IOP Publishing Journal of Optics
J. Opt. 18 (2016) 015605 (6pp) doi:10.1088/2040-8978/18/1/015605

Improved nanostructure reconstruction by performing data refinement in optical scatterometry

Jinlong Zhu[1], Hao Jiang[1,3], Yating Shi[1], Xiuguo Chen[1], Chuanwei Zhang[1,2] and Shiyuan Liu[1,2,3]

[1] State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, People's Republic of China
[2] Wuhan Eoptics Technology Co. Ltd., Wuhan, Hubei 430075, People's Republic of China

E-mail: hjiang@hust.edu.cn and shyliu@hust.edu.cn

Received 7 July 2015, revised 30 October 2015
Accepted for publication 3 November 2015
Published 11 December 2015

Abstract

Recently, we have indirectly demonstrated that nanostructure reconstruction accuracy is degraded by the outliers in optical scatterometry, and we have applied the robust estimation method to suppress these outliers. However, the existence of a possible heavy masking effect could result in the risk of low measurement accuracy, since the detection of outliers is simply based on the judgment of residual value. In this work, a novel method is introduced to directly detect outliers, which can provide the intuitional display of outliers in a two-dimensional coordinate system. Moreover, a robust correction step based on the principle of least trimmed squared estimator regression is proposed to replace the conventional Gauss–Newton iteration step, by which the more reliable and accurate nanostructure reconstruction is achieved. The improved reconstruction of a one-dimensional etched Si grating has demonstrated the feasibility of the proposed methods.

Keywords: optical scatterometry, Mueller matrix ellipsometry, inverse problem, least trimmed squared estimator, measurement accuracy

(Some figures may appear in colour only in the online journal)

1. Introduction

To meet the requirement of high-volume manufacturing in the semiconductor industry, it is of great importance to perform accurate and precise semiconductor nanostructure metrology [1, 2]. Among all the metrology techniques, optical scatterometry [3–5] can be regarded as the state-of-the-art one according to its inherent advantages such as nondestruction, high sampling rate, large aerial coverage, and low cost [6, 7].

Optical scatterometry is essentially a model-based technique whose success relies heavily on the solvent of the corresponding inverse problem [8]. The inverse problem of optical scatterometry is usually formulated as the nonlinear least square (LSQ) minimization with the target of finding the most similar theoretical signature to the measured one. The most widely used regression algorithms in optical scatterometry are the gradient-based deterministic ones, in consideration of their relatively higher efficiency when compared with the Monte Carlo algorithms [9, 10]. The gradient-based algorithms, such as the Gauss–Newton (GN) method, have iteration steps equivalent to an LSQ fitting with the objective of finding a hyperline that can best fit the data pairs, where each data pair is the combination of a current residual (observation) and a specific row of the linear operator (variable) [11]. It is usually assumed that the measurement errors contained in the residual are normally distributed with a small amplitude; therefore, the current iteration step of the GN method can usually give a reliable estimation of the next iteration result. However, we recently have indirectly demonstrated the irrationality of the normality assumption

[3] Authors to whom any correspondence should be addressed.

and the existence of large measurement errors due to the superposition of the errors from different sources. We have proposed a robust method based on the principle of robust estimation to suppress these outliers [12, 13]. Although the improved profile reconstruction can be achieved, the success of the robust estimation method highly relies on the detection of outliers, since its strategy is to allot the relatively small weights to the outliers. In the robust estimation method, the detection of outliers is simply based on the judgment of residual value; i.e., the data pair corresponding to the large residual value is treated as the outlier. Therefore, once there exists a heavy masking effect, which refers to the fact that some real outliers might be masked due to the distorted fitting curve when performing the curve-fitting procedure, the robust method still suffers from the risk of low measurement accuracy [14].

In this work, we stand on the viewpoint of least trimmed squared (LTS) estimator [15, 16] to solve the inverse problem in optical scatterometry. The LTS method treats each GN iteration as an LSQ fitting over only a subset of the whole data pairs, which means that the rest of the data pairs are qualified as outliers and are rejected in the LSQ fitting phase. Unlike the robust estimation method, in which the detection of outliers is simply based on the judgment of residual values within each GN iteration step, the LTS method dynamically adjusts and ranks these residuals, by which the reliable mapping between the outliers and the residual values can be constructed, and thus the masking effect can be avoided effectively.

2. Theory

The inverse problem in optical scatterometry is usually formulated as the minimization of an LSQ function

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\Omega} \left\{ [\mathbf{y} - \mathbf{f}(\mathbf{x})]^{\mathrm{T}} \mathbf{w} [\mathbf{y} - \mathbf{f}(\mathbf{x})] \right\}, \tag{1}$$

where $\mathbf{x}$, $\mathbf{\Omega}$, $\mathbf{y}$, $\mathbf{f}(\mathbf{x})$, and w represent the measurand, the parameter domain of $\mathbf{x}$, the measured signature, the simulated signature, and the weight matrix, respectively. To solve the inverse problem, the GN method is used to iteratively linearize the LSQ function, and the iteration result of the ith step is expressed as

$$\Delta\mathbf{x}^{(i)} = -\left[\mathbf{J}\left(\mathbf{x}^{(i)}\right)^{\mathrm{T}} \mathbf{w} \mathbf{J}\left(\mathbf{x}^{(i)}\right)\right]^{-1} \mathbf{J}\left(\mathbf{x}^{(i)}\right)^{\mathrm{T}} \mathbf{w} \Delta\mathbf{y}^{(i)}. \tag{2}$$

Here the terms $\Delta\mathbf{x}^{(i)}$, $\mathbf{J}(\mathbf{x}^{(i)})$, and $\Delta\mathbf{y}^{(i)}$ represent the parameter departure vector, the Jacobian, and the residual column vector of the ith iteration step, respectively, and the superscript T represents the transpose. By using the expressions $\tilde{\mathbf{J}}^{(i)} = \mathbf{w}^{1/2}\mathbf{J}(\mathbf{x}^{(i)})$ and $\Delta\tilde{\mathbf{y}}^{(i)} = \mathbf{w}^{1/2}\Delta\mathbf{y}^{(i)}$, we rewrite equation (2) as

$$\Delta\mathbf{x}^{(i)} = -\left[\tilde{\mathbf{J}}^{(i)\mathrm{T}} \tilde{\mathbf{J}}^{(i)}\right]^{-1} \tilde{\mathbf{J}}^{(i)\mathrm{T}} \Delta\tilde{\mathbf{y}}^{(i)}. \tag{3}$$

Here we should note that the GN method itself is one kind of iterative regularization method, and the iteration count plays the role of a regularization parameter the same as that in the well-known Tikhonov regularization [17]. Hence, if the iteration count is properly chosen, the GN method is usually convergent to the satisfactory result. Obviously, equation (3) is the LSQ solution of the compatible system of equations

$$-\tilde{\mathbf{J}}^{(i)} \Delta\mathbf{x}^{(i)} = \Delta\tilde{\mathbf{y}}^{(i)*}, \tag{4}$$

where $\Delta\tilde{\mathbf{y}}^{(i)*}$ is the optimal approximation of $\Delta\mathbf{y}^{(i)}$ in the subspace $\Re(-\tilde{\mathbf{J}}^{(i)})$ under 2-norm; namely, $\Delta\tilde{\mathbf{y}}^{(i)*}$ satisfies

$$\|\Delta\tilde{\mathbf{y}}^{(i)} - \Delta\tilde{\mathbf{y}}^{(i)*}\|_2 = \inf_{\Delta\mathbf{x}^{(i)}\in\Theta} \|\Delta\tilde{\mathbf{y}}^{(i)} + \tilde{\mathbf{J}}^{(i)}\Delta\mathbf{x}^{(i)}\|_2, \tag{5}$$

where $\|\cdot\|$ and Θ denote the 2-norm and space of $\Delta\mathbf{x}^{(i)}$, respectively. Equation (5) can be explained as the search of a hyperline with the optimal slope $\Delta\mathbf{x}^{(i)}$ that best fits the m data pairs $(-\tilde{\mathbf{J}}_k, \Delta\tilde{y}_k)$, where $-\tilde{\mathbf{J}}_k$ and $\Delta\tilde{y}_k$ represent the kth row of $-\tilde{\mathbf{J}}^{(i)}$ and kth element of $\Delta\tilde{\mathbf{y}}^{(i)}$, respectively. The row number of $-\tilde{\mathbf{J}}^{(i)}$, i.e., variable m, represents the number of wavelength points. If the measurement errors in $\Delta\mathbf{y}^{(i)}$ are relatively small, equation (3) usually ensures the relatively accurate value of $\Delta\mathbf{x}^{(i)}$. However, in consideration of the superimposed effect of different error sources, some of the m data pairs may obviously deviate from the majority. These deviated data pairs are called outliers, which will affect the estimation of $\Delta\mathbf{x}^{(i)}$ remarkably; therefore, equation (3) can only give the rough estimation of $\Delta\mathbf{x}^{(i)}$, which will reduce the measurement accuracy according to error accumulation. Here we propose to correct equation (3) based on the principle of LTS to eliminate h data pairs that are defined as outliers. The LTS method is proposed by Rousseeuw *et al* [14] and has demonstrated that it is capable of yielding a reliable analysis of regression data [18]. For simplicity we first express the m data pairs as a dataset $\mathbf{Z} = \left\{\left(-\tilde{\mathbf{J}}_k, \Delta\tilde{y}_k\right);\ k = 1, ..., m\right\}$ and represent the kth residual by $r_k = \Delta\tilde{y}_k + \tilde{\mathbf{J}}_k\Delta\mathbf{x}^{(i)}$. Them the LTS estimator is defined as the optimal $\Delta\mathbf{x}^{(i)}$ denoted by $\Delta\hat{\mathbf{x}}^{(i)}$, which minimizes [14],

$$\Delta\hat{\mathbf{x}}^{(i)} = \arg\min_{\Delta\mathbf{x}^{(i)}\in\Theta} \left[\sum_{j=1}^{m-h} \left(r^2\right)_{j:m}\right], \tag{6}$$

where $(r^2)_{1:m} \leqslant (r^2)_{2:m} \leqslant \ldots \leqslant (r^2)_{m:m}$ are the ordered squared residuals, and the term $(r^2)_{j:m}$ represents the jth squared residual in the ordered sequence containing m elements. This is equivalent to finding the $(m–h)$ subset with the smallest LSQ function values. The LTS estimate is then the LSQ fitting to these $(m–h)$ data pairs. For the choice of h value, it is related to the LTS breakdown value and the *a priori* knowledge of the nonlinear data pairs' number; for example, when h is set as 0.5 or 0.25, then this represents a belief that there is a relatively large or small number of outliers in the data pairs. Currently, there is no general accurate model to predict the value of h but only the empirical one for some specific cases, such as Rousseeuw's formula $h = (m - p - 1)/2$ [14], where p is the number of variables in $\mathbf{x}$; here in this article, p equals 3. Here we set h as 0.15. Note that equation (6) is also a regression problem, which can be solved effectively by gradient-based algorithms. Different from the robust estimation method that directly allots the

relatively small weights to the data pairs corresponding to the large residuals, the regression in equation (6) can effectively avoid the masking effect on the judgment of outliers [15]. Consequently, the result $\Delta\hat{\mathbf{x}}^{(i)}$ obtained by the proposed method is more robust and accurate than the conventional one, since the h outliers have been rejected, and only the 'pure' (m–h) data pairs are used for the LSQ fitting. The feasibility of the proposed method is built on top of the belief that some data pairs $(-\tilde{\mathbf{J}}_k, \Delta\tilde{y}_k)$ are not regular, and the masking effect is noticeable; therefore, we should demonstrate the existence of the outliers and masking effect firstly. Note that we cannot detect the outliers by visual inspection, since the problem considered in this article is in high dimensions; therefore, we will detect the outliers by plotting standardized LTS residual $\mathrm{SR}(r_k)$ versus robust distance $\mathrm{RD}(-\tilde{\mathbf{J}}_k)$. Here $\mathrm{SR}(r_k)$ and $\mathrm{RD}(-\tilde{\mathbf{J}}_k)$ are defined as

$$\mathrm{SR}(r_k) = \frac{r_k}{c_{h,m}\sqrt{\dfrac{1}{m-h}\sum_{j=1}^{m-h}\left(r^2\right)_{j:m}}}, \tag{7a}$$

$$\mathrm{RD}\left(-\tilde{\mathbf{J}}_k\right) = \sqrt{\left(\tilde{\mathbf{J}}_k + \hat{\boldsymbol{\mu}}\right)^{\mathrm{T}}\hat{\Sigma}^{-1}\left(\tilde{\mathbf{J}}_k + \hat{\boldsymbol{\mu}}\right)}, \tag{7b}$$

where r_k in equation (7a) is the kth residual corresponding to the kth wavelength point in a GN iteration step. The term $\sqrt{\frac{1}{m-h}\sum_{j=1}^{m-h}(r^2)_{j:m}}$ is calculated from the output of equation (6) that corresponds to the GN iteration step for the calculation of r_k, and $c_{h,m}$ makes the denominator part of equation (7a) consistent and unbiased at Gaussian error distributions. In equation (7b) $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are the minimum covariance determinant (MCD) estimator location and the scatter estimates [20]. The MCD method looks for the (m–h) observations (out of m) whose classical covariance matrix has the lowest possible determinant; thus $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are, respectively, the average of these (m–h) data pairs and a multiple of the covariance matrix of these (m–h) data pairs. To know more about MCD we refer the readers to reference [14].

3. Experiment

3.1. Measurement setup and sample description

The effectiveness of the proposed method is validated by the reconstruction of a one-dimensional (1D) etched Si grating, whose cross-section image, obtained by scanning electron microscopy (SEM), is shown in figure 1. We characterize such a grating by a symmetrical trapezoidal model with top critical dimension *TCD*, bottom critical dimension *BCD*, grating height *Hgt*, and period Λ. These parameters, obtained by SEM, are $TCD = 350$ nm, $Hgt = 472$ nm, and $BCD = 383$ nm. In the following experiments, parameters *TCD*, *Hgt*, and *BCD* are the ones that need to be extracted, while the grating period Λ is fixed at its nominal value 800 nm. A house-built dual-rotating compensator Mueller matrix ellipsometer [8], together with an in-house optical modeling software package based on rigorous coupled-wave

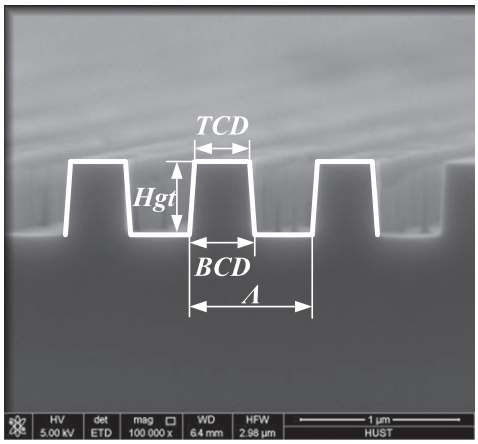

Figure 1. Cross-section image and parameterization of the one-dimensional etched Si grating.

analysis, is used to reconstruct the grating [19]. The Mueller matrices (normalized to m_{11}) are measured at 61 points over wavelengths ranging from 200–800 nm, which implies that the spectral resolution is selected as 10 nm. In order to simplify the formulation expression of the inverse problem in optical scatterometry, the measurands are unified and written as a vector $\mathbf{x}$; i.e., $\mathbf{x} = [TCD, Hgt, BCD]$.

3.2. Results and discussion

To demonstrate the existence of outliers and the masking effect at each iteration step of the GN algorithm experimentally, we obtain the measured Mueller matrix of the Si grating under a 50° incident angle and 0° azimuthal angle, which is different from the configuration shown in previous work [12, 13]. We present the regression outlier map for the m data pairs in figure 2 according to the above discussions, in which each circle represents a data pair. Figure 2 depicts the 1st $\sim$ 6th iteration steps of the GN algorithm. Data pairs for which the standardized absolute LS residual exceeds the cutoff $\sqrt{\chi^2_{1,0.975}}$ are considered to be outliers [14]. As can be seen in each sub-figure of figure 2, outliers do exist. Moreover, to demonstrate the existence of the masking effect, we pick out 20 outliers from the first GN iteration step and mark them by red squares, as shown in figure 2(a). With the increase of iteration numbers, we can find from figures 2(b)–(f) that some of the marked outliers from figure 2(a) enter the area between the two dotted lines, which means that the effect of true outliers is masked during the iterations. The masked outliers demonstrate that the fitting has already been distorted. Hence, to reduce the masking effect, it is of great importance to reject some outliers at the beginning of each iteration step, as presented in equation (6). Figure 3 presents the extraction results of *TCD*, *Hgt*, and *BCD* at each iteration step using the proposed method and the conventional GN method. As expected, the convergent values (350.0, 472.7, 386.0 nm) of *TCD*, *Hgt*, and *BCD* using the proposed method are closer to the SEM measured values when compared with those obtained by the conventional GN method, i.e., (348.0, 473.5,

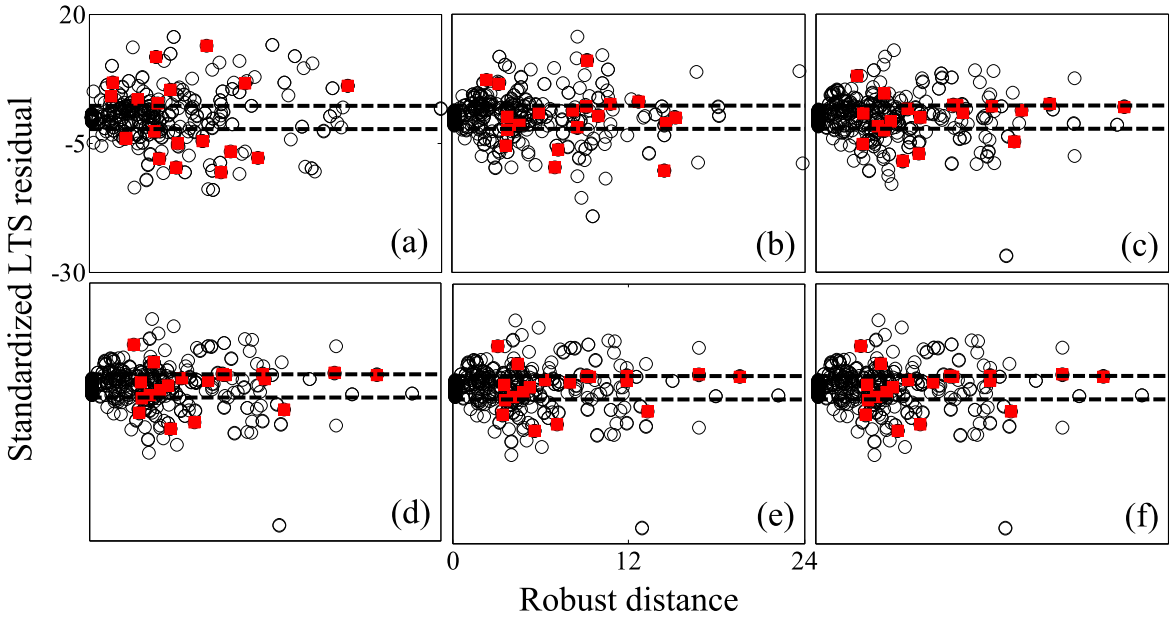

Figure 2. Regression outlier map of the m data pairs at different iteration steps. The red squares represent the artificially picked data pairs. Data pairs located outside of the area bounded by the dotted lines are treated as outliers. (a)–(f) denote the map of data pairs of the 1st to 6th iteration steps of the GN algorithm, respectively.

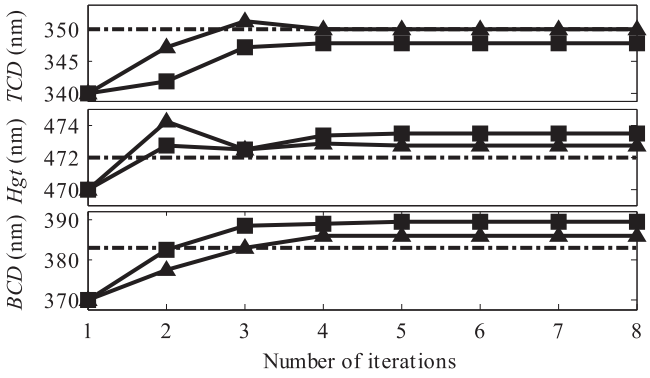

Figure 3. Iteration results of (a) *TCD*, (b) *Hgt*, and (c) *BCD*. The dash-dotted line, triangles, and squares represent the measured values by SEM, our proposed method, and the conventional GN method, respectively.

389.3 nm). Moreover, we can observe that the convergent result is obtained at the 4th iteration in the proposed method.

As is well known, it is extremely difficult to know the true values of the grating parameters in practice, which increases the difficulty of accuracy evaluation. Therefore the result presented in figure 3 is only a relative comparison of the extracted parameter accuracy. To further perform the accuracy evaluation and demonstrate the superiority of the proposed method, we extract the grating parameters under different incident angles θ and different azimuthal angles ϕ. Though such an evaluation method is a compromise indeed and cannot reflect the exact accuracy estimation, it has the capability to indirectly compare the accuracy among different measurement techniques or methods, as demonstrated by Novikova *et al* [21]. More specifically, we measure the Mueller matrices of the grating by fixing the azimuthal angle at 0° while varying the incident angle from 45° $\sim$ 65° with an increment of 5°. We also perform measurements by fixing the incident angle at 50° while varying the azimuthal angle from 10° $\sim$ 30° with an increment of 5°. All the measured signatures are inputted into the GN, the robust estimation [13], and the proposed methods to extract the geometrical parameters, respectively. We calculate the difference between the extracted geometrical values and the SEM-measured geometrical values. The absolute values of these differences are presented in figure 4.

As can be seen in figure 4(a), the proposed method presents a higher measurement accuracy than that of the conventional GN method. Moreover, we can also find that although most of the results obtained by robust estimation are better than those of the GN method, there still exist some violations, such as the values corresponding to incident angles 55° and 60° in figure 4(a). The reason for the lower accuracy obtained by the robust estimation method is that there exist relatively heavy masking effects, as emphasized above. The same trend can also be observed in figure 4(b). Moreover, we should point out that in figure 4(b) few extracted values of our proposed method are less accurate than that of the robust estimation method, such as the results under an azimuthal angle of 15° in figure 4(b). This phenomenon may be due to four reasons: first is the relatively small effect of masked outliers on the robust estimation; second, the value of factor h is underestimated so that some outliers are still kept at each iteration step; third, there is significant correlation between geometrical parameters in the nonplanar incidence mode; and last, the sampling multiple azimuthal angles when measuring line–space structures may increase sensitivity to line-edge roughness [22]. Here we evaluate the effect of factor h on the

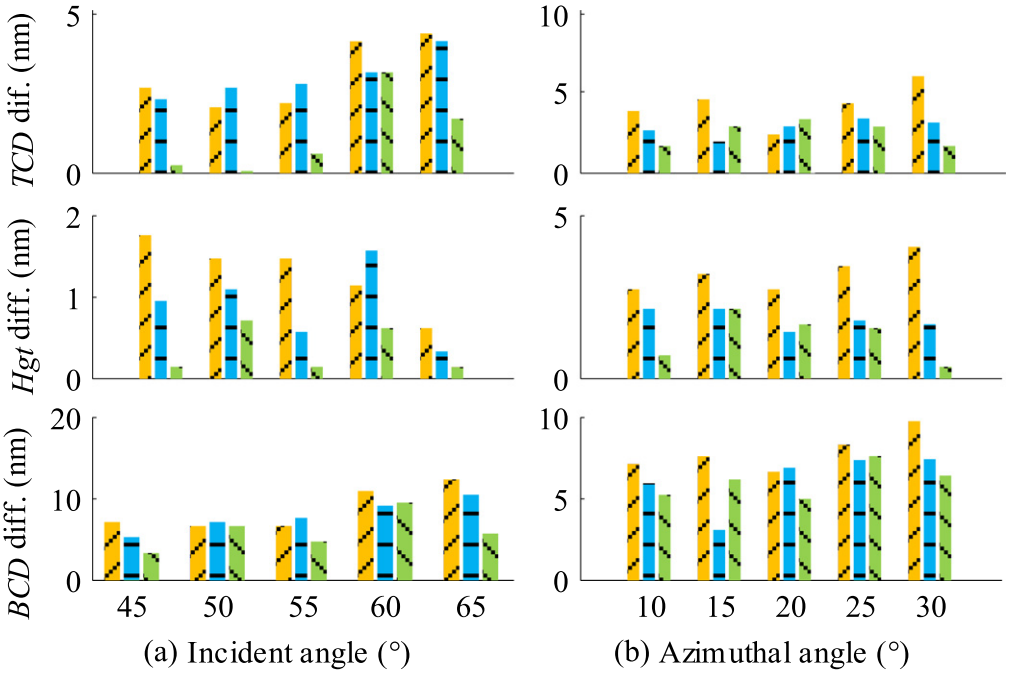

Figure 4. Absolute values of geometrical parameter differences between Mueller matrix ellipsometer and SEM under different (a) incident angles and (b) azimuthal angles. The bars marked with left, horizontal, and right slashes correspond to the values obtained by the conventional GN method, the robust estimation, and our proposed method, respectively.

Table 1. Comparison between the recalculated results by the proposed method and by the robust estimation.

	TCD (nm)			*Hgt* (nm)			*BCD* (nm)		
[θ (deg.), ϕ (deg.)]	SEM	Robust	Proposed	SEM	Robust	Proposed	SEM	Robust	Proposed
(50, 15)	350	348.2	347.7	472	474.1	474.6	383	385.9	387.5
(50, 20)	350	347.3	347.9	472	473.4	472.1	383	389.7	389.5
(50, 25)	350	346.6	347.1	472	473.7	473.2	383	390.3	388.3

measurement accuracy, since the underestimated h may be one of the major reasons accounting for the accuracy loss of the proposed method, as shown in figure 4(b). We set h as 0.35 and perform the inverse extraction using the proposed method corresponding to the azimuthal angles of 15°, 20°, and 25°, respectively. The increase of h indicates the belief that there may be more outliers in the measured data. The recalculated results using the proposed method are listed in table 1 with the results obtained by the robust estimation and SEM for comparison. It can be observed that the recalculated results by the proposed method are closer to the SEM measurements than by the robust estimation when the measurement configurations are (50°, 20°) and (50°, 25°), which to some extent has demonstrated that we can improve the accuracy by adjusting the factor h. However, it is worth noting that the robust estimation still presents the higher measurement accuracy when the measurement configuration (50°, 15°) is selected. This phenomenon may reiterate the facts that the effect of some other factors, such as the parameter correlation and the line-edge roughness, cannot be neglected. Thus, it is still of great importance to investigate all the potential reasons accounting for the loss of accuracy in the near future. Overall, the experimental results have demonstrated that the proposed method can realize the more accurate 1D grating reconstruction than that of the conventional GN method and presents superiority over the robust estimation method, since it can avoid the masking effect to some extent. Moreover, the proposed method also presents the potential in the reconstruction of more complex structures, such as the rectangles and L-shaped bidimensional gratings [23, 24], though further detailed investigation is necessary in the near future.

4. Conclusion

In summary, we have experimentally demonstrated that outliers reduce the accuracy of $\Delta\mathbf{x}^{(i)}$ estimation in solving the inverse problem in optical scatterometry. We have proposed a method based on the principle of LTS to directly detect and eliminate the outliers for estimating more accurate results in each iteration, by which not only the risk of the masking effect is avoided, but also the measurement accuracy has been improved.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (Grant Nos. 51475191, 51575214, and 51525502), the Natural Science Foundation of Hubei Province of China (Grant No. 2015CFA005), and the Program for Changjiang Scholars and Innovative Research Team in University of China (Grant No. IRT13017).

References

[1] Faruk M G, Zangooie S, Angyal M, Watts D K, Sendelbach M, Economikos L, Herrera P and Wilkins R 2011 Enabling scatterometry as an in-line measurement technique for 32 nm BEOL application *IEEE Trans. Semicond. Manuf.* **24** 499

[2] Postek M T and Lyons K W 2007 Instrumentation, metrology, and standards: key elements for the future of nanomanufacturing *Proc. SPIE* **6648** 664802

[3] Raymond C J, Murnane M R, Prins S L, Naqvi S S H, Hosch J W and McNeil J R 1997 Multiparameter grating metrology using optical scatterometry *J. Vac. Sci. Technol.* B **15** 361

[4] Niu X, Jakatdar N, Bao J and Spanos C J 2001 Specular spectroscopic scatterometry *IEEE Trans. Semicond. Manuf.* **14** 97

[5] Patrick H J, Gemer T A, Ding Y F, Ro H W, Richter L J and Soles C L 2008 Scatterometry for *in situ* measurement of pattern flow in nanoimprinted polymers *Appl. Phys. Lett.* **93** 233105

[6] Huang H and Terry F Jr 2004 Spectroscopic ellipsometry and reflectometry from gratings (scatterometry) for critical dimension measurement and *in situ*, real-time process monitoring *Thin Solid Films* **455–56** 828

[7] Kim Y N, Paek J S, Rabello S, Lee S, Hu J T, Liu Z, Hao Y D and McGahan W 2009 Device based in-chip critical dimension and overlay metrology *Opt. Express* **17** 21336

[8] Liu S Y, Chen X G and Zhang C W 2015 Development of a broadband Mueller matrix ellipsometer as a powerful tool for nanostructure metrology *Thin Solid Films* **584** 176

[9] Romeijn H E and Smith R L 1994 Simulated annealing for constrained global optimization *J. Global Optim.* **5** 101

[10] Deb K, Pratap A, Agarwal S and Meyarivan T 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II *IEEE Trans. Evol. Comput.* **6** 182

[11] Al-Assaad R M and Byrne D M 2007 Error analysis in inverse scatterometry: I. Modeling *J. Opt. Soc. Am.* A **24** 326

[12] Zhu J L, Liu S Y, Chen X G, Zhang C W and Jiang H 2014 Robust solution to the inverse problem in optical scatterometry *Opt. Express* **22** 22031

[13] Zhu J L, Liu S Y, Jiang H, Zhang C W and Chen X G 2015 Improved deep-etched multilayer grating reconstruction by considering etching anisotropy and abnormal errors in optical scatterometry *Opt. Lett.* **40** 471

[14] Hubert M, Rousseeuw P J and Aeist S V 2008 High-breakdown robust multivariate methods *Stat. Sci.* **23** 92

[15] Rousseeuw P J 1984 Least median of squares regression *J. Am. Stat. Assoc.* **79** 871

[16] Rousseeuw P J and Driessen K V 2006 Computing LTS regression for large data sets *Data Min. Knowl. Disc.* **12** 29

[17] Vogel C 2002 *Computational Methods for Inverse Problems* (Philadelphia: SIAM)

[18] Rousseeuw P J and Zoomeren B C V 1990 Unmasking multivariate outliers and leverage points *J. Am. Stat. Assoc.* **85** 633

[19] Liu S Y, Ma Y, Chen X G and Zhang C W 2012 Estimation of the convergence order of rigorous coupled-wave analysis for binary gratings in optical critical dimension metrology *Opt. Eng.* **51** 081504

[20] Rousseeuw P J and Driessen K V 1999 A fast algorithm for the minimum covariance determinant estimator *Technometrics* **41** 212

[21] Novikova T, De Martino A, Ben Hatit S and Drévillon B 2006 Application of Mueller polarimetry in conical diffraction for critical dimension measurements in microelectronics *Appl. Opt.* **45** 3688

[22] Dixit D *et al* 2015 Sensitivity analysis and line edge roughness determination of 28 nm pitch silicon fins using Mueller matrix spectroscopic ellipsometry-based optical critical dimension metrology *J. Micro/Nanolith. MEMS MOEMS* **14** 031208

[23] Kaplan B, Novikova T, Martino A D and Drevillon B 2004 Characterization of bidimensional gratings by spectroscopic ellipsometry and angle-resolved Mueller polarimetry *Appl. Opt.* **43** 1233

[24] Muthinti G R, Peterson B, Bonam R K and Diebold A C 2013 Characterization of e-beam patterned grating structures using Mueller matrix based scatterometry *J. Micro/Nanolith. MEMS MOEMS* **12** 013018