TEMPO: Fast Mask Topography Effect Modeling with Deep Learning

Wei Ye
ECE Department, UT Austin
weiye@utexas.edu

Mohamed Baker Alawieh
ECE Department, UT Austin
mohdbaker@utexas.edu

Yuki Watanabe
Kioxia Corporation
yuki9.watanabe@kioxia.com

Shigeki Nojima
Kioxia Corporation
shigeki.nojima@kioxia.com

Yibo Lin
CS Department, Peking University
yibolin@pku.edu.cn

David Z. Pan
ECE Department, UT Austin
dpan@ece.utexas.edu

ABSTRACT

With the continuous shrinking of the semiconductor device dimensions, mask topography effects stand out among the major factors influencing the lithography process. Including these effects in the lithography optimization procedure has become necessary for advanced technology nodes. However, conventional rigorous simulation for mask topography effects is extremely computationally expensive for high accuracy. In this work, we propose TEMPO as a novel generative learning-based framework for efficient and accurate 3D aerial image prediction. At its core, TEMPO comprises a generative adversarial network capable of predicting aerial image intensity at different resist heights. Compared to the default approach of building a unique model for each desired height, TEMPO takes as one of its inputs the desired height to produce the corresponding aerial image. In this way, the global model in TEMPO can capture the shared behavior among different heights, thus, resulting in smaller model size. Besides, across-height information sharing results in better model accuracy and generalization capability. Our experimental results demonstrate that TEMPO can obtain up to 1170× speedup compared with rigorous simulation while achieving satisfactory accuracy.

ACM Reference Format:
Wei Ye, Mohamed Baker Alawieh, Yuki Watanabe, Shigeki Nojima, Yibo Lin, and David Z. Pan. 2020. TEMPO: Fast Mask Topography Effect Modeling with Deep Learning . In *Proceedings of the 2020 International Symposium on Physical Design (ISPD '20), March 29-April 1, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3372780.3375565

1 INTRODUCTION

Lithography is a key step in the fabrication of nanoelectronic circuits. It is a patterning process through which a mask pattern is transferred into a thin photoresist (resist) layer on a substrate [1]. In practice, lithography simulations have been effectively used for process development, performance prediction and a number of other tasks including model-based optical proximity correction (OPC). These simulations are utilized to calculate correct resist shapes that can be used for physical verification such as hotspot detection. However, as the technology node continues scaling down, the trend to print features much smaller than the wavelength of light used has tremendously increased lithographic and manufacturing process complexity, as well as the lithography modeling complexity.

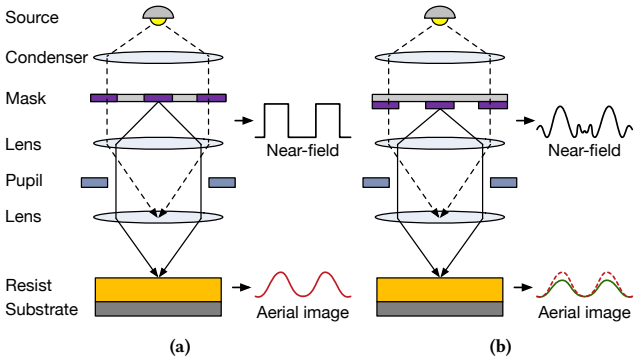


Figure 1: Imaging process of a lithography system. (a) Thin mask model and (b) thick mask model result in different near-fields and aerial images.

This continuous device scaling has posed the mask topography effects among the major challenges in lithography modeling. In the past, thin mask approximation, or so-called Kirchhoff approximation, was widely used in lithography simulation, as shown in Figure 1(a). With such an approximation, the three-dimensional structure of the mask is ignored despite its critical influence on the amplitudes, phases, and polarizations of the transmitted light, as demonstrated in Figure 1(b). When the feature sizes start to be comparable to the wavelength, the thin mask approximation is no longer adequate with the increasingly pronounced impacts of thick mask effects on the lithography imaging [2–4]. As a consequence, the failure to consider mask topography effects in lithography modeling could lead to critical dimension error and focus shift, resulting in the shrinkage of process window and the decrease of the image quality and the process robustness.

In the lithography process, many important properties, such as exposure and development latitude, can be derived from aerial images after optical simulation [5]. These images contain the intensity of the exposure radiation in the plane of the wafer; and hence, the topography effects of the mask can significantly impact their accuracy. Moreover, in lithography simulation, an accurate 3D view of aerial images at different resist heights is crucial to evaluate cross-section views of the resist pattern in order to find defects on the top

ISPD '20, March 29-April 1, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7091-2/20/03…$15.00
https://doi.org/10.1145/3372780.3375565

or bottom position. These defects, if gone undetected, can lead to catastrophic manufacturing failures. Therefore, accurate prediction of 3D aerial images with the mask topography effects considered is important in lithography development and verification.

Conventionally, rigorous simulators capable of capturing mask topography effects have been developed for aerial image calculation. Technically, the precise description of the mask diffraction spectrum in lithography is accomplished by using rigorous algorithms to solve Maxwell's equations for the electromagnetic field [6]. However, despite their superior accuracy, such rigorous methods are prohibitively expensive since performing rigorous calculations at the full-chip level, during OPC for example, is computationally intensive. Under the governing trade-off between accuracy and efficiency, different compact models were formulated as less accurate yet more efficient mask models [7, 8]. However, these compact models fail to maintain the accuracy level at advanced technology nodes since newly pronounced lithography effects invalidate several key assumptions in these models as shown in [9, 10].

Recently, advances in machine learning have been leveraged to devise new mask modeling techniques. In [11], an artificial neural network (ANN) was proposed to model the rigorous spectrum with respect to the feature vector containing the amplitude and the phase information of the scalar spectrum from different mask patterns. The output of the ANN is used to compute the aerial images using Abbe's method. In [10], for an arbitrary thick mask, its near-field is calculated using the nonparametric kernel regression model and the pre-calculated training libraries; then the aerial image is calculated using Abbe's method as well. The aforementioned machine learning approaches rely on conventional modeling techniques that require intensive feature engineering and depend heavily on post-processing methods which affect the model accuracy.

In the recent past, conditional generative adversarial networks (CGANs) have attracted attention due to their wide range of applications in image related tasks [12]. Among the state-of-the-art machine learning models, CGAN stands out due to its inherent capability to perform image translation tasks such as image colorization and background masking, where an image in one domain is mapped to a corresponding image in another domain. In practice, this model has been recently adopted to perform different lithography related tasks [13, 14]. Of particular significance is the application of CGAN in the end-to-end lithography simulation framework, LithoGAN [15]. While LithoGAN has demonstrated impressive efficiency, it only assumes a thin mask model which limits its capability of handling the mask topography effects. Besides, its output format is a monocolor image, while the desired output in the mask modeling task is the intensity map which has a higher accuracy requirement. Moreover, the aerial image estimation requires intensity map prediction at different resist heights. While the default approach is to train different CGAN models for prediction at different heights, such an approach is not efficient both in terms of training time and model size.

In this work, we propose TEMPO as a novel thick mask effect modeling framework using a single, one-fits-all model capable of predicting aerial image intensity at different resist heights. Besides the advantages in terms of the training cost and model size, incorporating the different modeling tasks into a single model can significantly improve the model accuracy. This is mainly due to the fact that various features and information are shared across all heights. Hence, having data from different heights available for training a single model results in a more robust model that has better generalization capabilities when compared to a set of models individually trained on a subset of the available data. To enable such a one-fits-all model, we propose a CGAN architecture that uses the desired prediction height as an additional input appended to the low-level latent representation in the model architecture. With such representation, the height information is efficiently incorporated at the CGAN bottleneck layer where it can have the most powerful impact on output generation.

The major contributions of this paper are highlighted as follows.

- We propose TEMPO as a novel framework for 3D aerial image generation considering mask topography effects.
- A one-fits-all CGAN model, with a novel target domain encoding, is proposed for aerial image prediction at multiple resist heights. The model is compact and can achieve superior accuracy by leveraging across-domain information sharing.
- Two schemes are presented in TEMPO to provide flexible trade-offs between accuracy and efficiency.
- Experimental results demonstrate the two schemes in TEMPO obtain 1170× and 27× speedup when compared with rigorous simulation while achieving satisfactory performance in aerial image quality and critical dimension fidelity.

The rest of this paper is organized as follows. Section 2 reviews the basic concepts and gives the problem formulation. Section 3 provides a detailed explanation of the proposed TEMPO framework. Section 4 demonstrates the effectiveness of our approaches with comprehensive results, followed by the conclusion in Section 5.

2 PRELIMINARIES

2.1 Mask Topography Effects

As shown in Figure 1, in an optical lithography system, the light source illuminates the mask and generates the near-field underneath the mask. Then, the light rays propagate through the projection lens and produce the aerial image on the wafer [10].

In the past, a mask in lithography was mostly considered as an infinitely thin object with homogeneously transparent and opaque areas as demonstrated in Figure 1(a). The conventional application of Kirchhoff's boundary conditions on the mask surface provides the so-called thin mask approximation of the near-field.

However, mask topography effects have been observed since the minimum feature size on the mask dropped below the exposure wavelength [4]. The light scattered by mask edges and corners changes the near-field of the light on the mask level. As shown in Figure 1(b), the scattering affects both the amplitude and the phase of the incident field, and thus not only changes the aerial image intensity on the wafer level, but also changes the resist profile after resist development. The failure to consider mask topography effects could lead to critical dimension error and focus shift, resulting in the shrinkage of the process window, and the decrease of the image quality and the process robustness. Therefore, mask topography models (thick mask models) have been indispensable since 28 nm tech node and below.

To precisely model the thick mask effects, rigorous simulators have been developed based on fundamental electromagnetism principles. However, they are rather slow and infeasible to apply on

full chips within acceptable runtime. Generally, the intensity distribution in the aerial image calculated by a rigorous thick mask simulation is lower than the calculation result by a thin mask simulation because of a waveguide effect due to the topographical structure of the mask [16]. Nevertheless, there is no simple transformation between the outputs of these two kinds of mask models since the magnitude of the mask topography effects varies at different locations on the wafer and is also affected by the design of mask patterns.

There are efforts attempting to construct fast compact models for approximating mask topography effects [7, 8]. However, as shown in [9, 10], newly pronounced lithography effects and conditions keep invalidating some simple assumptions in conventional compact models and render them inapplicable at advanced nodes. The impacts of key factors on the accuracy and efficiency of the compact models need further study and verification, and ad hoc compact model building is incapable of providing models that are adequate for advanced lithography.

2.2 3D Aerial Image

In order to simplify the analysis of a lithography process, the optical effects of the lithography tool are usually separated from the resist effects of the resist process. As one of the direct outputs of optical analysis, the aerial image is defined as the spatial intensity distribution at the wafer, and is simply the square of the magnitude of the electric field [17]. The aerial image is the source of information that is transferred into the resist, and therefore dictates the quality of the final resist profile. Moreover, from the aerial image, we can easily predict the performance of a given lithographic process in terms of depth of focus, exposure latitude, etc [5].

The spatial image intensity distribution inside the resist bulk is calculated up to the defined resist thickness, and henceforth will be referred to as 3D aerial image or 3D intensity map, as shown in Figure 2. 3D aerial image is valuable in evaluating cross-section views of the resist profile in order to find defects on the top or bottom position. Typically, an aerial image simulation extracts the 2D intensity at one specific resist height; thus, the calculation of the entire 3D image is distributed among different threads in rigorous simulation tools [18].

Note that for a pure aerial image setup where the substrate, stack and resist are all set as air, the extraction height of the 2D aerial image does not matter. However, the resist and the stacks are practically composed of one or several non-air like optical materials, which results in standing waves due to interference effects of the incoming and backscattered light in the resist [18]. For the systems where standing waves can be very pronounced, the evaluation of the image intensity at a certain extraction height h must be performed carefully. For example, consider the extraction height h = 10 nm and h = 70 nm in Figure 2. It is obvious that the extraction height h = 70 nm will yield a higher image contrast than h = 10 nm. Therefore, it is necessary to model 3D aerial images.

2.3 Problem Formulation

For image generation tasks, multiple evaluation metrics are typically used to judge upon model accuracy. Let I denote the golden aerial image and $\hat{I}$ denote the predicted aerial image, where $I, \hat{I} \in \mathbf{R}^{n \times n}$. One of the commonly used accuracy metrics is the root-mean-square error (RMSE), which is given by

$$\text{RMSE} = \frac{1}{n} \left\| \hat{I} - I \right\|_F, \tag{1}$$

where $\|A\|_F = (\sum_{i,j} A_{i,j}^2)^{1/2}$ represents the Frobenius norm.

Since the overall light intensity of different aerial image samples in the dataset could vary significantly, we also adopt the normalized root-mean-square error (NRMSE) to quantify model performance. The NRMSE between the predicted image and the golden image is defined as the RMSE normalized by the averaged Frobenius norm of the golden image:

$$\text{NRMSE} = \frac{\text{RMSE}}{\|I\|_F / n} = \frac{\left\| \hat{I} - I \right\|_F}{\|I\|_F}. \tag{2}$$

We define the problem studied in this work as follows.

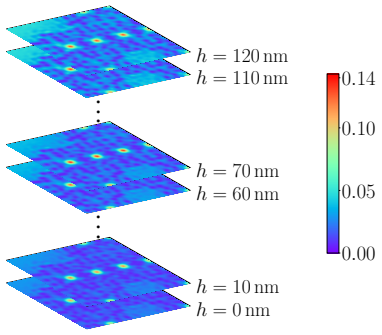


Figure 2: Example of the 2D aerial image slices inside a 3D aerial image.

Problem 1 (3D Aerial Image Learning)**.** Given a training dataset containing mask pattern samples and the corresponding 2D aerial images at m resist heights for each mask pattern sample, the objective is to train a model that can accurately predict the aerial images of a test mask pattern, where the accuracy is measured in terms of the RMSE and the NRMSE.

3 TEMPO FRAMEWORK

In a rigorous thick mask simulation flow, the simulator takes as input a mask pattern and generates the corresponding aerial image as shown in Figure 3(a). While such an approach is the common practice today, its inordinate runtime hinders its application in the early stages of the process development and mask optimizations. For example, simulating 1000 clips with mask topography effects could take up to 4 days. With this in mind, we propose TEMPO as a fast modeling framework that can significantly speed up the thick mask modeling task and hence, allow the consideration of the mask topography effects in the early stages of the process development. In practice, TEMPO provides in one of its schemes a CGAN model capable of mimicking the rigorous simulation process as shown in Figure 3(b). Under the same input/output set as in the rigorous simulation scheme shown in Figure 3(a), the CGAN model in TEMPO can translate the image from mask pattern to aerial images with orders of magnitude speedup. Hereafter, this direct translation using our proposed CGAN architecture is referred to as Scheme 1, and its details will be covered in Section 3.2.

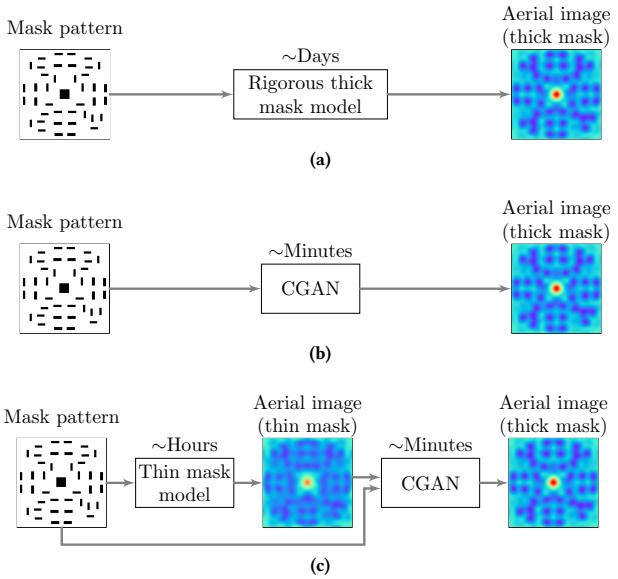


Figure 3: (a) Traditional rigorous thick mask simulation flow, (b) proposed Scheme 1 for high efficiency and (c) Scheme 2 for high accuracy in TEMPO.

It is evident that, compared to the rigorous simulation scheme, Scheme 1 in TEMPO is capable of achieving immense speedup at some compromise in accuracy. This accuracy compromise is due to the fact that the optical modeling inside the lithography system is a complicated process; aerial image is the outcome of the interactions among light source, mask pattern and the projection lens. Hence, given the limited information available in the input image containing only the mask patterns, the accuracy of Scheme 1 is not expected to be ideal but can still be acceptable for early exploration stages given its attractive efficiency.

For applications with high accuracy requirements, TEMPO provides an alternative framework, namely Scheme 2 shown in Figure 3(c), which represents a compromise between the accurate yet time-consuming rigorous simulation, and the efficient Scheme 1 with imperfect accuracy. Compared to Scheme 1, Scheme 2 sacrifices some additional runtime for better accuracy while still maintaining impressive speedup compared to the rigorous simulation. As a first step, TEMPO in Scheme 2 runs a fast thin mask model to generate aerial images assuming no mask topography effect, and the output aerial image is used along with the mask pattern as the input to the CGAN model. In this way, the aerial image given by the thin mask model provides the CGAN model with additional information not present in the mask pattern image, and hence improves its accuracy. In the next subsections, we first introduce the conventional CGAN model for image translation, then we present TEMPO for aerial image generation.

3.1 Generative Adversarial Networks

Generative adversarial networks (GANs) have demonstrated remarkable success in various computer vision tasks such as image generation [12], image translation [19, 20], and super-resolution imaging [21]. Originally, GANs were developed for the purpose of learning the distribution of a given dataset with the intent of generating new samples from it [22]. A typical GAN model consists of two modules: a generator and a discriminator. The generator is trained to produce samples that cannot be distinguished from real images by the adversarially trained discriminator which is trained to do as well as possible at detecting the generator fakes [22].

The conventional generator in a GAN is basically an encoder-decoder network where the input is passed through a series of layers that progressively downsample it (i.e., encoding), until a bottleneck layer, at which point the process is reversed (i.e,, decoding) [12, 22, 23]. On the other hand, the discriminator is a convolutional neural network whose objective is to classify fake and real images. Hence, its structure differs from that of the generator and resembles a typical two-class classification network [12, 22, 23]. This adversarial scheme is represented in the objective function given as:

$$\min_G \max_D \mathbb{E}_x[\log D(x)] + \mathbb{E}_z[\log(1 - D(G(z)))], \tag{3}$$

where $D(\cdot)$ represents the probability of a sample being real; i.e., not generated by G, $\mathbb{E}_x$ denotes the expectation over the input data x, and z is a random noise vector used as a seed for image generation.

A GAN model is typically trained with mini-batch stochastic gradient descent (SGD) [22]. The training alternates between one gradient descent step on the discriminator, and then one step on the generator. After training, the generator part of the GAN is used to generate new samples using random noise vectors while the discriminator is discarded as it is only needed for the training process [22].

Stemming from the core GAN model, different variants of generative neural networks were developed to address challenges in various fields of study, especially computer vision. Technically, many tasks in computer vision and graphics can be thought of as translation problems where an input image is to be translated from domain A to another domain B. Isola et al. [19] introduced an image-to-image translation framework that uses GANs in a conditional setting where the generator transforms images conditioned on the input image. Instead of randomly generating images from the learned distribution, it transfers an input image into another domain, hence, acting as an image translator. To train such a model, a paired training dataset is needed where each sample is a pair of an input image (i.e., image in the input domain) and its corresponding output image (i.e., translated image in the target domain).

Mathematically, the loss function used for training the CGAN can be given as [12, 19]:

$$\begin{aligned} L_{\text{CGAN}} &= \mathbb{E}_{x,y}[\log D(x,y)] \\ &\quad + \mathbb{E}_{x,z}[\log(1 - D(x, G(x,z)))] \\ &\quad + \lambda \cdot \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1], \end{aligned} \tag{4}$$

where x is a sample in the input domain, y is its corresponding sample in the output domain, and λ is the weight parameter. Comparing equations (3) and (4), one can notice the addition of the loss term which penalizes the difference between the generated sample $G(x,z)$ and its corresponding golden reference y.

3.2 TEMPO Architecture Design

Image translation using CGAN was proposed as a means for domain transfer between two distinct domains. However, different applications require more comprehensive translation schemes with one-to-many domain transfers. Aerial image generation requires

domain transfer from the single mask pattern domain to multiple resist height domains. Another popular application of such a scheme is facial image translation, where an input facial image is translated into different target domains representing different facial expressions or appearances [24]. For aerial image generation and other similar tasks, the most straightforward option is to train multiple domain-to-domain models. So, for m target domains, m such models are needed.

Clearly, the approach of building m individual models has multiple drawbacks. Most evident is the size of the model that scales with the number of target domains. This also requires a large dataset from all domains to train different independent models. Besides, when assuming that different target domains are independent, an opportunity for information sharing between those slightly different tasks is missed. In terms of the data, since we model the light intensity in a 3D continuous space in the aerial prediction task, the intensity values change continuously. The aerial images extracted from discrete resist heights should be highly correlated. In terms of the model, the input encoding performed by the generator's encoder is very similar across different domains in many applications. This is true in the aerial image generation as well as the facial translation scenario. Mainly, the important features for the translation tasks are common across different target domains, and the target specification is rather important in the decoder that generates the images. Hence, if an adequate information-sharing scheme is developed, the performance can be enhanced by exploiting the high correlation between images in different domains. Therefore, model scalability and information sharing render the setup of multiple individual models ineffective.

To overcome these two drawbacks, new variants of CGAN have been proposed, such as ComboGAN [25] and StarGAN [24]. In ComboGAN, information sharing is addressed through a joint training scheme for the m different 2-domain transfer models [25]. On the other hand, StarGAN tries to address the scalability issue by building a single generator and incorporating the target domain into its input. However, the target domain representation in StarGAN still carries high redundancy since it requires m additional channels in the input image to one-hot encode the chosen k-th target domain out of m domains. In other words, the size of the input image scales linearly with the number of target domains. Better scalability necessities a more compact input domain encoding scheme.

Towards the goal of a compact model with an information-sharing scheme, two important features of the one-to-many domain transfer task in this work should be noted. First, the target information is not necessary for the input encoding task. It is fair to assume that the features that are needed from the input image to generate the aerial image at different heights are the same. It is the way these features are later decoded that is impactful on the image generation. Hence, the target information is not needed as an input to the encoder in the generator network. The second feature is that the bottleneck layer in the generator carries the most critical information as it represents the latent representation of the input upon which the output image is generated; thus, the information in this layer is of significant impact on the result. Therefore, we propose within TEMPO a new *one-fits-all* model where a one-hot encoding vector of length m carrying the target domain information is appended to the latent space representation in the bottleneck layer, as shown in Figure 4. This way, the information is appended at a

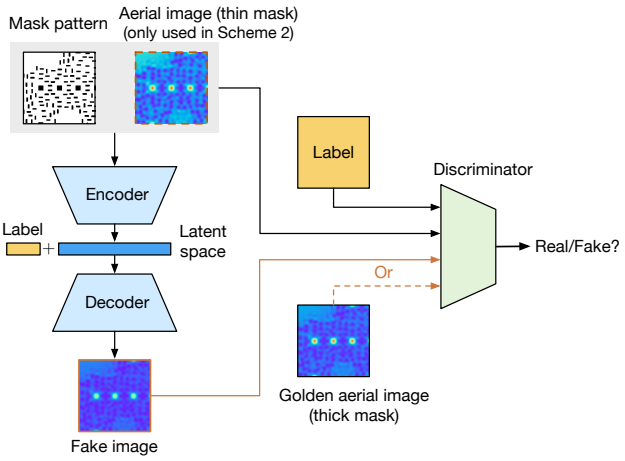


Figure 4: Overview of the TEMPO model.

critical location in the network where it can guide the output image generation while having a compact representation. Compared to that used in StarGAN where each one extra input channel is needed for each domain, the encoding scheme in TEMPO requires only a single channel for all the domains. This can significantly improve the scalability of TEMPO when faced with a significant increase in the number of target domains.

In the next subsections, the details of both the generator and discriminator used in TEMPO are shown. These implementations are adapted from the deep convolutional generative adversarial networks framework proposed in [23].

3.2.1 Generator. We adopt the encoder-decoder network which is commonly used to design a generator [12, 19, 22, 23]. The input is passed through a series of layers in the encoder that progressively downsamples it, until a bottleneck layer, at which point the process is reversed in the decoder. The details of the encoder and decoder are summarized in Table 1. Specifically, eight convolutional and deconvolutional layers are used for the encoder and decoder, respectively. In Table 1, the column "Size" and the column "Stride" give the size and stride of each filter, and the number of layers sharing the same filter setting is shown in the column "Count". "Additional" indicates the additional layers for normalization and activation function. Here, batch normalization (BN) [26] is selectively applied on certain convolutional layers both in the encoder and decoder. The encoder uses leaky ReLU (LReLU) as the activation function, whereas the decoder uses ReLU. The input of the generator is the images of 200×200 pixels, and can have single channel (mask pattern) in Scheme 1 or two channels (mask pattern and thin-mask aerial image) in Scheme 2. "Concat" denotes the concatenation of the one-hot label vector of size m and the latent space vector of size 512. For image translation tasks using CGAN, a significant amount of information is shared between the input and the output, and we followed the design of U-Net [27] with skip connections between encoder layers and decoder layers.

3.2.2 Discriminator. On the other hand, the discriminator is a convolutional neural network that performs classification to distinguish between the real image pairs and fake image pairs. Meanwhile, the target domain information is fed into the discriminator that is trained to discriminate image pairs from different target domains.

Here, the target information is encoded by appending to the input image a single channel whose pixel values reflect the target domain. In practice, since the different domains in this application correspond to different resist heights, there exists a true ordering for the target domains themselves. Therefore, an ordinal encoding scheme is used to encode the ID of the target domain k ($k \in \{0, 1, \ldots, m-1\}$) on the additional input channel whose pixel values are the same and are set as follows:

$$\frac{p_{\max} - p_{\min}}{m-1} \cdot k + p_{\min}, \tag{5}$$

where $p_{\min}$ and $p_{\max}$ denote the minimum and maximum possible values in the additional input channel. Commonly used settings include $p_{\min} = 0, p_{\max} = 255$ or $p_{\min} = -1, p_{\max} = 1$.

Table 1 summarizes the details of the discriminator which constitutes of four convolutional layers and one fully connected layer (FC) whose output is the binary classification results.

Table 1: Network architecture of the proposed TEMPO.

Network	Layer	Count	Channel	Size	Stride	Additional
Generator Encoder	Input	—	1 (2) [a]	—	—	—
	Conv	1	64	5	2	LReLU,BN
	Conv	1	128	5	2	LReLU,BN
	Conv	1	256	5	2	LReLU,BN
	Conv	5	512	5	2	LReLU,BN
Generator Decoder	Concat	1	$512 + m$	—	—	—
	Deconv	4	512	5	2	ReLU,BN
	Deconv	1	256	5	2	ReLU,BN
	Deconv	1	128	5	2	ReLU,BN
	Deconv	1	64	5	2	ReLU,BN
	Deconv	—	1	5	2	ReLU
Discriminator	Input	—	3 (4)	—	—	—
	Conv	1	64	5	2	LReLU
	Conv	1	128	5	2	LReLU
	Conv	1	256	5	2	LReLU
	Conv	1	512	5	2	LReLU
	FC	1	1	—	—	Sigmoid

[a] $(\cdot)$ denotes the number of channels in Scheme 2.

4 EXPERIMENTAL RESULTS

In this work, we explore mask topography effects on contacts as according to the existing studies and reports, the mask topography effect should be considered more carefully for contact hole patterns than line and space patterns [28]. We generate 966 clips of size 2×2 µm containing various contact patterns following the clip generation method described in [29]. Each contact is designed to be 60×60 nm, and the contact pitch is 128 nm. We perform sub-resolution assist feature (SRAF) insertion and OPC on contact patterns using Mentor Graphics Calibre [30].

We run rigorous optical simulation to generate 3D aerial images using Synopsys Sentaurus Lithography [18]. A quasar light source is used for this experiment. The wavelength of the light source is set to 193 nm, and the numerical aperture (NA) of the imaging system is 1.2. The simulation window of 1.5×1.5 µm is configured as nonperiodic and centers each of the clips. Since the resist thickness is 120 nm and simulation resolutions in X, Y and Z directions are set to 7.5 nm, 7.5 nm and 10 nm respectively, we got 2D aerial images of 200×200 pixels at 13 different resist heights for each clip, i.e., $n = 100$ in Equation (1) and Equation (2), and $m = 13$.

In this work, the aerial images generated by rigorous simulation considering mask topography effects are used as the golden data for TEMPO training. Each sample in the training set is a collection of the mask pattern image and the corresponding aerial images at 13 different resist heights. Note that the mask pattern clip within the simulation window is 1.5×1.5 µm and the grid unit in the original layout is 1 nm, so we size it down to a grayscale image of 200×200 pixels using average filtering. Each pixel in the aerial image is an intensity value stored in the 32-bit single-precision format.

The proposed TEMPO is implemented in Python with the Tensorflow library and validated on a Linux server with 3.3GHz Intel i9 CPU and Nvidia TITAN Xp GPU. In our experiments, we randomly sample 75% of the data for training the model and the remaining 25% clips are for testing. We set the batch size to 4 and the number of maximum training epochs to 70. The weight parameter λ in Equation (4) is set to 1000. We also build 13 individual models to predict 2D aerial images at each resist height separately which work as the baseline approach. Each of the individual models takes as input the dataset of aerial images at only one resist height and is trained with the same hyperparameter setting as TEMPO. Note that the 13 individual models have a total of 1.17×10^9 trainable parameters (weights and biases), whereas TEMPO has 1.03×10^8. Therefore, TEMPO effectively reduces the model size for the 3D aerial image prediction task.

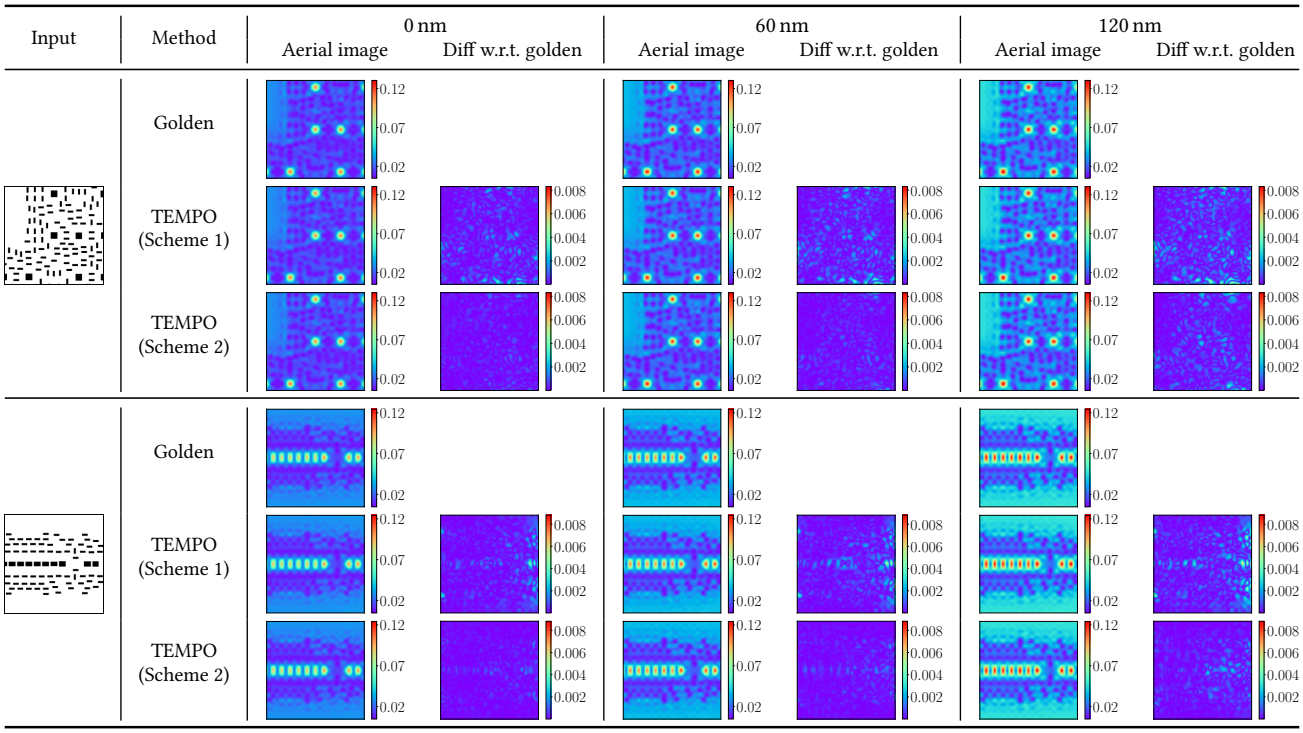
We first demonstrate the accuracy of our proposed TEMPO. Table 2 gives a detailed comparison between the individual models and our TEMPO under Scheme 1 and Scheme 2 using the proposed RMSE and NRMSE metrics in Section 2.3. The number shown in the table is the average of all the test samples on each resist height. One can easily see that TEMPO outperforms the individual modeling approach (denoted as Baseline) under both schemes. Besides, whether using the 13 individual GAN models or the proposed TEMPO approach, Scheme 2 always gives better accuracy than Scheme 1. Moreover, TEMPO improves the RMSE from 14.96×10^{-4} to 13.72×10^{-4} on average, and NRMSE from 4.63% to 4.23% in Scheme 1, while improving the RMSE from 7.3×10^{-4} to 5.81×10^{-4} and NRMSE from 2.27% to 1.79% in Scheme 2. Clearly, Scheme 2 in TEMPO can help gain better improvement in accuracy because the aerial image produced by the fast thin mask simulation, as an additional input in Scheme 2, provides more information about the lithography system, and hence TEMPO is able to achieve notable improvement under such situation. To visually examine the accuracy difference between the two schemes in TEMPO, the aerial images for two samples of distinct pattern designs are shown in Table 3.

As one of the most important outputs of optical models, the aerial image can be used together with resist models to simulate final resist profiles. Therefore, in addition to the direct comparison of aerial images, we also evaluate the effectiveness of our proposed methods based on the quality of generated resist patterns. We calculated the critical dimension (CD) value of the resist pattern for the center contact in each sample using the average of the aerial images at 13 resist heights. Using the CD values derived from the golden aerial images as reference, Table 4 shows the comparison of CD errors in the X and Y directions among different mask topography effect modeling methods. The row "thin mask sim." represents the CD errors when using the aerial images without considering mask topography effects, and the errors could go up to more than 20 nm. Our proposed TEMPO in Scheme 2 gives very small CD errors, for

Table 2: Comparison of evaluation metrics among different modeling methods.

Height (nm)	RMSE ($\times 10^{-4}$)				NRMSE (%)			
	Scheme 1		Scheme 2		Scheme 1		Scheme 2	
	Baseline	TEMPO	Baseline	TEMPO	Baseline	TEMPO	Baseline	TEMPO
0	11.88	10.87	4.96	4.12	4.55	4.15	1.96	1.62
10	12.53	11.48	5.41	4.21	4.55	4.15	2.03	1.57
20	13.50	12.63	5.24	4.50	4.51	4.19	1.79	1.54
30	15.30	13.26	6.11	4.74	4.97	4.25	2.02	1.56
40	14.26	13.32	5.53	4.79	4.63	4.31	1.82	1.58
50	14.36	13.11	5.96	4.93	4.71	4.29	1.98	1.63
60	14.37	13.22	7.99	5.23	4.63	4.24	2.63	1.70
70	15.18	13.61	7.32	5.71	4.62	4.13	2.27	1.76
80	15.58	14.52	7.71	6.24	4.48	4.17	2.26	1.81
90	16.42	15.25	8.00	6.79	4.57	4.23	2.26	1.90
100	16.79	15.59	8.40	7.42	4.62	4.28	2.34	2.05
110	17.16	15.75	8.96	8.17	4.68	4.29	2.46	2.23
120	17.11	15.74	13.27	8.66	4.63	4.26	3.67	2.34
Average	14.96	13.72	7.30	5.81	4.63	4.23	2.27	1.79
Max	17.16	15.75	13.27	8.66	4.97	4.31	3.67	2.34
Std. dev.	1.69	1.58	2.24	1.52	0.12	0.06	0.49	0.27

Table 3: Aerial image results for two test clips using Scheme 1 and Scheme 2 in TEMPO.



example, 0.38 nm in the X direction and 0.45 nm in the Y direction, which qualifies it for practical lithography usage. Besides, TEMPO gives smaller CD errors when compared with the baseline with 13 individual GAN models, which further demonstrates the advantages of our one-fits-all approach.

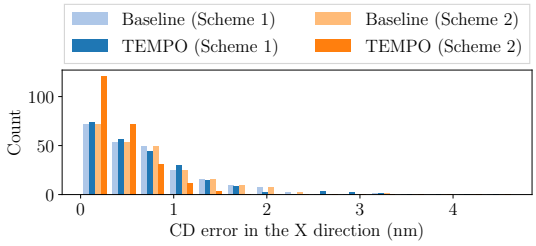
Last, we demonstrate the runtime comparison in Table 5, where the total runtime of generating the 3D aerial images for all the test samples, i.e., 242 samples, are shown. We can clearly see that the two schemes in TEMPO satisfy different needs for speed and accuracy at lithography development phases. Scheme 2 in TEMPO achieves $\sim$ 26.5× runtime reduction when compared to rigorous thick mask simulation while achieving satisfactory accuracy. Considering the acceptable CD degradation in Scheme 1 compared to Scheme 2 while being 50× faster, Scheme 1 in TEMPO is suitable for the early exploration stages where speed is favored over high accuracy.

5 CONCLUSION

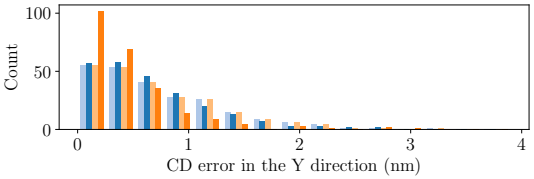
In this work, we have presented TEMPO, a novel and scalable framework which is capable of generating 3D aerial images efficiently

Table 4: Comparison of CD errors in the X and Y directions among different methods.

Method		CD error X (nm)		CD error Y (nm)	
		Average	Max	Average	Max
Thin mask sim.		2.77	20.67	3.93	33.49
Scheme 1	Baseline	0.75	4.64	0.73	3.19
	TEMPO	0.72	3.38	0.67	2.82
Scheme 2	Baseline	0.48	2.05	0.50	3.89
	TEMPO	0.38	1.88	0.45	3.11



(a)



(b)

Figure 5: Distribution of CD errors using different methods: (a) error in the X direction and (b) error in the Y direction.

Table 5: Runtime comparison between rigorous simulation and the proposed TEMPO framework.

	Rigorous mask sim.	TEMPO (Scheme 1)	TEMPO (Scheme 2)		
			Thin mask sim.	GAN	Total
Runtime	20.5 h	1.1 m	45.3 m	1.1 m	46.4 m
Ratio	26.51	0.02	—	—	1.00

and accurately for modeling mask topography effects. Essentially, TEMPO comprises a one-fits-all CGAN model for multi-domain image-to-image translation, with the accuracy and compactness further boosted by across-domain information sharing. Besides, the two flexible schemes of operations in TEMPO provide different trade-offs between accuracy and efficiency, which promotes the wider application of TEMPO in different stages of process development. The experimental results demonstrate that TEMPO can achieve superior performance in both speed and accuracy for advanced lithography usage.

ACKNOWLEDGMENT

This work is supported in part by NSF under Award No. 1718570 and Kioxia Corporation. The authors are thankful to Synopsys, Inc. for providing the Sentaurus Lithography software. The authors would like to thank Osamu Yamane of Kioxia Corporation and Dr. Zac Levinson, Dr. Peter Brooker, and Dr. Kevin Lucas of Synopsys, Inc. for the helpful discussions.

REFERENCES

[1] C. A. Mack, *Fundamental Principles of Optical Lithography: The Science of Microfabrication.* John Wiley & Sons, 2008.
[2] A. K. Wong and A. R. Neureuther, "Mask topography effects in projection printing of phase-shifting masks," *IEEE TED*, vol. 41, no. 6, pp. 895–902, 1994.
[3] R. L. Gordon and C. A. Mack, "Mask topography simulation for EUV lithography," in *Proc. SPIE*, vol. 3676, 1999, pp. 283–297.
[4] J. Ruoff, "Impact of mask topography and multilayer stack on high NA imaging of EUV masks," in *Photomask Technology 2010*, vol. 7823, 2010, p. 78231N.
[5] C. A. Mack, "Understanding focus effects in submicrometer optical lithography: a review," *Optical Engineering*, vol. 32, no. 10, pp. 2350–2363, 1993.
[6] M. Moharam, D. A. Pommet, E. B. Grann, and T. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," *Journal of the Optical Society of America A*, vol. 12, no. 5, pp. 1077–1086, 1995.
[7] K. Adam and A. R. Neureuther, "Domain decomposition methods for the rapid electromagnetic simulation of photomask scattering," *JM3*, vol. 1, no. 3, pp. 253–270, 2002.
[8] J. Tirapu-Azpiroz, P. Burchard, and E. Yablonovitch, "Boundary layer model to account for thick mask effects in photolithography," in *Proc. SPIE*, vol. 5040, 2003.
[9] P. Liu, Y. Cao, L. Chen, G. Chen, M. Feng, J. Jiang, H.-y. Liu, S. Suh, S.-W. Lee, and S. Lee, "Fast and accurate 3D mask model for full-chip OPC and verification," in *Proc. SPIE*, vol. 6520, 2007.
[10] X. Ma, X. Zhao, Z. Wang, Y. Li, S. Zhao, and L. Zhang, "Fast lithography aerial image calculation method based on machine learning," *Applied Optics*, vol. 56, no. 23, pp. 6485–6495, 2017.
[11] V. Agudelo, T. Fühner, A. Erdmann, and P. Evanschitzky, "Application of artificial neural networks to compact mask models in optical lithography simulation," *JM3*, vol. 13, no. 1, p. 011002, 2013.
[12] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
[13] Y. Lin, M. B. Alawieh, W. Ye, and D. Z. Pan, "Machine learning for yield learning and optimization," in *Proc. ITC*, 2018, pp. 1–10.
[14] M. B. Alawieh, Y. Lin, W. Ye, and D. Z. Pan, "Generative learning in VLSI design for manufacturability: Current status and future directions," *Journal of Microelectronic Manufacturing*, vol. 2, 2019.
[15] W. Ye, M. B. Alawieh, Y. Lin, and D. Z. Pan, "LithoGAN: End-to-end lithography modeling with generative adversarial networks," in *Proc. DAC*, 2019, p. 107.
[16] S.-b. Shim, Y.-c. Kim, S.-j. Lee, S.-w. Choi, and W.-s. Han, "Study of the mask topography effect on the OPC modeling of hole patterns," in *Proc. SPIE*, vol. 6924, 2008.
[17] B. W. Smith and K. Suzuki, *Microlithography: science and technology.* CRC press, 2018.
[18] Synopsys, "Sentaurus Lithography," https://www.synopsys.com/silicon/mask-synthesis/sentaurus-lithography.html, 2016.
[19] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017, pp. 5967–5976.
[20] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.
[21] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. CVPR*, vol. 2, no. 3, 2017, p. 4.
[22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 2672–2680.
[23] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
[24] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. ICCV*, 2018, pp. 8789–8797.
[25] A. Anoosheh, E. Agustsson, R. Timofte, and L. Van Gool, "ComboGAN: Unrestrained scalability for image domain translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 783–790.
[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015.
[27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.
[28] A. Mimotogi, M. Itoh, S. Mimotogi, K. Sato, T. Sato, and S. Tanaka, "Mask topography effects of hole patterns on hyper-na lithography," in *Proc. SPIE*, vol. 6607, 2007.
[29] Y. Lin, M. Li, Y. Watanabe, T. Kimura, T. Matsunawa, S. Nojima, and D. Z. Pan, "Data efficient lithography modeling with transfer learning and active data selection," *IEEE TCAD*, 2018.
[30] Mentor Graphics, "Calibre verification user's manual," 2008.