Accurate lithography simulation model based on convolutional neural networks

Watanabe, Yuki, Kimura, Taiki, Matsunawa, Tetsuaki, Nojima, Shigeki

Yuki Watanabe, Taiki Kimura, Tetsuaki Matsunawa, Shigeki Nojima, "Accurate lithography simulation model based on convolutional neural networks," Proc. SPIE 10454, Photomask Japan 2017: XXIV Symposium on Photomask and Next-Generation Lithography Mask Technology, 104540I (13 July 2017); doi: 10.1117/12.2279780

SPIE.

Event: Photomask Japan 2017, 2017, Yokohama, Japan

Accurate Lithography Simulation Model based on Convolutional Neural Networks

Yuki Watanabe[a], Taiki Kimura[a], Tetsuaki Matsunawa[a], and Shigeki Nojima[a]

[a]Toshiba Memory Corporation, Memory Div. Memory Design Dept.II, 2-5-1, Kasama, Sakae-Ku, Yokohama, 247-8585, Japan

ABSTRACT

Lithography simulation is an essential technique for today's semiconductor manufacturing process. In order to calculate an entire chip in realistic time, compact resist model is commonly used. The model is established for faster calculation. To have accurate compact resist model, it is necessary to fix a complicated non-linear model function. However, it is difficult to decide an appropriate function manually because there are many options. This paper proposes a new compact resist model using CNN (Convolutional Neural Networks) which is one of deep learning techniques. CNN model makes it possible to determine an appropriate model function and achieve accurate simulation. Experimental results show CNN model can reduce CD prediction errors by 70% compared with the conventional model.

Keywords: Lithography simulation, Compact resist model, Deep learning, Convolutional neural networks

1. INTRODUCTION

As semiconductor devices shrinking, patterns which designers draw cannot print on wafer as their original shapes. One solution to this issue is OPC (Optical Proximity Correction), which corrects mask patterns in advance to make the difference smaller.[1] Mask correction sizes are calculated using lithography simulation. After the OPC, the corrected mask patterns are verified using lithography simulation to detect hotspots, which cause patterning failures on wafer. Thus, lithography simulation is an indispensable technique for device manufacturing and it requires accuracy which means small difference between the simulated value and experimental value on wafer.

Resist model which simulates development behavior is one of keys to decide the accuracy of the simulation. There are two types of resist models; one is rigorous model and the other is called compact model. The former method simulates physical phenomena precisely, but it requires much computing time. This method is used to simulate a small region. The latter does not follow physical phenomena precisely. The model is established for faster calculation. In order to calculate an entire chip in realistic time, compact resist model is commonly used for OPC and verification.

The compact resist model function is calibrated to wafer CD (Critical Dimension) to minimize the difference between simulated CD and wafer CD.[3] To achieve an accurate simulation, it is necessary to fix a complicated non-linear model function. However, it is difficult to decide an appropriate function manually because there are many options.[4] It leads to insufficient accuracy of simulation. In this paper, in order to solve this issue, we propose a new method using CNN (Convolutional Neural Networks) which is one of deep learning techniques. The method can determine an appropriate model function automatically and achieve accurate simulation.

2. COMPACT RESIST MODEL

As shown in Figure 1, lithography simulation consists of two stages. One is optical simulation and the other is resist simulation. In optical simulation, an aerial image is calculated from a mask pattern using an optical model which is characterized by an illumination type and projection lenses of an exposure tool. In resist simulation, a resist pattern is calculated from the aerial image using compact resist model. In the compact resist model, resist development effect is simulated. The development effect is physical phenomenon such as photo chemical

Yuki Watanabe: E-mail: yuki9.watanabe@toshiba.co.jp

Photomask Japan 2017: XXIV Symposium on Photomask and Next-Generation Lithography
Mask Technology, edited by Kiwamu Takehisa, Proc. of SPIE Vol. 10454, 104540I
 · doi: 10.1117/12.2279780

reaction, dissolving and removing the resist material. The extent of the development effect is determined by the intensity of the aerial image. In the compact resist model, a threshold is set to an aerial image intensity and used to determine whether resist material is developed or not. This means the threshold determines CD after development in the compact resist model. Thus, the threshold has a dominant influence on the lithography simulation accuracy in the compact resist model. Although several compact resist models have been proposed, two of them are mainly used, namely CTR (Constant Threshold Resist) model and VTR (Variable Threshold Resist) model, which handle a threshold in different ways. In this section, we briefly introduce two models and discuss their issues.

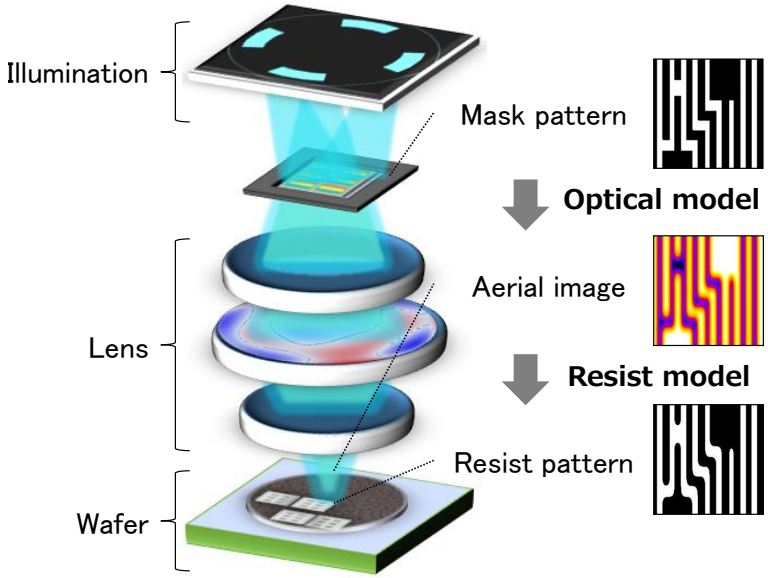


Figure 1. Lithography simulation.

2.1 CTR model

CTR model uses a fixed threshold and the process flowchart is shown in Figure 2. At first, an approximate formula which simulates photo chemical reaction of resist is defined and aerial images are modulated using the formula. As a simple example, given an aerial image I, the degree of resist chemical reaction D is calculated according to the following equation:

$$D = I * G \tag{1}$$

where * is convolution operation and G is Gaussian function to simulate chemical reaction.[2] After that, resist pattern R is calculated by the threshold processing:

$$R = \begin{cases} 1 & (D > threshold) \\ 0 & (otherwise) \end{cases} \tag{2}$$

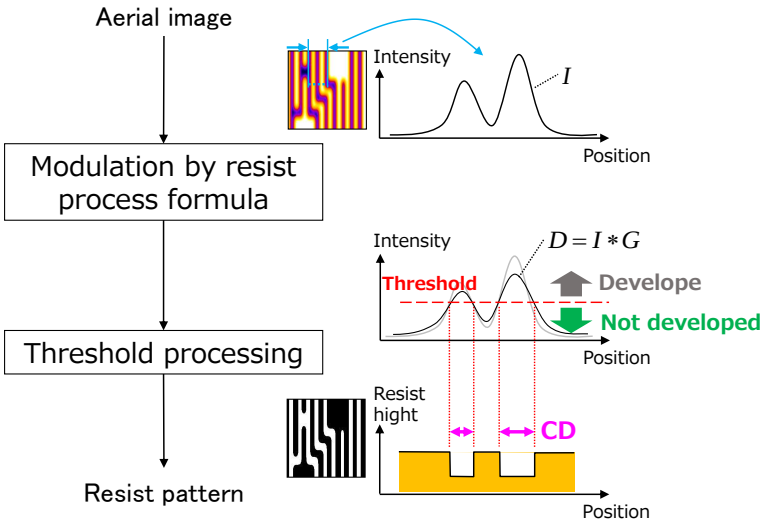


Figure 2. Process flowchart of CTR model.

CTR model executes the same operation for all aerial images, and there is an advantage in that the model formula is relatively simple. However, it is difficult to describe resist development effects with only simple formula operations. The reason is that resist process is complicated due to pattern shrinking and improvement of resist materials. Therefore, flexibility for models which can describe such complex processes is required to improve simulation.

2.2 VTR model

In VTR model, a threshold varies according to aerial images as shown in Figure 3.[3] In a typical case, a threshold to a pattern i is calculated using the following polynominal equation based on aerial image features such as maximum intensity $Imax_i$, minimum intensity $Imin_i$ and $Slope_i$ of the aerial image profile.

$$threshold_i = k_1 \cdot Imax_i + k_2 \cdot Imin_i + k_3 \cdot Slope_i \tag{3}$$

where $k1$, $k2$, $k3$ are coefficients. For example, the threshold can be set higher to the aerial image in Figure 4 (a) and lower to the aerial image (b), because $Imax_i$ and $Slope_i$ are different. Therefore, VTR model is expected to execute more accurate simulation than CTR model. However, it still has a week point that it is impossible to distinguish cases which have identical features. The aerial image in Figure 4 (c) has undershoots on the bottom of the peak, and that in Figure 4 (d) has multiple peaks. If aerial images are different, they sholud have different ideal thresholds which give accurate resist pattern CDs. However, thresholds cannot be changed because aerial image features in Figure 4 (b), (c) and (d) are identical.

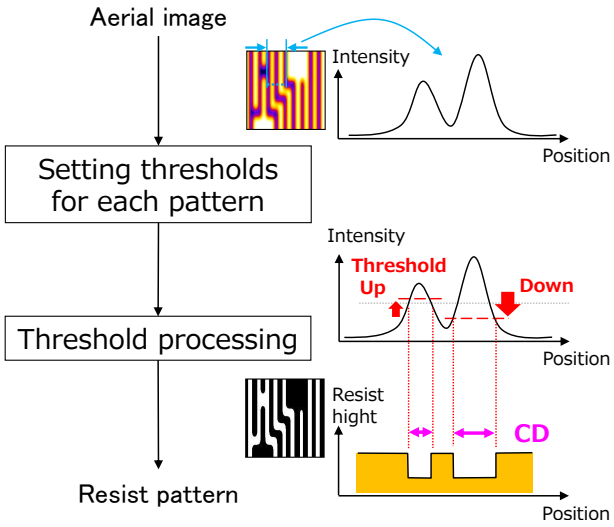
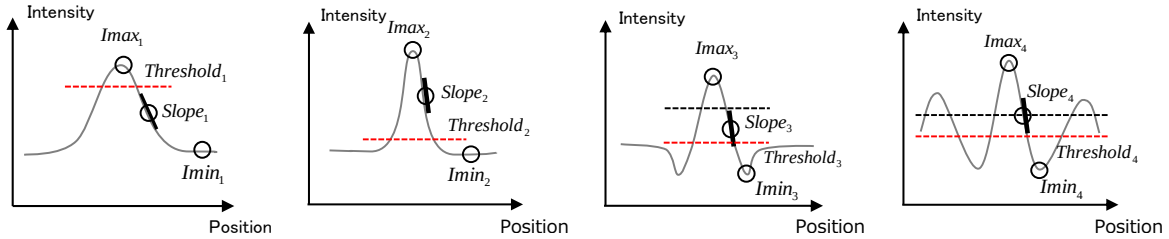


Figure 3. Process flowchart of VTR model.



(a) With gentle slope. (b) With steep slope. (c) With undershoots. (d) With multiple peaks.

Figure 4. Features of VTR model and thresholds on aerial image profiles. The red dotted lines represent calculated thresholds and the black dotted lines are ideal thresholds.

Thus the threshold predictive performance of VTR is not enough for some patterns which have identical features. In addition, using only limited features leads to an issue for extrapolation which means insufficient accuracy for unknown patterns which are not included in the modeling data.[4] Therefore, we need to decide thresholds based on appropriate features which can describe small differences in order to improve accuracy. Recently, a modeling method using non-linear function is proposed.[5] However, the issue for extrapolation remains open yet as same as conventional VTR model, because only limited information of aerial image features are used.

3. RESIST MODEL BASED ON CONVOLUTIONAL NEURAL NETWORKS

As an improved VTR model, we propose a new method which trains small differences and predicts appropriate thresholds using a machine learning method. Proposed method consists of two phases, a training phase and a testing phase as shown in Figure 5. In the training phase, the relationship between aerial images and thresholds are trained and a prediction model is created. In the testing phase, thresholds of aerial images are predicted using the model. Many machine learning methods are proposed such as, SVM (Support Vector Machine) and regular NN (Neural Network).[5] However, the both methods still have an issue that they need to decide appropriate input features in advance and it is difficult to decide the features manually. Because of this, we propose a new method using CNN which is one of deep learning. The method can extract aerial image features automatically and predict appropriate thresholds. CNN is an algorithm to calculate an output value in a hierarchical architecture and extracts features from input aerial images by convolution operations considering the relationship among pixels of input aerial images. Proposed CNN architecture is shown in Figure 6 and the architecture consists

of two layers as indicated by the dotted lines. One is the feature extraction layer and the other is the fully connected layer. The procedure of CNN architecture is optimized by training.

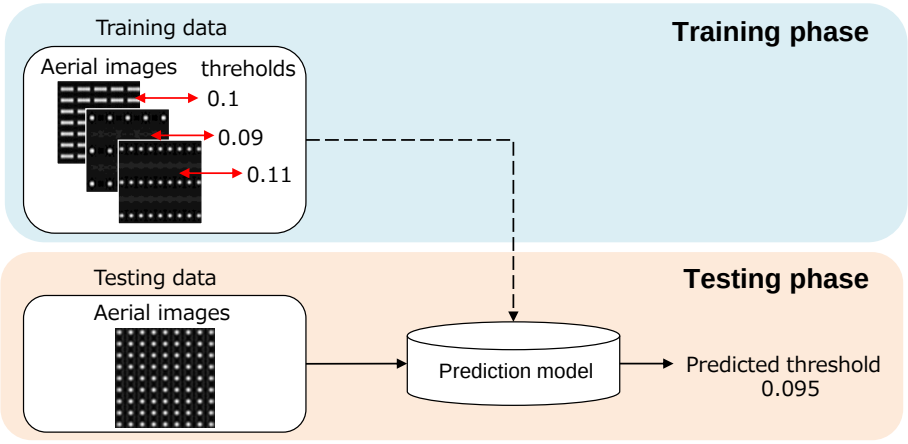


Figure 5. Training and testing phases.

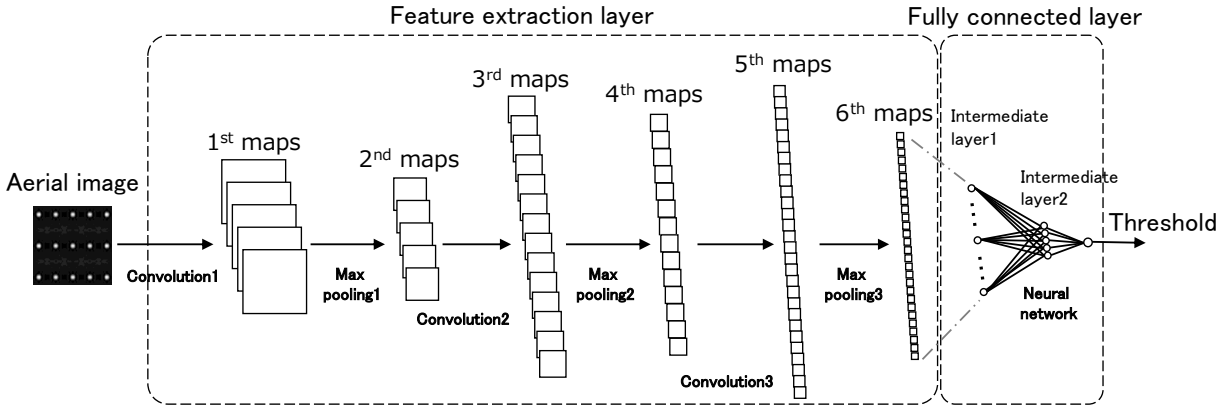


Figure 6. CNN architecture.

3.1 Feature extraction layer

The feature extraction layer consists of two operations, convolution and max pooling operation. Convolution operations emphasize image features and max pooling operations reduce the data size. In the proposed architecture, convolution 1-3 and max pooling 1-3 are executed as shown in Figure 6. Intermediate images are created by convolution and max pooling, which are called maps. In convolution processing, a number of maps are generated from input data, which are an aerial image or maps, by operations using many kinds of convolution filters as shown in Figure 7. In the each convolution operation, u_{ij} which is a pixel value of an output map at (i, j) is given by,

$$u_{ij} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{i+p,j+p} h_{pq} + b \tag{4}$$

where x_{ij} is a pixel value of the input aerial image or maps, h_{pq} is convolution filter weight of (p, q) which size is $H \times H$ and b is an offset value. On the convolution 2 and 3 in Figure 6, multiple filter operations are executed for a lot of maps. In addition, the output map value y is given from u using the equation (5), which is called activation function:

$$y = max(0, u) \tag{5}$$

The activation function is emulating behaviors of neurons and synapses. Neurons and synapses transfer information when they receive a certain signal or more. Sigmoid function is well known as an activation function, but the calculation cost is higher and the training speed is slower than those in the equation (5). So we choose the equation (5) as the activate function.

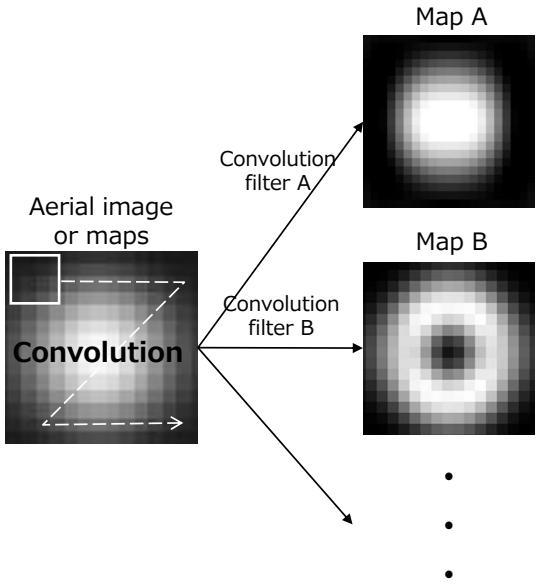


Figure 7. Convolution operation to emphasize image features.

On max pooling operation, the maximum pixel value is extracted from each pooling region and the extracted value is set for the output down scaled map as shown in Figure 8. The pixel value of the output map u_{ij} is given by,

$$u_{ij} = \max_{(p,q)\in P_{i,j}} x_{pq} \tag{6}$$

where x_{pq} is the pixel value of the input map at (p, q) and P_{ij} is the neighboring $I \times I$ region around (i, j)

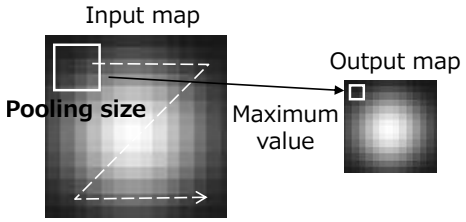


Figure 8. Max pooling operation to reduce the data size.

3.2 Fully connected layer

In the fully connected layer, a threshold is computed by NN from each pixel value of 6^{th} maps. NN is a mathematical model to emulate the human brain computing system consisting of a neuron and synapses. As Figure 9 shows, it consists of a number of units calculating output $y1$, $y2$, $\cdots$ from multiple inputs $x1$, $x2$,$\cdots$. Output of i th unit y_i calculated using equation (5) and u_{ij} is given by,

$$u_i = \mathbf{w_i x} + c_i \tag{7}$$

where $\mathbf{x} = (x1, x2, \cdots)^T$ is a input matrix, $\mathbf{w_i} = (w1, w2, \cdots)$ is a weight matrix and c_i is an offset value. Units which share the input matrix are defined as one intermediate layer. Complicated multi-layer architecture

can be built by setting the outputs of an intermediate layer as next layer's inputs. As shown in Figure 6, two intermediate layers are used in this proposed method.

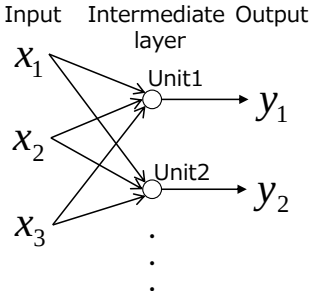


Figure 9. NN consists of multiple units.

3.3 Training method

A matrix whose elements are all convolution coefficients and weight of NN, such as h_{pq}, b $\mathbf{w_i}$ and c_i, are defined as $\mathbf{W}$. $\mathbf{W}$ is optimized by training. Training data of $n = 1, 2, \cdots N$ are written as (aerial image, threshold) $= (x_n, thr_n)$. A threshold value of an input x_n predicted by CNN architecture consisting of $\mathbf{W}$ is defined as $y(x_n; \mathbf{W})$. A prediction error function $E(\mathbf{W})$ is given by,

$$E(\mathbf{W}) = \frac{1}{2}\sum_{n=1}^{N} \|thr_n - y(\mathbf{x_n}; \mathbf{W})\|^2 \tag{8}$$

$\mathbf{W}$ is optimized so as to minimize function $E(\mathbf{W})$ by the iterative calculation according to the following equation,

$$\mathbf{W(t+1)} = \mathbf{W(t)} - \epsilon \frac{\partial E}{\partial \mathbf{W}} \tag{9}$$

where ϵ is a training coefficient to control an update rate and we used the method of Adam (Adaptive moment estimation) which adjusts the update rate efficiently corresponding to the gradient.[6] All layers gradient of $\mathbf{W}$ can be calculated by stochastic gradient descent method.[7,9] Calculation of the gradient with all training data simultaneously might lead to undesired local minimum value. To avoid this issue, we use a method called mini-batch training[9] and dropout method.[8]

4. EXPERIMENTAL RESULTS

The proposed method is evaluated with via patterns of total 1758 points. Figure 10 shows examples of pattern variations. 879 points are selected randomly and used as training data. The other 879 points are used as testing data to evaluate accuracy of predicting thresholds.

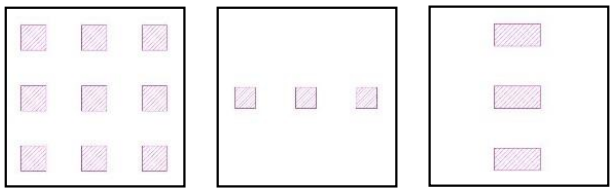


Figure 10. Examples of via patterns.

As shown in Figure 11, ideal thresholds are computed empirically from wafer CDs and aerial image profiles. The ideal thresholds are used as training thresholds and also used for evaluation of predicting accuracy on training

data and testing data. We calculated CDs from predicted thresholds and evaluated CD prediction errors using RMS (Root Mean Square) given by,

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} d_n^2} \tag{10}$$

where $N = 879$ is for both the number of training data and testing data, and d_n is a difference between the wafer CD and the predicted CD.

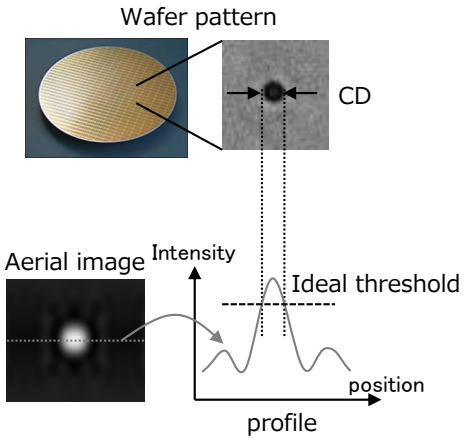
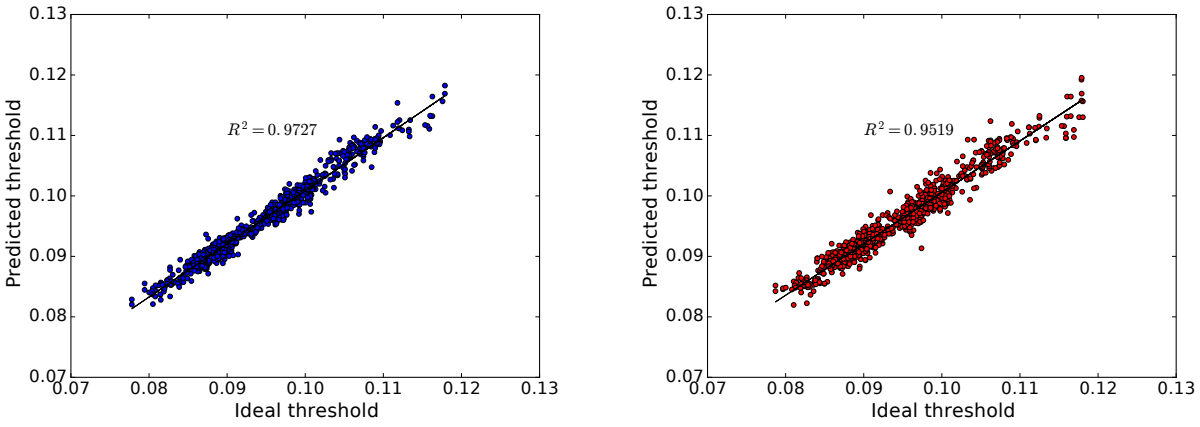


Figure 11. Ideal thresholds calculated empirically.

Relationship between ideal thresholds and predicted thresholds on training data and testing data are plotted respectively in Figure 12 (a) and (b). Both of squared correlation coefficients R^2 are more than 0.95 and it shows our method achieves highly accurate prediction performance. RMS of CD prediction errors on training data is 1.50nm and that on testing data is 1.64nm. This means we can rely on the extrapolation accuracy.



(a) Training data.

(b) Testing data.

Figure 12. Results of predicting thresholds.

We further compare our method with a conventional CTR and VTR model. We confirmed that RMS of predicted CD on both training and testing data are more than 5.01nm as shown in Figure 13, and CD prediction errors are reduced by about 70% compared with conventional VTR model.

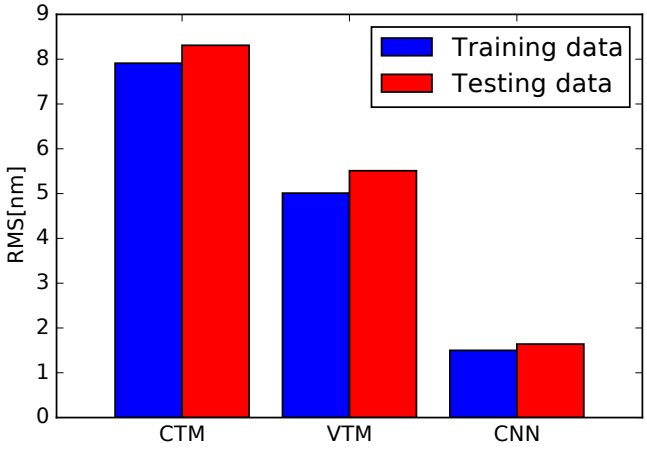


Figure 13. Comparison with conventional CTR and VTR model.

5. CONCLUSION

In this paper, we have proposed a new compact resist model based on CNN that is one of deep learning method to improve lithography simulation accuracy. By using a set of aerial images and ideal thresholds which define experimental CD for training, appropriate aerial image features are automatically extracted and parameters for CNN architecture are optimized. Experimental results for via patterns show squared correlation coefficients between ideal thresholds and predicted thresholds are more than 0.95 on both training data and testing data. We have evaluated CD prediction errors and confirmed that RMS is less than 1.64nm with the proposed method. This result indicates the accuracy of the proposed method is superior to the existing methods.

REFERENCES

[1] Cobb, N. and Zakhor, A., "Fast sparse aerial-image calculation for OPC," Proc. SPIE 2621, p534 (1995)
[2] Satake, M., Kariya, M., Tanaka, S., Hashimoto, K., and Inoue, S., "Evaluation of Lithography Simulation Model Accuracy for Hotspot-based Mask Quality Assurance," Proc SPIE 6607, 66701E (2007)
[3] Granik, Y., Cobb, N. and Do, T., "Universal process modeling with VTRE for OPC," Proc. SPIE 4691, 377-394 (2002)
[4] Mack, C., "A new fast resist model: the Gaussian LPM," Proc SPIE 7974, 79740B (2011)
[5] Zach, F., "Neural Network based approach to resist modeling and OPC," Proc. SPIE 5377, 670-679 (2004)
[6] Kingma, D. and Ba, J., "Adam: A Method for Stochastic Optimization," 3rd International Conference for Learning Representations, (2015)
[7] Matsunawa, T., Nojima, S. and Kotani, T., "Automatic Layout Feature Extraction for Lithography Hotspot Detection Based on Deep Neural Network," SPIE Vol. 9781, (2016)
[8] Dahl, G., Sainath, T., and Hinton, G., "Improving deep neural networks for LVCSR using rectified linear units and dropout," ICASSP, (2013)
[9] Le, Q., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B. and Ng, A., "On optimization methods for deep learning," ICML, (2011)