

Training procedure for scanning electron microscope 3D surface reconstruction using unsupervised domain adaptation with simulated data

Tim Houben^{a,b,*}, Thomas Huisman,^a Maxim Pisarenco^a,
Fons van der Sommen^b, and Peter de With^b

^aEindhoven University of Technology, Electrical Engineering,
Video Coding Architectures Group, Eindhoven, the Netherlands

^bASML Netherlands B.V., Veldhoven, the Netherlands

Abstract. Accurate metrology techniques for semiconductor devices are indispensable for controlling the manufacturing process. For instance, the dimensions of a transistor's current channel (fin) are an important indicator of the device's performance regarding switching voltages and parasitic capacities. We expand upon traditional 2D analysis by utilizing computer vision techniques for full-surface reconstruction. We propose a data-driven approach that predicts the dimensions, height and width (CD) values, of fin-like structures. During operation, the method solely requires experimental images from a scanning electron microscope of the patterns concerned. We introduce an unsupervised domain adaptation step to overcome the domain gap between experimental and simulated data. Our model is further fine-tuned with a height measurement from a second scatterometry sensor and optimized through a tailored training scheme for optimal performance. The proposed method results in accurate depth predictions, namely 100% accurate interwafer classification with a root-mean-squared error of 0.67 nm. The R^2 of the intrawafer performance on height is between 0.59 and 0.70. Qualitative results also indicate that detailed surface features, such as corners, are accurately predicted. Our study shows that accurate z -metrology techniques can be viable for high-volume manufacturing. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMM.22.3.031208](https://doi.org/10.1117/1.JMM.22.3.031208)]

Keywords: SEM; scatterometry; 3D metrology; synthetic data; domain adaptation; surface reconstruction.

Paper 23005SS received Jan. 19, 2023; revised manuscript received Apr. 6, 2023; accepted for publication Apr. 10, 2023; published online Apr. 28, 2023.

1 Introduction

Over the years, there is a continuous shrink of semiconductor devices, which increases the demand for more sophisticated metrology techniques. The measurements are used to control the production process in order to optimize production yield. As the shrinking of dimensions continues, local measurements become necessary, since stochastic behaviors of physical processes become more dominant in describing the shape of structures.¹ Moreover, the third dimension becomes increasingly important, as vertically dominated structures (e.g., FinFETs, gate-all-around, and 3D NAND) have grown into popular building blocks for semiconductor devices. In this work, we zoom in on the dimension of a fin, which is the transistor channel of a fin field-effect-transistor (FinFET), as a key description parameter for the height of the physical structures (see Fig. 1). Currently, metrology devices that are suitable for high-volume manufacturing, are optimized for 2D measurements. This work investigates if accurate reconstruction of a 3D surface of the chip is possible, by applying computer vision techniques.

Scanning electron microscopes (SEMs) are widely used in the production process to inspect defects and measure the dimensions of structures due to their ability to provide high-resolution,

*Address all correspondence to Tim Houben, t.houben@tue.nl

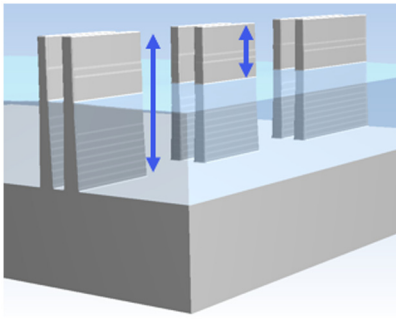


Fig. 1 Example drawing of fin structures on a wafer, indicating the main dimension of interest in this work.

localized measurements at a relatively fast rate. Other tools offer more local and precise measurements (e.g., transmission electron microscopy² and focused ion beam SEM³), but compromise in practicality, since the structures require first to be cut into slices, which is a destructive and cumbersome operation. Alternatively, scatterometry tools⁴ are nondestructive and faster than SEMs, but they are only suitable for global measurements on repetitive structures. On top of that, scatterometry suffers from long time-to-recipe and significant model complexity in thick stacks. In summary, various measurement tools are available, each with its own trade-offs in terms of accuracy and speed. In this work, we utilize a combination of these tools.

This research is focused on applying data-driven computer vision techniques for semiconductor systems. Most of the methods need a ground truth during training, which can be difficult to obtain for nanometer-sized structures due to the high cost and complexity of performing precise device measurements. This makes ground-truth information scarce, from a machine learning perspective. Instead, we need to explore alternative solutions.

One strategy to address the absence of topological ground truth for SEM images is to generate synthetic data by modeling an SEM as accurately as possible. There are various electron particle simulators⁵ available, which simulate the scattering of particles in semiconductor devices. The input required for these simulators is a 3D surface model of the structure, which can also serve as a reliable reference for training a neural network. Although these simulators have their advantages, they are not without limitations. Important components of the SEM tool, such as the electron column, are not fully modeled, leading to certain physical phenomena not being represented in the simulated SEM images, such as spherical and chromatic aberration effects, which can affect resolution. Additionally, effects such as charging,⁶ contamination, and material shrinkage, which impact the accuracy of the tool, are not always incorporated as well. Another point of consideration is that creating 3D meshes of on-chip structures requires significant effort.

Our goal in this research is to use a combination of different measurement tools to achieve optimal algorithm performance. We combine experimental and synthetic SEM images by utilizing an unsupervised domain adaptation method to bridge the gap between the two. Additionally, measurements from a scatterometry sensor are used to calibrate the height dimension. To achieve optimal results, we propose a novel, optimized training strategy, and workflow. The proposed method is applied to line-space gratings, which serve as a proxy structure for FinFET transistor gates. The combination of the line-space gratings and FinFET structures indicates the relevance of this use-case, i.e., the results of this experiment can be directly applied to other relevant transistor structures and construct.

The contributions of this paper are as follows:

- A reliable algorithm for accurately predicting the surface of the wafer, utilizing a combination of data sources, such as experimental and synthetic SEM images as well as scatterometry data.
- The use of a tailored unsupervised domain adaptation method to bridge synthetic and experimental images, while preserving geometrical information, in the field of SEM nanometer-sized structures.

2 Related Work

2.1 Depth Estimation from SEM Images

In the past, various methods have been developed for estimating depth in SEM images. Most techniques are based on homography,^{7–10} which require two images taken from a different angle. This can be impractical in the semiconductor industry, where speed is a critical factor and using multiple images for measurements significantly lowers throughput. Also physical constraints in the SEM column restrict the angle and resolution of the electron beam. Other methods use multiple detectors,^{11,12} where depth is obtained from differences in signal strengths between detectors. However, these methods are constrained to using smooth surfaces only. Another approach is to exploit the focus change of the electron beam,^{13,14} similar to the depth of field effect in optical systems. Unfortunately, this technique is only feasible for the micrometer range and insufficient for most current applications in the semiconductor industry. Instead, in this work, we investigate methods that only require one SEM measurement during inference. These methods are mostly based on synthetic modeling.^{15–17}

2.2 Synthetic Data Generation

Synthetic databases are commonly used in machine learning, as label generation incurs little cost. Two examples of popular datasets in the field of computer vision are GTA5¹⁸ and FlyingThings3D.¹⁹ Methods that rely on synthetic data have been applied to various computer vision tasks, such as semantic segmentation, and depth estimation in various application fields.^{20,21} For SEM, there are also different publicly available datasets.^{22–24} However, these datasets often lack relevant ground-truth data for depth estimation, or have limited images available, making them not suitable for deep learning techniques.

2.3 Synthetic-to-Real Transfer

Domain adaptation is a widely used technique for closing the gap between real and synthetic data for which many methods have been proposed. These methods include reconstruction-based techniques^{25–27} that use generative models to enhance synthetic data, adversarial-based methods that generate synthetic target data which is related to the source domain,^{28–30} and discrepancy-based methods that promote a domain-invariant feature space.^{31,32} Also non-learnable techniques³³ are capable of bridging the gap, which can be useful when limited data are available. In this work, we use a reconstruction-based method, which is derived from a method proposed by Atapour-Abarghouei and Breckon.³⁴

2.4 Deep Learning for SEM Data

Deep learning has already been applied to various tasks relevant to SEM images, such as line edge roughness (LER) estimation,³⁵ denoising,^{36,37} and defect inspection.³⁸ More specifically, similar to the method presented in this work, networks with cyclic losses have also been applied to SEM data. Examples of the latter approach aim at enhancing the quality of the SEM images³⁹ or mapping the chip layout designs to SEM images.⁴⁰ These works demonstrate that the use of deep learning can be effective in improving prediction results on SEM data.

3 Methodology

The proposed method builds upon previous work on depth estimation for SEM images,⁴¹ where it was demonstrated that the height of on-chip structures can be predicted using neural networks trained with simulated SEM images. Several improvements to this method are made with this work. First, the synthetic geometry generation is enhanced to generate more realistic structures. Second, an unsupervised domain adaptation method is added to accommodate for the domain gap between synthetic and experimental images, which requires a training procedure to preserve the important information present in the images. Third, a more realistic use case is employed,

where the design of experiments is significantly upgraded. Specifically, multiple wafers with different pattern-height setpoints are developed and measured, as described in Sec. 4.

3.1 Domain Adaptation

The goal of domain adaptation is to maximize the performance of a specific task (e.g., depth estimation) on a target domain (e.g., experimental SEM images), even when ground-truth data are not available, by leveraging well-labeled data from a source domain (e.g., synthetic SEM images). The domains are typically related but have dissimilarities, which is referred to as a domain gap in machine learning. Empirically, we have discovered that inference performance on experimental images from the network is unsatisfactory (i.e., no realistic structures as output) when trained with synthetic images. This effect can be accounted for by a large domain gap, as previous research⁴¹ has shown that inference performance can be satisfactory when the domain gap is less pronounced.

As can be seen in Fig. 2, structural dissimilarities are present between the synthetic and experimental images. These differences arise from the fact that not all physical effects resulting from using SEMs are incorporated into the simulator. For example, there is some asymmetric edge blooming present in the images because the charge accumulates differently at downward and upward flanks relative to the scan direction. Also, a clear intensity drop in the middle of the line structure occurs in the experimental signal, which is caused by a mismatch between contrast and/or edge blooming with the simulator. The simulator is verified with bulk yield measurements, but both effects are dependent on the geometry. Another discrepancy is a parabolic curvature effect appearing in the experimental data, suspected to be the result of how the electron beam propagates through the electron-optical column.

Since there are no dense height labels available in the experimental domain, an unsupervised domain adaptation technique is used to bridge the domain gap. In this work, we employ CycleGAN, an architecture introduced by Zhu et al.⁴² The cyclic loss in this architecture bypasses the need for defining exactly matched training pairs in the dataset. CycleGAN consists of two generative adversarial networks (GANs), one responsible for translating images from domain *A* to *B*, and one for the inverse translation. Domain *A* is defined as the experimental domain and *B* as the synthetic domain. During the final training procedure, only one of the trained generators is used for inference.

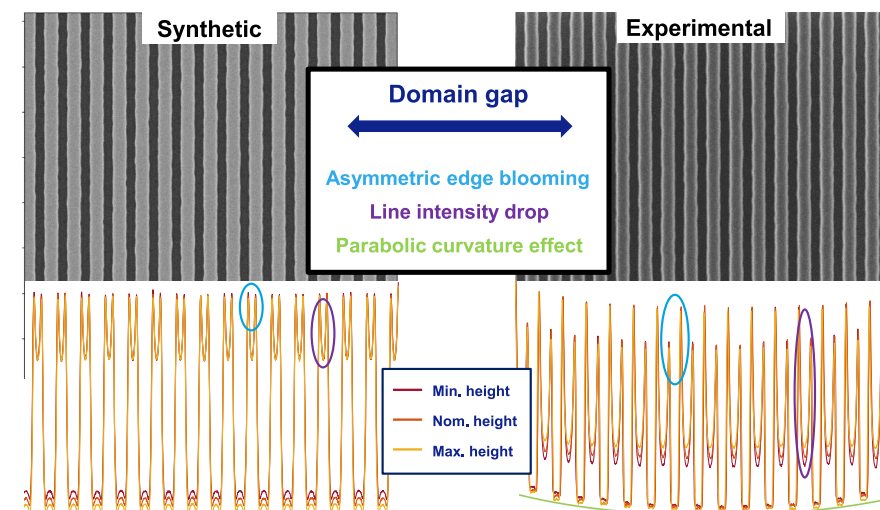


Fig. 2 Qualitative visualization of the domain gap present between the images. An example of a synthetic and experimental SEM image is depicted. The noiseless SEM profiles, averaged in the direction parallel to the lines, are displayed below. The main contributing effects of the domain gap are specified and highlighted.

3.2 Steps of the Method

An overview of all important steps is shown in Fig. 3. First, a synthetic dataset is generated that matches the experimental data as closely as possible. A small portion of the experimental data (in this case a separate test wafer) together with the synthetic data is used to train the domain adaptation network. The synthetic data are employed to train a surface-prediction network, which is a modified version of the generator network of the work of Wang et al.⁴³ Then a pipeline is constructed of both networks, where an experimental image is first converted to a synthetic image with the domain adaptation network and then into a height map with the surface-prediction network. The resulting height map is then calibrated with a corresponding scatterometry measurement H_{gt} . After this, the final prediction network is trained, using the weights of the previously trained surface-prediction network as a starting point. We train with the original experimental data, paired with the output from the described pipeline, i.e., the calibrated height maps. This approach combines the experimental input with the final output into a single network, which is able to establish a correct information mapping and account for the possible loss of information in the domain adaptation step. The next section elaborates further on this. Another benefit of this approach, compared with the method presented by Atapour-Abarghouei and Breckon,³⁴ is that the final inference path consists of a single network, which is beneficial for speed optimization during high-volume manufacturing.

3.3 Geometrical Information Preservation

The pipelined steps that generate the height maps used for training the final prediction network are designed to preserve known geometrical properties. However, the domain adaptation network does not have complete control over the image translation. Therefore, we focus on the feature preservation in both the 2D and height information, in the xy - and z -directions, respectively.

3.3.1 2D information

We have observed that edge information when applying CycleGAN roughly coincides during domain transfer, when learning from similar data in both domains. Therefore, we closely match

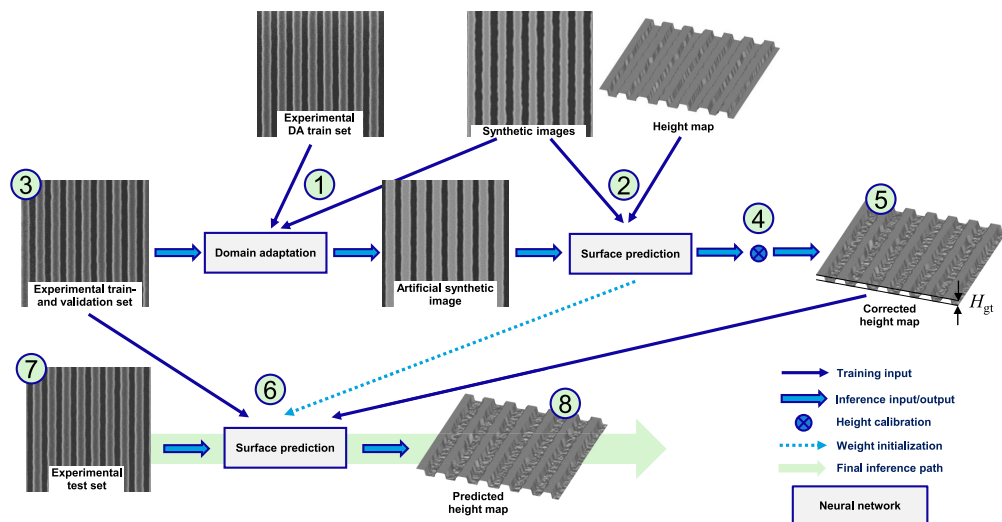


Fig. 3 Overview of the training method: First, the domain adaptation model (1) and surface-prediction model (2) are trained. Then experimental data are supplied into a pipeline of these models (3). The resulting depth maps are calibrated (4) with height H_{gt} from a scatterometry tool. Then a final prediction network is trained (6) with experimental data (3) in combination with the calibrated height maps (5). This network is used for the final inference of the 3D surface reconstruction of the chip (8).

the geometry edge-profile distribution of the synthetic images to the experimental images. This is achieved by generating the line-space roughness patterns using a statistical method, which is further explained in Sec. 4. However, in this case, we want to perfectly match the line-width (CD) information in the SEM image before and after domain adaptation. This is achieved by adding an extra edge-preserving loss to the network. The loss is placed over the generators (G) of the GAN because it uses the input image (I) and output image $G(I)$. This cross-domain loss is defined as

$$\mathcal{L}_{\text{edge}} = \|W_x(I - G(I))\|_1 + \|W_y(I - G(I))\|_1, \quad (1)$$

where $W_x = \frac{1}{e^{\tilde{I}^+ |\frac{dI}{dx}|}} - 1$ and $W_y = \frac{1}{e^{\tilde{I}^+ |\frac{dI}{dy}|}} - 1$. Furthermore, \tilde{I} is the input image (I) convolved with a Gaussian kernel ($\sigma = 3$), and \tilde{I}^+ , which is used in the denominator of the exponent fraction, is the maximum value of the gradient matrix that normalizes the exponent. This loss can be interpreted as a weighted L1-loss, where similarity is enforced at edges present in the input image, which results in strict 2D information preservation. This loss $\mathcal{L}_{\text{edge}}$ is added as a weighted sum to the other losses present in the network.

3.3.2 Height information

The height information is captured within the contrast of the image, which is texture information and is likely to be lost during the domain adaptation step. To correct this, we scale the resulting height maps in the z -direction with a value from a corresponding scatterometry measurement H_{gt} . Despite being an average value, this is valid to exploit, since the properties of a lithographic multilayer etch process implies that the structure height within the field of view of a scatterometry measurement is quite constant. The method used for this is part of pixel-wise fine-tuning,⁴¹ where the histogram of the height map is calculated. This results in two peaks at the bottom/top levels of the structure, which are scaled with H_{gt} distance in between.

4 Experiments

In the subsequent sections, we will outline the specific application, describe the methods used for data collection and processing, how the experiment was conducted, as well as some implementation details of the algorithm.

4.1 Application

The following use case is considered. During the front-end-of-line, individual transistors are patterned onto the chip. Modern transistors (FinFETs) consist of fins, which are rectangular structures mapped onto the substrate. The height of these structures is an important parameter, as it partly defines the switching voltage, stray fields, and parasitic capacities between transistors. We examine after-etch line space dielectric gratings, which serve as a proxy geometry for this use case.

The first step is to construct geometries on wafers, as depicted in Fig. 4. Initially, two layers of material are deposited onto the wafer. Then a lithography step is carried out, followed by an etching step, which yields the final structure. Several wafers are produced at around three different height levels with varying CD values.

4.2 Experimental Data

Initially, all wafers are measured at 60 or 240 different fields using a SEM tool. The measurement times of the sparsely and densely sampled wafers were around 40 min and 2 h, respectively. Furthermore, there is a gap of 12 days between the measurements of the sparsely and densely sampled wafers. Important SEM parameters are a landing energy of 800 eV, 3.3 nm full-width at half-maximum (FWHM) spotsize and a resolution of 0.8 nm per pixel. We have also validated one structure with a cross-sectional SEM (XSEM) measurement, useful for tuning the

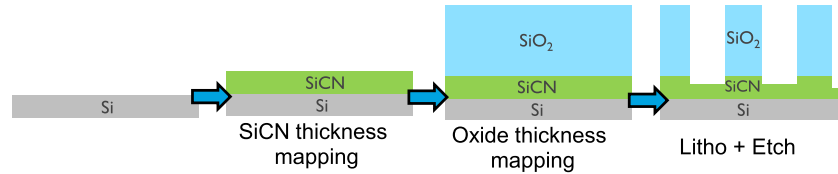


Fig. 4 Overview of the steps involved in placing a grating structure on the wafer.

parameters of the created geometry, which is discussed later. Subsequently, scatterometry measurements are captured at the same locations as the SEM measurements. Unfortunately, we could not realize to converge to a parametric grating model on the raw pupil data coming from the machine. Instead, surrogate labels have been used, where the heights are measured of an open pad (part on the measured chip where only the SiCN material is present) and a closed pad (part on the measured chip where the full-oxide layer is present). The total height label is then calculated by

$$H_{\text{gt}} = H_{\text{SiO}_2} + (H_{\text{SiCN closed}} - H_{\text{SiCN open}}). \quad (2)$$

By comparing this combination of measurements with the real grating height from a few TEM measurements, we have observed a small global offset of <1 nm, which can be compensated. We have also found that this offset is independent of the CD of the structure, which means that the measurements are correct relative to each other.

4.3 Synthetic SEM Data

After acquiring the experimental measurements, closely matching synthetic geometries are created. Upon examining the XSEM image, we noticed that the substrate layer (SiCN) is also partly etched. As a result, the parametric model displayed in Fig. 5 is used, where the etch height is divided into two parts, h_1 and h_2 . The observed dimensions from the experimental data, along with the corresponding generation parameters derived from these, are listed in Table 1. We allow for a small sidewall angle (sloped edge profile) and top and bottom corner roundings. We introduce LER to the geometries along the lines by adopting the Thorsos method from Mack.⁴⁴ This method is based on the power spectral density (PSD), where the autocorrelation is approximated by

$$R(x) = \sigma^2 e^{-(|x|/l_c)^{2\alpha}}. \quad (3)$$

This approach is used to make the geometries as realistic as possible. Parameters l_c , α , and σ represent the correlation length, roughness factor, and standard deviation, respectively. All parameter values are sampled uniformly from the defined ranges in Table 1. Even though a Gaussian distribution would be a more realistic model for some parameter distributions, a uniform distribution yields a more diverse dataset, with more edge cases as input to the network. We have created two geometries for every height in the specified range, with steps of 0.2 nm. The total amount of created geometries is 452, as depicted in Fig. 5.

For the simulation part, Nebula⁵ is utilized. The simulation parameters used include a landing energy of 800 eV, a pixel size of 0.8 nm, and a spot size of 3.3 nm FWHM, matched with the experimental SEM data. The applied materials are indicated in Fig. 5. Additionally, some pre-processing is applied on the SEM images prior to starting the training process. First, the image values are scaled between 0 and 255 to fully exploit the 8-bit dynamic range. Second, detector noise is added, which has a significant contribution to the image SNR in the present operating range. This noise is modeled as an additive Gaussian distribution and tuned to the experimental data.

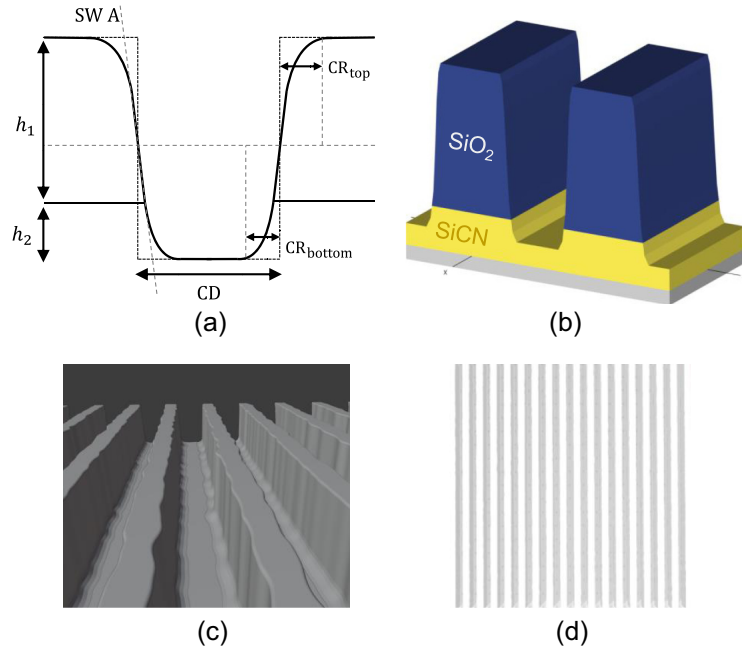


Fig. 5 Geometry creation process. (a) The parametric model used to generate the synthetic data. (b) 3D crop of the line pattern constructed with the materials used. (c) 3D rendering of the geometry created with visible roughness. (d) Top view of the full geometry.

Table 1 Measured and simulated ranges of the parametric model. The simulated range is a superset of the actual range. Rounding is defined as a ratio of the CD: $2 \frac{CR}{CD} \times 100\%$. In simulations, the thickness of the SiCN layer was set to infinity, as the maximum penetration depth of the electrons is less than the minimum thickness of the layer.

Parameter	Actual range	Simulated range
Trench height ($h_2 + h_1$) (nm)	108 to 137	100 to 145
Line mid-CD (nm)	56 to 64.5	53 to 67
Bottom rounding	Unknown	0% to 70%
Top rounding	Unknown	0% to 25%
SiCN thickness	29 to 36 nm	∞
Etch ratio (h_2/h_1)	0.28 to 0.37	0.25 to 0.40
Sidewall angle	Unknown	85 to 90 deg
Roughness amplitude (σ) (nm)	1.2 to 1.4	1 to 1.5
Roughness corr. length (l_c)	18 to 21	16 to 24
Roughness factor (α)	0.5	0.5

4.4 Dataset Construction

This experiment utilizes six wafers, each containing three reference heights. For each height, a densely (210 fields) and sparsely (60 fields) measured wafer is available. The densely measured wafers are selected for training, whereas the sparsely measured wafers are reserved for testing. The training and test sets are separated on a per-wafer basis to avoid data leakage in the model. Figure 6 represents an overview of the height and CD plot distributions in the dataset. It is important to note that there is no obvious correlation between height and CD in this dataset. This lack

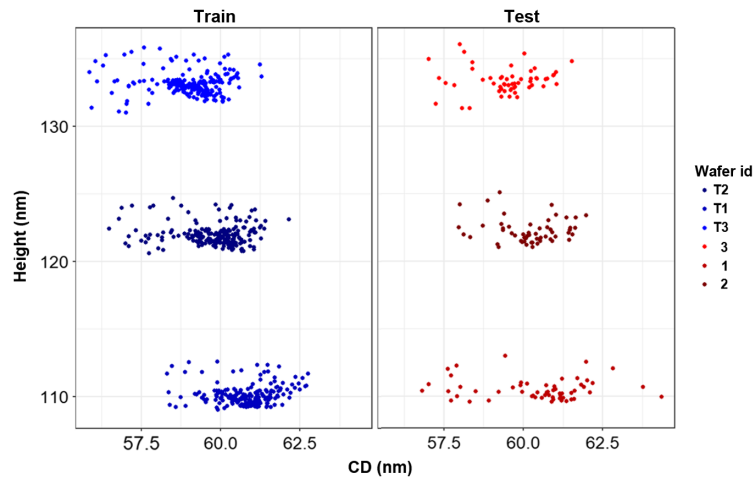


Fig. 6 Distributions of labels in the average CD and height variation in the dataset. Every dot represents a field.

of correlation is crucial, as it prevents the network from using CD as a basis for height prediction, which is undesirable.

4.5 Implementation Details

The domain adaptation network is implemented using the default settings of CycleGAN, with 64 generator filters. The identity loss is enabled with a weight of $\lambda = 0.5$. The weight of the reconstruction is set to $\lambda = 10$ and the weight of the edge-preserving loss is set to $\lambda = 0.1$. The network is trained for 400 epochs using 50 images from both domains. The learning rate is set to 0.0002 with a linear decay after 200 epochs.

The surface-prediction network is a modified version of the Pix2PixHD algorithm⁴³ that uses six Resblocks and a one-layer encoder and decoder, which downsamples and upsamples the 265×265 -pixel input to 128×128 pixels. The \mathcal{L}_1 loss is used only as the optimization objective, while the discriminator is dropped. During pretraining, the network is trained for 150 epochs, with an initial learning rate of 0.0002 and a linear decay after 100 epochs. In the final fine-tuning step, the network is trained for 200 epochs with a learning rate of 0.0001, and a linear decay after 100 epochs. Random cropping is always used as a data augmentation technique. When solely training with synthetic data, also random rotation and flipping are enabled. The reason for this is that experimental data does contain artifacts that are dependent on the scan direction. All networks are trained on a single Tesla V100 GPU.

5 Results

This section first provides the results of the domain adaptation method. These results compare the CycleGAN method with and without the edge-preserving loss, to demonstrate the significance of the adaptation to the training procedure. Then the qualitative and quantitative results of the overall method are presented. Afterward, a comparison of comparable training procedures from the literature is given.

5.1 Unsupervised Domain Adaptation

Results of the domain adaptation network after training with a separate wafer, are shown in Fig. 7. As can be observed, the domain gap is significantly reduced. The two main contributors, asymmetric edge blooming and the line-intensity drop, are no longer present. Since we use only the central 1024×1024 pixels of the image for further processing, the parabolic curvature effect is negligible. We have evaluated the preservation of 2D features during inference by applying a

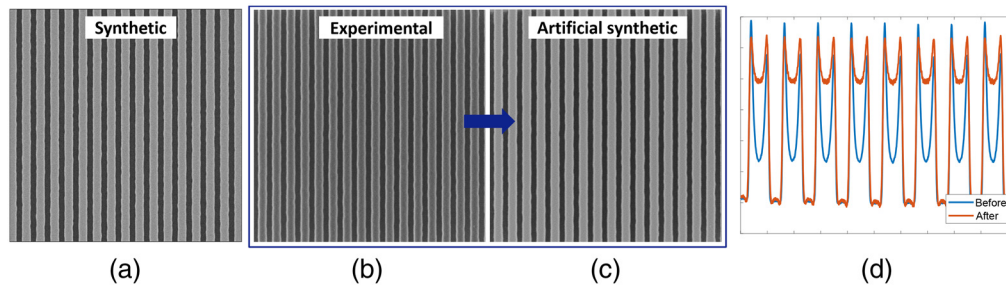


Fig. 7 Example result of the domain adaptation step: (a) arbitrary synthetic image for comparison; (b) experimental image; (c) generated synthetic image conditioned on the experimental image; and (d) average SEM signal before and after domain adaptation.

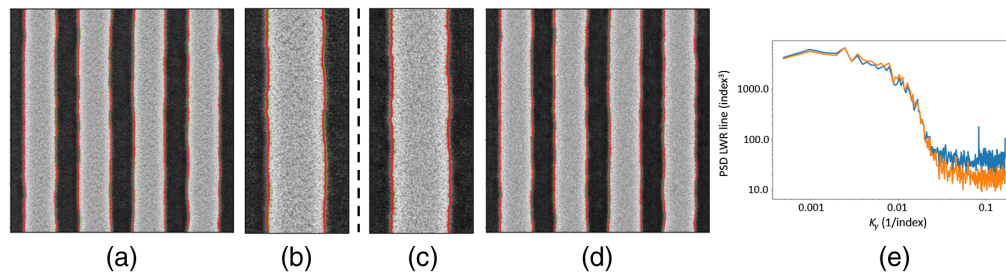


Fig. 8 Experimental results on 2D information preservation during the domain adaptation step. (a)–(d) The SEM edge detection algorithm results projected over the artificial synthetic image. A red line (before domain adaptation) is plotted over the green line (after domain adaptation). The more the green line is overwritten, the better. (a) CycleGAN results,⁴² (b) zoomed version of (a), (d) CycleGAN with edge-preserving loss Eq. (1), (c) zoomed version of (d), and (e) PSD analysis of the edges. Blue is before and orange is after domain adaptation.

contouring algorithm designed for SEM applications on both the experimental and the corresponding translated images from the synthetic domain. We have compared two implementations of CycleGAN, one with the original loss functions and one with the edge-preserving loss included. Then we have measured the average CD values of the validation set (~ 20 images) before and after conversion with and without the edge-preserving loss. The measured average CD difference is 0.59 and 0.09 nm, respectively, indicating that the 2D preservation performance is improved over 6 times. The results of the best-performing method are presented in Fig. 8. It can be observed that the borders before and after domain adaptation coincide with sub-pixel accuracy. Additionally, by analyzing the PSD signal generated with the method described by Pu et al.,⁴⁵ it is found that the noise floor only differs slightly.

5.2 Qualitative Results

Figure 9 shows an SEM image together with the height map from each wafer. Realistic linespace patterns are obtained. The height differences between the wafers (about 10 nm) are hardly noticeable in the SEM image by visual inspection. However, with the depth prediction, this can be clearly observed.

Figure 10 shows a comparison of a predicted geometry pattern, obtained by averaging in one direction, and a TEM measurement from a wafer cross section. The profile of the predicted pattern closely matches the reliable part of the image. Notably, the predicted profile is only adjusted through horizontal and vertical translation, without changes in size or shape. This indicates that detailed features, such as sidewall angle and corner rounding, will also be predicted realistically. This outcome highlights the value of using simulated data for pretraining and offers a promising perspective for detailed geometry reconstruction.

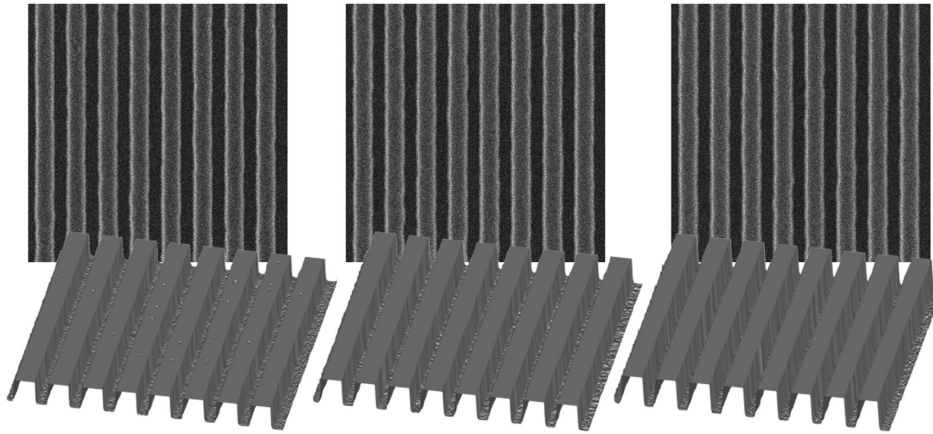


Fig. 9 SEM image from each test wafer with the corresponding reconstructed surface.

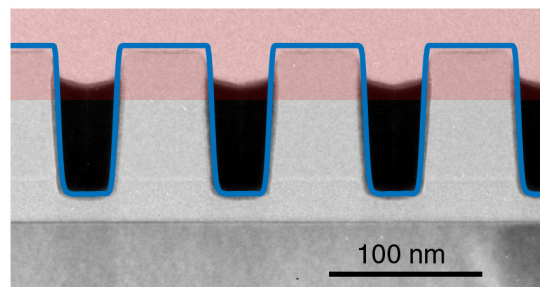


Fig. 10 Comparison of a predicted profile (blue line) projected on a TEM measurement of a sliced wafer. The comparison is performed on the same section of the wafer, not strictly at the same place. The top (red) part is not reliable to compare because of further processing steps, such as polishing, which are not considered in this work, have altered the top of the geometry after the SEM measurement.

5.3 Quantitative Results

Figure 11 presents the final prediction results of the average height of every SEM image measured on all three test wafers. The average height metric is obtained by calculating the distance between the peaks of the histogram of the resulting height map. The imposed height stems from the scatterometry measurement. Every SEM image is classified in the cloud of points of their wafer, which makes the interwafer classification score 100%. It is evident that correlation is present within a wafer, as the R^2 scores are between 0.76 and 0.81. Also the average error is about 0.5 nm.

Figure 12 shows the final prediction result of the average CD of every SEM image in the dataset for the final method. The CD metric is obtained by imposing a threshold on the depth map at 60% and then dividing the amount of pixels above the threshold by the total amount of pixels. The 60% is related to the threshold of common CD algorithms from the industry. The imposed CD is coming from the CD label, which is derived from a CD analysis on an SEM image. As can be noticed, a clear correlation between imposed and obtained CD is visible, indicating that the CD information is preserved during the training process. However, there is a small offset present between wafers, which remains unresolved at this point.

5.4 Performance of Other Methods

We develop a comparison between five different methods. (1) Pix2pixHD⁴³ trained on synthetic data, without any fine-tuning. (2) Pixel-wise fine-tuning,⁴¹ as presented in the previous work. (3) Pixel-wise fine-tuning with an additional domain adaptation step, as discussed by Atapour-Abarghouei and Breckon.³⁴ Here the method is supplemented with a default CycleGAN

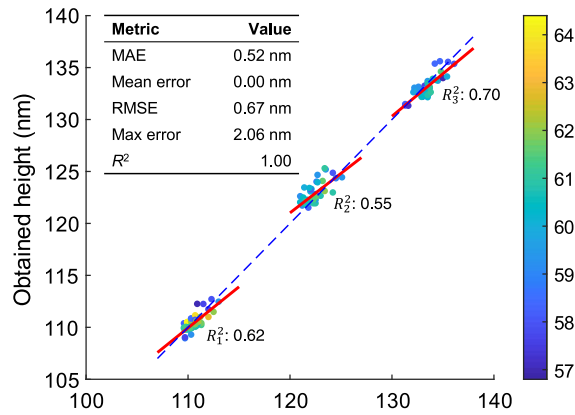


Fig. 11 Results of the three test wafers on the height metric, showing inter- and intra-wafer performance. The red line is a linear fit per wafer. The blue dotted line is exactly $y = x$, indicating perfect correspondence. The color bar indicates the CD.

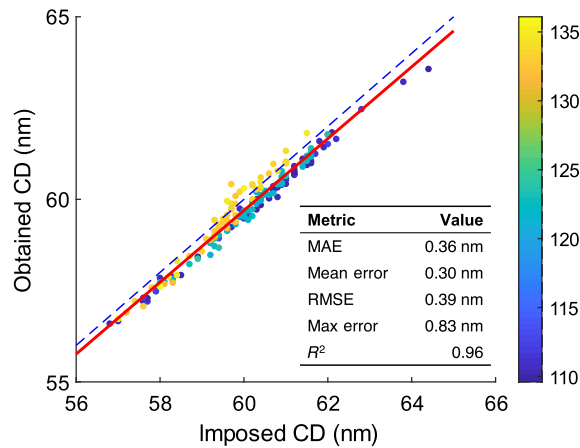


Fig. 12 CD performance of the entire dataset. Every dot represents a measurement. The red line is a linear fit over all measurements. The blue line is exactly $y = x$, indicating perfect correspondence. The color bar indicates the height.

network, which is used during training and inference. (4) The previous method (3) but extended with customized training procedure described. The emphasis of this procedure is to preserve the height information. (5) The proposed novel method. This procedure also adopts the edge-preserving loss in the domain adaptation network, which improves geometric information preservation in all directions. The results of this method have already been discussed in Figs. 9–12.

Figure 13 and Tables 2 and 3 present a comparison between the different methods, including the proposed technique. As can be observed, the first method, which does not employ fine-tuning, has very inaccurate height predictions. Additionally, since no domain adaptation method is used in the first two methods, realistic surfaces are only obtained in methods 3 and 4. For the other methods, it is visible that a correlation is present. However, because no domain adaptation is applied and shapes are unrealistic (method 2) and height information is lost in the domain adaptation step (method 3), results are not optimal. Method 3 seems to do a classification per wafer, where the majority of the points are predicted correctly (lowest MAE), but this method lacks height sensitivity, as R^2 statistics are poor. The final methods 4 and 5 show the best performance on height because the sensitivity (R^2) is closest to unity, which is the most important metric for process control.

The CD results are interesting since we observe a per-wafer bias when the final training step is directly performed on experimental data. This effect has been mitigated by incorporating an

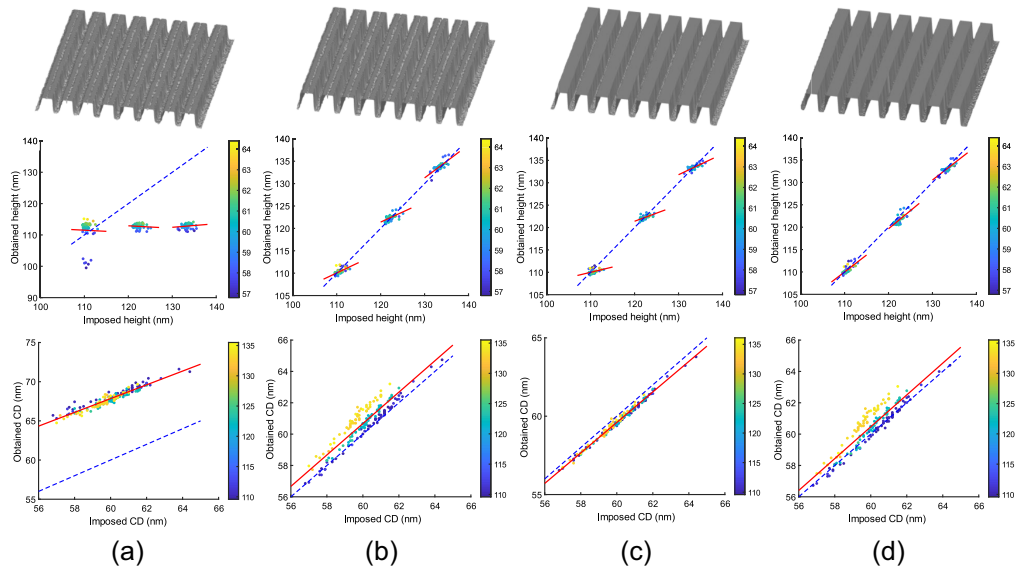


Fig. 13 Qualitative and quantitative results of other methods. Top row: Example of a predicted height map. Middle row: correlation plot of the average height. Bottom row: correlation plot of the average CD. (a) Pix2pixHD, (b) pixel-wise fine-tuning, (c) pixel-wise fine-tuning with Atapour et al., (d) pixel-wise fine-tuning with a custom training method, (e) and pixel-wise fine-tuning with custom training method and edge-preserving loss.

Table 2 Numeric comparison of all five methods on height metrics. MAE, mean error (ME), root-mean-squared error (RMSE), and maximum error (MAXE), all in nanometers. R^2 expresses the coefficient of determination of the linear fit of all measurements in the test set. Metrics R_1^2 , R_2^2 , and R_3^2 express the fit per wafer.

No.	Method	MAE ↓	ME ↓	RMSE ↓	MAXE ↓	R^2 ↑	R_1^2 ↑	R_2^2 ↑	R_3^2 ↑
1	Pix2PixHD	10.86	9.41	13.15	23.66	0.06	0.00	0.01	0.02
2	Pixel-wise fine-tuning	0.55	-0.11	0.71	2.19	1.00	0.36	0.36	0.58
3	2 + Atapour et al.	0.50	0.17	0.68	2.41	1.00	0.16	0.40	0.51
4	3 + custom training	0.57	0.37	0.73	2.05	1.00	0.57	0.59	0.69
5	4 + edge preservation	0.52	0.00	0.67	2.06	1.00	0.62	0.55	0.70

Table 3 Numeric comparison of all methods on all CD metrics. MAE, ME, RMSE, and MAXE, all in nanometers. R^2 expresses the coefficient of determination of the linear fit of all measurements in the test set. Metrics R_1^2 , R_2^2 , and R_3^2 express the fit per wafer.

No.	Method	MAE ↓	ME ↓	RMSE ↓	MAXE ↓	R^2 ↑	R_1^2 ↑	R_2^2 ↑	R_3^2 ↑
1	Pix2PixHD	7.84	-7.84	7.85	9.43	0.91	0.95	0.91	0.92
2	Pixel-wise fine-tuning	0.69	-0.67	0.87	2.02	0.85	1.00	0.99	0.98
3	2 + Atapour et al.	0.40	0.40	0.42	0.64	0.99	1.00	0.99	0.98
4	3 + custom training	0.54	-0.47	0.73	1.93	0.96	0.99	0.98	0.97
5	4 + edge preservation	0.36	0.30	0.39	0.83	0.96	0.99	0.98	0.97

edge-preserving loss in the proposed method. However, we consider the results from method 3 remarkably well, although this method shows clearly higher error metrics in the vertical direction (covered by R_1^2 to R_3^2).

6 Conclusion

This paper extends 2D to 3D using computer vision for full-surface reconstruction. We have presented a data-driven method that predicts the dimensions, height and width (CD) values, of fin-like structures. The results show that we can handle situations where a significant domain gap occurs between data domains, which allows extracting heights from SEM images on more diverse experimentally obtained data. The obtained results are closing the gap between the synthetic and experimental SEM images.

The intra-wafer height prediction results indicate that it is possible to estimate height values with a mean absolute error (MAE) of 0.5 nm, which is within specification for various modern applications. Also the height sensitivity is presented with the method proposed, as the R^2 values are between 0.55 and 0.70. The SNR values of SEM images and the calibration sensor may play a role in further improving the results. The resulting algorithm for interwafer height prediction, utilizing a combination of multimodal data sources, can produce realistic height values with a maximum absolute error of 2.06 nm and a MAE of 0.5 nm, which makes it a suitable contribution to the metrology application field. However, the available data are crucial as the prediction can become biased when the data samples are not homogeneously distributed, which is also partially presented in our case. We anticipate that when a more balanced dataset is available, the average error metrics could be further improved, or more general methods will be sufficient for solving the problem.

One important question that needs to be addressed is the generalizability of the proposed method across different scenarios and its impact on performance, as this algorithm is expected to generate accurate results under varying conditions. First, it is expected that this algorithm can handle morphological changes, especially when the simulated data are diverse enough. In fact, previous work⁴¹ demonstrated that the network can predict structures that were not included in the training set, such as defects. It is also known that the presented method is able to generalize across different wafers measured with multiple days in between. Nevertheless, a proper study of robustness of this method on machine drift is still an open point of research.

As previously discussed, the reliance on calibration data is still presented, mainly due to the domain shift between synthetic and experimental data. In this particular case, we use scatterometry, which limits the applicability of this method on repetitive structures. However, other local tools, such as TEM and AFM, could be employed, which would broaden the scope of the method considerably, but hands in on speed and practicability. Another possible option is to examine the extent to which grating measurements close to nonrepetitive parts (e.g., logic) are representative, which would potentially reduce the need for local calibration. Additionally, it is postulated that by utilizing better forward models of the SEM process, the dependence on an external calibration source could be decreased and eventually eliminated.

Compared to a basic average height regression algorithm, the proposed method has the significant advantage that local information and feature details can be derived from a full-depth map. This implies that not only CD and height but also LER, SWA, and corner-rounding values as well as the dimensions of defects can be measured. This is illustrated with a qualitative comparison of a TEM image with a prediction result. Further experiments for quantitative comparisons would be useful for future work.

Acknowledgments

We would like to thank the Interuniversity Microelectronic Centre (IMEC), Leuven, Belgium, for conducting all measurements for this experiment. We would also like to acknowledge several experts from ASML, Ilirjan Aliaj, Ton Kiers, Roy Anuciado, and Richard van Haren for arranging the experiments and useful discussions. The research leading to these results received funding from Topsector Holland High Tech (Grant No. TKI-HTSM/20.0396). The first three authors hold patents related to this work.

References

1. J. Kim et al., “An application study on the stochastic effect of EUV photons,” *Proc. SPIE* **11323**, 341–348 (2020).
2. C. Kübel et al., “Recent advances in electron tomography: TEM and HAADF-stem tomography for materials science and semiconductor applications,” *Microsc. Microanal.* **11**(5), 378–400 (2005).
3. K. Lepinay and F. Lorut, “Three-dimensional semiconductor device investigation using focused ion beam and scanning electron microscopy imaging (FIB/SEM tomography),” *Microsc. Microanal.* **19**(1), 85–92 (2013).
4. A. J. Den Boef, “Optical metrology of semiconductor wafers in lithography,” *Proc. SPIE* **8769**, 57–65 (2013).
5. L. van Kessel and C. Hagen, “Nebula: Monte Carlo simulator of electron–matter interaction,” *SoftwareX* **12**, 100605 (2020).
6. M. T. Postek and A. E. Vladár, “Does your SEM really tell the truth? How would you know? Part 4: charging and its mitigation,” *Proc. SPIE* **9636**, 963605 (2015).
7. G. Piazzesi, “Photogrammetry with the scanning electron microscope,” *J. Phys. E: Sci. Instrum.* **6**(4), 392 (1973).
8. S. Roy et al., “Automatic 3D reconstruction of quasi-planar stereo scanning electron microscopy (SEM) images,” in *Annu. Int. Conf. IEEE Eng. in Med. and Biol. Soc.*, IEEE, pp. 4361–4364 (2012).
9. C. Valade et al., “Tilted beam SEM, 3D metrology for industry,” *Proc. SPIE* **10959**, 171–188 (2019).
10. A. P. Tafti et al., “A comparative study on the application of SIFT, SURF, BRIEF and ORB for 3D surface reconstruction of electron microscopy images,” *Comput. Methods Biomed. Eng. Imaging Visual.* **6**(1), 17–30 (2018).
11. W. Sówko and M. Krysztof, “Detector system for three-dimensional imaging in the variable pressure/environmental SEM,” (2013).
12. M. Yoshikawa et al., “Novel three dimensional (3D) CD-SEM profile measurements,” in *China Semicond. Technol. Int. Conf.*, IEEE, pp. 1–3 (2015).
13. C. Dahmen, “Focus-based depth estimation in the SEM,” *Proc. SPIE* **7266**, 517–524 (2008).
14. N. Marturi, S. Dembélé, and N. Piat, “Depth and shape estimation from focus in scanning electron microscope for micromanipulation,” in *Int. Conf. Control, Autom., Robot. and Embedd. Syst. (CARE)*, IEEE, pp. 1–6 (2013).
15. A. E. Vladár et al., “10nm three-dimensional CD-SEM metrology,” *Proc. SPIE* **9050**, 53–63 (2014).
16. K. T. Arat et al., “Estimating step heights from top-down sem images,” *Microsc. Microanal.* **25**(4), 903–911 (2019).
17. J.-H. Seo et al., “Non-destructive depth measurement using SEM signal intensity,” *Proc. SPIE* **11611**, 699–706 (2021).
18. S. R. Richter et al., “Playing for data: ground truth from computer games,” *Lect. Notes Comput. Sci.* **9906**, 102–118 (2016).
19. N. Mayer et al., “A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 4040–4048 (2016).
20. R. Barth et al., “Synthetic bootstrapping of convolutional neural networks for semantic plant part segmentation,” *Comput. Electron. Agric.* **161**, 291–304 (2019).
21. M. Frid-Adar et al., “Synthetic data augmentation using gan for improved liver lesion classification,” in *IEEE 15th Int. Symp. Biomed. Imaging (ISBI 2018)*, IEEE, pp. 289–293 (2018).
22. A. P. Tafti et al., “3DSEM: a 3D microscopy dataset,” *Data Brief* **6**, 112–116 (2016).
23. R. Wilson et al., “REFICS: a step towards linking vision with hardware assurance,” in *Proc. IEEE/CVF Winter Conf. Appl. of Comput. Vis.*, pp. 4031–4040 (2022).
24. R. Aversa et al., “The first annotated set of scanning electron microscopy images for nanoscience,” *Sci. Data* **5**(1), 1–10 (2018).

25. J. Hoffman et al., "CYCADA: cycle-consistent adversarial domain adaptation," in *Int. Conf. Mach. Learn.*, PMLR, pp. 1989–1998 (2018).
26. A. Shrivastava et al., "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 2107–2116 (2017).
27. S. R. Richter, H. A. Al Haija, and V. Koltun, "Enhancing photorealism enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(2), 1700–1715 (2022).
28. Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.* **17**(1), 2096–2030 (2016).
29. Z. Ren and Y. J. Lee, "Cross-domain self-supervised multi-task feature learning using synthetic imagery," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 762–771 (2018).
30. E. Tzeng et al., "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 7167–7176 (2017).
31. A. Rozantsev, M. Salzmann, and P. Fua, "Beyond sharing weights for deep domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(4), 801–814 (2018).
32. Y. Zhang et al., "A curriculum domain adaptation approach to the semantic segmentation of urban scenes," *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(8), 1823–1841 (2019).
33. Y. Yang and S. Soatto, "FDA: Fourier domain adaptation for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 4085–4095 (2020).
34. A. Atapour-Abarghouei and T. P. Breckon, "Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 2800–2810 (2018).
35. N. Chaudhary, S. A. Savari, and S. S. Yeddulapalli, "Line roughness estimation and poisson denoising in scanning electron microscope images using deep learning," *J. Micro/Nanolithogr. MEMS MOEMS* **18**(2), 024001 (2019).
36. E. Giannatou et al., "Deep learning denoising of SEM images towards noise-reduced ler measurements," *Microelectron. Eng.* **216**, 111051 (2019).
37. B. Dey et al., "SEM image denoising with unsupervised machine learning for better defect inspection and metrology," *Proc. SPIE* **11611**, 245–254 (2021).
38. B. Dey et al., "Deep learning-based defect classification and detection in SEM images," *Proc. SPIE* **PC12053**, PC120530Y (2022).
39. Y. Midoh and K. Nakamae, "Image quality enhancement of a CD-SEM image using conditional generative adversarial networks," *Proc. SPIE* **10959**, 37–46 (2019).
40. H.-C. Shao et al., "From IC layout to die photograph: a CNN-based data-driven approach," *IEEE Trans. Comput.-Aid. Des. Integr. Circuits Syst.* **40**(5), 957–970 (2020).
41. T. Houben et al., "Depth estimation from a single SEM image using pixel-wise fine-tuning with multimodal data," *Mach. Vis. Appl.* **33**(4), 1–16 (2022).
42. J.-Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2223–2232 (2017).
43. T.-C. Wang et al., "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 8798–8807 (2018).
44. C. A. Mack, "Generating random rough edges, surfaces, and volumes," *Appl. Opt.* **52**(7), 1472–1480 (2013).
45. L. Pu et al., "Analyze line roughness sources using power spectral density (PSD)," *Proc. SPIE* **10959**, 624–631 (2019).

Tim Houben received his master's degree in electrical engineering from Eindhoven University of Technology. Currently, he is a PhD student at Eindhoven University of Technology and a visiting researcher in the Research Department of ASML. He works on 3D reconstruction for scanning electron microscopy.

Thomas Huisman received his master's degree in 2012 from the University of Twente and his PhD in 2016 from Radboud University, The Netherlands. In 2017, he joined ASML as a researcher.

Maxim Pisarenco earned his PhD in applied mathematics from Eindhoven University of Technology. Since 2011, he has been in the Research Department at ASML, where he is currently working as a senior scientist. His research interests include inverse problems, machine learning, model discovery, and holistic data- and physics-driven modeling.

Fons van der Sommen is an assistant professor at Eindhoven University of Technology and heads the Healthcare and High-Tech Cluster of the VCA Research Group. He has worked on a variety of image processing and computer vision applications, mainly in the medical domain, and strives to exploit signal processing and information theory methods to improve the robustness, efficiency, and interpretability of modern-day AI architectures.

Peter de With is a full professor and leads the VCA group at Eindhoven University of Technology. He has co-authored more than 70 refereed international book chapters and journal articles, more than 500 conference publications, and 40 international patents. He served as a technical committee member of the IEEE CES, ICIP, and SPIE and is a member of the Royal Holland Society of Academic Sciences and Humanities.