IEEE *Access*

Linewidth Reconstruction Employing a Radial Basis Function Network in Optical Scatterometry

HUNG-FEI KUO[1], MUHAMAD FAISAL[1], AND SHUN-FENG SU[2], (Fellow, IEEE)
[1]Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei 106, Taiwan
[2]Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan
Corresponding author: H.-F. Kuo (hfkuo@mail.ntust.edu.tw)

ABSTRACT This paper applied a radial basis function network (RBFN) in coherent Fourier scatterometry (CFS) to reconstruct the linewidth of periodic line/space (L/S) patterns. The fast, nondestructive, and repeatable measurement capability of CFS enables its integration with intelligent lithography systems. Two steps to reconstruct the linewidth of the L/S patterns were performed in this paper. The first step was to use the finite difference time domain numerical electromagnetic tool to rigorously establish the library of modeled diffraction signatures by using the L/S patterns. Each modeled signature was converted to an intensity vector as the training data to construct the RBFN. The trained RBFN has a simple architecture consisting of three layers: input, hidden, and output layers. The second step was to collect the experimental signatures and feed them into the trained RBFN model to predict the linewidth of L/S patterns. This paper used the transverse electric polarized incident beam at the wavelength of 632 nm in the experimental setup of the CFS. Five L/S patterns were used to test the constructed RBFN. The experimental results indicated that the maximal difference was 13 nm between the CFS and the atomic force microscopy (AFM) measurements for the sample D with an L/S of 200 nm. The minimum difference was 2 nm for the sample A with an L/S of 140 nm. The correlation coefficient between the CFS and AFM metrology measurement running through five samples was 0.972. The high correlation between the CFS with the proposed RBFN measurements and the AFM revealed the potential to implement the radial basis learning kernel in optical metrology to achieve intelligent lithography.

INDEX TERMS Coherent Fourier scatterometry (CFS), critical dimension (CD) reconstruction, holistic lithography, optical scatterometry, radial basis function network (RBFN).

I. INTRODUCTION

The newly developed holistic lithography technique is an intelligent method for manufacturing the advanced process nodes used in the semiconductor industry [1]–[3]. Holistic lithography is enabled through the integration of in-line metrology and computational tools, which increases the efficiency and quality of the lithography process. As stated in an International Technology Roadmap for Semiconductors report, the metrology tools for linewidth (or critical dimension, CD) measurement in the semiconductor process must be nondestructive and compatible with in-line metrology [4]. The CD-scanning electron microscope (SEM) tool has been commonly used to characterize device structures for decades. However, because the device structure designs have become increasingly complicated and diversified to realize multiple functions on chips, such conventional metrology tools have

become insufficient [5], [6]. Optical scatterometry is emerging as a complementary tool for CD-SEM [7]–[9] because it performs fast and nondestructive measurement of periodic line/space (L/S) patterns.

Several configurations of optical scatterometry exist, including specular [10], angular [11], polarimetry [12], and ellipsometry [13]. Recently, coherent Fourier scatterometry (CFS) has attracted considerable attention because of the high-speed and high-sensitivity measurement enabled through this optical configuration [14]–[17]. CFS measurement is based on diffraction by periodic L/S patterns [18], [19]. Two steps are required to characterize the linewidth of L/S patterns in CFS. The first step is a forward modeling process to obtain modeled diffraction signatures through the periodic L/S pattern, which is treated as the surface relief grating. Rigorous numerical electromagnetic

techniques such as rigorous coupled wave analysis, finite difference time domain (FDTD), and an integral equation solver can effectively compute the diffraction signatures from L/S patterns on a nanometer scale [20], [22]. The next step is to reconstruct the linewidth of the L/S pattern, which is widely known as the inverse problem [23], [24].

The inverse problem in optical scatterometry has attracted much attention and is considered essential to characterizing the linewidth of L/S patterns. The library search method has been proposed to solve the inverse problem [25], [26]. However, the library search method is limited by the time-consuming nature of its computation, because the method requires numerous theoretical signatures to complete the accuracy measurement [27]–[29]. Various algorithms, such as the recursive random search and genetic algorithm, can also be applied in the inverse problem [30]. However, the reconstruction results from these methods are less accurate than those from the library search method. Another research direction for improving the drawbacks of methods for solving the inverse problem is toward the use of an intelligent algorithm; for example, an artificial neural network (ANN) [31]–[33].

An ANN is trained by a collection of modeled signatures and estimates the linewidth of the L/S pattern according to experimental signatures. However, when an ANN has an excessive number of input terminals, its architecture becomes complicated and it tends to become instable, resulting in it being impractical for implementation in the semiconductor manufacturing process [34]–[36]. A radial basis function network (RBFN), an expanded version of the ANN, is an algorithm with the potential to solve the inverse problem in optical scatterometry. Compared with the ANN, the RBFN has architecture that is easy to implement [37]. A Gaussian activation function is employed in the hidden layer to map the nonlinear input data. Consequently, the RBFN is sensitive to noisy data [38]. Several papers have proposed and discussed the applications of the RBFN in various systems such as dipole antenna [38], robotic [39], earthquake prevention [40], and cardiology interval-valued data systems [41].

This paper proposes an RBFN based on the forward selection technique to reconstruct the linewidth of L/S patterns [42]. The forward selection is expected to obtain a parsimonious network. Section II of this paper reviews the diffraction theory used for CFS and presents the L/S pattern design for producing the theoretical signatures. Section III explains the proposed RBFN and the linewidth reconstruction procedure. Section IV demonstrates the experimental results, and Section V provides a conclusion.

II. THEORETICAL SIGNATURES BY THE DESIGNED PATTERNS

Fig. 1 explains the measurement principle of the optical scatterometry, with diffraction by the periodic L/S pattern as the surface relief grating. The grating is characterized by the following geometrical parameters: linewidth (w), pitch (p), height (h), and sidewall angle (SWA). When the plane wave,
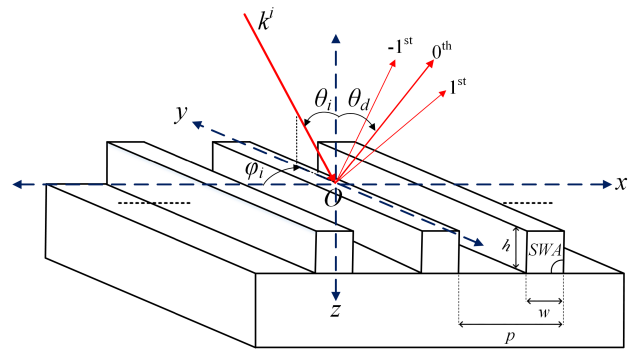

FIGURE 1. Incident beam on a surface relief grating by the polar θ_i and azimuthal φ_i angles being diffracted by the grating into the zeroth, negative first, and first orders.

as indicated by the wave vector k^i, illuminates the grating by the polar angle θ_i and azimuthal φ_i, the wave is diffracted into the directions of the zeroth order, $\pm$ first order, and higher orders. The incident angle θ_i and the diffracted angle θ_d for the nth diffraction order beam are related by the Braggs law as $\sin\theta_i + \sin\theta_d = n\lambda/p$, where λ is the wavelength and n is the diffraction order [47]. The total fields diffracted by the L/S pattern are described as follows:

$$U(x, y) = A_0 \sum \exp\left[i\left(k^i \sin\theta_i + \frac{2\pi n}{p}\right)x\right], \quad (1)$$

where n is an integer. The diffracted field is indicated by the wave vector $k^n = (k_x^n, 0, k_z^{n\prime})$, corresponding to the incident angle θ_i such that $k^n \sin\theta_d = k^i \sin\theta_i + 2\pi n/p$.
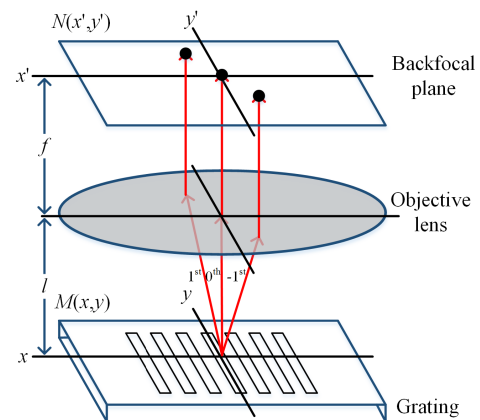

FIGURE 2. Diffracted plane waves on $M(x, y)$ plane are collected by an objective lens and propagated onto the backfocal plane $N(x\prime, y\prime)$.

An objective lens is placed above the L/S pattern to collect the diffraction fields, as shown in Fig. 2. The $M(x, y)$ plane is placed on the L/S pattern surface. The separation between the objective lens and the $M(x, y)$ plane is l. The plane $N(x\prime, y\prime)$ is defined as the backfocal plane of the objective lens with the focal length f. This study relates the optical field on the $M(x, y)$ plane with the field on the backfocal plane $N(x\prime, y\prime)$ in the optical system based on the Fourier optics theory.

The diffracted beams travel at the angles $\theta_x = \sin^{-1}\lambda v_x$ and $\theta_y = \sin^{-1}\lambda v_y$ in directions corresponding to the x-axis and y-axis, respectively. The diffracted beam is mapped using the lens onto the point (x', y') on the backfocal plane, where $x' = \theta_x f = \lambda v_x$, $y' = \theta_y f = \lambda v_y$. The complex amplitude on the backfocal plane $N(x', y')$ is formulated as follows:

$$N(x', y') \approx h_0 F\left(\frac{x}{\lambda l}, \frac{y}{\lambda l}\right), \tag{2}$$

where $F(v_x, v_y)$ is the Fourier transform of the $U(x, y)$, and $h_0 = (j/\lambda d)\exp(-jks)$ is the interference equation [46]. The intensity at each point (x', y') on the backfocal plane $N(x', y')$ results from the diffraction caused by the L/S pattern when the incident beam is used at the angles θ_i and φ_i. The diffracted angle θ_d for the nth order is limited by the numerical aperture (NA) of the objective lens. Therefore, only the diffracted beams within the NA are further propagated onto the backfocal plane.
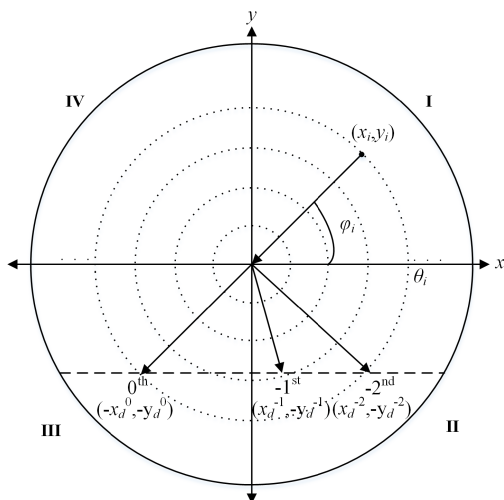

FIGURE 3. Locations of the diffracted beams at the zeroth and negative first, and second orders observed from the top view of the objective lens [47].

Fig. 3 illustrates the top view of the objective lens and the relation between the incident beam and the diffracted beam within the NA. The solid circle represents the effective lens area defined by the NA divided by four quadrants on an x-y Cartesian plane. Each circle with a dotted line represents a polar angle θ_i mapped to the Cartesian plane. The grating is placed parallel to the y-axis and perpendicular to the x-axis. The beam is incident on the grating by the polar angle θ_i and the azimuthal angle φ_i. The x_i and y_i on the Cartesian plane are related to the incident angles by the $x_i = \sin\theta_i \cos\varphi_i$ and $y_i = \sin\theta_i \sin\varphi_i$ [47]. Equation (5) describes the location of x_d^n and y_d^n standing for the nth diffracted beam on the x-y Cartesian plane. For example, the beam is incident on the grating by the angles of $\theta_i = 60°$ and $\varphi_i = 45°$ when in quadrant I. The zeroth order diffracted beam is located in quadrant III. The negative first- and second-order diffracted beams are located in quadrant II.
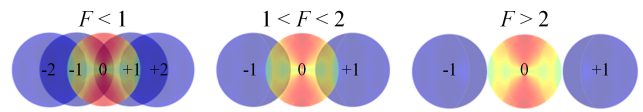

FIGURE 4. The degree of overlap, *F*, between the diffracted orders on the backfocal plane.

However, the first- and second-order diffracted beams are outside of the NA of the objective lens. Equation (3) is also applicable when the incident beam (x_i, y_i) is located in quadrants II, III, or IV. The intensity distribution map on the $N(x', y')$ plane is obtained by adding the diffracted beams from various incident angles. The diffraction intensity map indicates the overlap of diffracted beams, as depicted in Fig. 4. The degree of overlapping between the negative first, zeroth, and first diffraction orders is described in (4). Overlap exists between the negative first and first orders when $F < 1$; overlap also exists between the negative first and zeroth, and the zeroth and first orders when $1 < F < 2$. Otherwise, no overlap exists between diffraction orders when $F > 2$ [19].

$$\left(x_j^n, y_j^n\right) = \left(-x_i - n\frac{\lambda}{p}, -y_i\right), \tag{3}$$

$$F = \frac{\lambda}{NA \times p}. \tag{4}$$

TABLE 1. Geometrical parameter of designed grating.

	Grating Parameter			
	Period (nm)	Linewidth (nm)	Height (nm)	SWA (°)
A	700	140	300	90
B	800	160	300	90
C	900	180	300	90
D	1000	200	300	90
E	1100	220	300	90

Although (3) predicts the location of the nth diffracted beam, rigorous electromagnetic modeling is still required to calculate the intensity of the nth diffracted beams. The diffraction field of the L/S pattern was calculated using the Lumerical FDTD software in this study. Table 1 summarizes the L/S pattern design parameters used in the study. Five designed L/S patterns were used, with the $h = 300$ nm, $SWA = 90°$, and various w/p including 140 nm/700 nm, 160 nm/800 nm, 180 nm/900 nm, 200 nm/1000 nm, and 220 nm/1100 nm. The L/S patterns were formed on the Si substrates as one-dimensional surface relief gratings. The incident beam with transverse electric (TE) field polarization was at the wavelength of 632 nm. The incident polar angle θ_i was between 0° and 33°, including the NA effect. The azimuthal angle φ_i was between 0°–360°. The refractive index was 3.825 and k was −0.026 for the modeling of the Si gratings. The modeled diffraction intensity from each

L/S pattern in Table 1 was defined as the modeled signature for the pattern. The library of model signatures contained the diffraction intensity from the L/S patterns listed in Table 1. The library also contained the signatures of the various L/S patterns including the designed linewidth range of ±20 nm with a 1 nm step each, the designed height of ±20 nm with a 5 nm step each, and the designed *SWA* of 86°–90° with a 2° step. Therefore, the number of modeled signature maps was 1107 for each grating sample.
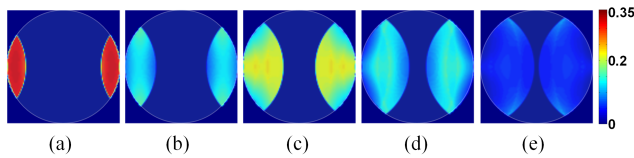

FIGURE 5. Modeled signatures of the designed gratings for *w/p* = (a) 140/700 nm; (b) 160/800 nm, (c) 180/900 nm; (d) 200/1000 nm; (e) 220/1100 nm.

This research only computed the signature map with the ± first diffraction orders because they were more sensitive to the grating structure variations than the zeroth order [9]. Fig. 5 illustrates the modeled signature maps with the ± first diffraction orders for the five designed cases listed in Table 1. The maximal diffraction efficiency was 0.3168 for the A pattern, 0.1654 for the B pattern, 0.1964 for the C pattern, 0.1361 for the D pattern, and 0.0852 for the E pattern. We further verified the degree of overlap for the modeled signatures, as displayed in Fig. 5. The calculated *F* was 1.64, 1.43, 1.27, 1.14, and 0.14 respective to the cases illustrated in Figs. 5(a), 5(b), 5(c), 5(d), and 5(e). The *F* results indicated that the degrees of overlap between the modeled first and zeroth diffraction orders, and the modeled zeroth and negative first diffraction orders, were consistent with the predictions achieved using (4).
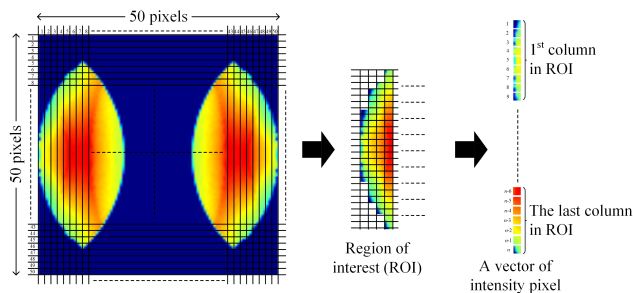

FIGURE 6. Process flow of formatting the modeled signature into an intensity vector.

Each modeled signature map illustrated in Fig. 5 was converted to a set of intensity vectors then used as the training data for constructing the subsequent RBFN. Fig. 6 depicts the transformation process flow from a modeled signature map to the sets of intensity vectors. First, each modeled signature map was formatted as an image with 50×50 pixels. Effective pixels among the 2500 pixels were assigned to minimize the computation resources used in this research.

In this study, the effective pixels were defined as the pixels for which the diffraction efficiency was not 0. The effective pixels of the image were grouped as a region of interest (ROI). Each column in the ROI was extracted and all columns were recombined in the ROI as an intensity vector. The number of intensity vectors was 1107 for each grating sample. Each element in the intensity vector corresponded to one node in the input layer. The number of input nodes was equal to the number of elements in the intensity vector. The intensity vector of the input layer and the linewidth of the output layer were used as the training data to construct the RBFN. Table 2 summarizes the number of intensity vectors and nodes of the input layer for each grating pattern listed in Table 1. The number of intensity vectors was 1107 for all patterns and the numbers of input nodes were 388, 760, 1084, 1380, and 1612 respective to grating patterns A, B, C, D, and E. A larger modeled signature map size was correlated with the generation of a higher number of input nodes.

TABLE 2. Number of intensity vectors and input nodes for constructing the RBFN.

	Number of Intensity Vector	Number of Input Node
A	1107	388
B	1107	760
C	1107	1084
D	1107	1380
E	1107	1612

III. RBFN ALGORITHM IN CFS

This research proposed using an RBFN to solve the inverse problem of CFS. The proposed RBFN applies the idea of a receptive field to the ANN. A receptive field is a region in which a sensory field of cells elicits a response [43]. Fig. 7 illustrates that the architecture of the proposed RBFN
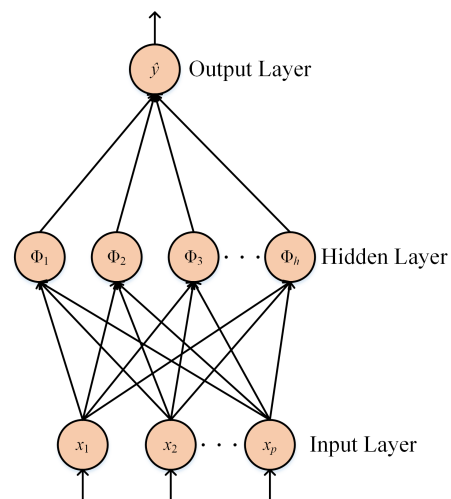

FIGURE 7. Architecture of radial basis function network.

consists of three layers: input, hidden, and output layers. The receptive field is represented as the Gaussian activation function located in the hidden layer. The nonlinear function by the Gaussian function converts the intensity vector in the input layer to a scalar number in a node in the hidden layer. The linear transformation maps the hidden layer to the output layer. The parameters in the proposed RBFN, of the number of hidden layers (h), the location of the Gaussian center ($\mathbf{c}$), the radii of the center (σ), and the weighting factors (ω) are closely related to the performance. The number of neurons in the hidden layer is determined when the RBFN with the neurons has a minimized root-mean-square error (RMSE). Through a conventional method, the location of the Gaussian center is determined using the training data ($\mathbf{c}_i = \mathbf{x}_i$, $1 \le i \le s$). However, such a center selection tends to be overly sensitive to the input vectors and results in an overfitting problem. To avoid this problem, we applied Tikhonov regularization and forward selection techniques [41]. The radii of the center were initially selected. The performance of the proposed RBFN was analyzed using two cross-validation methods, resubstitution and k-fold cross validation, to obtain the optimal network [42].

Equation 5 relates the output $\hat{y}_i$ and the inputs $\mathbf{x}_i$, $\in$, and R^n through the hidden layer function Φ_i for the proposed RBFN, in which the y_i corresponds to the grating linewidth w and the $\mathbf{x}_i$ corresponds to the ith column in the training data:

$$\hat{y}_i = \sum_{j=1}^{h} \omega_j \Phi_{ji}(\|\mathbf{x}_i - \mathbf{c}_j\|), \qquad (5)$$

where ω_j is the weighting factor and Φ_{ji} is the Gaussian activation function, denoted as follows:

$$\Phi_{ji} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{c}_j\|}{\sigma_j^2}\right), \quad i = 1, 2, \ldots, s, \ j = 1, 2, \ldots, h, \qquad (6)$$

where the norm is the Euclidean distance. The $\mathbf{c}_j$ and σ_j are the jth center and radii used in the function, respectively. Equation (5) can be formulated to a linear system equation to simplify the calculation of Tikhonov regularization and forward selection [41]:

$$\hat{y} = \Psi \omega + \mathbf{e}, \qquad (7)$$

where Ψ is the hidden layer response to the n-intensity vector and $\mathbf{y} = [y_1 \ldots y_p]^T$ is the set of the s linewidth. The vector $\mathbf{e}$ represents the s errors between the estimated $\hat{\mathbf{y}}$ and the desired $\mathbf{y}$. The purpose of using (7) is to determine the combination of Ψ and ω to obtain the estimated $\hat{\mathbf{y}}$ with the minimized RMSE of $\mathbf{e}$.

In Tikhonov regularization, the regularization penalty Γ is added to $E = \sum (\hat{\mathbf{y}} - y)^2 + \Gamma \omega^T \omega$ to reduce the value of the center. The regularization penalty parameter is used to adjust the balance between the estimated $\hat{\mathbf{y}}$ and the penalty $\Gamma \omega^T \omega$. If Γ decreases, the estimated $\hat{\mathbf{y}}$ is fitted tightly to the desired $\mathbf{y}$.

If Γ increases to cause the penalty, the estimated $\hat{\mathbf{y}}$ is nullified. The solution to (7) is as follows:

$$\omega = (\mathbf{F}^T \mathbf{F} + \Gamma \mathbf{I}_p)^{-1} \mathbf{F}^T \hat{\mathbf{y}}, \qquad (8)$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix, $\mathbf{F}$ is the full design matrix constructed using the vector intensity extracted from the training data, and each vector intensity corresponds to the selected radii.

In the forward selection technique, a subset center Ψ_{k-1} in (11a) is updated by adding a new center $\mathbf{f}_i$ from the full design matrix $\mathbf{F}$ in each iteration. The updated subset center Ψ_k is used to calculate the regularized weighting factor ω_k in (10). The energy E_k at the kth step used as the cost function to construct the RBFN is then calculated on the basis of the regularized weighting factor in (11a) or (11b). A new center $\mathbf{f}_i$ is determined using (12). This $\mathbf{f}_i$ is again added to the subset center Ψ_{k-1} until the maximal energy difference is achieved in (13):

$$\Psi_k = [\Psi_{k-1} \quad \mathbf{f}_i] \qquad (9)$$

$$\omega_k = (\Psi_k^T \Psi_k + \Gamma_k \mathbf{I})^{-1} \Psi_k^T \mathbf{y}, \qquad (10)$$

$$E_k^{(i)} = \mathbf{e}_k^T \mathbf{e}_k + \Gamma_k \omega_k^T \omega_k \qquad (11a)$$

$$E_k^{(i)} = \hat{\mathbf{y}}^T \mathbf{P}_k \hat{\mathbf{y}}, \qquad (11b)$$

where k is the kth iteration, i is the ith column, and

$$\mathbf{P}_k = \mathbf{I}_p - \Psi_k(\Psi_k^T \Psi_k + \Gamma_k \mathbf{I}_k)^{-1} \Psi_k^T;$$
$$E_k^{(i)} \le E_k^{(j)}, \quad 1 \le j \le p \qquad (12)$$

$$E_{k-1} - E_k^{(i)} = \frac{(\hat{\mathbf{y}}^T \mathbf{P}_{k-1} \mathbf{f}_i)^2}{\Gamma_k + \mathbf{f}_i^T \mathbf{P}_{k-1} \mathbf{f}_i}. \qquad (13)$$

For $\Gamma_k > 0$, the Γ_k is re-estimated using generalized cross-validation (GCV), given as follows:

$$\text{GCV}_k = \frac{1}{s} \frac{\|\mathbf{P}_k \hat{\mathbf{y}}\|^2}{((1/s) trace(\mathbf{P}_k))^2}, \qquad (14)$$

where s is the number of training data and

$$\mathbf{P}_k = \mathbf{I}_p - \sum_{j=1}^{k} \frac{\psi_j \psi_j^T}{\Gamma_k + \psi_j \psi_j^T}.$$

Here, GCV_k indicates the variance of the estimated $\hat{\mathbf{y}}$ in the RBFN. The ψ_k represents the new selected center $\mathbf{f}_i$ for updating the subset center Ψ_{k-1}. The re-estimated Γ_{k+1} is calculated using the differentiating (14) with respect to Γ_k, formulated as follows:

$$\Gamma_{k+1} = \frac{\hat{\mathbf{y}}_k^T \mathbf{P}_k^2 \hat{\mathbf{y}}_k \, trace(\mathbf{A}_k^{-1} - \Gamma_k \mathbf{A}_k^{-2})}{\omega_k^T \mathbf{A}_k^{-1} \omega_k \, trace(\mathbf{P}_k)}, \qquad (15)$$

where $\mathbf{A}_k = (\Psi_k^T \Psi_k + \Gamma_k \mathbf{I})$.

The derivation of (14) to (15) is discussed in [41].

This research developed the procedure listed in Tables III to construct the RBFN by using the equations from (5) to (13). The first step was to normalize the training data and initialize the radii of the center σ, the regularization penalty Γ_0, GCV, and the number of hidden layers h. The normalized

TABLE 3. RBFN procedure with the forward selection technique.

```
==========================================================
1   : Normalizing data training (x,y)
2   : Select σ, Γ₀, GCV, h
3   : Construct F
4   : While iteration k = 1 : h
5   :    Add one center fᵢ to Ψ in (9)
6   :    Calculate the regularized weighting in (10)
7   :    Calculate the current energy in (11)
8   :    Calculate the maximum difference of the energy in (13)
9   :    If k = h
10  :       Break
11  :    End If
12  :    If (13) is achieved
13  :       Break
14  :    Else
15  :       Estimate the regularization penalty Γₖ₊₁ in (15)
16  :       Back to 5
17  :    End If
18  : End While
==========================================================
```

training data and the defined radii were added to construct the full design matrix **F**. After the new center $\mathbf{f}_i$ was selected from the **F**, the new center was added to Ψ_k. The new subset center Ψ_k was used to calculate the regularized weighting in step 6. The current energy in each iteration was determined in step 7. The energy at the kth $-$ 1 iteration was compared with the energy at the kth iteration. If the maximal energy difference between the two iterations was achieved in step 8, the iteration was stopped. Otherwise, the iteration in the procedure returned to step 5 and the new center $\mathbf{f}_i$ was determined again. The regularization penalty Γ_k was re-estimated, and the iteration was continued until the number of iterations was equal to the h. Two methods exist to evaluate the RBFN developed in this research: resubstitution validation and k-fold cross validation [42].

The resubstitution validation was performed by varying the number of neurons in the hidden layer to verify that the developed RBFN was the optimal network with the minimized RMSE. The full design matrix **F** was constructed using the training data corresponding to the radii σ varying from 0.5 to 2.5.The number of neurons in the hidden layer was varied from 300 to 700 to verify the minimized RMSE. The regularization parameter was 0.01 and the GCV was 10. Fig. 8 illustrates that the RMSE was a function of the number of neurons in the hidden layer. When the trained RBFN that employed the theoretical signatures of sample A used 382 neurons, the error achieved the minimized RMSE of 0.083. This result meant that the maximal difference of energy was achieved by 382 neurons in the hidden layer. The RMSE for sample A was higher than that of the other samples. This result could have been caused by the limited vector intensity pixels of sample A. For samples B and C, the trained RBFN used 465 and 477 neurons to achieve the minimized RMSE of 0.009 and 0.008, respectively. This result indicated
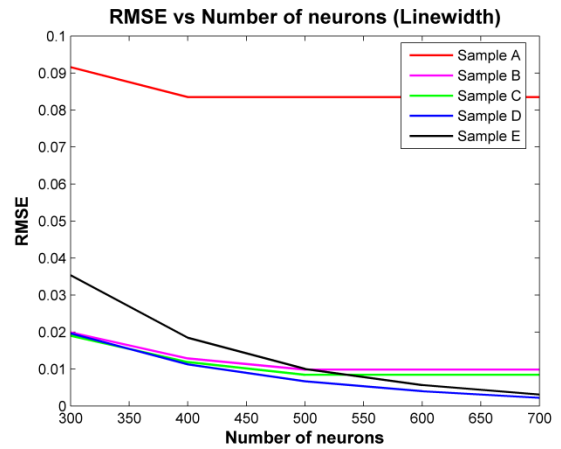

FIGURE 8. RMSE as a function of the number of neurons in the hidden layer.

TABLE 4. RBFN performance by k-fold cross validation.

	RMSE	
	Training Data	Testing Data
A	0.107	0.170
B	0.015	0.047
C	0.012	0.053
D	0.009	0.072
E	0.007	0.169

that the RBFN was optimized by the number of neurons in the hidden layer. For samples D and E, this study used 700 neurons because the RMSE for both cases was below 0.01 for the proposed number of neurons in the hidden layers.

After the number of neurons was validated, the RBFN performance was further characterized using the k-fold cross validation. The modeled signature data was divided into two parts: 90% of the data was used to train the network and 10% of the data was used to test the generalization of the network. The trained RBFN used the optimized number of neurons for the k-fold cross validation. Table 4 lists the validation results of the constructed RBFN. The Table 4 column indicated by the training data reports the comparison between the designed linewidth and the reconstructed linewidth determined using the trained RBFN. The Table 4 column indicated by the testing data reports the k-fold cross validation results. The RMSE values between the designed and the reconstructed linewidth determined using the training data were 0.107, 0.015, 0.012, 0.009, and 0.007 respective to samples A, B, C, D, and E. The constructed RBFN for sample A had an RMSE that was higher than the RMSE of the RBFNs for the other samples because it had a lower density of intensity pixels.

The RMSE was 0.170, 0.047, 0.053, 0.072, and 0.169 respective to samples A, B, C, D, and E when the k-fold cross validation was used for the network performance evaluation. The constructed networks with a higher RMSE for samples
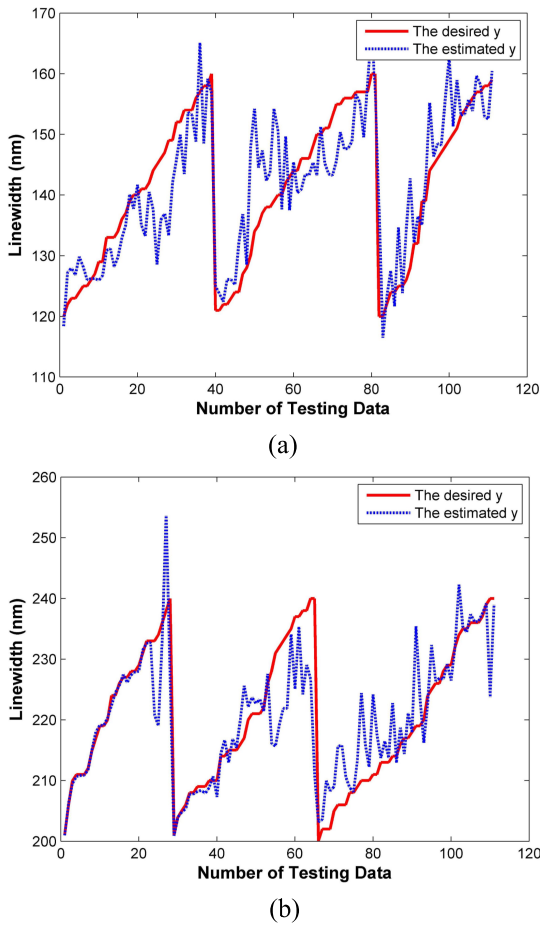
FIGURE 9. The *k*-fold cross validation used to test the constructed RBFN performance for (a) sample A; (b) sample E.

TABLE 5. Measurement results achieved using SEM and AFM metrology.

A and E implied a poorer generalization for the networks than did the networks of samples B, C, and D. This research further traced the performance of the RBFN for samples A and E during the linewidth reconstruction process, as shown in Fig. 9. In Fig. 9, the searching trend of the **ŷ** indicated by the blue dot line follows the trend of **y** as indicated by the red line. Because of this, the networks were still acceptable in this research. The RMSE for the networks corresponding to the B, C, and D samples was lower than 0.01, indicating that the linewidth could be adequately reconstructed.

IV. EXPERIMENTAL MEASUREMENT

The measurements performed using the constructed RBFN were conducted through five fabricated Si gratings by using e-beam lithography and reactive ion etching techniques. The designed profiles of the five gratings are summarized in Table 1. Each grating was patterned onto a 2×2 mm sample. Sample A had 2857 L/S pairs. Sample B had 2500 L/S pairs. Samples C, D, and E had 2222, 2000, and 1818 L/S pairs, respectively. The number of L/S pairs was sufficient in the CFS measurement because the measurement required at least 250 L/S pairs to achieve a stable diffraction map [8]. In Table 5, the second column lists the separate SEM images

of the top surfaces of the Si gratings on the five samples. The third column reports the linewidth measurements of the Si gratings on the five samples achieved using AFM metrology.

The experimental setup of CFS is depicted in Fig. 10. This research used a TE polarized He–Ne laser at the wavelength
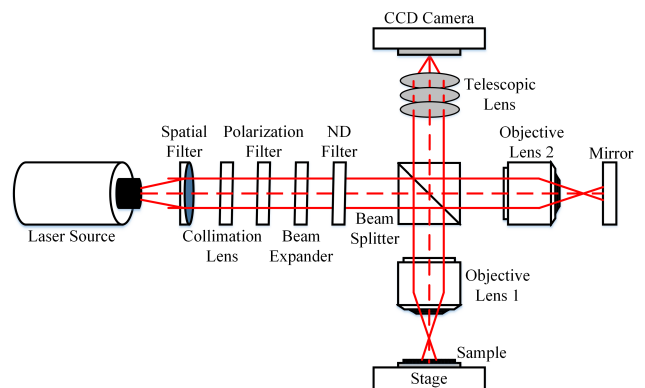

FIGURE 10. Experimental setup of CFS.

of 632 nm to acquire the experimental signatures. The laser output was 15 mW. The laser beam passed through a spatial filter of a collimation lens to generate the collimated beam. The polarization of the collimated beam was regulated by a TE polarization filter, followed by a beam expander to uniformly distribute the beam intensity. To avoid charge-coupled device (CCD) saturation, two neutral-density filters were installed on the optical path to enable beam intensity transmission rates of 25% and 10%. The beam was then divided by a beam splitter at a ratio of 50:50. The first beam was focused using an objective lens with an NA of 0.55. The focused beam illuminated the sample on a controllable stage capable of achieving the x-y-z translation and a 360° rotation. The diffracted beam caused by the sample was collected by the objective lens and refocused to the CCD camera on the backfocal plane through the telescopic lens consisting of several best-form lenses. The diffraction intensity map was captured by the high resolution CCD camera with 2560×1920 pixels and a $2.2 \times 2.2 \ \mu$m pixel size. This study used the second beam from the beam splitter as a reference beam to remove the image noise caused by the optics.
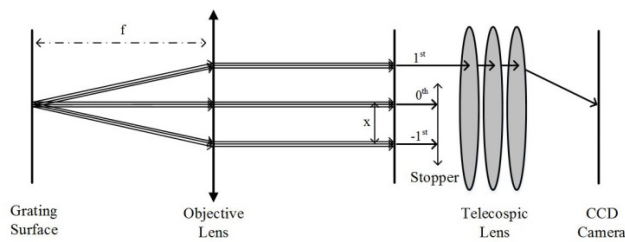

FIGURE 11. Illustration of the beam stopper inserted in the CFS experiment setup to extract the first diffraction order beam.
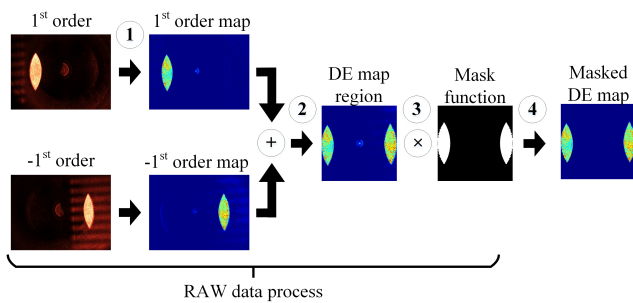

FIGURE 12. Process flow of converting the raw images of the ± first diffraction order to the masked diffraction efficiency map.

The experiment required acquiring the ± first diffraction orders separately. A stopper was inserted in front of the telescopic lens to block the unwanted diffraction orders, as shown in Fig. 11. To acquire the negative first diffraction orders, the zeroth and first diffraction beams were blocked. Similarly, each first diffraction order was acquired by blocking the negative first and zeroth diffraction beams. Fig. 12 illustrates the process flow of the transformation from the raw image data to the diffraction efficiency maps. After the raw images of the ± first diffraction orders were separately acquired,
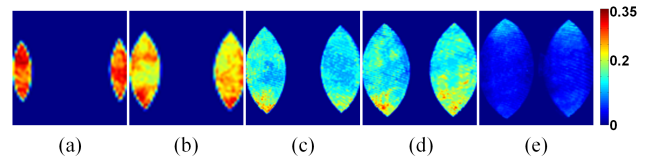

FIGURE 13. Masked diffraction efficiency maps from the five Si gratings with the designed linewidth w = (a) 140 nm; (b) 160 nm; (c) 180 nm; (d) 200 nm; (e) 220 nm.

the raw images were further converted into the diffraction efficiency maps on the basis of the $\eta = p_{diffs}/p_{inc}$, where p_{diffr} was the measured power for the diffracted beams and p_{inc} was the power for the incident beam. Next, the ± first diffraction efficiency map was masked using the mask function created from the modeled signatures. The mask function filtered the non-0 diffraction intensity pixels on the basis of the modeled signature data. Fig. 13 illustrates the masked diffraction efficiency maps corresponding to the five grating samples. The experimental images were not as smooth as the modeled signature images because of the noise that occurred in the measurement process. This research removed the noise pixels when the pixel intensity level was out of the range of the maximal intensity level in the modeled signature data. In addition, the removed noise pixels were used as references for the modeled intensity vectors to retrain the RBFN. Each experimental signature was then converted into an intensity vector as the input to the retrained RBFN to reconstruct the linewidth of each Si grating.

This research further compared the measurement results between the CFS with the constructed RBFN and the AFM. Each grating sample was measured five times by using the CFS with the constructed RBFN, and each measurement point was separated by 100 μm. The average of five measurement results was used as the CFS reconstruction results for each Si grating sample. Table 6 summarizes the measurements from the AFM in its fourth column. The CFS with the RBFN reconstructed the linewidth of the gratings as 139 nm, 153 nm, 176 nm, 181 nm, and 217 nm respective to samples A, B, C, D, and E. These results were in good agreement with the AFM measurements. The maximal difference between the two measurements was 12 nm for sample D, whereas the minimum difference was 2 nm for sample A. The CFS measurement was the average of the linewidth on the beam illumination area and the AFM measurement was conducted on an isolated spot as the local measurement. The error sources included the small NA being unable to collect full diffraction beams, insufficient imaging pixel resolution, fabricated gratings with linewidth roughness and line-edge roughness, and numerical errors in the calculation of theoretical signatures. The measurement of the CFS also revealed higher standard deviations for samples A and B (3.48 nm and 3.77 nm, respectively) and lower standard deviations for samples C, D, and E (less than ±1.5 nm). The higher than average standard deviation could be caused by the

TABLE 6. Comparison between the designed grating linewidth, the CFS measurement, and the AFM measurement.

	Linewidth (nm)		
	Design	CFS	AFM
A	140	139	137
B	160	153	156
C	180	176	183
D	200	181	193
E	220	217	220

lower than average number of inputs to the trained RBFN. The result of correlation coefficient $R^2 = 0.9720$ indicates that the linewidth measurements achieved using the CFS with the constructed RBFN were strongly correlated with the measurements achieved using AFM.

V. CONCLUSION

This research successfully demonstrated that the RBFN could reconstruct the linewidth of L/S patterns. Both the RBFN procedure and the optical scatterometry setup were presented for the measurement of five various L/S patterns on Si substrates. The study determined the experimental measurements by using optical scatterometry with the proposed RBFN. These results were highly correlated with the measurement results achieved using the AFM metrology tool. The contribution of this research is the integration of the Gaussian-based learning kernel into optical scatterometry for potential use in the process control of semiconductor device manufacturing on the nanometer scale. The proposed RBFN metrology application could be for measuring L/S patterns with uncertainties such as line-edge roughness and linewidth roughness in an intelligent lithography system. Future works should include the introduction of multiple wavelengths and polarizations into the constructed RBFN to enhance measurement accuracy and further reconstruct the completed three-dimensional geometrical parameters of the nanostructures.

REFERENCES

[1] M. V. D. Brink, "Holistic lithography and metrology's importance in driving paterning fidelity," *Proc. SPIE*, vol. 9778, pp. 977802-1–977802-28, Apr. 2016.

[2] K. Bhattacharyya *et al.*, "Holistic approach using accuracy of diffraction-base integrated metrology to improve on-product performance, reduce cycle time and cost at litho," *Proc. SPIE*, vol. 9424, pp. 94241E-1–94241E-8, Mar. 2015.

[3] A. Void *et al.*, "Holistic metrology approach: Hybrid metrology utilizing scatterometry, critical-atomic force micorscope and cirtical dimension- scanning electron microscope," *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 10, no. 4, pp. 043016-1–043016-13, Dec. 2011.

[4] (2015). *International Technology Roadmap for Semiconductor*. [Online]. Available: http://www.itrs.net/

[5] D. Dixit, M. Medikonda, A. C. Diebold, B. Peterson, and J. Race, "Mueller matrix optical scatterometry of Si fins patterned using directed self-assembly block copolymer line arrays," in *Proc. 25th Annu. SEMI ASMC*, May 2014, pp. 180–185.

[6] 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.

[7] A. Void *et al.*, "Hybrid metrolgy: From the lab into the fab," *J. Micro/Nanolithogr., MEMS, MOEMS*, vol. 13, no. 4, pp. 041410-1–041410-20, Nov. 2014.

[8] H. -F. Kuo, Frederick, "Gaussian beam incident on the one-dimensional diffraction gratings with the high-κ metal gate stack structures," *J. Nanosci. Nanotechnol.*, vol. 14, no. 4, pp. 2780–2785, Apr. 2014.

[9] V. Zhuang, S. Spielman, J. Fielden, D. C. Wack, L. Poslavsky, and B. D. Bunday, "Dark-field optical scatterometry for line-width-roughness metrology," *Proc. SPIE*, vol. 7272, pp. 72720L-1–72720L-12, Mar. 2009.

[10] V. P. Korolkov, A. S. Konchenko, V. V. Cherkashin, N. G. Mironnikov, and A. G. Poleshchuk, "Etch depth mapping of phase binary computer-generated holograms by means of specular spectrocopic scatterometry," *Opt. Eng.*, vol. 52, no. 9, pp. 091722-1–091722-6, Apr. 2013.

[11] L. Yang, L. Zeng, and L. Li, "Scatterometry specialized for highly asymmetric triangular grating on a transparent substrate," *Appl. Opt.*, vol. 53, no. 6, pp. 1143–1150, Feb. 2014.

[12] A. Peinado, A. Turpin, A. Lizana, E. Fernández, J. Mompart, and J. Campos, "Conical refraction as a tool for plus polarization metrology," *Opt. Lett.*, vol. 38, no. 20, pp. 4100–4103, Oct. 2013.

[13] M. Losurdo, "Spectroscopic ellipsometry and polarimetry for materials and systems analysis at the relax nanometer scale: State-of-the-art, potential, and perspectives," *J. Nanopart. Res.*, vol. 11, no. 7, pp. 1521–1554, 2009.

[14] V. F. Paz *et al.*, "Depth sensitive Fourier-scatterometry for the characterization of sub-100 nm periodic," *Proc. SPIE*, vol. 8083, pp. 80830M-1–80830M-9, May 2011.

[15] N. Kumar *et al.*, "Coherent Fourier scatterometry: Tool for improved sensitivity in semiconductor metrology," *Proc. SPIE*, vol. 8384, pp. 83240Q-1–83240Q-8, Feb. 2012.

[16] N. Kumar, O. El Gawhary, S. Roy, S. F. Pereira, and H. P. Urbach "Phase information in coherent Fourier scatterometry," *Proc. SPIE*, vol. 8788, pp. 87881P-1–87881P-8, May 2013.

[17] S. Roy, A. C. Assafrão, S. F. Pereira, and H. P. Urbach, "Coherent Fourier scatterometry for detection of nanometer-sized particles on a planar substrate surface," *Opt. Exp.*, vol. 22, no. 11, pp. 13250–13262, May 2014.

[18] V. F. Paz, A. Peterhansel, K. Frenner, and W. Osten, "Solving the inverse grating problem by white light interference Fourier scatterometry," *Light: Sci. Appl.*, vol. 1, no. 11, p. e36, Nov. 2012.

[19] N. Kumar *et al.*, "Reconstruction of sub-wavelength features and nano-positioning of gratings using coherent Fourier scatterometry," *Opt. Exp.*, vol. 22, no. 20, pp. 24678–24688, Oct. 2014.

[20] R. Silver *et al.*, "Fundamental limits of optical critical dimension metrology: A simulation study," *Proc. SPIE*, vol. 6518, pp. 65180U-1–65180U-17, Apr. 2007.

[21] M. Besbes *et al.*, "Numerical analysis of a slit-grrove diffraction problem," *J. Eur. Opt. Soc.- Rapid Publications*, vol. 2, p. 07022, Jul. 2007.

[22] A. Adhikari, K. Dev, and A. Asundi, "Subwavelength metrologival characterization by Muller matrix polarimeter and finite difference time domain method," *Opt. Lasers Eng.*, vol. 86, pp. 242–247, Jun. 2016.

[23] C. Lukianowicz, "Direct and inverse problem in scatterometry of rough surfaces," *Proc. SPIE*, vol. 4517, pp. 120–125, Aug. 2001.

[24] C. J. Raymond, "Overview of scatterometry applications in high volume silicon manufacturing," in *Proc. AIP Conf.*, 2005, vol. 788. no. 394, pp. 394–402.

[25] C. J. Raymond *et al.*, "Multiparameter grating metrology using optical scatterometry," *J. Vac. Sci. Technol. B*, vol. 15, no. 2, pp. 361–368, Mar. 1997.

[26] S. A. Coulombe, B. K. Minhas, C. J. Raymond, S. S. H. Naqvi, and J. R. McNeill, "Scatterometry measurement of sub-0.1 μm linewidth gratings," *J. Vac. Sci. Technol. B*, vol. 16, no. 1, pp. 80–87, 1998.

[27] E. M. Drege and D. M. Byrne, "Lithographic process monitoring using diffraction measurements," *Proc. SPIE*, vol. 3998, pp. 147–157, Jun. 2000.

[28] E. M. Drege, R. M. Al-Assad, and D. M. Byrne, "Mathematical analyses of inverse scatterometry," *Proc. SPIE*, vol. 4689, pp. 151–162, Jun. 2002.

[29] R. M. Al-Assad and D. M. Byrne, "Error analysis in inverse scatterometry. I. Modeling," *J. Opt. Soc. Amer. A*, vol. 24, no. 2, pp. 326–338, Feb. 2007.

[30] C. J. Raymond, M. E. Littau, A. Chuprin, and S. Ward, "Comparison of solutions to the scatterometry inverse problem," *Proc. SPIE*, vol. 5375, pp. 564–575, May 2004.

[31] S. Wei and L. Li, "Measurement of photoresist grating profiles based on multiwavelength scatterometry and artificial neural network," *Appl. Opt.*, vol. 47, no. 13, pp. 2524–2532, May 2008.

[32] N. Rana, Y. Zhang, D. Wall, B. Dirahoui, and T. C. Bailey, "Machine learning and predictive data analytics enabling metrology and process control in IC fabrication," *Proc. SPIE*, vol. 9424, pp. 94241I-1–94241I-11, Mar. 2015.

[33] H. G. Lee *et al.*, "Virtual overlay metrology for fault detection supported with integrated metrology and machine learning," *Proc. SPIE*, vol. 9424, pp. 94241T-1–94241T-6, Mar. 2015.

[34] J. Zhu, S. Liu, C. Zhang, X. Chen, and Z. Dong, "Classification and recognition of diffraction structures using support vector machine in optical scatterometry," *Proc. SPIE*, vol. 8324, pp. 83242S-1–83242S-11, Apr. 2012.

[35] J. Zhu, S. Liu, C. Zhang, X. Chen, and Z. Dong, "Identification and reconstruction of diffraction structures in optical scatterometry using support vector machine method," *J. Micro/Nanolithogr., MEMS, MOEMS*, vol. 12, no. 1, pp. 0130041-1–0130041-10, Jan. 2013.

[36] K. Ganapathy, V. Vaidehi, and J. B. Chandrasekar, "Optimum steepest descent higher level learning radial basis function network," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 8064–8077, Nov. 2015.

[37] D. Kosic, "Fast clustered radial basis function networks as an adaptive predictive controller," *Neural Netw.*, vol. 63, pp. 79–86, Mar. 2015.

[38] S. Mishra, R. N. Yadav, and R. P. Singh, "Directivity estimations for short dipole antenna arrays using radial basis function neural networks," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 1219–1222, Feb. 2015.

[39] R.-J. Lian, "Adaptive self-organizing fuzzy sliding-mode radial basis-function neural-network controller for robotic systems," *IEEE Trans. Ind. Electron.*, vol. 61, no. 3, pp. 1493–1503, Mar. 2014.

[40] A. Alexandridis, E. Chondrodima, E. Efthimiou, G. Papadakis, F. Vallianatos, and D. Triantis, "Large earthquake occurrence estimation based on radial basis function neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5443–5453, Sep. 2014.

[41] S. F. Su, C. C. Chuang, C. W. Tao, J. T. Jeng, and C. C. Hsiao, "Radial basis function networks with linear interval regression weights for symbolic interval data," *IEEE Trans. Syst., Man, Cybern., Part B Cybern.*, vol. 42, no. 1, pp. 69–80, Feb. 2012.

[42] M. J. L. Orr, "Regularisation in the selection of radial basis function centres," *Neural Comput.*, vol. 7, no. 3, pp. 606–623, May 1995.

[43] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, 2010, doi: 10.1214/09-SS054.

[44] S. S. Haykin, *Neural Networks and Learning Machines*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2009.

[45] T. Hastie, R. Tibshirani, and J. Friedman, *The Element of Statistical Learning*. New York, NY, USA: Springer, 2009.

[46] B. E. A. Saleh and M. C. Teich, *Fundamental of Photonics*. Hoboken, NJ, USA: Wiley, 2007.

[47] M. G. M. M. van Kraaij, *Forward Diffraction Modelling: Analysis and Application to Grating Reconstruction*. Eindhoven, The Netherlands: Eindhoven Univ. of Technology, 2011.

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 with GCC, Taiwan. From 2007 to 2009, he was an Application Development Engineer with ACE, ASML. In 2010, he joined the faculty of the Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include the development of next-generation lithography techniques and optical system integrations.

MUHAMAD FAISAL received the B.E. degree in electrical engineering from the Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, and the M. S. degree in the Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei, Taiwan. His current research interests include the intelligent control system, robust control system, neural networks, and optical metrology based on scatterometry. He received the NTUST Scholarship Award.

SHUN-FENG SU (S'89–M'91–SM'05–F'10) received the B.S. degree from National Taiwan University, Taiwan, and the M.S. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1983, 1989, and 1991, respectively, all in electrical engineering.

He is currently the Chair Professor with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taiwan. He has authored over 190 refereed journal and conference papers in the areas of robotics, intelligent control, fuzzy systems, neural networks, and nonderivative optimization. His current research interests include computational intelligence, machine learning, virtual reality simulation, intelligent transportation systems, smart home, robotics, and intelligent control. He is a CACS fellow.

Dr. Su is very active in various international/domestic professional societies. He is the President of the International Fuzzy Systems Association. He is also in the Boards of Governors of various academic societies, including the IEEE Systems, Man, and Cybernetics Society. He also acted as the Program Chair, the Program Co-Chair, or a PC members for various international and domestic conferences. He currently serves as an Associate editor of the IEEE Transactions on Cybernetics, the IEEE Transactions on Fuzzy Systems, and of the IEEE Access. He serves as a Subject Editor of the *Journal of the Chinese Institute of Engineers*. He is the Editor-in-Chief of the *International Journal of Fuzzy Systems*.

∙ ∙ ∙