Ensemble Learning-Fused Solution to the Inverse Problem in Integrated Optical Critical Dimension Metrology

Chunfu Guo, Jiamin Liu, Honggang Gu, Jinlong Zhu, Chuanwei Zhang, Xiuguo Chen, Hao Jiang, and Shiyuan Liu

***Abstract*— In sub-28-nm nodes, existing optical critical dimension (OCD) metrology tools are challenging to simultaneously and thoroughly meet the metrology requirements of the wafer-level nanostructures in terms of measurement precision, efficiency, and robustness. To tackle the challenge, we propose a Mueller matrix ellipsometry (MME)-based integrated OCD metrology (IM OCD) technique to achieve fast, accurate, and robust in-line metrology of wafer-level nanostructures at advanced nodes. The MME probe is integrated into the integrated circuit (IC) process platform via a polar-coordinate stage, allowing high-throughput measurement of the geometrical parameters of nanostructures with 144 wafers per hour. Meanwhile, we propose an ensemble learning-fused regression (ELFR) approach to improve the solution to the typical inverse problem in nanostructure metrology, effectively addressing the fluctuations in measurement sensitivity and precision caused by the measurement azimuth variations in the IM OCD. The essential of the ELFR approach is the weighted averaging of the respective results from ridge regression (RR) based on the Levenberg–Marquardt algorithm and from bagging-mode neural networks (BMNNs). It improves the metrology results of wafer-level nanostructures to an ultrahigh coefficient of determination better than 0.93, an average deviation, and a maximum deviation less than 1.0 and 2.0 nm, respectively. Characterizations of wafer-level shallow trench isolation (STI) and interconnect-layer nanostructures have demonstrated the feasibility and effectiveness of the proposed MME-based IM OCD method, especially the capability of the ELFR approach to suppress azimuthal-dependent errors.**

Received 9 April 2024; revised 1 October 2024; accepted 29 October 2024. Date of publication 27 November 2024; date of current version 5 December 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52130504, Grant 52305577, and Grant 52205592; in part by the Key Research and Development Program of Hubei Province under Grant 2022BAA013; in part by the Major Program [JianDao (JD)] of Hubei Province under Grant 2023BAA008-2; in part by the Interdisciplinary Research Program of Huazhong University of Science and Technology under Grant 2023JCYJ047; in part by the Innovation Project of Optics Valley Laboratory under Grant OVL2023PY003; in part by the Postdoctoral Fellowship Program (Grade B) of China Postdoctoral Science Foundation under Grant GZB20230244; and in part by the Fellowship from the China Postdoctoral Science Foundation under Grant 2024M750995. The Associate Editor coordinating the review process was Dr. Md. Moinul Hossain. *(Chunfu Guo and Jiamin Liu contributed equally to this work.) (Corresponding authors: Jiamin Liu; Shiyuan Liu.)*

Chunfu Guo, Jiamin Liu, Honggang Gu, Jinlong Zhu, Chuanwei Zhang, Xiuguo Chen, and Hao Jiang are with the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: chunfuguo@hust.edu.cn; jiaminliu@hust.edu.cn).

Shiyuan Liu is with the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the Optics Valley Laboratory, Wuhan 430074, China (e-mail: shyliu@hust.edu.cn).

Digital Object Identifier 10.1109/TIM.2024.3507037

***Index Terms*— Ensemble learning, integrated metrology, inverse problem solving, measurement precision, Mueller matrix ellipsometry (MME), optical critical dimension (OCD) metrology, optical scatterometry.**

I. Introduction

IN INTEGRATED circuit (IC) manufacturing, the size deviations or contour errors of the nanostructures on the patterned wafers can easily lead to final performance deviations, yield degradation, or functional failure of IC devices or chips [1], [2], [3]. Therefore, it is essential in IC manufacturing to accurately monitor the magnitude and consistency of the critical dimensions (CDs) of nanostructures in each die before the subsequent process of the wafers. With the semiconductor industry struggling to manufacture devices with increasingly smaller sizes and complex shapes, optical scatterometry has played an increasingly essential role in the inline metrology of CDs in IC manufacturing due to its outstanding advantages of high sensitivity, nondestructiveness, high efficiency, and ease of integration [4], [5], [6], [7].

Shallow trench isolation (STI) is a critical process in chip manufacturing at sub-250-nm nodes. It enables insulating isolation between active areas in the Si substrate by forming shallow trenches on the Si substrate and filling them with oxide [8], [9], [10]. The geometric features of STI structures [11], [12], [13], such as the trench height, the surface flatness, and the even coating height, are the most critical parameters that should be accurately inline monitored during the chemical mechanical polishing (CMP) process. However, the measurement throughput of existing stand-alone optical CD (OCD) tools based on optical scatterometry can only reach several wafers per hour, which cannot match the productivity of hundreds of wafers per hour for the CMP process. Therefore, it is of great significance to integrate the OCD metrology module into the CMP platform to form the integrated OCD metrology (IM OCD) tool, thereby well matching the metrology efficiency with the manufacturing productivity, for ensuring the full metrology of all STI wafers.

Spectroscopic reflectometry (SR) [14], spectroscopic ellipsometry (SE) [15], and Mueller matrix ellipsometry (MME) are three typical OCD metrology techniques [16], which all have the potential to be integrated into the IC process platform to form the IM OCD tool. The SR-based IM OCD tool has

been widely used to monitor the geometric parameters of typical nanostructure fabricated by CMP or other processes, such as trench height, top CD (TCD), and sidewall angle (SWA) [17]. However, it usually faces the challenge of limited accuracy in characterizing the complex nanostructures contained with multilayer stacks or irregular holes or steps, since it only captures the normal-incidence reflectance of samples. Since SE can measure the polarization change of probing light under an oblique incidence [15], [18], the SE-based IM OCD tool can theoretically achieve higher metrology sensitivity in characterizing complex multilayer stacks. However, SE cannot sufficiently consider the effects of the incidence-plane azimuth variation and the depolarization effects [19], which obstruct the actual integration of this technique into the IM OCD tool. Due to the ability to capture richer polarization scattering information in the form of a 4×4 Mueller matrix [16], [19], the MME typically offers higher measurement sensitivity and adaptability than SR and SE, allowing for more accurate measurement of complex nanostructures, especially the STI and metal interconnection layer nanostructures. This highlights the great potential of integrating the MME into IM-OCD.

However, the polar-coordinate wafer stage has to be utilized in the IM OCD tool due to the seriously restricted installation space, which will lead to different measurement azimuth angles of various pads over the whole wafer [20]. The variations among the azimuth angles cannot be identified by the IM OCD tool [20], [21], which further leads to an azimuth-related modeling error in the forward optical model [21]. Meanwhile, the model-based inverse problem solution invoked in the IM OCD technique usually leads to azimuth-dependent variation of the metrology sensitivity and precision when characterizing complex nanostructures [22], [23], [24], [25], [26]. Also, the first-order error propagation theory and the sensitivity-based configuration optimization theory have demonstrated the distinct dependency of the inverse problem solutions' consistency and accuracy on the azimuthal angle [27], [28], [29]. With IC process nodes further shrinking to the sub-28-nm nodes, the perturbations caused by the azimuth-dependent metrology variations in the measurement repeatability, reproducibility, and stability will be unacceptable. Correspondingly, some studies have tried to use various machine-learning techniques to extract structural parameters directly from the measured spectra [30], [31], [32], [33], which suppressed the measurement inconsistency caused by azimuth variations and improved the stability of the reconstructed parameters to some extent. However, it is still challenging to thoroughly eliminate the negative impact of azimuth variation on the accuracy of structural parameter reconstruction using these methods due to the random errors that always exist in the training set. Therefore, it is of great significance to improve the inverse problem solutions to alleviate significantly the influence of the azimuth-caused errors mentioned above, which ensures the measurement precision of the IM OCD techniques in nanostructure metrology.

To tackle the abovementioned issue, an MME-based IM OCD technique has been self-built to achieve fast, accurate, and robust in-line metrology of wafer-level nanostructures for advanced IC manufacturing. The MME probe is integrated into the IC process platform via an automated polar-coordinate stage, enabling the precise metrology of the wafer-level nanostructures at a high throughput matching the CMP's productivity. An ensemble learning fused regression (ELFR) method has been proposed to improve the solution of the typical inverse problem in the self-built MME-based IM OCD. The essential of the ELFR approach is the weighted averaging of the respective extraction results from ridge regression (RR) [34], [35] based on the Levenberg–Marquardt (LMRR) [36] algorithm and bagging-mode neural networks (BMNNs). It is also a fusion model of multiple machine-learning algorithms, which is the first attempt within our knowledge to improve the metrology precision of the IM OCD technique over the whole wafer's nanostructures.

The ELFR method procedure includes three steps: 1) using the Levenberg–Marquardt (LM) algorithm to extract the initial structural parameters from the measured Mueller ellipsometric spectra for each pad in the wafer, and establishing an RR-based mapping model from the initial structural parameters to the baseline values; 2) achieving five structural parameter sets using the BMNNs with cross-validation function; and 3) determining the final output using the weighted averaging of the respective extraction results of LMRR and BMNN. Besides, the optimal weight set [37] has been searched to ensure the optimal output results and the selectable configuration window.

The main contributions of this study are as follows.

1) To the best of our knowledge, the MME was first introduced into the IM-OCD tool to thoroughly meet the increasing inline metrology requirements for advanced IC manufacturing regarding measurement precision, efficiency, robustness, and adaptability.
2) The ELFR method has been proposed to improve the solution to the typical inverse problem in nanostructure metrology, effectively addressing the fluctuations in measurement sensitivity and precision caused by the measurement azimuth variations in the IM OCD. The proposed method improves the metrology results of wafer-level nanostructures reported by the MME-based IM OCD to an ultrahigh coefficient of determination better than 0.93, an average deviation, and a maximum deviation less than 1.0 and 2.0 nm, respectively.
3) It is feasible to achieve optimal metrology precision and a selectable weight-setting window by optimizing the critical systematic parameters such as the weights.

The rest of the article is organized as follows. The second part introduces the implementation details of the method, the results and discussions are shown in the third part, and the conclusion is summarized in the fourth part.

II. Methodology

A. MME-Based IM OCD Technique

As shown in Fig. 1(a), an MME probe has been integrated into the CMP process platform with a seriously restricted installation space via an automated polar-coordinate wafer stage, forming the first MME-based IM OCD tool to the best of our knowledge. The polar-coordinate wafer can scan every inspected site on the entire wafer using rotation and x-axis linear motions, leading to incident planes with different azimuthal

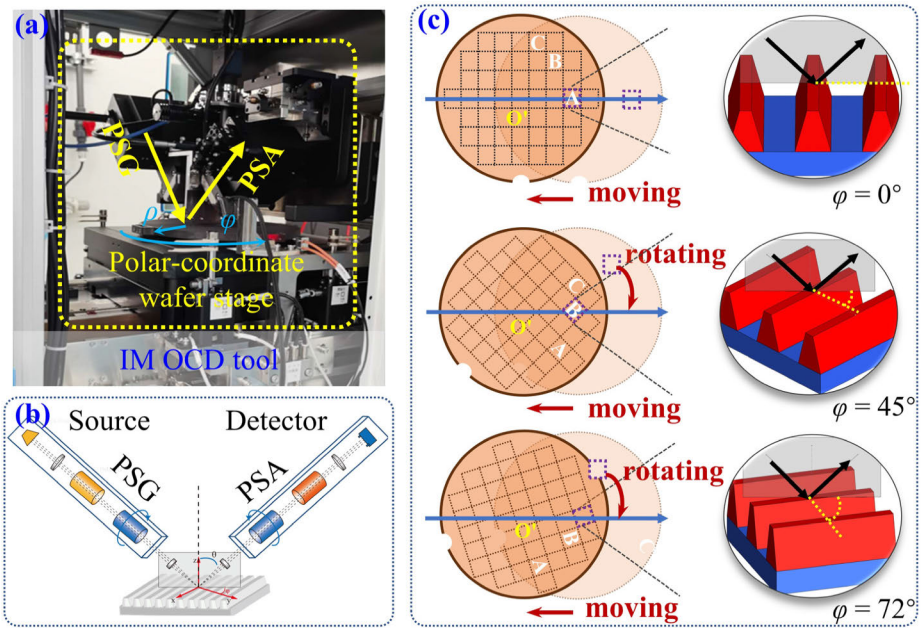

Fig. 1. (a) Schematic of the Mueller matrix-based IM OCD tool. (b) Schematic of the MME probe used in the IM OCD tool. (c) Azimuthal angles to the measurement plane at different probing sites.

angles for each probing site pad over the whole wafer. This integration strategy allows for the high-throughput measurement of wafer-scale nanostructures via the MME-based IM OCD technique at 144 wafers per hour. The MME probe consists of a light source, a polarization state generator (PSG) with a rotating compensator, a polarization state analyzer (PSA) with another rotating compensator, and a detector [38], [39], [40], which can capture the Mueller matrix spectra of every site via the modulation and demodulation of the probe beam's polarization state based on a double-rotating compensator framework. Correspondingly, the systematic principle is shown in Fig. 1(b). The captured Mueller matrix spectra are retrieved from the time-harmonic changes of the polarization state caused by the interaction between the probe beam and the nanostructure [41]. It should be noted that the Mueller matrix spectra measured by the self-built IM OCD tool are manifested as a 4×4 matrix, which will be converted into a k-dimensional vector $\mathbf{Y}_{\text{meas}} = [y_1, y_2, \ldots, y_k]^{\text{T}}$ to facilitate calculation. The dimensional number k satisfies the relation $k = 16 \times n_\lambda$, where n_λ represents the wavelength number. Correspondingly, the structural parameters are expressed as an n-dimensional vector $\mathbf{P} = [x_1, x_2, \ldots, x_n]^{\text{T}}$, which can be extracted by fitting the measured Mueller matrix spectra with the theoretical spectra [28]. The process is essentially a numerical solution to the inverse problem using the Levenberg–Marquardt (LM) nonlinear regression algorithm [16], [22], [23], [42].

Fig. 1(c) presents a schematic of detecting nominally identical STI nanostructures at different locations within the wafer. Since the MME probe is fixed and stationary, each probing site has to be moved into the detection light spot via the rotation and translation operations of the polar-coordinate wafer stage. Correspondingly, as shown in Fig. 1(c), nominally identical nanostructures "A," "B," and "C" would correspond to different incidence azimuthal angles, which cannot be accurately identified by the IM OCD tool. That is the origin of the azimuth-dependent error in the MME-based IM OCD tool. In order to ensure the measurement consistency of the instrument itself, the azimuthal angle-related error urgently needs to be corrected and suppressed.

Without loss of generality, the Mueller matrix spectra captured by the IM OCD tool usually contain systematic and random errors [42], leading to absolute bias and metrology consistency bias in the extraction parameters. Herein, the systematic error can be understood as the model error caused by the variation in azimuth angle, while the random error can be understood as the random noise in the light intensity. According to the first-order error propagation model [28], [29], [42], the impact of systematic errors $\boldsymbol{\mu}_{\text{M}}$ and random errors $\boldsymbol{\sigma}_{\text{M}}$ in the Mueller matrix spectra on extraction parameters would change with the azimuth angle of the incidence plane, which can be expressed as the following formula [29]:

$$\mathbf{J}_{\mathbf{P}}^{\text{T}}\mathbf{W}\mathbf{J}_{\mathbf{P}}\Delta\mathbf{P} = \mathbf{J}_{\mathbf{P}}^{\text{T}}\mathbf{W}\Delta\mathbf{Y} = \mathbf{J}_{\mathbf{P}}^{\text{T}}\mathbf{W}(\boldsymbol{\mu}_{\mathbf{M}} + \boldsymbol{\sigma}_{\mathbf{M}}) \tag{1.a}$$

$$[\mathbf{J}_{\mathbf{P}}]_{ij} = \frac{\partial \boldsymbol{f}_i(\mathbf{P}, \mathbf{a})}{\partial \boldsymbol{x}_j} \tag{1.b}$$

where $\mathbf{J}_{\mathbf{P}}$ represents the Jacobian matrix, in which each element is the partial derivation of the forward optical model $\mathbf{f}(\mathbf{P}, \mathbf{a})$ with respect to $\mathbf{P}$. Vectors $\Delta\mathbf{P}$ and $\Delta\mathbf{Y} = \boldsymbol{\mu}_{\mathbf{M}} + \boldsymbol{\sigma}_{\mathbf{M}}$ are the errors in the extracted parameters and the measured Mueller matrix, respectively. Vector $\mathbf{a} = [a_1, a_2, \ldots, a_m]^{\text{T}}$ represents the measurement configuration, which contains the azimuthal angle of the measurement plane. Diagonal matrix $\mathbf{W}$ consists of a series of weighing factors w_i, each of which is the variance of the measured Mueller matrix element at the ith wavelength.

Considering that other configuration parameters, except the azimuth angle of the incidence plane, will not change in actual

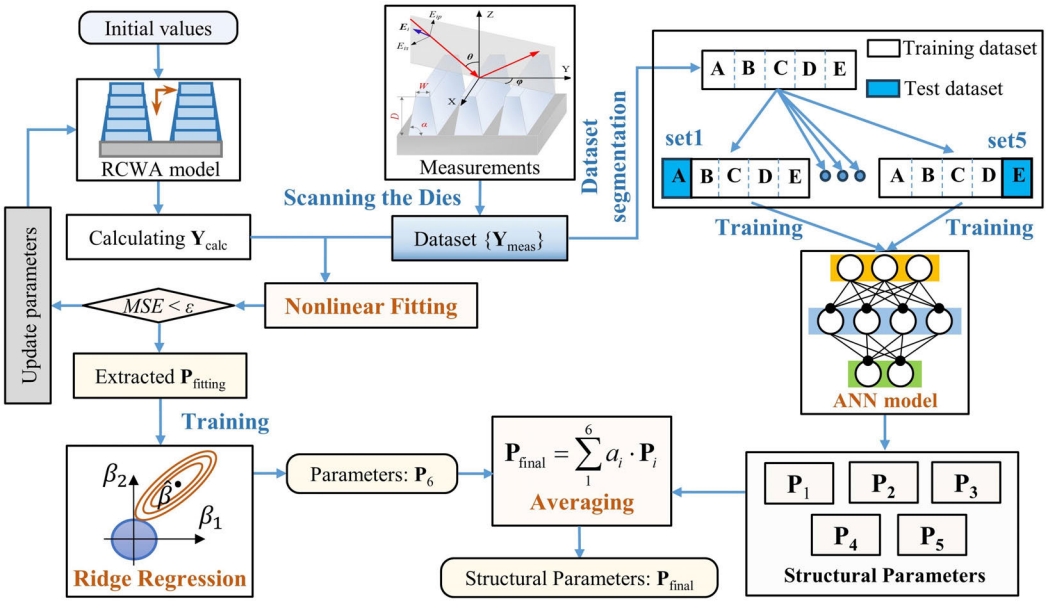

Fig. 2. Flow of ensemble learning-fused solution method for extracting the nanostructure parameters.

measurements, it can be easy to find that the Jacobian matrix $\mathbf{J_P}$ will show apparent dependence on the azimuth angle. Thus, the error propagation coefficient $(\mathbf{J}_\mathbf{P}^\mathrm{T}\mathbf{W}\mathbf{J}_\mathbf{P})^{-1}\mathbf{J}_\mathbf{P}^\mathrm{T}\mathbf{W}$ will also exhibit significant distinction at different azimuth angles, which serves as the theoretical basis for the azimuth-dependent metrology error in the nanostructure metrology using self-built IM OCD tool.

B. Ensemble Learning-Fused Regression Method for Parameter Extraction

In conventional IM OCD technique-based characterization of nanostructures, the parameter extracted by the LM algorithm is usually regarded as the final result. However, the LM algorithm often produces inaccurate results in the IM OCD due to the local optimal searching strategy and the strong dependence of extraction results on the azimuth angle. Herein, an ELFR method is proposed to improve the solution of the typical inverse problem in the self-built MME-based IM OCD. The essential of the ELFR approach is the weighted averaging of the respective extraction results from the LMRR and BMNN algorithms. The corresponding algorithm principle is shown in Fig. 2, and the detailed description is as follows.

First, as shown in the left part of Fig. 2, the OCD structural parameters are solved using the LM nonlinear regression algorithm, and these parameters are then used as the dataset $\mathbf{P}_{fitting}$ to train the mapping model to the baseline dataset $\mathbf{P}_{bsl}$ using the RR method.

This RR method-based mapping model, combined with the LM nonlinear regression algorithm, jointly generates structural parameters from the measured Mueller matrix spectra, denoted as $\mathbf{P}_6$. Unlike previous studies that map spectra to results [21], this method directly maps results to results and is referred to as the RR based on the LMRR. The detailed description of the RR method can be found in [35] and [36].

Second, as shown in the right part of Fig. 2, the Mueller matrix spectra are randomly divided into five datasets, and the BMNN can train five different mapping models through both the dataset segmentation and the cross-validation sampling, which will produce five structural parameter sets $\mathbf{P}_i$ with $i = 1$–5. In this process, the activation function used is the Sigmoid function, and the corresponding results are denoted as $\mathbf{P}_1$ to $\mathbf{P}_5$. The advantage of the BMNN-based models lies in their ability to effectively integrate multisource data and automatically learn complex features [43], thereby enhancing the expressive capability and robustness of the five models.

The BMNN based on the multilayer perceptron strategy is first constructed to realize the parameter extraction, in which the mapping relationship $\mathbf{P} = \mathbf{f}^{-1}(\mathbf{Y}_{meas})$ from the measured Mueller matrix spectra Y_{meas} to the structural parameter $\mathbf{P}$ will be learned. In this multilayer perceptron neural network, each component layer uses linear equations to calculate the output, accompanied by using the sigmoid activation function to quantify the nonlinear changes between adjacent layers. Thus, the layer output function O can be estimated by the input $\mathbf{I}$, the neural weight $\mathbf{w}$, the bias $\mathbf{b}$, and the sigmoid activation function $\sigma(x)$

$$o = \sigma(x) \cdot [\mathbf{w} \cdot \mathbf{I} + \mathbf{b}] = \frac{1}{1 + e^{-x}} \cdot [\mathbf{w} \cdot \mathbf{I} + \mathbf{b}]. \tag{2}$$

Then, the weighted averaging of the structural parameter set $\mathbf{P}_6$ reported from the LMRR algorithm and the parameter sets $\mathbf{P}_1$–$\mathbf{P}_5$ determined by the BMNN algorithm will output the ultimate results. Assigning a weight w_1 to the structural parameter set $\mathbf{P}_6$ and a weight w_2 to the structural parameter set $\mathbf{P}_1$–$\mathbf{P}_5$, the ultimate structural parameter can be obtained

$$\mathbf{P}_\mathrm{u} = w_1\mathbf{P}_6 + w_2 \sum_1^5 \mathbf{P}_i \tag{3.a}$$

$$w_1 + 5w_2 = 1. \tag{3.b}$$

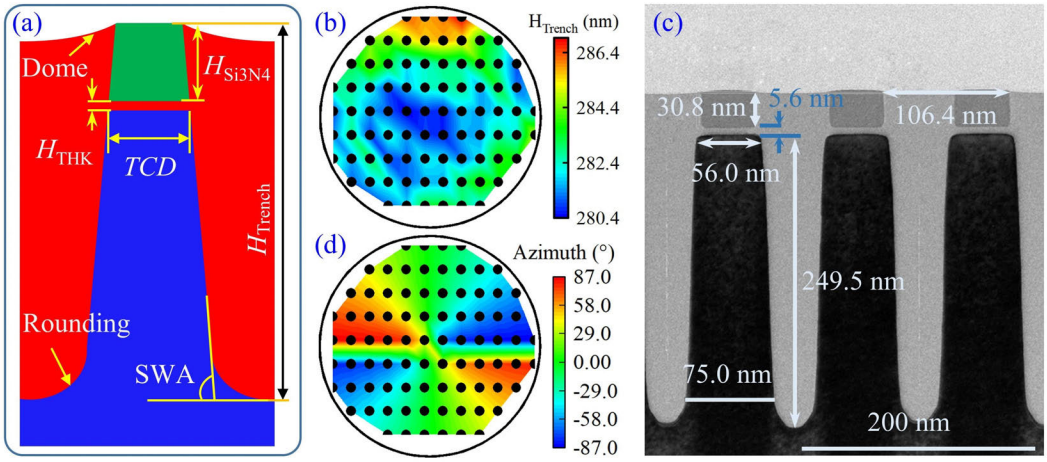

Fig. 3. Basic information about STI samples. (a) Schematic of STI nanostructure specimen. (b) Height distribution of the validation test specimen reported by the OCD metrology golden tool. (c) Cross-sectional TEM view of the STI nanostructure. (d) Azimuthal angle distribution in the wafer reported by the IM OCD tool.

Further optimizing the weights w_1 and w_2 under the constraint can improve the accuracy of $\mathbf{P}_u$, which will be explained in the subsequent discussion. The specific steps of weight optimization are as follows.

1) Traversing w_1 in the range of [0, 1] with an interval of 0.05. Correspondingly, w_2 can be calculated through (3.b).
2) Calculating the covariance R^2 and the absolute deviation ε_{bias} between the metrology and benchmark results under every weight set (w_1, w_2). The covariance R^2 is also referred to as the coefficient of determination, which not only characterizes the overall error between the measurement result $\mathbf{P}_u$ and the benchmark result $\mathbf{P}_{bsl}$ but also quantifies the measurement consistency

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left[p_u(\mathbf{r}_i) - p_{bsl}(\mathbf{r}_i)\right]^2}{\sum_{i=1}^{n}\left[p_u(\mathbf{r}_i) - \sum_{i=1}^{n} p_u(\mathbf{r}_i)/n\right]^2} \quad (4)$$

where p_u symbolizes any component element in $\mathbf{P}_u$, while p_{bsl} represents any component element in $\mathbf{P}_{bsl}$. Position vector $\mathbf{r}_i$ with $i = 1$–n scans measurement sites over the entire wafer. The maximum value of R^2 can reach one, and the larger the value of R^2, the smaller the overall error. The deviation ε_{bias} characterizes the absolute error between the measurement and the benchmark results, implying that the smaller the deviation ε_{bias}, the higher the accuracy of the measurement results

$$\boldsymbol{\varepsilon}_{bias} = \mathbf{P}_u - \mathbf{P}_{bsl}. \quad (5)$$

3) Using $R^2 \geq 0.93$, $\mathbf{mean}\{|\varepsilon_{bias}|\} < 1$ nm, and $\mathbf{max}\{|\varepsilon_{bias}|\} < 2$ nm as criteria for judgment, and select the optimal range of w_1.

Furthermore, the proposed method will train these two LMRR- and BMNN-based models using the Mueller matrix spectra and the corresponding structural parameter sets from standard wafer specimens. Subsequently, the metrology results obtained from validation test wafers will be employed to validate the effectiveness of the proposed method.

TABLE I
DETAILED MEASUREMENT CONFIGURATIONS

	Para.	Nominal values	Floating or fixing	Ranges
Main Meas. Conf.	TCD	56 nm	Floating	30~80 nm
	H_{Si3N4}	31 nm	Floating	10~50 nm
	H_{Trench}	282 nm	Floating	260~300 nm
	Dome	5 nm	Fixing	3~8 nm
	H_{THK}	5.6 nm	Fixing	3~8 nm
	Rounding	5 nm	Fixing	3~8 nm
	SWA	87°	Floating	83~95°
Other Conf.	φ		Floating	-90~90°
	θ	44.9°	Fixing	44.7~45.3°
	λ	—	—	340-800 nm

III. PROTOTYPE IMPLEMENTATION

The parameter extraction results of the ELFR method have been compared to those reported by the LM nonlinear regression, the LMRR method, and the BMNN method in the manner of the ablation study, highlighting the improved effect of the proposed method on the inverse extraction of the structural parameter. In the measurement experiment, a typical 1-D periodic multilayer grating reflecting the STI nanostructure feature was used as the measured specimen, which consisted of a Si grating layer, a SiO_2 film layer, and a Si_3N_4 grating layer, as shown in Fig. 3(a). In the metrology experiment, only the TCD, the Si_3N_4 grating height (H_{Si3N4}), the trench height (H_{Trench}), and the SWA were set as floating parameters, which were the key parameters of the STI structure. Other structural parameters, such as SiO_2 film thickness (H_{THK}), are fixed as constants due to their minimal impact on the measured Mueller matrix spectra.

The detailed configurations for characterizing the wafer specimens are shown in Table I. These simplified treatments can significantly reduce the complexity and ill-posedness of solving inverse problems [44], ensuring the effectiveness and expedience of structural parameter extraction. Moreover, these

simplifying are even more in line with the requirement of monitoring the STI process, which primarily focuses on trench height H_{Trench}.

It needs to be emphasized again that the MME probe (ME-L, Wuhan Eoptics Technology Company, China) of the IM OCD tool is responsible for capturing the Mueller matrix spectra covering a wavelength range of 210–800 nm, and the polar-coordinate wafer stage realizes the scanning of each measurement site over the entire wafer. Four 12-in wafers with 83 test sites were used as experimental measurement specimens, three of which served as standard specimens to train the extracted model and analysis method, and the remaining of which served as the validation test specimen to verify the proposed method. The trench height distribution of the validation test specimen was characterized using the OCD metrology golden tool (SpectralShapeTM 10k, KLA Corporation, USA) [45], whose measurement results have been extensively validated by global users through high-resolution scanning electron microscope (SEM) or transmission electron microscope (TEM) cross-sectional observations [45], [46], [47]. The corresponding results are shown in Fig. 3(b). From the contour map presented in Fig. 3(b), it can be easily found that the trench height H_{Trench} of 83 probing sites on the validation test specimen is distributed between 280 and 287 nm, mainly concentrated around 282.55 nm. The trench height H_{Trench} of several sites in the validation test specimen has been randomly verified using cross-sectional TEM observation (Tecnai G2 F30 S-TWIN, FEI Company, Netherlands), in which the representative characterization result at the sixth site in the second row on the specimen is presented in Fig. 3(c). It can be easily noticed that the trench height at this site is 285.9 nm, which is highly consistent with the observation results of 285.7 nm, as shown in Fig. 3(b). Given the OCD metrology golden tool's de facto benchmark capability and the high consistency demonstrated by our sampling testing based on the cross-sectional TEM viewing, the metrology results shown in Fig. 3(b) would serve as a benchmark to verify the feasibility and accuracy of the proposed method. Meanwhile, the geometrical profile of the STI nanostructure presented by cross-sectional TEM viewing also confirms the rationality of the geometric description of the nanostructure shown in Fig. 3(a) for the forward optical modeling [48]. Fig. 3(d) shows the azimuthal angle distribution of the validation test specimen, namely the angle between the incidence plane and the grating periodic orientation at each site, which is determined by the notch sensor attached to the self-built IM OCD tool. The azimuthal angles of each measurement site are dispersed in the range of $-90°$ to $90°$, which will make the measurement precision of the IM OCD tool strong azimuth dependence. Moreover, the azimuthal angle distribution will introduce apparent errors in the measurement consistency of the trench height H_{Trench}.

This study explicitly verifies the metrology accuracy of the proposed method in characterizing H_{Trench}, which is one of the most critical parameters of the CMP-processed STI structures. The primary experimental goal is to improve the coefficient of determination R^2 between the prediction result of the test set and the target result to make it better than 0.93, which is set by introducing an additional tolerance threshold into the conventional requirement of $R^2 = 0.90$ used for IC inline metrology. The second goal is to ensure the small magnitude of mean and maximum absolute deviations, requiring them to be less than 1 and 2 nm, respectively. Emphasizing again, by drawing on the ablation study strategy widely used in machine learning and comparing the coefficient of determination R^2, the mean and maximum values of $\varepsilon_{\text{bias}}$ of the results obtained by the ELFR, the LM, the RR, and the BMNN, it can be more objective to evaluate and reveal the mechanism of the ELFR method in improving measurement results.

IV. Experiments and Results

A. Influence of Measurement Azimuth Angle on Reconstruction Precision by MME

A typical simulation experiment for the metrology process based on the IM OCD tool also proved the influence of the measurement azimuth angle on the extracted trench height H_{Trench}. The simulation flow is shown in Fig. 4(a), in which both the forward optical model and the ellipsometry system model are derived from previous research work [19], [23], [28], [48]. Given the nanostructure and its geometrical parameters shown in Fig. 3(a), the theoretical Mueller matrix can be calculated using the forward optical model under different measurement configurations dominated by the azimuth angle. The essence of this process is the simulation of the interaction between the probe beam and the nanostructure. Subsequently, with the given systematic parameters of the PSG and PSA, different Hadamard components can be calculated using the ellipsometry system model [48], [49], which essentially simulates the MME probe's measurement function. By introducing a system error of 1% into a typical system parameter of the PSA, such as the phase retardance, the ellipsometry system model can be used to generate the "measured" Mueller matrix for the given nanostructure. Then, the LM nonlinear regression algorithm is used to solve the corresponding inverse problem to obtain the "measured" geometrical parameters, which will be compared with the previously input parameters to evaluate the "measurement bias." At this time, the measurement bias distribution under different azimuth angles will intuitively reveal the tool's artifacts caused by the azimuth angle in the measurement consistency of the IM OCD tool, which is what the proposed method strives to suppress or eliminate.

In the simulation experiment, the trench heights were set to 60 random values in the range of [273, 286 nm], and the azimuth angles were set to 60 random values at an interval of $[-90°, 90°]$. In contrast, other input parameters were assigned to the fixed values, as shown in Table I. The trench heights and azimuth angles were combined in pairs to form 60 sets of input conditions for the simulation flow shown in Fig. 4(a), which would output 60 "measured" trench heights. The correlation between the output values H_{out} and the input values H_{in} of the trench height is exhibited in Fig. 4(b). The linear regression result $y = 0.756x + 66.87$, accompanied by a coefficient of determination $R^2 = 0.142$, indicates fragile consistency between the input and output values, which can be attributed to the azimuth angle-dependent measurement error. The azimuth angle-dependent measurement biases with large magnitudes shown in Fig. 4(c) also confirm the above

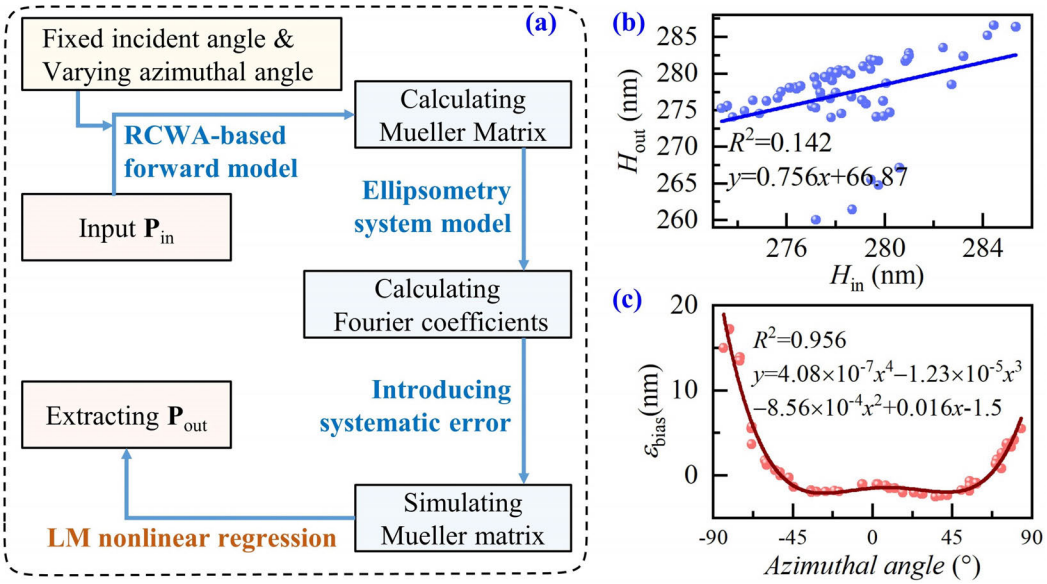

Fig. 4. Simulation flow and corresponding results for the IM OCD-based metrology experiment. (a) Simulation flow. (b) Consistency estimation of input and extracted H_{Trench}. (c) Azimuthal angle dependency of absolute bias $\varepsilon_{\text{bias}}$.

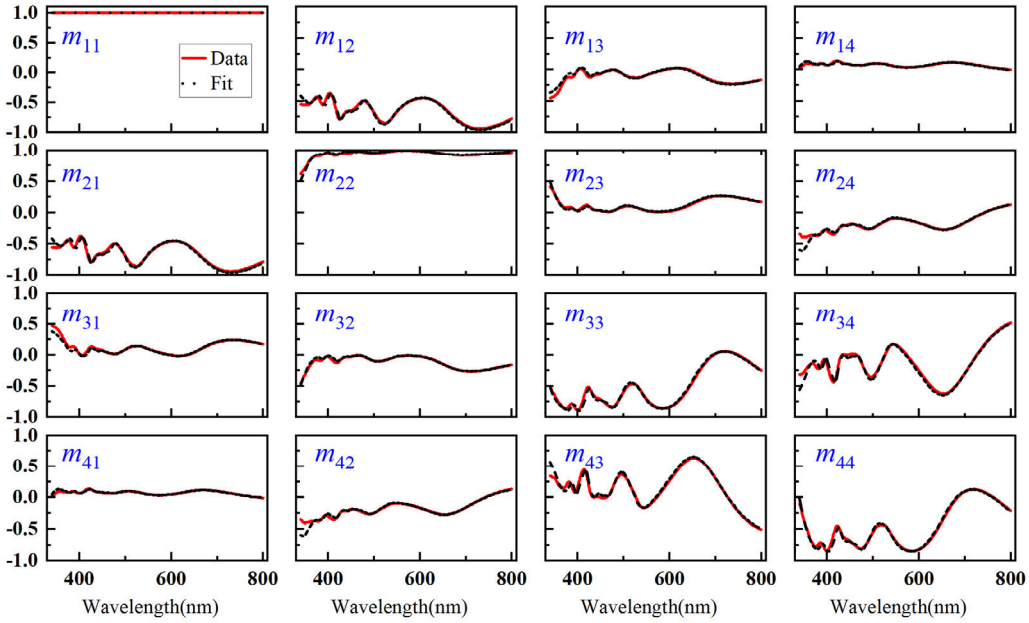

Fig. 5. Illustration of the fitting results of calculated and measured Mueller matrix spectra.

inference, in which the coefficient of determination $R^2 = 0.956$ shows extremely high credibility for the inference. The above observation suggests that the IM OCD tools will exhibit significant artifacts in the measurement consistency when measuring nanostructures at different sites, which cannot be thoroughly eliminated using measurement configuration optimization. Besides, it can be noticed that large deviation $\varepsilon_{\text{bias}}$ mainly appears at the azimuth angles of $-90°$ and $90°$ because of the sharply decreased metrology sensitivity at the two azimuthal angles. Moreover, the off-diagonal elements in the Mueller matrix also gradually disappear at the two azimuthal angles, which will undoubtedly reduce the effective scattering signal. Similar phenomena will also be observed in the actual metrology results of the validation test specimen, which will be explained in Sections IV-B and IV-F.

B. LM Nonlinear Regression Results

It is straightforward to determine the trench height H_{Trench} using the LM nonlinear regression method to fit the measured Mueller matrix spectra of the validation test specimen. The corresponding fitting result is shown in Fig. 5, with the trench height H_{Trench} determined as 283.55 nm. It is easy to notice that the match between the calculated and experimental spectra is satisfied, demonstrating the reasonability of the extracted structural parameters and the forward optical model describing the STI nanostructure.

Similar to the fitting results shown in Fig. 5, the Mueller matrix spectra measured at the other 82 probing sites can also be well-fit. Fig. 6(a) exhibits the mean squared error (MSE) values for all 83 probing sites on the validation test specimen clearly, showing a range of 21–32 and a mean value of 26.92.

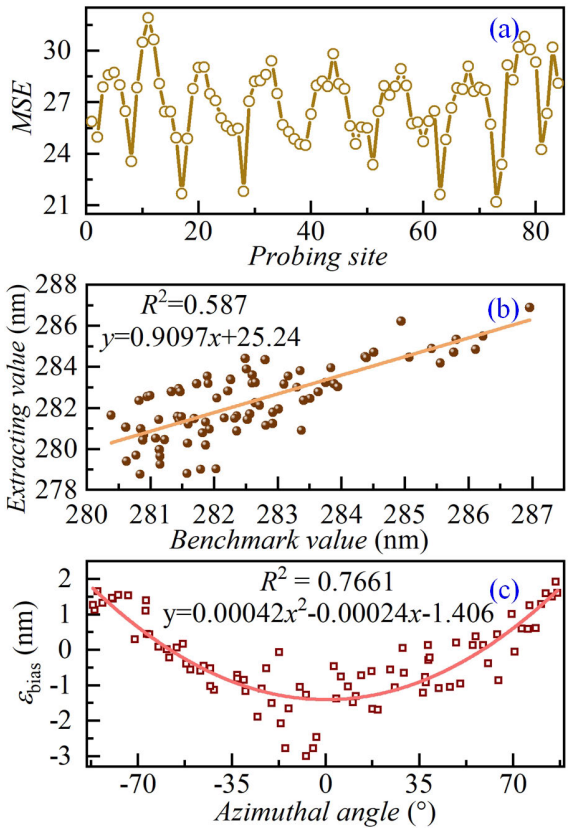

Fig. 6. Extraction results of LM nonlinear regression analysis. (a) MSE values for all 83 probing sites reported by the LM nonlinear regression analysis. (b) Comparison between the extraction and benchmark results. (c) Azimuth dependency of the absolute deviations ε_{bias}.

With the benchmark data reported by the OCD golden tool shown in Fig. 3(b) and the trench height H_{Trench} determined by the LM nonlinear regression used as the independent and dependent variables, respectively, the correlation between these two variables could be checked thoroughly. It would facilitate for revealing the measurement accuracy of the LM regression analysis based on the MME-based IM OCD tool.

Correspondingly, the correlation analysis results are shown in Fig. 6(b), in which their linear regression analysis result has also been presented. The dispersed distribution of discrete points and the coefficient of determination $R^2 = 0.587$ simultaneously reveal the weak linear correlation between the independent and the dependent variables, indicating insufficient accuracy of the trench depth extracted by LM nonlinear regression. Moreover, the linear regression result $y = 0.9097x + 25.24$ deviates from the straight-line equation $y = x$, especially involving the slopes of these two linear equations, which intuitively shows non-negligible deviations between the extracted results and the benchmark results. Ideally, the correlation between the extracted value and the baseline value should be approximately described by $y = x$ or $y = x + c$, in which the slope value of 1 accompanied by a calibratable constant intercept c corresponds to satisfactory consistency. This rule of thumb has been widely adopted for OCD inline metrology in the IC production lines. The linear equation $y = 0.9097x + 25.24$ suggests that the LM nonlinear regression results have not yet reached the OCD online measurement accuracy requirements of the CMP process segment used for the STI nanostructure.

Furthermore, the measurement deviations ε_{bias} between the extraction and benchmark results at each probing site, as a function of the measurement azimuth angle, are shown in Fig. 6(c). All deviations ε_{bias} are dispersedly distributed in the range of -3 to 2 nm, among which the overall mean and maximum values are 0.970 and 3.005 nm, respectively, as shown in Table II. This result shows that the maximum absolute error exceeds the preset threshold for ensuring the accurate metrology of the trench height. Meanwhile, Fig. 6(c) also shows the parabolic function fitting results of the deviation changing with the azimuthal angle, where the parabolic equation is $y = 0.00042x^2 - 0.00024x - 1.406$. The corresponding coefficient of determination $R^2 = 0.7661$ indicates that the parabolic dependence of the absolute deviation ε_{bias} on the azimuthal angle is reasonable to a certain extent. Namely, the parameter extraction error of the LM nonlinear regression varies with the measurement azimuth angle. Similar to the simulation result shown in Fig. 4(c), the absolute deviation ε_{bias} reaches the maximum values at the azimuthal angles of $\pm 90°$ and $0°$. The deviation ε_{bias} reaches a minimum value close to 0, as the measurement azimuth angle is $\pm 58°$. Thus, it can be inferred that the sharply decreased metrology sensitivity at the azimuthal angles of $\pm 90°$ and $0°$, combined with the gradual disappearance of the off-diagonal elements in the Mueller matrix, jointly explains why the LM nonlinear regression would introduce artifact errors in the measurement consistency of the trench depth.

C. LMRR Method Results

Fig. 7(a) exhibits the trench height extracted by the LMRR method, in which each discrete point implies the consistency between the extracted and benchmark results. Most points are distributed near the empirical equation $y = 0.9853x - 4.70$ produced by the linear regression, which presents high consistency between the extracted and benchmark values of trench depth. Both the slope of 0.9853 and the corresponding coefficient of determination $R^2 = 0.9178$ are better than the values determined by the LM nonlinear regression analysis, indicating significant improvement in the measurement precision. The absolute deviation ε_{bias} shown in Fig. 7(b) are primarily concentrated in the range of 0.5418 ± 0.4410, where 0.5418 and 0.4410 represent the mean value and standard deviation of the overall absolute deviations, respectively, as shown in Table II. It can also be noted that the maximum absolute error is only 1.29 nm, which is approximately 57.07% smaller than the corresponding value determined by LM nonlinear regression analysis.

Moreover, the deviation no longer exhibits apparent dependence on the measurement azimuthal angle, which shows that the RR method can effectively eliminate the artifact error in the measurement consistency of the trench height mentioned above. However, the coefficient of determination $R^2 = 0.9178$ has not yet reached the preset threshold $R^2 = 0.93$, which means that the LMRR method needs to be further improved to improve the selectivity of the measurement configuration for the IM OCD tool. Compared with the results

TABLE II
ABLATION STUDY OF EXTRACTION RESULTS FROM THE ELFR, THE LM, THE LMRR, AND THE BMNN ALGORITHMS

Mapping Model	R^2	Mean absolute deviation (nm)	Maximum absolute deviation (nm)
LM	0.578	0.970	3.006
LMRR	0.918	0.542	1.289
BMNN5	0.878	0.802	2.615
LMRR + BMNN5	0.937	0.631	1.592
LMRR + BMNN1	0.915	0.695	2.105
LMRR + BMNN2	0.922	0.713	1.698
LMRR + BMNN3	0.926	0.673	1.512
LMRR + BMNN4	0.931	0.661	1.459
LMRR + BMNN6	0.933	0.660	1.487
RR	0.891	0.761	2.116
LM + BMNN5	0.917	0.610	1.287
RR + BMNN5	0.878	0.804	2.613
Neural Network 1	0.891	0.795	2.514
Neural Network 2	0.876	0.835	2.843
Neural Network 3	0.861	0.816	2.355
Neural Network 4	0.877	0.807	2.626
Neural Network 5	0.871	0.824	2.750

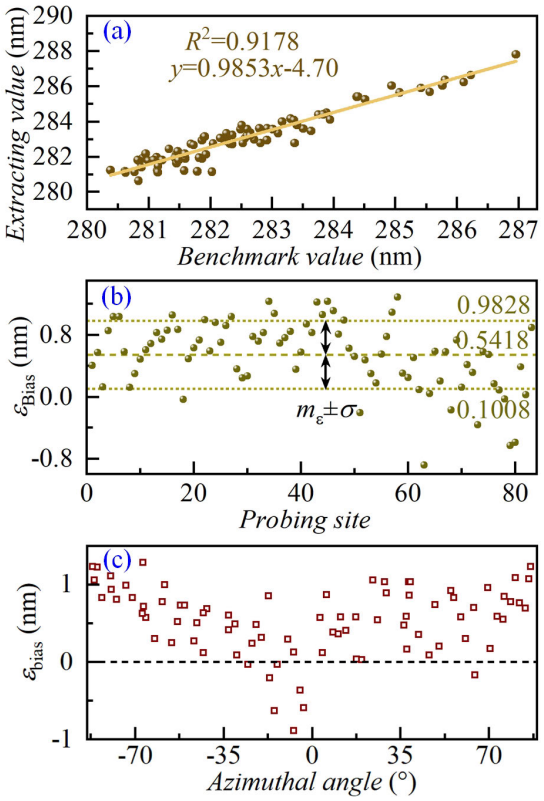

Fig. 7. Extraction results of the LMRR method. (a) Comparison between the extraction and benchmark results. (b) Deviation distribution. (c) Azimuth dependency of the absolute deviations ε_{bias}.

shown in Fig. 6(c), Fig. 7(c) indicates that the LMRR method narrows down the range of deviation ε_{bias}. Besides, at azimuth angles of ±90° and 0°, slightly larger deviation ε_{bias} close to ±1 can be still noticeable.

D. BMNN Method Results

The measured Mueller matrix spectral data of all probing points on the above three standard wafers were randomly divided into five subsets, four of which were used for training the neural network, and the remaining of which was used for testing the neural network. Such a post-processing strategy will produce five neural network models for cross-validation with each other. By applying these five neural network models to the extraction of trench height H_{Trench} of the validation test specimen, the corresponding analysis results can be obtained, as shown in Table II. According to the ablation study results shown in Table II, the coefficient of determination R^2 determined by each cross-validation neural network model is mostly around 0.86–0.89, and the maximum absolute deviation is always greater than 2 nm, which indicates that a single neural network-based extraction method cannot ensure the desired metrology accuracy of H_{Trench}. Besides, the five coefficients of determinations R^2 are highly consistent, as are the five maximum absolute deviations and the five average absolute deviations. This reveals that the cross-validated neural network is highly immune to random noise in the measured spectra.

Fig. 8 illustrates the deviation distribution with azimuthal angle obtained by the BMNN method across five different datasets. The results suggest that the consistency and accuracy of the extracted parameters are not significantly affected by the dataset selection. Furthermore, a comparison with the results presented in Fig. 6(c) reveals significant differences in the azimuthal distribution of the deviation ε_{bias} determined by the BMNN and LMRR methods. Notably, when the azimuthal angles are around ±90° and 0°, the deviation range (−1.0, 2.5 nm) determined by the BMNN method is smaller than that range (−3, 2 nm) determined by the LM algorithm, but is larger than the range (−1.0, 1.0 nm) determined by the

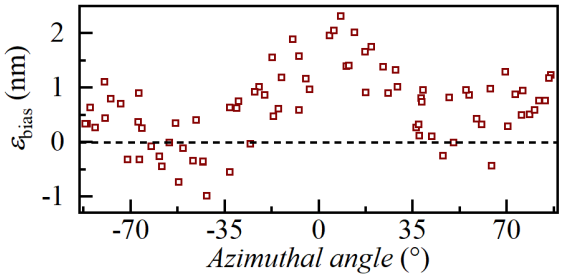

Fig. 8. Azimuthal dependent deviation ε_{bias} determined by the BMNN method.

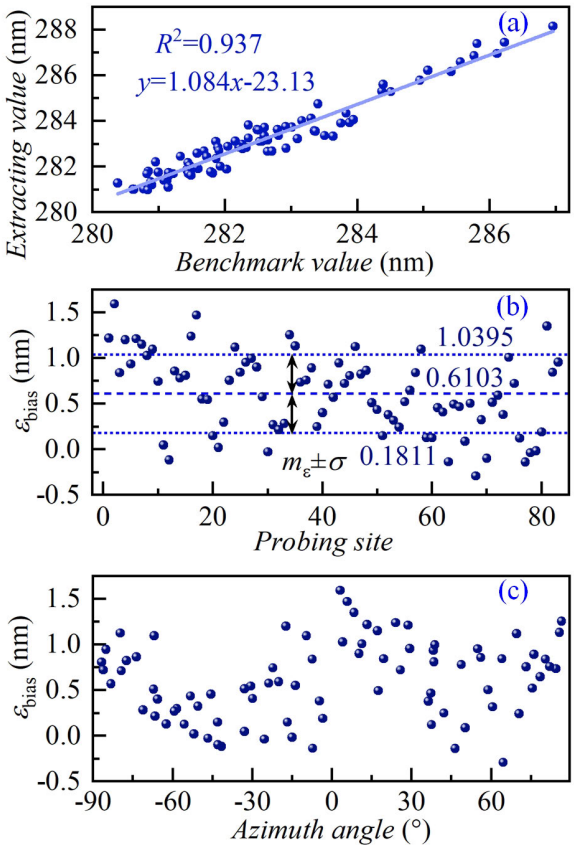

Fig. 9. Extraction results of ensemble learning-fused nonlinear regression. (a) Consistency comparison, (b) bias distribution ε_{bias}, and (c) azimuth dependency of deviation ε_{bias}.

LMRR method. Particularly near the azimuth angle of 0°, the absolute deviations determined by the LMRR and BMNN methods exhibit an opposite sign. These observations provide a rationale for optimizing the weighted average algorithm.

E. ELFR Method Results

Introducing a weighted summation strategy, that is, a weight factor $w_1 = 0.5$ is assigned to the extraction result of the LMRR method, and a weight factor $w_2 = 0.1$ is assigned to each of the five results from the BMNNs, the final extraction value of the trench height H_{Trench} can be determined. This is the extraction result determined by the proposed ELFR method, which will have the advantages of both the LMRR method and the BMNN model, such as high accuracy and robustness. The corresponding coefficient of determination R^2 and the bias distribution ε_{bias} are shown in Fig. 9(a) and (b), respectively. The R^2, **mean**$\{|\varepsilon_{bias}|\}$, and **max**$\{|\varepsilon_{bias}|\}$ are 0.937, 0.631 nm, and 1.592 nm, respectively, all of which satisfy the target requirements for the metrology of trench height of the STI nanostructure. It is easy to notice that the coefficient of determination and absolute deviations are significantly better than the results extracted by both LM nonlinear regression analysis (labeled as "LM" in Table II) and single neural network model (labeled as "Neural Network 1" to "Neural Network 5" in Table II). Compared with the LMRR analysis results (labeled as "LMRR" in Table II), the coefficient of determination R^2 and maximum absolute deviation of the trench height extracted by the ELFR method are significantly better than the results determined by the former, while the mean absolute deviation is roughly equivalent to the results determined by the former. These comparative results reveal that the ELFR method has more excellent parameter extraction capabilities than the LMRR and BMNN analysis. Moreover, by adapting the number of cross-validated neural networks in the BMNN algorithm, we can obtain other types of ELFR methods via the combination of the LMRR and the adapted BMNN algorithms. The corresponding ablation experimental results are presented in Table II. The proposed ELFR method (labeled as "LMRR + BMNN5" in Table II) has the highest coefficient of determination $R^2 = 0.937$, which implies that the optimal number of cross-validated neural networks is five. This might be attributed to the limited-sized dataset.

Furthermore, the self-consistency check of the ELFR method has been carried out. The trench heights of the 83 probing sites on the validation test specimen extracted by the proposed method are all within the range of [280.98, 288.16 nm], and most of them are distributed near the fitting straight-line function $y = 1.084x - 23.13$. The average trench height is 283.16 nm, highly consistent with the average value of the given benchmark results with a relative deviation of 0.22%. Moreover, the absolute deviation between the average height and the reference value of 285.9 nm determined by the TEM observation is only 2.74 nm, meaning that the relative deviation is less than 1%. This result also indicates the accuracy of the parameter extraction of the proposed ELFR method. The sites with slight significant differences between the extraction and the benchmark results are mainly distributed at the edge of the validation test wafer, possibly due to the low positioning accuracy of the R-theta stage at the border.

Fig. 9(c) shows the distribution of the deviation ε_{bias} of the trench height at 83 probing sites as the function of the measurement azimuth angle. The dispersed distribution of ε_{bias} shown in Fig. 9(c) indicates no clear correlation between ε_{bias} and the azimuth angles. The result suggests that the ELFR method can effectively eliminate the artifacts caused by the azimuthal angle distribution in the extracted parameters, which ensures the performance of the MME-based IM OCD tool regarding measurement consistency.

Meanwhile, the trench heights extracted by the ELFR method can be input into the forward optical model of the nanostructure at each site, accompanied by floating the other four parameters listed in Table I, which leads to the matching result between the calculated and measured Mueller matrix

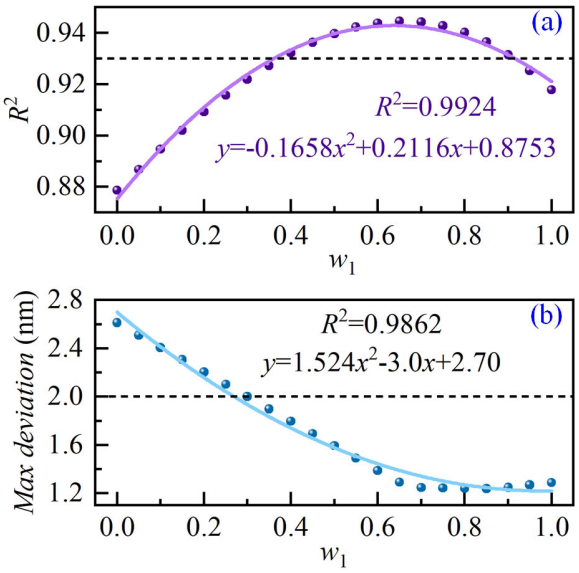

Fig. 10. (a) Coefficient of determination R^2 and (b) maximum absolute deviation determined by the ELFR method under different weight vector $[w_1, w_2]^T$ of STI sample.

spectra. The matching degree between the measured and calculated spectra of each element in the Mueller matrix within the entire wavelength range of [320, 800 nm] is satisfactory, which undoubtedly reveals that the fixed trench height H_{Trench} is physically reliable and accurate.

Furthermore, the influence of weighting factors on the measurement accuracy of the ELFR method has been investigated using the traversing method. Under the constraint condition $w_1 + 5w_2 = 1$, by traversing the value of weight w_1 in the range of [0, 1] with a step size of 0.05, the trench height H_{Trench} and their coefficient of determination R^2 and maximum absolute deviation under each weight vector $[w_1, w_2]^T$ can be calculated using the ELFR method. The calculated coefficient of determination R^2 and maximum absolute deviation are presented in Fig. 10(a) and (b), respectively.

According to the results shown in Fig. 10(a), as the weight factor w_1 increases from 0 to 1, the coefficient of determination R^2 will continue to increase from 0.879 and reach a peak of 0.945 at $w_1 = 0.65$ and then decrease to 0.918. With parabolic function-based regression analysis, the empirical function $y = -0.1658x^2 + 0.2116x + 0.8753$ with a coefficient of determination of 0.9924 can be obtained, which will describe the dependence between the coefficient of determination R^2 and the weight w_1 more accurately. By solving the parabolic equation $-0.1658x^2 + 0.2116x + 0.8753 = 0.93$, it can be found that when the weight w_1 takes a value within the range of [0.3601, 0.9161], the coefficient of determination will always satisfy the target condition $R^2 \geq 0.93$. Moreover, as the weight w_1 reaches 0.6381, the coefficient of determination R^2 will get the maximum peak of 0.9428.

As for the maximum absolute deviation shown in Fig. 10(b), it first decreases to the minimum value and then slowly grows, with the weight w_1 increasing. The empirical function $y = 1.524x^2 - 3.0x + 2.7$ with a coefficient of determination of 0.9862 can be determined using the parabolic function-based regression analysis, describing the weight dependency of the maximum absolute deviation reliably. Moreover, as the weight w_1 becomes larger than 0.2705, the maximum absolute deviation will always satisfy the target condition $\mathbf{max}(\varepsilon_{\text{bias}}) \leq 2$ nm.

Due to the fitting error of the parabolic function-based regression analysis, this empirical equation indicates that the maximum absolute deviation will reach the minimum value of 1.2236 nm at the weight $w_1 = 0.9843$. At the same time, the actual test shows the weight $w_1 = 0.65$ to get the minimum value of the maximum absolute deviation. By calculating the intersection between the selectable weights revealed in Fig. 10(a) and (b), it can be straightforward to determine the weight range as the [0.3601, 0.9161], which enables the coefficient of determination R^2 and the maximum absolute deviation simultaneously satisfying the target threshold requirements. Namely, by configuring the weight w_1 within the range of [0.3601, 0.9161], the ELFR method can extract more accurate trench height H_{Trench} than the original LM nonlinear regression method and the LMRR method. Meanwhile, the weights vector $[w_1, w_2]^T = [0.65, 0.07]^T$ is the optimal solution for reaching the peak value of R^2 and the valley value of the maximum absolute deviation, respectively, which implies the most satisfactory measurement precision of the IM OCD tool. It should be emphasized that the selectable range and optimal configuration of the above weight factors are of guiding significance to ensure the best performance of the IM OCD tool in the CMP process section of STI nanostructure and the adaptability to complex working conditions of advanced IC manufacturing.

F. Another Experimental Results

To assess the robustness of the ELFR method, a sample structure from the metal interconnection layer process, denoted as Metal 2 (M2), is chosen for validation. The structure of the M2 sample consists of an oxide layer deposited on the Si substrate, followed by a SiNx layer, and then another thicker oxide film layer. Voids are etched into this oxide layer to create a specific shape, after which metal tungsten (W) is filled into these voids to prepare the nanostructure. The structural details of the M2 sample are depicted in Fig. 11. To simplify the forward optical model and enhance the parameter extraction effectiveness, specific structural parameters that minimally affect the spectrum are set as fixed values. The primary variables include the TCD, bottom CD (BCD), height of the tungsten metal layer (H_{W}), and total height of silicon dioxide (H_{All}). Among these parameters, particular focus is placed on the structural parameter H_{W}. The detailed measurement configurations for characterizing specimens are shown in Table III.

The metrology process of the M2 experiment is the same as that of the STI structure. First, the H_{W} result is extracted using the LM method, and another H_{W} result is obtained using the ELFR method. Finally, the results are compared, which is shown in Fig. 12.

Fig. 12(a) illustrates the coefficient of determination R^2 of the results extracted by the LM algorithm, yielding an R^2 value of 0.4076, accompanied by more significant azimuth-dependent errors shown in Fig. 12(b). In contrast, the extracted results of the ELFR method have a coefficient of determination $R^2 = 0.929$ and a suppressed azimuth

TABLE III
DETAILED MEASUREMENT CONFIGURATIONS OF M2

	Para.	Nominal values	Floating or fixing	Ranges
Main Meas. Conf.	TCD	40 nm	Floating	0~60 nm
	BCD	32 nm	Floating	0~50 nm
	H_{W}	60 nm	Floating	0~100 nm
	H_{All}	940 nm	Fixing	0~1100 nm
Other Conf.	φ		Floating	-90°~90°
	θ	45.02°	Fixing	44.7°~45.3°
	$\cdot$	—	—	340~800 nm

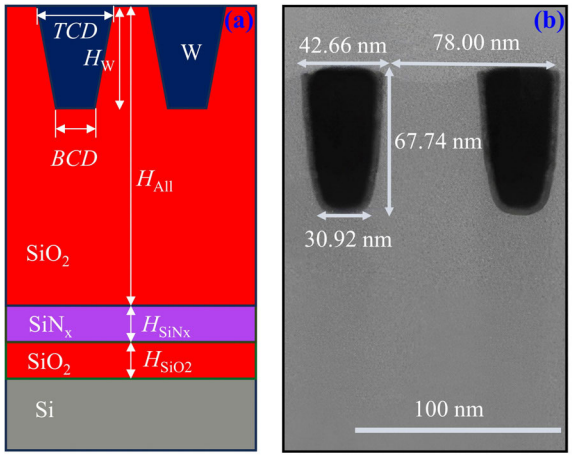

Fig. 11. Structural information of M2 samples. (a) Schematic of M2 nanostructure specimen. (b) Cross-sectional SEM view of the M2 nanostructure.

dependency of the deviation $\varepsilon_{\mathrm{bias}}$, as shown in Fig. 12(c) and (d). And the range of the deviation $\varepsilon_{\mathrm{bias}}$ shrinks from (−13, −4 nm) to (−2.0, 1.2 nm). The output range of H_{W} determined by the ELFR method is [62.25, 71.57 nm], with an average value of 67.24 nm, similar to the SEM characterization result of 67.74 nm. Optimization results of the weight factors are depicted in Fig. 13(a) and (b). It can be observed that the range of w_1, ensuring $R^2 \geq 0.93$ and $\mathbf{max}(\varepsilon_{\mathrm{bias}}) \leq 2$ nm, is [0.55, 0.9]. Remarkably, when the weight w_1 approaches 0.7, R^2 reaches the maximum value of 0.94, with the maximum deviation $\varepsilon_{\mathrm{bias}}$ being only 1.9 nm. These experimental results highlight the scalability of the ELFR method.

G. Discussion

The conventional IM-OCD tool primarily integrates the SR and SE probing module to characterize the geometrical parameters of nanostructures, such as linewidth, line height, and SWA. As for complex nanostructures containing multi-layer pattern stacks, grooves, and holes, it will be challenging to ensure measurement accuracy and adaptability using this method. Herein, it is full of practical significance to self-build the MME-based IM OCD tool to meet increasingly strict inline metrology requirements of the wafer-level nanostructure. However, its metrology precision suffers a non-negligible loss, due to the variations in incidence azimuth angles caused by the polar-coordinate wafer stage.

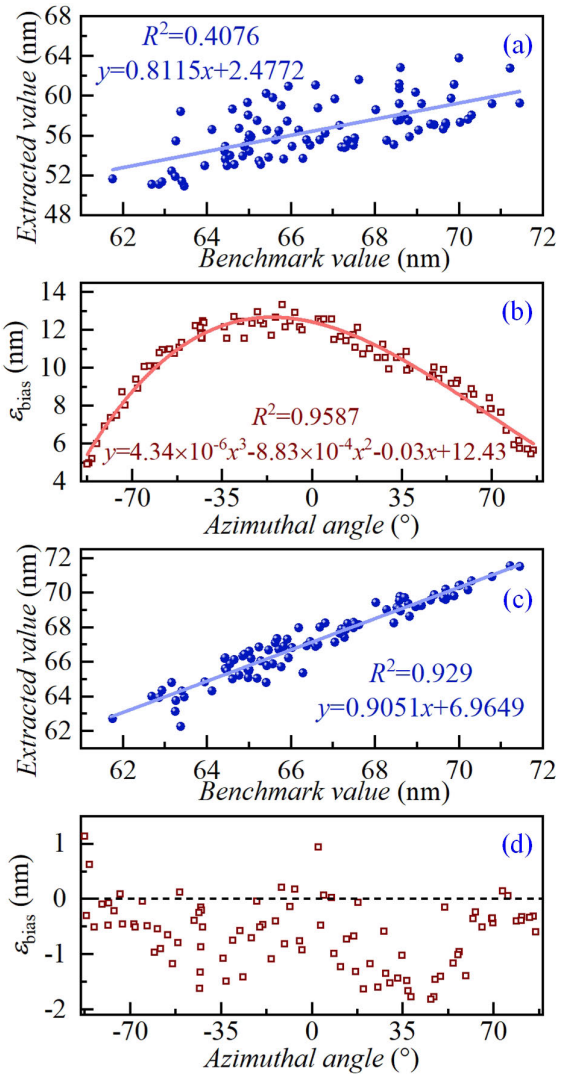

Fig. 12. Result of the ELFR method. (a) Coefficient of determination R^2 and (b) azimuthal-dependent deviation $\varepsilon_{\mathrm{bias}}$ of the results extracted by the LM method. (c) R^2 and (d) azimuthal-dependent deviation $\varepsilon_{\mathrm{bias}}$ of the results extracted by the ELFR method.

Differing from the LM algorithm [36], [49], library search strategy [22], and the typical machine learning algorithm [32], [50], the proposed ELFR method utilizes the weighted averaging of the respective extraction results from the LMRR and the BMNN algorithms to improve nanostructure reconstruction in the MME-based IM OCD technique. The LMRR algorithm adopts an RR strategy to construct and train a mapping model from the structural parameters extracted by the LM algorithm to the final structural parameters instead of directly mapping the measured signals to the final structural parameters, thereby enhancing the inverse extraction precision. As shown in Fig. 7, the corresponding coefficient of determination R^2 reaches 0.918, which is better than the $R^2 \leq 0.90$ obtained using a conventional neural network model and a cross-validation neural network model. Meanwhile, the absolute deviation $\varepsilon_{\mathrm{bias}} \leq$ 1.2 nm determined by the LMRR algorithm is better than that specified by the neural network model, which indicates that the LMRR rather than the BMNN dominates the metrology precision of the proposed ELFR method. Moreover, it can

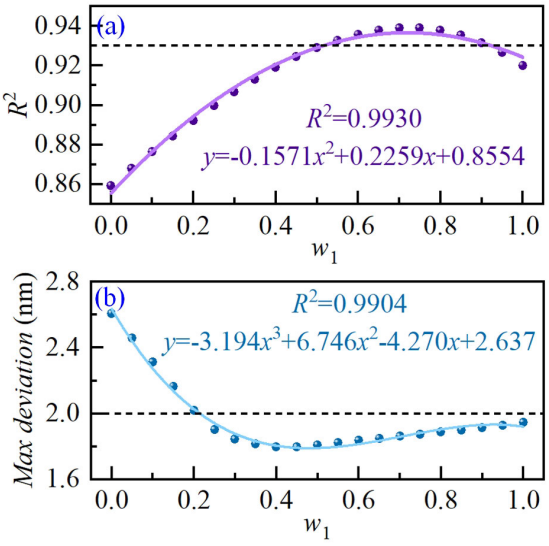

Fig. 13. (a) Coefficient of determination R^2 and (b) maximum absolute deviation determined by the ELFR method under different weight vector $[w_1, w_2]^T$ for the M2 sample.

be easily found that the azimuth dependency of the deviation ε_{bias} determined by the LMRR and BMNN algorithms exhibit complementarity at the azimuth angle of 0°, according to Figs. 7(c) and 8. These results lay the theoretical foundation for the weighted averaging between the results extracted by the LMRR and BRNN algorithms, which enables better metrology precision. By comparing the results obtained using the ELFR algorithm, the LM algorithm, the LMRR algorithm, the LM algorithm, the RR algorithm, the BMNN algorithm, and other revised ELFR algorithms, as shown in Table II, we can find that a single strategy among these algorithms cannot eliminate the dependency of extracted parameters on the incidence azimuth angle.

The MME exhibits high measurement sensitivity primarily through the captured 4 × 4 Mueller elements. However, with all the off-diagonal elements in the Mueller matrix being zero at azimuth angles of ±90° and 0°, the measurement sensitivity of the MME-based IM OCD tool would diminish near these three azimuth angles. Consequently, the LMRR and BMNN algorithms would yield significant absolute deviations near these azimuth angles. Since the BMNN algorithm cannot strictly distinguish the Mueller matrix spectra at the azimuth angles of ±90° and 0°, especially unable to distinguish the difference in the isotropic scattering features exhibited by the nanostructure at these three azimuth angles, the extracted results at these three azimuth angles are similar and have the same sign. Thus, the reversing sign of the extracted results at the azimuth angle of 0° enables the trade-off between the LMRR and BMNN algorithms, illustrating the feasibility of the weighted average fusion model algorithm.

Taking the trench height metrology for the STI nanostructure after the CMP processing as an example, the coefficient of determination R^2 larger than 0.93, the mean absolute deviation less than 1 nm, and the maximum absolute deviation less than 2 nm can be easily realized using the proposed ELFR method. Fig. 10 illustrates that the ELFR method eliminates the evolution trend of result parameters varying with azimuth angle. By optimizing the weight vector $[w_1, w_2]^T$ to the optimal configuration $[0.65, 0.07]^T$ using the traversing method, the coefficient of determination R^2 and the maximum absolute deviation can be improved to 0.945 and 1.292 nm, respectively, which together symbolize the optimal configuration for achieving the highest measurement precision and consistency. Meanwhile, selectable ranges of [0.3601, 0.9161] and [0.0168, 0.1280] for w_1 and w_2, respectively, have been suggested to ensure the R^2 and ε_{bias} meet the preset requirements for inline precision metrology. Both the optimal weights and the selectable ranges of weights are of guiding significance to provide the best performance of the MME-based IM OCD technique in the CMP-processed STI nanostructure and the adaptability to complex working conditions of advanced IC manufacturing. The robustness of the ELFR method is further demonstrated through measurement experiments on the metal interconnection M2 nanostructure.

V. Conclusion

In summary, this study proposes an MME-based IM OCD technique to meet the stringent inline metrology demands of advanced IC manufacturing at sub-28-nm nodes. The MME probe is integrated into the IC process platform via an automated polar-coordinate stage, enabling the precise metrology of the wafer-level nanostructures at a high throughput of 144 wafers per hour. We then propose an ELFR approach to improve nanostructure reconstruction in the MME-based IM OCD tool, specifically addressing metrology errors caused by variations in incidence azimuth angle. In the ELFR method, the weighted averaging of the respective extraction results from the LMRR and the BMNN algorithms dominates the final parameter extraction. The LMRR method is first introduced in the inverse extraction of structural parameters involved in the IM OCD tool, in which a mapping model from the structural parameters extracted by the LM algorithm to the structural parameters of a given benchmark sample has been constructed and trained. Then, the BMNN has been used to train five mapping models through cross-validation sampling. In addition, by optimizing the weight set (w_1, w_2), the ELFR method can achieve optimal metrology accuracy and consistency, which is much superior to the conventional LM algorithm.

Using the CMP-processed STI wafer and metal interconnection M2 wafer as the measurement object, the results reported by the ELFR method, the LM algorithm, the LMRR algorithm, and the BMNN algorithm have been compared in detail, which could highlight the underlying mechanism for improving the measurement accuracy and robustness using the proposed method. With the weights w_1 and w_2 located in the optimal setting window, like $w_1 \in [0.6, 0.9]$ and $w_2 \in [0.02, 0.08]$, the ELFR method can easily achieve a coefficient of determination R^2 greater than 0.93, a mean absolute deviation less than 1 nm, and a maximum absolute deviation less than 2 nm. Moreover, an additional comparative analysis of simulation and measurement results has been carried out, which indicates that the excellent metrology capability of the proposed ELFR method can be attributed to the balance and cancellation characteristics of the azimuth dependency of the results extracted by the LMRR and BMNN algorithms. It can

be expected that the proposed ELFR method, combined with the MME-based IM OCD tool, will be promising as a precise inline metrology approach that advanced IC manufacturing requires.

ACKNOWLEDGMENT

The authors thank the technical support from the Experiment Center for Advanced Manufacturing and Technology in School of Mechanical Science and Engineering at Huazhong University of Science and Technology.

REFERENCES

[1] N. G. Orji et al., "Metrology for the next generation of semiconductor devices," *Nat. Electron.*, vol. 1, no. 10, pp. 532–547, 2018.

[2] J. Park et al., "Microsphere-assisted hyperspectral imaging: Super-resolution, non-destructive metrology for semiconductor devices," *Light, Sci. Appl.*, vol. 13, no. 1, p. 122, May 2024.

[3] C.-H. Jen, S. S. Fan, and Y.-Y. Lin, "Data-driven virtual metrology and retraining systems for color filter processes of TFT-LCD manufacturing," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[4] Q. Sun et al., "Nondestructive monitoring of annealing and chemical–mechanical planarization behavior using ellipsometry and deep learning," *Microsystems Nanoengineering*, vol. 9, no. 1, p. 50, Apr. 2023.

[5] M. G. Faruk, M. S. Angyal, O. Ogunsola, D. K. Watts, and R. Wilkins, "Variability modeling and process optimization for the 32 nm BEOL using in-line scatterometry data," *IEEE Trans. Semicond. Manuf.*, vol. 27, no. 2, pp. 260–268, May 2014.

[6] A. C. Diebold, A. Antonelli, and N. Keller, "Perspective: Optical measurement of feature dimensions and shapes by scatterometry," *APL Mater.*, vol. 6, no. 5, May 2018, Art. no. 058201.

[7] T. Li, A. Chen, L. Fan, M. Zheng, J. Wang, and G. Lu, "Photonic-dispersion neural networks for inverse scattering problem," *Light Sci. Appl.*, vol. 10, p. 154, May 2021.

[8] M. Rizquez, A. Roussy, D. Pompier, J. Pinaton, and J. Pasquet, "Optimizing the critical dimension of the STI etch process by integrating inline scatterometry measurements and feedback R2R control," in *Proc. Int. Symp. Semicond. Manuf. (ISSM)*, Dec. 2016, pp. 1–4.

[9] Y.-C. Lu, J.-K. Huang, K.-Y. Chao, L.-J. Li, and V. P.-H. Hu, "Projected performance of Si- and 2D-material-based SRAM circuits ranging from 16 nm to 1 nm technology nodes," *Nature Nanotechnol.*, vol. 19, no. 7, pp. 1066–1072, Jun. 2024.

[10] Y. Lin et al., "Scaling aligned carbon nanotube transistors to a sub-10 nm node," *Nature Electron.*, vol. 6, no. 7, pp. 506–515, Jul. 2023.

[11] F. Khatkhatay et al., "Scribe line defect-induced yield loss in FINFET technology," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 387–392, Nov. 2019.

[12] W. Sun, A. Doi, M. Isawa, V. Vega-Gonzales, Z. Tokei, and G. Lorusso, "In-line metrology for vertical edge placement control of monolithic CFET using CD-SEM," in *Proc. Metrology, Inspection, Process Control*, Apr. 2023, p. 63.

[13] Y. Fu, T. Bi, Y. Chang, R. Xu, Y. Xu, and H. Kawarada, "Oxidized-silicon-terminated diamond p-FETs with SiO_2-filling shallow trench isolation structures," *IEEE Electron Device Lett.*, vol. 44, no. 11, pp. 1899–1902, Nov. 2023.

[14] P. Zhang, F. Peng, D. Yang, Z. Lei, and Y. Song, "A Laplace sensitivity operator enhances the calculation efficiency of OCD metrology," *Opt. Exp.*, vol. 31, no. 2, p. 2147, Jan. 2023.

[15] K. Fujita, K. Fujii, L. Zhang, Y. Azuma, S. Mizushima, and N. Kuramoto, "Investigating stability of Si sphere surface layer in Ambient–Vacuum cyclic measurements using ellipsometry," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.

[16] S. Liu et al., "Machine learning aided solution to the inverse problem in optical scatterometry," *Measurement*, vol. 191, Mar. 2022, Art. no. 110811.

[17] W. Yang, J. Hu, R. Lowe-Webb, R. Korlahalli, D. Shivaprasad, and H. Sasano, "Line-profile and critical-dimension monitoring using a normal incidence optical CD metrology," *IEEE T. Semiconduct. M.*, vol. 17, no. 4, pp. 1–26, Nov. 2004.

[18] S. Liu, X. Chen, T. Yang, J. Zhang, and S. Liu, "Inverse optical scatterometry using sketch-guided deep learning," *Opt. Exp.*, vol. 32, no. 11, p. 20303, May 2024.

[19] J. Weng, C. Gao, and B. Lei, "Deep learning enabled spatially polarization modulated Mueller matrix ellipsometer," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.

[20] P. Yu, W. Jiang, X. Wang, and Q. Wen, "A newly developed R-theta-Z motion stage for high-speed wafer inspection system," *Design, Manuf. Mechatronics*, vol. 3, pp. 652–662, Feb. 2017.

[21] C. Guo et al., "Integrated optical critical dimension metrology with Mueller matrix ellipsometry," *Thin Solid Films*, vol. 768, Mar. 2023, Art. no. 139695.

[22] J. Zhu, S. Liu, X. Chen, C. Zhang, and H. Jiang, "Robust solution to the inverse problem in optical scatterometry," *Opt. Exp.*, vol. 22, no. 18, p. 22031, Sep. 2014.

[23] T. Yang, X. Chen, S. Liu, J. Zhang, and S. Liu, "Condition-number-based measurement configuration optimization for nanostructure reconstruction by optical scatterometry," *Meas. Sci. Technol.*, vol. 34, no. 12, Aug. 2023, Art. no. 125001.

[24] 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.

[25] S. Liu, "Estimation of the convergence order of rigorous coupled-wave analysis for binary gratings in optical critical dimension metrology," *Opt. Eng.*, vol. 51, no. 8, May 2012, Art. no. 081504.

[26] Q. Wang, Y. Wei, Q. Wu, Y. Li, and X. Liu, "An optical critical dimension (OCD) model analysis on 3nm complementary FET (CFET) gate stacks," in *Proc. Metrology, Inspection, Process Control*, Apr. 2023, p. 97.

[27] N. L. Kazanskiy and P. G. Serafimovich, "Cloud computing for rigorous coupled-wave analysis," *Adv. Opt. Technol.*, vol. 2012, pp. 1–7, Jul. 2012.

[28] X. Chen, S. Liu, C. Zhang, and H. Jiang, "Measurement configuration optimization for accurate grating reconstruction by Mueller matrix polarimetry," *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 12, no. 3, Aug. 2013, Art. no. 033013.

[29] X. Chen, H. Gu, H. Jiang, C. Zhang, and S. Liu, "Probing optimal measurement configuration for optical scatterometry by the multi-objective genetic algorithm," *Meas. Sci. Technol.*, vol. 29, no. 4, Mar. 2018, Art. no. 045014.

[30] J. Zhu, Y. Shi, L. L. Goddard, and S. Liu, "Application of measurement configuration optimization for accurate metrology of sub-wavelength dimensions in multilayer gratings using optical scatterometry," *Appl. Opt.*, vol. 55, no. 25, p. 6844, Aug. 2016.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[32] K. A. Brown, S. Brittman, N. Maccaferri, D. Jariwala, and U. Celano, "Machine learning in nanoscience: Big data at small scales," *Nano Lett.*, vol. 20, no. 1, pp. 2–10, Dec. 2019.

[33] J. Liu, D. Zhang, D. Yu, M. Ren, and J. Xu, "Machine learning powered ellipsometry," *Light: Sci. Appl.*, vol. 10, no. 1, p. 55, Mar. 2021.

[34] S. Robert, A. Mure-Ravaud, and D. Lacour, "Characterization of optical diffraction gratings by use of a neural method," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 19, no. 1, p. 24, Jan. 2002.

[35] Z. Tang, Y. Liu, T. Liu, X. Xu, and J. Liu, "Adaptive weighted ridge regression estimator for time-varying sensitivity identification," *IEEE Trans. Power Syst.*, vol. 39, no. 1, pp. 2377–2380, Jan. 2024.

[36] X. Wang, X. Wang, B. Ma, Q. Li, and Y.-Q. Shi, "High precision error prediction algorithm based on ridge regression predictor for reversible data hiding," *IEEE Signal Process. Lett.*, vol. 28, pp. 1125–1129, 2021.

[37] Z. Li, S. Li, and X. Luo, "Efficient industrial robot calibration via a novel unscented Kalman filter-incorporated variable step-size Levenberg–Marquardt algorithm," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[38] P. Guo, H. Dai, Q. Yang, Q. Huang, and H. Yao, "An improved magnetic tracking approach based on ResNet-LM fusion algorithm," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[39] J. Liu et al., "Longitudinal magneto-optical Kerr effect in subwavelength thick ferromagnetic films investigated by Mueller matrix ellipsometry," *Opt. Lasers Eng.*, vol. 171, Dec. 2023, Art. no. 107803.

[40] M. Korde et al., "Nondestructive characterization of nanoscale subsurface features fabricated by selective etching of multilayered nanowire test structures using Mueller matrix spectroscopic ellipsometry based scatterometry," *J. Vac. Sci. Technol. B, Nanotechnol. Microelectronics: Mater., Process., Meas., Phenomena*, vol. 38, no. 2, Mar. 2020, Art. no. 024007.

[41] J. Liu, Z. Zhou, H. Gu, J. Zhu, H. Jiang, and S. Liu, "Asymmetric optical properties and bandgap shift of pre-strained flexible ZnO films," *APL Mater.*, vol. 12, no. 4, Apr. 2024, Art. no. 041126.
[42] J. Liu et al., "Measurement configuration optimization for dynamic metrology using Stokes polarimetry," *Meas. Sci. Technol.*, vol. 29, no. 5, Apr. 2018, Art. no. 054010.
[43] S. Trenn, "Multilayer perceptrons: Approximation order and necessary number of hidden units," *IEEE Trans. Neural Netw.*, vol. 19, no. 5, pp. 836–844, May 2008.
[44] M. Wurm, J. Endres, J. Probst, M. Schoengen, A. Diener, and B. Bodermann, "Metrology of nanoscale grating structures by UV scatterometry," *Opt. Exp.*, vol. 25, no. 3, p. 2460, Feb. 2017.
[45] F. Fang et al., "Improving OCD time to solution using signal response metrology," in *SPIE Proc.*, Mar. 2016, pp. 1–20.
[46] A. Vaid et al., "A holistic metrology approach: Hybrid metrology utilizing scatterometry, CD-AFM, and CD-SEM," in *SPIE Proc.*, vol. 7971, Mar. 2011, Art. no. 797103.
[47] B. Bunday et al., "Tool-to-tool matching issues due to photoresist shrinkage effects," in *SPIE Proc.*, vol. 7971, Mar. 2011, p. 79710.
[48] X. Chen, H. Gu, J. Liu, C. Chen, and S. Liu, "Advanced Mueller matrix ellipsometry: Instrumentation and emerging applications," *Sci. China Technological Sci.*, vol. 65, no. 9, pp. 2007–2030, Aug. 2022.
[49] H. Fujiwara, *Spectroscopic Ellipsometry: Principles and Applications*. Hoboken, NJ, USA: Wiley, 2007, pp. 87–129.
[50] X. Chen, S. Liu, C. Zhang, and J. Zhu, "Improved measurement accuracy in optical scatterometry using fitting error interpolation based library search," *Measurement*, vol. 46, no. 8, pp. 2638–2646, Oct. 2013.
[51] Z. Jiang, Z. Gan, C. Liang, and W.-D. Li, "Generic characterization method for nano-gratings using deep-neural-network-assisted ellipsometry," *Nanophotonics*, vol. 13, no. 7, pp. 1181–1189, Jan. 2024.

Chunfu Guo is currently pursuing the Ph.D. degree in mechanical engineering with Huazhong University of Science and Technology, Wuhan, China.

His research interests include optical critical dimension (OCD) metrology technology based on Mueller matrix ellipsometry, including standalone- and integrated-OCD metrology.

Jiamin Liu received the Ph.D. degree in mechanical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2020.

He has been a Post-Doctoral Researcher at the National Key Laboratory of Intelligent Manufacturing Equipment and Technology, HUST, since 2020. His research interests include in situ Mueller matrix ellipsometry and applications, patterned wafer inspection, and computational lithography modeling.

Honggang Gu received the B.S. and Ph.D. degrees from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2011 and 2016, respectively.

He is currently a Professor with the School of Mechanical Science and Engineering, HUST. His research interests include instrumentation and applications of computational imaging and advanced ellipsometry, optical metrology and defect inspection of integrated circuits, and characterization of low-dimensional materials and complex.

Jinlong Zhu received the Ph.D. degree in mechanical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2015.

From 2015 to 2020, he worked as a Post-Doctoral Researcher at the University of Illinois Urbana-Champaign, Champaign, IL, USA. He became a Professor with Huazhong University of Science and Technology in 2020. He conducts research on basic science and engineering at the intersection of optical physics, nanodevices, semiconductor integrated circuit (IC) chips in industrial robots, and machine learning.

Chuanwei Zhang received the B.E. and M.E. degrees in mechanical engineering from Wuhan University, Wuhan, China, in 2004 and 2006, respectively, and the Ph.D. degree in mechanical engineering from Huazhong University of Science and Technology (HUST), Wuhan, in 2009.

After two years of post-doctoral research, he is currently an Associate Professor at the School of Mechanical Science and Engineering, HUST. His research interests include Mueller matrix ellipsometry-based metrology techniques for optical critical dimension, overlay, and nanofilm profile in advanced nanomanufacturing.

Xiuguo Chen received the Ph.D. degree in mechanical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2013.

He worked as a JSPS Fellow at Tohoku University, Sendai, Japan, from 2016 to 2018. He is currently a Professor at the School of Mechanical Science and Engineering, HUST. His research interests include optics and X-ray-based nanometrology, and instrumentation for advanced semiconductor manufacturing.

Hao Jiang received the Ph.D. degree in mechanical engineering from Florida Institute of Technology, Melbourne, FL, USA, in 2011.

He is currently a Professor at the School of Mechanical Science and Engineering, HUST, Wuhan, China. He is interested in the research topics of polarization optics-based multiparameter measurement, micro- and nano-scale mechanical variable measurement, complex smart/bio material-based computational metrology, and computational lithography.

Shiyuan Liu received the Ph.D. degree in mechanical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1998.

He worked as a Visiting Scholar at the University of Manchester, Manchester, U.K., from 2000 to 2001. He is currently a Professor at the School of Mechanical Science and Engineering, HUST. His research interests include computational imaging and computational lithography, micro/nano measurement technology and instrumentation, optical metrology, and defect inspection for integrated circuit manufacturing.