Integrated optical critical dimension metrology with Mueller matrix ellipsometry

Chunfu Guo \(^a\), Yating Shi \(^b\), Huaxi Wu \(^b\), Weiqi Li \(^b\), Chuanwei Zhang \(^a,\ast\), Hao Jiang \(^a,\ast\), Shiyuan Liu \(^a,\ast\)

\(^a\) State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China
\(^b\) Wuhan Eoptics Technology Co. Ltd., Wuhan 430075, China
\(^\ast\) Optics Valley Laboratory, Wuhan 430074, China


c A R T I C L E   I N F O

Keywords:
Integrated metrology
Optical critical dimension
Mueller matrix ellipsometry
Sensitivity analysis
Ridge regression

A B S T R A C T

Mueller matrix ellipsometry (MME) is commonly applied by the standalone instruments in semiconductor manufacturing process for films and nanostructures characterization. However, MME is rarely used in the integrated metrology optical critical dimension (IM OCD) tools due to the difficulty in extracting the parameters under varied azimuth angle conditions, which is induced by the rotation of R-\(\theta\) wafer stage adapted to the restricted space. When the measurements on a same wafer are achieved under multiple azimuth angles, the measured nanostructure parameters usually mismatch the baseline values provided by the manufacturer, and therefore lead to the unacceptable accuracy loss. In this paper, we propose an azimuthal sensitivity analysis and ridge regression algorithm (ASA-RR) to enable the MME-based IM OCD. The sensitivity to variability in azimuth angles is calculated by the local sensitivity analysis algorithm, and the result of which is applied as the weight factor for the spectrum input in the ridge regression. Experiments demonstrate that the ASA-RR algorithm provides more accurate results for the IM OCD, which satisfy the requirements in manufacturing.

1. Introduction

Nanostructure morphology, including the film thickness, line width, line height, sidewall angle, etc. \([1–3]\) is necessary to be monitored to optimize and adjust the process parameters in time for the device quality control. Optical scatterometry, frequently referred to as optical critical dimension (OCD) metrology \([4–6]\), has significant advantages for in-line metrology as a fast, low-cost, non-contact, non-destructive, and easily being integrated technology. The spectrscopic reflectometry (SR) and spectroscopic ellipsometry (SE) are two typical OCD measurement technologies. Instead of measuring the intensities using normal incidence configuration by SR, SE adopts the oblique incidence configuration and measures the polarization change, which exhibits the advantage of characterizing the complex nanostructures \([7,8]\) due to the abundant scattered light information. Therefore, SE is more popularly used in the standalone OCD instruments. However, in the integrated metrology optical critical dimension (IM OCD) tools, the R-\(\theta\) wafer stage is usually adopted due to the restricted space. The rotation of the R-\(\theta\) wafer stage induces the changes of the azimuth angles among the measurements of different pads on the same wafer, which prohibits the application of the azimuth sensitive metrology in IM OCD, such as SE. Comparing with the conventional SE that can only acquire two ellipsometric angles, Mueller matrix ellipsometry (MME) can provide all 16 elements of the \(4 \times 4\) Mueller matrix, and consequently exhibits superior sensitivity and accuracy on nanostructure metrology due to the rich information such as anisotropy and depolarization acquired \([9–11]\). The same as SE, MME is a typical model-based method which consists of two important steps known as the forward modeling and inverse problem solving. The inverse problem is a nanostructure parameter extraction matching the measured spectrums and the theoretical spectrums calculated by the forward model, including the finite-different time-domain (FDTD) method \([12]\), the finite element method (FEM) \([13]\), and the rigorous coupled-wave analysis (RCWA) method \([14–16]\). The well-known methods for inverse problem solving include the library search method \([17–19]\), the Levenberg–Marquardt (LM) method \([20, 21]\), and the machine learning method \([22, 23]\).

Sensitivity analysis is usually applied to estimate the influence of the measurands variations on the calculated spectrum \([24]\). The methods
described in [25] involve calculating the partial derivatives of the modeled spectrum to each nanostructure parameter respectively, which can be summarized as the local sensitivity analysis (LSA). Another method of sensitivity analysis is the global sensitivity analysis (GSA) [26], in which all nanostructure parameters are floated instead of a single parameter in LSA. Several GSA techniques can be used to calculate the sensitivity, including Morris [27], Sobol [28], and the extended Fourier amplitude sensitivity test (EFAST) [29]. No matter which method is adopted, the objective of sensitivity analysis is in general limited in the applications such as the optimizing measurement configurations [30–33], while the utilizing the sensitivity change as the measured information instead of only as the reference information have been rarely explored.

Inspired by the superior sensitivity of Mueller matrix spectra to the azimuth angle change [34,35], we propose to use the sensitivity spectra encoded in the Mueller matrix spectra to overcome the accuracy degradation issue, and furthermore enable the application of MME in IM OCD. Besides the azimuthal sensitivity analysis, we propose a ridge regression method to solve the corresponding inverse problem with sensitivity change considered. The sensitivity of the Mueller matrix spectrums versus azimuth angles, which is applied to update the measurement spectrums in the ridge regression, is calculated by the partial derivatives. Then, a ridge regression model is used to map the spectrums to the nanostructure parameters from the training wafer. The test wafer is used to prove the accuracy and robustness of the proposed method by comparing the extracted parameters to the baseline values provided by the manufacturer. Experimental results show that the influence of varied azimuth angles is significantly weakened and consequently the measurement accuracy is improved, when the sensitivity is considered using the proposed method.

The paper is organized as follows. Section 2 first briefly introduces the proposed azimuthal sensitivity analysis and ridge regression (ASA-
RR) method. Section 3 provides the influence analysis of azimuth angles to Mueller matrix spectrums using simulations, then demonstrates the accuracy of the ASA-RR method using experimentally measured spectrums. Finally, we draw some conclusions in Section 4.

2. Method

2.1. Overview of the ASA-RR method

The flow chart shown in Fig. 1 is the proposed ASA-RR method that extracts the nanostructure parameters from the spectrum. The 4×4 Mueller matrix spectrum is reshaped as a 16-dimensional vector $X = [X_1, X_2, \ldots, X_{16}]$ in optical scatterometry, and the given nanostructure parameters as a n-dimensional vector $Y = [Y_1, Y_2, \ldots, Y_n]$ where $n$ represents the number of parameters. The inverse problem can be summarized as a minimization problem:

$$F = \arg\min_x \left\{ \sum (Y - F(x))^2 \right\}$$

(1)

The $F(x)$ in Eq.1 is a function that outputs nanostructure parameters from the input spectrum $X$, and which is learnt by the ridge regression method. Ridge regression is a biased estimation regression method which can be used for collinear data analysis in machine learning. By abandoning the unbiased estimator of the least square method, the regression coefficient is obtained at the cost of accuracy. Compared with the least square method, ridge regression is more practical and reliable in the few data and ill-posed problem. Furthermore, the influence of varied azimuth angles is weakened due to the application of sensitivity analysis. The procedure of the ASA-RR method is described as follows.

Step 1: Extract nanostructure parameters $y_{lm}$ from measurement spectrums $x$

Step 2: Calculate sensitivity $s$ to azimuth angles by the LSA technology.

Step 3: Modify the measurement spectrums $x_i$ by multiplying origin spectrums $x$ with sensitivity $s$

Step 4: Train the ridge regression $F(X)$ learnt from the measurement spectrums $x_i$ to target parameter $y_t$, and then compare the output $y_o$ with $y_t$ to calculate $R^2$ and the mean absolute bias.

The process of sensitivity analysis and the theory of the ridge regression are detailed in the following.

2.2. Azimuthal Sensitivity analysis

According to the RCWA model, the relationship between the Mueller matrix spectrum $X$ and the corresponding parameters $Y$ can be represented as:

$$X = f(Y, \varphi)$$

(1)

The varied azimuth angles $\varphi$ make significant influence on the spectrum and increase the difficulty in solving the inverse problem in the IM OCD, which can be solved by sensitivity analysis.

The corresponding nanostructure parameter $y_0$ and azimuth angle $\varphi_0$ can be extracted by the LM method in a measurement spectrum $x$. Given

Fig. 4. The range of parameters and azimuth angles baseline values in the training wafers.

Fig. 5. H_OX $R^2$ for the LM method compared with the baseline values in the test wafer.
the unavoidable noise in the measurement spectrum and the error of the LM method, the pure simulation spectrum \( x_0 \) is calculated by Eq.1:
\[
x_0 = f(y_0, \phi_0)
\] (2)

In LSA, with an assumption of slight departure vector \( \Delta \phi \) from \( \phi_0 \) to \( \phi' \) while the other nanostructure parameters are fixed, the new spectrum \( x' \) is calculated by:
\[
x' = f(y_0, \phi') = f(y_0, \phi_0 + \Delta \phi)
\] (3)

Hence, the sensitivity \( s_0 \) of spectrum \( x_0 \) to azimuth angle \( \phi_0 \) can be defined in the LSA as:
\[
s_0 = \frac{\Delta x}{\Delta \phi} = \frac{x' - x_0}{\Delta \phi}
\] (4)

The size of sensitivity \( S \) is a 16-dimensional vector \( S = [S_1, S_2, \ldots, S_{16}] \), which is equal to spectrum \( X \).

The modified spectrum \( x_s \) is from the multiplication of measurement spectrum \( x \) and sensitivity \( s \):
\[
x_s = x \ast s
\] (5)

where \( s \) can be calculated by the LM method. After modification on the spectrum, each Mueller matrix weight in varied azimuth angles is more balanced than on the unmodified spectrum. And the new spectrum is regarded as the input ridge regression.

2.3. Ridge regression

Ridge estimation is conducted on the linear regression model
\[
Y = X\beta + \epsilon
\] (6)

\( Y \) are the nanostructure parameters and \( X \) are the modified Mueller matrix spectrums in optical scatterometry. The regression weights and bias are represented by \( \beta \) and \( \epsilon \). The solution of minimization problem can be provided by the ordinary least squares (OLS) estimator:
\[
\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - x_i\beta)^2
\] (7)

When \( X \) have full rank \((N > K)\), the solution of the OLS problem is
\[
\hat{\beta} = (X^T X)^{-1} X^T Y
\] (8)

However, the number of measurement spectrums is limited in the inverse problem while wavelengths of spectrums are collected in wide range, since \( K \gg N \) in the case.

To address the problem, a ridge estimator \( \hat{\beta}_\lambda \) is introduced into modified minimization problem slightly
\[
\hat{\beta}_\lambda = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - x_i\beta)^2 + \lambda \sum_{k=1}^{K} \beta_k^2
\] (9)

where \( \lambda \) is a positive constant, highly frequently as a small number, the penalty factor that prevents the solution \( \beta \) from an unbounded value, with a plus of the squared norm of the vector of coefficients.

Therefore, the solution of ridge regression model is
where \( I \) is the \( K \times K \) identity matrix.

Therefore, the ridge regression in Eqs. 9 and 10 solve the overfitting in optical scatterometry in a certain degree and enhances robustness of ridge regression method.

3. Results

3.1. Experimental setup

A typical two-dimensional periodic structure is investigated, as shown in Fig. 2, to assess the theoretic feasibility of the proposed method. The sample of the structure consists of a silicon dioxide (SiO\(_2\)) layer filled with one silicon dioxide layer and one silicon nitride (SiN) layer. Some structural parameters with little effect on the spectrums are fixed as nominal values given the complexity of the model and the efficiency of parameter extraction. Therefore, the profile of the structure is characterized as top critical dimension (TCD), height of silicon nitride layer (H\(_{\text{SiN}}\)), height of silicon dioxide layer (H\(_{\text{OX}}\)), and sidewall angle (SWA). Nominal dimensions of the measurement sample are: TCD = 56 nm, H\(_{\text{SiN}}\) = 31 nm, H\(_{\text{OX}}\) = 279 nm, SWA = 87°. The structural parameter of the investigated sample that need to be extracted is H\(_{\text{OX}}\), while H\(_{\text{SiN}}\), TCD, and SWA are also floating in a reasonable range, which have little influence to result. The ranges of the structural parameters TCD, H\(_{\text{SiN}}\), H\(_{\text{OX}}\) and SWA are 30–80 nm, 10–50 nm, 250–300 nm and 83° to 95°. The azimuth angles are also floating in the IM OCD measurement, with the range from -90° to 90°. The other measurement conditions are set as follows: the incidence angle is fixed at

\[
\hat{\beta}_1 = (X^T X + \lambda I)^{-1} X^T Y \quad (10)
\]
65°, the spectral range is set to 350–990 nm with an interval of 5 nm. In order to identify the accuracy of the ASA-RR method in practical measurements, a dual-rotating compensator MME (ME-L, Wuhan Eop-tics Technology Co.) is employed to measure the Mueller matrix spectrums of wafers in the MME-based IM OCD instrument. The principle of MME is demonstrated in Fig. 3. Four wafers are prepared and each of them contains 83 dies with the same geometric structure shown in Fig. 2. Three of them are applied to be the training wafers, while the other is regarded as the test wafer. The range of the azimuth angles and H_OX for each training wafer is shown in Fig. 4. The experiments of the LM method, ridge regression method and azimuthal sensitivity analysis-based ridge regression method in bias and Correlation coefficients $R^2$ are conducted on the test wafer to verify the robustness of model. All experiments are performed on a computer workstation with an Intel Xeon(R) Sliver 4210R @ 2.4 GHz central processing unit (CPU) and the programming language that implements all algorithms is Python.

### 3.2. LM method results and ridge regression results

The $R^2$ of the LM method in the measurement test wafer is shown in Fig. 5, which is ineligible with 0.49 lower than the pass line 0.9. The results of the LM method and baseline values compared are shown in Fig. 6(a), and the bias and the fitting curve versus the azimuth angle are illustrated in Fig. 6(b). It is observed that the bias can be expressed by a second-order polynomial approximately versus azimuth angle, with the fitting curve $R^2$ reaches 0.76.

The ridge regression method is also implemented in the parameter extraction, with the $R^2$ of the ridge regression method is 0.84 in Fig. 7, and the $R^2$ of the second-order polynomial fitting curve reduces to 0.06 in Fig. 8(b), which indicates the correlation between H_OX bias and the azimuth angles is significantly compressed. The achieved low correlation between the bias and azimuth angles fitting suggests that the influence of azimuth angles has been limited when the ridge regression is employed, while the result of parameter extraction keeps disqualified yet. Therefore, incorporating the effects of varied azimuth angles in the LM algorithm into the ridge regression algorithm is a potential solution to improve the prediction accuracy of MME-based IM OCD.

### 3.3. Azimuthal sensitivity analysis results

In the LSA, the slight departure vector of azimuth angle is set at ±0.1°. The mean values of azimuth angle sensitivity for all wavelengths versus the azimuth angles are shown in Fig. 9. The sensitivity is axial symmetric in diagonal elements including $[M_{12}, M_{21}, M_{33}, M_{44}]$ while centrosymmetric in off-diagonal elements $[M_{13}, M_{14}, M_{23}, M_{24}, M_{31}, M_{32}, M_{41}, M_{42}]$. Furthermore, the value of sensitivity of the

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean bias(nm)</th>
<th>Max bias(nm)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM method</td>
<td>1.146</td>
<td>3.484</td>
<td>0.498</td>
</tr>
<tr>
<td>Ridge regression method</td>
<td>0.948</td>
<td>1.883</td>
<td>0.847</td>
</tr>
<tr>
<td>ASA-RR method</td>
<td>0.872</td>
<td>1.917</td>
<td>0.935</td>
</tr>
</tbody>
</table>

### Table 1
Result of LM method, ridge regression method and ASA-RR method comparison in the test wafer.

![Fig. 10. H_OX $R^2$ for the ASA-RR method compared with the baseline values in the test wafer.](image)

![Fig. 11. (a) H_OX result and (b) H_OX bias for the ASA-RR method compared with the baseline values in the test wafer.](image)
diagonal elements at azimuth angle $\phi=0^\circ$ is higher than at azimuth angle $\phi=45^\circ$, while it is on the verse corresponding to the off-diagonal elements. The high correlation of the bias and the azimuth angles in Fig. 6 (b) caused by the inconstant sensitivity in varied azimuth angles confines the direct application of the Mueller matrix spectrum in parameter extraction. Therefore, the sensitivity analysis result is introduced as a modified factor in the ridge regression in the MME-based IM OCD to reduce the influence of varied azimuth angles.

### 3.4. ASA-RR method results

With the azimuthal sensitivity analysis applied in the ridge regression method, the $R^2$ of the test wafer as shown in Fig. 10 enhances to 0.93 over the pass line 0.9. The results of the ASA-RR method and baseline values compared are shown in Fig. 11 (a), and the bias and the second-order polynomial fitting curve versus the azimuth angle are illustrated in Fig. 11 (b), which can be compared with the results shown in Figs. 6 and 8 straightforwardly. The correlation between the bias and the azimuth angles is further weakened, and the $R^2$ is only 0.05, indicating that the residual bias is not caused by the azimuth angle variations. The results of three methods are shown in Table 1. The $R^2$ of the ridge regression method is higher than that of the LM method, and a higher $R^2$ is achieved by means of the azimuth angle sensitivity analysis. As such, it is observed that the ASA-RR method provides a solution to the problems of the varied azimuth angles in the MME-based IM OCD.

### 4. Conclusion

A strategy of utilizing the sensitivity change as the measured information instead of only as the reference information has been proposed to compress the influence of the varied azimuth angles to the measurement accuracy, which can be expected to enable the application of MME in IM OCD. The sensitivity of spectrum versus azimuth angles is calculated by a local sensitivity analysis algorithm. The different symmetry of the positive azimuth angle and the negative azimuth angle is observed in the Mueller matrix, which indicates the necessity of spectrum correction. A ridge regression model that maps the modified spectrum to nanostructure parameters is used by comparing the predicted results with the baseline values provided by the manufacturer. The $R^2$ achieved in the measurement experiment demonstrate that the ASA-RR method outperforms the LM method and the ridge regression method, reaching 0.93 in the test wafer. Therefore, the obstacle of parameter extraction in MME-based IM OCD is removed.

**CRediT authorship contribution statement**

**Chunfu Guo:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Yating Shi:** Methodology, Writing – review & editing. **Huaxi Wu:** Methodology, Data curation. **Weiqi Li:** Writing – review & editing. **Chuanwei Zhang:** Conceptualization, Writing – review & editing, Supervision. **Hao Jiang:** Conceptualization, Supervision. **Shiyuan Liu:** Conceptualization, Writing – review & editing.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data Availability**

Data will be made available on request.

**Acknowledgments**

This work was funded by the National Key Research and Development Plan of China (2019YFB20050602), the National Natural Science Foundation of China (Grant No. 51727809, 52130504), and the National Major Science and Technology Projects of China (2017ZX02101006-004). The authors would like to thank the technical support from the Experiment Center for Advanced Manufacturing and Technology in School of Mechanical Science and Engineering, Huazhong University of Science and Technology (HUST).

**References**


