

# Few-shot AI Segmentation of Semiconductor Device FIB-SEM Tomography Data

Heiko Stegmann  
Carl Zeiss Microscopy GmbH

Flavio Cognigni  
Carl Zeiss SpA

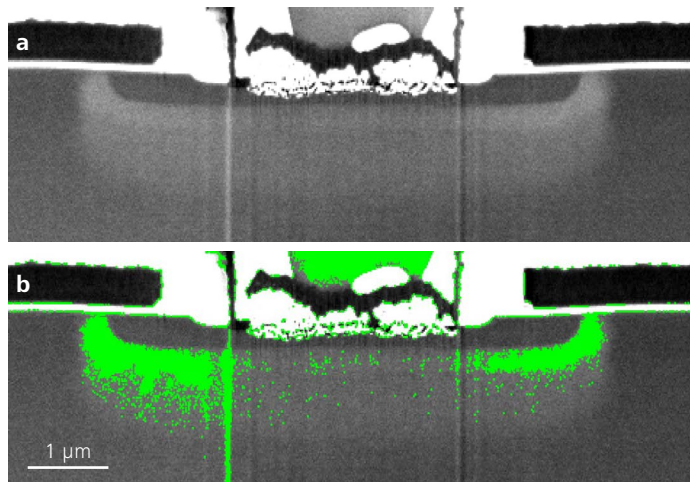
Seeing beyond

## Abstract

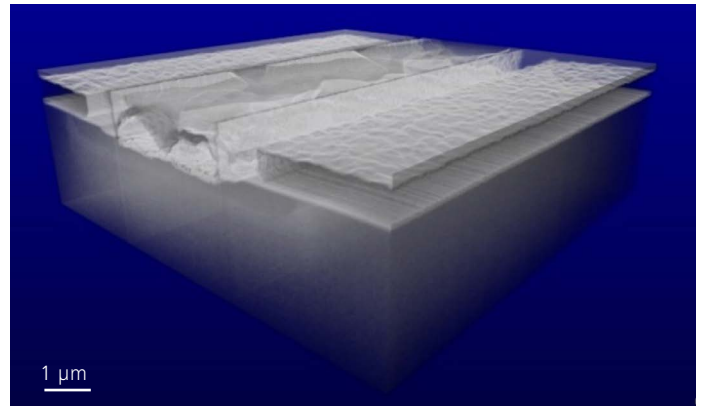
Image segmentation is a valuable tool for visual image data inspection of semiconductor device structures. For the large amounts of data provided by recent advancements in automated scanning electron microscope (SEM) and focused ion beam-scanning electron microscope (FIB-SEM) data acquisition, automatic segmentation becomes indispensable to fully exploit the information contained in the data in automated characterization workflows. Using two exemplary FIB-SEM tomography datasets, we explored artificial intelligence based image segmentation using only a minimum amount of training images annotated by a human user.

## Introduction

SEM images of semiconductor device structures are usually analyzed manually by experts using their specialist knowledge for identification and correct measurement of certain reference structures, or for detection and interpretation of defects. This approach can hardly be scaled to analyze a larger number of images and features than a human user can process. This is especially true for FIB-SEM tomography image series.



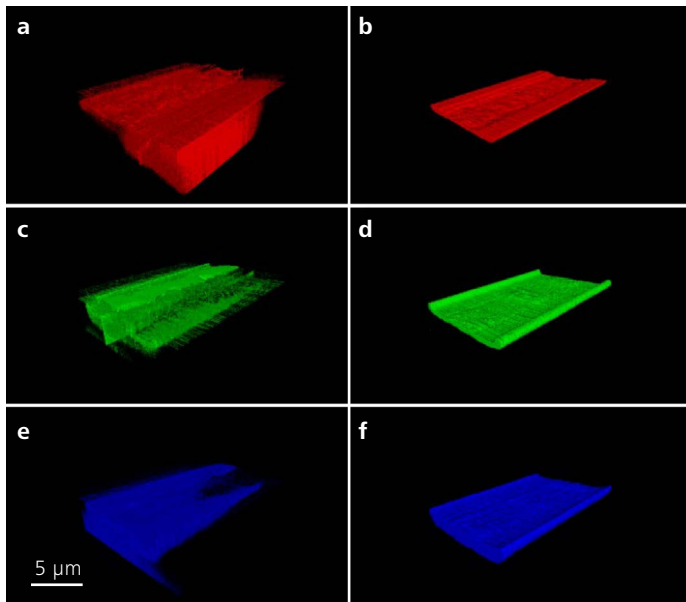
**Figure 1** (a) Single slice image from a FIB-SEM tomography dataset of a SiC MOSFET. (b) Same image showing an attempt of conventional grey value threshold segmentation of doping area P1.



**Figure 2** Projection image calculated from the SiC MOSFET dataset.

High spatial resolution 3D imaging of the interior of semiconductor devices by FIB-SEM tomography can be helpful for failure analysis, as it often provides better insight into a failure mechanism than cross-sectional 2D imaging alone. It also provides process engineers with clear visual feedback.<sup>[1,2]</sup> Defect signatures such as material patches remaining from previous process steps, misalignments, or cracks, and how they relate to their surroundings can be inspected in FIB-SEM tomography datasets by scrolling through the stack of 2D images, or by extracting virtual slices from it at a desired position and orientation to inspect structural details. However, this approach defeats the major advantage of 3D imaging, the ability to freely inspect the different device structures from any viewing angle in a 3D visualization to assess their spatial relationships.

A simple such 3D visualization can be generated by dynamically calculating projection images from the whole image stack in the required orientations and selecting some intensity-to-opacity relation. However, assessing all detail present in the dataset from projections is very difficult. Often, components that are spatially separated overlap with others in the projection image, making it difficult to distinguish them and to accurately determine their shape and position.



**Figure 3** Left column: Result of a conventional segmentation attempt to separate three different dopant regions. Right column: Result of the few-shot AI segmentation. (a and b) N region, (c and d) P1 region, (e and f) P2 region.

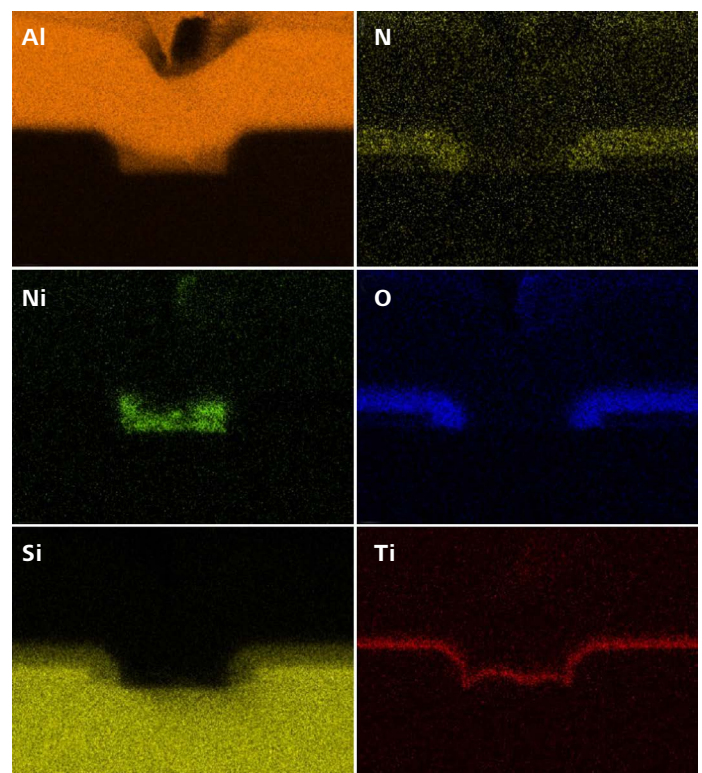
To display individual device components separately and without being overlaid by other components in a 3D FIB-SEM tomography data set, they must be clearly separated from each other. Furthermore, unique separation of different components in the 3D dataset is prerequisite to calculating their volume fractions or other quantitative geometrical structure properties. In the 2D case, the same applies to calculating area fractions, for example to calculate grain size distributions or to automatically measure layer thicknesses.

The process of assigning each pixel or voxel to a unique class, each class corresponding to one or a set of individual components in a 2D or 3D image dataset is called image segmentation. Segmentation by a human operator manually marking the structures of interest in datasets of hundreds or thousands of images is prohibitively expensive. Thus, automatic segmentation is required for productive use of the method. Conventional automatic segmentation assigns image pixels to unique classes according to their grey values by simple thresholding with optional filtering in pre- or post-processing, or by more advanced clustering algorithms that consider statistical properties of the pixel neighborhood.<sup>[3]</sup>

When manually segmenting features in SEM images, human operators do not only use grey values to decide which structural feature a pixel belongs to, but also their expert ability to identify what the feature looks like under the imaging conditions used.

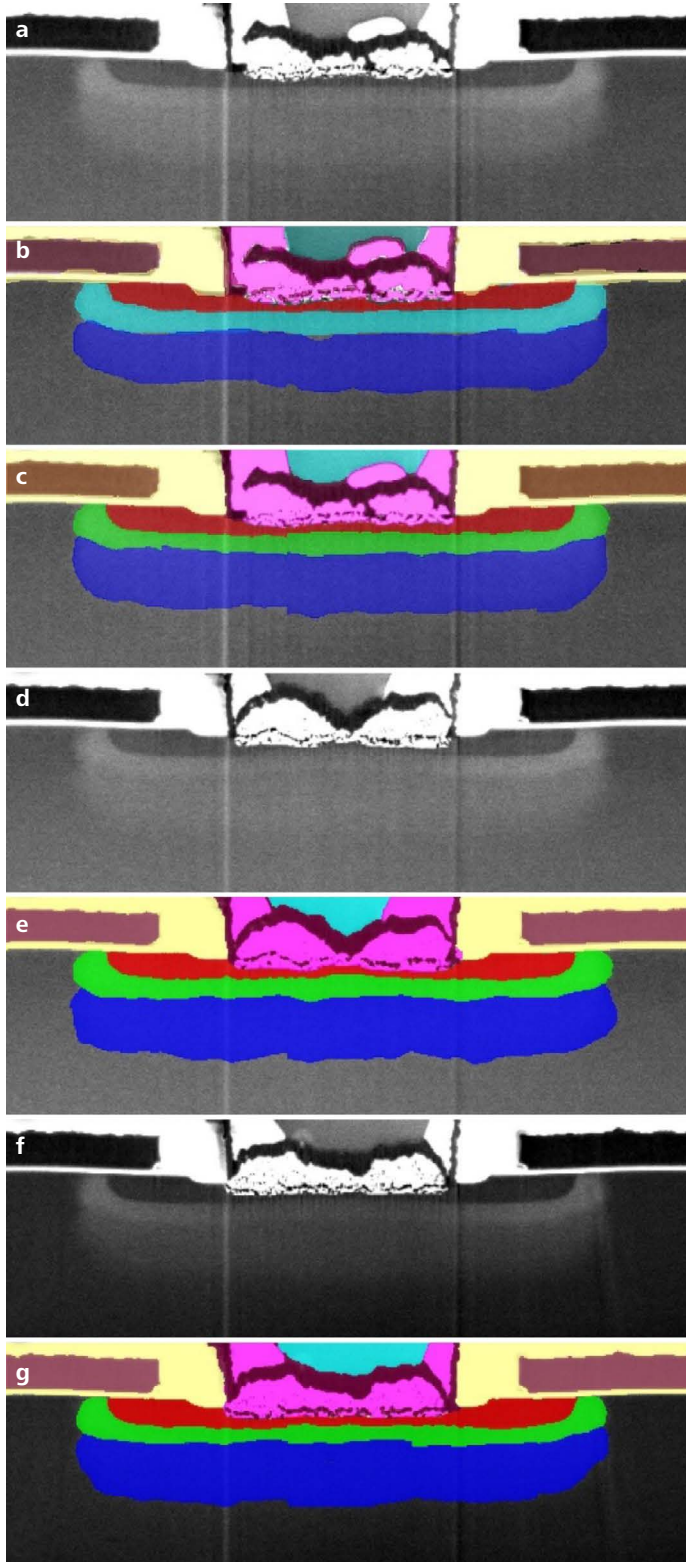
A corresponding approach in automatic image segmentation is the use of trainable segmentation. It has been used for defect segmentation<sup>[4]</sup> and reverse engineering.<sup>[5]</sup> These artificial intelligence (AI) methods learn domain knowledge by training a segmentation model from pixels labeled by an expert in a training dataset, then applying the model to segment other datasets acquired under the same conditions. In practice, labeling pixels consists of marking the target structures in the image using a suitable input device such as mouse, touchscreen, or graphics tablet. We refer to this process as annotating images. The model may use an artificial neural network or other machine learning algorithms to decide which pixels belong to which features.

AI is a term that refers to a wide range of *intelligent behaviors* exhibited by computers. Within the field of AI, there is a subset called machine learning (ML), which refers to methods where computers can learn and improve without explicit programming for each task. Deep learning (DL) is a further subset of ML that utilizes neural networks. These networks consist of interconnected nodes and layers that mimic the information encoding and decoding processes of the human brain. Each node, or neuron, focuses on a specific aspect of the data being processed.

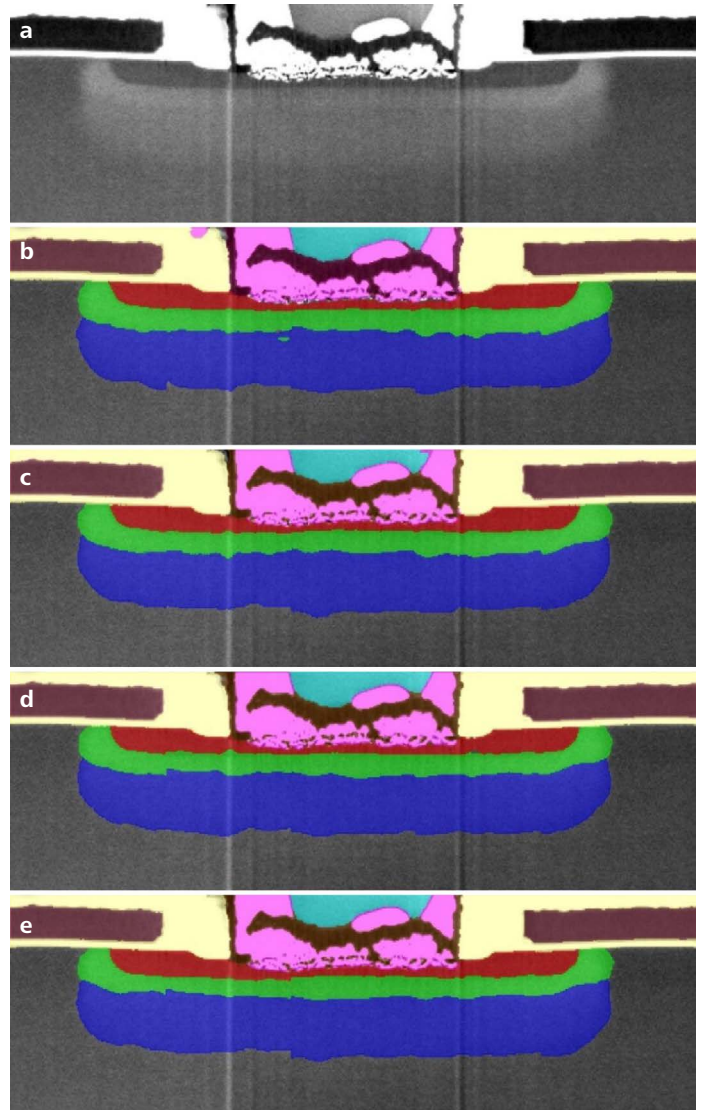


**Figure 4** EDX elemental maps obtained from the SiC MOSFET cross-section after FIB-SEM tomography acquisition.





**Figure 5** (a) One slice from the SiC MOSFET dataset. (b) Annotations manually drawn in the image to define the different components to be segmented. (c) Segmentation result on the same slice. (d and e) First slice of the dataset and its segmentation, (f and g) Last slice and its segmentation.



**Figure 6** (a) One slice from the SiC MOSFET dataset. (b-e) Same slice with segmentation result obtained from using 8, 12, 16 and 20 annotated images for model training.

In DL, chains of multiple neurons work together to learn and solve complex problems. While individual neurons cannot provide solutions on their own, their collective efforts contribute to accurate predictions in various domains. DL has demonstrated success in tasks such as biomedical image segmentation, image classification, and shape recognition for computer vision applications.

Convolutional neural networks (CNNs) are a type of DL neural network commonly used in image processing tasks. They are particularly effective in computer vision problems where there is a correlation among adjacent pixels in an image. Unlike considering the entire image at once, CNNs utilize small regions or patches that are moved over the image. This approach allows for localized analysis, leading to improved accuracy and efficiency.

One specific CNN architecture that has gained popularity in image segmentation problems is the U-net. Its success can be attributed to its ability to achieve remarkable results even with limited annotated inputs. The U-net architecture has been widely adopted due to its effectiveness in segmenting images and its potential for various applications.<sup>[6]</sup> Recent progress in this field made it possible to significantly reduce the amount of training images that need to be annotated by the user. Often it is sufficient to use only several individual images for training (“few-shot” learning).

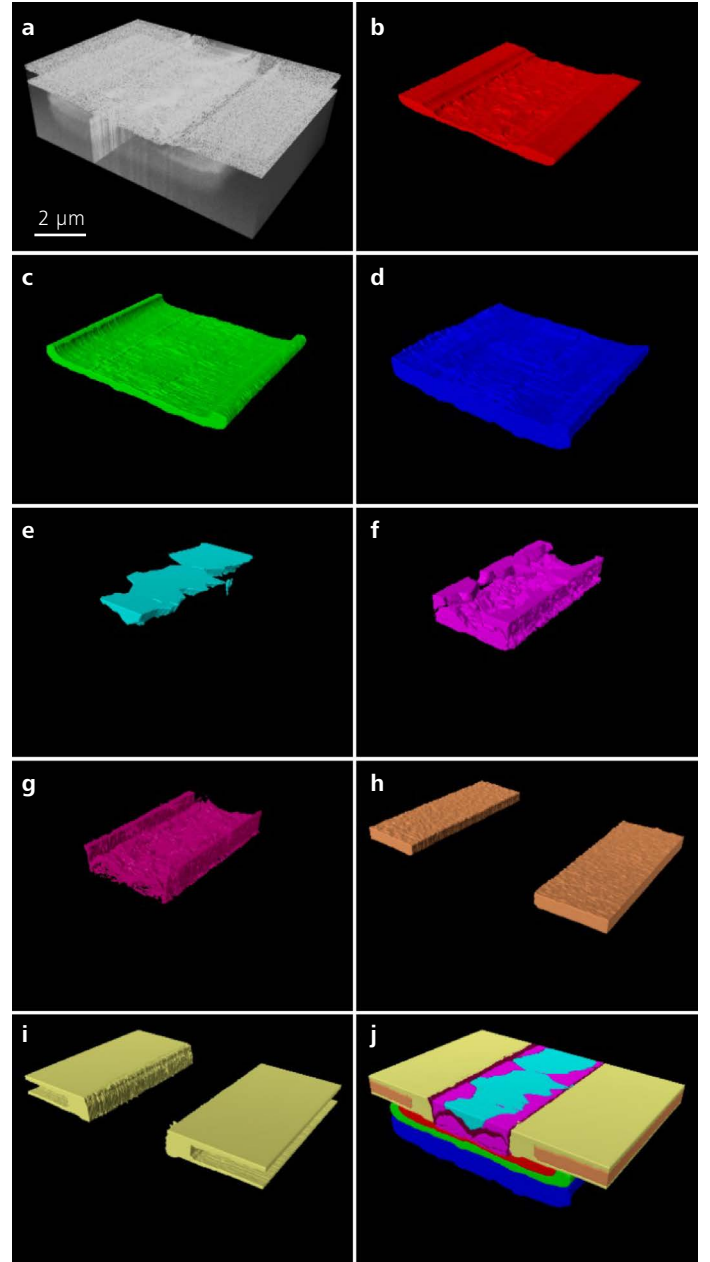
In this paper, we present few-shot AI semantic segmentation applied to FIB-SEM tomography datasets acquired from two different samples. We focused on investigating what the minimum number of manually segmented images would be to obtain a good segmentation, as well as how well such a segmentation model will generalize to segment FIB-SEM tomography images recorded from the same sample but using secondary electron (SE) versus backscatter electron (BSE) detection, different signal-to-noise ratio (SNR), and different FIB-SEM instrument configurations.

## Materials and Methods

The FIB-SEM tomography datasets used for segmentation were acquired from a SiC MOSFET power device (320 images, 20 nm pixel size, 30 nm slice thickness, volume size  $9.5 \times 3.1 \times 9.6 \mu\text{m}^3$ , subset of a dataset from <sup>[7]</sup>), and from a 5 nm technology node SRAM sample (6004 images, 1 nm pixel size, 0.5 nm slice thickness, volume size  $6.90 \times 0.40 \times 3.04 \mu\text{m}^3$ ).<sup>[8]</sup>

For the data acquisition, ZEISS Crossbeam FIB-SEM instruments with Atlas 3D software were used. See the corresponding references for further detail on the data acquisition. After tomography data acquisition, energy dispersive X-ray spectroscopy (EDX) elemental maps were recorded at 10 kV from the final cross-section using an Oxford Instruments Ultim Max 170 detector. From the SRAM sample, another dataset (1168 images, 1 nm pixel size, 2.5 nm slice thickness, volume size  $3.21 \times 0.29 \times 2.92 \mu\text{m}^3$ ) was acquired on a ZEISS GeminiSEM 560 with Ion-sculptor FIB column (“L-shape Crossbeam”). Unlike the conventional  $54^\circ$  angle between FIB and SEM columns, this instrument is configured with a  $90^\circ$  angle that allows to use smaller working distances. It is also equipped with a SEM objective lens optimized for low voltage operation. Both features result in higher imaging resolution than with a standard configuration. With all datasets, backscatter electron (BSE) and secondary electron (SE) images were recorded simultaneously using the respective in-column detectors.

The computation intensive AI segmentation model training was done with the ZEISS arivis Cloud service that uses a U-net with EfficientNet<sup>[9]</sup> as encoder and Pixelshuffle<sup>[10]</sup> as decoder.



**Figure 7** (a) Projection image of the SiC MOSFET dataset. (b) N, (c) P1, (d) P2, (e) Al, (f) Ni, (g) TiN, (h) Poly-Si, (i) SiON segmentations, and (j) all segments together.

Segmentation of the datasets using the trained models, as well as conventional segmentation and 3D visualization were done locally using the ZEISS arivis Pro software.

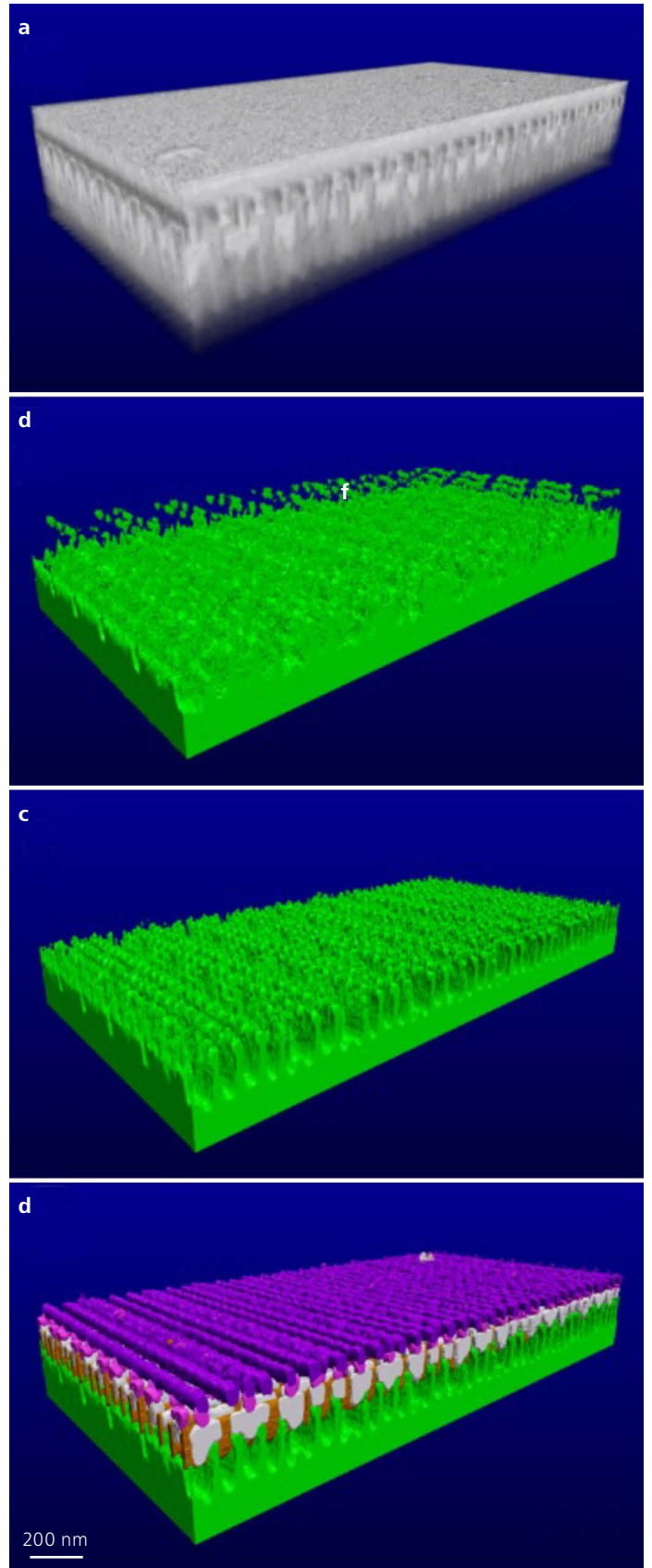
## Results and discussion

Because of the strong passive voltage contrast (PVC) achieved at the imaging conditions used,<sup>[7]</sup> the SiC MOSFET dataset exhibits three different implant zones: A dark  $N^+$  region below either end of the gates, and two different levels of bright P-well regions with a banana-shaped cross-section (hereinafter referred to as N, P1 and P2 regions). Figure 1a presents one of these images.

The dopant regions exhibit a relatively weak, noisy signal, and there is no clear separation between different groups of pixels belonging to different doping regions. Grey values of the same range as in a dopant zone can also be found in other parts of the dataset that do not represent doping zones. Beyond that, the dataset contains image artifacts such as grey value fluctuations originating from local sample charging, shadowing effects due to trench walls adjacent to the imaged cross-section, and residual vertical FIB milling artifacts (“curtaining”). Figure 2 shows a projection image of the dataset. It does not allow to assess shape and extent of the doping regions and their relationship to the overlying structures.

Achieving unambiguous segmentation of the three different doping levels using pixel grey value based segmentation was not possible. This is illustrated in Figure 1b, in which an attempt was made to segment the P1 dopant area by grey value thresholding. It is not possible to select a grey value range to separate P1, as other parts of the image contain pixels that fall into that range as well. Then, a more sophisticated conventional segmentation attempt was made with the complete dataset. The images were pre-processed by removal of inhomogeneous image background with a 250 pixel rolling ball filter, followed by 2 pixel median filter to decrease noise. Three different pairs of upper and lower gray value thresholds were chosen to select only the pixels belonging to either N, P1 or P2. The result is that pixels far away from the actual doping zones end up being included in the segmentation, obscuring the actual zones of interest, and rendering 3D visualization useless (Figure 3a, c, and e).

For AI segmentation of the SiC MOSFET dataset, eight classes were annotated: The three different doping zones N, P1 and P2 and five distinct components identified from the EDX maps shown in Figure 4 (Al, Ni, TiN, Poly-Si and SiON) were manually annotated by an expert user in the SE images. The BSE images were not considered here, as they do not show PVC. Figure 5b shows one slice with annotations to illustrate this. Areas not belonging to either of these were annotated as background. To assess how many images need to be annotated to achieve a good segmentation, first 8 images, then 12, 16, and finally 20 of the 320 images were annotated before training the model on the annotations. The default training time suggested by the software was used, we did not investigate the effect of increasing the training time. The segmentation obtained after training with eight annotated images showed only minor misclassifications (Figure 6b). These were absent when training with 12 annotated images (Figure 6c). Training with 16 or 20 annotated images showed no further improvement, apart from minor variations in the resulting segment shapes (Figure 6d, e). This indicates that indeed model training on only a few annotated images is sufficient for artifact-free segmentation of this dataset.

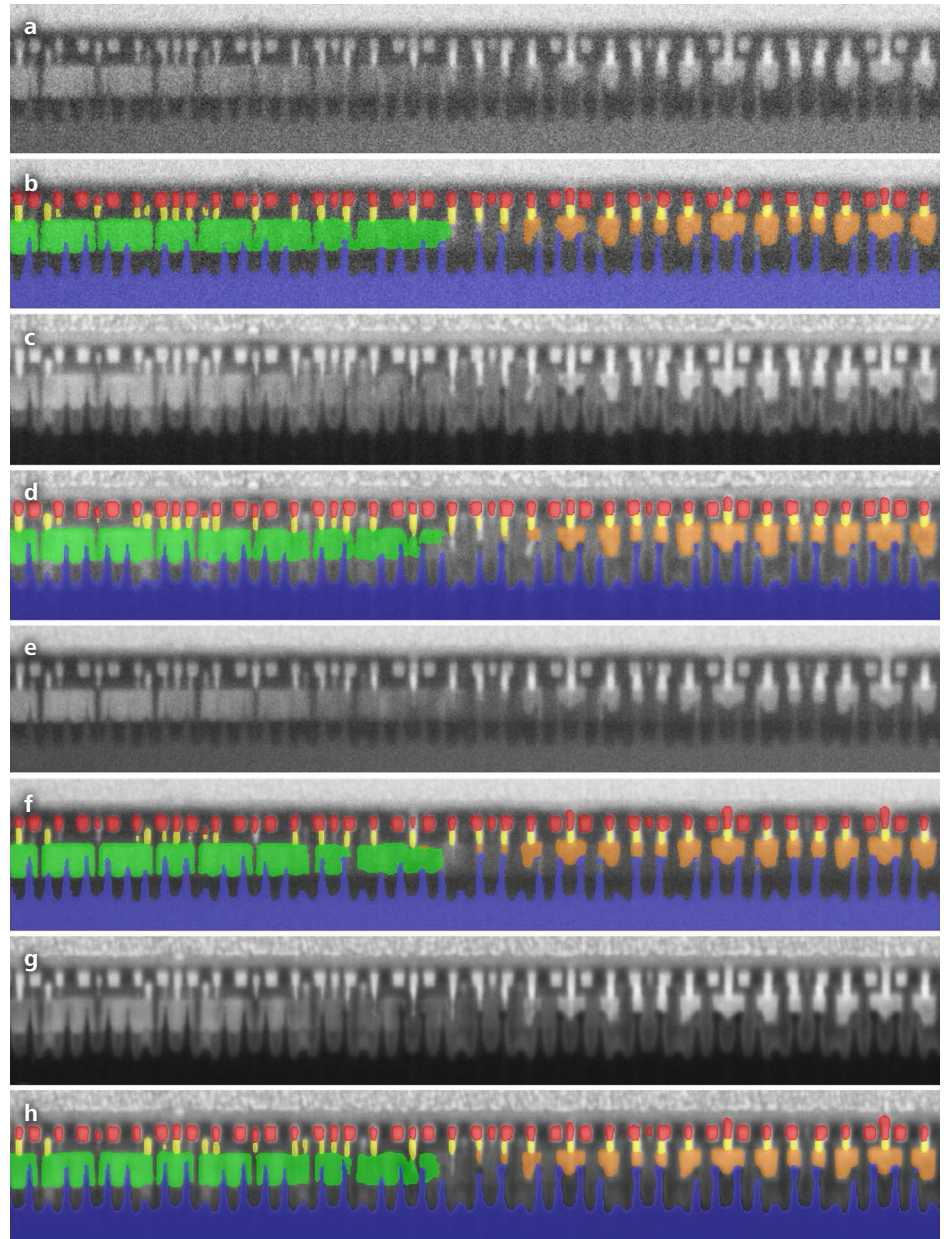


**Figure 8** Surface rendering of a FIB-SEM tomography dataset of a 5 nm technology node SRAM. (a) Projection. (b) Conventional segmentation of the transistor fins. (c) Few-shot AI segmentation of the fins. (d) Few-shot segmentation of fins (green), contacts (pink), gates (brown), M1 (purple), SD (white).



The model trained on 20 annotated images was finally used to segment all 320 images. The result is a clean segmentation of the three dopant zones and of the five components. Figure 5c-g shows the segmentation results for three different slices, representative for the whole dataset. Figure 3 presents a side-by-side comparison of 3D surface renderings of the N, P1 and P2 doping zones AI segmentation (b, d, f) to the conventional segmentation result (a, c, e), while Figure 7 shows an overview of all eight segmented classes. In contrast to the conventional segmentation, the AI segmentation approach provides a clear separation of the different device components.

For the 5 nm SRAM FIB-SEM tomography datasets, six classes were annotated: Fins, gates, epi-source/drain (SD), contacts, metal 1 (M1), and background. Starting with the standard 54° FIB-SEM dataset, we used the same approach of increasing numbers of annotated images to assess how many of them are required for good segmentation. Although the images were quite noisy due to the relatively short pixel dwell time used for acquisition, model training on only two annotated images was already sufficient for a segmentation of the SE data with only minor misclassifications, and on six annotated images for the BSE part. Figure 8 shows 3D renderings of a  $1.6 \times 0.4 \times 0.6 \mu\text{m}^3$  subset of the SE part of the data, a segmentation of the transistor fins by grey value thresholding and the AI segmentation of the fins based on two annotated images, and all six AI segmented classes together. Comparing the conventional and AI fin segmentation, the superiority of the AI method becomes obvious. More annotated images were then added to the training, up to 11 each for the SE and BSE part, but no further improvement was observed with more than eight annotated images.

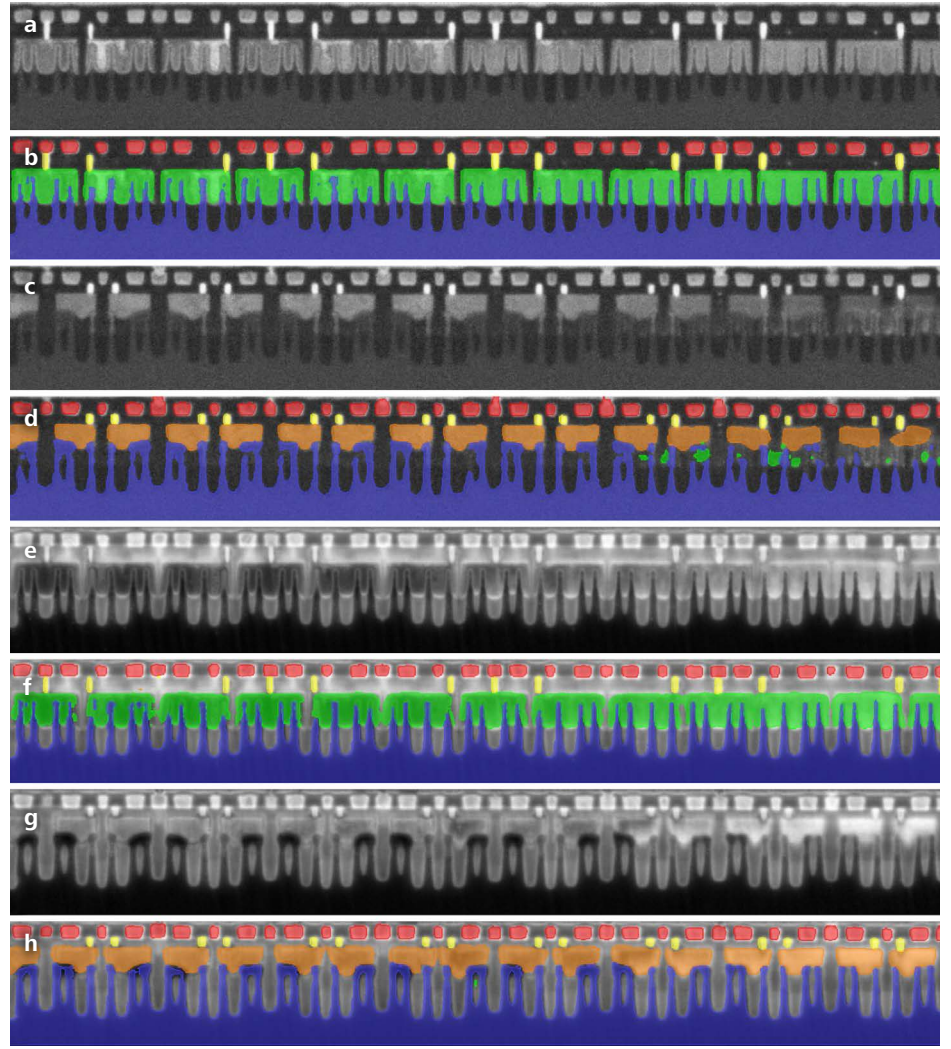


**Figure 9**  $2.45 \times 0.40 \mu\text{m}^2$  sections of a BSE (a) and an SE slice (c) from the 5 nm SRAM dataset and their corresponding AI segmentations (b, d). Slices obtained by averaging five neighboring slices of the original BSE (e) and SE (g) data, and their corresponding AI segmentations (f and h). Blue: Fins, Green: Gates, Orange: SD, Yellow: Contacts, Red: M1.

SE and BSE images render different structural components of the sample in a different way, which leads to slightly different manual annotation results. We reckoned that combining annotations obtained from SE and BSE images to train both dataset parts is advantageous, as it would reduce training bias originating from annotating only one kind of images.

Therefore, and to assess if it is even possible to use a single model for simultaneous segmentation of both the SE and the BSE data, the two image stacks were merged into one, and the annotations from the 11 annotated SE and 11 (different) annotated BSE slices were used to train a new model on the merged data. The segmentations thus obtained showed no obvious difference to the separate SE and BSE segmentations.

The original purpose of recording the SRAM data with short pixel dwell time and very small slice thickness was to minimize rift related image distortions at acquisition time, and to then calculate a moving average trough the image series to get a reduced, distortion-free dataset with good signal-to-noise ratio (SNR).<sup>[8]</sup> To assess the segmentation performance of the model trained so far on such a slice averaged dataset, moving five-slice averages were calculated both on the SE and BSE parts, resulting in a reduced dataset with 2.5 nm slice thickness and increased SNR. Segmentation of the averaged images with the previous model was not good enough but could be improved by adding only two annotated images each from the averaged SE and BSE images to the training. Figure 9 shows the segmentation results on representative individual BSE and SE slices. Segmentation accuracy of fins and gates improved with the better SNR of the averaged images, while the result for other components was the same.



**Figure 10**  $2.45 \times 0.29 \mu\text{m}^2$  sections of two BSE (a, c) and two SE slices (e, g) from the 5 nm SRAM L-shape dataset, and their corresponding AI segmentations (b, d and f, h). Blue: Fins, Green: Gates, Orange: SD, Yellow: Contacts, Red: M1.

Finally, we were interested in finding out whether the same model trained on the data acquired with the standard FIB-SEM configuration generalizes well enough to be suitable for segmenting the L-shape dataset. Due to higher resolution and different acceptance properties of its electron detection system, both SE and BSE images acquired with this instrument look different than those recorded with the standard FIB-SEM. They contain more detail and render the different device components in a slightly different way. As expected, with the existing model the result was an incomplete and erroneous

segmentation. However, adding only eight annotated SE and eight BSE images was enough to improve the segmentation of the L-shape dataset to a usable level. Figure 10 shows the results on four representative slice images. Corresponding 3D renderings are given in Figure 11. The gates are better segmented in the BSE images, while the fins are better segmented in the SE images. The other components are equally well segmented in both parts. We interpret this to mean that still more annotations and, possibly, more training time is needed to further improve the segmentation of this dataset.

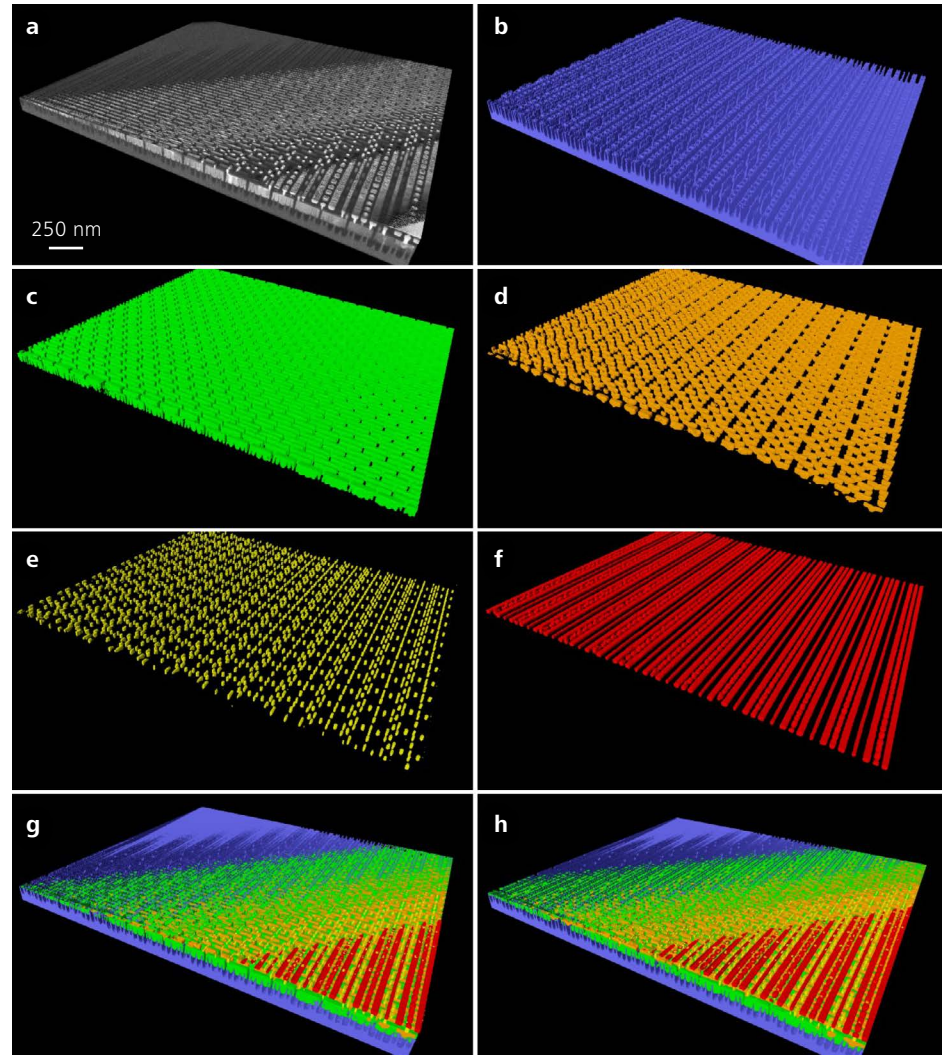


## Conclusions

Recent years have witnessed significant advancements in automated SEM and FIB-SEM image acquisition protocols. Notably, software solutions have enabled the acquisition of large multi-detector FIB-SEM tomography datasets along with 3D EDX maps. This progress is made possible by the automatic adjustment of parameters when transitioning from imaging to elemental analysis. Moreover, this approach includes the capability to automatically capture high resolution mosaic images of large specimen areas using customized recipes. Consequently, larger areas can be investigated with higher resolution.

Segmentation is often the first step in drawing manufacturing process related conclusions from such data. Segmented images allow to quantify geometrical properties of the different device structures, such as their volume fractions, deviations of their shapes from the target, or layer thicknesses. They provide a way of automatically locating candidates for defects by detecting outliers within a segmentation class. In contrast to the conventional segmentation approach, the AI based results presented in this paper were largely free from incorrect segmentation due to unclear pixel assignment and image artifacts. Therefore, more statistically relevant data become available in a shorter time by incorporating AI based image segmentation models in the data processing workflow, and it can potentially be fully automated.

Exploiting cloud-based services for training a DL segmentation model, while calculating the actual data segmentation with the trained model on a local workstation, as it was carried out here, has several advantages. First, modern cloud-based solutions provide the possibility to run batch processes in parallel,



**Figure 11** Projection rendering of the L-shape dataset (BSE part, a), and 3D renderings of fins (b), gates (c), SD (d), contacts (e), M1 (f), all together (g, h). (c, d, e, f, and g) From the BSE segmentation; (b and h) From the SE segmentation. a, g, h cut at an angle for better visibility of the different layers.

distributed computing resources making this a scalable approach that can minimize computation times and cost per dataset. Second, this approach allows physical analysis engineers to not rely solely on their local workstation resources that limit the overall throughput. When performed in the cloud, the segmentation process can be run in the background, so that local workstations can be used for other purposes. Third, data confidentiality concerns can potentially be dispelled by performing the predominant part of the model training in the cloud on data obtained from non-confidential devices

that are built similarly to the confidential ones, followed by comparatively short local re-training of the model with additional annotations made on the confidential data. We have shown a similar approach by training a model on data acquired on a particular type of FIB-SEM, then re-training the same model on data obtained from a different type. If this approach can be generalized to datasets from similar, but not identical devices remains to be explored.



## References

- [1] D. Mello, et al., "New Approach in Physical Failure Analysis Based on 3D Reconstruction." Proceedings of the ISTFA 2022, 201-205 (2022). doi.org/10.31399/asm.cp.istfa2022p0201.
- [2] F. Cognigni, et al., "A Multiscale and Multimodal Correlative Microscopy Workflow to Characterize Copper Segregations Identified in Epitaxial Layer of Power MOSFETs." Proceedings of the ISTFA 2023, 92-100 (2023). doi.org/10.31399/asm.cp.istfa2023p0092
- [3] Y. Yu, et.al., "Techniques and Challenges of Image Segmentation: A Review." Electronics 12, 1199 (2023). doi.org/10.3390/electronics12051199.
- [4] N. Ofir, et al., "Automatic defect segmentation by unsupervised anomaly learning." 2022 IEEE International Conference on Image Processing ICIP (2022). doi 10.1109/ICIP46576.2022.9898035
- [5] N. Rothaug, et al., "Towards Unsupervised SEM Image Segmentation for IC Layout Extraction." Proceedings of the 2023 Workshop on Attacks and Solutions in Hardware Security (ASHES '23) (2023). doi.org/10.1145/3605769.3624000.
- [6] O. Ronneberger, et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science Vol. 9351. Springer, (2015). Doi.org/10.1007/978-3-319-24574-4\_28
- [7] H. Stegmann, et al., "Combining Three-Dimensional FIB-SEM Imaging and EBIC to Characterize Power Semiconductor Junctions." Proceedings of the ISTFA 2023, 478-482 (2023). doi.org/10.31399/asm.cp.istfa2023p0478.
- [8] H. Stegmann, A. Laquerre, "FIB-SEM Tomography Acquisition and Data Processing Optimization for Logic and Memory Structures." Proceedings of the ISTFA 2023, 387-392 (2023). doi.org/10.31399/asm.cp.istfa2023p0387
- [9] M. Tan and Q.V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." Proceedings of the 36th International Conference on Machine Learning, PMLR 97:6105-6114 (2019). proceedings.mlr.press/v97/tan19a.html.
- [10] A.P. Aitken et al., "Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize." ArXiv abs/1707.02937 (2017). api.semanticscholar.org/CorpusID:21850448.



**Carl Zeiss Microscopy GmbH**

07745 Jena, Germany

microscopy@zeiss.com

[www.zeiss.com/semiconductor-microscopy](http://www.zeiss.com/semiconductor-microscopy)

**Follow us on social media:**

