

## Prediction of optical band gap of $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> using material informatics

Edward Swinnich, Yash Jayeshbhai Dave, E. Bruce Pitman, Scott Broderick,  
Baishakhi Mazumder, Jung-Hun Seo\*



Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY, 14260, USA

### ARTICLE INFO

**Keywords:**  
 $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub>  
 Optical bandgap  
 Electronic bandgap  
 Material informatics

### ABSTRACT

In this study, the optical band gap of  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> versus the Al composition x is predicted using principal component regression and a Gaussian stochastic process. Properties were sourced from other mature Al-alloyed compound semiconductors to form a band gap model. It is found that the electronic band gap, the thermal conductivity, and the Al composition have the greatest influences on the optical band gap. A final relation is generated from a hybrid informatics approach combining information gained from multiple models. The optical band gap of  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> versus the Al composition is predicted and agrees well with measured optical band gap.

### 1. Introduction

Wide band gap semiconductors have been intensively researched in recent years toward applications in power electronics and ultraviolet (UV) optoelectronics [1]. Traditionally, wide band gap semiconductors such as SiC and GaN, as well as compound semiconductors such as Al<sub>x</sub>Ga<sub>1-x</sub>N have been investigated for these applications [1]. However, these wide band gap semiconductors are expected to face their limitations in future applications that need a higher breakdown voltage or specific detection wavelength deeper into the UV spectrum. Therefore, the development of another class of wide bandgap semiconductors are urgently needed, in order to fulfill upcoming challenges and offer diverse technological solutions in wide range of future applications including electric vehicles, wireless communication systems, power switches, delicate and durable sensors for military and space applications [2].  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> which is based on  $\beta$ -Ga<sub>2</sub>O<sub>3</sub> and alloyed with Al, has the potential to become the next generation of wide band gap semiconductor and fill this need for a new material. With a breakdown field more than three times larger than current GaN or SiC devices [1], a high high Baliga's figure of merit of 3444 ( $\epsilon\mu E_c$  [3]) which is much higher than GaN and SiC [3], as well as stable operating temperatures of up to 250 °C [4],  $\beta$ -Ga<sub>2</sub>O<sub>3</sub> has several advantages over conventional power semiconductors. Alloying with Al allows the band gap to be tuned from 4.8 eV ( $\beta$ -Ga<sub>2</sub>O<sub>3</sub>) [5] to 8.7 eV ( $\alpha$ -Al<sub>2</sub>O<sub>3</sub>) [6],  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> can be used for deep UV applications that greatly eclipse the capabilities of SiC, GaN and even  $\beta$ -Ga<sub>2</sub>O<sub>3</sub> materials [1].  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> can also be fabricated via multiple methods such as chemical vapor deposition (CVD) [7], pulsed laser deposition (PLD) [8],

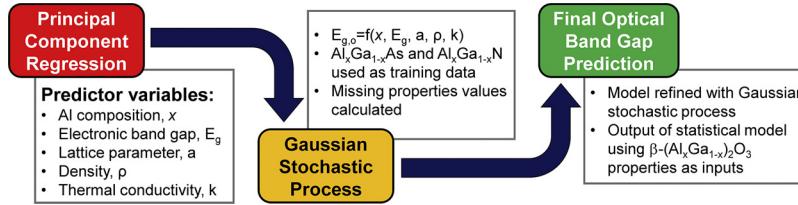
molecular beam epitaxy (MBE) [9], and solution combustion synthesis (SCS) [1], which allows the material to be processes for various applications and with different crystallinities and phases dependent on the fabrication method, allowing for further tuning of the band gap [1].

Although  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> is a promising wide band gap semiconductor, the  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> development is still in an early stage of development and many material properties such as optical, thermal, and mechanical properties have not been fully characterized. Particularly, to fully utilize the ultra-wide band gap property of  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> toward UV optoelectronics, such as solar blind UV detection and UV photon generation, it is still necessary to determine how the optical band gap relates to the Al composition. The optical band gap is the energy level at which photons can be absorbed, which may be different from the electronic band gap. At the optical band gap, the electron-hole pair is not completely separated when a photon is absorbed, whereas at the electronic band gap, the electron and hole are separated which allows transport to occur. It is necessary to know the optical band gap in order to design optoelectronic devices. While the electronic band gap of  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> with a small Al molar fraction is partially studied, its optical band gap is currently unknown due to technical difficulties in growing crystalline  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> with various Al compositions.

In this study, the optical band gap of  $\beta$ -(Al<sub>x</sub>Ga<sub>1-x</sub>)<sub>2</sub>O<sub>3</sub> is predicted using statistical methods which are able to capture the underlying physics driving similar semiconductors (particularly, Al<sub>x</sub>Ga<sub>1-x</sub>N and Al<sub>x</sub>Ga<sub>1-x</sub>As) and allow us to efficiently and robustly extend this physics to unexplored compounds. These experimental values of properties and their respective Al compositions were used as the training data in the

\* Corresponding author.

E-mail address: [junghuns@buffalo.edu](mailto:junghuns@buffalo.edu) (J.-H. Seo).



statistical methods such as principal component regression, a Gaussian stochastic process and linear regression. From these models, a relationship between the optical band gap and the Al composition  $x$  in  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  were generated and then used to predict the optical band gap over the possible range of Al compositions. Fig. 1 illustrates the steps used to construct the optical band gap model and make the final prediction. This hybrid material informatics approach is used to predict unknown materials properties. While similar logic has been successfully applied to properties of III-V semiconductor materials such as  $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{Sb}_{1-y}$  [10,11] and  $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{Sb}_z\text{As}_{1-y-z}$  [12], and various chalcopyrite semiconductors such as  $\text{CdGaAs}_2$  and  $\text{ZnGaAs}_2$  [13,14], much work had been done on these materials. Similar techniques have also been applied in related areas such as materials synthesis [15]. In contrast, the optical band gap of  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ , which has not been significantly studied, can be predicted with high accuracy.

## 2. Data analysis and statistical models

Fig. 2 shows both electronic and optical band gaps versus Al composition  $x$  of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  [16–18] and  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  [19,20], as well as preliminary data for the electronic band gap of  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  [1,8] and the optical band gap of  $\alpha\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  [7] from various literature. These data points were used as the input training data. It should be noted that  $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  undergoes a transition from the  $\beta$ -phase to the  $\alpha$ -phase at Al compositions greater than 0.68–0.8 [1], depending on the synthesis method and crystallinity of the film, which is undesirable as the phase should remain constant for all Al compositions to facilitate device fabrication. This transition is due to the preference of Al for the tetragonal site over the octahedral site [1]. The Al composition stability limit of  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  is still not fully understood, thus we assume Al composition of  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  covers 0% to 100% in this study. The electronic band gap of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  also undergoes a transition from direct to indirect above  $x = 0.45$  [18], and as such the band gap data is not shown for Al compositions above 0.45. Upon analysis of the existing data for each material shown in Fig. 2, it is clear that linear trends exist for both the electronic and optical band gaps versus the Al composition. As previously mentioned, the band gap of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  transitions into an indirect band gap above Al composition equal to 0.45 [18], and accordingly only the direct band gap region was taken into account due to the applications considered for these materials, which preferred the direct band gap (Fig. 2(a)). The most obvious difference between the

reference materials is that for  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  the optical band gap is equal to the electronic band gap, while for  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  the optical band gap is greater than the electronic band gap (Fig. 2(b)). As noted previously, due to the lack of data for  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ , data only exists for the electronic band gap of the  $\beta$ -phase and only exists for the optical band gap of the  $\alpha$ -phase (Fig. 2(c)). While building the statistical models, the  $\alpha$ -phase optical band gap was not taken into consideration but was included as a comparison for the final prediction of the optical band gap. In Fig. 2(c), it should also be noted that for the electronic band gap of  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ , the lower endpoint corresponding to an Al composition equal to 0, which is  $\beta\text{-Ga}_2\text{O}_3$ , the experimental electronic band gap of 4.81 eV agrees with the theoretical value of 4.8 eV [5]. However, the electronic band gap upper endpoint corresponding to  $\text{Al}_2\text{O}_3$  is equal to 6.69 eV, which suggests that the material has undergone a structural transition to  $\theta\text{-Al}_2\text{O}_3$  at the 100 percent Al composition endpoint, which has a theoretical band gap of 6.62 eV [1]. The experimental optical band gap values were measured from a sample of  $\alpha\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  [7] and is provided as a reference, as previously discussed. Although the band gap is phase dependent, it is not possible to predict the phase of  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  versus the Al composition with this model. Instead, only the electronic band gap of the  $\beta$ -phase is considered for this model due to data on other phases being unavailable. If other materials that undergo phase transitions are included in the training data, it may be possible to predict the Al composition at which the phase transitions.

The first step in analyzing the data was to find the relation between the band gap and Al composition using more than 80 data points from various literature sources. The data points were analyzed by simple linear regression as displayed in Table 1. During this process, in order to improve the correlation coefficient ( $R^2$ ) values of certain relations, any material with an  $R^2$  value of less than 0.99 was post-processed using a Cook's distance method [13], with the exception of the optical band gap relation of  $\alpha\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  due to a small number of existing data points. For  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  and the electronic band gap of  $\beta\text{-}(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ , select data points were removed in order to improve the fit using the Cook's distance method [21]. Thus, this method set a threshold of deviation, in this case, plus or minus the average of all points, to determine potential data points. Any points that lie outside of this threshold were removed. Removal of data points does not necessarily indicate that the information is wrong, but only serves to improve the correlation of the linear regression to be used in further steps and increase accuracy going forward. For each material, few data points were

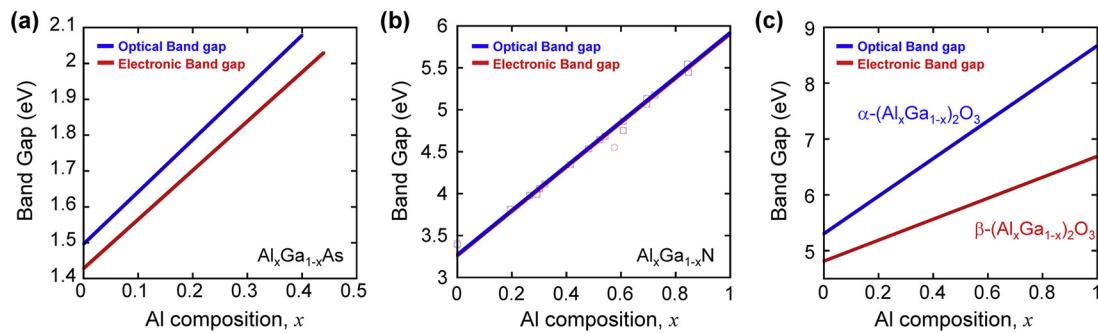


Fig. 2. A comparison of electronic and optical band gaps for (a)  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ , (b)  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  and (c)  $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ . In (c), the optical band gap refers to the  $\alpha$ -phase  $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  and the electronic band gap refers to the  $\beta$ -phase  $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ .

**Table 1**

Linear relations for band gap of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ,  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  and  $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  with their correlation coefficient values.  $E_{g,o}$  denotes optical band gap.

Material	Relation	$R^2$	Number of Unique Compositions
$\text{Al}_x\text{Ga}_{1-x}\text{As}$ (electronic)	$E_g = 1.37 \times x_{\text{Al}} + 1.43$	0.9999	44
$\text{Al}_x\text{Ga}_{1-x}\text{As}$ (optical)	$E_{g,o} = 1.47 \times x_{\text{Al}} + 1.49$	0.9929	10
$\text{Al}_x\text{Ga}_{1-x}\text{N}$ (electronic)	$E_g = 2.67 \times x_{\text{Al}} + 3.26$	0.9852	19
$\text{Al}_x\text{Ga}_{1-x}\text{N}$ (optical)	$E_{g,o} = 2.67 \times x_{\text{Al}} + 3.26$	0.9852	8
$(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ (electronic)	$E_g = 1.88 \times x_{\text{Al}} + 4.81$	0.9368	43
$(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ (optical)	$E_{g,o} = 3.37 \times x_{\text{Al}} + 5.31$	0.9874	7

actually removed, meaning that the robustness of the model is not significantly compromised. After this step, the  $R^2$  values for the electronic and optical band gaps of  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  and the electronic band gap of  $\beta-(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  improved to 0.9947 and 0.9669 respectively.

With the relations in [Table 1](#), the principal component regression (PCR) method was used to calculate the electronic and optical band gaps as a function of Al compositions. In order to perform the PCR method, it is necessary to have various key material property data such as lattice parameter, density, and thermal conductivity for the same Al compositions for all properties in the source data. These are properties for which there is abundant data for both  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  and  $\text{Al}_x\text{Ga}_{1-x}\text{N}$ . However, for each property, there are different values used for the Al composition, resulting in missing values for some Al compositions. Furthermore, some data does not yet exist for certain properties of  $\beta-(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ , namely the density and thermal conductivity as a function of the Al composition. A complete list of all properties found for each material and the number of data points of each property is found in [Table 2](#).

Because data does not exist for all properties for all materials, only five were selected to construct the statistical models. The selected properties are found in [Table 3](#). To fill in missing values, the relation between the property and Al composition needs to be identified and then extrapolated over the range of Al compositions of the experimental values of the electronic band gap for each material. For these missing properties, the PCR method was used to calculate their values with respect to the Al composition using the reference materials, namely  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  and  $\text{Al}_x\text{Ga}_{1-x}\text{N}$ , as training data and a scaling law was applied to relate the property of the element, such as Al, Ga, As, or N, to the property of the compound based on the fraction of each element in the compound. The complete list of all properties versus Al composition  $x$  is found in [Table 2](#). Using the linear relations listed in [Table 3](#), it was possible to extrapolate the property values for the same Al composition values as the electronic band gap. These relations were determined after outliers were removed by the Cook's distance method [21]. For  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ , only the Al compositions  $x < 0.45$  are considered to maintain a direct band gap.

**Table 2**

All properties considered for the training data and statistical model. For each material, the number of data points sourced from literature for a given property is listed. The properties are electronic band gap ( $E_g$ ), optical band gap ( $E_{g,o}$ ), lattice parameter  $a$ , lattice parameter  $c$ , density ( $\rho$ ), thermal conductivity ( $k$ ), volume ( $V$ ), absorption coefficient ( $\alpha$ ), enthalpy of formation ( $\Delta H_f$ ), entropy of formation ( $\Delta S_f^\circ$ ), and equilibrium temperature ( $T_{\text{equil}}$ ).

Material	$E_g$	$E_{g,o}$	$a$	$c$	$\rho$	$k$	$V$	$\alpha$	$\Delta H_f$	$\Delta S_f^\circ$	$T_{\text{equil}}$
$\text{Al}_x\text{Ga}_{1-x}\text{As}$	45 [16]	10 [17,18]	6 [22]	0	6 [22]	11 [22]	0	0	0	0	0
$\text{Al}_x\text{Ga}_{1-x}\text{N}$	16 [19]	8 [20]	27 [23]	27 [23]	6 [23]	11 [24]	6 [23]	6 [23]	20 [25]	20 [25]	11 [25]
$\beta-(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$	35 [1,8]	7 [7]	12 [1]	10 [1]	0	0	0	0	0	0	0

**Table 3**

Linear relations used for finding properties values at missing data points.

Property	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	$\text{Al}_x\text{Ga}_{1-x}\text{N}$	$(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$
El. Band Gap [1,16,18,20]	$1.37 \times x + 3.26$	$2.62 \times x + 3.27$	$1.73 \times x + 4.85$
Opt. Band Gap [7,8,17,19]	$1.47 \times x + 1.49$	$2.78 \times x + 3.20$	$3.37 \times x + 5.31$
Lattice Param [1,22,23]	$0.0073 \times x + 5.65$	$-0.076 \times x + 3.19$	$-0.156 \times x + 5.80$
Density [22,23]	$-1.56 \times x + 5.32$	$-2.81 \times x + 6.14$	$-6.42 \times x + 11.8$
Thermal Conduct [22,24]	Nonlinear	Nonlinear	$3.29 \times x + 6.29$

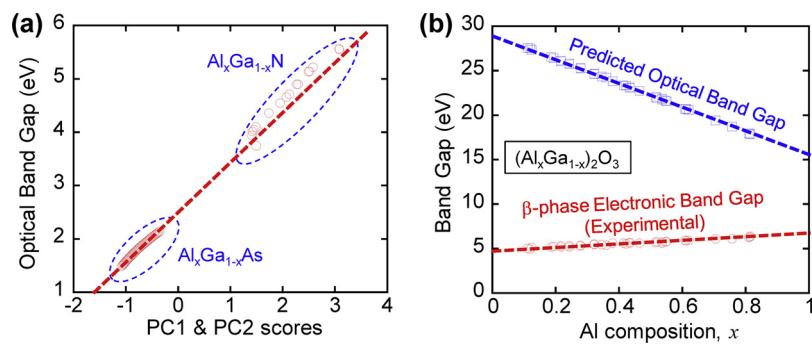
### 3. Calculations and results

The PCR method was used to predict the band gap using the training data prepared from the relationships listed in [Table 2](#). The goal of the PCR process was to develop an empirical relation between one property and a linear combination of other properties, derived from the coefficients determined by principal component analysis. Prior to the PCR step, the principal component analysis (PCA) was performed on the training data, excluding the optical band gap values from the computation. After principal component analysis (PCA), it was determined that the first two principal components capture 99.7% of the data, thus it is unnecessary to include more principal components in the calculation. The next step was to find a linear relation between the target property, the optical band gap, and the scores values of the training data which indicates the validation of the predicted values.

[Fig. 3\(a\)](#) displays the plot of the optical band gap versus the scores of the first two principal components. To find the empirical equation, the principal component weights were distributed into the loadings of each principal component, and denormalized, since the data was initially normalized for the PCA. For the PCR step, a model was developed and the results were displayed in [Fig. 3\(b\)](#). [Fig. 3\(b\)](#) shows that PCR produces inaccurate predictions of the optical band gap. The determined relation is shown in Eq. (1).

$$E_{g,o} = -0.344 \times x + 0.661 \times E_g + 1.63 \times a + 1.71 \times \rho - 0.309 \times k + 4.59 \quad (1)$$

where  $E_{g,o}$  is the optical band gap,  $x$  is the Al composition,  $E_g$  is the electronic band gap,  $a$  is the lattice parameter,  $\rho$  is the density, and  $k$  is the thermal conductivity. With all properties included in the training data, the band gap was far too high to be considered valid. To refine this model, the method used to calculate the values of density and thermal conductivity for  $\beta-(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  was changed from PCR to a scaling law. Previously, the PCR method was used to identify linear relations between the property and the Al composition (as seen in [Table 2](#)) and to find values of the missing properties for  $\beta-(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  using the reference materials as training data. Since using PCR to fill in missing values for properties before the final calculation can introduce more uncertainty into the prediction, an alternate scaling law was used by basing the value of the property at a specific Al composition from the fraction of each element, Al, Ga, and O, in the compound at that composition and the property of the singular elements. This scaling law



**Fig. 3.** (a) Recalculated Optical band gap of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  and  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  using a training data with their density, thermal conductivity. (b) Predicted optical band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  with properties calculated by scaling law. Experimental values of electronic band gap from literatures are shown in red in Fig. 2(b) and (d).

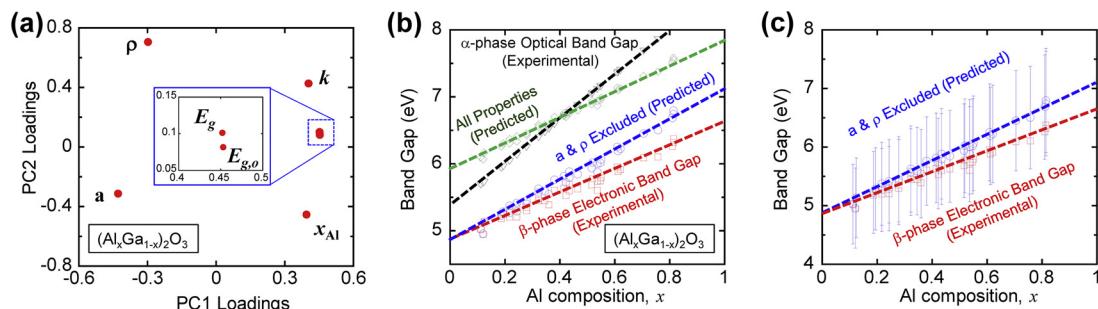
was then used to calculate the missing property values for the same Al compositions as the experimental electronic band gap data, which were the same Al compositions as used previously, and those property values are the ones used in this PCR model. From Fig. 3(b), it is clear that the PCR model is inaccurate in predicting the optical band gap due to the extremely high values produced. Discrepancies such as the nonlinear thermal conductivity of the  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  and  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  systems versus the linear nature of the  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  system could account for this error.

Because the PCR step yielded inaccurate results, the Gaussian stochastic process (GaSP) prediction was implemented. A GaSP can deliver a prediction of a desired variable, in this case the optical band gap, derived from the correlations between the input variables in the training data, in this case the properties of  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  and  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ . From the PCA loadings plot, shown in Fig. 4(a), it is apparent that there are no strong positive correlations between any of the properties used in the PCR, with the exception of the electronic and optical band gaps. Even though there is separation between data points for the electronic and optical band gaps, they are still highly correlated and both increase linearly with Al composition for the range of Al composition considered. The separation is due to the electronic band gap being equal to the optical band gap of  $\text{Al}_x\text{Ga}_{1-x}\text{N}$ , but not equal to that of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ . Because there are no strong correlations besides electrical and optical band gaps, all properties were included in the training data for the first attempt. The training data input into the computational model consists of inputs, all properties minus the optical band gap, and outputs, the optical band gap. Test data was also input including the same properties as the training data, and optical band gap predictions were made for the Al compositions included in the test data inputs. Based on the PCA result, density and lattice parameter were excluded from the calculation because those properties are not positively correlated to the rest of the properties. The Al composition is still considered due to its direct effect on the band gap and its strong positive correlation. The thermal conductivity is also considered due to the effect of temperature on the band

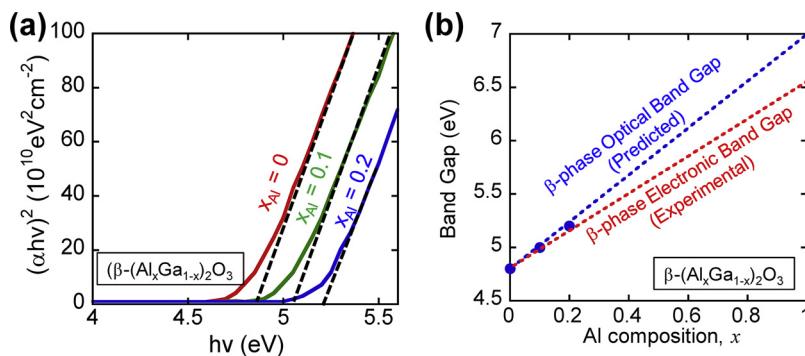
gap, which is directly related to thermal conductivity. This relation between temperature and band gap is also shown by the close proximity of the thermal conductivity and band gap data points of the PCA loadings plot in Fig. 4(a). After these properties were excluded, it was possible to perform the GaSP again with new training data to obtain a prediction of the optical band gap as shown in Fig. 4(b). For this GaSP, the Gaussian correlation was used for the highest accuracy in results. Both the predictions for the model with all properties in the training data and the model with density and lattice parameter excluded are displayed. The optical band gap of the  $\alpha$ -phase was also included as a reference for the possible values of the predicted optical band gap. Based on the results displayed in Fig. 4(b), the most realistic model is the one that excludes density and lattice parameter, due to considerations from the PCA that density and lattice parameter are uncorrelated and due to the relatively close matching of the predicted optical band gap and the known electronic band gap. For example, the lower endpoint of the model excluding density and lattice parameter matches the true endpoint from experimental data, whereas the lower endpoint of the model with all properties included does not match the true endpoint. Considering the standard error presented in Fig. 4(c), it is also clear that the experimental electronic band gap falls within the range of uncertainty of the predicted optical band gap of the model excluding density and lattice parameter. From the results of the Gaussian stochastic process predictions, the optical band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  is predicted to have the relation shown in Eq. (2),

$$E_{g,o} = 2.34 \times x + 4.78 \quad (2)$$

In order to validate the predicted optical band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ , the transmittance properties of three different 100 nm thick  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  on  $\beta$ - $\text{Ga}_2\text{O}_3$  substrates whose Al composition  $x$  are 0, 0.1, and 0.25 were investigated by conventional UV-VIS spectrophotometer (Shimadzu 3600 UV-vis-NIR spectrophotometer). To calculate the optical band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ , absorption coefficient ( $\alpha$ ) was calculated using Eq. (3),



**Fig. 4.** (a) PCA loadings plot of all properties of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ . As shown in inset, high correlations exist for only electronic band gap and optical band gap. (b) Gaussian-correlated Gaussian stochastic process prediction of the optical band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$ . Predicted optical band gap with lattice parameter (a) and density ( $\rho$ ) excluded shown in blue and electronic band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  is shown in red. Optical band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  is shown in black as a comparison. (c) Final predicted optical band gap showed along with the experimental optical band gap. The calculated standard error of each predicted data point is also shown.



$$\alpha = -\ln(T) / d \quad (3)$$

where  $T$  and  $d$  are the transmittance and the film thickness, respectively. Then, the optical bandgap was derived from  $\alpha^{1/2}$  versus  $h\nu$  curve using Eq. (4),

$$\alpha = (h\nu - E_g)^2 \quad (4)$$

As shown in Fig. 5(a), optical band gap was derived from  $\alpha^{1/2}$  versus  $h\nu$  curve by extrapolating the linear part of the curve. The optical band gaps of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  for Al composition  $x$  of 0, 0.1, and 0.2 were measured to 4.8 eV, 5.0 eV, and 5.2 eV and plotted in Fig. 5(b) with the predicted values which agree well with the measured optical values.

#### 4. Conclusion

In conclusion, the optical band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  was predicted using statistic-based material informatics approaches such as principal component regression and a Gaussian stochastic process by taking the large amount of material properties in similar mature semiconductors. The Gaussian stochastic process models yield more accurate results, especially when some properties are excluded from the model based on the PCA results. From the results of the Gaussian stochastic process predictions, the optical band gap of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  is successfully predicted and validated.

#### Acknowledgements

This work is supported by the New York State Center of Excellence

**Fig. 5.** (a) The plot of  $(\alpha h\nu)$  [2] versus  $h\nu$  extracted from UV-vis transmittance spectrum of the  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  film with  $x = 0$ , 0.1, and 0.2. (b) A comparison of measured (blue dots) and predicted optical band gap (blue dash) and experimental electronic band gap (red dash) of  $\beta$ - $(\text{Al}_x\text{Ga}_{1-x})_2\text{O}_3$  from Figs. 4(a) and 3(b), respectively.

in Materials Informatics (CMI) program and the seed grant by Research and Education in energy, Environment and Water (RENEW) institute at the University at Buffalo.

#### References

- [1] B.W. Krueger, et al., *J. Am. Ceram. Soc.* 99 (7) (2016) 2467.
- [2] V.R. Moorthi, *Power Electronics: Devices, Circuits and Industrial Applications*, Oxford University Press, Oxford; Delhi, 2005.
- [3] M. Kim, et al., *J. Mater. Chem. C* 5 (33) (2017) 8338.
- [4] M. Higashiwaki, et al., *Appl. Phys. Lett.* 103 (12) (2013) 123511.
- [5] H. Zhou, et al., *IEEE Electron Device Lett.* 38 (1) (2017) 103.
- [6] E. Gillet, B. Ealet, *Surf. Sci.* 273 (3) (1992) 427.
- [7] H. Ito, et al., *Jpn. J. Appl. Phys.* 51 (10) (2012).
- [8] F. Zhang, et al., *Appl. Phys. Lett.* 105 (16) (2014) 162107.
- [9] T. Oshima, et al., *Jpn. J. Appl. Phys.* 48 (7) (2009).
- [10] S. Srinivasan, K. Rajan, *Materials* 6 (1) (2013) 279.
- [11] S. Srinivasan, K. Rajan, *Comput. Sci. Eng.* 15 (5) (2013) 22.
- [12] H. Rabitz, K. Shim, *J. Chem. Phys.* 111 (23) (1999) 10640.
- [13] C. Suh, K. Rajan, *Appl. Surf. Sci.* 223 (1-3) (2004) 148.
- [14] P. Dey, et al., *Comp. Mater. Sci.* 83 (2014) 185.
- [15] M. Plumlee, et al., *Stat* 2 (1) (2013) 159.
- [16] M. El Allali, et al., *Phys. Rev. B* 48 (7) (1993) 4398.
- [17] B. Fluegel, et al., *Jpn. J. Appl. Phys.* 54 (2015) 4.
- [18] B. Lambert, et al., *Semicond. Sci. Technol.* 2 (8) (1987) 491.
- [19] Q.S. Paduano, et al., *Jpn. J. Appl. Phys. Part. 1-Regul. Pap. Short. Notes Rev. Pap.* 41 (4A) (2002) 1936.
- [20] S. Zhao, et al., *APL Mater.* 4 (8) (2016) 86115.
- [21] J.A. Díaz-García, G. González-Farías, *J. Stat. Plan. Inference* 120 (1) (2004) 119.
- [22] S. Adachi, *J. Appl. Phys.* 58 (3) (1985) R1.
- [23] A. Belousov, et al., *J. Cryst. Growth* 312 (18) (2010) 2585.
- [24] W. Liu, A.A. Balandin, *J. Appl. Phys.* 97 (7) (2005) 73710.
- [25] A. Belousov, et al., *J. Cryst. Growth* 312 (18) (2010) 2579.