

# Robust solution to the inverse problem in optical scatterometry

Jinlong Zhu,<sup>1</sup> Shiyuan Liu,<sup>1,2\*</sup> Xiuguo Chen,<sup>1</sup> Chuanwei Zhang,<sup>2</sup> and Hao Jiang<sup>2</sup>

<sup>1</sup> Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China

<sup>2</sup> State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China  
\*shyliu@mail.hust.edu.cn

**Abstract:** In optical scatterometry, the least squares (LSQ) function is usually used as the objective function to quantify the difference between the calculated and measured signatures, which is based on the belief that the actual measurement errors are normally distributed with zero mean. However, in practice the normal distribution assumption of measurement errors is oversimplified since these errors come from the superimposed effect of different error sources. Biased or inaccurate results may be induced when the traditional LSQ function based Gauss-Newton (GN) method is used in optical scatterometry. In this paper, we propose a robust method based on the principle of robust estimation to deal with the abnormal distributed errors. An additional robust regression procedure is used at the end of each iteration of the GN method to obtain the more accurate parameter departure vector. Simulations and experiments have demonstrated the feasibility of our proposed method.

©2014 Optical Society of America

**OCIS codes:** (120.0120) Instrumentation, measurement, and metrology; (120.2130) Ellipsometry and polarimetry; (290.3200) Inverse scattering; (050.1950) Diffraction gratings; (000.5490) Probability theory, stochastic processes, and statistics.

---

## References and links

1. C. J. Raymond, M. R. Murnane, S. L. Prins, S. S. H. Naqvi, J. W. Hosch, and J. R. McNeil, "Multiparameter grating metrology using optical scatterometry," *J. Vac. Sci. Technol.* **15**(2), 361–368 (1997).
2. X. Niu, N. Jakatdar, J. Bao, and C. J. Spanos, "Specular spectroscopic scatterometry," *IEEE Trans. Semicond. Manuf.* **14**(2), 97–111 (2001).
3. 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).
4. 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**(4), 499–512 (2011).
5. 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**(23), 21336–21343 (2009).
6. 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**(23), 233105 (2008).
7. Y. Ohkawa, Y. Tsuji, and M. Koshiba, "Analysis of anisotropic dielectric grating diffraction using the finite-element method," *J. Opt. Soc. Am. A* **13**(5), 1006–1012 (1996).
8. 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**(8), 1494–1502 (1990).
9. H. Ichikawa, "Electromagnetic analysis of diffraction gratings by the finite-difference time-domain method," *J. Opt. Soc. Am. A* **15**(1), 152–157 (1998).
10. M. G. Moharam, E. B. Grann, D. A. Pommel, 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**(5), 1068–1076 (1995).
11. L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A* **13**(9), 1870–1876 (1996).
12. S. Y. Liu, Y. Ma, X. G. Chen, and C. W. Zhang, "Estimation of the convergence order of rigorous coupled-wave analysis for binary gratings in optical critical dimension metrology," *Opt. Eng.* **51**(8), 081504 (2012).

13. J. L. Zhu, S. Y. Liu, C. W. Zhang, X. G. Chen, and Z. Q. Dong, "Identification and reconstruction of diffraction structures in optical scatterometry using support vector machine method," *J. Micro/Nanolith.* **12**(1), 013004 (2013).
14. X. G. Chen, S. Y. Liu, C. W. Zhang, and H. Jiang, "Improved measurement accuracy in optical scatterometry using correction-based library search," *Appl. Opt.* **52**(27), 6726–6734 (2013).
15. X. G. Chen, S. Y. Liu, C. W. Zhang, and J. L. Zhu, "Improved measurement accuracy in optical scatterometry using fitting error interpolation based library search," *Measurement* **46**(8), 2638–2646 (2013).
16. R. M. Al-Assaad and D. M. Byrne, "Error analysis in inverse scatterometry. I. Modeling," *J. Opt. Soc. Am. A* **24**(2), 326–338 (2007).
17. C. W. Zhang, S. Y. Liu, T. L. Shi, and Z. R. Tang, "Improved model-based infrared reflectrometry for measuring deep trench structures," *J. Opt. Soc. Am. A* **26**(11), 2327–2335 (2009).
18. H. Gross, A. Rathsfeld, F. Scholze, and M. Bar, "Profile reconstruction in extreme ultraviolet (EUV) scatterometry modeling and uncertainty estimates," *Meas. Sci. Technol.* **20**(10), 105102 (2009).
19. 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**(12), 12771–12786 (2012).
20. R. C. Geary, "Testing for normality," *Biometrika* **34**(3-4), 209–242 (1947).
21. A. Kato and F. Scholze, "Effect of line roughness on the diffraction intensities in angular resolved scatterometry," *Appl. Opt.* **49**(31), 6102–6110 (2010).
22. X. G. Chen, C. W. Zhang, and S. Y. Liu, "Depolarization effects from nanoimprinted grating structures as measured by Mueller matrix polarimetry," *Appl. Phys. Lett.* **103**(15), 151605 (2013).
23. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes – the art of science computing*, 3rd ed. (Cambridge University, 2007).
24. J. Hald and K. Madsen, "Combined LP and Quasi-Newton methods for nonlinear l1 optimization," *J. Numer. Anal.* **22**(1), 68–80 (1985).
25. E. J. Schlossmacher, "An iterative technique for absolute deviations curve fitting," *J. Am. Stat. Assoc.* **68**(344), 857–859 (1973).
26. R. A. El-Attar, M. Vidyasagar, and S. R. K. Dutta, "An algorithm for l1-norm minimization with application to nonlinear l1-approximation," *J. Numer. Anal.* **16**(1), 70–86 (1979).
27. H. Eklblom and K. Madsen, "Algorithms for non-linear Huber estimation," *BIT Numer. Math.* **29**(1), 60–76 (1989).
28. R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*, 1st ed. (Academic, 2005).
29. A. E. Beaton and J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics* **16**(2), 147–185 (1974).
30. D. Andrews, F. Bickel, F. Hampel, P. Huber, W. Rogers, and J. Tukey, *Robust Estimation of Location: Survey and Advances*, 1st ed. (Princeton University, 1972).
31. S. Y. Liu, X. G. Chen, and C. W. Zhang, "Mueller matrix polarimetry: A powerful tool for nanostructure metrology," *ECS Trans.* **60**(1), 237–242 (2014).
32. X. G. Chen, S. Y. Liu, C. W. Zhang, H. Jiang, Z. C. Ma, T. Y. Sun, and Z. M. Xu, "Accurate characterization of nanoimprinted resist patterns using Mueller matrix ellipsometry," *Opt. Express* **22**(12), 15165–15177 (2014).
33. 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.* **12**(3), 033013 (2013).
34. C. M. Herzinger, B. Johs, W. A. McGahan, J. A. Woollam, and W. Paulson, "Ellipsometric determination of optical constants for silicon and thermally grown silicon dioxide via a multi-sample, multi-wavelength, multi-angle investigation," *J. Appl. Phys.* **83**(6), 3323–3336 (1998).
35. T. Novikova, A. De Martino, S. Ben Htit, and B. Drévillon, "Application of Mueller polarimetry in conical diffraction for critical dimension measurements in microelectronics," *Appl. Opt.* **45**(16), 3688–3697 (2006).

## 1. Introduction

Optical scatterometry [1–3], also referred to as optical critical dimension (OCD) metrology, has been applied in many scopes with great success, such as the process control for back-end-of-the-line (BEOL) [4], the in-chip critical dimension (CD) and overlay metrology [5], and in situ measurement of pattern reflow in nanoimprinting [6]. In general, there are two main procedures in optical scatterometry. The first one involves calculation of the theoretical signature from a diffractive structure using reliable forward-modeling techniques, such as the finite element method (FEM) [7], the boundary element method (BEM) [8], the finite-difference time-domain (FDTD) method [9], and the rigorous coupled-wave analysis (RCWA) [10–12]. Here the general term signature contains the scattered light information of the diffractive structure, which can be in the form of reflectance, ellisometric angles, Stokes elements, and Mueller matrix elements. The second procedure involves the extraction of profile parameters from the measured signature, which is a typical inverse problem with the objective of finding the profile parameters whose calculated signature can best match the measured one.

To solve the inverse problem in optical scatterometry, an objective function, usually the least squares (LSQ) function, is used to quantify the difference between the measured and calculated signatures. To increase the precision and accuracy of the profile parameter extraction, the signature is usually obtained as a vector consisting of several data points at multiple wavelengths or incident angles. Then by using either a library search [13–15] or an iterative optimization such as the Gauss-Newton (GN) method [16, 17] to minimize the objective function, a set of optimal profile parameters corresponding to the best matched calculated signature can be achieved. The general form of the LSQ function contains a set of weight factors representing the reciprocal of variances. The weight factors can be either fixed [18] or varying [19], both of which are built on top of the belief that the measurement errors are normally distributed with zero mean. Consequently, the weighted measurement errors, defined as the measurement errors at all the wavelengths or incident angles divided by their corresponding estimated standard deviations, are considered to follow standardized normal distribution. This belief indicates that the weighted measurement errors at all the wavelengths or incident angles fall into the range  $-3 \sim 3$  with a probability of 99.7%. However, in an actual measurement system, normality is only a myth, as Geary pointed out [20]. In optical scatterometry, the superimposed effect from different error sources such as the power fluctuation of the incident beam [19], the imperfect modeling [21], the depolarization effect [22], and even human operation errors will bias the actual statistical property of measurement error from normal distribution. In addition, considering the inevitably inaccurate estimation of variances of measurement errors, some of the weighted measurement errors might way off from the range  $-3 \sim 3$ , which are called outliers in this paper. Once the outliers exist, the traditional iteration method like GN is willing to distort the calculated signature curve [23] to reduce the outliers, which in turn enlarges the other weighted measurement errors, and thus leads to a biased or even wrong estimation of profile parameters. Therefore, it is highly desirable to develop a method that can effectively reduce the effect of outliers in the profile parameter estimation.

To reduce the outlier effect, one may consider simply rejecting those wavelengths or incident angles that correspond to the outliers. However, it is difficult or even impossible to sort out the outliers from a measured signature directly. In the present paper, based on the principle of robust statistics [24–27], we introduce an additional regression procedure in each iteration of the LSQ-based GN method. Without modifying the experimental measurement setups, the proposed method achieves more robust estimation of profile parameters.

The remainder of this paper is organized as follows. Section 2 introduces the formulation of the inverse problem in optical scatterometry followed by description of the traditional LSQ-based GN iteration method and our proposed robust method. Section 3 presents the measurement setup, sample description, and numerical and experimental results. Then, we draw some conclusions in Section 4.

## 2. Methods

### 2.1 Inverse problem in optical scatterometry

Usually, the inverse problem in optical scatterometry is 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  $\chi^2$ . The use of  $\chi^2$  is built on top of the belief that the measurement errors are normally distributed with zero mean, namely, the measured signature  $\mathbf{y}$  can be expressed as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\epsilon}, \quad (2)$$

where  $\boldsymbol{\epsilon}$  is a vector of multiple random variables with each element corresponding to a normal distribution characterized by a specific standard deviation. Consequently, without losing generality, the inverse problem in optical scatterometry can be formulated as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \Omega} \{[\mathbf{y} - \mathbf{f}(\mathbf{x})]^T \mathbf{w} [\mathbf{y} - \mathbf{f}(\mathbf{x})]\}, \quad (3)$$

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

## 2.2 Solution to the inverse problem with GN iteration 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 as shown in Eq. (3) 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

$$\nabla F(\mathbf{x}^{(i+1)}) = \nabla F(\mathbf{x}^{(i)} + \Delta \mathbf{x}^{(i)}) = \nabla F(\mathbf{x}^{(i)}) + \nabla^2 F(\mathbf{x}^{(i)}) \Delta \mathbf{x}^{(i)}, \quad (4)$$

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

$$\nabla F(\mathbf{x}^{(i)}) = 2\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \Delta \mathbf{y}^{(i)}, \quad (5)$$

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

$$\Delta \mathbf{y}^{(i)} = \mathbf{y} - \mathbf{f}(\mathbf{x}^{(i)}). \quad (6)$$

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

$$\nabla^2 F(\mathbf{x}^{(i)}) \approx 2\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)}). \quad (7)$$

Substituting Eq. (5) and Eq. (7) into Eq. (4) and letting Eq. (4) equal zero, we will have

$$\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)}) \Delta \mathbf{x}^{(i)} = -\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \Delta \mathbf{y}^{(i)}. \quad (8)$$

Further we will have

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

which assumes that the expression  $\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)})$  is nonsingular. By conducting Eq. (4) ~ Eq. (9) iteratively, we can obtain an optimal result  $\hat{\mathbf{x}}$ , which is usually attributed to the GN iteration [28].

## 2.3 Robust solution to the inverse problem with robust statistics method

Since in Eq. (9) the Jacobian  $\mathbf{J}(\mathbf{x}^{(i)})$  is only calculated from the forward modeling operator at the current  $\mathbf{x}^{(i)}$ , and  $\mathbf{w}$  is the known matrix, the estimation of  $\Delta \mathbf{x}^{(i)}$  is dominated by  $\Delta \mathbf{y}^{(i)}$ . Here  $\Delta \mathbf{y}^{(i)}$  is the difference between the measured signature and the  $i$ th calculated signature of the

corresponding profile parameters. If the measurement errors follow the normal distribution, Eq. (9) usually can ensure a good estimation of  $\Delta\mathbf{x}^{(i)}$ . However, as mentioned in the introduction section, the assumption of normal distribution with zero mean for the measurement error is oversimplified. Actually, the measured signature  $\mathbf{y}$  should be formulated as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon} + \boldsymbol{\mu}, \quad (10)$$

where  $\boldsymbol{\mu}$  is a vector of multiple deterministic errors with each element corresponding to a specific value. The deterministic error vector  $\boldsymbol{\mu}$  is the superimposed result by errors from different error sources such as the imperfect modeling [21] and the depolarization effect [22]. Because of the existence of  $\boldsymbol{\mu}$ , the direct use of Eq. (9) may lead to a biased or even inaccurate estimation of  $\Delta\mathbf{x}^{(i)}$ . As such, it is of great importance to give a robust estimation for the  $i$ th parameter departure vector  $\Delta\mathbf{x}^{(i)}$ .

To achieve a robust estimation of parameter departure vector  $\Delta\mathbf{x}^{(i)}$ , we first use the Moore-Penrose pseudo-inverse of  $\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w}$  to transform Eq. (8) into

$$\Delta\mathbf{y}^{(i)} = -\{[\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w}]^T \mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w}\}^{-1} [\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w}]^T \mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)}) \Delta\mathbf{x}^{(i)} = -\mathbf{J}(\mathbf{x}^{(i)}) \Delta\mathbf{x}^{(i)}. \quad (11)$$

As shown in Eq. (11), the estimation of  $\Delta\mathbf{x}^{(i)}$  can be treated as a problem to estimate the slope of a hyperline, which can be described as estimating the slope of a hyperline  $\Delta\mathbf{x}^{(i)}$  in the  $n$ -dimensional space with  $m$  given input data pairs  $(\mathbf{J}(\mathbf{x}^{(i)})_j, \Delta y_j^{(i)})$ . Here  $\Delta y_j^{(i)}$  is the  $j$ th element of  $\Delta\mathbf{y}^{(i)}$ , and  $\mathbf{J}(\mathbf{x}^{(i)})_j$  is the  $j$ th row of  $\mathbf{J}(\mathbf{x}^{(i)})$ . Instead of simply using Eq. (9), we can achieve a robust estimation of  $\Delta\mathbf{x}^{(i)}$  by solving the following regression problem

$$\hat{\Delta\mathbf{x}}^{*(i)} = \arg \min_{\Delta\mathbf{x}^{*(i)} \in \Omega^*} \left\{ \sum_{j=1}^m \rho[\Delta y_j^{(i)} + \mathbf{J}(\mathbf{x}^{(i)})_j \Delta\mathbf{x}^{*(i)}] \right\} = \arg \min_{\Delta\mathbf{x}^{*(i)} \in \Omega^*} \left\{ \sum_{j=1}^m \rho[r_j^{(i)}] \right\}, \quad (12)$$

where  $\rho$  is a predefined robust objective function,  $\Omega^*$  is the parameter domain of  $\Delta\mathbf{x}^{*(i)}$ ,  $\Delta\mathbf{x}^{*(i)}$  is the estimated result of the parameter departure vector, and the asterisk is used to distinguish the result obtained by using the robust statistics method from that obtained by Eq. (9).  $r_j^{(i)}$  is the residual term equal to  $\Delta y_j^{(i)} + \mathbf{J}(\mathbf{x}^{(i)})_j \Delta\mathbf{x}^{*(i)}$ . To solve Eq. (12), the GN iteration could be used. For simplicity, we substitute  $\Delta\tilde{y}_j$ ,  $\Delta\tilde{\mathbf{y}}$ ,  $\Delta\tilde{\mathbf{x}}$ ,  $\tilde{\mathbf{J}}$ , and  $\tilde{r}_j$  for  $\Delta y_j^{(i)}$ ,  $\Delta\mathbf{y}^{(i)}$ ,  $\Delta\mathbf{x}^{*(i)}$ ,  $\mathbf{J}(\mathbf{x}^{(i)})$ , and  $r_j^{(i)}$  respectively. Consequently, the problem described in Eq. (12) can be rewritten as

$$\hat{\Delta\mathbf{x}}^{*(i)} = \arg \min_{\Delta\tilde{\mathbf{x}} \in \Omega^*} \left[ \sum_{j=1}^m \rho(\Delta\tilde{y}_j + \tilde{\mathbf{J}}_j \Delta\tilde{\mathbf{x}}) \right] = \arg \min_{\Delta\tilde{\mathbf{x}} \in \Omega^*} \left[ \sum_{j=1}^m \rho(\tilde{r}_j) \right]. \quad (13)$$

Starting with an initial guess  $\Delta\tilde{\mathbf{x}}^{(0)}$ , the GN iteration can be written in this context as

$$\Delta\tilde{\mathbf{x}}^{(k+1)} = \Delta\tilde{\mathbf{x}}^{(k)} - (\mathbf{J}^T \mathbf{Q}^{(k)} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{P}^{(k)}. \quad (14)$$

Here  $k$  indicates the  $k$ th iteration,  $\mathbf{P}^{(k)}$  is the gradient vector for the term  $\sum_{j=1}^m \rho[\tilde{r}_j^{(k)}]$  with respect to  $\tilde{r}_j^{(k)}$ , and  $\mathbf{Q}^{(k)}$  is a  $m \times m$  diagonal matrix with entries

$$Q_{jj} = \rho''[\tilde{r}_j^{(k)}], \quad (15)$$

where  $\rho''$  represents the second derivative of the robust objective function with respect to  $\tilde{r}_j^{(k)}$ . Defining the weight function as  $\omega(p) = \rho'(p)/p$  where  $p$  is an arbitrary scalar variable [29], we can get another iteration scheme

$$\Delta\tilde{\mathbf{x}}^{(k+1)} = \Delta\tilde{\mathbf{x}}^{(k)} - (\mathbf{J}^T \mathbf{S}^{(k)} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{S}^{(k)} (\Delta\tilde{\mathbf{y}} + \mathbf{J}\Delta\tilde{\mathbf{x}}^{(k)}), \quad (16)$$

where  $\mathbf{S}^{(k)}$  is a diagonal matrix with entries

$$S_{jj} = \omega(\tilde{r}_j^{(k)}) = \omega(\Delta\tilde{y}_j + \mathbf{J}_j \Delta\tilde{\mathbf{x}}^{(k)}). \quad (17)$$

As can be seen in Eq. (16) and Eq. (17), this method only requires knowledge of the weight function  $\omega(p)$ . In the present paper, the weight function  $\omega(p)$  is chosen as the Andrews's hard redescender [30],

$$\omega(p) = \begin{cases} \frac{c_A}{p} \sin pc_A, & |p| \leq \pi c_A \\ 0, & |p| > \pi c_A \end{cases}, \quad (18)$$

where  $c_A$  is a tuning constant and  $p$  is an arbitrary scalar variable. The definition of  $\omega(p)$  in (18) indicates that  $\rho'(p) \rightarrow 0$  when  $|p|$  is sufficiently large.  $c_A$  is set as 1.339 to ensure the robust regression is approximately 95% as statistically efficient as the LSQ estimates, provided the measurement error follows a normal distribution with no outliers. Since Andrews's hard redescender is applied, the weight factor corresponds to each residual  $\tilde{r}_j^{(k)}$  is adjusted at each iteration, which is different from Eq. (1) in which the weight factors are set as constant during the whole regression. By repeatedly conducting Eq. (13) ~ Eq. (17) we can get the estimate  $\hat{\Delta\mathbf{x}}^{*(i)}$ , which is more robust than the estimate  $\Delta\mathbf{x}^{(i)}$  obtained by using Eq. (9).

In summary, the proposed method introduces an additional “inner” GN iteration in the  $i$ th iteration of the “outer” GN iteration using Andrews's hard redescender. Instead of using Eq. (9) directly, a robust parameter departure vector  $\hat{\Delta\mathbf{x}}^{*(i)}$  can be achieved. We shall point out that this robust statistics method can be applied not only in scatterometry, but also in the general problem of model fitting in ellipsometry and related thin film measurements, provided that the expression  $\mathbf{J}(\mathbf{x}^{(i)})^T \mathbf{w} \mathbf{J}(\mathbf{x}^{(i)})$  is nonsingular. This implies that the Jacobian  $\mathbf{J}(\mathbf{x}^{(i)})$  is of full column rank and thus leads to  $\Delta\mathbf{y}^{(i)} = -\mathbf{J}(\mathbf{x}^{(i)}) \Delta\mathbf{x}^{(i)}$ .

### 3. Results

#### 3.1 Measurement setup and sample description

A self-developed dual-rotating compensator Mueller matrix ellipsometer (DRC-MME) prototype suitable from ultraviolet to infrared spectrum [31, 32] is used for demonstration. Data analysis is performed using the in-house developed optical modeling software based on rigorous coupled-wave analysis (RCWA) [10–12]. As schematically shown in Fig. 1(a), the system setting of the DRC-MME in order of light propagation is PC<sub>r1</sub>SC<sub>r2</sub>A, where P and A stand for the fixed polarizer and analyzer, respectively, C<sub>r1</sub> and C<sub>r2</sub> refer to the first and second frequency-coupled rotating compensators, respectively, and S stands for the sample. With the light source used in this ellipsometer, the available wavelengths are in 200 to 1000 nm range, covering the spectral range of 200 to 800 nm used in this paper. With the dual-rotating compensator setting, we can obtain the full Mueller matrix elements of the sample.

The investigated sample is a one-dimensional (1-D) etched Si grating, whose scanning electron microscope (SEM) cross-section image is shown in Fig. 1(b). The etched Si grating is chosen for this study due to its long-term dimensional stability, high refractive index contrast, and relevance to the semiconductor industry [33]. Optical properties of Si are taken from [34]. As depicted in Fig. 1(b), a cross-section of the Si grating is characterized by a symmetrical trapezoidal model with top critical dimension *TCD*, bottom critical dimension *BCD*, grating height *Hgt*, and period *Pitch*. Dimensions of the structural parameters that obtained by SEM are *TCD* = 350 nm, *Hgt* = 472 nm, and *BCD* = 383 nm. In the following experiments, profile parameters of the Si grating that need to be extracted include *TCD*, *Hgt*, and *BCD*, while the grating period is fixed at its nominal dimension, i.e., *Pitch* = 800 nm.

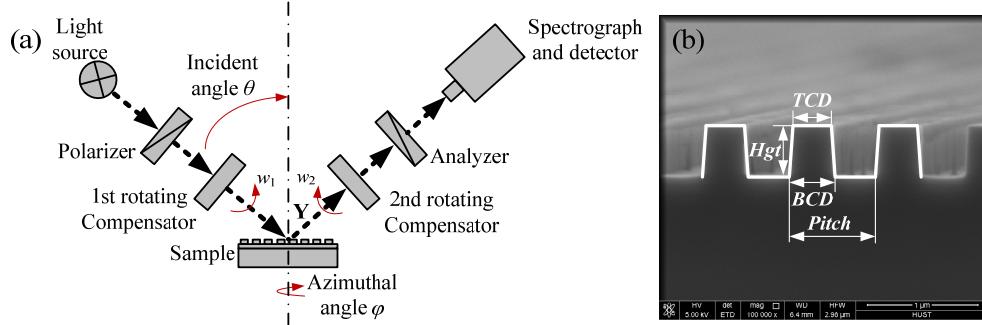


Fig. 1. (a) Measurement setup of the dual-rotating-compensator Mueller matrix ellipsometer, and (b) cross-section SEM image of the one-dimensional trapezoidal etched silicon grating.

### 3.2 Numerical results

In this section, we simulate the measurement process of Si grating by firstly calculating its corresponding Mueller matrix. The values of profile parameters *TCD*, *Hgt* and *BCD* of the Si grating are set as 350 nm, 472 nm and 383 nm respectively. The measurement configuration is 45° incident angle and 10° azimuthal angle. Then the simulated normal distribution errors are added into the calculated Mueller matrix to form the “measurement” signature. The standard deviation or noise level of the simulated normal distribution error 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 0 ~5%. To know more about the modeling of a specific normal distribution error, we refer the readers to [16]. Finally, the GN method and our proposed method are used to extract the profile parameters from the “measurement” signature. The values of *TCD*, *Hgt* and *BCD* extracted by our proposed method are 350.4 nm, 471.9 nm and 382.6 nm, respectively, which are in coincidence with the extracted values 350.5 nm, 471.4 nm, and 383.9 nm obtained by the GN method. The above simulation has demonstrated that our proposed method is effective in dealing with the normal distributed errors.

As described in the Introduction section, there are many different types of error sources that make contribution to the measurement errors. Since simulating all these types of error sources is impossible, we add some known simulated error sources to the “measured” signature described in the above paragraph instead. The error sources simulated here include the finite numerical aperture of the DRC-MME, the spectral resolution of the monochromator, and the biases of the estimated standard deviations of measurement errors. The simulation method of the former two error sources were directly taken from [22], while for the last one the bias of each estimated standard deviation was randomly chosen within the range of -0.003 ~0.003, which ensures the absolute bias is less than 20% of the estimated standard deviation. After injection of the errors above, the LSQ based GN method and the proposed robust method are used to extract the profile parameters from the “measured” signature respectively. The iteration results of *TCD*, *Hgt* and *BCD* for both the two methods are presented in Fig. 2, in which we can obviously find that the proposed robust method converges to the more accurate solution. For more clarity, Fig. 3 displays the weighted “measurement” errors, in which we can find that some data are obviously out of the range -3 ~3.

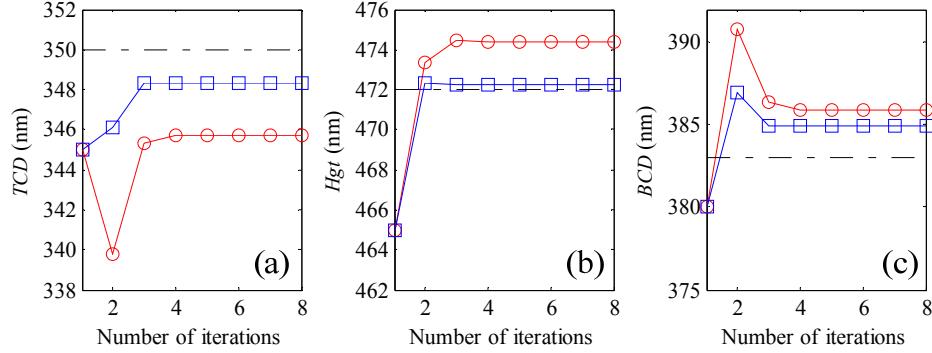


Fig. 2. Iteration results of (a)  $TCD$ , (b)  $Hgt$ , and (c)  $BCD$ . The black dash dotted line, blue solid line with squares, and red solid line with circles represent the true values, the extracted values by our proposed method, and those by the traditional GN method respectively.

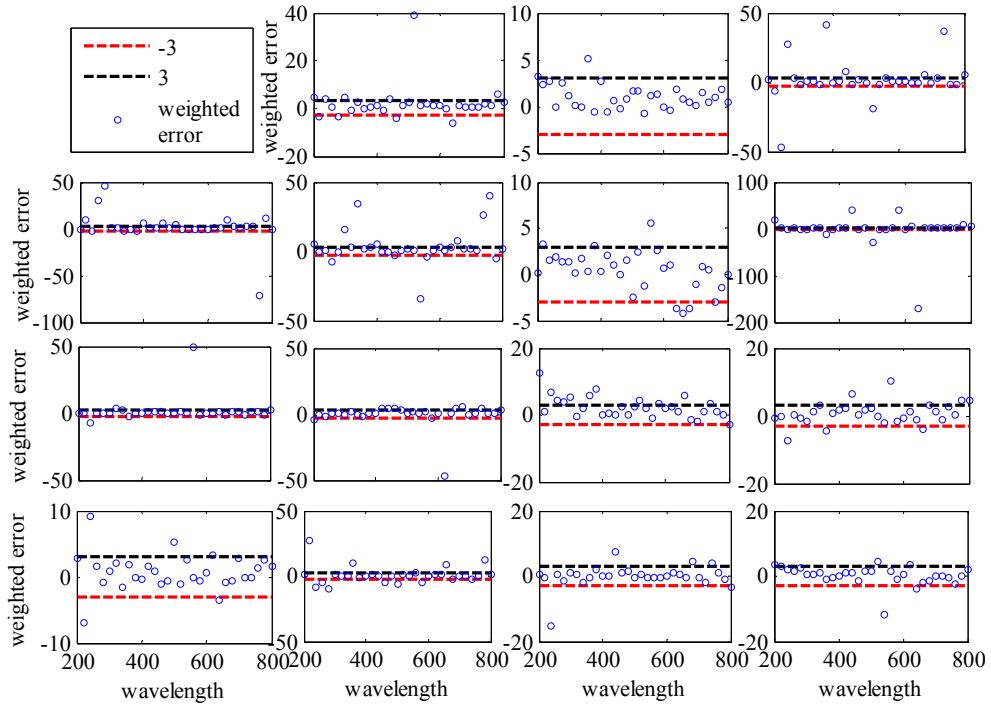


Fig. 3. Weighted “measurement” errors are represented by blue squares. The red and black dotted lines in each sub-figure represent  $-3$  and  $3$  respectively.

The above simulation has demonstrated that our proposed method is more robust than the GN method. Moreover, the results of the extracted values of profile parameters under different incident angles (with fixed  $10^\circ$  azimuthal angle) and azimuthal angles (with fixed  $45^\circ$  incident angle) are presented in Fig. 4, in which we can find that nearly all the extracted values achieved by our proposed method are closer to the true values than those obtained by the traditional GN method.

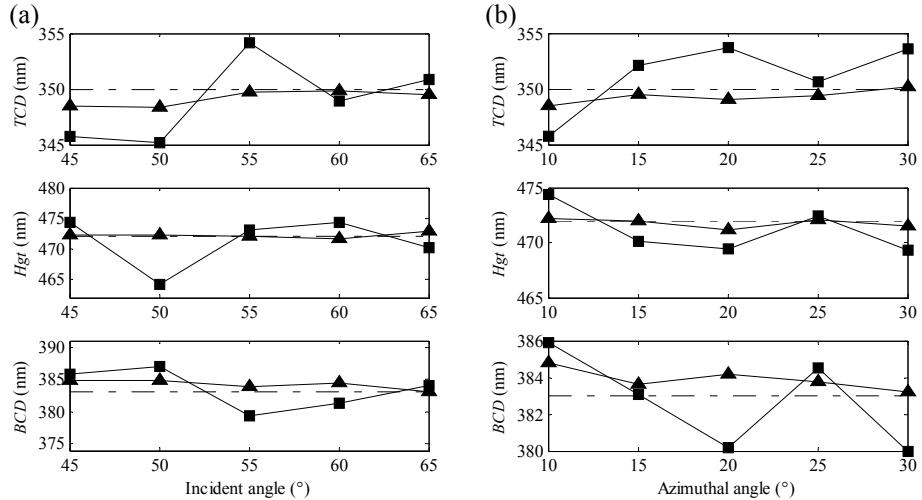


Fig. 4. Numerically extracted *TCDs*, *Hgts* and *BCDs* by varying (a) the incident angle and (b) the azimuthal angle. Black squares and triangles represent the values obtained by the LSQ-based GN method and our proposed method respectively. Black dash dotted lines represent the true values.

### 3.3 Experimental results

In this section, we perform experiments to further compare our proposed method with the GN method for extracting profile parameters. The measured Mueller matrix is obtained under an incident angle of  $45^\circ$  and an azimuthal angle of  $10^\circ$ . The extracted profile parameters are presented in Table 1, in which we can observe that the values of *TCD*, *Hgt*, and *BCD* extracted by our proposed method are 3.1 nm, 1.6 nm and 5.3 nm closer to the SEM measurements than those obtained by the GN method.

**Table 1. Mean values of *TCD*, *Hgt*, and *BCD* obtained by different methods.**

Method	<i>TCD</i> (nm)	<i>Hgt</i> (nm)	<i>BCD</i> (nm)
SEM	350	472	383
GN	343.5	476.4	393
Proposed	346.6	474.8	387.7

To explain the differences of profile parameters extracted by our proposed and the GN methods, we shall demonstrate that the measurement errors are not normally distributed with zero mean. Note that in practice to accurately obtain the measurement errors is not possible; instead, we calculate the fitting differences between the measured Mueller matrix and the best calculated Mueller matrix by the GN method. Figure 5 depicts the fitting differences together with their corresponding standard deviations. Each signature data point corresponds to a standard deviation of a measurement error, thus the standard deviations at different wavelengths also differ from each other. We then calculate the weighted fitting differences of the Mueller matrix elements obtained by the GN method, as shown in Fig. 6, where the weighted fitting differences are the fitting differences divided by the standard deviations.

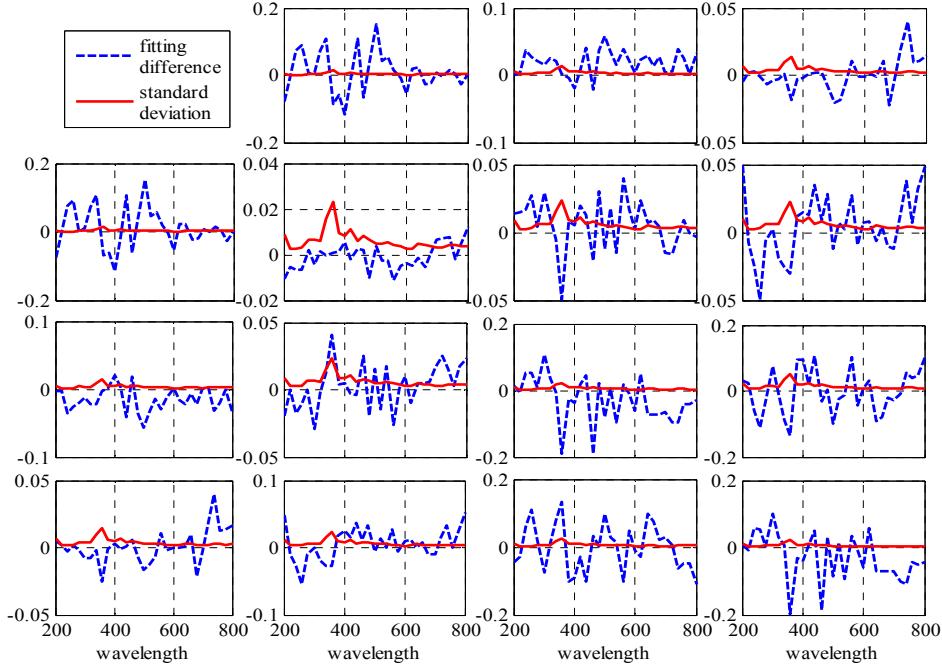


Fig. 5. Fitting differences between the measured and the best calculated Mueller matrices, and the standard deviations of the measurement errors used in the LSQ function. The best calculated Mueller matrix is obtained by the GN method.

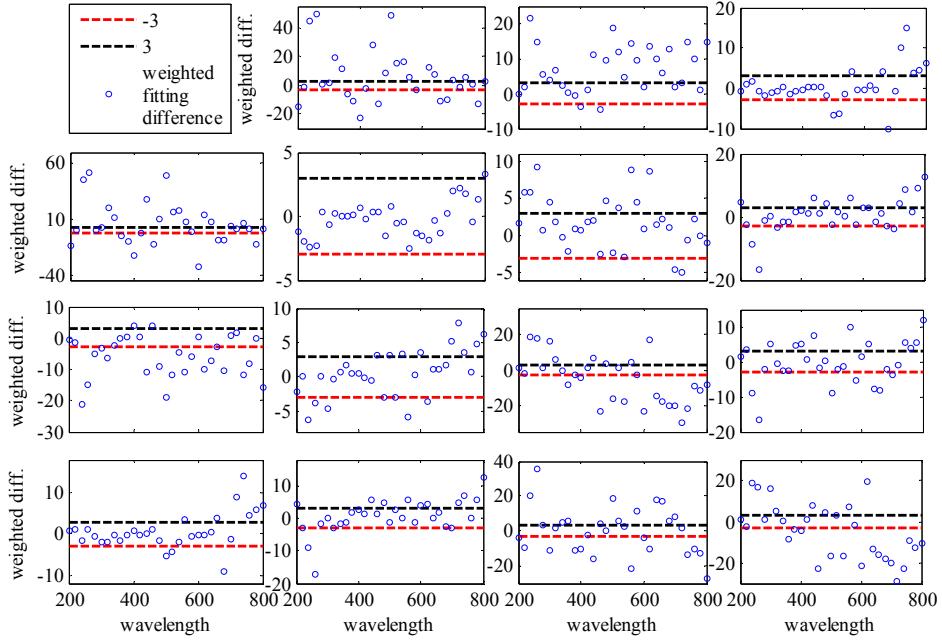


Fig. 6. Weighted fitting differences of Mueller matrix elements. The blue squares represent the weighted fitting differences at different wavelengths, and the red and black dotted lines in each sub-figure represent  $-3$  and  $3$  respectively.

From Fig. 6, we indeed observe many large data points that are out of the range  $-3$  to  $3$ . These outliers have demonstrated that the normal distribution assumption of measurement

errors is not valid. Actually, the measurement errors contain not only the random errors, but also the deterministic errors [16], as shown in Eq. (10). The deterministic error at a specific wavelength has no influence on the standard deviation of the measurement error. However, the deterministic error does have non-ignorable influence on the value of the weighted fitting differences, since the GN method cannot exclude the deterministic error effects. Consequently, the weighted fitting differences may be dragged out of the range  $-3 \sim 3$  by the deterministic error effects. In consideration of the fact that the deterministic errors differ with each other at different wavelengths, the weighted fitting differences then present the diversity, as can be seen in Fig. 6.

Unfortunately, the true values of the profile parameters are unknown in practice, which increases the difficulty of accuracy estimation. As an alternative, we can indirectly estimate the accuracy by varying the incident and azimuthal angles [35]. From this point of view, we measure the Mueller matrices of the 1-D etched Si grating by fixing the azimuthal angle at  $10^\circ$  while varying the incident angle from  $45^\circ \sim 65^\circ$  with an increment of  $5^\circ$ . We also perform measurements by fixing the incident angle at  $45^\circ$  while varying the azimuthal angle from  $10^\circ \sim 30^\circ$  with an increment of  $5^\circ$ . Figure 7 depicts the extracted values of  $TCD$ ,  $Hgt$  and  $BCD$  from these two sets of measurements. As shown in Fig. 7(a), at different incident angles the extracted  $TCDs$ ,  $Hgts$  and  $BCDs$  obtained by our proposed method are all closer to the SEM measured values compared with those obtained by the GN method. The same trend is also observed in Fig. 7(b) for different azimuthal angles. These experiments and discussions have demonstrated again that the normal distribution assumption of measurement errors is not always valid, and outliers are inevitable. They also have confirmed that our proposed method outperforms the traditional LSQ-based GN method when dealing with the outliers.

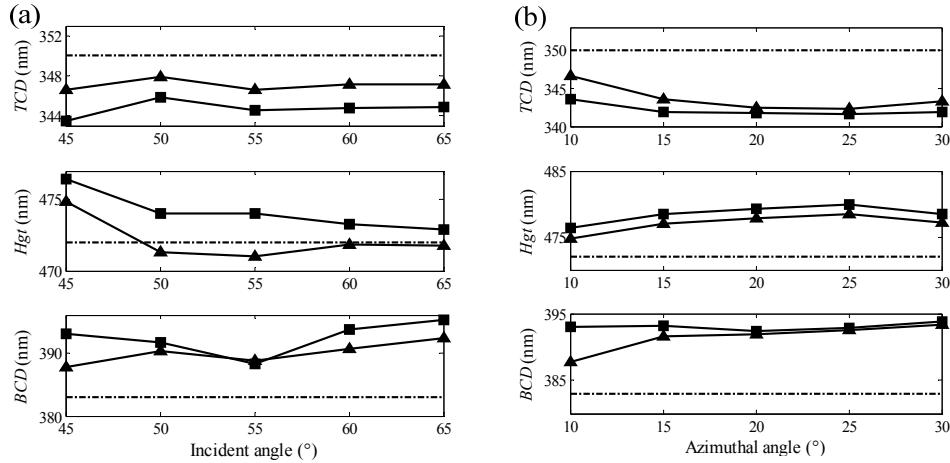


Fig. 7. Experimentally extracted  $TCDs$ ,  $Hgts$  and  $BCDs$  by varying (a) the incident angle and (b) the azimuthal angle. Black squares and triangles represent the values obtained by the LSQ-based GN method and our proposed method respectively. Black dash dotted lines represent the SEM measured values.

#### 4. Conclusions

In summary, we have numerically and experimentally demonstrated that it is not always valid to assume that the measurement errors in optical scatterometry are normally distributed. In this case, if the objective function is set as the LSQ function, the traditional iterative methods like the GN method leads to a biased estimation of profile parameters. To reduce the effect due to the abnormal distribution, we have proposed a robust parameter estimation method based on the principle of robust statistics, which has been proved to achieve more accuracy than the traditional GN method.

### **Acknowledgments**

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. The authors would like to thank the facility support of the Center for Nanoscale Characterization and Devices, Wuhan National Laboratory for Optoelectronics (WNLO) (Wuhan, China).