Accurate prediction of EUV lithographic images and 3D mask effects using generative networks

Abdalaziz Awad,[a,b,*] Philipp Brendel,[b] Peter Evanschitzky,[b] Dereje S. Woldeamanual,[c] Andreas Rosskopf,[b] and Andreas Erdmann[a,b]

[a]Friedrich Alexander University, Chair of Electron Devices, Erlangen, Germany
[b]Fraunhofer Institute for Integrated Systems and Device Technology, Erlangen, Germany
[c]Synopsys GmbH, Aschheim, Germany

Abstract

Background: As extreme ultraviolet lithography (EUV) lithography has progressed toward feature dimensions smaller than the wavelength, electromagnetic field (EMF) solvers have become indispensable for EUV simulations. Although numerous approximations such as the Kirchhoff method and compact mask models exist, computationally heavy EMF simulations have been largely the sole viable method of accurately representing the process variations dictated by mask topography effects in EUV lithography.

Aim: Accurately modeling EUV lithographic imaging using deep learning while taking into account 3D mask effects and EUV process variations, to surpass the computational bottleneck posed by EMF simulations.

Approach: Train an efficient generative network model on 2D and 3D model aerial images of a variety of mask layouts in a manner that highlights the discrepancies and non-linearities caused by the mask topography.

Results: The trained model is capable of predicting 3D mask model aerial images from a given 2D model aerial image for varied mask layout patterns. Moreover, the model accurately predicts the EUV process variations as dictated by the mask topography effects.

Conclusions: The utilization of such deep learning frameworks to supplement or ultimately substitute rigorous EMF simulations unlocks possibilities of more efficient process optimizations and advancements in EUV lithography.

© *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.20.4.043201]

Keywords: extreme ultraviolet lithography; generative networks; deep learning; mask topography effects.

Paper 21050 received Jun. 9, 2021; accepted for publication Oct. 20, 2021; published online Nov. 16, 2021.

1 Introduction

Extreme ultraviolet lithography (EUV) has become responsible for the majority of the advances in semiconductor technology toward the 10-nm technology node and beyond. Simulations play a crucial role in the advancement of EUV lithography and they are often the most efficient means of process development. However, accurate aerial image simulations of EUV lithography incur a significant computational load, which is largely attributed to the requirement of a rigorous solution of the electromagnetic fields (EMFs) with a 3D representation of the thick EUV mask.[1,2] The requirement of rigorous simulations poses a hindrance to optimizations and explorations of the parameter space. A 2D representation of the mask can be utilized to drastically speed up the

*Address all correspondence to Abdalaziz Awad, aziz.awad@fau.de

process while compromising accuracy using the so-called Kirchhoff approximation, which assumes the thick mask to be infinitely thin. The infinitely thin mask model has been commonly used to simulate lithographic aerial images. Nonetheless, such approximation overestimates the amount of light reaching the wafer, and it fails to represent numerous effects attributed to the 3D profile of the mask, often referred to as the mask topography effects. Furthermore, various compact mask models have been developed to increase the computational efficiency while maintaining sufficient accuracy in representing 3D mask effects, although the main assumptions of such models are invalidated at more advanced technology nodes.[3–5] Certain 3D mask effects or mask topography effects, such as asymmetric process windows, occur in both deep UV (DUV) and EUV lithography processes. Nonetheless, 3D mask effects manifest at a greater degree as the wavelength becomes larger in comparison to the feature sizes, and smaller in comparison to the thickness of the mask's absorber.[6] The aggregation of the 3D mask effects in EUV systems due to the smaller feature sizes makes them consume a greater part of the tolerable error budget compared to the DUV 3D mask effects.

Increasing research efforts in recent years have been committed to utilizing machine learning to efficiently perform a wide range of tasks in lithography. A diverse range of neural network architectures has been employed to surpass the computational bottleneck of EUV aerial image computation with rigorous mask models. Fully convolutional networks (FCNs) were used to efficiently calculate the near field spectrum of 3D EUV masks,[7] which is a highly time-consuming step of aerial image computation. Artificial Neural Networks were implemented to model the spectrum of the 3D mask, which is then used to compute the aerial image using the Abbe method.[8] Machine learning with non-parametric kernel regression has been utilized to compute aerial images with the aid of training libraries of thick-mask diffraction near-fields[9] and data-fusion and image-synthesis from small patches.[5] More recently, a convolutional neural network (CNN) is implemented to reconstruct the amplitudes of the 3D mask diffraction spectrum, which are then used to compute the aerial images.[10] Many of these machine learning implementations require additional computational steps to obtain the aerial images. Additionally, due to the data-intensive nature of such approaches, they often require copious amounts of training data to be trained effectively. The data requisites of machine learning can largely impede the efficiency when the training data involves rigorously simulated 3D mask model aerial images. Lin et al.[11] proposed an approach that tackles the data inefficiency by implementing active data selection in combination with transfer learning, where data from a simpler technology node is utilized to reduce the amount of training data the network required from the more advanced technology node.

Conditional generative adversarial networks (cGANs) are an archetype of generative networks that have been utilized in recent years for efficient aerial image generation, in addition to other lithographic image generation tasks such as generating optimized mask layouts in Gan-OPC.[12] The appeal of cGANs is their versatility in image-to-image translation problems as they have demonstrated their superior efficacy for a wide range range of such tasks.[13,14] LithoGAN was developed for an end-to-end lithography framework, using the cGAN architecture for aerial image generation.[15] Although LithoGAN has demonstrated high accuracy and efficiency, it is restricted to the 2D mask model for imaging, which is incapable of capturing the 3D mask effects and is therefore mostly invalid for various mask layout settings with feature sizes close to the wavelength. However, the incentive to include 3D mask effects with such efficient approaches grows. TEMPO is a framework proposed to predict lithographic aerial images at different resist heights with the mask topography taken into consideration.[3] The approach in the TEMPO framework involves appending an encoding vector that includes information of the target domain to the bottleneck layer of the network. TEMPO has demonstrated its effectiveness and accuracy in predicting the aerial images for different resist heights, however, the demonstrated accuracy of the framework does not take into account predicting the EUV process variations with the mask topography effects.

In this work, we propose a framework for efficiently training a generative network to generate aerial images that accurately represent the 3D mask model imaging, in addition to predicting the non-linearities and asymmetries in the behavior of an EUV process through focus and pitch. The framework is then capable of predicting the iso-dense bias, process window, best focus shift, and position shift for a given pattern layout within a varied range of industrially relevant mask

patterns. The framework entails providing the generative network with aerial images computed using the 2D mask model as inputs and aerial images computed using the 3D mask model as outputs. Additionally, we utilize the defocus from the projector as additional information for the network in a similar manner to the implementation of the one-hot encoding vector in the TEMPO framework.[3] In summary, the main contributions of our proposed framework to recent and similar efforts for 3D model aerial image generation are the following.

1. Framework to train cGANs to generate 3D mask model aerial images with high accuracy in terms of local lithographic metrics such as critical dimensions (CDs) as well as global imaging metrics such as the mean absolute error and the edge distance error (EDE).
2. Ability to generalize and predict aerial images of mask patterns that were not included in the training data.
3. Predictions of process variations of EUV images with 3D mask effects, such as best focus shifts, asymmetric process windows, horizontal–vertical feature bias, isolated-dense feature bias, in addition to focus-dependent feature position shifts and telecentricity.
4. A wide application range, given that a varied range of mask layout patterns such as contact holes, lines and spaces, line ends, and other layouts can be implemented in this framework.
5. Proposing means of increasing the effectiveness of cGANs for lithographic aerial image generation tasks by adjusting the loss ratios, and proposing another generative architecture that can outperform cGANs in such tasks.

In the following section, we present details of the physical phenomena associated with the 3D mask effects and their significance, in addition to details associated with the implemented network architectures. Section 3 details our proposed approach, specifics and important network parameters, and the training strategies that have been employed. Section 4 presents our results that are achieved by the network and compares them to simulated results using a rigorous EMF solver, followed by our conclusions in Sec. 5.

2 Preliminaries

2.1 *3D Mask Effects*

A characteristic of masks in EUV lithography is their thickness relative to the projection wavelength. EUV masks are composed of a mask absorber, which constitutes the mask's features, placed on top of a multi-layer mirror of 40 bi-layers. With the standard EUV wavelength being 13.5 nm, EUV mask absorbers have a thickness in the range of 60 nm, making them more than four wavelengths thick. The material properties of the bi-layer and absorber, in addition to the 3D profile of the thick EUV mask ensue several aberration-like effects and deformations to the wavefront and phase of the light as it interacts with the edges of the mask's features.[6] Additionally, the multitude of reflections of the diffracted light from the multilayer stack leads to further deformations at the masks near field and consequently at the resulting image.[6] The reflective nature of EUV masks necessitates an oblique incidence onto the mask, which causes an asymmetry across certain aspects of EUV imaging. The mask absorber being larger than the wavelength increases the severity of these effects. Another factor that aggregates the extent of these effects is the advancement of the feature sizes toward sizes approaching the wavelength or smaller than it. These effects are commonly referred to as the 3D mask effects, and they play a fundamental role in understanding and predicting the often non-linear behavior of an EUV lithography process.

One of the main 3D mask effects is an asymmetry in process windows. Process windows are indications of the operable ranges of defocus and dose (or threshold) variation that yield changes in the imaged feature sizes or CDs within a specified tolerance range. In other words, a process window is an indication of the dose and defocus latitudes of a given lithographic process. A symmetric process window for a given process means that shifts of the projector from the focal

plane that are either toward the mask or away from the mask yield the same change in the printed feature size. However, certain cases in both DUV and EUV lithography lead to asymmetric process windows, namely cases with semi-dense or isolated features. The degree of the asymmetry of a process window is dependant on the feature type and pitch. A primary consequence of the process window asymmetry is a layout-dependent shift in the best focus position which yields the image with the sharpest contrast. Orientation dependency is a 3D mask effect that is a result of the oblique incidence on the EUV mask. EUV projection systems presently employ a chief ray angle of incidence (CRA) of 6 deg. The CRA is defined in the yz-plane, whereas the mask surface is the xy-plane. Consequently, features that are oriented perpendicular to the incidence plane (horizontal lines) will experience different shadowing effects from features that are parallel to the incidence plane (vertical lines). This results in a pronounced asymmetric shadowing that occurs only with imaging horizontal features and not vertical ones.[6,16]

Non-telecentricity is a characteristic effect of EUV lithographic imaging that is attributed to the thick masks and the oblique incidence. Non-telecentricity is a focus-dependent shift of the feature position. The extent of the feature position variation through focus increases in magnitude as the angle of incidence (CRA) increases.[6] This 3D mask effect is numerically described by the gradient of the feature position versus focus curve at the best focus position.

A bias in imaging for dense features versus more isolated features as a result of the discrepancy in the number of contributing diffraction orders is a standard feature of projection lithography. In other words, isolated and dense features of the same size are projected with varying fidelities due to the diffraction limitation of the system. The iso-dense feature bias is one of the optical proximity effects that can be, to a certain extent, remedied by OPC techniques. EUV masks exacerbate the extent of the iso-dense feature bias. Furthermore, isolated EUV mask features are more sensitive to focus variation, which in turn further exacerbates the iso-dense bias in EUV lithography.

Accurate representations of the aforementioned 3D mask effects necessitate a rigorous numerical solution of the EMFs as they reflect from the thick EUV mask. Although 2D representations of masks have been widely used, at smaller technology nodes they become increasingly inaccurate. Therefore, the formulation of a viable deep learning model to be utilized for EUV systems requires the consideration of such 3D mask effects. Recent deep learning approaches for EUV lithography have tackled specific 3D mask effects by involving additional modeling or learning steps, such as in LithoGAN where an additional CNN was implemented to predict the position shift of the feature that is predicted by a cGAN.[15] In our approach, we aim to encompass all the mentioned 3D effects in the training of an efficient generative network.

2.2 *Generative Networks*

2.2.1 *Generative adversarial networks*

GANs are neural network models that are capable of generating and predicting instances of data as outputs that resemble those in the inputs of the network. The main characteristic of GAN models is that they consist of two competing neural networks with different, yet interlaced objectives. The two competing networks are a generator model and a discriminator model. The generator model is the one responsible for generating the data, and in the regular GAN case, it involves an unsupervised learning process. The discriminator model is tasked with classifying the data and determining whether they are real data from the input domain, or if they are fake data generated by the generator. These two models are trained alongside each other in an adversarially balanced manner, where a loss increase for one model contributes as a loss reduction for the other. The discriminator's objective is to accurately identify "real" and generated or "fake" data, while the generator's objective is to generate data that are as close as possible to the "real data" in the input domain, and to make the discriminator fail to differentiate the real and generated data.[17] The loss function of a general GAN model is defined in Eq. (1) as described by Goodfellow et al.[18]

$$\mathcal{L}_{GAN}(G,D) = E_x[\log(D(x))] - E_z[\log(1 - D(G(z)))], \tag{1}$$

where the generator (G) tries to minimize this loss while competing with the discriminator (D) that tries to maximize it. The term $D(x)$ is the discriminator's estimate of the probability that a data instance is real, x is the data from the real or input domain, $G(z)$ is the generator's output when given noise (z), and $D(G(z))$ is the discriminators estimate of the probability that a generated data instance is from the real. E_x and E_z are the expected value over all real data instances and over the random inputs to the generator, respectively. The generator works to minimize the overall loss through minimizing the term $\log(1 - D(G(z)))$.

2.2.2 *Conditional generative adversarial networks*

cGANs are ones where the generation is conditioned to a certain input structure, as opposed to regular GANs where the generation is based primarily on random input vectors. The generators of cGANs receive input data on which they perform a translation onto, with the aim of replicating the translation from inputs to outputs of the training data. In our case, the input data is an aerial image computed using the 2D mask model or thin-mask approximation, and the output is the aerial image of the same mask computed using the waveguide method, which is representative of typical simulation methods with a 3D mask description.[1,2]

cGANs provide a more effective general solution for image translation tasks (image-to-image mapping) as demonstrated by the paper on Image-to-Image translation with conditional adversarial networks.[13] This improved performance could be attributed to the conditioning of the output of the generator to the desired mapping provided in the input, which frames the problem as a supervised training task

The loss function for cGANs is analogous with the one for regular GANs; however, the definition of the inputs differs.

$$\mathcal{L}_{cGAN}(G, D) = E_x[\log(D(x, y))] - E_{x,z}[\log(1 - D(x, G(x, z)))]. \tag{2}$$

While x is the input data, and y is the real translation of x, y is also part of the input data. The objective function, which is shown in Eq. (4) below, is optimized using the same methodology discussed in the previous section. Here $G(x, z)$ is the image generated by the generator given a certain input image x, and a noise distribution z. The expression $D(x, y)$ denotes the discriminator's probability estimate that a data pair (x, y) is a real mapping or translation, as in a pair that is provided in the input data. The term $D(x, G(x, z))$ refers to the probability estimated by the discriminator that the data pair $(x, G(x, z))$ is a fake translation. Recent approaches mixed the GAN's adversarial loss with a more traditional loss such as the L1 distance shown in Eq. (3), as this is demonstrated to provide a reduced blur in the images.[13] An L1 loss is a mean absolute error and its inclusion incentivizes the generation of data that is closer to the target.

$$\mathcal{L}_{L1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1]. \tag{3}$$

This leads to the final objective in Eq. (4), considering that G is the model which needs to be optimized to enable generating accurate image translations:[13]

$$G = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G), \tag{4}$$

where λ is a hyperparameter representing the weighing ratio of the L1 loss to the adversarial loss. The first term in Eq. (4) is the adversarial loss, and it indicates the generator's task of fooling the discriminator and increasing its loss. Whereas the second term is the L1 loss, and it indicates the generator's task to provide generations that are as close as possible to the real or target translations of the input data.[13] In this work, we investigate the effect of the weight ratios of the aforementioned two losses.

Lines and spaces	Contact array	Butting line ends	Attacking line ends	Bar in U
CD $\in [16, 116]$ Space $\in [20, 166]$	Width $\in (16, 166)$ Height $\in [16, 166]$ Gap $\in [20, 170]$	CD $\in [16, 36]$ Space $\in [20, 60]$ Gap $\in [18, 50]$	CD $\in [16, 46]$ Space $\in [20, 60]$ Gap $\in [16, 66]$	CD $\in [15, 80]$ Gap-H$\in [8, 48]$ Gap-V$\in [6, 46]$ Height$\in [20, 138]$

Fig. 1 Mask pattern layouts utilized (top row) and samples thereof (middle row) and their ranges of geometry variation (bottom row). All units are nm on mask scale.

3 Proposed Approach

3.1 *Training Data*

The mask layout patterns used in the training data are horizontal and vertical lines, contact arrays, attacking, and butting line ends. Additionally, the pattern bar-in-U is utilized as a validation pattern to verify the model's ability to generalize. The implemented layouts are industrially relevant ones specified by Synopsys Inc. The pattern layouts and the ranges of variations in their geometries are shown in Fig. 1. The network's training data consist of 2D model aerial images provided as inputs, and rigorously computed 3D model aerial images provided as outputs. The 3D model aerial images are obtained by a rigorous solution of the diffracted light from the thick mask using the waveguide method, which solves the EMFs in the spatial frequency domain.[1] This method decomposes the thick mask into slices that are homogeneous in the z-direction. Similar to the analysis of optical waveguides and optical fibers, the waveguide method transfers Maxwell's equations to the Helmholtz wave equation shown in Eq. (5).

$$\nabla^2 \vec{A} + \tilde{k}_0^2 \tilde{\epsilon} \vec{A} = 0, \tag{5}$$

where $\vec{A}$ is the vector field, k_0 is the wave-number in vacuum and $\tilde{\epsilon}$ is the complex permittivity. The Helmholtz equation is then solved in the z-homogeneous slices. The solution is obtained by performing Fourier series expansions on the fields. The Fourier series expansions yield a system of linear equations that can be solved using the boundary conditions by connecting the components of consecutive slices.[1,2] The number of Fourier expansion terms is called the waveguide order. For our training data, we implement the optimal number of waveguide orders based on the pitch, as determined by Eq. (6).

$$\text{Waveguide orders} = \frac{p}{2\lambda}. \tag{6}$$

We perform the waveguide method's simulations without the Hopkins approach. The Hopkins approach is an approximation that is commonly utilized for image simulations, which assumes that the diffraction orders from the mask at small angles of incidence relative to the normal are equal to the diffraction orders obtained at the normal incidence. Accurate EUV simulations require a definition of the source without the Hopkins approximation, in which the mask diffraction spectrum is computed from a number of representative source points. In our training data, we use four non-Hopkins orders (1 per pole) referring to the representative source points. The 2D model aerial images are obtained via the Kirchhoff approach, which is based on the Kirchhoff diffraction and boundary condition. This approach ignores the 3D extent of the mask in the z-direction and only considers the geometry of the mask's features in the xy-plane. In other words, the mask is assumed to be infinitely thin and is simply defined as a 2D array of binary transmission values. The aerial image is then computed using a Fourier transformation of the mask layout.

The optical settings for the training data of both the 2D model and 3D model are shown in Table 1.

Table 1 Optical settings employed in simulations for the generation of training data.

Setting	Value
Sampling (x, y)	512×512 pixels
Image size (x, y)	$300 \text{ nm} \times 300 \text{ nm}$
Non-Hopkins orders (x, y)	2,2
Absorber material	Nickel
Absorber thickness	30 nm
Reduction (x, y)	$4\times, 4\times$
Illumination shape	Quasar
Chief ray angle	6 deg

3.2 *Network Architecture*

Pix2Pix-cGAN is the variant of cGANs proposed by Isola et al.,[13] which we implement to train a model to translate lithographic aerial image data from the 2D model to the 3D model. We propose a strategy for training this network to predict the 3D mask effects, which involves formulating sufficiently varied training data in a way that highlights the 3D mask effects, in addition to concatenating the defocus information to the inputs at the network's bottleneck layer, similar to the concatenation of the one-hot encoding vector done in TEMPO.[3] The generator and discriminator models for this cGAN architecture both involve enhancements over traditional cGAN generators and discriminators. Both the generator and discriminator employ modules of the form convolution-batch normalization-rectified linear unit (ReLU). As the name indicates, these modules are composed of a convolutional layer, followed by a batch normalization layer, followed by a ReLU activation function. This module formation has proven to be efficient for training deep neural networks.[19]

The generator model of the Pix2Pix-cGAN is based on a series of downsampling or decoding layers followed by an equal number of encoding or upsampling layers, where each encoding layer is connected to the reciprocal decoding layer that has the same dimensions. This encoder-decoder architecture with skip connections follows the structure of U-Nets.[20] Those skip connections serve the purpose of transferring low-level information that is shared between the input and output images.[13] Considering that for our application, the structures of the input and output images are largely aligned, therefore, a significant amount of low-level information is shared between mask patterns and aerial images.

The discriminator model consists of six down-sampling blocks involving convolutional layers. The special consideration with this discriminator is the focus on local image patches. This means this discriminator works to classify if each $N \times N$ patch in the image corresponds to a real or fake mapping, and generating a decision probability thereof. This localized focus of the discriminator serves to improve the learning of high-frequency features.[13] The corresponding patch size $(N \times N)$ is determined by number of convolutional layers in the discriminator and their filter sizes, in addition to the size of the input images. In this implementation, the input to the discriminator is a concatenation of two 512×512 images (2D and 3D model aerial image) and the patch size is 70×70 pixels.

The filter sizes and layer sequence for both the generator and discriminator models are shown in Figures 2 and 3, respectively. An encoder indicates a 2D-convolution layer followed by a batch normalization layer followed by a leaky ReLU activation. A decoder indicates a 2D-deconvolution layer followed by a batch normalization layer followed by a concatenation with the skip connection of the reciprocal encoder, followed by a ReLU activation. All convolutional and deconvolutional layers have filter sizes of 4×4 and stride sizes of 2×2. The last four decoders in the generator include an additional 50% dropout layer following the batch

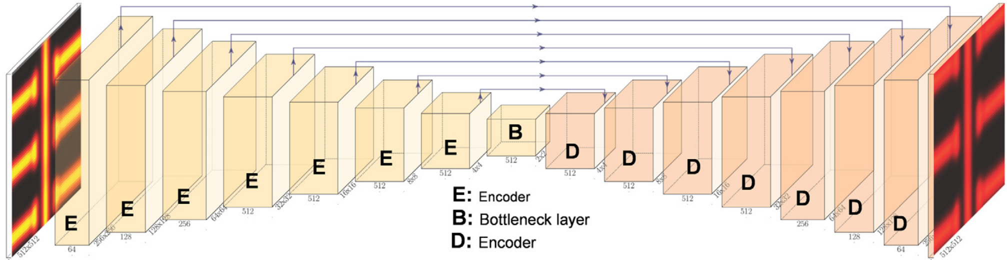

Fig. 2 Generator architecture. An encoder indicates a 2D-convolution layer followed by a batch normalization layer followed by a leaky ReLU activation. A decoder indicates a 2D-deconvolution layer followed by a batch normalization layer. The number of filters for each encoder/decoder is shown on the bottom of each block. The dimensions of the layers are shown on the bottom corner of the blocks. Plotted using PlotNeuralNet framework.[21]

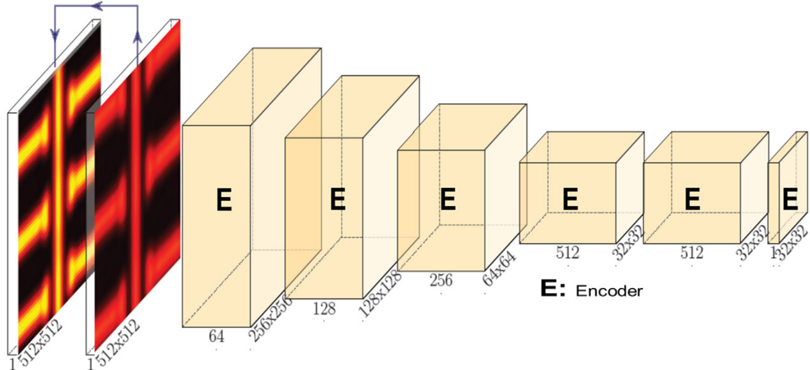

Fig. 3 Discriminator architecture. An encoder indicates a 2D-convolution layer followed by a batch normalization layer followed by a leaky ReLU activation. The number of filters for the convolutional layers for each encoder is shown on the bottom of each block. The dimensions of the layers are shown on the bottom corner of the blocks. Plotted using PlotNeuralNet framework.[21]

normalization layers. The inclusion of the dropouts allows the generator model to achieve such depth with minimized over-fitting. The activation function following the final layer is a sigmoid for the discriminator and a tanh for the generator, as proposed by Isola et al.[13]

3.3 *Cost Functions and Update Scheme*

The loss of the cGAN described in Eq. (4) is addressed by updates of both the discriminator and generator models. The discriminator's updates are based only on the accuracy of its estimate of whether the image translations are real or fake, and this is achieved using a binary cross-entropy loss. The updates of the generator loss are based on minimizing an L1 loss, in addition to an adversarial loss that aims to maximize the loss of the discriminator by minimizing its ability to distinguish real or fake translations.[17] The generator loss is defined as a weighted sum of the adversarial loss and the L1 loss. This weighing is determined by the parameter λ in Eq. (4). A λ of 1/100 means a weighing of 100:1 in favor of the L1 loss to encourage image generations that are closer to the target is recommended by the authors of the model.[13] In the following section, we discuss the impact of the λ on the accuracy of the generator models in predicting the 3D mask effects. The metrics we use to evaluate the network's training are the mean absolute error (MAE) of the generated images, in addition to the errors of the CDs in comparison to those of the 3D model images. Moreover, we further evaluate the model by assessing its predictions of 3D mask model process metrics such as telecentricity, best focus shifts, and iso-dense bias.

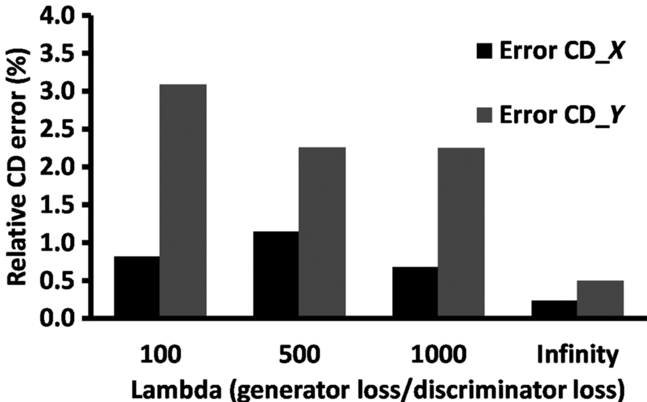

Fig. 4 Effect of the generator-to-discriminator loss weighing ratios on the prediction accuracy of the network. Training data are 3600 pairs of contact array 3D model and 2D model aerial images. Contact array layout variation is shown in Fig. 1. The model architecture is shown in Figs. 2 and 3, optical settings shown in Table 1.

3.4 *U-Net Model and Generator-to-Discriminator Weighing*

To identify optimum ratios of generator-to-discriminator loss weighing for lithographic aerial image translations, we compare the performance of different Pix2Pix cGAN models that have varying λ values. The λ value is the ratio of generator loss weight divided by the discriminator loss weight, and this dictates the contribution of each of their losses to the overall loss of the composite model. We benchmark the performance of the models based on the CD-fidelity of contact holes for a test dataset of 500 pairs of 2D model and 3D model aerial images. Figure 4 shows the average relative CD errors of generated images for models with λ values of 100, 500, 1000, and infinity. A lambda value of infinity means that only the generator's loss is taken into account, which effectively turns the cGAN into a pure generator U-Net. Relative CD errors indicate the deviation of the measured CD in the model predicted aerial image from the CD of the simulated 3D mask model aerial image. We utilize the mean of relative CD errors for predictions of a test dataset to represent the accuracy of a given model. Errors in Fig. 4 are computed from predictions of a test dataset of 500 contact array layouts.

Figure 4 shows that increasing the ratio in favor of the generator provides favorable results for this application range. Predictably, the pure generator U-Net provides a training runtime speedup of more than 2× compared to the full cGAN with a discriminator. For a test dataset of 2800 2D and 3D model image pairs, the U-Net generator trained for 50 epochs in under 50 min, while the full cGAN trained for 50 epochs in over 110 min. The hardware and setup utilized for the training are discussed in Sec. 4.1. The rest of the applications and investigations in this work are implemented using the pure generator U-Net architecture that is shown in Fig. 5.

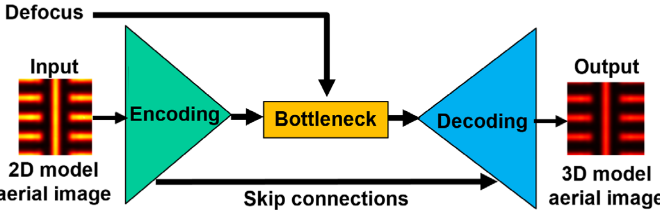

Fig. 5 Input-output framework of the generator U-Net model implemented. The value of the defocus position is adapted (expanded) to match the dimensions of the network's bottleneck layer and then concatenated to it. The decoding and encoding portions of the model are composed of seven blocks with 2D convolutional layers and seven blocks with 2D deconvolutional layers, respectively. Details of the generator's layers are shown in Fig. 2.

3.5 *Training Strategy*

The motivation for using the 2D model images as inputs instead of the mask layouts is their computational efficiency and their high structural similarity to the 3D model aerial images, which would allow for a higher prediction accuracy with minimal computational overhead. The defocus of the images is concatenated as an additional input at the bottleneck layer of the generator or U-Net, considering that this defocus information cannot be accurately represented with the input 2D images alone. The concept behind our training strategy is to incorporate sufficiently varied imaging scenarios that highlight the different 3D mask effects in the training data. This would allow the network to learn process variation trends from these scenarios which are represented in the output 3D model aerial images. This is achieved by including a variety of mask pattern layouts that involve different feature orientations to highlight the horizontal-vertical feature bias, different pitches to highlight the isolated-dense feature bias, in addition to different defocus ranges to learn the process variations and tendencies through focus. For each mask layout pattern, we simulate 2D and 3D model images at varying defocus positions. The network then trains on the 2D model images at the input, the defocus offset value concatenated at the bottleneck layer, and the 3D model images as the targets at the output layer of the generator.

To optimize the learning of the defocus-based effects, we first train the network on images generated at five fixed defocus positions with a relatively large separation, namely −100, −50, 0, +50, and +100 nm. Then we retrain this network on images that are generated at randomly selected non-fixed defocus values within the same range of defocus offsets (−100 to +100 nm). This would allow the network to learn the more pronounced structural patterns and trends from the large defocus steps in the first training iteration. The retraining serves as a fine-tuning stage to allow the network to learn the less pronounced trends from the intermediate and smaller defocus steps.

4 Results

4.1 *Training Details*

The training runs and tests are performed on an Nvidia Tesla V100 GPU and an Intel Xeon Gold 6134 CPU. For the first training, the generator U-Net model is trained on a set of 7500 pairs of aerial images obtained by the 2D model and the 3D model, generated at defocus positions of −100, −50, 0, 50, and 100 nm. The model is then retrained on another set of 7500 aerial image pairs generated at randomly selected defocus positions ranging from −100 to 100 nm for fine-tuning. It is worth noting that relevant accuracy levels can be achieved with training data that are less than half the amount of this dataset in cases where the range of data is less varied as shown in Fig. 4. The optimizer utilized is of the type Adam, with a batch size of 4 and a step size of 1. A set of an aerial image pair for each mask layout pattern is used as a test set to evaluate the performance of the training. The MAE between the 3D model aerial images and the network-generated images are utilized as the main metric to evaluate the training. Each training spans 200 epochs. The generator model is then extracted at the epoch where it generates images with the lowest MAE values. The optimum MAE point was reached in the first training at 124 epochs and in the retraining at 180 epochs.

4.1.1 *Image predictions and image fidelities*

The network is then capable of generating images that are demonstrably similar to the 3D model simulated aerial images, with median MAE values consistently below 0.005. Using an Intel Xeon CPU with 2×8 cores @3.20 GHz and an Nvidia Tesla V100 GPU with 32 GB of VRAM, the average runtime for an aerial image prediction by the model is 0.017 s, compared to the 3D model EMF simulations which can last from 9 s to more than 1 min, depending on the mask layout. Therefore, the speedup achieved by the U-Net model ranges from 500× to more than 3000× compared to the rigorous simulations. On the same hardware, the computation time of the 2D model aerial images ranges from 0.15 to 0.65 s. Therefore, the total end-to-end speedup achieved by the model is 100× on average, while it ranges from a minimum of 3× up to 400×. The accuracy of the network's prediction of the 3D model is verifiable via the

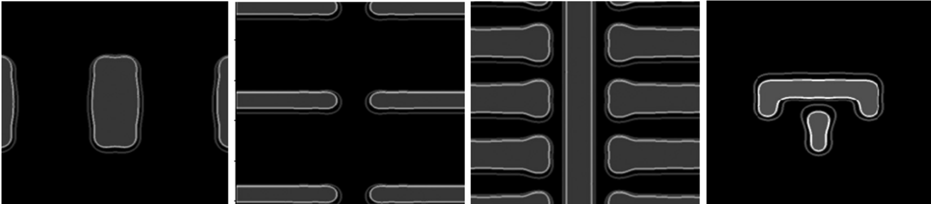

Fig. 6 Comparison of the 3D and 2D model and predicted aerial images using contors extracted at the same intensity threshold. The filled gray line represents the 3D model image, the hollow gray line represents the 2D model image and the hollow white line represents the predicted aerial image. The rightmost contor plot is from the validation layout pattern bar-in-U, which was not included in the model's training data.

predicted CDs and the EDEs. The EDE is a metric originally proposed to guide mask synthesis in inverse lithography applications.[22] The similarity between the model-predicted images and the target 3D model images is visualized via the aerial images contors in Fig. 6. Furthermore, the model's ability to generalize is demonstrated by its ability to predict images from a pattern that was not included in the training, as shown in the rightmost contor plot in Fig. 6. The ability to generalize indicates a minimal presence of overfitting and can be attributed to the utilization of 50% dropout layers in the last four decoders of the generator.

To extrapolate a more global representation of the image fidelity from the CDs, we define the CD error as the average relative CD error of multiple features for each image. In contact arrays, the error is an average of the x- and y-direction CD values of 5 features, the contact at the center, top, bottom, left, and right. In the attacking and butting line end patterns, the error is defined as the average of the CD error at the center gap, top, and bottom gaps. Considering that the bar-in-U pattern is not associated with a particular CD of interest, we do not include it in the CD error evaluations, rather in the global evaluations. Figure 7 demonstrates the mean values of relative CD errors obtained by the model for a test dataset of 1200 image pairs in the case of training with the proposed retraining framework.

The model's prediction accuracy is also evaluated using the EDE as a fully global metric based on the image contors, given the shortcomings of the CD when it comes to describing more complex mask layouts. The EDE is computed as per Eq. (7) and is measured by units of distance. While the EDE is not sufficient as a standalone metric since it may fail to represent feature shifts, it provides valuable information on the reciprocity of the imaged feature sizes in a global manner. Therefore, the EDE can be a viable metric to be used alongside others such as MSE, CD errors, telecentricity errors, etc. Average EDE values for a test dataset of 600 images are shown in Fig. 8.

$$\mathrm{EDE} = \frac{\text{Generated area} - \text{Target area}}{\text{Target perimeter}}. \tag{7}$$

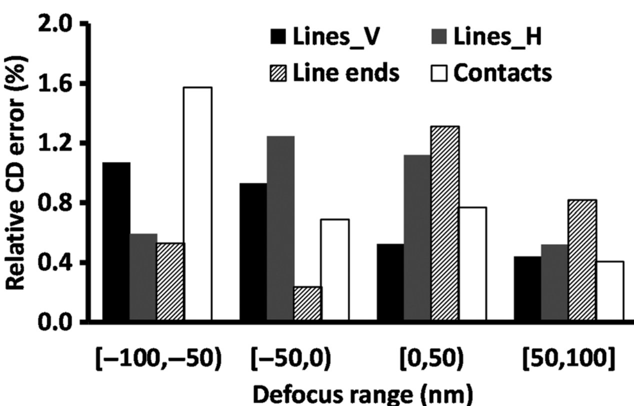

Fig. 7 Average CD errors for the models prediction of 3D model aerial images for different mask pattern layouts using a test dataset of 1200 3D and 2D model aerial images of the patterns shown in Fig. 1. Line ends include both attacking line ends and butting line ends.

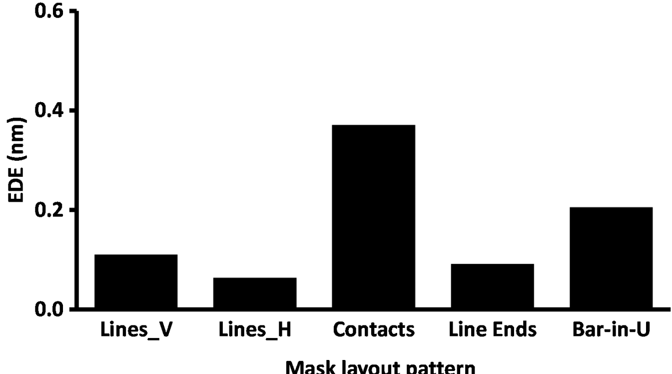

Fig. 8 Average EDE values for the models prediction of 3D model aerial images for different mask pattern layouts using a test dataset of 600 3D and 2D model aerial images of the patterns shown in Fig. 1. Line ends include both attacking line ends and butting line ends.

The generator U-Net is able to predict 3D model images accurately with relative CD errors averaging consistently sub-1% across various mask layout patterns, in addition to EDE values averaging below 0.4 nm for all the mask patterns that are implemented, including the pattern bar-in-U, which was not included in the training data. The relative CD errors from Fig. 7 are shown excluding two extreme outlier values from each selection, while the EDE values from Fig. 8 involve no outlier exclusion.

4.1.2 *Prediction of 3D model process variations*

The trained model is capable of accurately predicting the process variation tendencies as dictated per the 3D mask model simulations with sufficient accuracy for a variety of mask layouts and settings. As detailed in Sec. 2.1, EUV process windows of semi-dense and isolated features involve a significant degree of asymmetry and shift of the best focus position. The 2D mask model is incapable of capturing such asymmetric process windows. Figure 9 shows process windows simulated with the 2D and 3D mask model in addition to the process windows from the network's predicted images. The process windows are computed by calculating the threshold-to-size that provides a CD value on target, +10% and −10% from the target CD for 31 aerial images simulated with 2D and 3D model and predicted by the network for defocus values from −150 to 150 nm.

As shown in Fig. 9, the network's ability to more accurately predict the 3D mask model process windows is highlighted in the cases of isolated features. With a simple re-scaling, the 2D model may appear to accurately represent the 3D model process window for certain feature pitches. However, upon further investigation, the shapes of the 2D model process windows shift further away from those of the 3D model at larger pitches. The noisy behavior of the predicted curves is attributed to the fact that the curves are formulated from network predictions of numerous images across a defocus range, while the 3D model curves are smooth because they are formulated from images following the physical model. Although a complete match between the physical and predictive models across a range of focus variations is practically impossible, the network approaches the physical models' curves with a certain level of noise. However, the noise in the curves corresponds to relative errors fall within the achievable error ranges of the network shown in Fig. 7. Moreover, the noisy curves can be efficiently smoothed using a spline or curve-fitting functionality in practical scenarios.

The optical proximity effect (OPE) curves refer to the variations of the printed feature size across different pitches. While the 2D mask model is capable of representing some trends of the optical proximity effects, the OPE curves differ significantly between the 2D and 3D mask models. Our investigations of the predicted aerial images across a range of pitches have shown that the model is capable of representing the OPE curves of the 3D mask model with demonstrable accuracy, as shown in Fig. 10. Additionally, the network-predicted OPE curves capture the discrepancy between the behavior of the horizontal and vertical features as described by the 3D

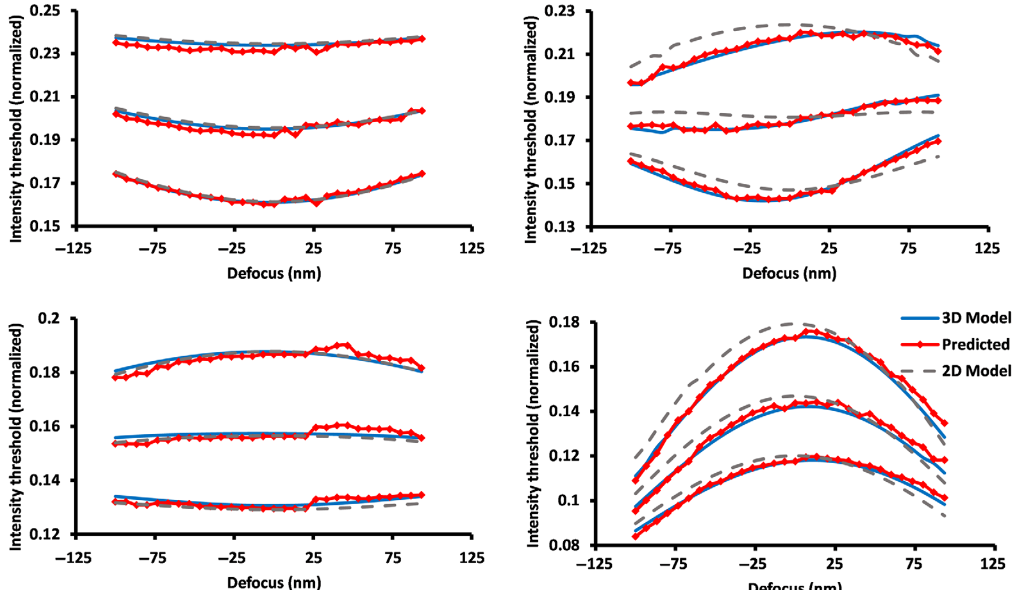

Fig. 9 Model-predicted process windows compared to 3D and 2D model process windows for horizontal lines and spaces patterns (top row) and contact array patterns in the *y*-direction (bottom row). The feature sizes are 32 nm for lines and $32 \times 32\ \mathrm{nm}^2$ for contacts. The left column shows process windows for dense feature settings (pitch = 50 nm), and the right column shows process windows for isolated features (lines pitch = 110 nm, contacts pitch = 150 nm). Other optical settings are shown in Table 1. Legend is in the bottom right plot. 2D model process windows are rescaled with factors from 55% to 70%, to take into account the transmission losses from the 3D mask.

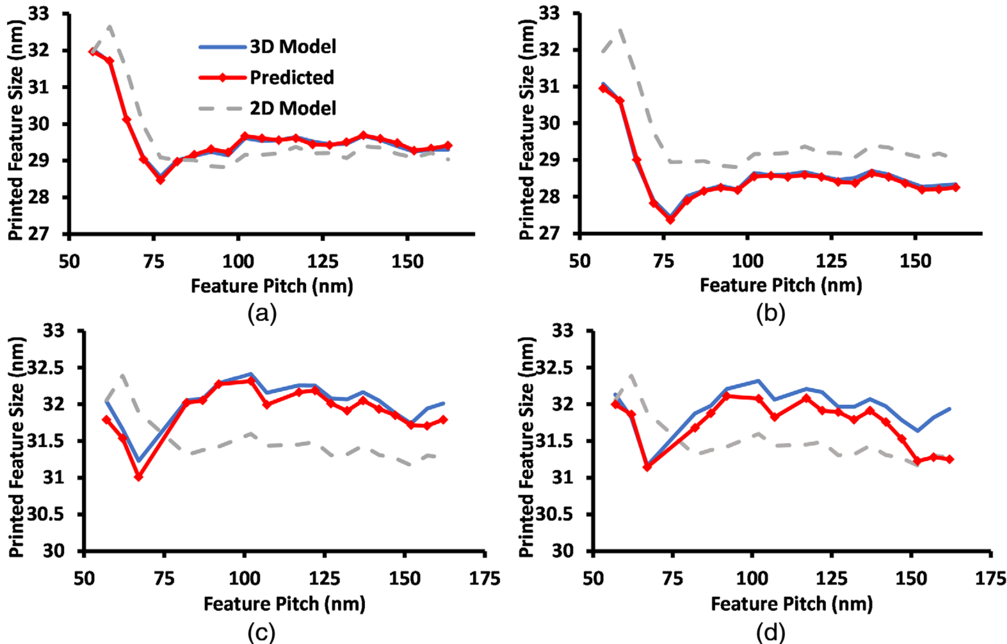

Fig. 10. Model-predicted OPE curves compared to 3D and 2D model process windows for lines and spaces and contact array masks. The top row shows OPE curves of (a) 32-nm horizontal lines and (b) 32-nm vertical lines. The bottom rows show the OPE curves for $32 \times 32\ \mathrm{nm}^2$ contact arrays in (c) *y*-direction and (d) *x*-direction. Optical settings are shown in Table 1. Legend is in the top-left plot. 2D model process windows are rescaled with factors from 55% to 70%, to take into account the reflectivity losses of the 3D model.

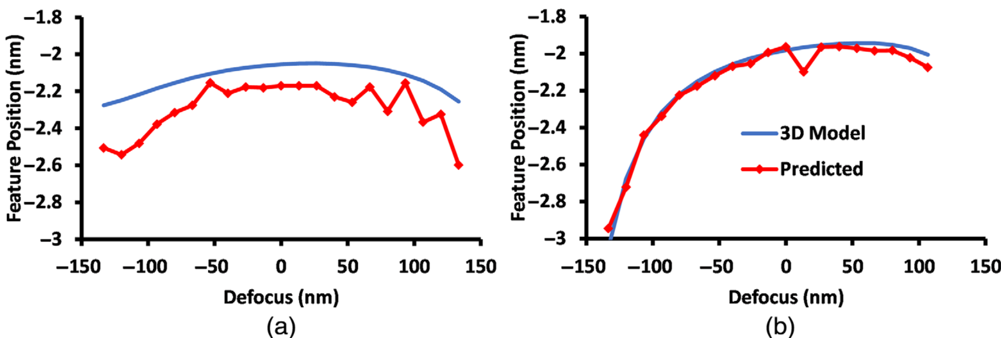

Fig. 11 Feature position behavior through focus using the 3D mask model and the network's predictions for 32-nm horizontal lines at (a) pitch of 70 nm and (b) pitch of 110 nm. Optical settings are shown in Table 1. Legend is on the right. 2D model curves are not demonstrated, as they are constant.

model's curves. On the other hand, OPE curves of the horizontal and vertical features as per the 2D model are identical. It can also be observed from the bottom row of Fig. 10 that the model may exhibit better accuracy predicting an orientation for certain mask patterns better than others. For example, CDs of vertical lines are predicted with smaller errors compared to vertical CDs of contact arrays. This prediction discrepancy is likely a result of the unequal representation of the different patterns in the dataset, as the layouts in the dataset were distributed in a random manner. The oscillatory behavior in the curves can be attributed to the sub-optimal sampling of the illumination source.

Telecentricity, or focus-dependent position shifts, is another mask topography effect that the 2D mask model fails to represent. The U-Net generator predicts the position shifts through focus accurately for mask layouts that involve a significant degree of telecentricity such as horizontal lines. Figure 11 shows that the network's predictions of the position shifts are more accurate for the more isolated case, in which the telecentricity error is larger in magnitude. Although the position shifts are predicted slightly less accurately in the dense case, the trend through focus is adequately captured and the relative error of the position shift remains within the previously demonstrated range of relative CD errors of $\sim$1%.

5 Conclusions

In this work, we proposed an approach that efficiently trains a generative network to generate accurate aerial images that take into account the thick EUV mask and its topography effects. We also compared the performance of cGANs of varying weight loss ratios to pure U-Net generators and observed that the cGAN's discriminator might be detrimental to the network's aerial image prediction ability. Our approach involves training a U-Net generator model on varied mask layouts that highlight the 3D mask effects. We train the network on 2D mask model images as inputs and 3D mask model images as outputs, in addition to the defocus information appended at the bottleneck of the U-Net. The network is then capable of predicting 3D mask model aerial images with demonstrable accuracy, not only in terms of MAE or CD errors but also in terms of contour-based global metrics such as the EDE. More importantly, the network's predictions accurately demonstrate the tendencies of the 3D mask model's process variations in terms of focus, pitch, dose, and position. Such accurate aerial image predictions using an efficient generative model present an ability to substitute rigorous field simulations. This would allow for a substantial speed-up of optimizations and process characterizations, considering that the model's predictions in addition to the input (2D model image) generation can be performed up to 400× faster than rigorous EMF simulations.

Acknowledgments

This project is funded by the German Federal Ministry of Education and Research (BMBF, project number 5M20WEC, siMLopt).

References

1. P. Evanschitzky and A. Erdmann, "Fast near field simulation of optical and EUV masks using the waveguide method," *Proc. SPIE* **6533**, 65330Y (2007).
2. F. Shao et al., "Fast rigorous simulation of mask diffraction using the waveguide method with parallelized decomposition technique," *Proc. SPIE* **6792**, 679206 (2008).
3. W. Ye et al., "TEMPO: fast mask topography effect modeling with deep learning," in *Int. Symp. Phys. Design '20*, pp. 127–134 (2020).
4. P. Liu et al., "Fast and accurate 3D mask model for full-chip OPC and verification," *Proc. SPIE* **6520**, 65200R (2007).
5. X. Ma et al., "Fast lithography aerial image calculation method based on machine learning," *Appl. Opt.* **56**, 6485–6495 (2017).
6. A. Erdmann et al., "Characterization and mitigation of 3D mask effects in extreme ultraviolet lithography," *Adv. Opt. Technol.* **6**(3–4), 187–201 (2017).
7. J. Lin et al., "Fast mask near-field calculation using fully convolution network," in *Int. Workshop Adv. Patterning Solut.*, pp. 1–4 (2020).
8. V. Agudelo et al., "Application of artificial neural networks to compact mask models in optical lithography simulation," *J. Micro/Nanolithogr. MEMS MOEMS* **13**, 011002 (2013).
9. J. Lin et al., "Fast extreme ultraviolet lithography mask near-field calculation method based on machine learning," *Appl. Opt.* **59**, 2829–2838 (2020).
10. H. Tanabe, S. Sato, and A. Takahashi, "Fast 3D lithography simulation by convolutional neural network," *Proc. SPIE* **11614**, 116140M (2021).
11. Y. Lin et al., "Data efficient lithography modeling with transfer learning and active data selection," *Trans. Comp.-Aided Des. Integr. Ciruits Syst.* **38**, 1900–1913 (2019).
12. H. Yang et al., "GAN-OPC: mask optimization with lithography-guided generative adversarial nets," in *Annu. Design Autom. Conf.*, Vol. 56, pp. 1–6 (2018).
13. P. Isola et al., "Image-to-image translation with conditional adversarial networks," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 5967–5976 (2017).
14. M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv:1411.1784 (2014).
15. W. Ye et al., "LithoGAN: end-to-end lithography modeling with generative adversarial networks," in *56th ACM/IEEE Design Autom. Conf.*, pp. 1–6 (2019).
16. A. Erdmann, *Optical and EUV Lithography: A Modeling Perspective*, SPIE Press (2021).
17. J. Brownlee, *Generative Adversarial Networks with Python. Deep Learning Generative Models for Image Synthesis and Image Translation*, Vol. 1.5, Machine Learning Mastery (2019).
18. I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM* **63**, 139–144 (2020).
19. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, Vol. 37, pp. 448–456 (2015).
20. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
21. H. Iqbal, "PlotNeuralNet," 2020, github.com/HarisIqbal88/PlotNeuralNet.
22. W. Lv, Q. Xia, and S. Liu, "Mask-filtering-based inverse lithography," *J. Micro/Nanolithogr. MEMS MOEMS* **12**, 043003 (2013).

Abdalaziz Awad is a PhD student at the University of Erlangen-Nuremberg working with the lithography simulation research group in the Fraunhofer Institute for Integrated Systems and Device Technology (IISB). He received his BSc degree in electrical engineering from Khalifa University in Abu Dhabi and an MSc degree in advanced optical technologies from the University of Erlangen-Nuremberg (FAU). His current research interests include EUV lithography, deep learning, generative networks, and interference lithography. He is a member of SPIE.

Biographies of the other authors are not available.