

Implementation of Large Scale Deep Learning Non-Destructive Methods for Characterizing 4H-SiC Materials

Robert T. Leonard^{1,a*}, Matthew Conrad^{1,b}, Edward VanBrunt^{1,c}, Jason Witry^{1,d}, Elif Balkas^{1,e}

¹Wolfspeed, Inc., 4600 Silicon Dr. Durham, NC 27703 USA

^aRobert.Leonard@wolfspeed.com, ^bMatthew.Conrad@wolfspeed.com,

^cEdward.VanBrunt@wolfspeed.com, ^dJason.Witry@wolfspeed.com, ^eElif.Balkas@wolfspeed.com

Keywords: KOH Etching, Machine Learning (ML), Deep Learning (DL), Non-Destructive Technique (NDT), DL-NDT, Photoluminescence Imaging (PL), Dislocation Characterization, Labelling.

Abstract. A whole wafer method for industrial high volume, non-destructive characterizing of extended defects is demonstrated for 150 mm and 200 mm 4H-SiC wafers. Deep learning (DL) coupled with non-destructive techniques (NDT, DL-NDT) involving high volume, fast optical microscopy methods correlates industry accepted chemistry and physics-based etch and diffraction techniques for defect characterization. The application of the DL-NDT method is shown to reproduce defect distributions achieved by accepted etch techniques for extended defects of threading dislocations (TD), basal plane dislocations (BPD), and threading screw dislocations (TSD). An example of algorithm development is described to show progress toward implementing the method, as well as DL-NDT defect density compared to etch density for multiple wafers. The development status for implementing this technique for large-scale industrial wafer production includes etch validation of the results to ensure the technique is consistent and reliable. The ability to use this non-destructive technique ultimately will result in better correlation with device behavior and provide feedback to crystal growth processes to improve substrate wafers, while reducing the need for etch methods.

Introduction

The use of silicon carbide accelerates the automotive industry's transformation to electric vehicles, enabling greater system efficiency and performance in electric cars with longer range and faster charging, while reducing cost, lowering weight, and conserving space [1]. The continuing advancement of state-of-the-art SiC substrates fuels this transformation. Increased device yield and manufacturing efficiency require continual feedback of extended defects, relying on consistent and high throughput detection methods. Defect information feedback for both success and failure, from crystal growth through device fabrication is key to process improvement.

Extended defects such as dislocations and micropipes traditionally have been characterized using destructive etching methods. Implementation of non-destructive techniques (NDT) not only avoids destroying the substrate wafer but enables the ability to observe the same substrate at subsequent processing steps. This also allows for more learning-based on self-consistent data sets. NDT techniques, such as synchrotron x-ray topography (SXRT), have been a staple for research and development with physics-based definitions of defects correlating to etch features, but throughput can be slow and confined to somewhat small sample sets based on tool availability [2]. High throughput scanning optical microscopy equipped for surface and photoluminescence (PL) imaging has been successfully used both to observe a non-destructive PL signal and to image etched surfaces of the same substrates with resolutions on the <5 µm scale, producing large volumes of images to observe the substrate [3]. With the rapidly advancing state of machine and deep learning (DL) techniques, large amounts of information can be used to train highly sophisticated algorithms to identify both subtle and obvious features in images and data sets providing feedback for improving substrates and processing [4]. Wolfspeed reported on combining deep convolutional neural networks (DCNN) correlating PL images of dislocations labeled from etch wafer data to non-destructively identify

dislocations in 4H-SiC [3]. This algorithm was based on previous work detailing the correlation between SiC etch pit and SXRT features and an extension of these features to abstract definitions for image matching based counting methods [5]. Threading mixed dislocations (TMD), containing both screw and edge components, have been observed in 4H-SiC [6], but were not included in the differentiation between threading edge dislocations (TED) and threading screw dislocations (TSD) in the counting definitions.

The success of this initial work enabled the transition of DL-NDT extended defect counting into a fully deployed and sustained deep learning (DL) defect detection system for 4H-SiC wafers to inform research and development endeavors. Defect counting in SiC is not perfect due to the complex nature of dislocations and how they present themselves with different etch conditions, optical or x-ray diffraction conditions. Deep learning non-destructive techniques are uniquely suited for classification of defects based on subtle shape or contrast variations with local contextual features. Precision and recall in the DL-NDT algorithm is dependent on how well the defect is initially defined and the ability to consistently label the defect, along with number of labeled images, architecture used, model size, and training method. Validation of the algorithm is needed to determine the overall reliability and whether more training is needed. Deep learning coupled with non-destructive fast optical microscopy methods is shown for whole wafer 150 mm and 200 mm 4H-SiC wafers to correlate industry accepted chemistry and physics based etch and diffraction techniques for defect characterization.

Methods

Implementing high volume DL-NDT for detecting dislocation defects involves training a robust ML algorithm, consistently imaging the wafers to stay within the algorithm conditions, and managing the output to both report the characterized defects and to ensure a validated algorithm.

ML algorithm development. Reducing the DL-NDT method into practice requires the following steps [4]. Acquisition of PL images of CMP surfaces for both Si- and C-face 4° off axis 4H-SiC n-type wafers by Lasertec Sica88 photoluminescence and confocal optical microscopy. Etching the surfaces of the wafers in KOH (Si-face) revealing TD and BPD etch features and NaOH/ KOH (C-face) to reveal TSD etch features [7]. Sica88 Nomarski imaging of both etched faces with automatic feature detection produced categorical defect labels. The unetched and etched images are then aligned to produce a labeled dataset of images selected with multiple features of interest, which is split into a training set and validation set. The PL images in the training set are correlated with etch labels using deep convolutional neural net (DCNN) machine learning (ML) methods to produce the ML algorithm. Finally, the validation set of images are used to evaluate the algorithm fidelity of the labeled image data.

Imaging conditions/ application of algorithm. Imaging conditions are managed routinely by measuring gauge wafers to ensure proper operating conditions of the Lasertec PL microscopy and replacing the optical components as needed. CMP wafer surfaces are used to reduce noise in the signal and to ensure measurement of bulk defects, differentiated from surface defects. After all of the PL images of each wafer are digitally captured and saved independently of the microscope hardware and software, the Wolfsped algorithm is applied to the images for defect extraction. The detected defects are then recorded by category and position to produce wafer density and defect maps.

Algorithm management. At this point in implementing NDT for dislocations defects, the defect density is validated by conventional etching to maintain high confidence in the development of DL-NDT algorithmic defect detection. Resultant detected defects produced through the application of the algorithm to sampled images apart from the dataset are manually compared for precision in labeling. Manual validation of the NDT method is carried out to ensure that the correct defects are detected. The associated ground truth (etched image labels) are compared to the algorithmic output to determine true positives (tp), false positives (fp) and false negatives (fn). These, in turn are used to determine the precision = $tp/(tp+fp)$ and recall = $tp/(tp+fn)$ of the feature detection. Finally, the

associated F1 score Eq. (1) is used to assess the effectiveness of the algorithm [8]. Re-training the ML algorithm is done as needed to ensure/ increase defect labeling reliability.

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{tp}{tp + \frac{1}{2}(fp + fn)} \quad (1)$$

Results and Discussion

A successful example of applying the described methods for full 4H-SiC 150 mm wafer DL-NDT dislocation detection and counting is shown in Fig. 1, illustrating the ability of the technique to determine distribution and densities of defects over the entire wafer. For clarity in seeing the different defect types, the images in Fig. 1 (a)-(c) represent the TD, BPD and TSD's detected by the ML algorithm on the Si face of the wafer. For the threading dislocation defects, image contrast in Fig. 1(a) and (b) is from constant size white dots representing the plotted output of the ML algorithm defect position on a black background, referenced to the wafer center position. For the TSD image, Fig. 1(c), the defect markers are slightly larger circles to increase the ability to show the defect position in the output image. The output image can also be generated to show all defects at once, or any combination of defect channel as needed.

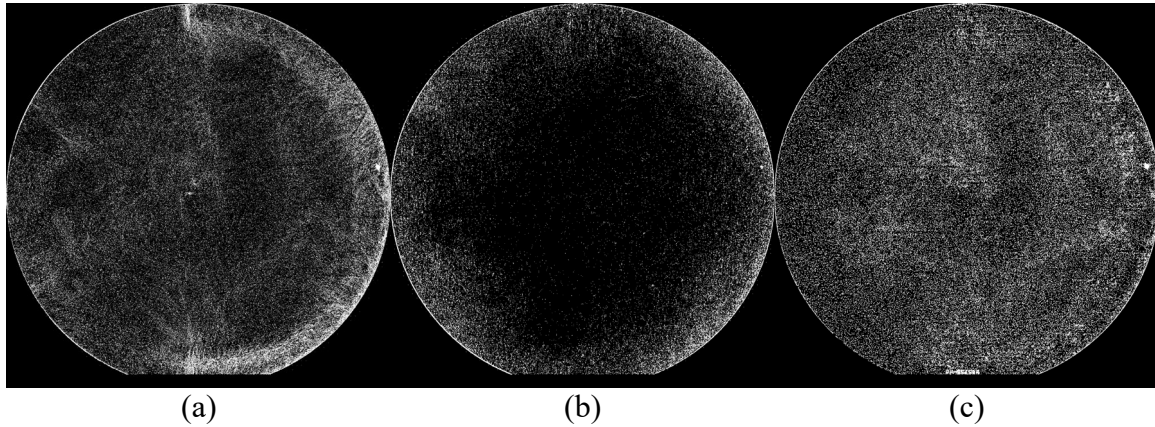


Fig. 1. 150 mm 4H-SiC PL-NDT dislocation defect detection and densities for a) TD, b) BPD and c) TSD. PL imaging on the Si-face.

For the individual maps of Fig. 1, the TD's in Fig. 1(a) are easily distinguished over the entire wafer, the BPD's of Fig. 1(b) are at lower densities distributed as shown, and the TSD's of Fig. 1(c) are distributed over the whole wafer.

Defect densities from DL-NDT method compared with etch data for the same wafer are shown in Table 1. The DL-NDT TD and BPD densities are slightly overcounted with respect to etch values, which may be affected by etch bath condition, wafer cleanliness, automatic defect labelling algorithm and size of etch defect. For the TSD density, the table shows the application of two different algorithms: A and B. The algorithm A was initially used to count TSD's, resulting in overcounting of the defects. To produce a more reliable model, images from the overcounted examples coupled with their respective optical microscopy auto-labeled etched features were added to the dataset and the algorithm was re-trained, resulting in algorithm B. Comparing the 2 algorithms for the validation dataset, the F1 score for algorithm A was 0.59 vs. 0.84 for algorithm B, resulting in a substantial improvement in the training, and ability to-detect the TSD defect.

Table 1. Values comparing the DL-NDT defect density to an etch defect count density for the Fig. 1. whole wafer. TD and BPD values are from algorithm A, whereas the TSD's are shown with results updated with a new detection algorithm B. ** indicates no change in algorithm.

Defect Type	NDT A [cm^{-2}]	NDT B [cm^{-2}]	Etch [cm^{-2}]
TD	1616	**	1183
BPD	580	**	345
TSD	653	370	304

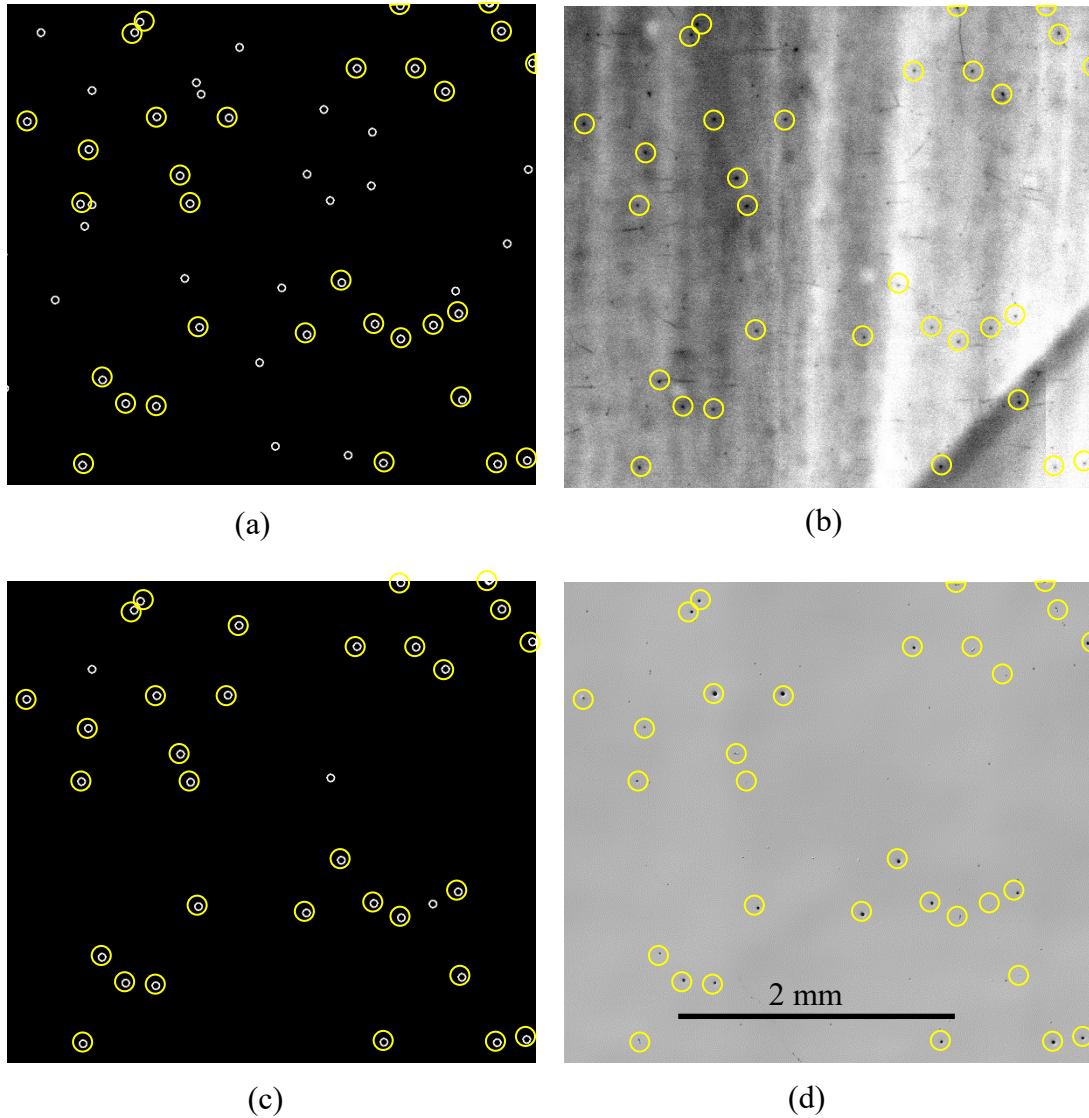


Fig. 2. Portion of full wafer DL-NDT detection comparing (a) original algorithm A with the (b) PL image signal used to detect TSD feature, (c) improved algorithm B, and (d) TSD etch features. The images are spatially aligned and taken from the C-face. The larger circles correspond to the TSD features in the etch image.

As an example of algorithm improvement, Fig. 2 illustrates the association of the 2 algorithms A and B and the respective PL image signal that corresponds to C-face KOH/ NaOH TSD etch features. Fig. 2(a) represents the initial algorithm A feature detection from the associated PL signal, Fig. 2(b). The circled features in the PL signal corresponds to circled features in the etch image of Fig. 2(d). Compared with the improved algorithm B, Fig. 2(c), of the 30 etch features, the original algorithm A counted 51 defects, whereas algorithm B counted 33, a significant improvement over algorithm A. The overcounted features correspond to threading edge defects, not screw features with C-face etch

behavior. The power of DL-NDT is the ability to use contextual information to determine which dot is a TSD (with respect to C-face etch) and which is not. This example compares qualitatively with the values of Table 1.

Fig. 3 shows a 1x1 mm defect density map of (a) NDT (algorithm B) TSD density versus the (b) etched wafer, a typical alternative depiction to the defect position map of Fig. 1. The two maps have the same scale and are nearly indistinguishable with defect position distribution, indicating the fidelity of the DL-NDT technique.

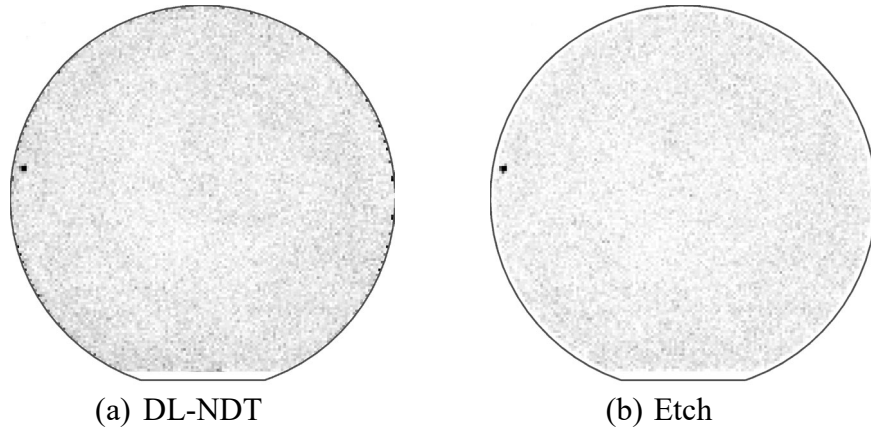


Fig. 3. Representative 1x1 mm bin 150 mm substrate defect density map indicating the correlation between the DL-NDT TSD distribution compared to etch TSD distribution across an example substrate: (a) DL-NDT PL imaging on the C-face, (b) Etch on C-face, scale is same in both images.

For multiple wafers, defect densities obtained from DL-NDT are shown in Fig. 4 versus the corresponding etch values, with orthogonal regression line statistics given in Table 2. In comparing the NDT density with corresponding etch density for TD and BPD's (Fig. 4 (a), (b)) the slope of the line and correlation coefficient is close to 1, indicating the DL-NDT algorithm infers the defect structure from the PL signal as labeled and trained with etch features. For the TSD correlation in Fig. 4(c), the slope of the line is closer to 2, with more variation in etch vs. DL-NDT as seen from the correlation coefficient (and the plot points). This is consistent with overcounting from the DL-NDT algorithm as discussed previously, as well as the error associated with etch features. The overcount may also be due to PL signal from threading mixed a+c Burgers vector dislocations (TMD), which were not differentiated with the etch labeling. The orthogonal regression fit of the data suggests both the precision in defining the feature used to label defect, and the uniqueness of the signal feature being generalized by the DL-NDT algorithm: a more well defined or unique feature is easier to accurately label. Some difficulties in the etch features can be size of pit, auto-labeling, cleanliness of wafer (before and after etch) and initial wafer conditions prior to etch. The algorithm can be improved further by reducing the inconsistencies in the etch and labeling, and recursively re-training with updated image datasets. The intent of this work is to produce a reliable and consistent method for defect counting so that etching is not needed for high volume wafer defect characterization; however, validation by etch continues to be the best way to ensure a high-fidelity DL-NDT algorithm.

Table 2. Fit information for the images in Fig. 4.

Defect Type	Slope	Intercept	Correlation
TD	1.07	-136	0.93
BPD	1.04	-28.2	0.95
TSD	1.70	76.0	0.68

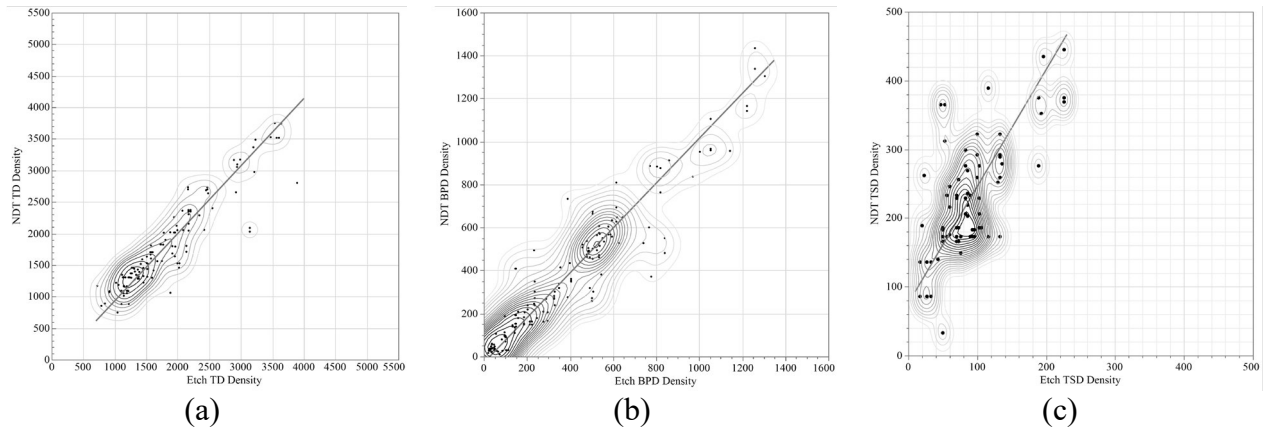


Fig. 4. These graphs represent the capability of DL-NDT dislocation density determination versus validated densities from conventional etching for (a) Total threading dislocations (TD), (b) Basal plane dislocation (BPD) and (c) Threading screw dislocations (TSD). The contours represent the data population density showing most of the data along the line fit.

The DL-NDT method algorithm is extensible to other diameter substrates as well. For a 200 mm 4H-SiC substrate, Fig. 5 shows the distribution of device dependent BPD and TSD detected features with densities of 552 and 277 cm^{-2} respectively. The ability to quickly characterize the defects in SiC allows for faster improvement in the crystal growth processes, while still being able to use the substrate to determine performance.

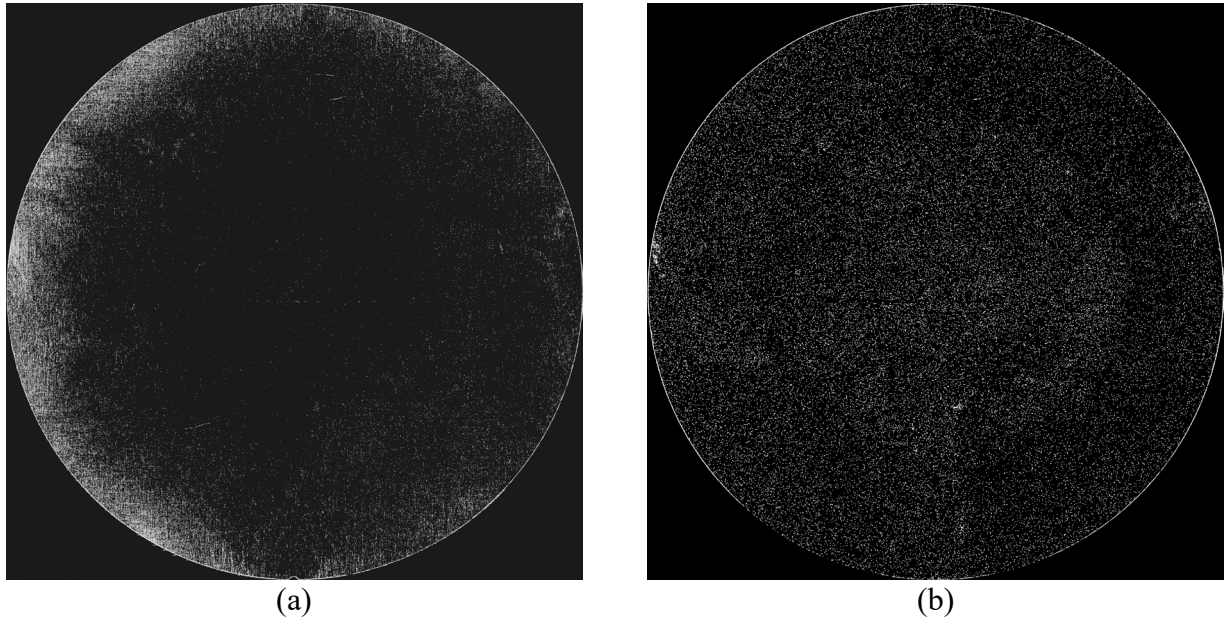


Fig. 5. a) DL-NDT BPD density of 552 cm^{-2} ; b) DL-NDT TSD density of 277 cm^{-2} .

Summary/ Conclusion

A whole wafer method for high volume non-destructive characterizing of dislocations has been demonstrated for 150 mm and 200 mm 4H-SiC wafers. Deep learning coupled with non-destructive high volume fast optical microscopy methods is shown to correlate industry accepted chemistry and physics based etch and diffraction techniques for defect characterization. The computational power for algorithm inference to detect features in the acquired PL optical confocal microscopy images as well as storage of the images and data adds to the complexity of systematic methods necessary to make this a viable process. Etch validation of the results ensures the technique is dependable, with the ultimate goal of reducing etch to minimal levels for algorithm maintenance. The ability to use this

non-destructive technique will result in better correlation with device behaviors, and fast feedback to crystal growth processes to improve substrate wafers.

References

- [1] Wolfspeed, Inc., “Lucid Motors Integrates Wolfspeed’s Silicon Carbide Semiconductors into the Award-Winning Lucid Air”, 5 April 2022, <https://investor.wolfspeed.com/>
- [2] X.R. Huang et. al., Appl. Phys. Lett. 91, 231903 (2007).
- [3] R. T. Leonard et. al, Materials Science Forum, Vol. 1004, pp. 321-327 (2020).
- [4] P. Isola et. al., arXiv:1611.07004v3 [cs.CV] 26 Nov 2018.
- [5] J.J. Sumakeris et. al., Materials Science Forum, Vol. 858, pp. 393-396 (2016).
- [6] Guo, J., Yang, Y., Wu, F. et al., J. Electron. Mater. 45, 2045–2050 (2016).
- [7] R.T. Leonard et. al., Materials Science Forum, vol. 600–603, pp. 7–10 (2008).
- [8] Nancy Chinchor. 1992. MUC-4 evaluation metrics. In Proceedings of the 4th conference on Message understanding (MUC4 '92). Association for Computational Linguistics, USA, 22–29. <https://doi.org/10.3115/1072064.1072067>.