IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 9, 2016, accepted October 5, 2016, date of publication October 19, 2016, date of current version November 18, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2618350

Artificial Neural Network for Diffraction Based Overlay Measurement

HUNG-FEI KUO AND ANIFATUL FARICHA
[2]Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei 106, Taiwan
Corresponding author: H.-F. Kuo (hfkuo@mail.ntust.edu.tw)

This work was supported by the Ministry of Science and Technology under Project 104-3011-E-011-001.

ABSTRACT Diffraction-based overlay (DBO) accuracy is critical to the intelligent nanolithography process control for producing advanced semiconductor fabrication nodes. Optical gratings located on various layers are commonly used as the targets for the detection of the overlay displacement offset in DBO measurement. The asymmetry in intensity between the 1st and -1st beams diffracted by the targets is used for the prediction of grating displacement offset. This paper describes the effect of grating targets with sidewall angles (SWAs) on asymmetry in intensity and proposes an artificial neural network (ANN) method for enhancing the accuracy of grating displacement offset prediction. Grating targets with a 1:3 line-to-pitch ratio and SWA profiles varying from 86° to 90° were employed in this paper. The asymmetry in the intensity of the designed targets was computed for incident beams with transverse-electric and transverse-magnetic polarization at visible wavelength. An ANN feed-forward model was developed for the displacement offset prediction. The ANN, the conventional linear model, and the regression models were evaluated using diffraction data calculated by a numerical electromagnetic solver. The mean square error and the mean of the residual indicated that using the ANN model and incident beams at wavelengths of 600, 650, and 750 nm is substantially more effective for prediction than the conventional linear model is.

INDEX TERMS Artificial neural network (ANN), diffraction based overlay (DBO), nanolithography, optical scatterometry, sidewall angle (SWA).

I. INTRODUCTION

The intelligent manufacturing known as Industry 4.0 will radically change how manufacturing is completed. The usage of greater connectivity and information sharing enabled by new capabilities in data analytics and modeling will lead to increased efficiency, raised quality, and reduced costs, for example, the intelligent nanolithography in the semiconductor manufacturing process on the 193nm immersion lithography systems [1]–[3]. A nanolithography technique has been proposed to realize the intelligent manufacturing to push Si-based semiconductor process toward to the single-digital nanometer scale. The newly developed process for the sub-20 nm nodes on the 193-immersion platforms, such as self-aligned double patterning process, multiple patterning process, require the stringent control of placement errors between the two process layers to maintain yield and performance [4]–[6]. The overlay parameter is used to define the placement error between the two layers. According to the international technology roadmap for semiconductors (ITRS) report, the major contributors to overlay errors are overlay metrology tools, lithography systems, and processing; metrology tool errors contribute to 20% of error budgets [7]–[9].

In overlay measurement schemes, overlay targets are arranged on the scribed lanes of the wafers for overlay detection by the metrology tool. For advanced technology nodes such as those of 20-nm and below, the overlay target locations are expected to be close to the device patterns and more specialized in-die targets are required for the control of on-product overlay. In this contex, smaller targets are essential for a better representation of device overlay. However, the overlay precision and uncertainty depend upon $1/\sqrt{L}$ where L is the target pattern length [10]–[13]. These requirements resulted the conventional critical dimension scanning electron microscope (CD-SEM) tool suffering from characterizing overlay in the advanced node [14].

Diffraction based overlay (DBO) metrology is favored because it has fast, non-destructive, stable, and repeatable measurement capability compared-with the CD-SEM [15]–[17]. In addition, DBO metrology is well-suited to being integrated with a 193-immersion lithography system for controlling the lithography process variations from die-to-die, wafer-to-wafer, and lot-to-lot [18]. DBO metrology is based on an optical scatterometry setup and records the diffraction response according to the designed overlay targets on the process layers. Each overlay target consists of a pair of binary grating stacks [19]. The first grating on the layer before the layers to be measured is treated as a reference. The lithography engineer sets the line-to-pitch ratio for the second grating on the layer to be measured. The lateral displacement between the two gratings results in an intensity difference between the +1st and −1st diffraction orders. The diffraction intensity map is recorded in the CCD and converted into the overlay at the overlay target location [20], [21].

In the lithography process, resist patterns exhibit profiles that have varying sidewall angles (SWAs) because of the exposure and development process. The gratings on the target layer inherit the SWA profiles from the lithography process. The grating's SWA plays a critical role in the intensity distribution of the diffraction beams. Researchers have reported that the performance of unsymmetrical SWAs differs from that of symmetrical ones; in this case, an incident light is divided into unequal distribution intensities according to the symmetrical diffraction orders [22], [23]. Therefore, this SWA profile causes DBO measurement errors and results in asymmetry uncertainty in the intensity distribution within the linear window [24]–[27].

This study investigated the effect of SWA profiles on the overlay target and proposed a correction scheme for enhanceing the overlay measurement accuracy. In Section II, we describe the construction of an overlay target by using a stack grating model and an evaluation performed using a numerical tool. In Section III, we present the correction scheme and demonstrate the numerical results and verification. In Section IV, we present our conclusion.

II. DBO TARGET THEORY BASED ON FIRST ORDER APPROXIMATION

DBO measurement in the nanolithography process is based on the intensity of backward diffraction by a pair of stack gratings used as an overlay target, as shown in Fig. 1 [28]. In Fig. 1, the overlay target profile is an arrangement of two

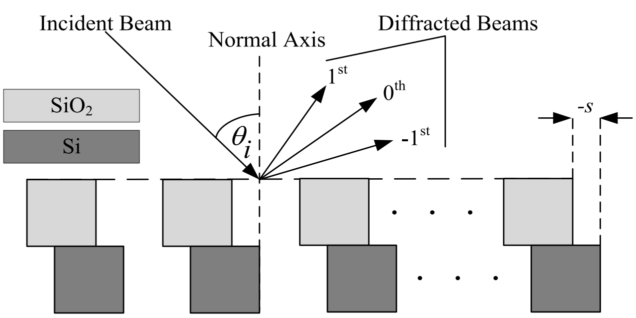
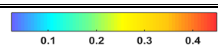


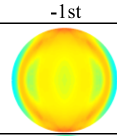
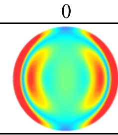
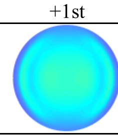
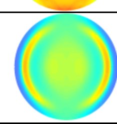
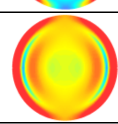
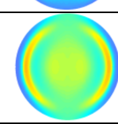
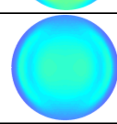
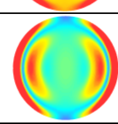
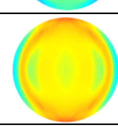
FIGURE 1. Overlay target consisting of a pair of stacked gratings.

gratings that is displacement offset s. One grating is on the first layer and the other one is on the second layer. Each grating has a 90° SWA profile. A plane wave illuminates the overlay target at an angle θ_i. The diffracted beams from the overlay target comprised of the 0^{th}, $+1^{st}$, and -1^{st} orders. A grating equation describes the relationship between the pitch and the diffraction orders as $p(\sin\theta_i + \sin\theta_r) = m\lambda$, where m is the diffraction order, p is the pitch, λ is the incident beam wavelength, θ_i is the incident angle, and θ_r is the reflection angle [29]–[31]. The Si grating is placed on the first layer and the SiO_2 grating on the second layer shifts away from the normal axis horizontally when the overlay is induced during the lithography process. A negative sign in s denotes the grating sifting to the left side of the normal axis.

The finite difference time domain (FDTD) method by lumerical commercial software was used to analyze the diffraction response by the overlay target described in Fig.1. The transverse-electric (TE) polarized beam parameters for the backward diffraction calculation were an incident wavelength of 532 nm, a 1:3 line-to-pitch ratio, and a grating height of 94 nm. The normalized diffraction intensity maps are depicted in Table 1 with respect to the displacement offsets $+s$, 0, and $-s$, respectively. When the displacement offset is 0, the diffraction intensities of the −1st and 1st orders are equal to each other. When the second grating shifts left from the normal axis by $-s$, the diffraction intensity of the +1st order is stronger than that of the −1st order. When the second grating shifts to the right from the normal axis by s, the diffraction intensity of the −1st order is stronger than that of the +1st order. The displacement of two grating offsets causes unequal distribution intensity between the +1st and −1st order.

TABLE 1. Normalized backward diffraction intensity maps with and without overlay displacement offset.



Offset	-1st	0	+1st
$+s$			
0			
$-s$			

The asymmetry in intensity between the +1st and −1st orders having the displacement offset s is characterized by the subtraction of normalized intensity between the +1st and −1st orders and defined as $\Delta I = I_{+1} - I_{-1}$. Fig. 2 presents the asymmetry in intensity ΔI between the two first orders as a function of the displacement offset s. In the linear window

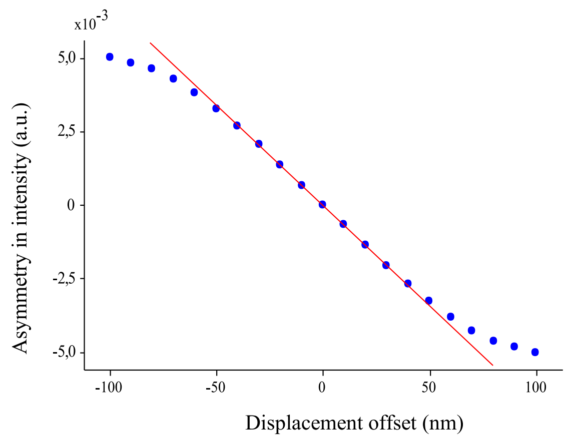


FIGURE 2. Asymmetry in intensity ΔI as a function of displacement offset s.

between +50 and −50 nm, the asymmetry in intensity ΔI is proportional to the displacement offset s and can be denoted as $\Delta I = K \cdot s$, where K is a proportional factor. The induced overlay OV, and the designed bias d are combined as the s, where the bias d is intentionally introduced to eliminate the K factor for OV retreving. Equation (1) describes the asymmetry in intensity with the positive bias $+d$, and Equation (2) describes the asymmetry in intensity with the negative bias $-d$. Therefore OV as a function of asymmetry in intensity can be derived in (3) [32], [33].

$$\Delta I^{+d} = K * (OV + d) \tag{1}$$

$$\Delta I^{-d} = K * (OV - d) \tag{2}$$

$$OV = d * \left(\frac{\Delta I^{+d} + \Delta I^{-d}}{\Delta I^{+d} - \Delta I^{-d}} \right), \tag{3}$$

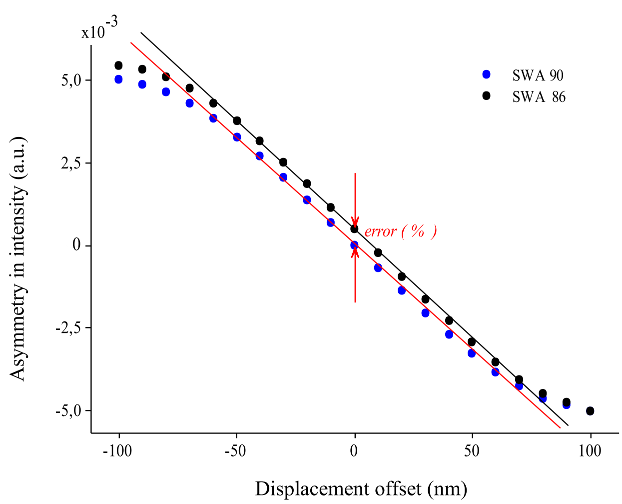


FIGURE 3. Effect of a target with an SWA profile on the asymmetry in intensity ΔI as a function of displacement offset s.

Fig. 3 illustrates the effect of varying SWAs profiles on the overlay target, it shows the ΔI discrepancy between the grating with 90° SWA and that with 86° SWA. This discrepancy affects the accuracy of the overlay estimation in (3). Therefore, a revision of the conventional linear model for the grating target with finite SWA profiles is necessary to enhance the overlay estimation accuracy of the DBO measurement.

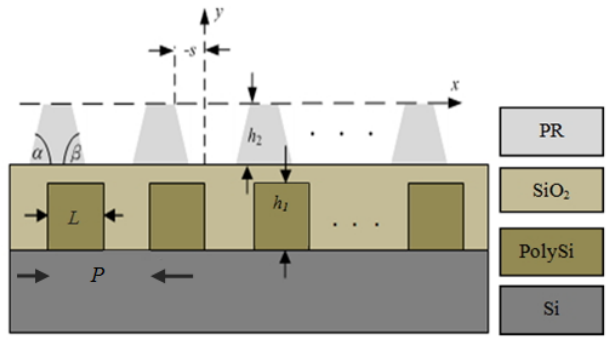


FIGURE 4. Stacked grating target with an SWA profile characterized by α and β.

III. ANN MODEL FOR GRATING TARGETS WITH SWA PROFILES

A schematic layout of a grating target with a finite SWA profile is shown in Fig. 4. The displacement offset s describes the lateral shift between the two layers and includes the designed bias and induced overlay, a negative sign in s denotes the grating shift being to the left side of the normal axis. The gratings on the first and second layers were made of PolySi and PR respectively. Each grating had a line-to-pitch ratio ($L : P$) of 1:3. The inter-layer between the two gratings was SiO_2, the h_1 and h_2 were 200 nm, and the left and right SWAs for the PR gratings were characterized by α and β respectively. The SWA α and β were varied from 90° to 86°. The incident beam for the study is at the wavelength from 400 nm to 700 nm. The normally incident beam was applied at 400 to 700 nm wavelengths. TE and transverse-magnetic (TM) polarization were used. Fig. 5 displays the numerical results for the asymmetry in intensity-displacement offset data sets for SWAs with α of 90°, 89°, 88°, 87°, and 86° and an SWA β of 90°. A SWA α of 90° indicates a grating target without an SWA profile in which the conventional linear model can be adopted for the overlay estimation. In Fig. 5, the TE beam is set at a wavelength of 400 nm for the computation of ΔI. When the displacement offset increased, the ΔI discrepancy between gratings with SWA profiles and those without SWA profiles increased. The ΔI for the grating with an SWA α of 86° showed the largest deviation from the grating with an α of 90°; this uncertainty regarding the ΔI distribution affects the grating displacement offset prediction in overlay metrology.

The ΔI sensitivity to s variation was analyzed to identify the most sensitive wavelengths for overlay detection, as shown in Fig. 6. The ΔI sensitivity was determined according to *slope* in (4) for TE and TM polarization at wavelengths

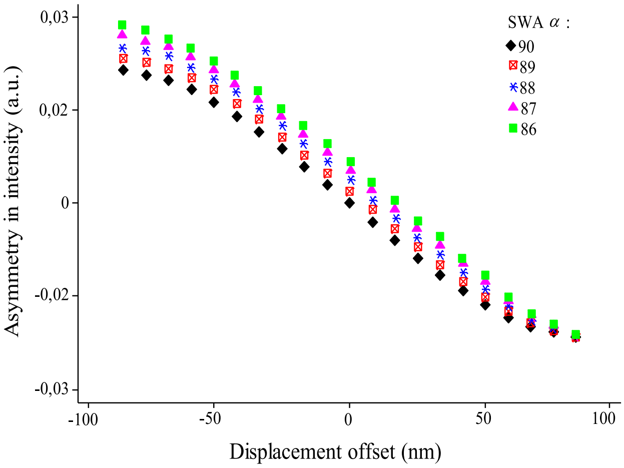


FIGURE 5. Asymmetry in intensity distribution ΔI with respect to various α when $\beta = 90°$ and the incident TE beam wavelength is 400 nm.

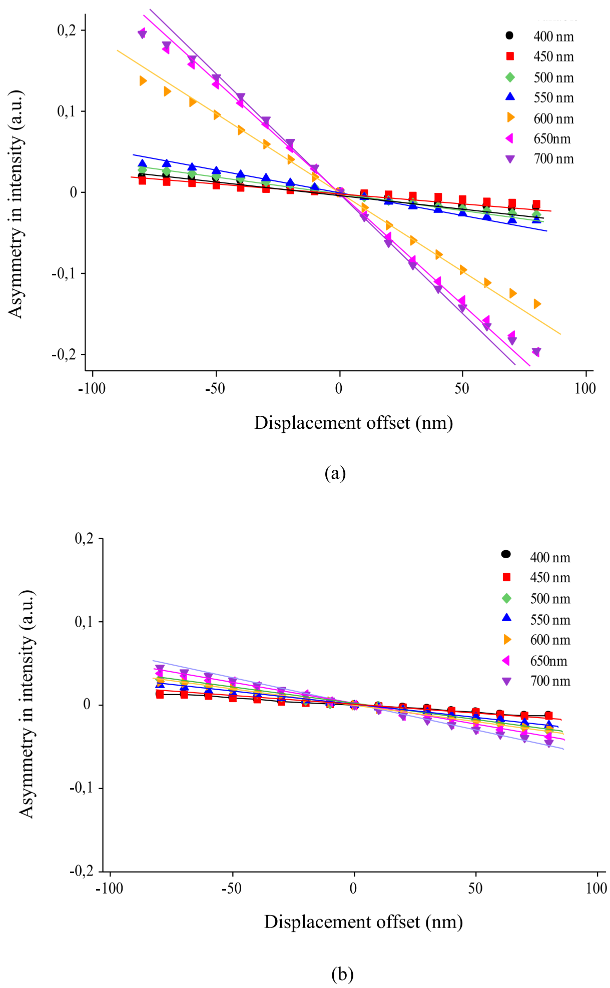


FIGURE 6. Asymmetry in intensity ΔI distribution with respect to α and β of 90° at various wavelengths when using (a) TE polarization and (b) TM polarization.

from 400 to 700 nm. The term ΔI^{+d} describes the asymmetry in intensity with the positive bias $+d$; ΔI^{-d} describes the asymmetry in intensity with the negative bias $-d$. A higher *slope* is more advantageous because it can distinguish s variation more clearly [34], [35]. According to Table 2, wavelengths of 600 nm, 650 nm, and 700 nm are the most sensitive wavelengths to the s variation for both TE and TM polarization.

$$slope = \frac{\Delta I^{+d} - \Delta I^{-d}}{(+d) - (-d)}. \tag{4}$$

TABLE 2. ΔI **sensitivity analysis for displacement offset variation using TE and TM polarization.**

Polarization	Wavelength (nm)	*Slope* ($\times 10^{-4}$)
TE	400	-3.762
	450	-3.120
	500	-1.766
	550	-4.947
	600	-19.92
	650	-27.66
	700	-12.43
TM	400	-3.760
	450	-0.0824
	500	-1.437
	550	-1.934
	600	-3.805
	650	-5.041
	700	-5.929

The ΔI sensitivity to the s variation was further analyzed by investigating the presence of the SWA profiles in the grating target at wavelengths of 600, 650, and 700 nm with TE and TM polarization. The results in Table 3 demonstrate that TE polarization is more effective for distinguishing s variation in various SWA profiles than TM polarization is (Table 4). Thus, a TE polarized beam incident on the grating target was used for the correction scheme study.

TABLE 3. The ΔI **sensitivity to the displacement offset including side wall angle profiles for the TE polarization beam.**

Wavelength (nm)		*Slope* ($\times 10^{-4}$)				
		α				
	β	86°	87°	88°	89°	90°
600	86°	-19.9	-20.3	-20.7	-21.1	-21.4
	87°	-19.5	-19.9	-20.3	-20.7	-21.1
	88°	-19.1	-19.5	-19.9	-20.3	-20.7
	89°	-18.7	-19.1	-19.5	-19.9	-20.3
	90°	-18.3	-18.7	-19.1	-19.5	-19.9
650	86°	-27.6	-28.0	-28.5	-28.9	-29.2
	87°	-27.2	-27.6	-28.0	-28.5	-28.9
	88°	-26.7	-27.2	-27.6	-28.0	-28.5
	89°	-26.3	-26.8	-27.2	-27.6	-28.0
	90°	-25.8	-26.3	-26.7	-27.2	-27.6
700	86°	-12.4	-12.2	-12.0	-11.8	-11.6
	87°	-12.6	-12.4	-12.2	-11.8	-11.8
	88°	-12.8	-12.6	-12.4	-12.2	-12.0
	89°	-13.0	-18.0	-12.6	-12.4	-12.2
	90°	-13.2	-13.0	-12.8	-12.6	-12.4

The artificial neural network (ANN) technique is proposed as the correction scheme for the prediction of the grating displacement offset and associated uncertainty of the SWA profiles in the DBO target. A neural network consists of

TABLE 4. The ΔI sensitivity to the displacement offset including side wall angle profiles for the TM polarization beam.

Wavelength (nm)		Slope (×10⁻⁴)				
		α				
	β	86°	87°	88°	89°	90°
600	86°	-4.27	-4.28	-4.28	-4.29	-4.30
	87°	-4.07	-4.07	-4.06	-4.08	-4.10
	88°	-3.86	-3.90	-3.89	-3.97	-3.98
	89°	-3.88	-3.90	-3.89	-3.90	-3.90
	90°	-3.92	-3.87	-3.88	-3.83	-3.80
650	86°	-5.45	-5.47	-5.48	-5.49	-5.51
	87°	-5.43	-5.44	-5.46	-5.45	-5.45
	88°	-5.27	-5.28	-5.30	-5.30	-5.33
	89°	-5.10	-5.13	-5.17	-5.20	-5.23
	90°	-4.90	-4.93	-4.94	-5.00	-5.04
700	86°	-5.67	-5.67	-5.67	-5.67	-5.67
	87°	-5.70	-5.70	-5.70	-5.70	-5.70
	88°	-5.72	-5.72	-5.72	-5.72	-5.72
	89°	-5.88	-5.88	-5.88	-5.88	-5.88
	90°	-5.92	-5.90	-5.91	-5.91	-5.92

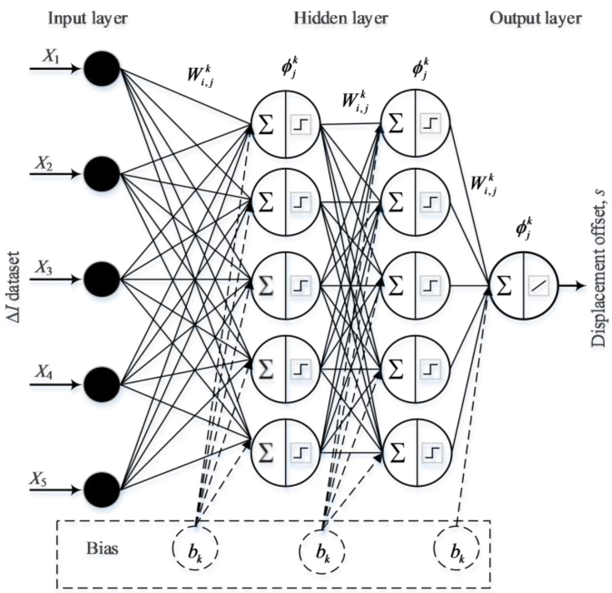


FIGURE 7. Architecture of neural network having one input layer, two hidden layers, and an output layer. The ΔI dataset is considered as the data input and the displacement offset s is presented as the output.

an array of interconnected nodes (nominally representing neurons) that are arranged as an input layer, hidden layer, and output layer. A node in a given layer is generally connected to all nodes in the subsequent layer [36]. The architecture of the neural network applied in this study consisted of four layers: one input layer, two hidden layers, and one output layer, as shown in Fig. 7. The ΔI dataset in Table 3 was introduced in the input layer, which contained five nodes. Those input nodes X_1, X_2, X_3, X_4, and X_5 were constructed as the ΔI by using a grating with SWA α of 90°, 89°, 88°, 87°, and 86°. $W_{i,j}^k$ is the weighting coefficient, b_k is the bias, ϕ_j^k is the neuron function, i is the index of neuron inputs, and j and k are the indices of the neurons. The proposed ANN uses a feed-forward model having two hidden layers followed by an output layer. The hidden layers apply a sigmoid transfer function and the output layer applies a linear transfer function. This setup enables the proposed network to learn relationships between the input ΔI and the output s [37].

The proposed feed-forward ANN model was trained using the Levenberg-Marquardt (LM) backpropagation learning process from the output layer backward to the input layer. The LM algorithm combines the steepest descent and Gauss–Newton algorithms during the ANN model training process [38], [39]. Equation (5) describes the difference ε between the actual displacement offset Y_j based on the FDTD numerical computation and the predicted displacement offset $\hat{Y}_j$ calculated by the proposed ANN model. The mean square error (MSE, e) in (6) and the mean of the residual $\bar{\varepsilon}$ in (7) were employed to evaluate the training process, where N is the total number of displacement offset data. In total, 565 offset data points were used for the ANN model training process. The values of 282 offset data points among the 565 data points were calculated according to the positive displacement offset. The values of 283 offset data points among the 565 data points were calculated according to the negative displacement offset. In addition, the values of 240 displacement offset data from the FDTD computation were used for verification. The implemented LM algorithm in the ANN model updated the weighting coefficients according to (8), where $\boldsymbol{J}$ represents the Jacobian matrix, $\boldsymbol{W}$ is the weighting coefficient matrix, $\boldsymbol{I}$ denotes the identity matrix, μ is the combination coefficient, and n is the index of iteration [40].

$$\varepsilon_j = Y_j - \hat{Y}_j \tag{5}$$

$$e = \frac{1}{N}\sum_{j=1}^{N}(\varepsilon_j)^2 \tag{6}$$

$$\bar{\varepsilon} = \frac{1}{N}\sum_{j=1}^{N}|\varepsilon_j| \tag{7}$$

$$\boldsymbol{W}_{(n+1)} = \boldsymbol{W}_{(n)} - (\boldsymbol{J}_{(n)}^T\boldsymbol{J}_{(n)} + \mu\boldsymbol{I})^{-1}\boldsymbol{J}_{(n)}^T\boldsymbol{\varepsilon}_{(n)}. \tag{8}$$

Table 5 describes the implementation of the proposed feed-forward ANN model trained using the LM backpropagation algorithm. The procedure first involves initializing parameters used in the proposed ANN model: the neuron function ϕ_j^k, the weight $W_{i,j}^k$, the input $X_{i,j}$, and the bias b_k. The term *ite* is introduced as the looping parameter and *maxite* is the iteration criterion for terminating the learning procedure. The feed-forward ANN model is constructed from the 4th to 14th steps, which propagate the input X_j forward through the network. The sigmoid transfer function is used in the hidden layer at the ninth step, and the linear transfer function is used in the output layer at the seventh step. The section from the 15th to 25th steps in the procedure is for the error control of the constructed ANN model, where the actual output Y_j from FDTD is used to verify the training error $\boldsymbol{\varepsilon}_j$. The error propagates from the output layer backward through the network to the input layer for further updating the weighting matrix $\boldsymbol{W}$. The matrix $\boldsymbol{W}$ is updated during

TABLE 5. Procedures for the artifical network model with the LM backpropagation algorithm.

1 : Initialize $i, j, k = 1$; $ni = 5$; W; b; $ite = 1$; and $maxite = 300$.
2 : Load
    $X = \Delta I$ dataset
    $Y =$ actual displacement offset
3 : Repeat
4 : For $k = 1$ to 4
5 :   For $j = 1$ to 5
6 :   If $k = 4$
7 : $\phi_j^k = \sum_{i=1}^{ni} W_{i,j}^k X_{i,j} + b_k$
8 :   ElseIf
9 : $\phi_j^k = \dfrac{1}{1 + e^{-\sum_{i=1}^{ni} W_{i,j}^k X_{i,j} + b_k}}$
10 :   End if
11 : $X_j = f_j(\phi_j^k)$;
12 : $\hat{Y}_j = X_j$
13 :   End For
14 : End For
15 : For $k = 4$
16 :   For $j = 1$
17 :   Calculate error using Eq (5)
18 :   End for
19 : End for
20 : For $k = 3$ to 1
21 :   For $j = 1$ to 5
22 :   Update each node's weight using Eq (8)
23 :   End For
24 : End For
25 : Calculate MSE using Eq (6) and $\bar{\varepsilon}$ using Eq (7)
26 : $ite = ite + 1$
27 : While $ite < maxite$.

the training process at the 22nd step according to the LM algorithm in (8). The 25th step entails defining the MSE and $\bar{\varepsilon}$ to evaluate the training process. The convergence of the proposed ANN model was verified through the MSE as a function of iterations, as shown by Fig. 8. An MSE of 0.0534 at the 134th iteration achieves an error percentage of less than 1%; the MSE in the proposed ANN model showed stable behavior after the 150th iteration. In addition, this research validated the proposed ANN architecture. Table 6 and Table 7 report the MSE results when using various number of neurons and hidden layers respectively. Table 8 reports the run-time when using various number of hidden layers.

TABLE 6. Summary of MSE when using various number of neurons in one hidden layer.

Neuron	1	2	3	4	5	6	7
MSE	15.8	3.45	1.75	0.96	0.16	0.16	0.16

To compare the performance of the proposed ANN model, a multiple regression model for the prediction of the displacement offset was used, as shown in (9) [41]. In (9), $\hat{Y}i$ is the i^{th} predicted displacement offset, X_{1i}, X_{2i}, X_{3i}, X_{4i}, and X_{5i} are the ΔI of the gratings with SWA α of 90°, 89°, 88°, 87°, and 86°, respectively; b_1, b_2, b_3, b_4, and b_5 are the coefficients to be calculated. The multiple regression model was developed using 565 offset data points in the positive and negative displacement offset regions. Table 9 summarizes the procedure applied to construct the multiple regression model. When the order equals 1, the multiple regression model reverts to the conventional linear model for the offset estimation. Two quality factors were used to evaluate the capability of the proposed ANN and multiple regression models for predicting the grating offset: MSE, e in (6), and the mean of the residual $\bar{\varepsilon}$ in (7). The e in (6) denotes the standard deviation between the actual displacement offset value Y_i and the predicted displacement offset $\hat{Y}i$. The $\bar{\varepsilon}$ in (7) describes the average of the absolute residual between Y_i (as the referenced displacement offset calculated by the numerical FDTD tool) and the predicted displacement offset $\hat{Y}_i$ evaluated by the regression model. The values of 240 displacement offset data points from FDTD were used for the verification.

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + b_4 X_{4i} + b_5 X_{5i}. \quad (9)$$

Tables 10, 11 and 12 summarize the quality factor results with respect to various wavelengths when using the ANN and multiple regression models. The numerical results in Table 10 demonstrate more effective prediction of the displacement offset by the ANN model when using an incident beam with a longer wavelength. This is a consequence of the incident beam with a longer wavelength being more sensitive to the

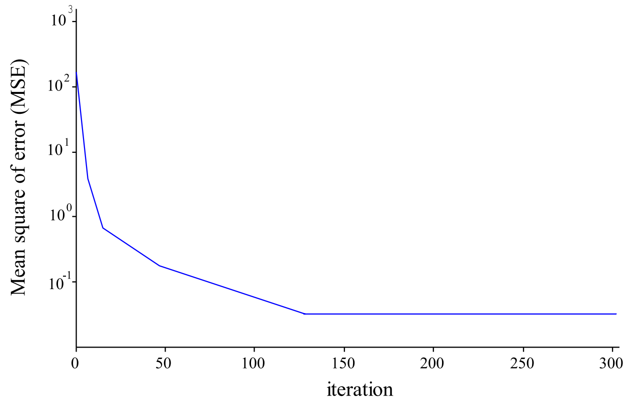


FIGURE 8. Convergence test for the proposed ANN model with the LM backpropagation algorithm.

TABLE 7. Summary of MSE when using five neurons in various number of hidden layers.

Layer	1	2	3	4	5
MSE	0.17	0.03	0.03	0.02	0.02

TABLE 8. The run-time when using various number of hidden layers.

Layer	1	2	3	4	5
Time	37 sec	63 sec	334 sec	672 sec	1087 sec

TABLE 9. Multiple regression procedure for the prediction of the displacement offset.

1 : Initialize $i = 1$; $j = 1$; $N = 565$.
2 : For $i = 1$ to N
3 : Load data
$X_{1i}, X_{2i}, X_{3i}, X_{4i}, X_{5i}$ and Y_i = actual displacement offset
4 : $i = i+1$
5 : End For
6 : Calculate
7 : $\hat{Y}_i^{(1)} = b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + b_4 X_{4i} + b_5 X_{5i}$
8 : $\hat{Y}_i^{(2)} = Y_i^{(1)} + b_6 X_{1i}^2 + b_7 X_{2i}^2 + b_8 X_{3i}^2 + b_9 X_{4i}^2 + b_{10} X_{5i}^2$
9 : $\hat{Y}_i^{(3)} = Y_i^{(2)} + b_{11} X_{1i}^3 + b_{12} X_{2i}^3 + b_{13} X_{3i}^3 + b_{14} X_{4i}^3 + b_{15} X_{5i}^3$
10 : $\hat{Y}_i^{(4)} = Y_i^{(3)} + b_{16} X_{1i}^4 + b_{17} X_{2i}^4 + b_{18} X_{3i}^4 + b_{19} X_{4i}^4 + b_{20} X_{5i}^4$
11 : $\hat{Y}_i^{(5)} = Y_i^{(4)} + b_{21} X_{1i}^5 + b_{22} X_{2i}^5 + b_{23} X_{3i}^5 + b_{24} X_{4i}^5 + b_{25} X_{5i}^5$
12 : If order = 1
Calculate b_0
$\hat{Y}_i = b_0 + Y_i^{(1)}$
13 : If order = 2
Calculate b_0
$\hat{Y}_i = b_0 + Y_i^{(2)}$
14 : If order = 3
Calculate b_0
$\hat{Y}_i = b_0 + Y_i^{(3)}$
15 : If order = 4
Calculate b_0
$\hat{Y}_i = b_0 + Y_i^{(4)}$
16 : Elseif order = 5
Calculate b_0
$\hat{Y}_i = b_0 + Y_i^{(5)}$
17 : End If
18 : For j=1 to k
$j = j+1$
Calculate MSE using Eq (6), and $\bar{\varepsilon}$ using Eq (7)
19 : End For.

TABLE 10. Quality factors for the displacement offst estimation using the ANN model.

λ (nm)	Process: Training e	Training $\bar{\varepsilon}$	Verification e	Verification $\bar{\varepsilon}$
600	0.0275	0.124	0.0416	0.148
650	0.0209	0.094	0.0331	0.135
700	0.0258	0.108	0.0342	0.142

TABLE 11. The training Proces for the displacement offset estimation using the multiple regression model with various orders.

Quality factor	λ (nm)	Order 1	2	3	4	5
e	600	3.756	3.501	0.158	0.156	0.156
	650	3.543	3.356	0.233	0.222	0.222
	700	3.721	3.512	0.093	0.090	0.090
$\bar{\varepsilon}$	600	1.558	1.157	0.371	0.353	0.350
	650	1.467	1.133	0.425	0.422	0.420
	700	1.474	1.165	0.296	0.287	0.287

grating offset than a beam with a shorter wavelength is. The results also reveal that the ANN model has more favorable prediction capability than the regression model does. Table 11 and 12 present the numerical results of e and $\bar{\varepsilon}$ for the training and verification process using multiple regression

TABLE 12. The verification process for the displacement offset estimation using the multiple regression model with various orders

Quality factor	λ (nm)	Order 1	2	3	4	5
e	600	4.937	4.704	0.279	0.270	0.269
	650	4.605	4.343	0.324	0.322	0.320
	700	4.809	4.726	0.198	0.193	0.193
$\bar{\varepsilon}$	600	1.957	1.483	0.971	0.954	0.950
	650	1.865	1.135	0.896	0.887	0.887
	700	1.897	1.340	0.925	0.922	0.921

model. Lower e and $\bar{\varepsilon}$ occur when the regression models are at higher orders than the first and second orders (i.e., the third, fourth, and fifth orders). The factors e and $\bar{\varepsilon}$ in Table 11 and 12 decreased substantially for the third order and converged for the fourth and fifth orders. The results of e and $\bar{\varepsilon}$ show that the ANN model predicts the displacement offset more effectively than the multiple regression model does.

IV. CONCLUSION

The feed-forward ANN model for predicting displacement offset of the stack grating targets with various SWAs was successfully demonstrated. The MSE and mean of the residual results showed that the ANN model provides more effective displacement offset prediction than the linear and high order regression models do. The ANN model substantially reduced the impact of the asymmetrical intensity variation within the linear window caused by the SWAs in the overlay targets. In addition, the results demonstrated that the regression model provides a lower MSE at the third and higher orders than at the lower orders. These findings indicate that the combination of the ANN model and optical beams with a longer wavelength can be integrated with the optical scatterometry tool for in-die overlay measurement. Analysis of the line width roughness of the overlay target will be included in the ANN self-learning process in future.

REFERENCES

[1] K. R. Khorram *et al.*, "Cost of ownership/yield enhancement of high volume immersion lithography using topcoat-less resists," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 2, pp. 173–181, May 2011.

[2] M. van den Brink, "Holistic lithography and metrology's importance in driving patterning fidelity," *Proc. SPIE*, vol. 9778, p. 977802, Apr. 2016.

[3] E. Schmitt-Weaver *et al.*, "An integrated approach to holistic metrology qualification for multi-patterning process layers: AM: Advanced metrology," in *Proc. IEEE Annu. SEMI Adv. Semiconductor Manuf. Conf. (ASMC)*, May 2016, pp. 16–19.

[4] Y. Jiao and D. Djurdjanovic, "Stochastic control of multilayer overlay in lithography processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 404–417, Aug. 2011.

[5] R. S. Ghaida, G. Torres, and P. Gupta, "Single-mask double-patterning lithography for reduced cost and improved overlay control," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 1, pp. 93–103, Feb. 2011.

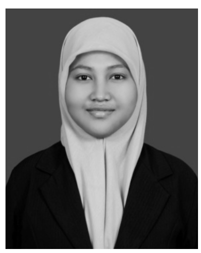
[6] S. S. Mehta *et al.*, "Process variation challenges and resolution in the negative-tone develop double patterning for 20nm and below technology node," *Proc. SPIE*, vol. 9425, p. 94250B, Mar. 2015.

[7] S.-C. Horng and S.-Y. Wu, "Compensating the overlay modeling errors in lithography process of wafer stepper," in *Proc. IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2010, pp. 1399–1404.

[8] H. Lee *et al.*, "Device overlay method for high volume manufacturing," *Proc. SPIE*, vol. 9778, p. 97781F, Mar. 2016.

[9] R. S. Ghaida and P. Gupta, ''Within-layer overlay impact for design in metal double patterning,'' *IEEE Trans. Semicond. Manuf.*, vol. 23, no. 3, pp. 381–390, Aug. 2010.
[10] H.-J. H. Smilde *et al.*, ''Target design optimization for overlay scatterometry to improve on-product overlay,'' *Proc. SPIE*, vol. 9424, p. 942412, Mar. 2015.
[11] T. Kato, Y. Okagawa, O. Inoue, K. Arai, and S. Yamaguchi, ''Fundamentals of overlay measurement and inspection using scanning electron-microscope,'' *Proc. SPIE*, vol. 8681, p. 86812Q, Apr. 2013.
[12] Z. Shi, V. Kochergin, and F. Wang, ''193nm superlens imaging structure for 20nm lithography node,'' *Opt. Exp.*, vol. 17, no. 14, pp. 11309–11314, 2009.
[13] M. Masahara *et al.*, ''Advanced FinFET process technology for 20 nm node and beyond,'' in *Proc. IEEE Nanoelectron. Conf.*, Jun. 2011, pp. 1–2.
[14] C. Chen *et al.*, ''Device-correlated metrology for overlay measurements,'' *J. Micro/Nanolithogr., MEMS, MOEMS*, vol. 13, no. 4, p. 041404, Oct. 2014.
[15] H.-C. Chin *et al.*, ''Metrology solutions for high performance germanium multi-gate field-effect transistors using optical scatterometry,'' *Proc. SPIE*, vol. 8681, p. 86813D, Apr. 2013.
[16] Y.-N. Kim *et al.*, ''Device based in-chip critical dimension and overlay metrology,'' *Opt. Exp.*, vol. 17, no. 23, pp. 21336–21343, 2009.
[17] A. Vaid *et al.*, ''A holistic metrology approach: Hybrid metrology utilizing scatterometry, CD-AFM, and CD-SEM,'' *Proc. SPIE*, vol. 7971, p. 797103, Mar. 2011.
[18] M. G. Faruk *et al.*, ''Enabling scatterometry as an in-line measurement technique for 32 nm BEOL application,'' *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 4, pp. 499–512, Nov. 2011.
[19] R.-J. Kou *et al.*, ''Immersion scanner proximity matching using angle resolving scatterometry metrology,'' *Proc. SPIE*, vol. 7272, p. 727240, Mar. 2009.
[20] S. van Haver *et al.*, ''Wafer-based aberration metrology for lithographic systems using overlay measurements on targets imaged from phase-shift gratings,'' *Appl. Opt.*, vol. 53, no. 12, pp. 2562–2582, 2014.
[21] S. Peterhänsel, M. L. Gödecke, V. F. Paz, K. Frenner, and W. Osten, ''Detection of overlay error in double patterning gratings using phase-structured illumination,'' *Opt. Exp.*, vol. 23, no. 19, pp. 24246–24256, 2015.
[22] J. Mulkens, P. C. Hinnen, M. Kubis, A. V. Padiy, and J. P. Benschop, ''Holistic optimization architecture enabling sub-14-nm projection lithography,'' *J. Micro/Nanolithogr., MEMS, MOEMS*, vol. 13, no. 1, pp. 011006-1–011006-7, 2014.
[23] D. L. Voronov, L. I. Goray, T. Warwick, V. V. Yashchuk, and H. A. Padmore, ''High-order multilayer coated blazed gratings for high resolution soft X-ray spectroscopy,'' *Opt. Exp.*, vol. 23, no. 4, pp. 4771–4790, 2015.
[24] X. Yang, Q. Huang, I. V. Kozhevnikov, Z. Wang, J. Zhao, and Y. Wu, ''Development of XUV multilayer gratings with high resolution and high efficiency,'' *Proc. SPIE*, vol. 9510, p. 95100Y, May 2015.
[25] K. Cho, S. Tarutani, N. Inoue, H. Tsubaki, and M. Neisser, ''LWR study on resist formulation parameters,'' *Proc. SPIE*, vol. 8682, p. 868217, Mar. 2013.
[26] H. Hody, V. Paraschiv, D. Hellin, T. Vandeweyer, G. Boccardi, and K. Xu, ''Gate double patterning strategies for 10nm node FinFET devices,'' *Proc. SPIE*, vol. 9054, p. 905407, Mar. 2014.
[27] S. Takei, T. Ogawa, and C. G. Willson, ''Study of fluorinated silicon-based resist material and photoreactive underlayer for defect reduction in step and repeat ultraviolet nanoimprint lithography,'' *IET Micro Nano Lett.*, vol. 6, no. 6, pp. 422–424, Jun. 2011.
[28] W.-T. Hsu, Y.-S. Ku, and D.-M. Shyu, ''A novel method for overlay measurement by scatterometry,'' *Proc. SPIE*, vol. 7432, p. 74320F, Sep. 2009.
[29] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[30] C. Palmer, *Diffraction Grating Handbook*, 6th ed. Irvine, CA, USA: Newport Corporation, 2005.
[31] J. T. Verdeyen, *Laser Electronics*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.
[32] Y. Blancquaert and C. Dezauzier, ''Diffraction based overlay and image based overlay on production flow for advanced technology node,'' *Proc. SPIE*, vol. 8681, p. 86812O, Apr. 2013.
[33] H. J. Jeong, S. J. Rabello, and T. A. Casavant, ''Dark field diffraction based overlay,'' U.S. Patent 0 278 942 A1, Oct. 24, 2013.
[34] W. T. Hsu, Y. S. Ku, H. L. Pang, and D. M. Shyu, ''Method for designing two-dimensional array overlay target sets and method and system for measuring overlay errors using the same,'' U.S. Patent 8 250 497 B2, Aug. 21, 2012.
[35] P. Dasari, J. Li, J. Hu, N. Smith, and O. Kritsun, ''Diffraction based overlay metrology for double patterning technologies,'' in *Nanofabrication Techniques and Applications*. Rijeka, Croatia: InTech, 2011.
[36] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.
[37] S. Razavi and B. A. Tolson, ''A new formulation for feedforward neural networks,'' *IEEE Trans. Neural Netw.*, vol. 22, no. 10, pp. 1588–1598, Oct. 2011.
[38] M. Fairbank and E. Alonso, ''Efficient calculation of the Gauss–Newton approximation of the Hessian matrix in neural networks,'' *Neural Comput.*, vol. 24, no. 3, pp. 607–610, Mar. 2012.
[39] R. Zhang, Z.-B. Xu, G.-B. Huang, and D. Wang, ''Global convergence of online BP training with dynamic learning rate,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 2, pp. 330–341, Feb. 2012.
[40] K. Malllikarjuna, G. R. C. K. Sarma, M. V. Subramanyam, and K. S. Prasad, ''EBP based GK-LM method for neural network training,'' in *Proc. Int. Conf. Comput. Res. Develop. (ICCRD)*, Mar. 2011, pp. 504–507.
[41] Q. Zhao *et al.*, ''Higher order partial least squares (HOPLS): A generalized multilinear regression method,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1660–1673, Jul. 2013.

HUNG-FEI KUO received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2004. From 2005 to 2006, he served as an Optical Department Manager, GCC, Taiwan. From 2007 to 2009, he was an Application Development Engineer with ACE, ASML. Since 2010, he has been with the Faculty, Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei, China. His research interests include the development of next generation lithography techniques and optical system integrations.



ANIFATUL FARICHA received the B.A. degree in electrical engineering from the Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. She is currently pursuing the M.S. degree with the Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei, Taiwan. Her research interests include the diffraction-based overlay measurement and the artificial neural network.