Machine-Learning-Assisted First-Principles Calculations of Strained $InAs_{1-x}Sb_x$ Alloys for Curved Focal-Plane Arrays

Alexandros Kyrtsos,[1,*] John Glennon,[2] Andreu Glasmann,[1] Mark R. O'Masta,[3] Binh-Minh Nguyen,[3] and Enrico Bellotti[1,2]

[1] *Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts 02215, USA*
[2] *Division of Materials Science and Engineering, Boston University, Boston, Massachusetts 02215, USA*
[3] *HRL Laboratories, LLC, 3011 Malibu Canyon Road, Malibu, California 90265, USA*

(Received 24 March 2021; revised 29 April 2021; accepted 7 May 2021; published 3 June 2021)

Curved focal-plane arrays offer significant advantages compared with their flat counterparts. However, the curving of the material induces strains which alter its optoelectronic properties. A comprehensive framework is presented for the computational investigation of the band gaps of $InAs_{1-x}Sb_x$ alloys under various strain conditions that are relevant for manufacturing curved focal-plane arrays. The framework consists of both standard and hybrid functional density functional theory (DFT) calculations, finite-element analysis (FEA) calculations, and Gaussian process (GP) regression. The DFT calculations are used for investigating the change of the band gap of the material under various strain conditions. This dataset is then used for training a GP model which is utilized to assess the effects of strain across the device, based on the FEA simulations. The results show excellent predictive capabilities for the machine-learning model at a significantly reduced computational cost and are directly transferable to other device designs.

DOI: 10.1103/PhysRevApplied.15.064008

I. INTRODUCTION

The alloying of InAs and InSb has led to the smallest band gaps among the III–V semiconductors [1], attracting attention both for fundamental research and relevant applications in the infrared (IR) regime. Specifically, the pure phases of InAs and InSb exhibit band gaps of 0.415 and 0.235 eV, respectively, whereas their ternary alloys demonstrate strong positive (downward) band gap bowing [2,3]. In particular, $InAs_{0.91}Sb_{0.09}$ can be grown lattice matched to GaSb [4], enabling mid-wavelength infrared (MWIR, i.e., 3–5 μm) applications [5]. Nonetheless, long-wavelength infrared (LWIR, i.e., 8–12 μm) applications demand even smaller band gaps. The use of $InAs_{1-x}Sb_x$ alloys, hereafter called InAsSb for the sake of brevity, in LWIR optoelectronic applications is of particular importance and has been discussed in the literature for decades [6–8]. However, the adoption of this material for such applications has been met with limited success. The two main reasons for that are the initially accepted values of the band gap bowing parameter, which did not allow for sufficiently small band gaps, and the absence of high-quality substrates, which enable the epitaxial growth of the material at relevant compositions [6].

A typical approach for obtaining suitable substrates for this material is the use of thick monolithic buffer layers that allow the relaxation through threading and misfit dislocations. Grading schemes [9,10] and blocking layers [11] techniques have been developed in order to minimize the vertical propagation of the mismatch related defects into the active region. These techniques, although successful in greatly reducing the defect densities, lead to residual strain, which affects the band gap and other optical properties of the material [5,6]. Further improvement to the quality of the samples has been achieved by the combination of compositional grading and a virtual substrate (VS) with the interfacial misfit (IMF) approach [5,6,12–14], resulting in unrelaxed and unstrained materials.

Early experimental data on samples grown on InAs substrates indicated bowing parameters smaller than 0.7 eV [2,3], which would imply that the band gaps are not small enough for LWIR applications. However, the development of methods to produce high-quality substrates has led to materials with reduced band gaps [6,13,15], where the bowing parameter reaches 0.87 eV. Therefore, InAsSb has become a promising platform that supports applications in the whole IR regime.

One emerging application is the creation of InAsSb-based curved focal-plane arrays (FPAs). The idea stems from the recent success in curving photographic sensors of visible-light cameras, which led to enhanced clarity and near-uniform illumination from the center to the corners of the image [16]. Typical imaging systems require numerous optical elements to flatten the Petzval field curvature to

*akyrtsos@bu.edu

match a planar image sensor. The complexity of the optics to achieve field flattening, while minimizing optical aberrations, increases the size, weight, and cost of the whole system [17]. On the other hand, a curved image sensor overcomes this limitation, significantly reducing off-axis aberrations. Improving the image quality at extreme angles is a desirable feature for various applications, such as situational awareness.

The curving of the die induces strain fields in the semiconducting material, which alter its optoelectronic properties. Hence, the study of the response of the optoelectronic properties of the material under different strain configurations is important for the design and creation of curved sensors. Therefore, it is the aim of this work to present a comprehensive and efficient computational framework for the study of the change of the physical properties of the strained material in the die. Density functional theory (DFT) calculations present an attractive method for capturing band shifts in strained semiconductors. However, this approach is limited by its high computational cost, especially when considering the numerous strain configurations in a curved die. Alternatively, machine-learning algorithms have shown excellent predictive capabilities in a variety of applications, accelerating various computational and experimental tasks. In particular, the use of machine learning in recent years has accelerated the simulation and design of devices [18–20]. Consequently, a machine-learning-assisted approach is presented here, aiming to overcome the limitations of the high computational cost of a purely first-principles calculations approach. The framework is based on a limited number of computationally expensive, yet high-fidelity, DFT calculations for obtaining the band gap of the strained material under different strain configurations, coupled with finite-element analysis (FEA) modeling of the stresses and strains in the die. Consequently, the use of machine learning enables the efficient coupling of the different methods, and significantly reduces the overall computational cost, while preserving the high-fidelity of the results.

The manuscript is organized as follows: Sec. II describes the various methods and models employed in this work; Sec. III presents and discusses the results; finally, Sec. IV summarizes and concludes the manuscript.

II. METHOD

ABAQUS/STANDARD [21] is used to perform finite-element simulations to estimate the strain induced in creating a curved FPA. The model dictates the displacement of the die surface directly, which forces the initially flat die into a spherical segment with a prescribed radius of curvature (ROC). The die is modeled as a 45 mm by 50 mm rectangular deformable part with a thickness of 100 μm. A 50 $\times$ 50 rectangular mesh is implemented with five integration points through the thickness of reduced integration shell elements (S4R in ABAQUS notation). Composition-dependent elastic constants are utilized, as determined by linear interpolation of the pure InAs and pure InSb values. Adopting the cubic symmetry of the materials, these are $C_{11} = 8.33 \times 10^{11}$ dyn/cm^2, $C_{12} = 4.53 \times 10^{11}$ dyn/cm^2, $C_{44} = 3.96 \times 10^{11}$ dyn/cm^2 for pure InAs [22] and $C_{11} = 6.67 \times 10^{11}$ dyn/cm^2, $C_{12} = 3.65 \times 10^{11}$ dyn/cm^2, $C_{44} = 3.02 \times 10^{11}$ dyn/cm^2 for pure InSb [23]. Only one quadrant of the die is modeled explicitly, taking advantage of symmetries about the x and y axes, respectively (z being through the thickness). In order to explore the applicability of our approach in both low- and high-strain conditions, the die is curved into two different ROCs of 140 and 70 mm, respectively. Spherical curvature is established by applying a spherically varying displacement condition to the shell.

Our DFT calculations employ the projector augmented wave (PAW) [24,25] method as implemented in the Vienna *ab initio* simulation package (VASP) [26]. The exchange-correlation energy is treated using both standard and hybrid functionals. In the case of standard DFT, the generalized gradient approximation (GGA) in the flavor of the PBEsol [27] functional is used for the optimization of the atomic structure. The hybrid functional calculations are performed using the parametrization of Heyd, Scuseria, and Ernzerhof (HSE) [28,29] in order to obtain the band gaps. By combining standard and hybrid functional calculations, we are able to reduce the computational cost while maintaining the high fidelity of our data. A composition-dependent mixing parameter for the Hartree-Fock exchange is used for the HSE calculation. The mixing parameter changes linearly from 0.286 for pure InAs ($x = 0$) to 0.264 for pure InSb ($x = 1$). Further details for the use of a variable mixing parameter can be found elsewhere [15,30,31].

The known shortcoming of standard DFT, which is the underestimation of the band gap [32], becomes a particularly important hurdle in narrow gap systems, where standard DFT yields even negative band gaps. Meta-GGA functionals help alleviate this issue to a certain extent at a computational cost comparable to standard DFT calculations. The Tran-Blaha modified Becke-Johnson (mBJ) [33,34] meta-GGA functional yields band gaps at the same degree of accuracy as hybrid functional or GW calculations and in very good agreement with experimental data [35,36]. The excellent agreement of the band gaps in the InAsSb system between HSE and mBJ calculations has been demonstrated before [15]. In the present work, the mBJ functional is used only for testing purposes and all the results presented here are obtained using HSE calculations at PBEsol relaxed structures.

The supercells for the calculations are created from 64-atom InAs supercells, where three and six As atoms are randomly replaced by Sb in order to obtain alloy compositions of 9.375% and 18.75%, respectively. For the sake of brevity, we refer to them as 9% and 19%,

respectively. These supercells are generated as a $2 \times 2 \times 2$ replication of the 8-atom conventional cubic cell. A cutoff energy of 350 eV is used for the plane-wave basis set, and the Brillouin zone is sampled using a $2 \times 2 \times 2$ Γ-centered k-mesh. Each structure is relaxed until the forces are reduced to less than 0.005 eV/Å. Furthermore, the In $3d$ electrons are treated as valence electrons and spin-orbit coupling (SOC) is taken into account only for the band gap calculations.

The atomic configuration plays an important role on the electronic properties of the alloys [15,31]. In order to determine the effect of the atomic configuration on the band gaps of the unstrained random alloys, a set of 200 distinct atomic configurations for each composition is used. Thereafter, a subset of 21 configurations is compiled for each composition based on the relevance of their band gaps and is used for the subsequent strained calculations. The application of strain in the supercells used in this work is straightforward because the faces of the unstrained cubic supercells correspond to the {100} directions of the crystal.

The generalized Hooke's law, relating the normal stresses σ_{ii} to the resulting normal strains ε_{ii} for $i = x, y, z$, can be expressed as

$$\begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \end{bmatrix} = -\frac{1}{E} \begin{bmatrix} -1 & \nu & \nu \\ \nu & -1 & \nu \\ \nu & \nu & -1 \end{bmatrix} \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \end{bmatrix}, \tag{1}$$

where E is the Young's modulus and ν is the Poisson's ratio. The Young's modulus and the actual numerical values of the stresses are not required for building the supercells for the DFT calculations. In fact, employing Eq. (1), only the stress condition needs to be specified in order to obtain the relationship among the strains. We distinguish the two special cases of uniaxial and biaxial stress. These cases are obtained when the applied stress is nonzero in one and two directions, respectively. For instance, in the case of uniaxial stress with $\sigma_{xx} \neq 0$, the relationship among the normal strains is given by

$$\varepsilon_{yy} = \varepsilon_{zz} = -\nu \varepsilon_{xx}, \tag{2}$$

Alternatively, in the case of biaxial stress with $\sigma_{xx} \neq 0$, $\sigma_{yy} \neq 0$, and $\sigma_{zz} = 0$, the relationship among the normal strains is given by

$$\varepsilon_{zz} = -\frac{\nu}{1-\nu}(\varepsilon_{xx} + \varepsilon_{yy}). \tag{3}$$

The special biaxial stress condition for which the two nonzero stresses are equal ($\sigma_{xx} = \sigma_{yy}$) is typically called equibiaxial. The majority of the calculations in this work are done under the uniaxial and equibiaxial stress conditions. In the uniaxial case, the choice of one strain ($\varepsilon_{\mathrm{uxl}}$) determines the rest. In the equibiaxial case, the choice of a pair of strains ($\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{\mathrm{bxl}}$) determines the third. Hence, Eqs. (2) and (3) can be written as $\varepsilon_{yy} = \varepsilon_{zz} = -\nu\varepsilon_{\mathrm{uxl}}$ and $\varepsilon_{zz} = -2\nu\varepsilon_{\mathrm{bxl}}/(1-\nu)$, respectively. It is important to note that the choice of directions is arbitrary and any permutation is valid. Consequently, specifying the values of $\varepsilon_{\mathrm{uxl}}$ and $\varepsilon_{\mathrm{bxl}}$ determines the supercells for the DFT calculations under uniaxial and equibiaxial stress conditions, respectively.

Furthermore, shear deformations are decoupled from the normal stresses and depend only on shear stresses. As a result, a shear deformation $\gamma_{xy} = \gamma$ can be included as well, simply by changing the orthogonality of the supercells. Therefore, all the strains can be determined and the corresponding strained supercells can be obtained for the various cases. For the DFT calculations, the Poisson's ratio is set to $\nu = 0.35$ in accordance with the elastic constants of the materials.

As a final comment for the construction of the strained supercells of the random alloys, it should be noted that the three different {100} directions are not equivalent because of the randomness of the alloy. Therefore, taking the uniaxial case as an example, any permutation of the $\{\varepsilon_{\mathrm{uxl}}, -\nu\varepsilon_{\mathrm{uxl}}, -\nu\varepsilon_{\mathrm{uxl}}\}$ strain configuration yields a different strained supercell. As a result, the set of 21 unstrained supercells per composition discussed earlier yields 63 distinct supercells per composition and strain configuration.

The active material is subject to a complex set of strains $\{\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{zz}, \gamma\}$ where each configuration can be viewed as a four-dimensional strain vector $\vec{S}$, which depends on the physical x, y, and z coordinates of the die, given by $\vec{R}$. The FEA calculations provide the strain vectors $\vec{S}$ as a function of the coordinates $\vec{R}$ of the material. At the same time, the DFT calculations link physical properties of the material, such as the band gap, to different strain configurations. Therefore, using a suitable regression tool, one can obtain the physical properties of the material as a function of the actual positions in the active region of the device. This tool has to have enough predictive capabilities so that it minimizes the amount of DFT calculations that are required for an accurate fit, in order to mitigate their high computational cost.

Gaussian processes (GPs) [37] are a powerful machine-learning algorithm which can be used, among other things, for nonparametric Bayesian regression. The main advantage of GP regression is that it provides a reliable estimate of the model's uncertainty. The basis of this method relies on conditioning a prior distribution of functions on the observed data to obtain the posterior distribution. Eventually, the mean value and standard deviation of this distribution can be used for providing predictions and their uncertainties, respectively. We use the GP implementation provided in the PYTHON SCIKIT-LEARN package [38]. The kernel we used combined a radial basis function (RBF) kernel and a white noise kernel.

Specifically, an original dataset is collected consisting of strain vectors as inputs and the corresponding band

gap values obtained through DFT calculations as outputs. This dataset is then used to train the GP model. Using the trained model, the predictions of the band gap differences along the die are obtained based on the strain configurations that are determined through the FEA simulations. As the GP regression inherently provides the uncertainty of the model for any given strain configuration, this property can be utilized to further augment the original dataset appropriately for any device design. Specifically, in our case we chose to include three additional strain configurations as described later.

III. RESULTS

The ε_{xx}, ε_{yy}, ε_{zz}, and γ components of the strain vector $\vec{S}$ are extracted from each unique node at the top of the surface of the die from the FEA calculations, along with their respective physical location on the die. As an example case, Fig. 1 shows the colormaps of the four different strain components obtained for the high curvature (high strain) die of the $InAs_{0.91}Sb_{0.09}$ alloy. The results are identical for all practical purposes for the $InAs_{0.81}Sb_{0.19}$ alloy. For the low-strain configuration, the results are similar with the maximum strains reaching only 0.6% instead.

As mentioned earlier, the atomic relaxations are performed using the PBEsol functional in order to reduce the computational cost. Consequently, the HSE functional is used to obtain the band gaps of the 200 distinct random structures in each composition. Figure 2 shows the distribution of band gaps as violin plots for the two compositions of the unstrained material. The width of the violin is scaled by the number of observations in that bin. One immediate observation is that the band gap spread of the $InAs_{0.91}Sb_{0.09}$ alloys is narrower compared with the

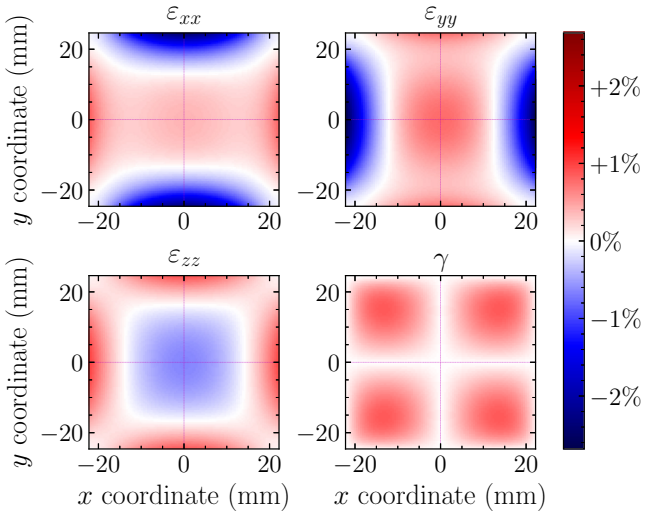

FIG. 1. The four strains obtained by the FEA calculations of the $InAs_{0.91}Sb_{0.09}$ alloy curved to a 70 mm ROC.

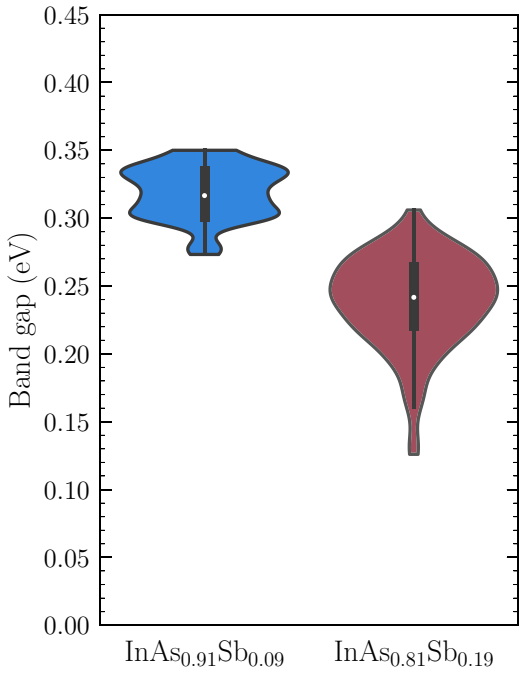

FIG. 2. Boxplots of the band gaps, including their distribution for the two InAsSb alloys.

$InAs_{0.81}Sb_{0.19}$ alloys, with band gaps ranging from 0.27 to 0.35 eV and 0.13 to 0.31 eV, respectively. The violin plots also include boxplots, which indicate the quartiles and the median value for each distribution.

Choosing a single metric to describe the band gaps is not a trivial matter because either the median or the mean value could be considered. In the latter case, the way to obtain the average is again not straightforward. The mean value can be obtained as the simple arithmetic average, or the weighted average by considering the energy of each configuration. In the latter case, the weight for each configuration i is given by

$$p_i = \frac{e^{-\beta \Delta G_i}}{\sum_j e^{-\beta \Delta G_j}}, \tag{4}$$

where $\beta = 1/k_B T$, ΔG_i refers to the Gibbs free energy and the summation occurs over all the configurations. Any constant shift in the free energy cancels out. Therefore, the configuration entropy can be neglected considering that it is the same for all the configurations because they are all random and of the same composition. Furthermore, the formation enthalpy would normally require knowing the total energies of InAs and InSb, so that

$$\Delta H = E_{\text{tot}}^{\text{InAsSb}} - (1-x)E_{\text{tot}}^{\text{InAs}} - xE_{\text{tot}}^{\text{InSb}}. \tag{5}$$

However, for a given composition, the total energies of the pure phases are also a constant shift to the total energy of each configuration. Hence, eventually it suffices to use just

TABLE I. The median and weighted average values of the band gaps for the two different alloy compositions. The weighting factors of the weighted average are obtained at various temperatures.

		Weighted average (eV)				
x	Median (eV)	$T = 50$ K	$T = 400$ K	$T = 700$ K	$T = 1000$ K	$T \to \infty$
9%	0.317	0.341	0.329	0.326	0.324	0.319
19%	0.242	0.297	0.267	0.257	0.252	0.239

the total energy of each configuration (E_i) to determine the weighted average, with weights given by

$$p_i = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}}. \tag{6}$$

Table I summarizes the band gaps for the two compositions, using the median and the mean value in each case. The temperature of $T \to \infty$ essentially corresponds to the arithmetic average. A trend that is observed in Table I is that the average band gap increases with decreasing temperature in which the weights are computed. This can be explained by the fact that the band gaps and the total energies are correlated in a way that lower total energy configurations yield higher band gaps. Therefore, because low temperatures favor the configurations with lower total energies, the band gaps are larger. However, a more realistic weighted average of the band gaps can be obtained for temperatures that are relevant during growth, namely closer to 700 K. As neither the 9% nor the 19% compositions exhibit band gap distributions that are normal, the median value is more appropriate. Another advantage of the median is that it is less sensitive to outliers. Regardless, the results of the median and the mean values are only marginally different for temperatures that are experimentally relevant.

The results of the band gaps can be compared to the model provided by [39] where the values of the band gap for the two compositions of 9% and 19% are 0.336 and 0.273 eV, respectively. Therefore, our median values are in good agreement with the said model. As a final note, it is important to mention that typically the weighted average obtained by the total energies is not sufficient to provide a good metric because it relies only on bulk thermodynamics to determine the occurrence probability of each structure. This is not adequate because kinetics and surface phenomena that occur during growth and play a significant role on the formation of the structure are not accounted. Nonetheless, in this particular case, because all the structures are random, the total energies are sufficient for obtaining the weighted average.

Performing calculations in strained supercells for the whole set of 200 configurations per composition would be very computationally expensive and ineffective. Furthermore, not all these configurations yield band gaps that are available experimentally. Therefore, the strained calculations are performed in a subset of 21 configurations for each composition. The choice of the subset for each composition is made based on the band gaps of the configurations. The band gaps of the configurations that constituted the subsets are similar to within approximately 15 meV and the mean value of each subset is similar to that obtained for the whole dataset. As mentioned in the previous section, owing to the equivalent permutations of the different strain configurations and the fact that the alloys are random, these 21 atomic configurations actually yield 63 distinct strained supercells. This keeps the computational cost reasonable, while composing a representative sample.

Ideally, for the computational framework to be as predictive and transferable as possible, the DFT calculations should be decoupled from the FEA modeling of the stresses in the die. In other words, the DFT dataset should be agnostic to the particular stresses in the die. Indeed, this is followed in this work. Therefore, three independent strain conditions are used in order to obtain an initial dataset of band gaps for strained configurations. These conditions refer to uniaxial, biaxial, and pure shear stress, respectively, and will serve as a transferable dataset for strained calculations. The resulting deformations from these conditions are discussed in Sec. II. The strains $\varepsilon_{\mathrm{uxl}}$ of -2.5%, -1.5%, -0.5%, $+0.2\%$, $+0.5\%$, $+1.5\%$, and $+2.5\%$ are used for the uniaxial stress condition. The strain in the two other directions that are free of stress is given by Eq. (2). In the case of equibiaxial loading, the strains $\varepsilon_{\mathrm{bxl}}$ of -2.0%, -1.0%, -0.5%, $+0.2\%$, $+0.8\%$, $+1.5\%$, and $+2.0\%$ are used. These values refer to the two strains in the directions of the applied stresses. The strain in the third direction is given by Eq. (3). Finally, in the case of pure shear stress, the strains γ of $+0.5\%$, $+1.0\%$, $+1.5\%$, $+2.0\%$, and $+4.0\%$ are used.

Figure 3 shows the band gap difference compared with the unstrained material under the different strains for both compositions. The solid line shows the mean value of the strained configurations referenced to the mean value of the unstrained configurations. The shaded areas indicate the minimum and maximum values obtained for each strain configuration. The results are similar for both alloy compositions. The differences are mainly attributed to the different strain conditions. The largest difference in the band gap is observed for the case of biaxial stress and particularly for positive (tensile) loading, where the band gap

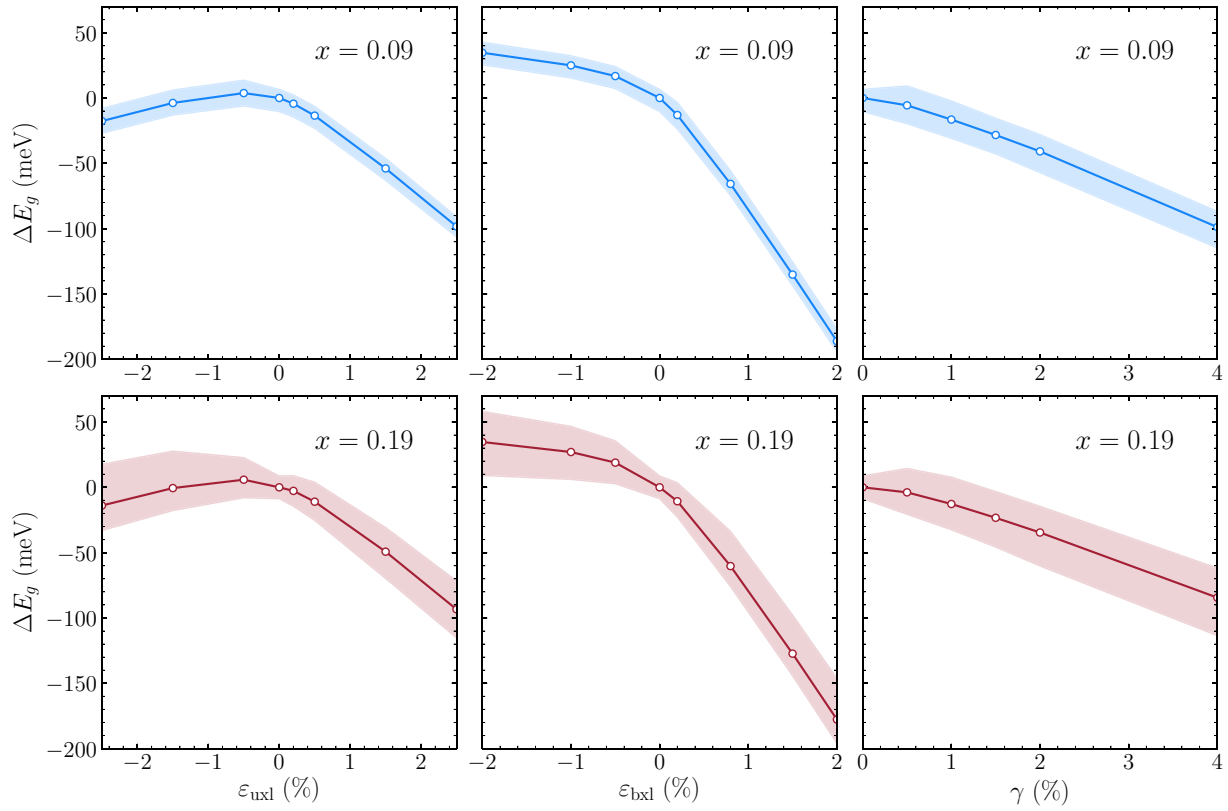

FIG. 3. Band gap difference compared with the unstrained material for the case of uniaxial (left), equibiaxial (middle), and pure shear (right) loading conditions for the $InAs_{0.91}Sb_{0.09}$ (top) and $InAs_{0.81}Sb_{0.19}$ (bottom) alloys. The shaded areas indicate the range of the minimum and maximum values obtained for each strain configuration.

drops by more than 170 meV for $\varepsilon_{\text{bxl}} = +2\%$. The band gaps are affected less under uniaxial loading, dropping by almost 100 meV for $\varepsilon_{\text{uxl}} = +2.5\%$ and less than 20 meV for $\varepsilon_{\text{uxl}} = -2.5\%$. The smallest effect is caused by the pure shear stresses with a band gap drop close to 40 meV for $\gamma = +2.0\%$.

The mean value of the band gap difference in every strain configuration is used in the training set for the GP regression model. There are seven strain configurations for both uniaxial and biaxial stress conditions, as well as five pure shear configurations. Owing to the three distinct permutations in the first two cases, the total number of strain configurations reaches 48, including also the unstrained case. Therefore, 48 $\vec{S}$ vectors are used for the training of the model for each composition. It is important to note that these strain vectors are completely agnostic to the actual strain vectors that exist in the curved die. The kernel of the GP consists of a RBF and a white noise kernel. The L-BFGS-B algorithm is used for the optimization. Finally, the input data are standardized by removing the mean and scaling to unit variance. The whole dataset is used for training, taking advantage of the fact that the validation can be performed directly on the dataset of the actual strains by using the standard deviation of the predictive distribution at the query points.

Before delving into the results obtained by the GP regression, it is important to discuss in more detail one significant advantage of this model. The predictions for the band gaps or the band gap differences in this particular case are obtained as the mean values of the posterior distribution of the GP, while the uncertainties are derived by its standard deviation. The latter is an extremely powerful feature of GP regression because it allows for the identification of the regions with the highest uncertainty. Consequently, these regions can be sampled with some additional DFT calculations, augmenting the original dataset. In our case, in order to reduce even further the uncertainty of our predictions, we augmented our initial dataset with three additional strain configurations. The additional strain configurations are given by {0.3%, −1.1%, 0.4%, 0.5%}, {0.2%, 0.8%, −0.6%, 0.0%}, and {0.2%, 0.4%, −0.3%, 0.5%}. As can be seen, the three strain configurations that are used for augmenting the original dataset consist of asymmetric biaxial strain and combinations with shear strain. This is not surprising because none of the above existed in the original dataset.

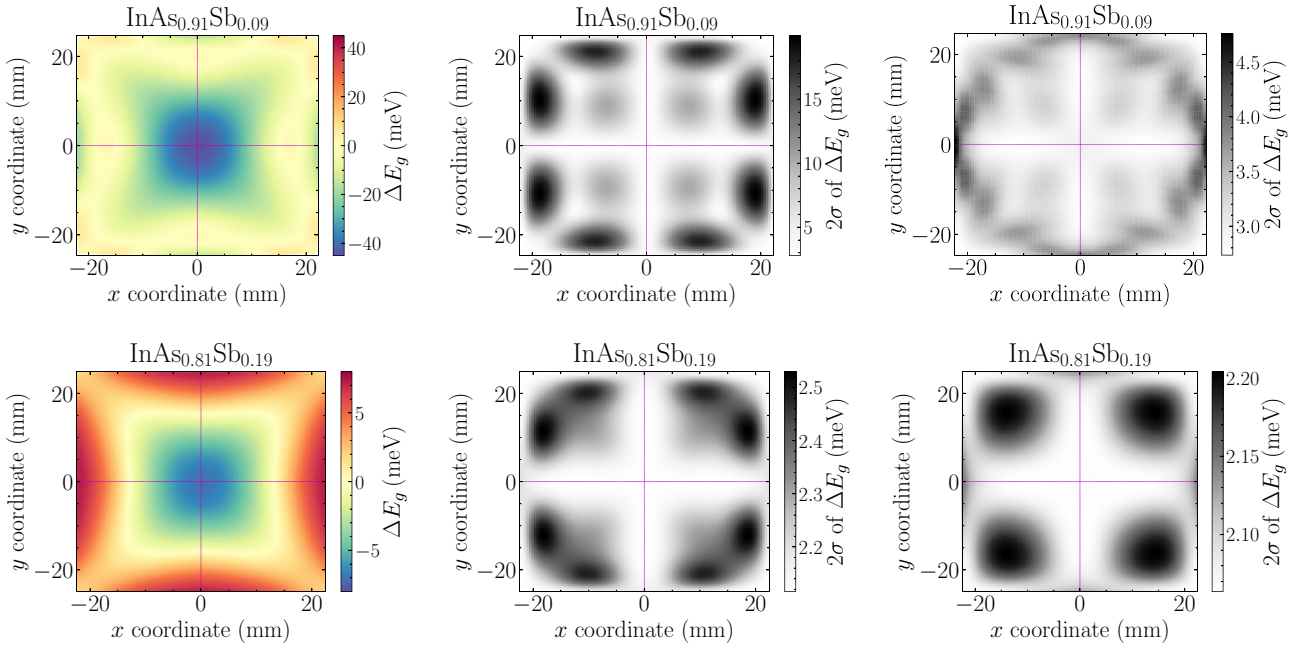

FIG. 4. Band gap change (left) compared to the unstrained material in different regions of the die under high curvature (70 mm ROC) for the $InAs_{0.91}Sb_{0.09}$ alloy (top) and low curvature (140 mm ROC) for the $InAs_{0.81}Sb_{0.19}$ alloy (bottom). The 2σ uncertainties, indicating the 95% confidence interval for the fit, are shown for the original (middle) and augmented (right) datasets.

Figure 4 shows the predicted band gap differences compared with the unstrained material, in the whole region of the die. The 2σ uncertainties are also shown as a function of the physical coordinates in order to obtain the 95% confidence interval for our results. The top and bottom row of the figure presents the results for highly and lowly curved dies, respectively, using the $InAs_{0.91}Sb_{0.09}$ and the $InAs_{0.81}Sb_{0.19}$ alloys. The results for each alloy are similar in each curvature. The middle and right columns present the 95% confidence interval for the predictions using the original and the augmented datasets, respectively. It is important to note that only the uncertainty of the model improved with the augmented dataset. The predictions remained the same for all practical purposes, showing the excellent performance and predictive capabilities of the original dataset. Furthermore, even though the uncertainty dropped for both dies, the change is more evident in the case of the highly strained die where the uncertainty dropped by more than a factor of four. This is due to the fact that the augmentation of the dataset is targeted to help the sampling of the large strains. Nonetheless, the low-strain die results also benefited.

It is common to assess the quality of the fit of a regression model with respect to the training data by the coefficient of determination R^2. However, this metric is not applicable for GP regression. For some kernels that are typically used in GP regression, the fit is guaranteed to predict the provided datapoints exactly, therefore yielding $R^2 = 1$ by definition. This is not the case when a white noise kernel is included as in this work. In this case, due to the noise, the model is not guaranteed to fit the training set exactly, therefore reducing slightly the value of R^2. The inclusion of a white kernel with a small noise variance of the order of 10^{-4} along with a RBF kernel is found to drastically improve the predictive capabilities of our GP regression and significantly decrease the uncertainties in the actual strain data. Evidence of this is provided in the Supplemental Material [40]. Furthermore, by including the white noise kernel changed only marginally the R^2 coefficient for the $InAs_{0.91}Sb_{0.09}$ and $InAs_{0.81}Sb_{0.19}$ alloys to 0.9997 and 0.9998, respectively. Therefore, the inclusion of the white kernel did not practically sacrifice the inherent property of the GP regression to fit exactly the training data.

At this point it would be useful to comment on the uncertainties of the model and its applicability on different datasets. As mentioned before, the uncertainties are obtained through the standard deviation of the posterior distribution of the GP. Naturally, the uncertainty of the model increases for strain configurations that are not included in the training. However, even though the uncertainty increases, the prediction, obtained by the mean of the posterior distribution, is not necessarily bad. As an example, let us consider the model trained on the original dataset which yields 2σ uncertainties of more than 18 meV in some regions of the highly strained die, as seen in Fig. 4. This region is the first to be sampled in order to generate the augmented dataset and, in fact, the values predicted by the original model and the actual DFT

values that are calculated for the augmentation differed by less than 2 meV. This is a testament to the excellent predictability of the GP model. It is important to mention that the choice of sampling points for augmentation should be made iteratively because after the inclusion of every additional datapoint in the training set, the region of highest uncertainty changes. Furthermore, because the GP model is actually agnostic of the way the training dataset is collected, it is expected to perform equally well regardless of whether the training set is obtained experimentally or computationally as in this work.

As a final remark, we would like to discuss the choice of using the band gap differences rather than the actual band gap values. The absolute value of the band gap, regardless of whether it is obtained experimentally or computationally, is subject to slight variations. The differences are used in this case with the hope that any systematic errors can be absorbed into them, making the results and the method more transferable. Nevertheless, this choice is just a preference and the proposed framework is not affected by it. Another significant aspect for the transferability of the results is the fact that the original DFT dataset, which is found to perform well, is completely agnostic to the actual strains that are induced present in the device. Therefore, it can be employed in other device designs as well, significantly accelerating the device design process. However, the strains should not exceed the limits of the strains included in the dataset.

IV. CONCLUSION

In conclusion, a computational framework constituted by DFT calculations, FEA modeling, and machine learning is presented as a comprehensive computational solution for the study of the properties of strained alloys which are used in various imaging applications. In particular, the framework is showcased for the study of the band gap changes caused by strain in the case of curved FPAs under both low and high curvatures in two distinct alloy compositions. An original dataset of band gaps under various basic strain configurations is obtained by DFT calculations. Consequently, the actual strains in the material are obtained by modeling the die via FEA. Finally, GP regression is employed to couple the results and predict the band gap changes in the actual die, as well as to predict the uncertainties of the model. Even though the original DFT dataset is completely agnostic to the actual strains present in the die, the results exhibit excellent predictive capabilities. An augmented DFT dataset with minimal additional computational cost is created based on the uncertainties of the initial model and is able to achieve even smaller uncertainties without any practical change in the predicted values though. Furthermore, this approach is expected to be easily transferable to other applications with different strain conditions. Finally, the results obtained by this approach can be utilized for further analysis, such as the study of the transport properties of the material.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the Defense Advanced Research Projects Agency (DARPA) through the FOcal arrays for Curved Infrared Imagers (FOCII) program (Grant No. N66001-20-C-4011) managed by Dr. Whitney Mason and the Center for Semiconductor Materials and Devices Modeling (Grant No. W911NF-18-2-0027) at Boston University. The computational resources were provided by the DoD HPC Systems and the 2019 Army Research Office DURIP Award (Grant No. W911NF-19-1-0161) made to Dr. E. Bellotti.

The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

[1] A. G. Thompson and J. C. Woolley, Energy-gap variation in mixed III–V alloys, Can. J. Phys. **45**, 255 (1967).

[2] Z. M. Fang, K. Y. Ma, D. H. Jaw, R. M. Cohen, and G. B. Stringfellow, Photoluminescence of InSb, InAs, and InAsSb grown by organometallic vapor phase epitaxy, J. Appl. Phys. **67**, 7034 (1990).

[3] M. Y. Yen, R. People, K. W. Wecht, and A. Y. Cho, Long-wavelength photoluminescence of $InAs_{1-x}Sb_x$ ($0 < x < 1$) grown by molecular beam epitaxy on (100) InAs, Appl. Phys. Lett. **52**, 489 (1988).

[4] S. Elies, A. Krier, I. R. Cleverley, and K. Singer, Photoluminescence of MBE-grown $InAs_{1-x}Sb_x$ lattice matched to GaSb, J. Phys. D: Appl. Phys. **26**, 159 (1993).

[5] W. L. Sarney, S. P. Svensson, Y. Xu, D. Donetsky, and G. Belenky, Bulk InAsSb with 0.1 eV bandgap on GaAs, J. Appl. Phys. **122**, 025705 (2017).

[6] S. P. Svensson, W. L. Sarney, H. Hier, Y. Lin, D. Wang, D. Donetsky, L. Shterengas, G. Kipshidze, and G. Belenky, Band gap of $InAs_{1-x}Sb_x$ with native lattice constant, Phys. Rev. B **86**, 245205 (2012).

[7] J. D. Kim, D. Wu, J. Wojkowski, J. Piotrowski, J. Xu, and M. Razeghi, Long-wavelength InAsSb photoconductors operated at near room temperatures (200–300 K), Appl. Phys. Lett. **68**, 99 (1996).

[8] W. M. Coderre and J. C. Woolley, Electrical properties of $InAs_{1-x}Sb_x$ alloys, Can. J. Phys. **46**, 1207 (1968).

[9] E. C. Young, F. Wu, A. E. Romanov, D. A. Haeger, S. Nakamura, S. P. Denbaars, D. A. Cohen, and J. S. Speck, Compositionally graded relaxed AlGaN buffers on semipolar Gan for mid-ultraviolet emission, Appl. Phys. Lett. **101**, 142109 (2012).

[10] M. A. Hossain, M. R. Islam, M. K. Hossain, A. Hashimoto, and A. Yamamoto, Dislocation reduction in heteroepitaxial $In_xGa_{1-x}N$ using step-graded interlayer for future solar cells, Mater. Renew. Sustain. Energy **3**, 20 (2013).

[11] M. B. Katz, M. E. Twigg, N. A. Mahadik, C. L. Canedy, and C. A. Affouda, Threading and near-surface dislocations in

InGaSb/AlSb films with blocking and anti-blocking layers, J. Electron. Mater. **45**, 2102 (2016).

[12] A. P. Craig, A. R. J. Marshall, Z.-B. Tian, S. Krishna, and A. Krier, Mid-infrared $InAs_{0.79}Sb_{0.21}$-based Nbn photodetectors with $Al_{0.9}Ga_{0.2}As_{0.1}sb_{0.9}$ barrier layers, and comparisons with $InAs_{0.87}Sb_{0.13}$ p-i-n diodes, both grown on GaAs using interfacial misfit arrays, Appl. Phys. Lett. **103**, 253502 (2013).

[13] G. Belenky, D. Donetsky, G. Kipshidze, D. Wang, L. Shterengas, W. L. Sarney, and S. P. Svensson, Properties of unrelaxed $InAs_{1-x}Sb_x$ alloys grown on compositionally graded buffers, Appl. Phys. Lett. **99**, 141116 (2011).

[14] W. L. Sarney, S. P. Svensson, H. Hier, G. Kipshidze, D. Donetsky, D. Wang, L. Shterengas, and G. Belenky, Structural and luminescent properties of bulk InAsSb, J. Vac. Sci. Technol. B **30**, 02B105 (2012).

[15] A. Kyrtsos, M. Matsubara, and E. Bellotti, Investigation of the band gaps and bowing parameter of $InAs_{1-x}Sb_x$ alloys using the modified Becke-Johnson potential, Phys. Rev. Mater. **4**, 014603 (2020).

[16] B. Guenter, N. Joshi, R. Stoakley, A. Keefe, K. Geary, R. Freeman, J. Hundley, P. Patterson, D. Hammon, G. Herrera, E. Sherman, A. Nowak, R. Schubert, P. Brewer, L. Yang, R. Mott, and G. McKnight, Highly curved image sensors: A practical approach for improved optical performance, Opt. Express **25**, 13010 (2017).

[17] A. Rogalski, P. Martyniuk, and M. Kopytko, Challenges of small-pixel infrared detectors: a review, Rep. Progress Phys. **79**, 046501 (2016).

[18] D. Melati, Y. Grinberg, M. Kamandar Dezfouli, S. Janz, P. Cheben, J. H. Schmid, A. Sánchez-Postigo, and D.-X. Xu, Mapping the global design space of nanophotonic components using machine learning pattern recognition, Nat. Commun. **10**, 4775 (2019).

[19] W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, Deep learning for the design of photonic structures, Nat. Photonics **15**, 77 (2021).

[20] A. Glasmann, A. Kyrtsos, and E. Bellotti, Machine learning for analyzing and characterizing InAsSb-based Nbn photodetectors, Mach. Learn.: Sci. Technol. **2**, 025006 (2021).

[21] M. Smith, *ABAQUS/Standard User's Manual, Version 6.9* (Dassault Systèmes Simulia Corp, United States, 2009).

[22] D. Gerlich, Elastic constants of single-crystal indium arsenide, J. Appl. Phys. **34**, 2915 (1963).

[23] L. J. Slutsky and C. W. Garland, Elastic constants of indium antimonide from 4.2 K to 300 K, Phys. Rev. **113**, 167 (1959).

[24] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994).

[25] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[26] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[27] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces, Phys. Rev. Lett. **100**, 136406 (2008).

[28] J. Heyd, G. E. Scuseria, and M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential, J. Chem. Phys. **118**, 8207 (2003).

[29] J. Heyd, G. E. Scuseria, and M. Ernzerhof, Erratum: "Hybrid functionals based on a screened Coulomb potential" [j. chem. phys. 118, 8207 (2003)], J. Chem. Phys. **124**, 219906 (2006).

[30] A. Alkauskas, P. Broqvist, F. Devynck, and A. Pasquarello, Band Offsets at Semiconductor-Oxide Interfaces from Hybrid Density-Functional Calculations, Phys. Rev. Lett. **101**, 106802 (2008).

[31] A. Kyrtsos, M. Matsubara, and E. Bellotti, First-principles study of the impact of the atomic configuration on the electronic properties of $Al_xGa_{1-x}N$ alloys, Phys. Rev. B **99**, 035201 (2019).

[32] J. P. Perdew and A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, Phys. Rev. B **23**, 5048 (1981).

[33] A. D. Becke and E. R. Johnson, A simple effective potential for exchange, J. Chem. Phys. **124**, 221101 (2006).

[34] F. Tran and P. Blaha, Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential, Phys. Rev. Lett. **102**, 226401 (2009).

[35] D. Koller, F. Tran, and P. Blaha, Improving the modified Becke-Johnson exchange potential, Phys. Rev. B **85**, 155109 (2012).

[36] H. Jiang, Band gaps from the Tran-Blaha modified Becke-Johnson approach: A systematic investigation, J. Chem. Phys. **138**, 134115 (2013).

[37] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (The MIT Press, Cambridge, MA, USA, 2005).

[38] 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 E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. **12**, 2825 (2011).

[39] A. I. D'Souza, E. Robinson, A. C. Ionescu, D. Okerlund, T. J. de Lyon, H. Sharifi, M. Roebuck, D. Yap, R. D. Rajavel, N. Dhar, P. S. Wijewarnasuriya, and C. Grein, Electrooptical characterization of MWIR InAsSb detectors, J. Electron. Mater. **41**, 2671 (2012).

[40] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.15.064008 for further details on the effect of the white noise kernel on the results.