

Wafer scale assessment of dislocation defects on SiC wafers using a convolutional neural network

James C Gallagher^{*}, Nadeemullah A Mahadik, Robert E Stahlbush,
Karl D Hobart and Michael A Mastro

Naval Research Laboratory, Washington DC, United States of America

E-mail: james.c.gallagher26.civ@us.navy.mil

Received 24 February 2025, revised 2 June 2025

Accepted for publication 16 June 2025

Published 1 July 2025



Abstract

Wide bandgap semiconductors, such as SiC, are expected to surpass traditional silicon-based technologies in high-power electronics applications, owing to their potential for reduced size, weight, and enhanced power efficiency. SiC, in particular, has evolved significantly, reaching a maturity that enables its adoption in high-voltage applications including power modules. However, the current state-of-the-art SiC epitaxial wafers face significant challenges, primarily due to the presence of threading dislocations, basal plane dislocations, stacking faults, polytype inclusions, and screw dislocations. These structural imperfections lead to electron and hole carrier scattering, which impairs device performance. Addressing this, the present research introduces a novel approach employing a convolutional neural network (CNN) trained to identify and quantify dislocation defects with high precision. Analysis of the models shows it both highlights the features of interest and delineates the shape, size, and number of the dislocations with higher accuracy than the manual labeling used to train the model, suggesting that CNN has successfully captured the subtle details needed to identify dislocations defects.

Keywords: SiC, wide bandgap semiconductor, machine learning, convolutional neural network (CNN)

1. Introduction

Silicon carbide (SiC) is a wide bandgap semiconductor that has emerged as a superior alternative to silicon in high power applications [2, 8, 12, 13, 38, 39]. Its high critical electric field and electron mobility allow it to operate at high voltages and frequencies [3, 17], and its high thermal conductivity, chemical stability, radiation hardness, and wide bandgap

allow it to operate in high temperature and chemically harsh environments [8, 20, 25]. Although basal plane dislocations (BPDs), micropipes, stacking faults, and polytype inclusions in the epitaxial layers have shown to have a more drastic influence on device performance, these defect types have been significantly reduced in SiC wafers [32]. However, threading mixed dislocations (TMDs) and threading screw dislocations (TSDs) still persist with higher densities and their mitigation has yet to be optimized [14, 27, 31]. TMDs and TSDs may cause decreased breakdown voltage, increased leakage, and premature device failure [9, 28, 42], and they may promote the multiplication of stacking faults via complex interaction mechanisms with existing BPDs [23]. Dislocation type defects are potentially the most common defect in vertical PiN SiC wafers used in high power electronics [2, 5, 19, 32, 37, 38]. Vertical PiN SiC structures typically comprise a heavily doped

^{*} Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

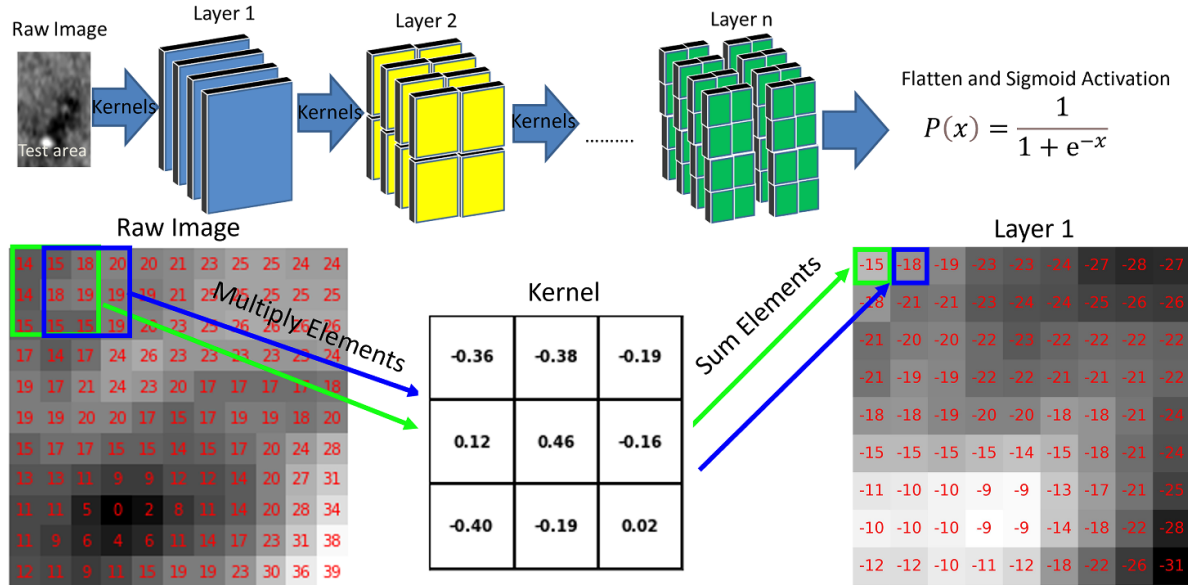


Figure 1. A basic diagram illustrating the operation of a CNN is presented. The process begins with a raw image, which undergoes transformations through multiple layers via the application of several convolutional kernels operating in parallel. These kernels slide across the entire image to perform element-wise multiplication and summation at each location. The output from these convolutional layers results in numerous feature maps (subimages), which capture specific features that the network aims to detect. Each feature map is subsequently flattened and processed through one or more fully connected layers. The final layer often includes a sigmoid activation function for binary classification tasks, outputting probabilities. The diagram specifically demonstrates the application of a 3×3 kernel on an 11×11 pixel image, resulting in a 9×9 feature map in the first convolutional layer. Note that this depiction is a simplification; actual CNN architectures incorporate additional elements such as activation functions (e.g. ReLU) and may use kernels of various sizes. Optimal performance of a CNN involves tuning several hyperparameters.

N-type substrate upon which a lightly doped N-type drift layer (sometimes referred to as an intrinsic layer) is epitaxially grown, followed by a heavily doped P-type cap layer. The presence of BPDs in the substrate can cause the formation of TMDs and TSDs within the drift and P-type layers [24]. This results in materials exhibiting lower conduction loss and reduce the on-state resistance, resulting in decreased switching losses [12]. Despite the advantages, widespread adoption of wide bandgap semiconductor technology is hampered by the presence of several different types of defects that significantly affect device performance including stacking faults, dislocations, inclusions, point defects, and grain boundaries [10]. Though methods for detecting these defects are well known, wafer scale quantification requires a large, computationally burdensome datasets [5]. Therefore, the most practical inspection algorithms use measurement techniques that are quick and are computationally efficient. Techniques that require a set of measurement values at every spatial point of measurement on the wafer are not as ideal as wafer scale imaging techniques requiring only a single point of measurement, such as optical profilometry, luminescence imaging, or x-ray topography (XRT).

With imaging techniques, the computational quantification of these dislocations poses a significant challenge, as the variety of defect shapes and sizes can make counting defects with traditional computational image analysis difficult unless the defects produce a high signal-to-noise ratio, such as BPDs in

ultraviolet photoluminescence (UVPL) images of SiC [16]. Many defects, such as threading dislocations in XRT images, exhibit subtle and inconsistent changes in pixel values across different images. This limitation has prompted the exploration of advanced machine learning techniques [36]. In particular, deep learning classifications often outperform traditional methods [21, 22, 29, 33, 34]. Specifically, convolutional neural networks (CNNs) have shown promise for faster detection and counting of such defects with higher accuracy and has been used to detect imperfections in semiconductor devices [4, 7, 11, 26, 35, 39, 40]. This method is effective at image recognition because it fits several kernels that act as filters which highlight desired features in an image. A detailed diagram of a basic CNN is shown in figure 1. Previous studies have developed CNNs that detect defects in semiconductors using both scanning electron microscope images and optical images [6, 30]. This research aims to advance wide bandgap semiconducting technology by training a CNN to accurately locate and quantify the screw and threading dislocations. This was achieved by training CNNs on wafer-scale images to highlight defects on the wafer with a particular focus on detecting dislocation-type defects in SiC. The methodology demonstrates the potential to accelerate the development and significantly improve the quality control process in wide bandgap semiconductor manufacturing, thereby enhancing the performance and reliability of future power electronics.

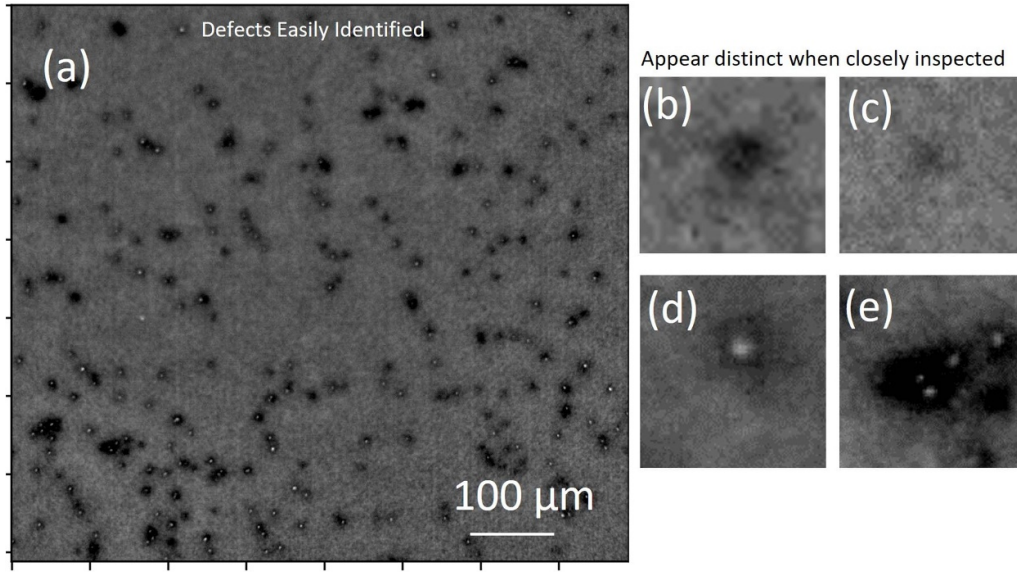


Figure 2. (a) A UVPL image of SiC containing several dislocation defects is shown. (b)–(e) Close up views of four of the defects shows large variations in appearance that are difficult for a tradition pattern matching techniques to address.

2. Experimental techniques

TMDs and TSDs in SiC can be imaged with both UVPL and XRT. This paper explores the application of machine learning models utilizing a combination of both these techniques. UVPL images were obtained by leveraging a custom-designed apparatus tailored for high-resolution spectral imaging. UV excitation was achieved using a 4 W, 355 nm laser. This specific wavelength effectively excites carriers in the material, causing photoluminescence (PL), and allowing for detailed examination of extended defects. The resulting PL was captured using a CCD detector with 665 nm long pass filter. The UV laser spot has an approximate Gaussian diameter of 4 mm, which is large enough overfill the area imaged by the CCD 2.6 mm \times 2.6 mm field of view for uniform illumination. The CCD has an approximate . The UV irradiation power density is $\sim 100 \text{ W cm}^{-2}$. The imaging process followed a step-and-repeat pattern across the entire wafer. This method allowed for the acquisition of discrete PL images, which were subsequently processed and assembled into a complete, high-resolution image of the wafer. The system's high-resolution capability, approximately (2 μm used in this study), enabled the detection of fine structural details across wafers with diameters up to 200 mm. TMD/TSDs are observed as dark spots in the near UVPL images. Figure 2(a) shows a typical UVPL image containing several TMD/TSDs. Though the defects can be readily identified by an expert examining enlarged images of the TMDs and TSDs in figures 2(b)–(e) shows the defects have different contrasts, shapes, and sizes. Therefore using a traditional pattern matching technique, e.g. based on intensity variations or gradients, would encounter many challenges because they require accounting for every possible configuration a defect could present itself. Additionally, if conditions change slightly it could require extensive modification to the algorithm while with machine learning models only a

sufficient number of examples of each case needs to be shown in training data and updates to the model need only require additional training with the new examples.

The Rigaku XRT Micron System used in this study is equipped with a high brilliance 70 μm microfocus dual wavelength Cu/Mo rotating anode, multifunctional goniometer capable of 200 mm diameter wafers. It features a high-resolution detector with a measurement down to a pixel size of 2.16 μm . This system is versatile in its imaging modes, offering reflection geometry, transmission geometry, and specialized slit adapters for depth-resolved section topography. These capabilities make it a robust tool for detailed material characterization. The (0008) reflection of SiC highlights TMDs and TSDs, as demonstrated through Burgers vector analysis [15], which is further examined in this paper. XRT images are less intense near regions where the intensity is lower, however the XRT images are inverted so white spots are the TMD/TSDs. Similar to UVPL, TMD/TSD are readily identified by human eye, however, the defects may look very different depending on the contrast levels (see figure 3), which can also make traditional pattern matching difficult.

3. Computation analysis

Training a CNN requires organization of input and output data. This project utilized 10 s of high-resolution UVPL and XRT images of SiC wafers, with a resolution of 2–5 $\mu\text{m}/\text{pixel}$, covering areas from ~ 1 to 10 square millimeters obtained from random sections of wafers. When the model is trained 80% of the images chosen at random were used to train the CNN, while the other 20% were used to validate it and calculate the accuracy. For each image, every single permutation of a 40 \times 40 pixel area was chosen creating a large dataset. The x – y coordinates of defects and a characteristic

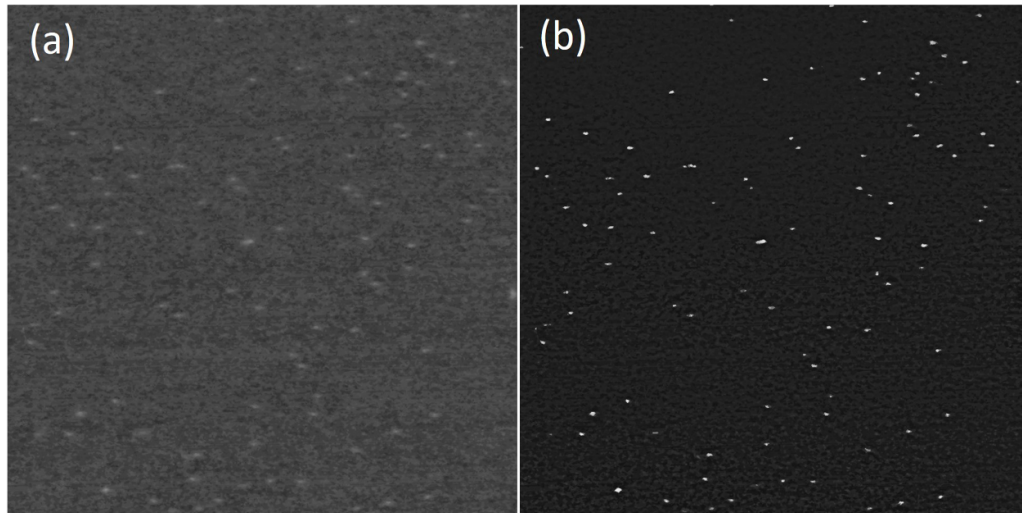


Figure 3. An XRT image of a section taken at low contrast (a) and high contrast (b). The high contrast is obtained using various custom designed image processing algorithms.

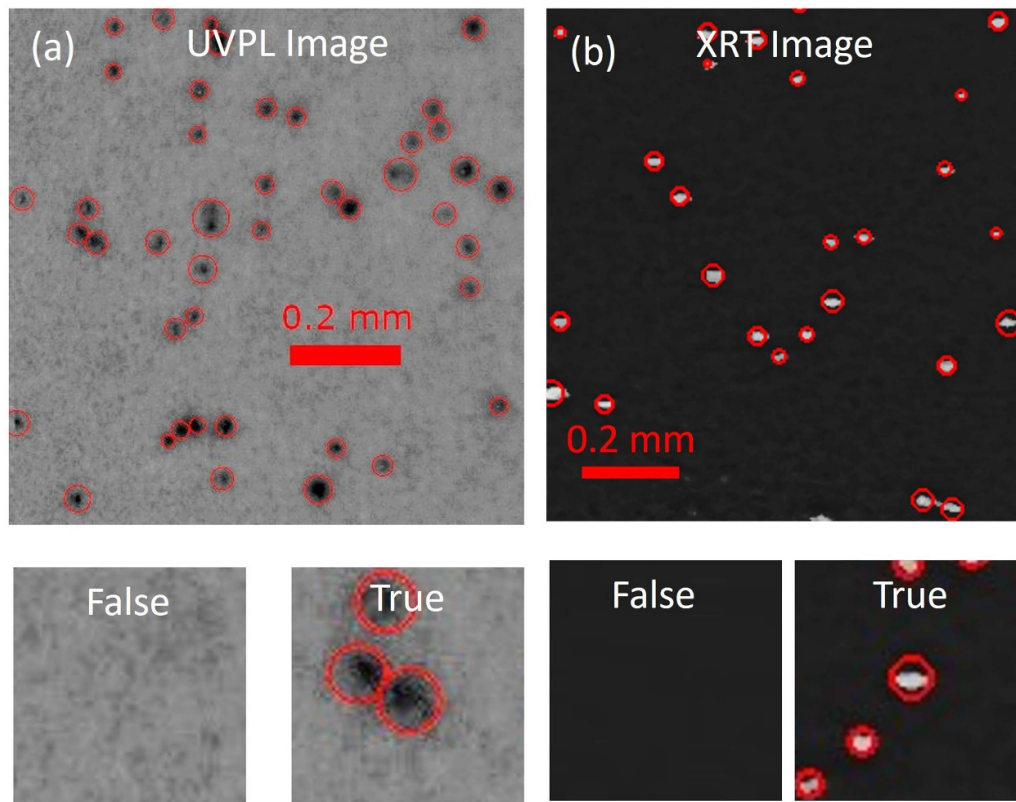


Figure 4. An example UVPL image (a) and XRT image (b) from the training dataset with dislocation defects as black dots of variable intensity with defects manually cataloged using red circles. Below the large images are two small 40×40 pixel regions. Showing an example of a ‘true’ and ‘false’ label. A ‘true’ label is assigned if a defect is close to the center of the image.

radius were manually recorded (as shown in figure 4) as ‘true’ (positive results) if the center pixels were inside the circle, creating a binary classification problem for the CNN. Since ‘false’ values vastly outnumbered ‘true’ values, every ‘true’ value was augmented by rotating and reflecting the image to

increase the number of ‘true’ datasets by a factor of 8. The CNN was trained and used with the TensorFlow [1] package in Python using two NVIDIA RTX™ 4090 graphics cards. TensorFlow optimizes the CNN’s kernels (see figure 1) by using the Adaptive Moment Estimation (Adam) to minimize

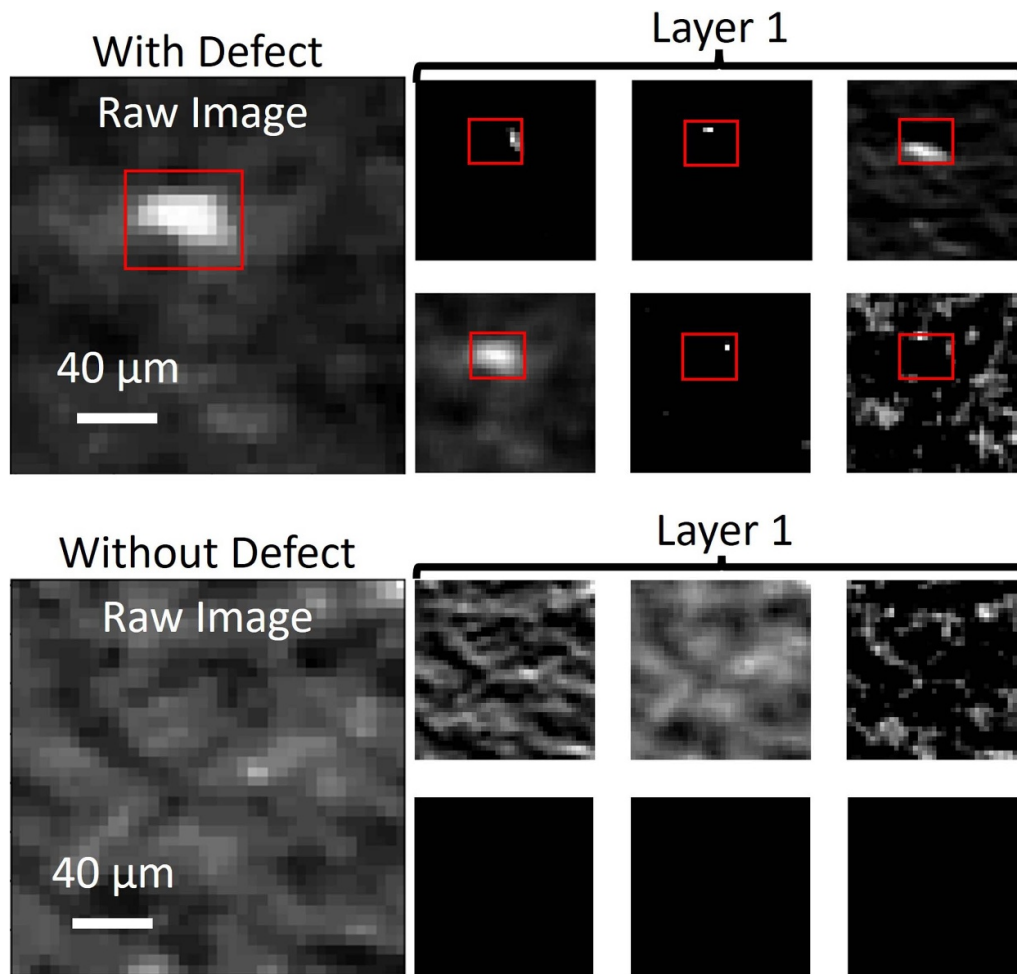


Figure 5. Two XRT images are shown one with defects and another without. The results of applying 6 kernels in the first layer of the CNN shows the defect is being highlighted by the kernels, with regions without defects being set to zero by the kernels. The defect is highlighted with a red box which is translated into the same location in the first layer images.

the binary cross entropy loss [18], which consistently produced an accuracy of the validation data set greater than 95%. The model can map every pixel on a 8 in. wafer in approximately four hours if a resolution of $2\ \mu\text{m}/\text{pixel}$ is used.

4. Results and discussion

Though the accuracy of the model was $>95\%$, effective evaluation of a machine learning model's effectiveness requires in depth analysis because accuracy is probability of a pixel being inside a manually drawn circle on the dislocations. Since 'false' (contains no defect) values are much more common than 'true' values, the accuracy is inflated by a high true negative rate. The CNN primary purpose is to locate, count the number of TMDs and TSDs, and estimate the diameter and shape of the defects; therefore, its effectiveness will be related to its utility in those applications.

Since a CNN is fitted by optimizing several kernels, it is important to examine the dominate kernels in the first convolutional layer [41]. Figure 5 displays two small regions of a wafer: one with a TMD/TSD and the other contains no dislocations. Included are the first layer feature maps constructed using the six kernels with the highest impact on the prediction. In the image containing a TMD/TSD, the defect is highlighted in the first layer images. The three in the top row and a fourth in the middle of the bottom row are mostly black images with small light regions near the edge of the box. These four kernels are likely fit to identify the boundaries of the defect. The other two kernels highlight the defect which a light or dark region and are likely used to further refine the defect's boundaries in other layers of the CNN. When the defect is absent, nothing is highlighted, and only three kernels produce a non blank image. This suggests the kernels in the first layer are optimized to quickly classify a 'false' reading using the first layer alone. Applying all the layers of the CNN eventually produces a probability map at each pixel. Figure 6 shows an example of the models being applied to a UVPL image (figure 6(a)) and an

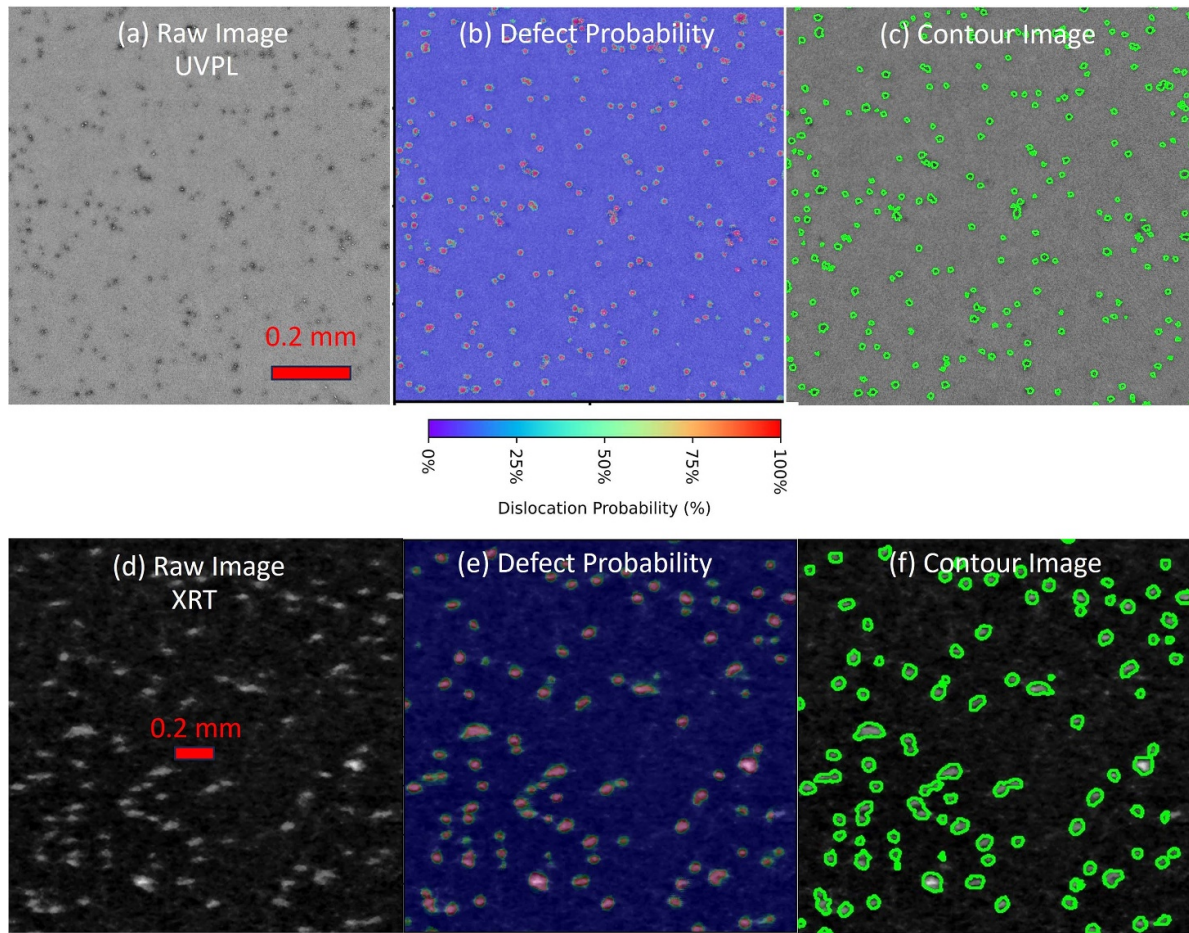


Figure 6. Examples of the CNN highlighting defects on both UVPL images (a)–(c) and XRT images (d)–(f) a raw UVPL (a) and XRT (d) images with several threading dislocations. Applying the CNN can calculate the probability of each pixel being part of the defect, as displayed with a semitransparent color overlay on the UVPL (b) and XRT (e). Applying a contouring algorithm set at 50% dislocation probability surrounds individual defects in both UVPL (c) and XRT (f) images. These typically surround exactly one individual defect, thus counting the number of contour regions should yield an accurate count of the number of dislocations.

XRT image (figure 6(d)). Both these images have a large number of TMD/TSDs. The model can calculate the probability of each pixel belonging to a defect. Figures 6(b) and (e), demonstrate that consistent higher probabilities are assigned to pixels on the defects. To quantify the number of defects and classify each defect's size and shape, an image segmentation technique is needed. In this work, the probability map was converted to a segmentation map by application of a contouring algorithm to draw contour lines at 50% probability (figures 6(c) and (f)) to highlight most of the defects and classifies their sizes and shapes accurately.

Since the primary goal is to examine if the model can count the number of dislocations, the model was applied to distinct test data in figure 7. The figure shows contour maps drawn with the model (figures 7(a)–(d)), with the defects manually counted (figures 7(e)–(h)). Investigating the accuracy of the model is performed by analyzing the location of False Positives and False Negatives that are highlighted in figures 7(i)–(l). The discrepancy between the manually counted value and the model-predicted value arises from two main factors: the challenge of distinguishing overlapping defects and inaccuracies in the manual labeling of the data. This is highlighted by the images

with lower defect concentrations (figures 7(a) and (b)) that have a smaller percent difference as the overlapping defects are less common and errors in manual counting are less frequent. This is especially highlighted in figures 7(d), (h) and (l) where green dots labeled as false positives by the model are actually defects that the model detected but were missed in manual labeling, and several defects labeled as a cluster are highlighted as a single large defect. Most of the false positives and false negatives are not from mislabeled defects as they are near true positive regions but instead from a difference in the size of the defects determined manually vs. by the machine learning model. Studying a close-up region with four defects (figure 8(a)) demonstrates where the discrepancy is in more detail. Since the manual labels are circular, they do not capture the size and shape perfectly as demonstrated in figure 8(b). Because the model picks up the shapes more accurately, when evaluating the accuracy (figure 8(c)) all points inside the manual circles, which are not part of the defects are incorrectly labeled as false negatives (red) during testing; likewise, the portions of the defect outside of the manually drawn circles are incorrectly labeled as false positives (green). This was also observed when on samples from the training dataset

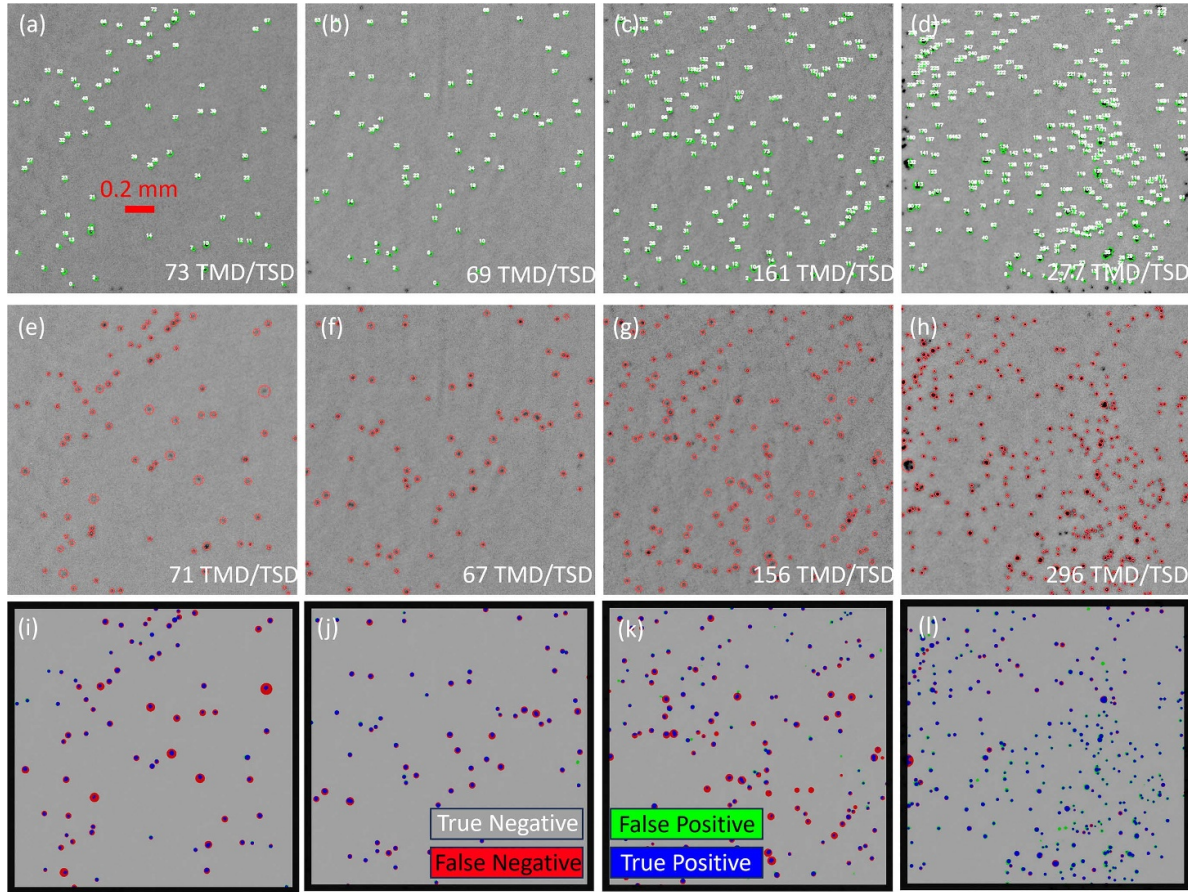


Figure 7. Contour images of four UVPL (all 2×2 mm regions) images in a validation dataset with the TMD/TSD counted by the algorithm circled in green (a)–(d). The corresponding manually counted labeled are highlighted using red circles (e)–(h). All images have the number of TMD/TSD counted is listed in the bottom right corner. Note: dislocations within 25 pixels ($50 \mu\text{m}$) of the edge are excluded from the count in all images. The direct results of the CNN applied to each pixel are (i)–(l) highlighting the pixels that are true negatives (gray), true positives (blue), false positives (green), and false negatives (red).

(figures 8(d)–(f)). This is an interesting result as the defect assignment region in the training datasets for the machine learning models were circular demonstrating that the CNN based model is more accurate than the manual labeling used in training, thus providing evidence that the accuracy tested of the model is higher than the computation test shows.

Because the methods are able to accurately model the size of the defects, a histogram plot (figure 9) was constructed. The Burgers vector of TMDs and TSD are nanometers

wide but perturb the crystal at a distance of several microns from the core. The machine learning model shows that these dislocations show an influence in crystalline properties (detected from XRT) over a $53 \pm 14 \mu\text{m}$ diameter, while the optical properties (determined from UVPL) have an influence of $24 \pm 8 \mu\text{m}$. Results were obtained from the analysis of several SiC wafers for obtaining good statistics. This highlights one potential benefit for utilizing a well-trained CNN for defect detection in semiconductor wafers.

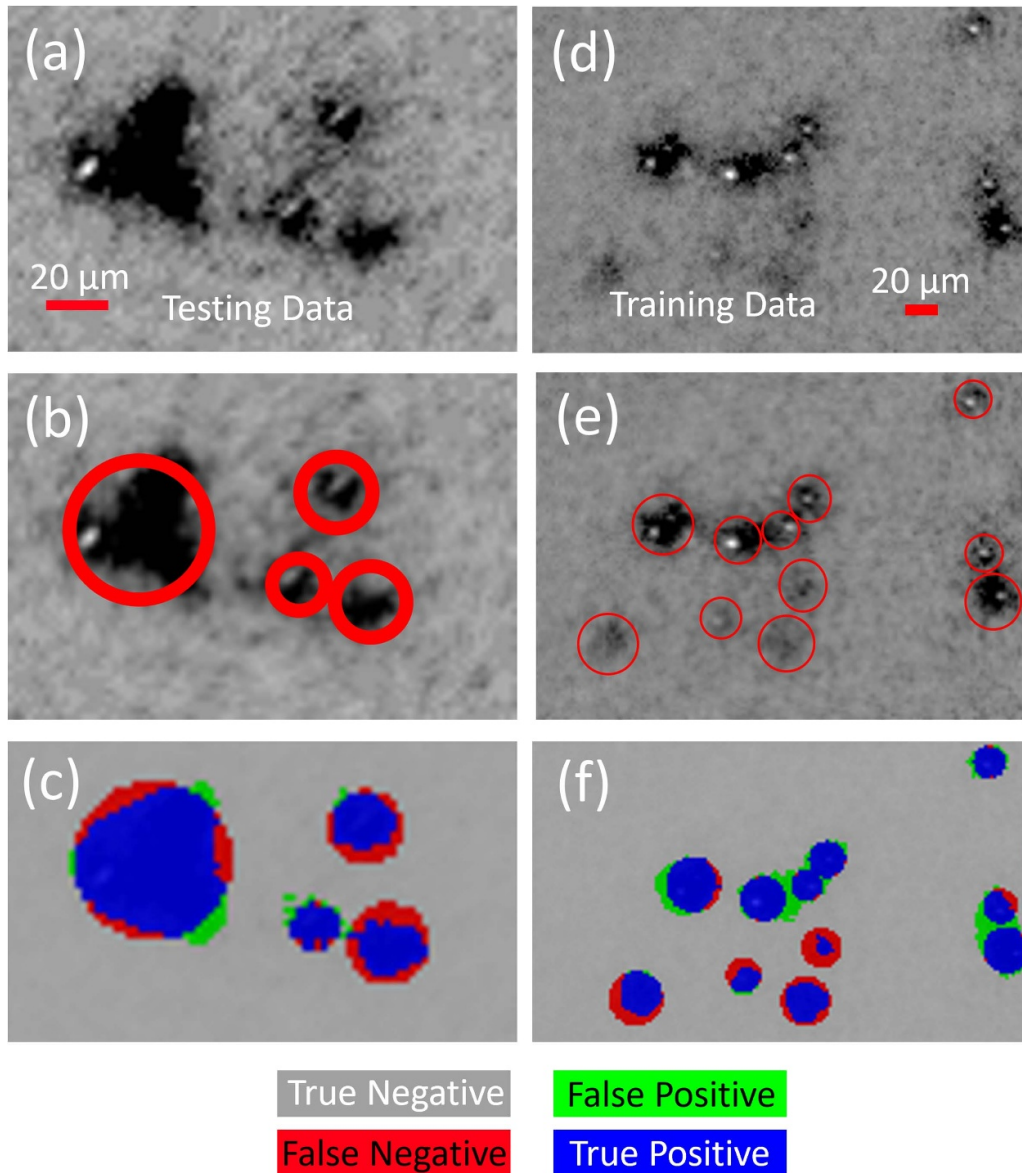


Figure 8. (a) A close up view of 4 dislocations. The size and shape of the dislocations are manually labeled in (b). Every pixel is tested using the machine learning model (c) shows which points are registered as false positives, false negatives, true positives, and true negatives. (d)–(f) A similar close up view of 11 dislocations on a training dataset is applied to show the model is consistent and defines the shape of the dislocations capturing all the intricacies even though the manual model used to train the dataset assumed all defects are circular.

5. Conclusion

This study demonstrated that a CNN can count the number of threading and screw dislocations using either XRT or UVPL images with high accuracy. From a qualitative standpoint, the models highlights most of the defects visible to an expert. From a quantitative standpoint, the computed quantity was within 7% of the manually counted value, with errors arising from the challenge of differentiating overlapping defects and imperfections in the manual labeling of images. Because the model finds dislocations and defines the shape of the defects

with superior performance to manual labeling, the accuracy of the model is likely higher than demonstrated by the program.

One striking feature of this study is that it achieved precise predictions using less than 50 images of manually cataloged dislocations. Additionally, though UVPL and XRT were used in this study on SiC, this research could be expanded to study similar types of defects on other semiconductors including Si, GaAs, GaN, Ga_2O_3 , diamond, AlN, and BN providing a method for detecting the defects in those materials is well established and a training dataset is labeled.

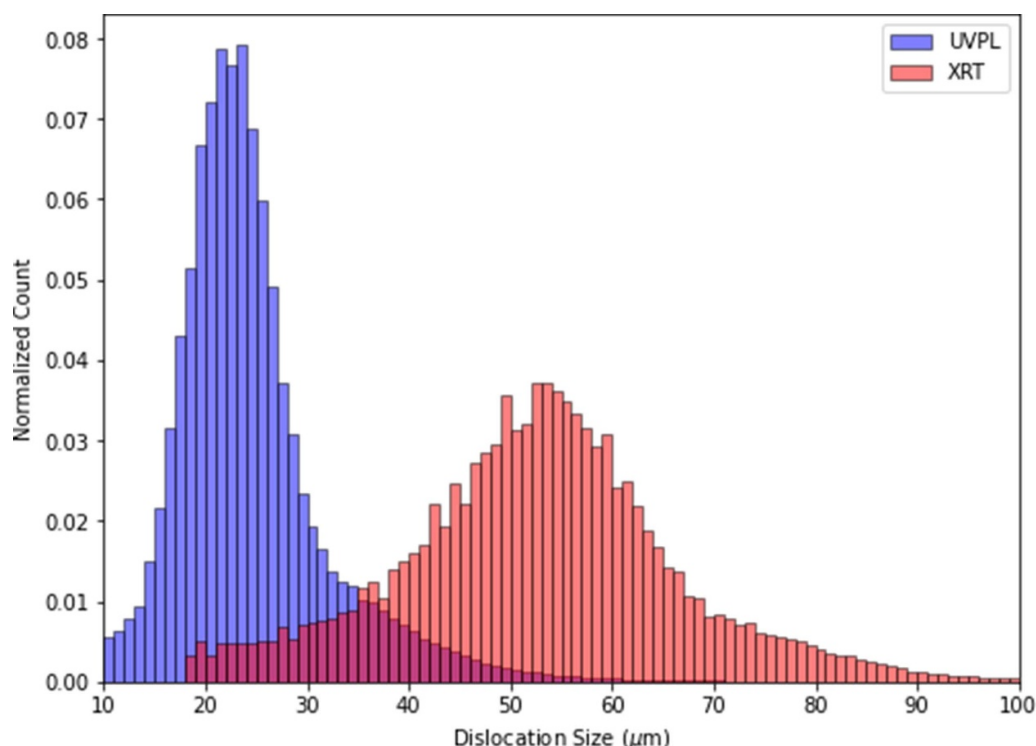


Figure 9. A histogram plot showing the distribution of size of TMD/TSD defects highlighted using the machine learning model. Size was determined by approximating each contour (see figures 6(c) and (f) or figures 7 (a)–(d) as a circle and recording the diameter. Approximately 6.8 million dislocations detected using UVPL and 0.8 million dislocations detected from XRT are included. The plots were normalized so the area under the curve is equal to one.

Data availability statement

The source code is a trade secret of the U S Naval Research Laboratory and cannot be legally disclosed. The data is controlled unclassified information and is not cleared for public release.

Acknowledgments

Work performed at the U S Naval Research Lab was sponsored by the Office of Naval Research and the DOD Microelectronics Commons program.

References

- [1] Abadi M et al 2016 44 arXiv:1605.08695
- [2] Armstrong K O, Das S and Cresko J 2016 2016 IEEE 4th Workshop on Wide Bandgap Power Devices and Applications (WiPDA) (IEEE) (available at: <http://ieeexplore.ieee.org/document/7799949/>) pp 259–64
- [3] Baliga B 1989 IEEE Electron Device Lett. **10** 455–7
- [4] Bimrose M V, Hu T, McGregor D J, Wang J, Tawfik S, Shao C, Liu Z and King W P 2024 Manuf. Lett. **41** 1216–24
- [5] Chen P C, Miao W C, Ahmed T, Pan Y Y, Lin C L, Chen S C, Kuo H C, Tsui B Y and Lien D H 2022 Nanoscale Res. Lett. **17** 30
- [6] de la Rosa F L, Sanchez-Reolid R, Gomez-Sirvent J L, Morales R and Fernandez-Caballero A 2021 Appl. Sci. **11** 9508
- [7] Devika B and George N 2019 2019 10th Int. Conf. on Computing, Communication and Networking Technologies (ICCCNT) (IEEE) pp 1–6
- [8] Papanasam E, Prashanth Kumar B, Chanthini B, Manikandan E and Agarwal L 2022 Silicon **14** 12887–900
- [9] Fiorenza P, Alessandrino M S, Carbone B, Martino C D, Russo A, Saggio M, Venuto C, Zanetti E, Giannazzo F and Roccaforte F 2020 Nanotechnology **31** 125203
- [10] Setera B and Christou A 2022 Electron. **11** 10
- [11] Fukushima K 1969 IEEE Trans. Syst. Sci. Cybern. **5** 322–33
- [12] Gachovska T K and Hudgins J L 2024 Power Electronics Handbook 5th edn ed M H Rashid (Elsevier) pp 87–150
- [13] Giovanni F D 2023 Chips **2** 209–22
- [14] Ha S, Mieszkowski P, Skowronski M and Rowland L 2002 J. Cryst. Growth **244** 257–66
- [15] Harada S, Matsubara Y, Hayashi S, Kawase M, Seo K, Mizutani S, Mizutani Y, Mizutani S and Murayama K 2025 AIP Adv. **15** 025208
- [16] Harada S, Tsujimori K and Matsushita Y 2022 J. Electron. Mater. **51** 243–8
- [17] Ivkovic M 2022 The undeniable advantages of sic technology over si. (available at: www.avnet.com/wps/portal/us/resources/article/the-undeniable-advantages-of-sic-technology-over-si/)
- [18] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)
- [19] Kizilyalli I C, Edwards A P, Aktas O, Prunty T and Bour D 2015 IEEE Trans. Electron Devices **62** 414–22
- [20] Lebedev A A, Kozlovski V V, Davydovskaya K S and Levinshtein M E 2021 Materials **14** 4976
- [21] Leonard R, Conrad M, van Brunt E, Giles J, Hutchins E and Balkas E 2020 Mater. Sci. Forum **1004** 321–7
- [22] Leonard R, Conrad M, van Brunt E, Witry J and Balkas E 2023 Defect Diffus. Forum **426** 3–9

- [23] Mahadik N, Dudley M, Raghothamachar B, Chen Z, Stahlbush R, Hinojosa M, Lelis A and Sung W 2024 *Mater. Des.* **248** 113435
- [24] Myers-Ward R L, Mahadik N A, Wheeler V D, Nyakiti L O, Stahlbush R E, Imhoff E A, Hobart K D, Eddy C R and Gaskill D K 2014 *Cryst. Growth Des.* **14** 5331–8
- [25] Ojha G P, Kang G W, Kuk Y S, Hwang Y E, Kwon O H, Pant B, Acharya J, Park Y W and Park M 2022 *Nanomaterials* **13** 150
- [26] OShea K and Nash R 2015 arXiv:1511.08458
- [27] ou Lu S, Xu B, Chen H, Hang W, Wang R, Yuan J, Pi X, Yang D and Han X 2024 *J. Cryst. Growth* **627** 127526
- [28] Senzaki J, Kojima K, Kato T, Shimozato A and Fukuda K 2006 *Appl. Phys. Lett.* **89** 022909
- [29] Sheng H, Cheng K, Jin X, Jiang X, Dong C and Han T 2024 *AIP Adv.* **14** 045329
- [30] Shi H, Jin Z, Tang W, Wang J, Jiang K, Xu M, Xia W and Xu X 2023 *Knowl.-Based Syst.* **280** 110994
- [31] Stahlbush R E, VanMil B L, Myers-Ward R L, Lew K K, Gaskill D K and Eddy C R 2009 *Appl. Phys. Lett.* **94** 041916
- [32] Stahlbush R, Mahadik N, Bonanno P, Soto J, Odekirk B, Sung W and Agarwal A 2022 *IEEE Int. Reliability Physics Symp. (IRPS)* vol 2022 p 65–1– 65–6 (IEEE) (available at: <https://ieeexplore.ieee.org/document/9764473/>)
- [33] Taha K 2023 arXiv:2310.10705
- [34] Taha K 2025 *J. Intell. Manuf.* (<https://doi.org/10.1007/s10845-024-02521-0>)
- [35] Tajudin P N A and Shapiai M I 2021 *IOP Conf. Series: Materials Science and Engineering* 1176 012034
- [36] Theodosiou T, Rapti A, Papageorgiou K, Tziolas T, Papageorgiou E, Dimitriou N, Margetis G and Tzovaras D 2023 *Proc. Comput. Sci.* **217** 570–83
- [37] Tsao J Y et al 2018 *Adv. Electron. Mater.* **4** 1600501
- [38] Varley J B, Shen B and Higashiwaki M 2022 *J. Appl. Phys.* **131** 230401
- [39] Wang Y, Gao L, Gao Y and Li X 2021 *Robot. Comput.-Integr. Manuf.* **68** 102083
- [40] Wang Z, Li L and Yao Y 2021 *IEEE Trans. Electron Devices* **68** 2212–9
- [41] Yamashita R, Nishio M, Do R K G and Togashi K 2018 *Insights Into Imaging* **9** 611–29
- [42] Zhao F, Islam M M, Daas B K and Sudarshan T S 2010 *Mater. Lett.* **64** 281–3