RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2024, **14**, 6385

Predicting band gaps of ABN_3 perovskites: an account from machine learning and first-principle DFT studies†

Swarup Ghosh and Joydeep Chowdhury *

The present paper is primarily focused on predicting the band gaps of nitride perovskites from machine learning (ML) models. The ML models have been framed from the feature descriptors and band gap values of 1563 inorganic nitride perovskites having formation energies <−0.026 eV and band gaps ranging from ∼1.0 to 3.1 eV. Four supervised ML models such as multi-layer perceptron (MLP), gradient boosted decision tree (GBDT), support vector regression (SVR) and random forest regression (RFR) have been considered to predict the band gaps of the said systems. The accuracy of each model has been tested from mean absolute error, root-mean-square error and determination coefficient R^2 values. The bivariate plots between the predicted and input band gaps of the compounds for both the training and test datasets have also been estimated. Additionally, two ABN_3-type nitride perovskites $CeBN_3$ (B = Mo, W) have been selected and their electronic band structures and optoelectronic properties have been studied from density functional theory (DFT) calculations. The band gap values of the said compounds have been estimated from DFT calculations at PBE, HSE06, G_0W_0@PBE, G_0W_0@HSE06 level of theories. The present study will be helpful in exploring the ML models in predicting the band gaps of nitride perovskites which in turn may bear potential applications in photovoltaic cells and optical luminescent devices.

Received 15th January 2024
Accepted 14th February 2024

DOI: 10.1039/d4ra00402g

rsc.li/rsc-advances

1. Introduction

Perovskites are ternary compounds with general chemical formula ABX_3 where "A" and "B" are cationic elements and "X" can be either oxygen or halogens in anionic forms. While "A" atoms are located in the cuboctahedral cavities of the crystal, "X" atoms on the other hand form corner sharing BX_6 octahedra. These compounds are known to show fascinating physical and chemical properties and find extensive applications in the fabrications of solar, fuel cells, energy conversion and optoelectronic devices.[1–10] Moreover, perovskites containing heavy fermionic elements, that show Rashba splittings, find promising applications in spintronics.[11–14]

Despite considerable successes of oxide and halide perovskites, recently much attention has been focused on the synthesis and first-principle calculations of nitride perovskites having general formula ABN_3.[15–19] Of late, ABN_3 find potential applications as topological insulators, photovoltaic cells and are being successfully used as optical luminescent materials.[18,20,21] The primary driving force behind the realization of ABN_3 stems from their extraordinary electronic, optical and magnetic properties of nitride compounds. For example, 2D hexagonal boron nitride is used in 5G wireless technology.[22] Vanadium nitride (VN), niobium nitride (NbN) and tantalum nitride (TaN) compounds are commonly employed in the fabrications of high – T_c superconductors, supercapacitors and batteries.[23–27]

Barring several attributes, syntheses of nitride perovskites are experimentally challenging as their synthetic protocols demand oxygen free environment.[28] Despite this limitation, successful preparations of $ThTaN_3$, $LaWN_3$, $LaReN_3$, $LaMoN_3$, $CeMoN_3$, $CeWN_3$, $YMoN_3$ and YWN_3 systems have so far been reported.[28–31] Readers who are interested in their synthesis procedures are referred to the literature[28–31] as depicted under reference section. Interestingly, these nitride perovskites find wide applications as topological insulators, spintronics and microelectromechanical devices.[21,30,32]

Considering the enormous applications of nitride perovskites, *in silico* approaches alone or in conjunction with experiments have been successfully applied to predict the structure–function relationships of these compounds. A combined minima hopping method and density functional theory (DFT) calculations have been performed to predict the crystal structures and band gaps (E_g) of twenty-one ABN_3 perovskites such as $LaBN_3$ (B = Re, W, Mo, Tc, Os, Cr, Mn, Co), $SrBN_3$ (B = Re, W, Tc), $CaBN_3$ (B = Re, W, Mo, Tc), $BaBN_3$ (B = Re, W) and YBN_3 (B = Re, W, Mo, Tc).[19] Those systems were observed to be thermodynamically stable

Department of Physics, Jadavpur University, 188, Raja S.C. Mallick Road, Kolkata 700032, India. E-mail: joydeep72_c@rediffmail.com; joydeep.chowdhury@jadavpuruniversity.in

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ra00402g

and proposed to have excellent chances of being experimentally accessible.[19] Recently, the first-principle DFT calculations have been carried out to calculate the magnetic moments and thermodynamic stabilities of some rare-earth nitride perovskites ABN_3 (A = La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu and B = Re, W) and suggested their numerous technological applications in the domain of nitride materials.[16] In this regard DFT is now recognized to be an elegant approach to estimate the electronic and optoelectronic properties of the materials under study. Both the electronic as well the optoelectronic properties are primarily guided by the band gaps of the materials. While DFT calculations with local density approximation (LDA) and generalized gradient approximation (GGA) underestimate E_g values,[33–36] the unscreened hybrid and Perdew–Burke–Ernzerhof–Hartree–Fock exchange (PBE0) functionals overestimate band gap energies of the compounds relative to their experimental counterparts.[37–39] In this regard DFT calculations, as accomplished from single-shot GW (G_0W_0) approximation using hybrid exchange-correlation (XC) functionals such as Heyd–Scuseria–Ernzerhof (HSE), Becke-3-parameter-Lee-Yang-Parr (B3LYP) and B3PW91 are known to predict the E_g values of the compounds close to the experimental results.[14,33,40–48] The major pitfalls of such calculations are that they are computationally demanding and need high end servers to run them. In this context, the machine learning (ML) is now considered as an effective alternative route to avoid the inherent computational costs linked with DFT calculations and helps in establishing a simple model between the characteristics features of materials and the target variable (here E_g).[49–60] Although ML approach is successfully implemented recently to predict the band gaps of oxide, halide perovskites and double perovskite compounds,[61–66] no such report is found in predicting the band gaps of nitride perovskites.

Considering the above issues in mind, the present paper is aimed to predict the band gaps of ABN_3 perovskites from ML models. The DFT studies have been performed to estimate the electronic band structures, E_g values and optoelectronic properties of two new nitride perovskites $CeBN_3$ (B = Mo, W). The manuscript has been organized in the following manner. In Section 2, the computational methodology, which includes ML methods and first-principle DFT calculations, has been discussed. Cleaning and preprocessing of data for ABN_3 perovskites have been shared in Section 3.1. In section 3.2, the training and validation of ML models have been deliberated. Section 3.3 is devoted to understand the structural properties and stabilities of two newly discovered nitride perovskite compounds $CeBN_3$ (B = Mo, W). Section 3.4 is framed with the calculations of electronic band structures and band gaps of $CeBN_3$ compounds using different level of theories in DFT calculations. The corresponding optoelectronic properties have also been highlighted in the Section 3.5. Overall conclusions of this work have been discussed under Section 4.

2. Computational methodology

2.1. Machine learning models

To predict the band gaps of nitride perovskite compounds, the ML models such as support vector regression (SVR), gradient boosted decision tree (GBDT), random forest regression (RFR) and multi-layer perceptron (MLP) algorithms have been considered. The input data consist of material descriptors of nitride perovskites and their corresponding target variables E_g. The algorithms have been realized using the Scikit-learn software package within the Python 3.9 framework.[67] The parameters of the algorithms have been optimized and the model performances have been estimated using the grid search of the average root-mean-square error (RMSE) of each model validation set. The E_g values have been predicted from SVR model using the following mathematical relation:[53]

$$f_{\mathrm{SVR}}(y) = \sum_{i=1}^{N}\left(a_i - a_i^*\right)K(y_i, y) + b \tag{1}$$

where a_i and a_i^* are non-negative multipliers for each observation y_i. K is the kernel or radial basis function which is used to calculate the difference between training (y_i) and predicted (y) values of band gaps. "b" can be estimated from the Lagrange function.

The E_g values, so predicted from the GBDT algorithm, are represented using the following expression:[53]

$$f_{\mathrm{GBDT}}(y) = f_0(y) + \sum_{m=1}^{M}\sum_{j=1}^{J}\Theta_{m,j}I\left(y \in R_{m,j}\right) \tag{2}$$

where $f_0(y)$ is the initial single decision tree. M,J and $\Theta_{m,j}$ are the number of regression trees, number of leaf nodes of the trees and the best fitting value respectively for each set of leaf nodes $R_{m,j}$.

Using RFR model the predicted E_g can be represented as:[53]

$$f_{\mathrm{RFR}}(y) = \sum_{t=1}^{T}\sum_{n=1}^{N} C_n I(y \in D_n)/t \tag{3}$$

where t runs from 1 to T and T represents the number of formed classification or regression trees. C_n signifies the mean value of dataset D_n for randomly selecting N training samples from the said dataset.

The E_g values, so predicted from neural network based MLP framework, can be expressed using the following mathematical function:[53]

$$f_{\mathrm{MLP}}(y) = \sum_{j=1}^{N} {W_0}^{ij}\theta\left({W_1}^{ij}y_j + b_1\right) + b_0 \tag{4}$$

where ${W_0}^{ij}$ and ${W_l}^{ij}$ are the weights of ith neurons in the output and lth layers respectively. b_0 and b_1 represent the respective output layer and hidden biases respectively.

The characteristics of each model have been estimated from the Pearson correlation coefficient (p) which is expressed as:[68]

$$p = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{5}$$

where x_i and y_i are the comparative features of x and y respectively, $\bar{x}$ and $\bar{y}$ are their respective average values. N represent total number of samples in the training set. The accuracy of each ML model has been determined from RMSE, mean absolute error

(MAE) and evaluation coefficient (R^2) values. RMSE, MAE and R^2 have been calculated using the following relations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y)^2} \quad (6)$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - y| \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - y)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (8)$$

respectively. Where y_i and y are the respective input and predicted values of the dataset.

2.2. First-principle DFT calculations

The first-principle calculations have been carried out under the DFT framework within Quantum ESPRESSO (QE) software.[69–71] The crystal structures of $CeBN_3$ (B = Mo, W) compounds have been primarily optimized using "variable-cell relaxation" method followed by Broyden–Fletcher–Goldfarb–Shanno scheme.[72–75] The Born Oppenheimer molecular dynamics (BOMD) simulations have then been performed at $T = 300$ K on the supercell structures ($2 \times 2 \times 2$) of the compounds, so obtained from the optimized unitcell geometries of the said systems. BOMD simulations have been accomplished from the NPT ensemble whose temperature is controlled by the Nose–Hoover thermostat. The projector augmented wave (PAW) pseudopotentials[76] have been included to consider the electron–ion interactions and the exchange-correlation (XC) term of the pseudopotential has been implemented from Perdew–Burke–Ernzerhof (PBE) XC functional[77] which is known as a parameterization form of GGA. The valence electrons of Ce, B (B = Mo, W) and *N* atoms have been considered as plane wave with kinetic energy cut-off of 80 Ry. Each crystal structure is allowed to optimize until the Hellmann–Feynman force attains a value $<10^{-3}$ Ry Bohr^{-1} for total electronic energy difference $<10^{-8}$ Ry at $P = 0$ GPa. A gamma-centered *k*-point mesh of $10 \times 10 \times 10$ grid under Monkhorst–Pack scheme has been set for geometry optimization and self-consistent-field (SCF) calculations.

The electronic band structures of $CeBN_3$ crystal systems have been estimated from different level of theories such as PBE, HSE06, G_0W_0@PBE and G_0W_0@HSE06 in the DFT calculations. The optoelectronic properties of the said compounds have been calculated from SCF method as employed in QE software.[78] The optical broadening of 0.15 eV has been considered for estimating the optoelectronic parameters which include complex dielectric function [$\varepsilon(\omega)$], absorption coefficient [$\alpha(\omega)$] and optical conductivity [$\sigma(\omega)$].

3. Results and discussions

3.1. Data cleaning and pre-processing for machine learning

The raw data of inorganic ABN_3-type perovskite compounds have been obtained from the available literature.[15–19,21,28–31,79,80] By selecting all possible combinations of cations as "A" and "B", a set of 5566 inorganic ABN_3-type perovskites have been initially selected for model evaluation. Lead (Pb) containing perovskites in all combinations, due their inherent toxic characters, are not considered in the above mentioned dataset. In the next step, the samples have been primarily identified according to their respective formation energies (E_f). Using support vector classification (SVC), the samples are then categorized by eliminating those candidates which show $E_f >$ thermal excitation energy ($E_{Th} \sim -0.026$ eV) at room temperature ($T = 300$ K).[57] To implement SVC, initially 5566 samples have been randomly divided into 4866 (~87%) and 700 (~13%) samples as training and test set, respectively. Fig. 1(a) shows the E_f as a function of sample number of the test dataset as obtained from SVC algorithm. From Fig. 1(a), it is observed that the samples are segregated into $E_f < E_{Th}$ and $E_f > E_{Th}$ classes. The merit of this SVC algorithm is then tested from confusion matrix and receiver operating characteristics (ROC) curve.[55] The results are shown in Fig. 1(b) and (c) respectively. From Fig. 1(b), the correctly identified compounds with $E_f < E_{Th}$ and $E_f > E_{Th}$ classes are observed to be 482 [true positive (TP)] and 181 [true negative (TP)], respectively. However, the number of incorrectly identified samples under $E_f < E_{Th}$ and $E_f > E_{Th}$ classes are found to be 18 [false positive (FP)] and 19 [false negative (FN)], respectively. Out of 700 test dataset, it is found that only 5% samples (37 out of 700) have been misclassified, which in turn suggests the reliability of the SVC model to segregate the samples into their respective classes. The F1 score of the model, as attained from confusion matrix, is mathematically expressed as:

$$\text{F1 score} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (9)$$

$$\text{where precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (10)$$

$$\text{and recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (11)$$

The F1 score is calculated to be 96.2% which also suggests the better performance of the said model. The ROC curve, as depicted in Fig. 1(c), represents the relation between the TP rate (sensitivity) and FP rate (1 – specificity). From Fig. 1(c), the area under curve (AUC) is estimated to be 0.98. The AUC closer to 1 further signifies the excellent segregation between the $E_f < E_{Th}$ and $E_f > E_{Th}$ classes.

Here we primarily focused on ABN_3 compounds which bear potential applications as photovoltaic and optical luminescent materials, the selection space has thus been further narrowed down to 1563 samples which exhibit E_g values spanning in the range from 1.0 to 3.1 eV. The E_g of the compounds in general are linked with the proposed 145 feature descriptors which include electronegativity, atomic weight, covalent radius, d valence electrons *etc.*[57,81] Out of 145 feature descriptors, 117 feature attributes of ABN_3 perovskites have been initially framed for the target variable E_g. The list of 117 features such as Mendeleev number, electronegativity, covalent radius, d valence electrons

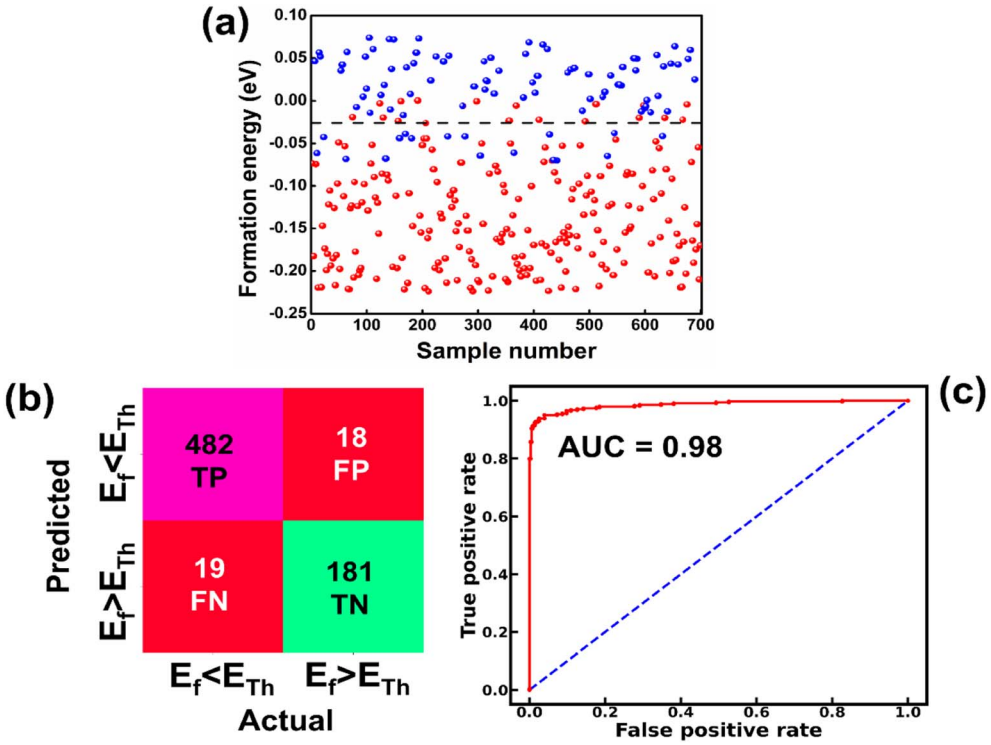

Fig. 1 (a) Formation energy as a function of sample number for 700 test sets as estimated from SVC model. [The horizontal dashed line represents formation energy (E_f) = −0.026 eV]. (b) Confusion matrix representing the correlation between predicted and actual values in form of true positive (TP), false negative (FN), false positive (FP) and true negative (TN) values for the tested samples. [$E_f < E_{Th}$ and $E_f > E_{Th}$ symbolize the formation energies below and above E_{Th} (=−0.026 eV), respectively]. (c) The ROC curve illustrating the TP and FP rates of the test datasets.

etc. are shown in Table S1 (ESI).† With these 117 features, the Pearson correlation heatmap is constructed and the result is shown in Fig. 2(a). The features having absolute correlation values <0.89 and with multiple collinearities between them have been eliminated. The final top 10 feature descriptors which include electronegativity, d valence electrons, formation energy, p valence electrons, Mendeleev number, specific volume of ground state, mean covalent radius, space group number, melting temperature and atomic weight, have been selected using the de-correlation method for further model evaluation. The heatmap portraying the evaluated top 10 feature descriptors and their order of importance are shown in Fig. 2(b) and (c), respectively. From Fig. 2(c), it is clearly seen that the electronegativity, d valence electrons and mean covalent radius show predominant contributions in predicting E_g values of the ABN_3 compounds.

The correlations between E_g and electronegativity, mean covalent radius and d valence electrons can be rationalized. Electronegativity is a measure of ability to attract electrons of two bonded atoms towards their valence electrons. This ability results the delocalized distribution of valence electrons, which in turn influences the nature of the bonding as well as the E_g values of the compounds.[82–84] The mean covalent radii of the constituent atoms are intrinsically linked with the electronegativity of the system. Systems with constituent atoms having larger mean covalent radii thus result in decrease in electronegativity of the compounds in general.[85] Hence, the mean covalent radius is considered as an important feature descriptor in modulating the E_g values of the systems. The d valence electrons play major roles in the formation of energy bands in the electronic band structure and has direct impact on the E_g of the materials. Larger number of d valence electrons can realign the Fermi energy level (E_F) through p-d and s-d hybridizations, which in turn may alter the E_g values of the compounds.[57]

3.2. Machine learning model training and validation

To precisely predict the band gaps of ABN_3-ype perovskite compounds, four ML models such as SVR, MLP, GBDT and RFR have been considered. The accuracy of each model has been tested by selecting their respective hyperparameters. The list of hyperparameters of each model are shown in Table 1. From the dataset of 1563 ABN_3 perovskites and their top 10 feature descriptors (*vide supra*), 1363 (∼87%) samples have been randomly selected for training and the rest 200 (∼13%) samples have been set for testing of each model. The random selection has been considered two times for each model to confirm the statistical validity of the results. The performance of ML model has been estimated by calculating the RMSE, MAE and R^2 values. The RMSE, MAE and R^2 values of each model for the test dataset, so attained from the 10-dimensional feature space, are illustrated in Table 1.

From Table 1, it is observed that while RFR predicts the E_g values of the compound with 3% MAE; SVR, GBDT and MLP

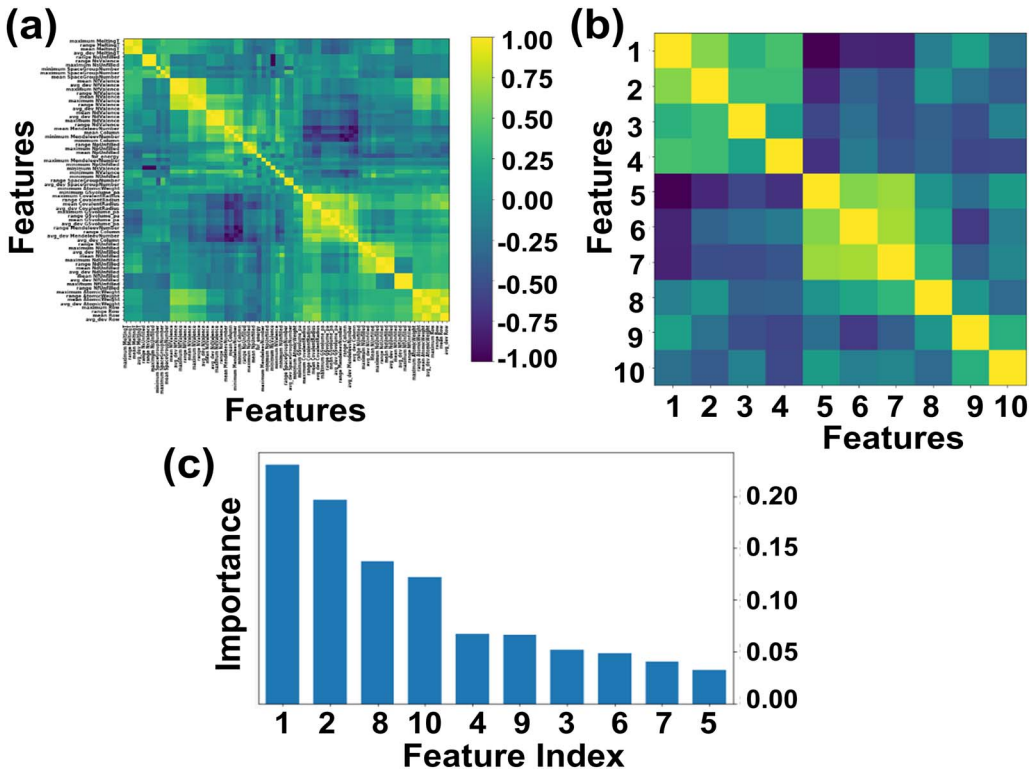

Fig. 2 (a and b) Heatmap representing the initial 117 and final top 10 features respectively of ABN_3 perovskites [yellow and dark blue colors in the heatmap represent the respective strong and weak correlations among the features]. (c) The relative importance of top 10 features of the compounds. [Feature indices (in the order of importance): 1-electronegativity, 2-d valence electrons, 3-formation energy, 4-p valence electrons, 5-Mendeleev number, 6-specific volume of ground state, 7-atomic weight, 8-mean covalent radius, 9-melting temperature and 10-space group number].

predict the same E_g within MAEs of 8%, 10% and 16% respectively. From Table 1, the RFR algorithm shows lowest RMSE (=0.11 eV) and highest R^2 (=0.94) values in contrast to other models. R^2 (RMSE) values of GBDT, SVR and MLP models are estimated to be 0.90 (0.15 eV), 0.91 (0.13 eV) and 0.74 (0.22 eV) respectively. These results collectively suggest the reliability of the RFR model in predicting the band gaps of ABN_3-type perovskite compounds. The bivariate plots from the RFR model showing both the predicted, input band gap values for the training and testing datasets are illustrated in Fig. 3(a). From Fig. 3(a), strong linear correlation [with $R^2 = 0.98$ (training set) and 0.94 (test set)] between the predicted and input E_g values have been observed in both the training and test datasets. This result further justifies the accuracy of RFR model in predicting the band gaps of ABN_3 perovskites. The corresponding bivariate plots between the predicted and input values of E_g, as accomplished from SVR, GBDT and MLP ML models, are also shown in Fig. 3(b–d) respectively. From Fig. 3(b–d), moderate to weak linear correlations between the predicted and input E_g values have been observed as the model goes from SVR, GBDT to MLP algorithms. The above results as a whole suggest the superiority of the RFR algorithm as an effective ML model is predicting the band gaps of ABN_3 compounds in contrast to SVR, GBDT and MLP algorithms.

Besides, two newly synthesized nitride perovskites $CeMoN_3$ and $CeWN_3$, as reported elsewhere[28] and whose band gaps are yet to be explored, have been selected for first-principle DFT calculations. Their electronic band structures, E_g values and optoelectronic properties have been studied from DFT calculations. The E_g values of the said systems have been predicted from both the DFT as well as RFR ML model to verify the

Table 1 Statistics of different ML models and their corresponding hyperparameters for predicting bandgaps of ABN_3 perovskites

Models	Hyperparameters	MAE	RMSE	R^2
MLP	Solver = adam, alpha = 1×10^{-8}, tol = 1×10^{-6}, max_iter = 5000, random_state = 0	0.16	0.22	0.74
GBDT	n_estimators = 2000, max_depth = 30, min_samples_split = 2	0.10	0.15	0.90
SVR	C = 50, gamma = 10, epsilon = 0.05, kernel = 'rbf'	0.08	0.13	0.91
RFR	max_depth = 30, min_samples_split = 2, n_estimators = 2000, min_samples_leaf = 1, random_state = 0	0.03	0.11	0.94

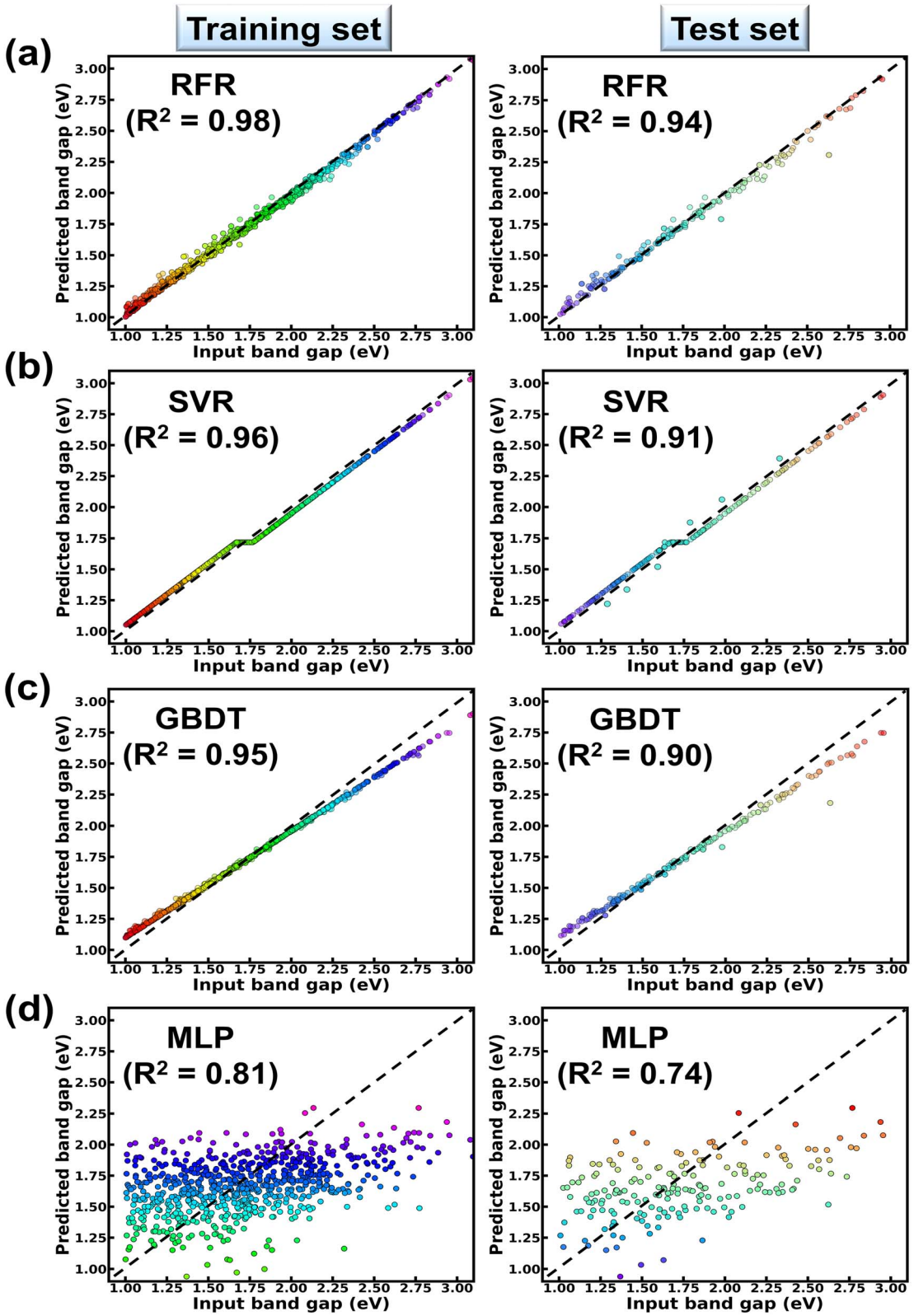

Fig. 3 Fitted bivariate plots showing the variations of predicted and input band gaps of ABN_3 perovskites for both the training and test datasets as obtained from (a) RFR (b) SVR (c) GBDT and (d) MLP models.

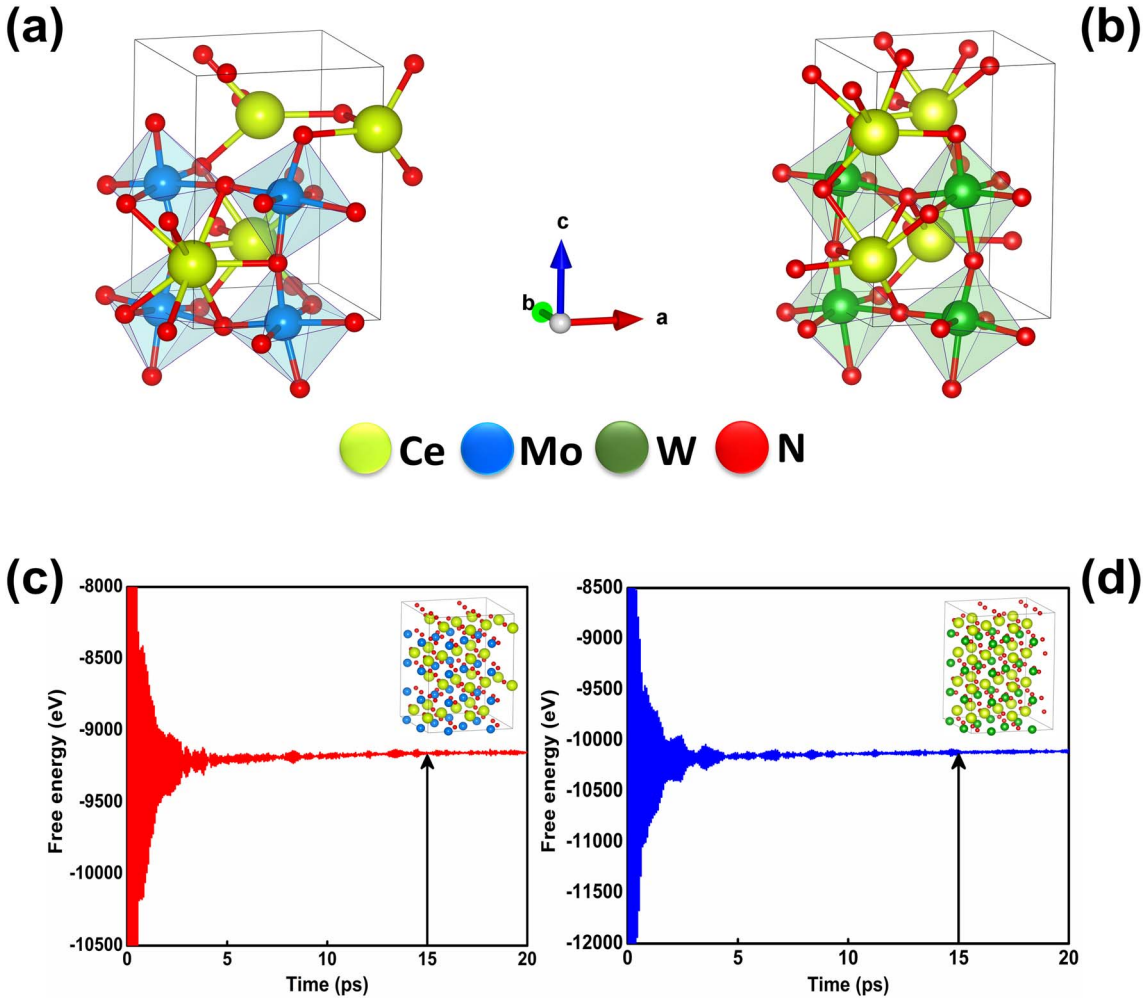

Fig. 4 Optimized crystal structures of (a) $CeMoN_3$ and (b) $CeWN_3$ compounds as obtained from DFT calculations. Free energy as a function of time for (c) $CeMoN_3$ and (d) $CeWN_3$ systems as attained from BOMD simulations. [insets of (c and d) show the respective supercell geometries of $CeMoN_3$ and $CeWN_3$ so obtained from the BOMD simulation run at 15 ps which is marked with vertical arrows].

accuracy of ML model in predicting E_g values of ABN_3 compounds.

3.3. Crystal structures and stabilities of the $CeBN_3$ (B = Mo, W) compounds

At room temperature ($T = 300$ K) and under ambient pressure ($P = 0$ GPa), $CeBN_3$ (B = Mo, W) crystalize to primitive orthorhombic phase and belong to $Pmc2_1$ space group symmetry with space group no. 26. The optimized unitcell geometries of the compounds, as obtained from first-principle DFT calculations with GGA-PBE level of theory, are shown in the upper panel of Fig. 4. The lattice parameters of the $CeMoN_3$ ($CeWN_3$) crystal system have been estimated to be $a = 5.789$ Å (5.686 Å), $b = 5.854$ Å (5.667 Å) and $c = 7.776$ Å (8.027 Å) which are in close agreement with the experimentally determined X-ray diffraction data as reported elsewhere.[28] From Fig. 4(a and b), it is observed that B (=Mo, N) and N atoms form several distorted octahedral environments within the crystal structure and B atom, which is located at the center of each BN_6 octahedron, is found to be shifted towards the apical N atom. The distorted octahedron, aka pyramidal coordination and large c/a ratio of $CeBN_3$ may result large spontaneous ferroelectric polarization in the systems under study.[86,87] The ferroelectric polarization values of $CeMoN_3$ and $CeWN_3$ compounds are estimated to be ~47.24 and 49.68 μC cm^{-2} respectively along the direction of crystallographic c-axis. To comprehend the thermal stabilities of the compounds at $T = 300$ K, the BOMD simulations have been performed over a time span of 20 ps under time step of 1 fs. Temporal variations of free energies of $CeBN_3$ systems at $T = 300$ K are shown in the lower panel of Fig. 4. The free energies (F) of the compounds have been estimated using the following relation:[88]

$$F(T,\nu) = E(\nu) + F_{el}(T,\nu) + F_{ph}(T,\nu) \quad (12)$$

where E measures the total ground state energy of the supercell, ν signifies the supercell volume at temperature T, $F_{el}(T,\nu)$ and $F_{ph}(T,\nu)$ are the free energies of the electrons and phonons respectively. $F_{el}(T,\nu)$ can be estimated using the following mathematical relation:[89,90]

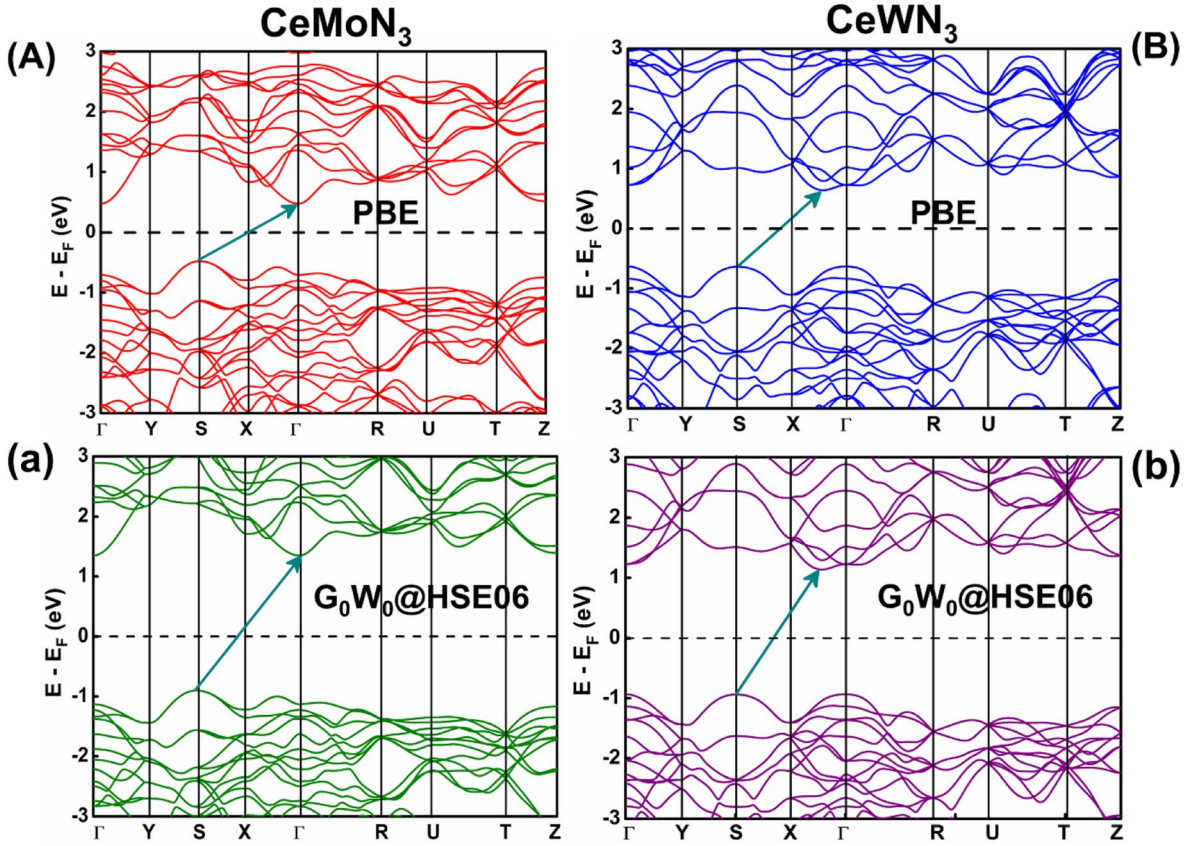

Fig. 5 E–k diagrams of $CeMoN_3$ (left panel) and $CeWN_3$ (right panel) compounds along Γ → Y → S → X → Γ → R → U → T → Z high-symmetry direction as obtained from PBE and G_0W_0@HSE06 level of theories. [Horizontal dashed line and inclined arrow represent Fermi energy level and band gap respectively].

$$F_{el}(T,\nu) = k_B T \int d\upsilon N_{el}(\upsilon)[n\ln n + (1-n)\ln(1-n)] \quad (13)$$

where $N_{el}(\nu)$ is the electronic density of states and n is the Fermi occupation number. $F_{ph}(T,\nu)$ has been calculated under the thermodynamic integration approach[90,91] and is expressed as:

$$F_{ph}(T,\nu) = \int_0^1 d\upsilon \left\langle F_{el}(\{R_I\}) - F_{el}(\{R_I{}^0\}) - \sum_{i,j}\frac{m}{2}u_i u_j D_{ij}(T,\nu) \right\rangle_\upsilon \quad (14)$$

where "$F_{el}(\{R_I\}) - F_{el}(\{R_I{}^0\})$" corresponds to the difference in electronic free energies at a certain point of ionic coordinates $\{R_I\}$. $u_i = \{R_i\} - \{R_i{}^0\}$, m is the atomic mass and D represents the dynamical matrix of the system. "$\langle \ldots\ldots \rangle_\nu$" represents the thermodynamic average and has been estimated using the BOMD simulations. From Fig. 4(c and d) it is found that both the crystal structures attain their respective stabilizing free energy values after a time lapse of ~5 ps. The refined geometries of the compounds, so obtained from the BOMD simulation run at 15 ps, are shown in the inset of Fig. 4(c and d).

3.4. Estimations of E_g values for $CeBN_3$ (B = Mo, W) compounds

The E–k diagrams of $CeBN_3$ systems for B = Mo, W along Γ → Y → S → X → Γ → R → U → T → Z high-symmetry direction have been estimated from DFT calculations at PBE, HSE06, G_0W_0@PBE and G_0W_0@HSE06 level of theories. The results are shown in Fig. 5 and S1 (ESI).† The estimated E_g values of the compounds from the corresponding E–k diagrams are shown in Table 2. Interestingly the E–k diagrams for both $CeMoN_3$ and $CeWN_3$ compounds show the existence of indirect band gaps ranging from 0.94–1.60 and 1.00–1.81 eV, respectively. Presence of indirect band gaps with gap openings in the range 0.94–1.81 eV may render their applications in photovoltaics.[92–96] In this connection it may be relevant to mention that while the PBE functional is known to underestimate the E_g values of the systems in general,[14,33–36,40,41,43] the HSE06 functional on the other hand can predict the E_g values of the compounds in close agreement with the experimental observations.[14,33,40,45,97] The

Table 2 Band gaps of $CeBN_3$ (B = Mo, W) compounds as obtained from DFT calculations at different level of theories

	Band gaps (eV)	
Level of theories	$CeMoN_3$	$CeWN_3$
PBE	0.94	1.00
HSE06	1.57	1.80
G_0W_0@PBE	1.46	1.71
G_0W_0@HSE06	1.60	1.81

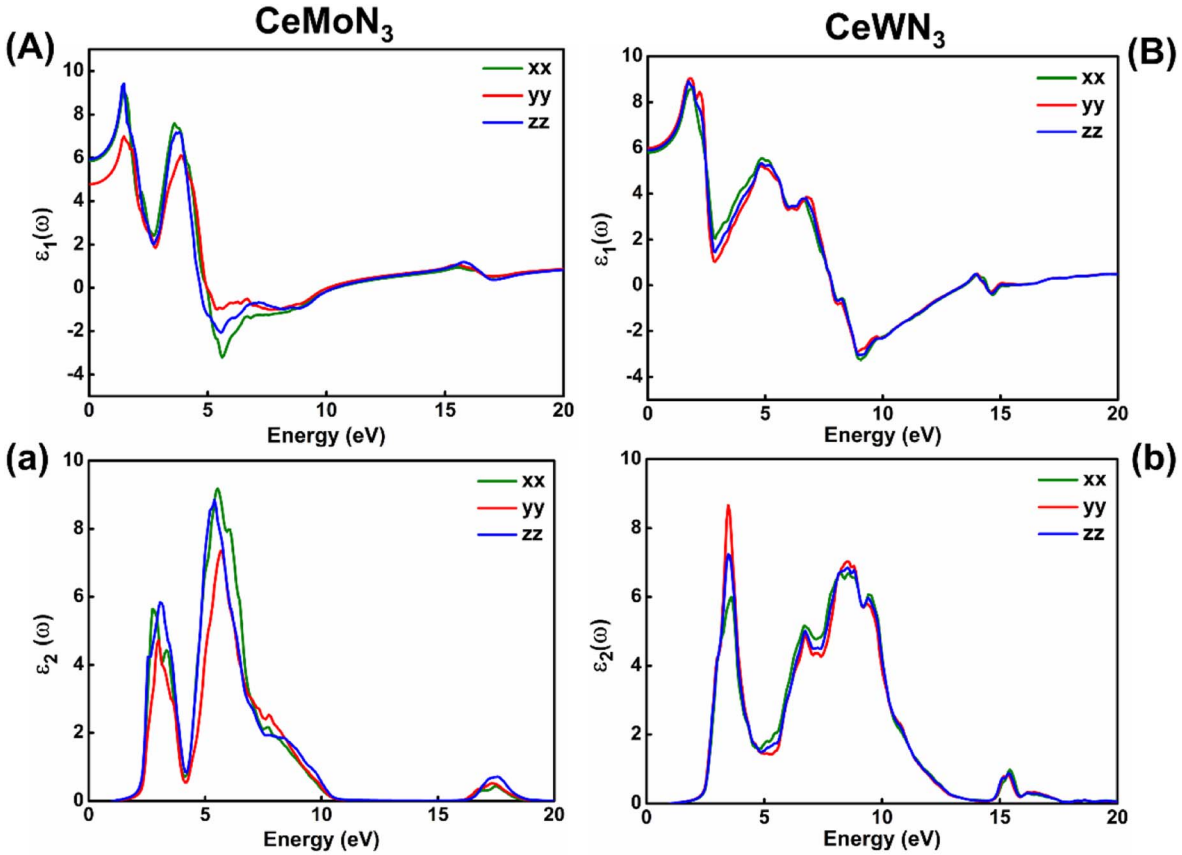

Fig. 6 Left panel: (A) Real and (a) imaginary parts of dielectric function *versus* incident EM wave energy for $CeMoN_3$ compound. Right panel: (B) Real and (b) imaginary parts of $\varepsilon(\omega)$ as a function of EM wave energy for $CeWN_3$ compound.

HSE06 XC energy functional (E_{XC}^{HSE06}) is mathematically expressed as:[98]

$$E_{XC}^{HSE06} = 1/4E_X^{HF,\ SR}(\omega) + 3/4E_X^{PBE,\ SR}(\omega) + E_X^{PBE,\ LR}(\omega) + E_C^{PBE} \quad (15)$$

where "X", "C", "SR" and "LR" symbolize the exchange, correlation, short range and long range terms respectively in the energy functional. "ω (=0.2)" governs the extent of short range interactions. The difference of E_g values between HSE06 and PBE functionals can be rationalized from the fact that the HSE06 functional, which consists of screened Coulomb potential for Hartree–Fock (HF) exchange, considers the exact mixing of both the HF and PBE exchange only for short-range interactions. This mixing allows the exchange hole to become delocalized among its nearest neighbors, which in turn results in E_g values which are reported to be closer to experimental observations in contrast to the same being performed with PBE functional.[98]

The E_g values of $CeMoN_3$ and $CeWN_3$ compounds are predicted to be ~1.55 and 1.76 eV from RFR ML model for the G_0W_0@HSE06 input of 1.60 and 1.81 eV respectively. While the DFT calculations at HSE06 and G_0W_0@HSE06 level of theories can closely predict the experimentally observed E_g values of the systems within <1% error,[40,42,45,48,97,99,100] the predicted E_g values of $CeBN_3$ (B = Mo, W) compounds as obtained from RFR ML model are estimated within an error of ~3% with standard deviation of ~0.035. The selective choice of G_0W_0@HSE06 level of theory in the DFT calculation as input for the prediction of band gaps is thus meaningful, given its efficacy in reproducing the experimental band gaps of many such compounds as reported elsewhere.[14,101–103]

3.5. Optical properties of $CeBN_3$ (B = Mo, W) systems

Electronic band structures are intrinsically linked with the optoelectronic properties of the systems in general. The optoelectronic properties of $CeBN_3$ (B = Mo, W) crystal systems have been estimated from the frequency (ω) dependent complex dielectric function [$\varepsilon(\omega)$]. Since, the $Pmc2_1$ phase of the compounds is structurally anisotropic, the incident electromagnetic (EM) field components along the direction of *a*, *b* and *c*-crystallographic axes are designated as *xx*, *yy* and *zz*, respectively in the consequent calculations. Fig. 6(A) and (B) show the variations of real part of $\varepsilon(\omega)$ [$\varepsilon_1(\omega)$] as a function of incident EM wave energy ($\tilde{E}_M$), so obtained from G_0W_0@HSE06 level of theory, for $CeMoN_3$ and $CeWN_3$ compounds respectively. From the upper panel of Fig. 6, static values of $\varepsilon_1(\omega)$ [$\varepsilon_1(0)$] have been estimated to be ~5.85 (5.80), 4.77 (5.99) and 5.93 (5.90) for $CeMoN_3$ ($CeWN_3$) system along xx, yy and zz polarization directions respectively of the incident EM wave. These non-zero

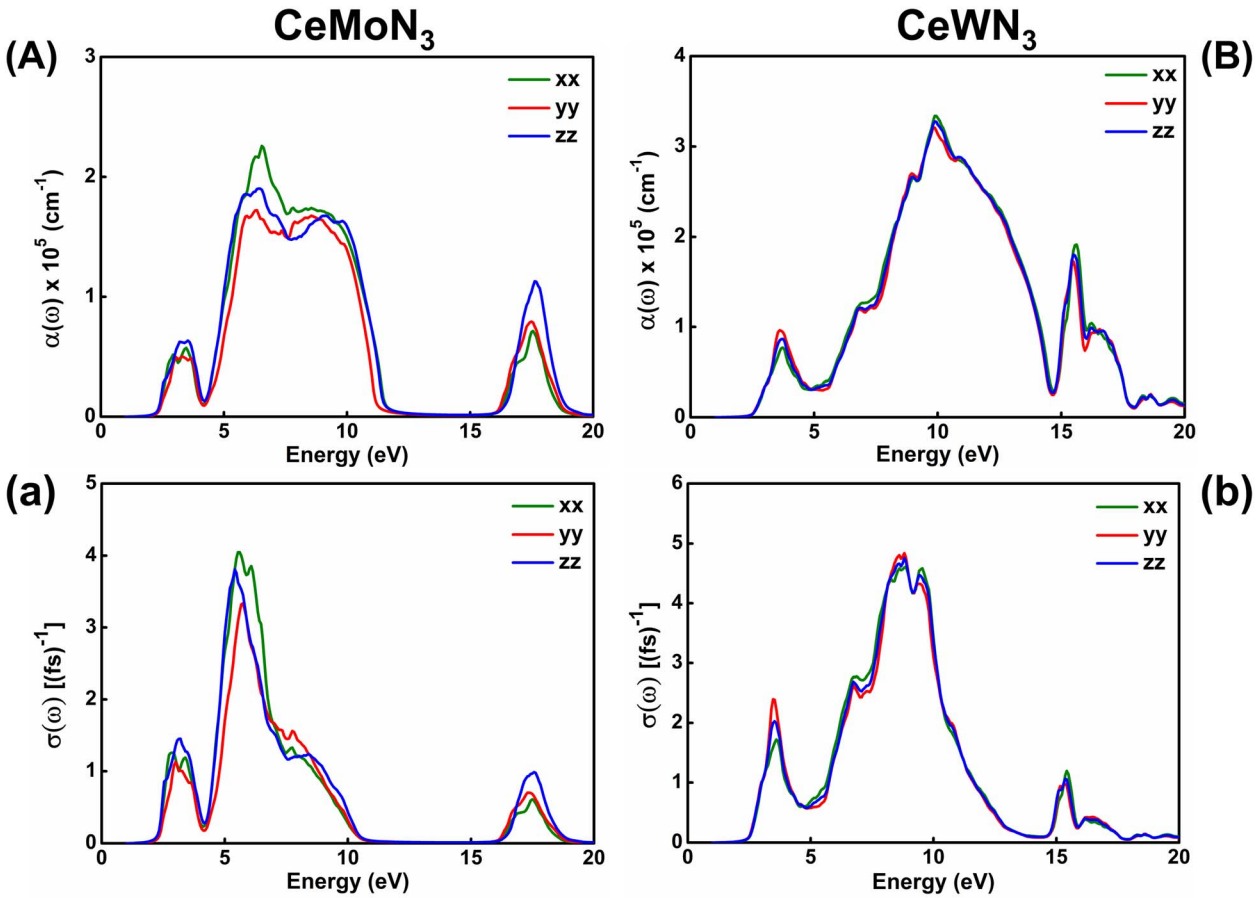

Fig. 7 Left panel: (A) Optical absorption and (a) conductivity as a function of incident EM wave energy for $CeMoN_3$ compound. Right panel: (B) Optical absorption and (b) conductivity as a function of EM wave energy for $CeWN_3$ compound.

values of $\varepsilon_1(0)$ for all the three polarization directions indicate their semiconducting nature so attained from the E–k diagrams [*vide ante*, Fig. 5]. The most intense structure band of $\varepsilon_1(\omega)$ for $CeMoN_3$ ($CeWN_3$) compound is obtained at $\tilde{E}_M$ ~1.45 (1.85) eV which corresponds to maximum dispersion of the EM wave. However, beyond $\tilde{E}_M$ ~3.75 (6.86) eV, $\varepsilon_1(\omega)$ values of $CeMoN_3$ ($CeWN_3$) system gradually decrease and attain negative values within the energy window $\tilde{E}_M$ ~4.72 (7.86)–10.42 (13.39) eV. The negative values of $\varepsilon_1(\omega)$ signify metallic Drude – like behaviour of the compounds within the said energy windows.[42,43,104]

$\varepsilon_2(\omega)$ is known to correlate directly with the band gaps of the compounds in general. Fig. 6(a) and (b) depict the variations of $\varepsilon_2(\omega)$ as a function of $\tilde{E}_M$ for $CeMoN_3$ and $CeWN_3$ systems respectively, so attained from G_0W_0@HSE06 level of theory. From Fig. 6(a and b) the critical or take-off values of $\varepsilon_2(\omega)$ have been observed at $\tilde{E}_M$ ~1.60 (1.81) eV for $CeMoN_3$ ($CeWN_3$) compound for all the three polarization directions. The critical or take-off values of $\varepsilon_2(\omega)$ are in exact agreement with the E_g values of the systems, as obtained from the G_0W_0@HSE06 level of theory (*vide ante*, Table 2). Interestingly, the take-off values of $\varepsilon_2(\omega)$ are found to be independent of polarization directions of EM wave. Moreover, finite values of $\varepsilon_2(\omega)$ between $\tilde{E}_M$ ~2.30 (2.40) and 10.01 (12.50) eV for $CeMoN_3$ ($CeWN_3$) crystal corroborate strong attenuation of the incident EM wave inside the crystal systems within the said energy window.[42,43]

The response of strong attenuation of the incident EM wave inside the $CeBN_3$ (B = Mo, W) compounds, so reflected in the $\varepsilon_2(\omega)$–$\tilde{E}_M$ plots (Fig. 6), can also be attributed from their absorption coefficients [$\alpha(\omega)$] and optical conductivities [$\sigma(\omega)$]. Fig. 7 shows the variations of $\alpha(\omega)$ and $\sigma(\omega)$ as function of $\tilde{E}_M$ for the referred systems. Akin to $\varepsilon_2(\omega)$–$\tilde{E}_M$ plots, the critical values of both $\alpha(\omega)$ and $\sigma(\omega)$ are found at $\tilde{E}_M$ ~1.60 (1.81) eV for $CeMoN_3$ ($CeWN_3$) crystal system for all the three polarization directions. Hence the estimations of $\alpha(\omega)$ and $\sigma(\omega)$ help to recheck the E_g values as obtained from electronic band structures and $\varepsilon_2(\omega)$–$\tilde{E}_M$ plots. From Fig. 7, finite values of $\alpha(\omega)/\sigma(\omega)$ have been observed within $\tilde{E}_M$ ~2.30/2.30 to 11.43/10.01 eV, 16.47/16.55 to 18.77/18.64 eV and 2.40/2.40 to 17.90/12.50 eV for $CeMoN_3$ and $CeWN_3$ compounds, respectively. The finite values of $\alpha(\omega)$ and/or $\sigma(\omega)$ within the specified energy window signify large opacity and high optical conductivity of the said systems.[42,43] Moreover, the high transparency of $CeMoN_3$ ($CeWN_3$) compound within the energy window ~0–1.60 eV (0–1.81 eV) has also been noticed in $\alpha(\omega)$–$\tilde{E}_M$ and $\sigma(\omega)$–$\tilde{E}_M$ plots (Fig. 7). The high transparency of the said systems covering visible and Infrared (IR) regions of the EM wave may promote them as potential candidates for optical luminescent materials and IR detectors.[43,105,106]

4. Conclusions

The band gaps of ABN_3 perovskites have been predicted from ML models based on the feature descriptors. The dataset of 1563 ABN_3 perovskites, which show $E_f < -0.026$ eV and E_g values spanning in the range from 1.0 to 3.1 eV, have been selected from initial 5566 samples for ML model evaluation. The E_g values of the compounds are found to be linked with 117 feature descriptors. The top 10 feature descriptors have been selected by eliminating those features which show absolute correlation values <0.89 and low rank of importance. Correlations between the band gap and the topmost 3 features such as electronegativity, mean covalent radius and d valence electrons have also been discussed. Four supervised ML models such as MLP, GBDT, SVR and RFR have been considered to predict the band gaps of these nitride perovskite compounds. The accuracy of each model has been tested from MAE, RMSE and R^2 values. The RFR algorithm shows lowest RMSE (=0.11 eV) and highest R^2 (=0.94) values in comparison to other models. R^2 values and the corresponding bivariate plots suggest the superiority of RFR algorithm as an effective ML model is predicting the band gaps of ABN_3 perovskites in contrast to SVR, GBDT and MLP algorithms. Two newly synthesized ABN_3-type perovskites $CeBN_3$ (B = Mo, W) have been further chosen and their electronic band structures and optoelectronic properties have been studied from DFT calculations. The E_g values of $CeBN_3$ (B = Mo, W) compounds have been estimated from DFT calculations at PBE, HSE06, G_0W_0@PBE and G_0W_0@HSE06 level of theories. The E_g values of $CeMoN_3$ and $CeWN_3$ compounds have been predicted to be 1.55 and 1.76 eV from RFR algorithm for the G_0W_0@HSE06 input of 1.60 and 1.81 eV respectively. The E_g values of the systems have also been verified from their respective optoelectronic parameters such as $\varepsilon_2(\omega)$, $\alpha(\omega)$ and $\sigma(\omega)$. We believe that this study will help to unveil the efficacy of ML models in predicting the band gaps of nitride perovskites which in turn will bear potential applications as photovoltaic and optical luminescent materials.

Data availability

The data and codes that support the findings of this study are available on reasonable request from corresponding author.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

Authors would like to thank the National Supercomputing Mission (NSM) for providing computing resources of 'PARAM Kamrupa' at IIT Guwahati, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India. Swarup Ghosh sincerely acknowledges the University Grants Commission (UGC), Government of India for providing the senior research fellowship (NET) award.

References

1 Z. Shi and A. H. Jayatissa, *Materials*, 2018, **11**, 729.
2 P. Tonui, S. O. Oseni, G. Sharma, Q. Yan and G. Tessema Mola, *Renewable Sustainable Energy Rev.*, 2018, **91**, 1025–1044.
3 M. Aldamasy, Z. Iqbal, G. Li, J. Pascual, F. Alharthi, A. Abate and M. Li, *Phys. Chem. Chem. Phys.*, 2021, **23**, 23413–23427.
4 B. V. Politov, E. P. Antonova, E. S. Tropin, D. A. Osinkin, A. Y. Suntsov and V. L. Kozhevnikov, *Renewable Energy*, 2023, **206**, 872–878.
5 Y. Zhou, X. Guan, H. Zhou, K. Ramadoss, S. Adam, H. Liu, S. Lee, J. Shi, M. Tsuchiya, D. D. Fong and S. Ramanathan, *Nature*, 2016, **534**, 231–234.
6 A. Raj, S. Sharma, D. V. Singh, A. Kumar, R. K. Chourasia, J. M. Siqueiros, O. R. Herrera, A. Anshul and M. Kumar, *Phys. B*, 2024, **673**, 415504.
7 L. Polavarapu, M. A. Loi, H. Zeng and J. M. Luther, *Nanoscale*, 2023, **15**, 15075–15078.
8 A. Balilonda, Z. Li, Y. Fu, F. Zabihi, S. Yang, X. Huang, X. Tao and W. Chen, *J. Mater. Chem. C*, 2022, **10**, 6957–6991.
9 R. Wang, J. Wang, S. Tan, Y. Duan, Z.-K. Wang and Y. Yang, *Trends Chem.*, 2019, **1**, 368–379.
10 D. Luo, X. Li, A. Dumont, H. Yu and Z.-H. Lu, *Adv. Mater.*, 2021, **33**, 2006004.
11 H. C. Koo, S. B. Kim, H. Kim, T.-E. Park, J. W. Choi, K.-W. Kim, G. Go, J. H. Oh, D.-K. Lee, E.-S. Park, I.-S. Hong and K.-J. Lee, *Adv. Mater.*, 2020, **32**, 2002117.
12 M. Kepenekian and J. Even, *J. Phys. Chem. Lett.*, 2017, **8**, 3362–3370.
13 M.-Y. Liu, Q.-Y. Chen, C. Cao and Y. He, *Phys. Chem. Chem. Phys.*, 2019, **21**, 2899–2909.
14 S. Ghosh and J. Chowdhury, *Mod. Phys. Lett. B*, 2023, **38**, 2330003.
15 C. Gui, J. Chen and S. Dong, *Phys. Rev. B*, 2022, **106**, 184418.
16 J. A. Flores-Livas, R. Sarmiento-Pérez, S. Botti, S. Goedecker and M. A. L. Marques, *JPhys Mater.*, 2019, **2**, 025003.
17 B. F. Grosso, D. W. Davies, B. Zhu, A. Walsh and D. O. Scanlon, *Chem. Sci.*, 2023, **14**, 9175–9185.
18 V.-A. Ha, H. Lee and F. Giustino, *Chem. Mater.*, 2022, **34**, 2107–2122.
19 R. Sarmiento-Pérez, T. F. T. Cerqueira, S. Körbel, S. Botti and M. A. L. Marques, *Chem. Mater.*, 2015, **27**, 5957–5963.
20 T. K. Ng, J. A. Holguin-Lerma, C. H. Kang, I. Ashry, H. Zhang, G. Bucci and B. S. Ooi, *J. Phys. D: Appl. Phys.*, 2021, **54**, 143001.
21 M.-C. Jung, K.-W. Lee and W. E. Pickett, *Phys. Rev. B*, 2018, **97**, 121104.
22 M. Kim, E. Pallecchi, R. Ge, X. Wu, G. Ducournau, J. C. Lee, H. Happy and D. Akinwande, *Nat. Electron.*, 2020, **3**, 479–485.
23 I. Ashraf, S. Rizwan and M. Iqbal, *Front. Mater.*, 2020, **7**, 181.
24 W. Lengauer, *J. Phys. Chem. Solids*, 1988, **49**, 59–63.
25 Y. M. Shy, L. E. Toth and R. Somasundaram, *J. Appl. Phys.*, 2003, **44**, 5539–5545.

26 S. P. Chockalingam, M. Chand, J. Jesudasan, V. Tripathi and P. Raychaudhuri, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2008, **77**, 214503.
27 V. M. Pan, V. G. Prokhorov, V. A. Komashko, G. G. Kaminsky, M. A. Kousenetsov and C. G. Tretiatchenko, *IEEE Trans. Magn.*, 1989, **25**, 2000–2003.
28 R. Sherbondy, R. W. Smaha, C. J. Bartel, M. E. Holtz, K. R. Talley, B. Levy-Wendt, C. L. Perkins, S. Eley, A. Zakutayev and G. L. Brennecka, *Chem. Mater.*, 2022, **34**, 6883–6893.
29 N. E. Brese and F. J. DiSalvo, *J. Solid State Chem.*, 1995, **120**, 378–380.
30 K. R. Talley, C. L. Perkins, D. R. Diercks, G. L. Brennecka and A. Zakutayev, *Science*, 2021, **374**, 1488–1491.
31 S. D. Kloß, M. L. Weidemann and J. P. Attfield, *Angew. Chem., Int. Ed.*, 2021, **60**, 22260–22264.
32 S. Bandyopadhyay, A. Paul and I. Dasgupta, *Phys. Rev. B*, 2020, **101**, 014109.
33 S. Ghosh, R. R. Banik and J. Chowdhury, *Indian J. Theor. Phys.*, 2021, **69**, 43–68.
34 V. I. Anisimov, J. Zaanen and O. K. Andersen, *Phys. Rev. B*, 1991, **44**, 943–954.
35 H. Xiao, J. Tahir-Kheli and W. A. Goddard III, *J. Phys. Chem. Lett.*, 2011, **2**, 212–217.
36 I. N. Yakovkin and P. A. Dowben, *Surf. Rev. Lett.*, 2007, **14**, 481–487.
37 D. S. Lambert and D. D. O'Regan, *Phys. Rev. Res.*, 2023, **5**, 013160.
38 S. Tomić and N. M. Harrison, *AIP Conf. Proc.*, 2010, **1199**, 65–66.
39 D. Wing, G. Ohad, J. B. Haber, M. R. Filip, S. E. Gant, J. B. Neaton and L. Kronik, *Proc. Natl. Acad. Sci. U.S.A.*, 2021, **118**, e2104556118.
40 S. Ghosh, S. Sarkar and J. Chowdhury, *Mater. Chem. Phys.*, 2022, **276**, 125379.
41 R. Ray Banik, S. Ghosh and J. Chowdhury, *Phys. Scr.*, 2023, **98**, 045920.
42 S. Ghosh and J. Chowdhury, *Mater. Sci. Eng., B*, 2022, **284**, 115903.
43 R. R. Banik, S. Ghosh and J. Chowdhury, *Phys. Scr.*, 2023, **98**, 105914.
44 S. Śmiga and L. A. Constantin, *J. Phys. Chem. A*, 2020, **124**, 5606–5614.
45 A. J. Garza and G. E. Scuseria, *J. Phys. Chem. Lett.*, 2016, **7**, 4165–4170.
46 H. Einollahzadeh, R. S. Dariani and S. M. Fazeli, *Solid State Commun.*, 2016, **229**, 1–4.
47 J. M. Crowley, J. Tahir-Kheli and W. A. Goddard III, *J. Phys. Chem. Lett.*, 2016, **7**, 1198–1203.
48 Y. Duan, L. Qin, L. Shi and G. Tang, *Comput. Mater. Sci.*, 2015, **101**, 56–61.
49 S. Chowdhury, S. Ghosal, D. Mondal and D. Jana, *J. Phys. Chem. Solids*, 2022, **170**, 110909.
50 E. Bedolla, L. C. Padierna and R. Castañeda-Priego, *J. Phys.: Condens. Matter*, 2021, **33**, 053001.
51 S. Chibani and F.-X. Coudert, *APL Mater.*, 2020, **8**, 080701.
52 C. Gao, X. Yang, M. Jiang, L. Chen, Z. Chen and C. V. Singh, *Phys. Chem. Chem. Phys.*, 2022, **24**, 4653–4665.
53 Y. Zhang, W. Xu, G. Liu, Z. Zhang, J. Zhu and M. Li, *PLoS One*, 2021, **16**, e0255637.
54 A. Halder, A. Ghosh and T. S. Dasgupta, *Phys. Rev. Mater.*, 2019, **3**, 084418.
55 Y. Zhuo, A. Mansouri Tehrani and J. Brgoch, *J. Phys. Chem. Lett.*, 2018, **9**, 1668–1673.
56 Y. Tang, H. Chen, J. Wang and X. Niu, *Phys. Chem. Chem. Phys.*, 2023, **25**, 18086–18094.
57 M. Gao, B. Cai, G. Liu, L. Xu, S. Zhang and H. Zeng, *Phys. Chem. Chem. Phys.*, 2023, **25**, 9123–9130.
58 H. J. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. L. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, A. Smolyanyuk, S. Curtarolo, A. Tkatchenko, A. P. Bartók, S. Manzhos, M. Ihara, T. Carrington, J. Behler, O. Isayev, M. Veit, A. Grisafi, J. Nigam, M. Ceriotti, K. T. Schütt, J. Westermayr, M. Gastegger, R. J. Maurer, B. Kalita, K. Burke, R. Nagai, R. Akashi, O. Sugino, J. Hermann, F. Noé, S. Pilati, C. Draxl, M. Kuban, S. Rigamonti, M. Scheidgen, M. Esters, D. Hicks, C. Toher, P. V. Balachandran, I. Tamblyn, S. Whitelam, C. Bellinger and L. M. Ghiringhelli, *Electron. Struct.*, 2022, **4**, 023004.
59 O. Allam, B. W. Cho, K. C. Kim and S. S. Jang, *RSC Adv.*, 2018, **8**, 39414–39420.
60 F. Wang, Z. Yang, F. Li, J.-L. Shao and L.-C. Xu, *RSC Adv.*, 2023, **13**, 31728–31737.
61 V. Gladkikh, D. Y. Kim, A. Hajibabaei, A. Jana, C. W. Myung and K. S. Kim, *J. Phys. Chem. C*, 2020, **124**, 8905–8918.
62 G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis and T. Lookman, *Sci. Rep.*, 2016, **6**, 19375.
63 A. Talapatra, B. P. Uberuaga, C. R. Stanek and G. Pilania, *Commun. Mater.*, 2023, **4**, 46.
64 V. Vakharia, I. E. Castelli, K. Bhavsar and A. Solanki, *Phys. Lett. A*, 2022, **422**, 127800.
65 S. Priyanga G, M. N. Mattur, N. Nagappan, S. Rath and T. Thomas, *J. Mater.*, 2022, **8**, 937–948.
66 Z. Wan, Q.-D. Wang, D. Liu and J. Liang, *New J. Chem.*, 2021, **45**, 9427–9433.
67 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, *Mach. Learn. Res.*, 2011, **12**, 2825–2830.
68 J. Hauke and T. Kossowski, *Quaest. Geogr.*, 2011, **30**, 87–93.
69 P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru and S. Baroni, *J. Chem. Phys.*, 2020, **152**, 154105.
70 P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia,

M. Kawamura, H. Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H. V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu and S. Baroni, *J. Phys.: Condens. Matter*, 2017, **29**, 465901.

71 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari and R. M. Wentzcovitch, *J. Phys.: Condens. Matter*, 2009, **21**, 395502.

72 D. F. Shanno, *Math. Comput.*, 1970, **24**, 647–656.

73 D. Goldfarb, *Math. Comput.*, 1970, **24**, 23–26.

74 R. Fletcher, *Comput. J.*, 1970, **13**, 317–322.

75 C. G. Broyden, *IMA J. Appl. Math.*, 1970, **6**, 76–90.

76 A. Dal Corso, *Comput. Mater. Sci.*, 2014, **95**, 337–350.

77 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys. Rev. Lett.*, 2008, **100**, 136406.

78 A. Benassi, *PWSCF 's epsilon.x user's manual*, **https://www.researchgate.net/profile/Nowzar-Soltani/post/Can_anyone_answer_this_question_on_the_optical_properties_of_the_Epsilon_calculations/attachment/59d623046cda7b8083a1d75b/AS%3A310495943299072%401451039409313/download/eps_man.pdf**.

79 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.

80 J. Hu, S. Stefanov, Y. Song, S. S. Omee, S.-Y. Louis, E. M. D. Siriwardane, Y. Zhao and L. Wei, *npj Comput. Mater.*, 2022, **8**, 65.

81 L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, *npj Comput. Mater.*, 2016, **2**, 16028.

82 K. Li, Y. Li and D. Xue, *Funct. Mater. Lett.*, 2011, **04**, 217–219.

83 A.-S. Om Kumar, V. Shukla and S. K. Srivastava, *J. Sci.: Adv. Mater. Devices*, 2019, **4**, 158–162.

84 K. Li, C. Kang and D. Xue, *Mater. Res. Bull.*, 2012, **47**, 2902–2905.

85 M. Rahm, P. Erhart and R. Cammi, *Chem. Sci.*, 2021, **12**, 2397–2403.

86 S. Bandyopadhyay and I. Dasgupta, *Phys. Rev. B*, 2021, **103**, 014105.

87 T. Ahmad, K. Jindal, M. Tomar and P. K. Jha, *Phys. Chem. Chem. Phys.*, 2023, **25**, 5857–5868.

88 S. Ghosh and J. Chowdhury, *Phase Transitions*, 2023, **96**, 446–463.

89 R. Phillips, *Crystals, Defects and Microstructures: Modeling across Scales*, Cambridge University Press, Cambridge, 2001.

90 B. Grabowski, P. Söderlind, T. Hickel and J. Neugebauer, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 214107.

91 B. Grabowski, T. Hickel and J. Neugebauer, *Phys. Status Solidi B*, 2011, **248**, 1295–1308.

92 J. Kangsabanik, M. K. Svendsen, A. Taghizadeh, A. Crovetto and K. S. Thygesen, *J. Am. Chem. Soc.*, 2022, **144**, 19872–19883.

93 S. Halder, R. A. Kumar, R. Maity and T. P. Sinha, *Ceram. Int.*, 2023, **49**, 8634–8645.

94 O. I. Malyi and C. M. Acosta, *J. Phys. Chem. C*, 2020, **124**, 14432–14438.

95 R. Anbarasan, J. K. Sundar, M. Srinivasan and P. Ramasamy, *Comput. Condens. Matter*, 2021, **28**, e00581.

96 Y. Kang, Y. Youn, S. Han, J. Park and C.-S. Oh, *Chem. Mater.*, 2019, **31**, 4072–4080.

97 R. R. Pela, M. Marques and L. K. Teles, *J. Phys.: Condens. Matter*, 2015, **27**, 505502.

98 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.

99 P. Borlido, J. Schmidt, A. W. Huran, F. Tran, M. A. L. Marques and S. Botti, *npj Comput. Mater.*, 2020, **6**, 96.

100 S. Matsuishi, D. Iwasaki and H. Hosono, *J. Solid State Chem.*, 2022, **315**, 123508.

101 P. Bhumla, D. Gill, S. Sheoran and S. Bhattacharya, *J. Phys. Chem. Lett.*, 2021, **12**, 9539–9546.

102 K. Demmouche and J. Coutinho, *Int. J. Mod. Phys. B*, 2018, **32**, 1850328.

103 Y. Kang, G. Kang, H.-H. Nahm, S.-H. Cho, Y. S. Park and S. Han, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 165130.

104 M. I. Naher and S. H. Naqib, *Sci. Rep.*, 2021, **11**, 5592.

105 R. Reisfeld, *Opt. Mater.*, 2010, **32**, 850–856.

106 L. Su, X. Fan, T. Yin, H. Wang, Y. Li, F. Liu, J. Li, H. Zhang and H. Xie, *Adv. Opt. Mater.*, 2020, **8**, 1900978.