (VBM) in isolated materials, E_C ($E_{C'}$) is the core energy level in isolated materials, $\Delta E_{C'/C}$ is the core energy level difference between two materials at both sides of an interface model constructed and A_V is the valence band deformation potential. The core level has been set to be the 1s level of the nitrogen atom owing to the adequately low energy, which is around −370 eV. Considering the geometric similarity among constructed nitrides in the design space, for simplicity, it was assumed that the valence band deformation potential of each compound when connected with wurtzite GaN was neglected, *i.e.*, $A_V(GaN) + A_V(XN) = 0$. Consequently, the band offset between wurtzite XN and GaN can be written as $\Delta E_{V,C'}(XN) - \Delta E_{V,C}(GaN) + \Delta E_{C'/C}(XN/GaN)$. Calculated band offsets were compared with the widely accepted results reported by Wei[25] in Table 1, which shows satisfying consistency.

c. Machine learning

Machine learning work in this paper was implemented in Python 2.7 code with frameworks Scikit-learn[27] for SVR and Tensorflow[28] for linear regression and neural networks. Three types of machine learning models were used: support vector regression (SVR) with linear, polynomial and radial kernels, linear regression and neural network (NN) with single hidden layer (ANN) and two hidden layers (DNN). All hyper-parameters were optimized. In order to prevent overfitting, the L^2 regularization term was added to the loss function of the NN models. According to previous work, the covalent radius, electronegativity and valence of each component element are three of the most common elemental features chosen for machine learning predictions of band gap properties. For example, electronegativity, ionic radius, and row in the periodic table have been used by Weston *et al.* in the prediction of bandgap of I_2–II–IV–VI_4 kesterite compounds.[18] Elemental information including absolute value of the formal ionic charge, period in the periodic table, atomic number, atomic mass, van der Waals radius, electronegativity, and the first ionization energy were chosen as features for bandgap predictions in binary and ternary compounds by Lee *et al.*[17] From the view of physical intuition, covalent radius, electronegativity and valence are the most important electronic properties of an element and should be the most relevant and effective descriptors for electronic band gap and alignment

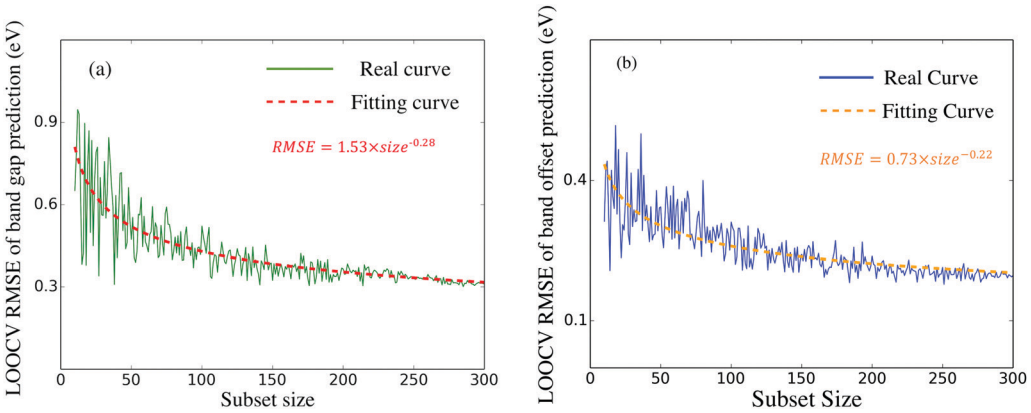

Fig. 2 (a) LOOCV RMSE of band gap prediction *versus* subset size (green curve). The red dashed curve is the fitting curve fitted by the power function shown in red. (b) LOOCV RMSE of band offset prediction *versus* subset size (blue curve). The orange dashed curve is the fitting curve fitted by the power function shown in orange.

predictions in this study. Therefore, these three elemental features have been chosen as an initial trial. After removing symmetrically repeated values, an 18-dimensional feature space was built and used first. Model performance was evaluated by averaged RMSE of the validation set in 10-fold cross validation.

Results & discussion

In order to check if the 300 randomly selected nitrides are adequate to effectively learn band gap and band offset, the leave-out-one-cross-validation (LOOCV) RMSEs of radial kernel SVR models trained and tested on randomly selected subsets of the 300 nitride samples were calculated. Band gap and band offset prediction RMSE with subset size are plotted in Fig. 2(a) and (b), respectively. The curves were fitted with power functions[29] and it is shown that with 300 nitrides, the prediction capacity is adequately stable and almost reaches its limit. Therefore, a sample set with a size of 300 should be large enough for a machine learning model to learn in this work.

a. Band offset regressor

For band offset prediction, by using the 18-dimensional elemental feature space, RMSEs of all models are shown in Table 2. Optimized hyper-parameters are listed in Table S2 in the ESI.†

The smallest RMSE, 0.183 eV, suggests that SVR with radial kernel is the best model for band offset prediction. In order to intuitively show the accuracy of band offset prediction, predicted band offset values as a function of calculated band offset

Table 2 RMSE of band offset and band gap prediction for different machine learning models. The RMSE values listed are the averaged RMSE of the validation set in 10-fold cross validation

	Linear SVR	Poly SVR	Radial SVR	Linear regression	ANN	DNN
Band offset RMSE (eV)	0.256	0.239	0.183	0.256	0.219	0.230
Band gap RMSE (eV)	0.412	0.335	0.298	0.474	0.385	0.379

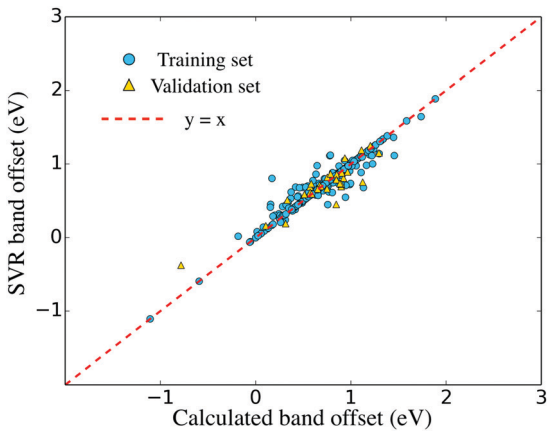

Fig. 3 SVR predicted band offset *versus* calculated band offset. The blue circles represent the training set, the gold triangles represent the validation set and the red dashed line is the guidance line on which prediction error is zero.

in both the training and validation sets are plotted in Fig. 3. The excellent prediction performance in both the training set and the validation set indicates that the model is neither under-fitting nor over-fitting.

b. Band gap regressor

For band gap prediction, the RMSEs of all optimized models trained with the elemental 18-feature space are shown in Table 2. Optimized hyper-parameters are listed in Table S2 in the ESI.†

SVR with radial kernel again performs best for band gap prediction and came up with a RMSE of 0.298 eV. From a perspective of first-principles calculations, the HSE calculated band gap is sensitive to the exchange parameter with empirically selected values, which gives rise to a band gap calculation uncertainty reaching up to 0.4 eV.[24] Since the RMSE of 0.298 eV is within the HSE calculation uncertainty, the model's performance is satisfactory. When the band gap of each nitride compound calculated by the DFT-PBE method was added into the 18-dimensional feature space, a radial kernel-based SVR model

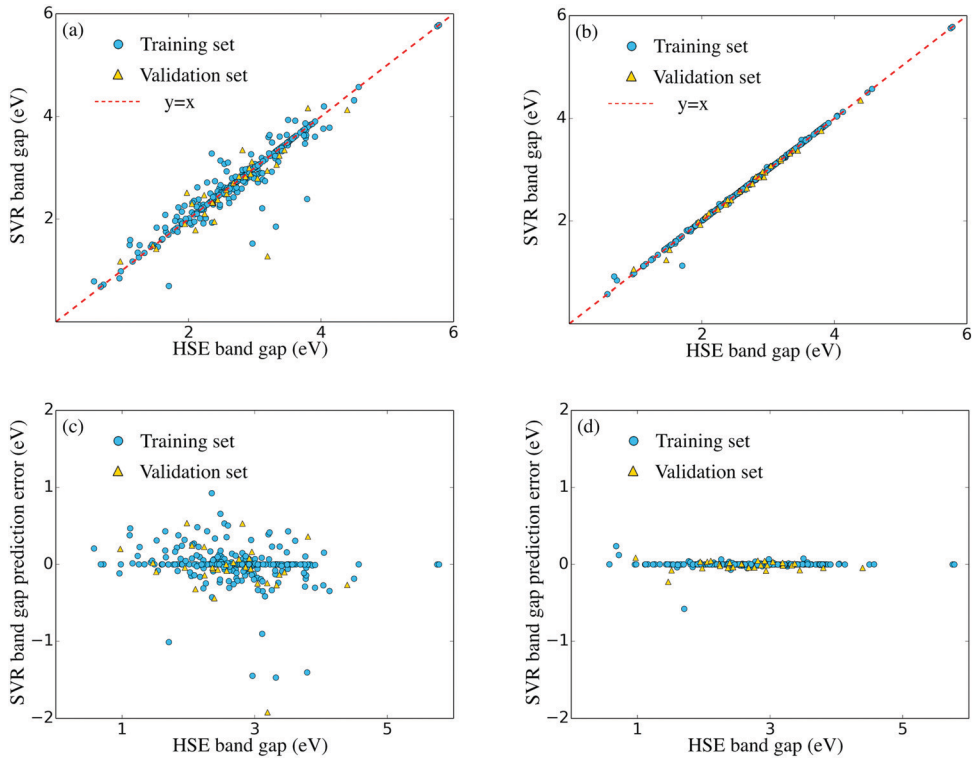

Fig. 4 (a) SVR predicted band gap *versus* HSE calculated band gap in 18-dimensional elemental property-based feature space. (b) SVR predicted band gap *versus* HSE calculated band gap in 19-dimensional PBE band-gap-included feature space. (c) SVR band gap prediction error (the difference between SVR predicted band gap and HSE calculated band gap) *versus* HSE calculated band gap in 18-dimensional elemental property-based feature space. (d) SVR band gap prediction error *versus* HSE calculated band gap in 19-dimensional PBE-band-gap-included feature space. Blue circles represent the training set and gold triangles represent the validation set.

was trained under the new 19-dimensional feature space and the validation RMSE becomes as low as 0.099 eV. A performance comparison of radial kernel SVR models with the 18-dimensional and 19-dimensional feature space is shown in Fig. 4. The tremendous accuracy enhancement by introducing the PBE band gap can be explained by the approximately linear relationship between the PBE band gap and the HSE band gap.[30]

As a trial to further improve the performance of SVR with radial kernel for band gap prediction based on elemental properties, feature space expansion implemented by an elemental property-based recursive feature extraction (EPRFE) algorithm was conducted. In EPRFE, firstly, a larger feature space that includes all accessible and reportedly-band-gap-related elemental properties was built. After removing symmetrically repeating features, a new 58-dimensional feature space was established that includes eight elemental properties: covalent radius, electronegativity, valence, atomic number, periodic number, atomic weight, first ionization energy and melting point. Secondly, models were trained and tested with the feature space that is the subset of the 58-dimensional space based on all possible combinations of the eight elemental properties and the validation RMSEs of all 255 combinations were compared. The lowest RMSEs with the corresponding number of properties selected are shown in Fig. 5. Interestingly, it was found that the lowest RMSE when three properties are selected corresponds to the covalent radius, electronegativity and valence, exactly the three

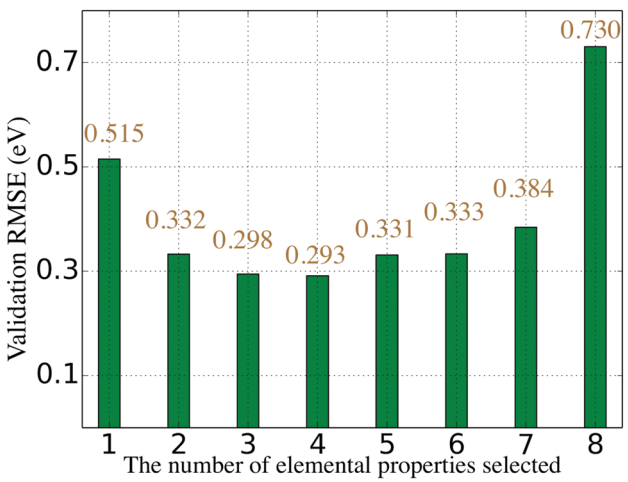

Fig. 5 Lowest validation RMSEs in 10-fold cross validation with the number of elemental properties selected.

properties in the original 18-dimensional elemental feature space. RMSE can be further decreased slightly by around 0.005 eV when the first ionization energies were introduced as new features, which corresponds to the case of four properties selected in Fig. 5. Besides, in Fig. 5, the lowest RMSE of one property corresponds to electronegativity and the lowest RMSE of two properties corresponds to electronegativity and covalent radius, which indicates

Table 3 Comparison of predicted band gaps of previously explored nitrides with reported and HSE values

	Reported E_g (eV)	Predicted E_g (eV)	HSE calculated E_g (eV)
$AlGaN_2$	4.650[7]	4.570	4.569
$InGaN_2$	1.795[7]	2.272	1.925
$AlInN_2$	2.890[7]	3.445	2.976
$ZnGeN_2$	3.420[31]	3.405	3.406
$ZnSnN_2$	2.020[31]	2.155	1.566
$MgGeN_2$	5.140[32]	4.857	4.304
$MgSiN_2$	5.840[32]	5.753	5.755
$CaSiN_2$	4.500[33]	4.701	5.072

Table 4 Predicted band gaps of selected previously unexplored nitrides in three domains of applications: infrared detectors, solar cell absorbers and ultraviolet LEDs

Infrared detector	E_g (eV)	Solar cell absorber	E_g (eV)	Ultraviolet LED	E_g (eV)
$BeBaSn_2In_4N_8$	0.016	$CaSnGa_2N_4$	1.226	$BeMgSiTiN_4$	4.862
$CdSnIn_2N_4$	0.044	$CdSiSn_2N_4$	1.306	$BeMg_3Si_2Ge_2N_8$	4.964
$CdSnGaInN_4$	0.345	$BaCdSn_2N_4$	1.341	$Mg_4GeTi_3N_8$	5.008
$SrSnIn_2N_4$	0.371	$SrCdSn_2N_4$	1.368	Mg_2SiGeN_4	5.020
$BaSnIn_2N_4$	0.420	$BaGeGa_2N_4$	1.483	$BeSiAl_2N_4$	5.106
$CdGeIn_2N_4$	0.554	$CdGeGa_2N_4$	1.491	$BeMgSi_2N_4$	5.457
$CaSnIn_2N_4$	0.561	$SrGeGa_2N_4$	1.499	$BeMg_3Si_2TiZrN_8$	5.020
$ZnSnIn_2N_4$	0.596	$SrSnGaYN_4$	1.191	$BeMg_3Si_2GeTiN_8$	5.056

that the relative importance of each property for SVR-based band gap prediction from high to low is electronegativity, covalent radius, valence and first ionization energy. Other more complex feature engineering methods such as different orders of polynomial feature combinations with filtering method for large-scale feature selection were tried, but no improvement was observed in the model performance.

c. Predicted results

The band gap and band offset against wurtzite GaN of all 68 115 constructed nitrides were predicted by using radial SVR with 26-dimensional (original 18 features plus eight first ionization energies) and the original 18-dimensional feature space, respectively. The predicted band gaps and band offsets of all 68 115 nitrides are listed in Table S4 in the ESI.† The predicted band gaps of several nitrides that have been previously investigated are listed

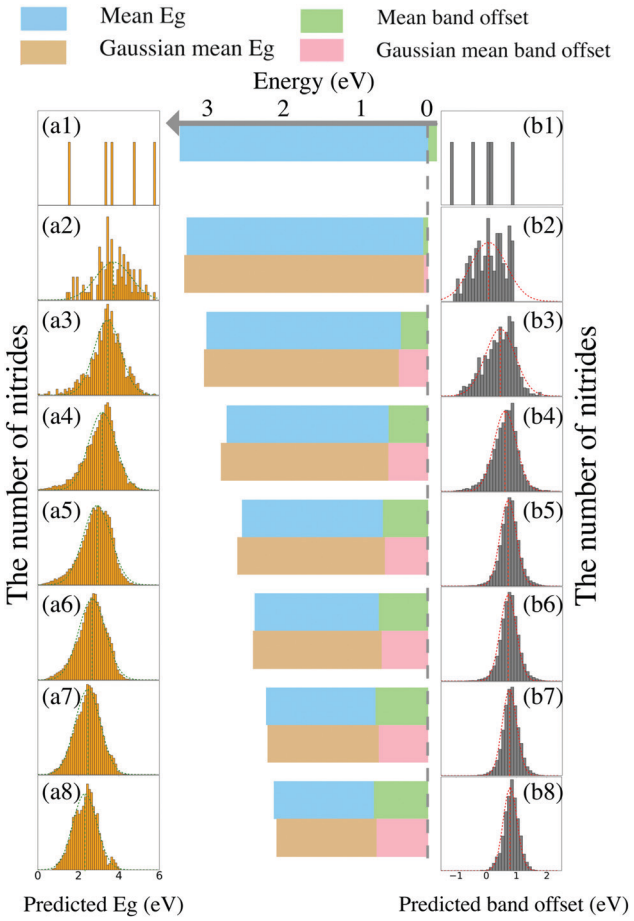

Fig. 7 Left: Distribution of predicted band gap of all designed nitrides with 1–8 types ((a1)–(a8)) of cations. Right: Distribution of predicted band offset against wurtzite GaN of all designed nitrides with 1–8 types ((b1)–(b8)) of cations. Middle: Mean and Gaussian mean of predicted band gaps and band offsets *versus* the number of types of cations, horizontally matched with left and right figures. Gaussian fittings for one type of cation were not made due to small sample size.

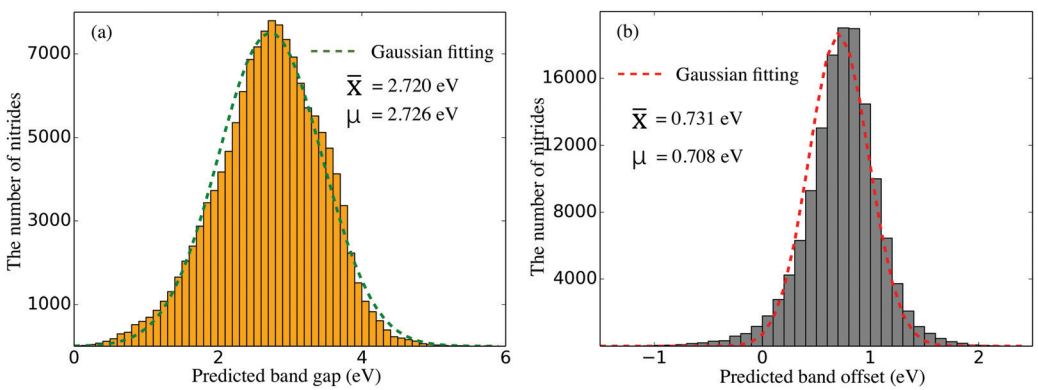

Fig. 6 (a) Distribution of predicted band gaps of all designed nitrides. (b) Distribution of predicted band offset (against wurtzite GaN) of all designed nitrides. The green and red dashed curve is the fitting curve fitted by Gaussian function. $\bar{x}$ is the mean value of predicted results. μ is the mean value of the Gaussian fitting curves.

in Table 3 in a comparison with reported and HSE calculated results. In Table 4, some previously unexplored nitrides are listed with band gaps categorized into three application domains: infrared detector, solar cell absorber and ultraviolet LED. The overall distributions of predicted band gap and band offset with Gaussian fitting curves are shown in Fig. 6. The distributions of band gap and band offset by different numbers of cation types with Gaussian fitting curves are shown in Fig. 7. It was found that the designed nitrides exist in all valuable band gap ranges and, pretty interestingly, as the number of cation types increases, both the mean and Gaussian mean band gap tend to decrease while both the mean and Gaussian mean band offset tend to increase. Mean and Gaussian mean values with the number of cation types are listed in Table S3 in the ESI.† It is suggested that both theorists and experimentalists can make further investigations on their nitrides of interest included among the predicted results in this work. Specifically, people can find their targeted materials by looking for satisfactory band gaps in the band gap database predicted through the band gap regressor. When searching for materials to make heterojunctions, targeted materials can be found by screening both band gap and band offset data generated by both regressors. Furthermore, when looking for promising materials for specific device applications, various device parameters need to be taken into consideration. If device parameter regressors were built for targeted applications, such as infrared detectors, solar cell absorbers and ultraviolet LEDs, then combined with band gap and band offset regressors, materials with potential for excellent device performance can be found from a materials database based on the three sorts of regressors.

Conclusions

In this work, machine learning models trained on first-principles calculation results were utilized to successfully provide accurate predictions for band gap and band alignment of nitrides in a large design set. After model comparison, SVR with the radial kernel function came up with the lowest RMSE of 0.183 eV for band offset prediction and, through feature engineering, a RMSE of 0.293 eV for band gap prediction. It was found that when the DFT-PBE-calculated band gap was introduced into the feature space, the band gap prediction RMSE could jump down to 0.099 eV. Eventually, the band gap and band offset were predicted on all of the 68 115 nitrides in the design space and nitrides with useful band gaps and alignment were discovered. The prediction results also indicate that the more types of cations a nitride includes, the smaller the band gap and the larger the band offset it tends to have. Along with the predicted results, further investigations can be conducted on new nitride semiconductor materials with desired applications.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

This work was jointly supported by the National Natural Science Foundation of China (Grant No. 11774078), the National Key R&D Program of China (2016YFE0118400), and the Outstanding Young Talent Research Fund of Zhengzhou University (Grant No. 1521317008). The calculations were performed in the high-performance computational center of Zhengzhou University.

References

1 K. T. Butler, D. W. Davies and H. Cartwright, *et al.*, Machine learning for molecular and materials science, *Nature*, 2018, **559**(7715), 547–555.
2 Y. Liu, T. Zhao and W. Ju, *et al.*, Materials discovery and design using machine learning, *J. Materiomics*, 2017, **3**(3), 159–177.
3 S. Curtarolo, G. L. W. Hart and M. B. Nardelli, *et al.*, The high-throughput highway to computational materials design, *Nat. Mater.*, 2013, **12**(3), 191–201.
4 Y. Zhu, X. Kong and T. D. Rhone, *et al.*, Systematic search for two-dimensional ferromagnetic materials, *Phys. Rev. Mater.*, 2018, **2**(8), 081001.
5 T. D. Moustakas and R. Paiella, Optoelectronic device physics and technology of nitride semiconductors from the UV to the terahertz, *Rep. Prog. Phys.*, 2017, **80**(10), 106501.
6 A. Zakutayev, Design of nitride semiconductors for solar energy conversion, *J. Mater. Chem. A*, 2016, **4**(18), 6742–6754.
7 I. Vurgaftman and J. R. Meyer, Band parameters for nitrogen-containing semiconductors, *J. Appl. Phys.*, 2003, **94**(6), 3675–3696.
8 J. Wu, When group-III nitrides go infrared: new properties and perspectives, *J. Appl. Phys.*, 2009, **106**(01), 011101.
9 E. T. Yu, X. Z. Dang and P. M. Asbeck, *et al.*, Spontaneous and piezoelectric polarization effects in III–V nitride heterostructures, *J. Vac. Sci. Technol., B: Microelectron. Nanometer Struct.–Process., Meas., Phenom.*, 1999, **17**(4), 1742–1749.
10 A. Woessner, M. B. Lundeberg and Y. Gao, *et al.*, Highly confined low-loss plasmons in graphene–boron nitride heterostructures, *Nat. Mater.*, 2015, **14**(4), 421.
11 H. Kroemer, Heterostructure bipolar transistors and integrated circuits, *Proc. IEEE*, 1982, **70**(1), 13–25.
12 P. Mori-Sánchez, A. J. Cohen and W. Yang, Localization and delocalization errors in density functional theory and implications for band-gap prediction, *Phys. Rev. Lett.*, 2008, **100**(14), 146401.
13 J. Heyd, J. E. Peralta and G. E. Scuseria, *et al.*, Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional, *J. Chem. Phys.*, 2005, **123**(17), 174101.
14 M. Shishkin and G. Kresse, Self-consistent G W calculations for semiconductors and insulators, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2007, **75**(23), 235102.
15 L. Weston, H. Tailor and K. Krishnaswamy, *et al.*, Accurate and efficient band-offset calculations from density functional theory, *Comput. Mater. Sci.*, 2018, **151**, 174–180.

16 Y. Zhuo, A. Mansouri Tehrani and J. Brgoch, Predicting the Band Gaps of Inorganic Solids by Machine Learning, *J. Phys. Chem. Lett.*, 2018, **9**(7), 1668–1673.
17 J. Lee, A. Seko and K. Shitara, *et al.*, Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques, *Phys. Rev. B*, 2016, **93**(11), 115104.
18 L. Weston and C. Stampfl, Machine learning the band gap properties of kesterite I_2–II–IV–V_4 quaternary compounds for photovoltaics applications, *Phys. Rev. Mater.*, 2018, **2**(8), 085407.
19 G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**(16), 11169.
20 G. Kresse and J. Furthmüller, Efficiency of *ab initio* total energy calculations for metals and semiconductors using a plane-wave basis set, *Comput. Mater. Sci.*, 1996, **6**(1), 15–50.
21 P. E. Blöchl, Projector augmented-wave method, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**(24), 17953.
22 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1996, **77**(18), 3865.
23 H. J. Monkhorst and J. D. Pack, Special points for Brillouin-zone integrations, *Phys. Rev. B: Solid State*, 1976, **13**(12), 5188.
24 T. Wang, C. Ni and A. Janotti, Band alignment and p-type doping of $ZnSnN_2$, *Phys. Rev. B*, 2017, **95**(20), 205205.
25 Y. H. Li, A. Walsh and S. Chen, *et al.*, Revised *ab initio* natural band offsets of all group IV, II–VI, and III–V semiconductors, *Appl. Phys. Lett.*, 2009, **94**(21), 212109.
26 S. H. Wei and A. Zunger, Calculated natural band offsets of all II–VI and III–V semiconductors: Chemical trends and the role of cation d orbitals, *Appl. Phys. Lett.*, 1998, **72**(16), 2011–2013.
27 https://http://scikit-learn.org/.
28 M. Abadi, P. Barham and J. Chen, *et al.*, Tensorflow: a system for large-scale machine learning, *OSDI*, 2016, **16**, 265–283.
29 Y. Zhang and C. Ling, A strategy to apply machine learning to small datasets in materials science, *npj Comput. Mater.*, 2018, **4**(1), 25.
30 Z. Zhu, B. Dong, T. Yang, *et al.*, Fundamental Band Gap and Alignment of Two-Dimensional Semiconductors Explored by Machine Learning, 2017, arXiv preprint arXiv:1708.04766.
31 A. Punya, W. R. L. Lambrecht and M. van Schilfgaarde, Quasiparticle band structure of Zn-IV-N_2 compounds, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**(16), 165204.
32 A. P. Jaroenjittichai and W. R. L. Lambrecht, Electronic band structure of Mg–IV–N_2 compounds in the quasiparticle-self-consistent G W approximation, *Phys. Rev. B*, 2016, **94**(12), 125201.
33 W. A. Groen and M. J. Kraan, New ternary nitride ceramics: $CaSiN_2$, *J. Mater. Sci.*, 1994, **29**(12), 3161–3166.