

Data refinement for robust solution to the inverse problem in optical scatterometry

Jinlong Zhu, Hao Jiang, Chuanwei Zhang, Xiuguo Chen, and Shiyuan Liu *

State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

ABSTRACT

Optical scatterometry is widely used in the process control of integrated circuits (IC) manufacturing due to its inherent advantages such as nondestruction, high sampling rate, large aerial coverage and low-cost. However, in the conventional inverse problem solvent of optical scatterometry, the measurement errors are usually excessively simplified as normally distributed errors, which deviate from the actual complex ones. In this work, we will demonstrate that there exist typical outlying measurement errors in the measurement signature, and these outlying measurement errors will notably affect the result of each iteration step in the conventional Gauss-Newton (GN) method. By performing a method based on the principle of least trimmed squared estimator (LTS) regression instead of each GN iteration step, the higher measurement accuracy of a nanostructure can be achieved. The remarkably improved reconstruction of a deep-etched multilayer grating has demonstrated the feasibility of the proposed method.

Keywords: optical scatterometry, Mueller matrix ellipsometry, inverse problem, data refinement, least trimmed squared estimator.

1. INTRODUCTION

With the continuing shrinkage of critical dimension (CD) in integrated circuits (IC) and the requirements of high volume and high precision manufacturing ^[1], nanometrology is now facing up to tremendous challenges. Among all the metrology techniques, the optics-based ones, such as critical dimension small-angle x-ray scattering (CD-SAXS) ^[2], specular x-ray reflectivity (SXR) ^[3] and ellipsometric scatterometry ^[4, 5], are regarded as having the potential to meet these strict demands. Moreover, optical scatterometry, can be regarded as the optimal optics-based technique according to its inherent advantages such as nondestruction, high sampling rate, large aerial coverage and low-cost ^[6]. Based on these superiorities, optical scatterometry has been successfully applied in the in-chip CDs and overlay displacement errors metrology of the dynamic random access memory (DRAM) manufacturing ^[7] and process control for back-end-of-the-line (BEOL) ^[1], which have driven it to become the state-of-the-art nanometrology technique in IC manufacturing.

Optical scatterometry is essentially a model-based technique, whose success highly relies on two procedures, i.e., forward modeling and inverse problem solvent. For the forward modeling, it is the procedure of using electromagnetic modeling algorithms such as finite element method (FEM) ^[8], the boundary element method (BEM) ^[9], the finite-difference time-domain (FDTD) ^[10] or rigorous coupled-wave analysis (RCWA) ^[11, 12] to calculate the theoretical signature for a nanostructure. The forward modeling is relative simple and intuitive according to its inherent well-posedness ^[13]. While for the inverse problem, it is inherently ill-posed and has attracted much more research attentions.

Conventionally, the inverse problem in optical scatterometry is formulated as a nonlinear least square (LSQ) minimization with the object of finding the model parameters corresponding to the minimal LSQ function value. Under the normal distribution assumption of measurement errors, the LSQ function is the natural logarithm of the maximum likelihood estimation (MLE) function, whose minimization either by nonlinear regression with strict terminal condition or by library search with small enough grid can result in the optimal solution. However, as pointed out by Geary ^[14], in an actual measurement system, normality is only a myth. In the field of optical scatterometry, the measurement errors are never normally distributed in consideration of the multiple system error sources such as the finite bandwidth ^[15], the

* Corresponding author: shyliu@hust.edu.cn; phone: +86 27 87559543; webpage: <http://www2.hust.edu.cn/nom>.

power fluctuation of the incident beam^[16] and the collimation of light source^[15]. The superimposed effect of these different kinds of measurement errors may lead to some outlying ones whose scales are obviously larger than the rest. These outlying measurement errors will remarkably affect the final nanostructure measurement accuracy^[17, 18].

In the present paper, we will stand on the view of least trimmed squared estimator (LTS)^[19, 20] instead of LSQ fitting to reduce the effect of outlying measurement errors on the measurement accuracy. Specifically, we will firstly demonstrate that the result of each iteration step in the conventional LSQ function based Gauss-Newton (GN) method is notably affected by the outlying measurement errors. Then by using a method based on the principle of LTS we can eliminate these data pairs suffering from outlying measurement errors directly, after which the more robust and accurate iteration result can be achieved. A data pair consists of a current residual and a specific row of the linear operator, where the current residual is defined as the measured signature minus the theoretical signature of the current estimated model parameters, and the linear operator is the negative Jacobian matrix that is obtained by calculating the partial differential of signatures with respect to model parameters. Here those data pairs suffering from outlying measurement errors are uniformly called outliers in this manuscript. The proposed method offers a different way of thinking that only requires some mathematical methods to correct the intermediate procedures, and therefore the conventionally complicated and time-consuming processes such as measurement configuration optimization^[21] and correction of limited NA^[15] effect can be avoided. The improved reconstruction of a deep-etched multilayer grating has demonstrated the feasibility of the proposed method.

2. METHODOLOGY

2.1 Inverse problem in optical scatterometry

The inverse problem in optical scatterometry is usually described as an object to minimize an LSQ function, which can be generally expressed as

$$F(\mathbf{x}) = \sum_{j=1}^m w_j [y_j - f_j(\mathbf{x})]^2 = [\mathbf{y} - \mathbf{f}(\mathbf{x})]^T \mathbf{w} [\mathbf{y} - \mathbf{f}(\mathbf{x})]. \quad (1)$$

Here y_j is the j th measured data point, and \mathbf{y} is the measured signature as a vector containing m data points. $f_j(\mathbf{x})$ is the j th calculated data point with respect to the profile parameters under measurement as an n -dimensional vector $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, and $\mathbf{f}(\mathbf{x})$ is the calculated signature as a vector containing m data points. w_j is the j th weight factor, and \mathbf{w} is an $m \times m$ diagonal matrix with diagonal elements $\{w_j\}$. If the diagonal element w_j of matrix \mathbf{w} is given by $w_j = 1/\sigma^2(y_j)$, where $\sigma(y_j)$ is the standard deviation of the measurement error that follows the normal distribution, Eq. (1) relates to the commonly used chi-square statistic χ^2 . Consequently, without losing generality, the inverse problem in optical scatterometry can be formulated as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \Omega} \{[\mathbf{y} - \mathbf{f}(\mathbf{x})]^T \mathbf{w} [\mathbf{y} - \mathbf{f}(\mathbf{x})]\}, \quad (2)$$

where $\hat{\mathbf{x}}$ is the solution of the inverse problem, and Ω is the associated parameter domain.

2.2 Solution to the inverse problem with GN method

In consideration of the highly nonlinear relationship between the n -dimensional vector \mathbf{x} and the calculated signature $\mathbf{f}(\mathbf{x})$, the inverse problem shown in Eq. (2) is usually solved iteratively until an optimal result $\hat{\mathbf{x}}$ is reached. If the result of the current i th iteration is represented by $\mathbf{x}^{(i)}$, the necessary condition for $\mathbf{x}^{(i+1)}$ to minimize $F(\mathbf{x})$ as shown in Eq. (1) is that $\nabla F(\mathbf{x}^{(i+1)}) = \mathbf{0}$. The relationship between $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(i+1)}$ can be expressed as $\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \Delta \mathbf{x}^{(i)}$, where $\Delta \mathbf{x}^{(i)}$ is the parameter departure vector. By using the Taylor expansion, we can approximate the gradient in the vicinity of $\mathbf{x}^{(i)}$ as

$$\nabla F(\mathbf{x}^{(i+1)}) = \nabla F(\mathbf{x}^{(i)} + \Delta \mathbf{x}^{(i)}) = \nabla F(\mathbf{x}^{(i)}) + \nabla^2 F(\mathbf{x}^{(i)}) \Delta \mathbf{x}^{(i)}, \quad (3)$$

where $\nabla F(\mathbf{x}^{(i)})$ is the gradient of $F(\mathbf{x})$ at $\mathbf{x}^{(i)}$, which can be written in matrix notation as

$$\nabla F(\mathbf{x}^{(i)}) = 2\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \Delta \mathbf{y}^{(i)}, \quad (4)$$

where $\mathbf{J}(\mathbf{x}^{(i)})$ is the $m \times n$ Jacobian with respect to $\mathbf{x}^{(i)}$, and $\Delta \mathbf{y}^{(i)}$ is the residual column vector given by

$$\Delta \mathbf{y}^{(i)} = \mathbf{y} - \mathbf{f}(\mathbf{x}^{(i)}). \quad (5)$$

The term $\nabla^2 F(\mathbf{x}^{(i)})$ in Eq. (4) is the $n \times n$ Hessian matrix at $\mathbf{x}^{(i)}$. Since the calculation of the Hessian matrix is very time-consuming, it is usually approximated as

$$\nabla^2 F(\mathbf{x}^{(i)}) \approx 2\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)}). \quad (6)$$

Substituting Eq. (4) and Eq. (6) into Eq. (3) and letting Eq. (3) equal zero, we will have

$$\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)}) \Delta \mathbf{x}^{(i)} = -\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \Delta \mathbf{y}^{(i)}. \quad (7)$$

Further we will have

$$\Delta \mathbf{x}^{(i)} = -[\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)})]^{-1} \mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \Delta \mathbf{y}^{(i)}, \quad (8)$$

which assumes that the expression $\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)})$ is nonsingular. By conducting Eq. (3) ~ Eq. (8) iteratively, we can obtain an optimal result $\hat{\mathbf{x}}$.

2.3 Robust solution to the inverse problem based on the principle of LTS

By using the expressions of $\tilde{\mathbf{J}}^{(i)} = \mathbf{w}^{1/2} \mathbf{J}(\mathbf{x}^{(i)})$ and $\tilde{\Delta \mathbf{y}}^{(i)} = \mathbf{w}^{1/2} \Delta \mathbf{y}^{(i)}$, we can rewritten Eq. (8) as

$$\Delta \mathbf{x}^{(i)} = -[\tilde{\mathbf{J}}^{(i)T} \tilde{\mathbf{J}}^{(i)}]^{-1} \tilde{\mathbf{J}}^{(i)T} \tilde{\Delta \mathbf{y}}^{(i)}. \quad (9)$$

Obviously, Eq. (9) is the LSQ solution of the compatible system of equations:

$$-\tilde{\mathbf{J}}^{(i)} \Delta \mathbf{x}^{(i)} = \tilde{\Delta \mathbf{y}}^{(i)*}, \quad (10)$$

where $\tilde{\Delta \mathbf{y}}^{(i)*}$ is the optimal approximation of $\tilde{\Delta \mathbf{y}}^{(i)}$ in the subspace $\mathfrak{R}(-\tilde{\mathbf{J}}^{(i)})$, namely, $\tilde{\Delta \mathbf{y}}^{(i)*}$ satisfies

$$\|\tilde{\Delta \mathbf{y}}^{(i)} - \tilde{\Delta \mathbf{y}}^{(i)*}\|_2 = \inf_{\Delta \mathbf{x}^{(i)} \in \Theta} \|\tilde{\Delta \mathbf{y}}^{(i)} + \tilde{\mathbf{J}}^{(i)} \Delta \mathbf{x}^{(i)}\|_2. \quad (11)$$

Here $\|\cdot\|_2$ and Θ denote the 2-norm and space of $\Delta \mathbf{x}^{(i)}$, respectively. Expression (10) can be explained as the search of a hyperline with the optimal slope $\Delta \mathbf{x}^{(i)}$ that can best fits the m data pairs $(-\tilde{\mathbf{J}}_k, \tilde{\Delta \mathbf{y}}_k)$, where $\tilde{\mathbf{J}}_k$ and $\tilde{\Delta \mathbf{y}}_k$ represent the k th row of $-\tilde{\mathbf{J}}^{(i)}$ and k th element of $\tilde{\Delta \mathbf{y}}^{(i)}$, respectively. The row number of $-\tilde{\mathbf{J}}^{(i)}$, i.e., variable m , represents the number of wavelength points. If the measurement errors in $\tilde{\Delta \mathbf{y}}^{(i)}$ are relatively small, Eq. (8) usually ensures the relative accurate value of $\Delta \mathbf{x}^{(i)}$. However, in consideration of the superimposed effect of different error sources, the expected high linearity of the m data pairs does not always hold. The outliers will largely affect the estimation of $\Delta \mathbf{x}^{(i)}$, therefore the iterative phase of GN algorithm shown in Eq. (8) can only give the rough estimation of $\Delta \mathbf{x}^{(i)}$, which will reduce the measurement accuracy according to error accumulation. In this paper, we propose to use a correction based on the principle of LTS to eliminate h data pairs that are defined as outliers. LTS method is proposed by Rousseeuw^[19] and has been demonstrated that can yield a reliable analysis of regression data^[22]. For simplicity we first express the m data pairs as a dataset $\mathbf{Z} = \{(-\tilde{\mathbf{J}}_k, \tilde{\Delta \mathbf{y}}_k); k=1, \dots, m\}$ and represents the k th residual by $r_k = \tilde{\Delta \mathbf{y}}_k + \tilde{\mathbf{J}}_k \Delta \mathbf{x}^{(i)}$, then the LTS estimator is defined as the optimal $\Delta \mathbf{x}^{(i)}$ denoted by $\Delta \hat{\mathbf{x}}^{(i)}$ in this paper that minimizes

$$\Delta \hat{\mathbf{x}}^{(i)} = \arg \min_{\Delta \mathbf{x}^{(i)} \in \Theta} \left[\sum_{k=1}^{m-h} (r^2)_{k:m} \right]. \quad (12)$$

where $(r^2)_{1:m} \leq (r^2)_{2:m} \leq \dots \leq (r^2)_{m:m}$ are the ordered squared residuals. This is equivalent to finding the $(m - h)$ data pairs with smallest LSQ function values. The LTS estimate is the LSQ fitting to these $(m - h)$ data pairs. For the choice of h value, it is related to the LTS breakdown value and our priori-knowledge of the nonlinear data pairs number, in this paper we set h as 0.15. Since expression (6) is a regression problem, which can be solved effectively by gradient-based

algorithms. It is expected that the result $\Delta\hat{\mathbf{x}}^{(i)}$ obtained by the proposed method is more robust and accurate than the conventional one.

In summary, the proposed method introduces an additional “inner” LTS regression to replace the conventional GN iteration step as presented in Eq. (8) to directly eliminate the outliers that affect the LSQ fitting. Since the LTS regression is regardless of the forward modeling simulator, it can be done very quickly and presents no remarkable influence on the time cost.

3. RESULTS

3.1 Sample description and measurement setup

The sample under measurement is a deep-etched multilayer grating, whose cross-section image obtained by a transmission electron microscopy (TEM) (TE20, TEM.FEI Co.) and the corresponding geometrical model are presented in Fig. 1(a) and Fig. 1(b), respectively. As can be seen in Fig. 1(b), this sample consists of Si, SiO₂, and nitride Si₃N₄ trapezoidal gratings from bottom to top, and it can be characterized by 7 geometrical parameters that are D_1 , H_1 , D_2 , H_2 , D_3 , H_3 and pitch P . A self-developed dual-rotating compensator Mueller matrix ellipsometer (DRC-MME) prototype suitable from ultraviolet to infrared spectrum^[23] is used for demonstration. Data analysis is performed using the in-house developed optical modeling software based on RCWA. In the following contents, D_1 , H_1 , D_2 , H_2 , D_3 , and H_3 are measurands, while the pitch P is fixed at 154 nm.

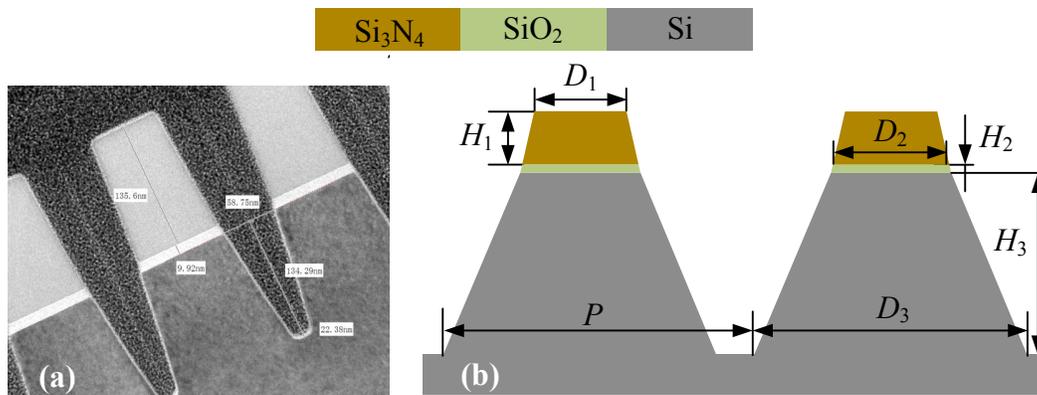


Fig. 1 The (a) cross-section TEM image and (b) geometrical model of the investigated deep-etched multilayer grating.

3.2 Numerical results

In this section, we simulate the measurement process of the multilayer grating by firstly calculating its corresponding Mueller matrix. The true values of profile parameters D_1 , H_1 , D_2 , H_2 , D_3 , and H_3 of the grating are set as 75 nm, 135 nm, 90 nm, 10 nm, 130 nm and 135 nm respectively. The incident and azimuthal angles are set as 55° and 0° respectively, and the wavelength varies from 200 nm to 800 nm with the resolution set as 10 nm. Then the simulated normal distributed errors are added into the calculated Mueller matrix to form the “measurement” signature. The standard deviation or noise level of the simulated normal distributed errors at a specific wavelength is set as a fraction of root-mean-square (rms) in the Mueller matrix over the full wavelength range of interest. The fractions of the wavelengths differ with each other, but are all within the range of 1% ~ 5%. Besides the random errors, system errors such as the limited spectral resolution of the monochromator and finite numerical aperture may also have an influence on the measurement accuracy^[15], it is therefore their effect should be taken into consideration in this paper. Moreover, since the standard deviations of measurement errors at all the wavelength points are obtained by a pre-established noise model, the biases between the actual standard deviations and the estimated ones are inevitable. Hence, we will add a perturbation at each true standard deviation of a wavelength point to simulate the estimated standard deviation. Each perturbation scale is randomly chosen within the range of $\pm 20\%$ of the corresponding true standard deviation.

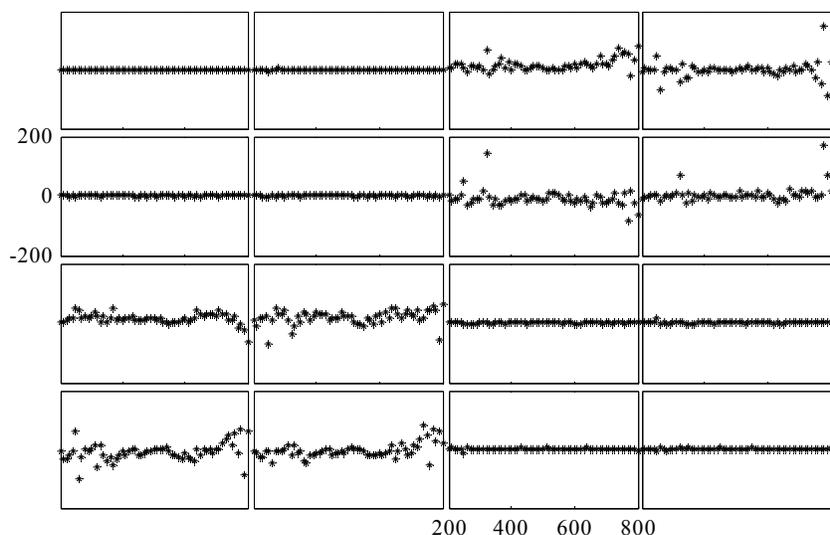


Fig. 2 Weighted measurement errors of Mueller matrix elements. The Mueller matrix elements are normalized to m11. The horizontal axes, varying from 200 to 800 nm with an increment of 10 nm, denote the wavelengths, and the vertical axes, varying from -200 to 200, denotes the associate weighted measurement errors.

TABLE 1. Comparison of parameters extracted from GN and our proposed methods

	D_1 (nm)	H_1 (nm)	D_2 (nm)	H_2 (nm)	D_3 (nm)	H_3 (nm)
True	75	135	90	10	130	135
GN	76.6	136.6	87.1	8.5	133.2	133.6
Proposed	75.0	134.3	90.0	11.1	130.0	134.6

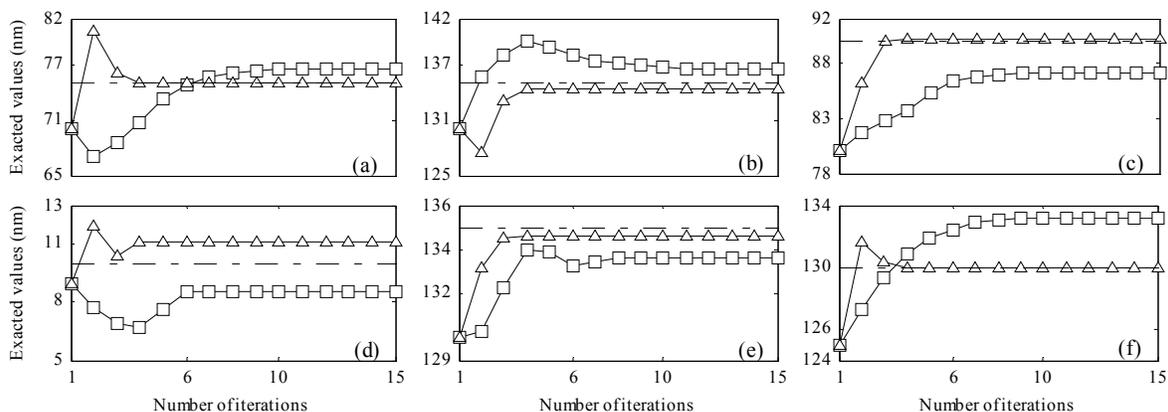


Fig. 3 Iteration results of (a) D_1 , (b) H_1 , (c) D_2 , (d) H_2 , (e) D_3 , and (f) H_3 . Black squares and triangles represent the values obtained by the GN method and our proposed method respectively. Black dash dotted lines represent the true values.

To demonstrate the existence of outlying “measurement” errors, we first present the weighted “measurement” errors in Fig. 2. The weighted “measurement” errors are the “measurement” errors divided by their corresponding standard deviations at all the wavelength points. Obviously, we can find there are some outlying “measurement” errors in the off-diagonal Mueller matrix elements of Fig. 2, and the some absolute values of whom are even as large as 200. Then the GN method and our proposed method are used to extract the profile parameters from the “measurement” signature. We present the extracted results in Table 1, in which we can clearly find that the proposed method leads to the more accurate results than that of GN method. We also present the iteration results of GN and our proposed methods in Fig. 3. As expected, we can find that at each iteration step the iterative result of our proposed method is different from that of the conventional GN method. The above simulation has demonstrated the correctness of our proposed method.

3.3 Experimental results

In the following content we will experimentally demonstrate the improved grating reconstruction by our proposed method. The Mueller matrices (normalized to m11) were measured at 61 points over wavelengths ranging from 200 to 800 nm with the resolution set as 10 nm, and the incidence angle was fixed at 55°. While for the azimuthal angle, we treated it as a floating parameter since in our DRC-MME prototype there is no positioning device on the rotating stage yet and the zero angle cannot be guaranteed accurately by the manual mode. The pitch P is fixed at 154 nm, and the profile parameters as well as the azimuthal angle φ are under measurement. The initial values of geometrical parameters and azimuthal angle are chosen as the nominal ones and zero, respectively.

TABLE 2. Comparison of parameters extracted from MME and TEM measurement

	D_1 (nm)	H_1 (nm)	D_2 (nm)	H_2 (nm)	D_3 (nm)	H_3 (nm)	φ (°)
TEM	75.01	135.60	86.90	9.92	123.13	134.29	
GN	73.11	133.45	88.44	13.42	127.34	130.21	-2.30
Proposed	75.13	136.33	87.79	8.48	127.21	132.83	-2.75

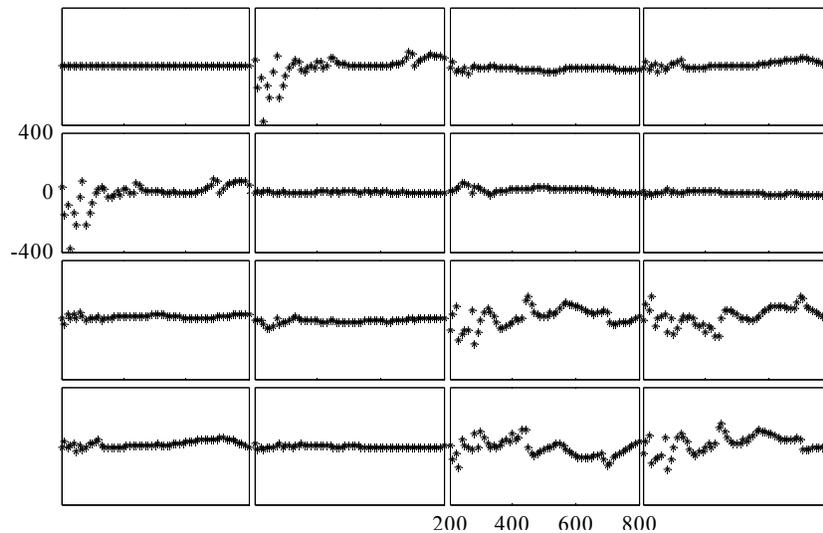


Fig. 4 Experimental weighted fitting differences of Mueller matrix elements. The Mueller matrix elements are normalized to m11. The horizontal axes, varying from 200 to 800 nm with an increment of 10 nm, denote the wavelengths, and the vertical axes, varying from -400 to 400, denotes the associate weighted fitting differences.

Firstly we will validate some of the measurement errors in the measured Mueller matrix of the deep-etched multilayer grating are abnormally distributed. We calculate the weighted fitting differences of the measured and best fitted Mueller matrix elements, as shown in Fig. 4. The best fitted Mueller matrix elements correspond to the extracted values of geometrical parameters that are obtained by GN method. Note that in practice to accurately obtain the measurement errors is impossible, thus the fitting differences between the measured Mueller matrix and the best fitted one is a compromise, but it is still able to indirectly reflect the actual statistics property of the measurement errors in the measured signature. As can be seen in Fig. 4, many data points in the diagonal elements of Mueller matrix are obviously out of the range of $-3 \sim 3$, and a part of which are even larger than 300. Moreover, we could also notice that most of the large abnormally distributed data points locate in the Mueller matrix elements of m12, m21, m33, m34, m43 and m44.

We use the proposed method and the conventional GN method to extract the geometrical parameters of the deep-etched multilayer grating respectively, and present the fitted results in Table 2. As expected, all of the fitted geometrical parameters obtained by our proposed method are closer to the TEM measured ones than that obtained by GN method. Specifically, we can find that the values of D_1 , H_1 , D_2 , H_2 , D_3 , and H_3 obtained by our proposed method are 1.82 nm, 1.45 nm, 0.64 nm, 2.08 nm, 0.12 nm and 2.59 nm closer to the TEM measured values than that by GN method, respectively. Moreover, the relative accuracy improvements of parameters D_1 , H_1 , D_2 , H_2 , D_3 , and H_3 are 95%, 67%,

41%, 59%, 3% and 62% respectively. The relative accuracy improvement is defined as $e = \frac{\left| |x_{\text{GN}} - x_{\text{TEM}}| - |x_{\text{Robust}} - x_{\text{TEM}}| \right|}{|x_{\text{GN}} - x_{\text{TEM}}|}$,

where x_{GN} , x_{TEM} and x_{Robust} are the measured values by GN method, TEM and our proposed method respectively. The above results have demonstrated that the measurement errors in the measured signature cannot be treated as normally distributed, and it is necessary to use the principle of LTS to suppress the effect of outlying measurement errors.

4. CONCLUSIONS

In summary, we have demonstrated that outlying measurement errors will reduce the accuracy of $\Delta \mathbf{x}^{(i)}$ estimation by simulation and experiment. Moreover, we have proposed a method based on the principle of LTS to detect and eliminate outliers for estimating the more accurate result of each iteration step, by which the final measurement accuracy has been improved when compared with the conventional GN method. The proposed method enables us to obtain the nanostructure reconstruction with high accuracy without perform the complicated and time-consuming processes such as measurement configuration optimization and limited NA correction. We believe the present work will provide a different point of view for the accurate nanostructure reconstruction in IC manufacturing.

ACKNOWLEDGMENT

This work was funded by the National Natural Science Foundation of China (Grant Nos. 51475191 and 51405172), the National Instrument Development Specific Project of China (Grant No. 2011YQ160002), and the Program for Changjiang Scholars and Innovative Research Team in University of China (Grant No. IRT13017). The authors would like to thank the facility support of the Center for Nanoscale Characterization and Devices, Wuhan National Laboratory for Optoelectronics (WLNO) (Wuhan, China).

REFERENCES

- [1] M. G. Faruk, S. Zangoie, M. Angyal, D. K. Watts, M. Sendelbach, L. Economikos, P. Herrera, and R. Wilkins, "Enabling scatterometry as an in-line measurement technique for 32nm BEOL application," *IEEE Trans. Semicond. Manuf.* **24**, 499–512 (2011).
- [2] R. L. Jones, T. Hu, C. L. Soles, E. K. Lin, R. M. Reano, S. W. Pang, and D. M. Casa, "Real-time shape evolution of nanoimprinted polymer structures during thermal annealing," *Nano Lett.* **6**, 1723–1728 (2006).
- [3] H.-J. Lee, C. L. Soles, H. W. Ro, R. L. Jones, E. K. Lin, W. Wu, and D. R. Hines, "Nanoimprint pattern transfer quality from specular x-ray reflectivity," *Appl. Phys. Lett.* **87**, 263111 (2005).
- [4] D. Fuard, C. Perret, V. Farys, C. Gourgon, and P. Schiavone, "Measurement of residual thickness using scatterometry," *J. Vac. Sci. Technol. B* **23**, 3069–3074 (2005).
- [5] H. J. Patrick, T. A. Gemer, Y. F. Ding, H. W. Ro, L. J. Richter, and C. L. Soles, "Scatterometry for in situ measurement of pattern flow in nanoimprinted polymers," *Appl. Phys. Lett.* **93**, 233105 (2008).
- [6] H. Huang and F. Terry, Jr., "Spectroscopic ellipsometry and reflectometry from gratings (scatterometry) for critical dimension measurement and in situ, real-time process monitoring," *Thin Solid Films* **455-456**, 828–836 (2004).
- [7] Y. N. Kim, J. S. Paek, S. Rabello, S. Lee, J. T. Hu, Z. Liu, Y. D. Hao, and W. McGahan, "Device based in-chip critical dimension and overlay metrology," *Opt. Express* **17**, 21336–21343 (2009).
- [8] Y. Ohkawa, Y. Tsuji, and M. Koshiba, "Analysis of anisotropic dielectric grating diffraction using the finite element method," *J. Opt. Soc. Am. A* **13**, 1006–1012 (1996).
- [9] Y. Nakata and M. Koshiba, "Boundary-element analysis of plane-wave diffraction from groove-type dielectric and metallic gratings," *J. Opt. Soc. Am. A* **7**, 1494–1502 (1990).
- [10] H. Ichikawa, "Electromagnetic analysis of diffraction gratings by the finite-difference time-domain method," *J. Opt. Soc. Am. A* **15**, 152–157 (1998).

- [11] M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *J. Opt. Soc. Am. A* **12**, 1068–1076 (1995).
- [12] L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A* **13**, 1870–1876 (1996).
- [13] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, 3rd ed. (Springer, 2013).
- [14] R. C. Geary, "Testing for normality," *Biometrika* **34**, 209–242 (1947).
- [15] T. A. Germer, and H. J. Patrick, "Effect of Bandwidth and Numerical Aperture in Optical Scatterometry," *Proc. SPIE* **7638**, 76381F (2010).
- [16] M. A. Henn, H. Gross, F. Scholze, M. Wurm, C. Elster, and M. Bär, "A maximum likelihood approach to the inverse problem of scatterometry," *Opt. Express* **20**, 12771–12786 (2012).
- [17] J. L. Zhu, S. Y. Liu, X. G. Chen, C. W. Zhang, and H. Jiang, "Robust solution to the inverse problem in optical scatterometry," *Opt. Express* **22**, 22031–22042 (2014).
- [18] J. L. Zhu, S. Y. Liu, H. Jiang, C. W. Zhang, and X. G. Chen, "Improved deep-etched multilayer grating reconstruction by considering etching anisotropy and abnormal errors in optical scatterometry," *Opt. Lett.* **40**, 471–474 (2015).
- [19] P. J. Rousseeuw, "Least median of squares regression," *J. Amer. Statist. Assoc.* **79**, 871–880 (1984).
- [20] P. J. Rousseeuw and K. V. Driessen, "Computing LTS regression for large data sets," *Data Min. Knowl. Disc.* **12**, 29–45 (2006).
- [21] X. G. Chen, S. Y. Liu, C. W. Zhang, and H. Jiang, "Measurement configuration optimization for accurate grating reconstruction by Mueller matrix polarimetry," *J. Micro/Nanolith. MEMS MOEMS* **12**, 033013 (2013).
- [22] P. J. Rousseeuw and B. C. V. Zoomeren, "Unmasking multivariate outliers and leverage points," *J. Amer. Statist. Assoc.* **85**, 633–651 (1990).
- [23] S. Y. Liu, X. G. Chen, and C. W. Zhang, "Mueller matrix polarimetry: A powerful tool for nanostructure metrology," *ECS Trans.* **60**, 237–242 (2014).