Fast and accurate solution of inverse problem in optical scatterometry using heuristic search and robust correction

Jinlong Zhu,[a] Hao Jiang,[a] Yating Shi, Chuanwei Zhang, Xiuguo Chen, and Shiyuan Liu[b]
State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

(Received 10 February 2015; accepted 22 April 2015; published 5 May 2015)

Library search is one of the most commonly used methods in optical scatterometry, which consists of the beforehand construction of a signature library and the grid search. The efficiency of existing search algorithms such as k-dimensional tree method and locality-sensitive hashing heavily depends on the size of the signature library and usually is inversely proportional to the library scale. Additionally, since the two-norm based objective function is quite sensitive to the outliers, the abnormally distributed measurement errors will bias the solution of the traditional chi-square or maximum likelihood function. In the present paper, the authors propose a heuristic search algorithm and a robust correction method to realize the fast library search and to achieve the more accurate results, respectively. Instead of searching in the signature library, the authors perform the search procedure in an extra constructed Jacobian library using the principle of gradient-based iteration algorithms, by which the fast search speed can be achieved for an arbitrary scale library. After the search, a robust correction procedure is performed on the basis of the searched optimal parameter set to obtain the more accurate results. Simulations and experiments performed on an etched silicon grating have demonstrated the feasibility of the proposed heuristic search algorithm and robust correction method. © *2015 American Vacuum Society*.
[http://dx.doi.org/10.1116/1.4919713]

I. INTRODUCTION

Optical scatterometry, also called optical critical dimension metrology, is one of the state-of-the-art techniques for nanostructure metrology in semiconductor industry because it meets most of the requirements of high-volume manufacturing and in-line monitoring[1,2] when compared with the conventional measurement tools such as scanning electron microscopy (SEM) or atomic force microscopy. Generally, two main procedures are involved in optical scatterometry. The first procedure is the simulation of the optical signature from a nanostructure using reliable forward modeling techniques, such as the finite element method,[3] the boundary element method,[4] the finite-difference time-domain method,[5] and the rigorous coupled-wave analysis (RCWA) method.[6–8] The term signature is a vector consisting of several data points at different wavelengths or incident angles and contains the scattered light information from the diffractive structure, which can be in the form of reflectance, ellipsometric angles, Stokes vector elements, or Mueller matrix elements. The second procedure involves the extraction of nanostructure profile parameters from the measured signature, which usually involves the use of numerical algorithms to find a profile with whose theoretical signature can best match the measured one.

To solve the inverse problem in optical scatterometry, two kinds of methods, which are the nonlinear regression[9,10] and the library search,[11,12] are widely used. The inherent essence of nonlinear regression is that the profile parameters are achieved through an iterative procedure, which repeatedly requires computation of the forward modeling.[13] Since the relationship between the optical signature and the profile parameters is highly nonlinear, the forward modeling procedure is usually time-consuming, and it will be even worse and unacceptable for two-dimensional or more complex nanostructure measurements. Hence, the nonlinear regression method is rarely used in the in-line monitoring of the complex nanostructure fabrications. A feasible way to meet the speed requirement, known as library search, is to beforehand construct a signature library and then to search the best match with the measured signature in the library. Although the offline generation of the signature library is time-consuming, the search itself during the online measurement can be done quickly with a global solution guaranteed.[14] Therefore, the library search has been widely used in semiconductor industry.[15]

Generally, in the library search method, the search speed and measurement accuracy are the two most concerned issues. To realize the fast search in a huge signature library, some search algorithms, such as the k-dimensional (k-d) tree method[16,17] and locality-sensitive hashing (LSH),[18] have been introduced. The main principle of k-d tree method is using the key node of nodes in k-d tree and the strategy of dividing and conquering to divide the elements set of searching space into subsets according to their position in space so as to reduce the number of elements to be visited.[17] However, the nodes number in the constructed k-d tree is in proportion to the number of signatures in the library, which indicates the time consumption increases with the expansion of library scale. The basic idea of LSH is to hash the library signatures by hash functions so that similar signatures are

[a]J. Zhu and H. Jiang contributed equally to this work.
[b]Author to whom correspondence should be addressed; electronic mail: shyliu@hust.edu.cn

mapped to the same buckets with high probability, then a query signature is mapped by the hash function to the most related bucket together with the retrieve of the most similar library signature in the bucket. However, as a randomized algorithm, the LSH does not guarantee an exact result but guarantees a high probability for a correct result or one close to it.[11] Moreover, the time consumption of LSH is also in proportion to the number of signatures in the library since the contained signature number in the bucket increases with the expansion of library scale, supposing the number of bucket keeps constant. Thus, it is of great importance to propose an algorithm that can realize the fast and reliable search in a library with arbitrary size.

Besides the search speed, the measurement accuracy is another concerned issue. To achieve the high measurement accuracy, we can either decrease the grid interval of the library or using the interpolation.[12] However, as will be explained in the following content, the measurement accuracy obtained by these methods is limited since the commonly used least square (LSQ) function[19] cannot effectively deal with the abnormal distributed measurement errors in the measured signature. In optical scatterometry, the objective function is usually selected as LSQ, in consideration of its formation simplicity and the differentiable property. From the statistical perspective, the LSQ function is equivalent to the maximum likelihood function,[20] which is built on top of the belief that the measurement errors are normally distributed. If the measurement errors are normally distributed, the weighted measurement errors obtained by dividing the measurement errors by weighted factors are then standard normally distributed and fall into the range of -3 to 3 with the probability of 99.7%. Here, the term measurement error and weight factor are the difference between a best numerical data point and a measured one, and the variance at a specific wavelength or incident angle, respectively. However, in a real measurement system, normality is only a myth, as Geary pointed out.[21] In optical scatterometry, the superimposed effect from different error sources such as the power fluctuation of the incident beam,[20] the imperfect modeling,[22] the depolarization effect,[23] and even human error will bias the actual statistical property of measurement error from the normal distribution. In addition, considering the inevitably inaccurate estimation of measurement error variances, some weighted measurement errors might way off from the range of -3 to 3, which are called outliers in this paper. Once the outliers exist, the solution corresponds to the minimal value of LSQ function may biased the true values of profile parameters, and the bias cannot be eliminated by simply reducing the grid interval or using interpolation. Hence, it is highly desirable to develop a method that can effectively reduce the effect of outliers on the estimation of profile parameters, for the purpose of achieving higher measurement accuracy.

Recently, we have proposed a robust regression method based on the principle of robust estimation to deal with the outliers in the measurement signature[24] and have demonstrated its feasibility in the improvement of a deep-etched multilayer grating reconstruction.[25] Though the proposed method is capable to realize the higher measurement accuracy of a nanostructure compared with that of the conventional Gauss–Newton (GN) method, it is inherent a regression-based method, and the extreme time-consumption cannot be avoided. In the present paper, we propose a heuristic search algorithm combined with a robust correction method to realize the fast and accurate reconstruction of profile parameters. We off-line generate a Jacobian library together with a signature library. We derive a formulation which is able to heuristically determine a search direction based on the signature and its corresponding Jacobian with respect to the current searched profile. By repeating the calculation of search direction, we can quickly find the solution in the signature library. Once the solution is found, an additional robust correction step based on the principle of robust statistics is used to correct the actually searched solution in the library. The corrected solution will be much closer to the true solution and therefore improves the final measurement accuracy. Since the time consuming generation of Jacobian library is off-line and the robust correction method in essence is regardless of the forward modeling, the proposed method will have no remarkable influence on its time cost.

The remainder of this paper is organized as follows. Section II briefly introduces the formulation of inverse problem in optical scatterometry and presents the heuristic search algorithm as well as the robust correction in detail. Section III presents the measurement setup, sample description, and numerical and experimental results. Then, we draw some conclusions in Sec. IV.

II. METHOD

A. Inverse problem in optical scatterometry

Usually, the inverse problem in optical scatterometry is described as an object of the minimization of an LSQ function, which can be generally formulated as

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

where the w_j is the weight factor, y_j is the jth measured data point, $\mathbf{y}$ is the measured signature, which is a vector and contains m data points. $f_j(\mathrm{x})$ is the jth numerical data point with respect to the profile parameters under measurement as an n-dimensional vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$, and $\mathbf{f}(\mathbf{x})$ is the numerical signature, which is a vector and contains m data points. $\mathbf{w}$ 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 the $\sigma(y_j)$ is the standard deviation of the measurement error that follows the normal distribution, then Eq. (1) relates to the chi-square statistic χ^2. Hence, without loss of generality, the inverse problem in optical scatterometry is typically formulated as

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

where $\hat{\mathbf{x}}$ is the solution of the inverse problem, and Ω is the associated parameter domain. In library search, the set of profile parameters corresponds to the numerical signature,

which ensures the minimization of Eq. (1) is treated as the desired solution $\hat{\mathbf{x}}$.

B. Heuristic search algorithm and robust correction method

To solve the inverse problem presented in Eq. (2), the library search method first discretizes the parameter domain Ω, then the forward modeling procedure is used to calculate the corresponding signature for each discretized grid point to construct the signature library. The search algorithm is used to search for the grid point whose simulated signature minimizes Eq. (2). However, as emphasized in the Introduction, the conventional library search algorithm cannot effectively deal with the case of large signature library. In this paper, we propose a heuristic search algorithm, which is based on the principle of gradient-based iteration and can guarantee the fast search speed for an arbitrary scale library. To describe the details of our proposed algorithm, we first present the GN algorithm. Supposing the current ith iteration result is represented by $\mathbf{x}^{(i)}$, the goal of GN algorithm is to calculate the iteration direction to obtain the next iteration result $\mathbf{x}^{(i+1)}$. It is well known that a necessary condition for $\mathbf{x}^{(i+1)}$ being the minimum of $\mathbf{F}(\mathbf{x})$ is $\nabla F(\mathbf{x}^{(i+1)}) = 0$. 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 $\mathbf{x}^{(i)}$ by

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

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

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

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

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

$\nabla^2 F(\mathbf{x}^{(i)})$ is the $n \times n$ Hessian matrix at $\mathbf{x}^{(i)}$. Usually, the calculation of Hessian matrix is very time consuming; hence for simplicity, it is usually approximated by

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

We substitute Eqs. (4) and (6) into Eq. (3) and let Eq. (3) equals zero, we will have

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

Further, we will have

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

supposing the $\mathbf{J}(\mathbf{x}^{(i)})^{\mathrm{T}}\mathbf{w}\mathbf{J}(\mathbf{x}^{(i)})$ is nonsingular. By conducting Eqs. (3)–(8) iteratively, we can get the optimum result $\hat{\mathbf{x}}$. Compared with the conventional library search method which only contains the numerical signature library used to calculate the $\Delta\mathbf{y}^{(i)}$, Eq. (8) provides a search direction for next iteration result and contains the more important information—the Jacobian $\mathbf{J}(\mathbf{x}^{(i)})$, which is only the function of current iteration result $\mathbf{x}^{(i)}$. Hence, we can use Eq. (8) to give the search direction in the library search, on the premise of knowing the Jacobian of all the sets of profile parameters in the library.

Based on the above discussion, we propose a heuristic search algorithm to realize the fast search of global solution. The basic procedure of the proposed algorithm is described as follows.

Step 1: Linearize the range of the jth profile parameter into K_j discrete values; thus, for n profile parameters, we will have $\prod_{j=1}^{n} K_j$ sets of profile parameters in all. Then, the forward modeling is used to generate $\prod_{j=1}^{n} K_j$ simulated signatures to construct the signature library. Besides, a Jacobian library, each element of which is the Jacobian of the corresponding set of profile parameters in the signature library, is also constructed off-line.

Step 2: For the in-line search procedure, an initial set of profile parameters $\mathbf{x}^{(0)}$ is selected in the $\prod_{j=1}^{n} K_j$ sets. Then, the corresponding simulated signature $\mathbf{f}(\mathbf{x}^{(0)})$ and Jacobian $\mathbf{J}(\mathbf{x}^{(0)})$ of $\mathbf{x}^{(0)}$ can be obtained immediately in the signature and Jacobian libraries, respectively. By using Eq. (8), we can obtain the parameter departure vector $\Delta\mathbf{x}^{(0)}$; then, the newly searched result $\mathbf{x}^{(1)}$ can be obtained by $\tilde{\mathbf{x}}^{(0)} = \mathbf{x}^{(0)} + \Delta\mathbf{x}^{(0)}$.

Step 3: Notice that the newly searched result $\tilde{\mathbf{x}}^{(0)}$ may not be one of the $\prod_{j=1}^{n} K_j$ sets of profile parameters; thus, we should find the set of profile parameters from the $\prod_{j=1}^{n} K_j$ sets that is closest to $\tilde{\mathbf{x}}^{(0)}$. The term "closest" means that each profile parameter in this set is closest to the corresponding element in $\tilde{\mathbf{x}}^{(0)}$. This closest set of profile parameters is taken as the newly searched result $\mathbf{x}^{(1)}$.

Step 4: Repeating steps 2 and 3 until the difference between each profile parameter of the current searched set $\mathbf{x}^{(k)}$ and the corresponding element in the newly searched one is less than half of the grid interval. Then, stop the search, and output $\mathbf{x}^{(k)}$ as the global solution of the inverse problem presented in Eq. (2).

By following the procedures from step 1 to step 4, we can quickly find the solution in the library. Here, we should emphasize that the principle of heuristic search algorithm is built on top of the GN algorithm, which is one kind of locally nonlinear regression. Hence, to enable the heuristic search algorithm, it is important to give an appropriate initialization in advance. Fortunately, for most cases in optical scatterometry, setting the nominal values as the initialization can usually result in a reasonable solution. As mentioned in the Introduction, in practice, the abnormally distributed measurement errors in the measured signature will lead the solution obtained by LSQ function bias the actual values. Thus, it is of great importance to develop a method to suppress the influence of abnormally distributed measurement errors on the solution. In this paper, we propose a robust correction method based on the principle of *robust statistics* to correct the solution $\hat{\mathbf{x}}$ obtained in step 4. The additional correction procedure is taken as step 5.

Step 5: For simplicity, we use $\hat{\mathbf{J}}$ and $\hat{\mathbf{f}}$ to represent the corresponding Jacobian and simulated signature of the global

solution $\hat{\mathbf{x}}$, respectively, which can be picked out from the Jacobian and signature libraries, respectively. The correction parameter departure vector $\Delta\mathbf{x}^*$ can be obtained by solving the following regression problem, which is:

$$\Delta\mathbf{x}^* = \arg\min_{\Delta\hat{\mathbf{x}}\in\hat{\Omega}}\left\{\sum_{j=1}^{m}\rho[\Delta\hat{y}_j + \hat{\mathbf{J}}_j\Delta\hat{\mathbf{x}}]\right\} = \arg\min_{\Delta\hat{\mathbf{x}}\in\hat{\Omega}}\left\{\sum_{j=1}^{m}\rho[\hat{r}_j]\right\}, \tag{9}$$

where $\hat{\Omega}$ is the parameter domain of $\Delta\hat{\mathbf{x}}$, $\Delta\hat{y}_j$ is the jth element of $\Delta\hat{\mathbf{y}}$, which is $\Delta\hat{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{f}}$, and $\hat{r}_j$ represents the term $\Delta\hat{y}_j + \hat{\mathbf{J}}_j\Delta\mathbf{x}$. ρ is a robust objective function whose first derivative fulfills $\rho'(p) = \omega(p)\times p$, where p is an arbitrary scalar variable, and $\omega(p)$ is the weighted function that will be defined and discussed in the following contents. To solve the problem in Eq. (9), the GN algorithm could be used. Starting with an initial guess $\Delta\hat{\mathbf{x}}^{(0)}$, GN algorithm can be written in the context as

$$\Delta\hat{\mathbf{x}}^{(k+1)} = \Delta\hat{\mathbf{x}}^{(k)} - (\hat{\mathbf{J}}^{\mathrm{T}}\mathbf{Q}^{(k)}\hat{\mathbf{J}})^{-1}\hat{\mathbf{J}}^{\mathrm{T}}\mathbf{P}^{(k)}, \tag{10}$$

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

$$Q_{jj} = \rho''[\hat{r}_j^{(k)}], \tag{11}$$

in which ρ'' represent the second derivative with respect to $\hat{r}_j^{(k)}$. Defining the weight function $\omega(p) = \rho'(p)/p$ (where p is an arbitrary scalar variable), we can get another iteration scheme

$$\Delta\hat{\mathbf{x}}^{(k+1)} = \Delta\hat{\mathbf{x}}^{(k)} - (\hat{\mathbf{J}}^{\mathrm{T}}\mathbf{S}^{(k)}\hat{\mathbf{J}})^{-1}\hat{\mathbf{J}}^{\mathrm{T}}\mathbf{S}^{(k)}(\Delta\hat{\mathbf{y}} + \hat{\mathbf{J}}\Delta\hat{\mathbf{x}}^{(k)}), \tag{12}$$

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

$$S_{jj} = \omega(\hat{r}_j^{(k)}) = \omega(\Delta\hat{y}_j + \hat{\mathbf{J}}_j\Delta\hat{\mathbf{x}}^{(k)}). \tag{13}$$

Equation (12) is usually attributed to Beaton and Tukey.[26] This method only requires the knowledge of the weight function $\omega(p)$. In this paper, Andrews's hard redescender[27] is used, which is

$$\omega(p) = \begin{cases} \dfrac{c_{\mathrm{A}}}{p}\sin pc_{\mathrm{A}}, & |p| \le \pi c_{\mathrm{A}} \\ 0, & |p| > \pi c_{\mathrm{A}}, \end{cases} \tag{14}$$

where c_{A} is the tuning constant and p is an arbitrary scalar variable. The definition of $\omega(p)$ in Eq. (14) means $\omega(p)\to 0$ for sufficiently large $|p|$. In this paper, the c_{A} is set as 1.339 to give coefficient estimates that are approximately 95% as statistically efficient as the LSQ estimates, provided the measurement error follows a normal distribution with no outliers. By using the above robust regression, we can obtain the corrected parameter departure vector $\Delta\mathbf{x}^*$, then the searched global solution $\hat{\mathbf{x}}$ can be corrected by using

$$\mathbf{x}^* = \Delta\mathbf{x}^* + \hat{\mathbf{x}}. \tag{15}$$

To give a detailed explanation of why the final result $\mathbf{x}^*$ obtained by Eq. (15) in step 5 is more accurate than the searched result $\hat{\mathbf{x}}$, we present the geometric illustration in Fig. 1. As depicted in Fig. 1, the true values of profile parameters $\mathbf{x}_{\mathrm{true}}$ locate out of the parameter domain (represented by a hypersurface S), which is caused by the measurement errors. From the perspective of functional analysis, the optimal solution that can be obtained within the framework of LSQ function is the one on the hypersurface S that is closest to $\mathbf{x}_{\mathrm{true}}$.[28] Thus, the optimal solution represented by $\mathbf{x}_{\mathrm{map}}$ in Fig. 1 is the intersection point of the hypersurface S and its normal, which goes through the point $\mathbf{x}_{\mathrm{true}}$. $\mathbf{x}_{\mathrm{map}}$ can be infinitely approximated either by setting a extremely terminal condition of nonlinear regression or by setting the infinitesimal grid interval in library search. However, this is not for real in consideration of the limited computing resources. Actually, the search solution of library search method $\hat{\mathbf{x}}$ rarely equals the optimal solution $\mathbf{x}_{\mathrm{map}}$ since the value of grid interval is limited (but one should notice that there is still a possibility that $\mathbf{x}_{\mathrm{map}}$ locates on a grid point by coincidence). Moreover, the gap between $\mathbf{x}_{\mathrm{true}}$ and $\mathbf{x}_{\mathrm{map}}$ caused by measurement errors is another important limited factor. This gap cannot be effectively eliminated since the two-norm based function is quite sensitive to the outliers, as emphasized in the Introduction. To break the bottleneck of LSQ function, we recently proposed a robust regression method based on the principle of robust statistics to effectively reduce the gap. Step 5 in the present paper is an extension of the method of Ref. 24 that uses an additional robust regression procedure on the basis of the search solution $\hat{\mathbf{x}}$ to give a corrected solution $\mathbf{x}^*$. Consequently, $\mathbf{x}^*$ may be more accurate than $\mathbf{x}_{\mathrm{map}}$ if the grid interval is small enough, as shown in Fig. 1.

III. RESULTS

A. Experimental setup and sample description

The experimental setup used in this article is a dual-rotating compensator Mueller matrix ellipsometer (DRC-MME) prototype suitable from ultraviolet to infrared spectral bands and developed in our lab.[29,30] As schematically shown in Fig. 2(a), the system setting of the

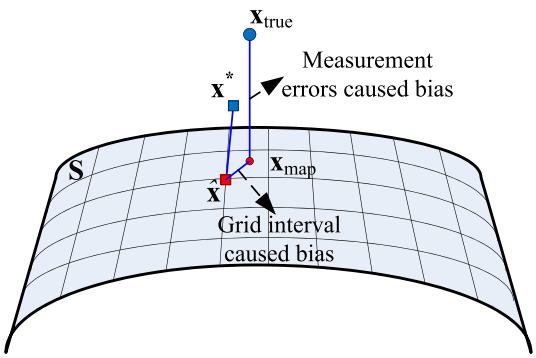

FIG. 1. (Color online) Geometrical illustration of the corrected solution $\mathbf{x}^*$ to the inverse problem in optical scatterometry.

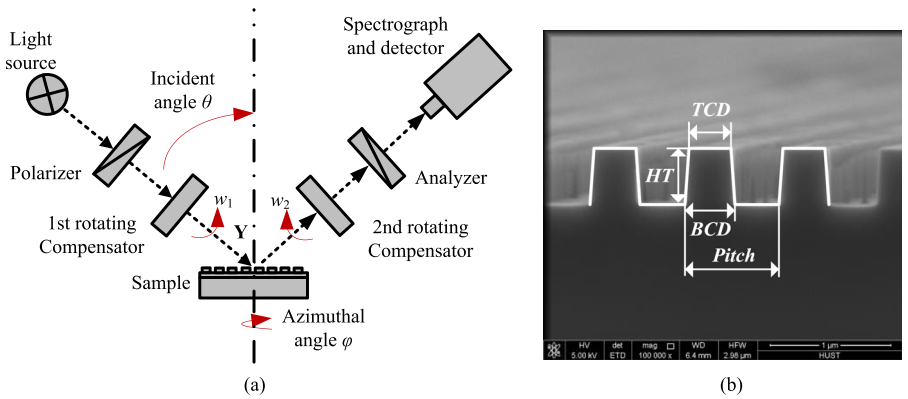

FIG. 2. (Color online) (a) Measurement setup of dual-rotating-compensator ellipsometer and (b) cross-section SEM image of one-dimensional trapezoidal etched silicon grating.

DRC-MME in order of light propagation is PCr_1SCr_2A, where P and A stand for the fixed polarizer and analyzer, Cr_1 and Cr_2 refer to the first and second frequency-coupled rotating compensators, and S stands for the sample. With the light source used in this ellipsometer, the wavelengths available are in the 200 and 1000 nm range, covering the spectral range of 200–800 nm with 20 nm increment used in this article. In this paper, the signature is set as the form of Mueller matrix elements. With this dual-rotating compensator setting, we can obtain the full Mueller matrix elements of the sample under measurement.

The investigated sample is a one-dimensional (1D) etched Si grating, whose SEM cross-section image is shown in Fig. 2(b). The etched Si grating is chosen for this study due to its long-term dimensional stability, higher refractive index contrast, and relevance to the semiconductor industry.[31] Optical properties of Si are taken from Ref. 32. As depicted in Fig. 2(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 *HT*, and period *Pitch*. Dimensions of the structural parameters that obtained by SEM are $TCD = 350$ nm, $HT = 472$ nm, and $BCD = 383$ nm. In the following experiments, profile parameters of the Si grating that need to be extracted include *TCD*, *HT*, and *BCD*, while the grating period is fixed at its nominal dimension, i.e., $Pitch = 800$ nm. The in-house computer program used for data analysis is based on MATLAB® high-level language and interactive environment [version (R2013b), The MathWorks, Inc., Natick, MA, USA], which will be run on a 2.3 GHz Intel i5–2410M personal computer. Here, the in-house computer program mainly contains two parts, which are the forward modeling based on RCWA algorithm[8] and the inverse extraction based on least square regression[10,13] and robust estimation.[24,25]

B. Numerical results

In this section, the heuristic search algorithm including step 1 to step 4 is first performed to obtain the solution in the library. The grid intervals for *TCD*, *HT*, and *BCD* with parameter ranges 330–370, 450–490, and 360–400 nm, respectively, are all set as 2 nm, which means there are 9261 (21^3) sets of profile parameters in all. The in-house computer program is used to calculate the corresponding Mueller matrices together with the Jacobians of the 9261 sets under 45° incident angle and 10° azimuthal angle to construct the signature and Jacobian library, respectively. Since there are three profile parameters to be extracted, the least time consumption of the Jacobian calculation triples that of the signature calculation. Supposing the true values of the extracted *TCD*, *HT*, and *BCD* are 350, 470, and 380 nm, respectively, the in-house computer program is used to generate the corresponding Mueller matrix. Then, the measurement errors, which consist of the random errors and system errors, are simulated and added into the Mueller matrix. The random errors are treated as normally distributed errors, whose standard deviation or noise level at a data point is set as a fraction of root-mean-square (rms) in the Mueller matrix over all of the data points.[33] Thus, all the weighted random errors (random errors divided by the standard deviations) are nearly in the range of −3 to 3. The fractions of the wavelengths differ from each other, but are all within the range of 0%–5%. For the system errors, we simulate some known error sources here, which are 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 was directly taken from Ref. 23, while for the last one the bias of each estimated standard deviation was randomly chosen within the range of ±20% of the standard deviation.

The heuristic search algorithm with an initialization (342, 464, and 384 nm) and the linear search method are performed for the above "measured" signature, respectively. The searched results together with the time consumptions are presented in Table I. The time consumption of heuristic search is only 0.0025 s, which is much less than the 0.0510 s of linear search. The result (352, 470, and 378 nm) obtained by heuristic search is slightly different to the result (354, 470, and 376 nm) by linear search. The small difference is due to the fact that the grid point obtained at each iteration is not the actual iterative value but its approximation, and the Jacobian of the grid point also will lead to the approximate gradient vector while not the actual one. Moreover, the

TABLE I. Searched values of *TCD*, *HT*, and *BCD* together with the time consumptions by different methods.

Method	Library scale	*TCD* (nm)	*HT* (nm)	*BCD* (nm)	Time (s)
Linear search	9261	354 ±0.80	470 ±0.46	376 ±0.84	0.0510
	103 823	354 ±0.80	470 ±0.46	376 ±0.84	0.5824
Heuristic search	9261	352 ±0.72	470 ±0.41	378 ±0.77	0.0025
	103 823	352 ±0.72	470 ±0.41	378 ±0.77	0.0027

parameter uncertainties obtained by heuristic search are ±0.72, ±0.41, and ±0.77 nm, respectively, which are not much different with that obtained by linear search. On all accounts, this difference is smaller or equal to one grid interval. To demonstrate the proposed heuristic search algorithm is able to realize the fast search for arbitrary large library, we generate a larger library that contains 103 823 (47^3) sets of profile parameters (the parameter ranges of *TCD*, *HT*, and *BCD* are enlarged to 304–396, 424–516, and 334–426 nm, respectively, while the grid intervals are all kept as 2 nm). The heuristic search with the same initialization and the linear search are performed, and the search results and the time consumptions are also presented in Table I. In Table I, we can find that the heuristic search achieves the same result within nearly the same time (0.0027 vs 0.0025 s) in a much larger library. As a contrast, the time consumption of linear search in the large library is nearly 11 times larger than that in the small library.

We present the iterative values of *TCD*, *HT*, and *BCD* in the heuristic search algorithm for large and small libraries, respectively. As presented in Fig. 3, only after three iterations, the solution of *TCD*, *HT*, and *BCD* is achieved. Moreover, as expected, the iterative results for large and small libraries coincide with each other, which have demonstrated the proposed heuristic search algorithm is able to realize the fast search for a library with arbitrary scale.

The robust correction method presented in step 5 is used to correct the searched results (352, 470, and 378 nm) obtained by heuristic search algorithm in Sec. II. To make a comparison, the Levenberg–Marquardz (LM) algorithm is also used to extract the parameter values. The LM is one kind of gradient-based optimization algorithm, by using the strict terminal condition the LM can achieve the minimization as presented in Eq. (2). Thus, the extracted result of LM can be regarded as $\mathbf{x}_{\text{map}}$, as presented in Fig. 1.

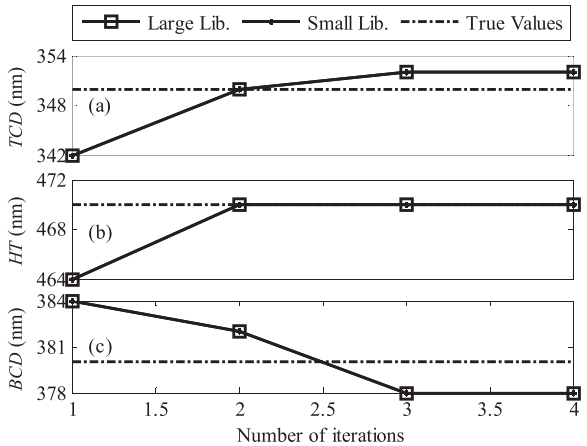

FIG. 3. Iterative values of (a) *TCD*, (b) *HT*, and (c) *BCD* in the heuristic search procedure for the large and small libraries, respectively.

TABLE II. Extracted values of *TCD*, *HT*, and *BCD* by different methods, respectively.

Methods	*TCD* (nm)	*HT* (nm)	*BCD* (nm)
LM algorithm	352 ±0.72	469.9 ±0.41	378.3 ±0.76
Robust correction	350.2 ±0.68	470 ±0.39	380.1 ±0.73

The extracted results of *TCD*, *HT*, and *BCD* by different methods are presented in Table II. It can be observed that the robust correction method achieves the more accurate results than the LM algorithm. Moreover, the parameter uncertainties obtained by robust correction are ±0.68, ±0.39, and ±0.73 nm, respectively, which are not much different with that obtained by linear search. Further, we generated ten more measured Mueller matrices with each Mueller matrix corresponds to the specific measurement errors by randomly choosing the bias of each estimated standard deviation as well as the noise level of normal distributed errors. The results obtained by robust correction (on the basis of heuristic searched result) and LM methods are presented in Fig. 4, in which we can observe that the robust correction results are significantly more accurate than that of the LM. To demonstrate the existence of outliers, we calculated the weighted "measurement" errors and present them in Fig. 5, in which we can find that many data are out of the range −3 to 3. The above simulation has demonstrated that the robust correction method presented in step 5 goes beyond the framework of LSQ and is able to deal with the outliers effectively.

To further verify the generality of the proposed heuristic search algorithm and robust correction method, we perform the simulations for two additional examples, i.e., a Si grating

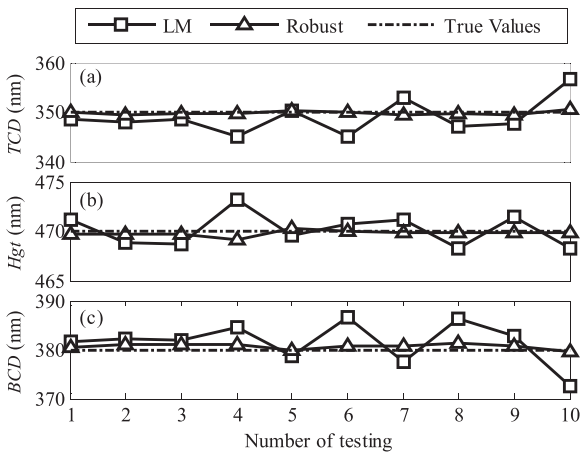

FIG. 4. Numerical extracted (a) *TCD*s, (b) *HT*s, and (c) *BCD*s for ten more generated measured Mueller matrices by LM and robust correction, respectively.

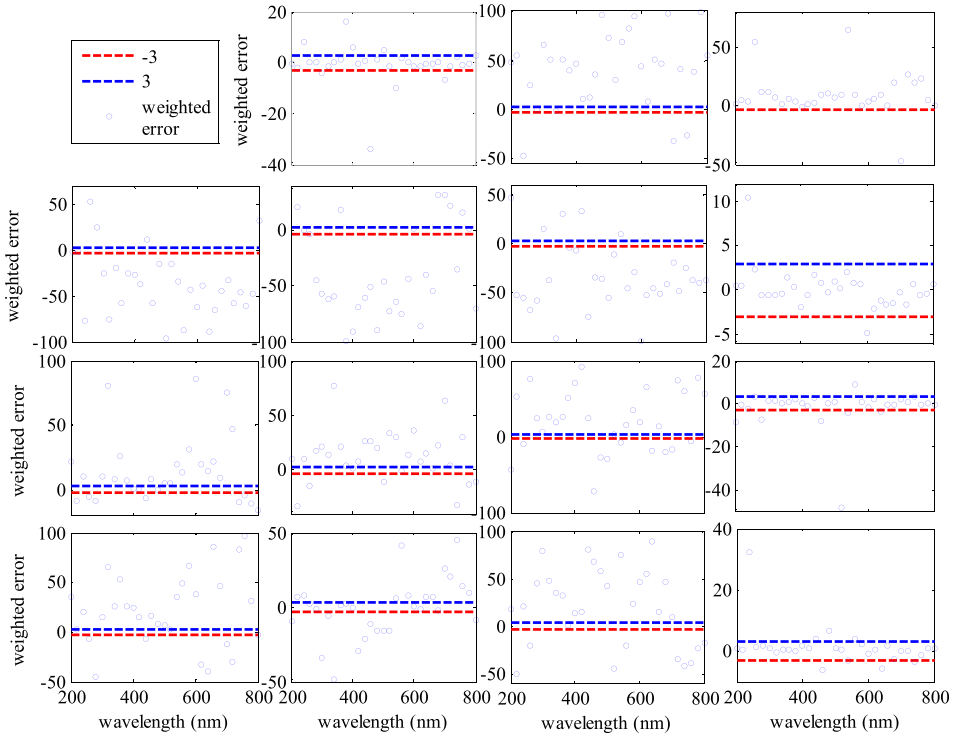

FIG. 5. (Color online) Weighted measurement errors of Mueller matrix elements.

with the true *HT* of 800 nm at current *TCD* (350 nm) and *BCD* (380 nm), and a Si grating with the true *TCD* of 420 nm at current *HT* (470 nm) and *BCD* (380 nm). In case 1, we set the parameter ranges of *TCD*, *HT*, and *BCD* as 304–396, 754–846, and 334–426 nm, respectively. By choosing the grid intervals of these ranges as 2 nm and using our in-house program, we can finally obtain a signature library and a Jacobian library with element numbers of both 103 823 (47^3). The heuristic search and robust correction are then performed on the basis of the signature and Jacobian libraries, whose corresponding results are presented in Table III. As expected, the time consumption of linear search is about 335 times larger than that of heuristic search. After a robust correction step, the corrected result (349, 800, and 381.8 nm) becomes closer to the true values (350, 800, and 380 nm). Moreover, we can also find that the parameter uncertainties obtained by different methods are quite similar with each other. We then perform the simulation for the case 2. We set the parameter ranges of *TCD*, *HT*, and *BCD* as 374–466, 424–516, and 334–426 nm, respectively. The grid interval is also set as 2 nm; thus, we can obtain the signature and Jacobian libraries with elements number of both 103 823 (47^3). The extracted values of geometrical parameters as well as the time consumption are also presented in Table III, in which we can find, which is the same as that of case 1, by performing the proposed heuristic search and robust correction, the more accurate result can be obtained with a faster speed than using the conventional library search method.

C. Experimental results

In this section, we perform experiments to further demonstrate the effectiveness of the proposed heuristic search algorithm and robust correction method. The measured Mueller matrix is obtained under an incident angle of 45° and an azimuthal angle of 10°.

The generated large signature library (contains 103 823 grid points) together with the corresponding Jacobian library in Sec. III B is used here. To show the details of the proposed heuristic search and robust correction, we present the extracted parameter values at each procedure of this method in Fig. 6. Starting with the initialization (342, 464, and 384 nm), only after three iterations the convergent set (344,

TABLE III. Extracted values of *TCD*, *HT*, and *BCD* together with the time consumptions for different samples by different methods.

Example	Method	*TCD* (nm)	*HT* (nm)	*BCD* (nm)	Time (s)
Case 1	Linear search	348 ±2.66	800 ±2.09	382 ±2.73	0.6033
	Heuristic search	348 ±2.66	800 ±2.09	382 ±2.73	0.0018
	Robust correction	349 ±2.68	800 ±2.09	381.8 ±2.74	
Case 2	Linear search	420 ±3.96	474 ±4.27	378 ±4.66	0.5916
	Heuristic search	418 ±4.20	474 ±4.28	378 ±4.74	0.0025
	Robust correction	420 ±4.02	469.7 ±4.19	379.1 ±4.72	

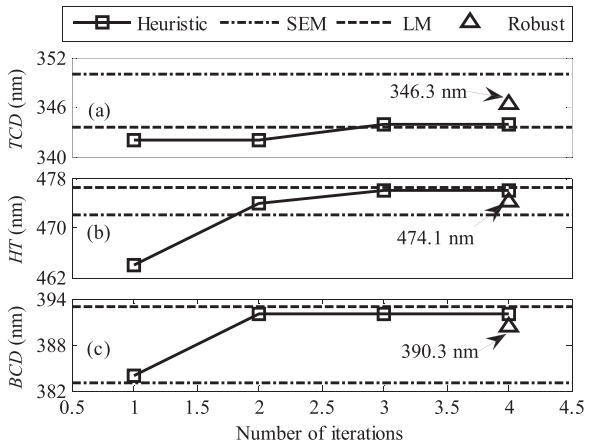

FIG. 6. Values of (a) *TCD*, (b) *HT*, and (c) *BCD* obtained by the heuristic search, the robust correction, the SEM, and LM, respectively.

476, and 392 nm) is obtained. Moreover, we can find that the convergent set is not only the same as that obtained by linear search method but also it corresponds to the far less time consumption, as listed in Table IV.

After the heuristic search, the obtained convergent set (344, 476, and 392 nm) is then corrected by the robust correction method. This correction is based on the set (344, 476, and 392 nm) itself as well as its corresponding Jacobian in the Jacobian library. After the correction, we can find that the corrected results (346.3, 474.1, and 390.3 nm) represented by up-triangles in Fig. 6 are closer to the SEM measured results. Moreover, the corrected result is also more accurate than the result (343.5, 476.4, and 393 nm) obtained by LM algorithm, as presented in Fig. 6 and Table IV. The

TABLE IV. Experimentally extracted values of *TCD*, *HT*, and *BCD* together with the time consumptions by different methods.

Method	*TCD* (nm)	*HT* (nm)	*BCD* (nm)	Time (s)
Linear search	344 ±1.45	476 ±0.88	392 ±1.58	0.5937
Heuristic search	344 ±1.45	476 ±0.88	392 ±1.58	0.0023
LM algorithm	343.5 ±1.44	476.4 ±0.89	393 ±1.57	
Robust correction	346.3 ±1.43	474.1 ±0.87	390.3 ±1.57	

above experimental results have verified the validity of explanation about Fig. 1.

Moreover, to explain why the robust correction method is able to achieve the more accurate result, 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 weighted fitting differences between the measured Mueller matrix and the best calculated Mueller matrix by the LM algorithm, as shown in Fig. 7. The weighted fitting differences are the fitting differences divided by the standard deviations. From Fig. 7, we indeed observe many large data points that are out of the range −3 to 3. From the statistics point of view, the value of a standard normal variable (a normal variable divided by its standard deviation) locates in the range of −3 to 3 with a possibility larger than 99.7%, which means it is almost impossible to obtain a value out of the range of −3 to 3. However, as Fig. 7 shows, many data points are out of the range −3 to 3, which indirectly demonstrate that the measurement errors corresponding to these data points are abnormally distributed.

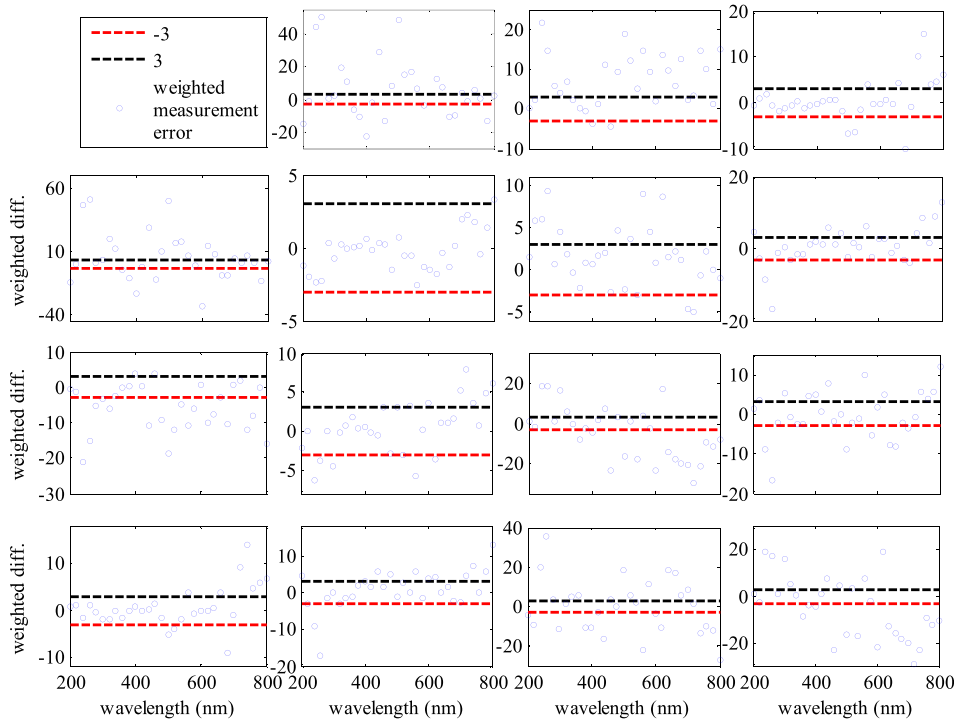

FIG. 7. (Color online) Weighted fitting differences of Mueller matrix elements.

In summary, from the above experiments, we can conclude that the proposed heuristic search algorithm and robust correction method are capable of realizing the fast and accurate parameter extraction.

IV. CONCLUSIONS

In the present paper, to realize the fast retrieve of best matched simulated signature for the measured one and to reduce the effect of outliers on the parameter extraction, we proposed a heuristic search algorithm and a robust correction method, respectively. Instead of searching in the signature library, the heuristic search algorithm takes advantage of an extra constructed Jacobian library using the principle of gradient-based iteration algorithms, by which the fast search speed can be guaranteed. Once the best matched simulated signature is found, the robust correction is applied to achieve the more accurate result using the best matched signature as well as its corresponding Jacobian in the Jacobian library. Numerical and experimental results based on a 1D etched Si grating have verified the validity of the proposed heuristic search algorithm and robust correction 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 (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).

[1]C. J. Raymond, M. R. Murnane, S. L. Prins, S. S. H. Naqvi, J. W. Hosch, and J. R. McNeil, J. Vac. Sci. Technol., B **15**, 361 (1997).
[2]C. J. Raymond, *Handbook of Silicon Semiconductor Metrology* (Academic, New York, 2001).
[3]Y. Ohkawa, Y. Tsuji, and M. Kashiba, J. Opt. Soc. Am. A **13**, 1006 (1996).
[4]Y. Nakata and M. Kashiba, J. Opt. Soc. Am. A **7**, 1494 (1990).
[5]H. Ichikawa, J. Opt. Soc. Am. A **15**, 152 (1998).
[6]M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, J. Opt. Soc. Am. A **12**, 1068 (1995).
[7]L. Li, J. Opt. Soc. Am. A **13**, 1870 (1996).
[8]S. Y. Liu, Y. Ma, X. G. Chen, and C. W. Zhang, Opt. Eng. **51**, 081504 (2012).
[9]H. T. Huang, W. Kong, and F. L. Terry, Appl. Phys. Lett. **78**, 3983 (2001).
[10]C. W. Zhang, S. Y. Liu, T. L. Shi, and Z. R. Tang, J. Opt. Soc. Am. A **26**, 2327 (2009).
[11]J. L. Zhu, S. Y. Liu, C. W. Zhang, X. G. Chen, and Z. Q. Dong, J. Micro/Nanolithogr., MEMS, MOEMS **12**, 013004 (2013).
[12]X. G. Chen, S. Y. Liu, C. W. Zhang, and J. L. Zhu, Measurement **46**, 2638 (2013).
[13]C. W. Zhang, S. Y. Liu, T. L. Shi, and Z. R. Tang, J. Opt. Soc. Am. A **28**, 263 (2011).
[14]X. G. Chen, S. Y. Liu, C. W. Zhang, and J. Hao, Appl. Opt. **52**, 6726 (2013).
[15]X. Niu, N. Jakatdar, J. Bao, and C. J. Spanos, IEEE Trans. Semicond. Manuf. **14**, 97 (2001).
[16]J. L. Bentley, Commun. ACM **18**, 509 (1975).
[17]D. T. Lee and C. K. Wong, Acta Inf. **9**, 23 (1977).
[18]A. Gionis, P. Indyk, and P. Motwani, *Proceedings of the 25th International Conference on VLDB* (Morgan Kaufmann Publishers Inc., San Francisco, 1999), p. 518.
[19]G. E. Jellison, Thin Solid Films **234**, 416 (1993).
[20]M. A. Henn, H. Gross, F. Scholze, M. Wurm, C. Elster, and M. Bar, Opt. Express **20**, 12771 (2012).
[21]R. C. Geary, Biometrika **34**, 209 (1947).
[22]A. Kato and F. Scholze, Appl. Opt. **49**, 6102 (2010).
[23]X. G. Chen, C. W. Zhang, and S. Y. Liu, Appl. Phys. Lett. **103**, 151605 (2013).
[24]J. L. Zhu, S. Y. Liu, X. G. Chen, C. W. Zhang, and H. Jiang, Opt. Express **22**, 22031 (2014).
[25]J. L. Zhu, S. Y. Liu, H. Jiang, C. W. Zhang, and X. G. Chen, Opt. Lett. **40**, 471 (2015).
[26]A. E. Beaton and J. W. Tukey, Technometrics **16**, 147 (1974).
[27]D. Andrews, F. Bickel, F. Hampel, P. Huber, W. Rogers, and J. Tukey, *Robust Estimation of Location: Survey and Advances*, 1st ed. (Princeton University, NJ, 1972).
[28]W. Rudin, *Functional Analysis*, 2nd ed. (McGraw-Hill, New York, 1991).
[29]S. Y. Liu, X. G. Chen, and C. W. Zhang, Thin Solid Films **584**, 176 (2015).
[30]X. G. Chen, S. Y. Liu, C. W. Zhang, H. Jiang, Z. C. Ma, T. Y. Sun, and Z. M. Xu, Opt. Express **22**, 15165 (2014).
[31]X. G. Chen, S. Y. Liu, C. W. Zhang, and J. Hao, J. Micro/Nanolithogr., MEMS, MOEMS **12**, 033013 (2013).
[32]C. M. Herzinger, B. Johs, W. A. McGahan, and J. A. Woollam, J. Appl. Phys. **83**, 3323 (1998).
[33]R. M. Al-Assaad and D. M. Byrne, J. Opt. Soc. Am. A **24**, 326 (2007).