IOP Publishing

Measurement Science and Technology

Meas. Sci. Technol. 29 (2018) 045014 (9pp)

https://doi.org/10.1088/1361-6501/aaa8af

Probing optimal measurement configuration for optical scatterometry by the multi-objective genetic algorithm

Xiuguo Chen[1], **Honggang Gu**[1], **Hao Jiang**[1], **Chuanwei Zhang**[1,2] **and Shiyuan Liu**[1,2]

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

E-mail: xiuguochen@hust.edu.cn and chuanweizhang@hust.edu.cn

Received 7 November 2017, revised 12 January 2018
Accepted for publication 18 January 2018
Published 14 March 2018

Abstract

Measurement configuration optimization (MCO) is a ubiquitous and important issue in optical scatterometry, whose aim is to probe the optimal combination of measurement conditions, such as wavelength, incidence angle, azimuthal angle, and/or polarization directions, to achieve a higher measurement precision for a given measuring instrument. In this paper, the MCO problem is investigated and formulated as a multi-objective optimization problem, which is then solved by the multi-objective genetic algorithm (MOGA). The case study on the Mueller matrix scatterometry for the measurement of a Si grating verifies the feasibility of the MOGA in handling the MCO problem in optical scatterometry by making a comparison with the Monte Carlo simulations. Experiments performed at the achieved optimal measurement configuration also show good agreement between the measured and calculated best-fit Mueller matrix spectra. The proposed MCO method based on MOGA is expected to provide a more general and practical means to solve the MCO problem in the state-of-the-art optical scatterometry.

Keywords: optical scatterometry, inverse problem, measurement configuration, multi-objective optimization, genetic algorithm

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

1. Introduction

Optical scatterometry, also referred to as optical critical dimension metrology, has become one of the most important techniques for critical dimension (CD) and overlay metrology in semiconductor industry due to its inherent speed, noncontact, nondestructive, and inexpensive merits over other techniques, such as scanning electron microscopy (SEM) and atomic force microscopy [1–9]. As a model-based metrology technique, the implementation of optical scatterometry involves two steps [10, 11]. First, in what is known as the forward problem, the signature of a nanostructure (typically a periodic sub-wavelength structure) under test is measured using a proper instrument. Here, the term *signature* represents the optical response of the nanostructure under test, which can be in forms of reflectance, ellipsometric angles, Stokes vector elements, or Mueller matrix elements etc. The second step involves establishment of a light-nanostructure interaction model to relate the signature with the structural parameters and extraction of the structural parameters under measurement from the measured signature by solving an inverse problem with the objective of finding an optimal input to the established model whose simulated signature can best match the measured one. Previous studies revealed that the measurement uncertainty of optical scatterometry depends on not only the quality (e.g. signal-to-noise ratio) of measured signatures but also the measurement configuration [12–19]. The latter is defined as the combination of measurement conditions in

© 2018 IOP Publishing Ltd Printed in the UK

optical scatterometry, such as wavelength, incidence angle, azimuthal angle, and/or polarization direction. Moreover, it was reported that some measurement configurations could help decorrelate the structural parameters [6], which is beneficial to avoid the solution of the inverse problem fall into local optima, especially when the number of floating parameters increases. For a given measuring instrument, a proper choice of the measurement configuration is a significant approach of low cost to achieve a higher measurement precision.

Several approaches have been proposed to optimize the measurement configuration for optical scatterometry over the past decades. Logofătu proposed a sensitivity analysis for fitting method by defining the sensitivity as the estimated standard uncertainties of structural parameters to optimize the measurement configuration for angle-resolved rotating-analyzer and angle-resolved phase-modulation scatterometers, respectively [12, 13]. Littau *et al* investigated several optimal diffraction signature scan path selection techniques to improve scatterometry precision [14]. Gross *et al* proposed an algorithm to determine the optimal measurement data set by minimizing the condition number of a Jacobian matrix with elements defined as partial derivatives of the diffraction signature with respect to the structural parameters [15]. Vagos *et al* developed an uncertainty and sensitivity analysis package that can be used to guide the optimization of the measurement model and azimuthal angle [16]. Foldyna *et al* proposed to choose the measurement configurations with small parameter correlations and small estimated standard uncertainties of structural parameters for Mueller matrix scatterometry (MMS, also called Mueller matrix ellipsometry based scatterometry or Mueller matrix based scatterometry) [17]. In order to achieve a higher measurement accuracy in MMS, we proposed to optimize the measurement configuration by minimizing the Frobenius norm an introduced configuration error propagating matrix [18]. In addition, we also introduced the global sensitivity analysis for guiding the choice of measurement configurations and meanwhile avoiding the local property of traditional partial derivative-based sensitivity analysis [19].

Since the semiconductor industry is primarily interested in reproducibility, how to choose measurement configurations that could yield a higher measurement precision is thus much preferred and also the focus of this research. Based on the reported MCO approaches in the literature [12–19], the most straightforward approach to fulfill the above goal is to directly optimize the estimated uncertainties in the extracted structural parameters, which typically vary at different measurement configurations and can be treated as the objective functions in optimization. We should note that since each structural parameter corresponds to an estimated uncertainty, the MCO problem will contain M objectives if there are M structural parameters under measurement. If we further take parameter correlations into account, there will be much more objectives (M objectives about estimated uncertainties and $M(M-1)/2$ objectives about parameter correlations) that need to be optimized. Obviously, the MCO problem based on the optimization of the estimated uncertainties in the extracted structural parameters is a multi-objective optimization problem and should be addressed with care. However, to the best of our knowledge, the MCO problem in optical scatterometry has never been surveyed from the point of view of multi-objective optimization. The commonly adopted approach in the literature is to optimize the objective functions (i.e. the estimated uncertainties of the structural parameters) one by one [12, 13, 16, 17]. Since the optimal measurement configurations are typically different for different structural parameters, a trivial compromise among the selected measurement configurations has to be made finally.

In this paper, the multi-objective genetic algorithm (MOGA) is introduced to deal with the MCO problem in optical scatterometry. The classical approach to solve a multi-objective optimization problem is to assign a weight to each objective function so that the problem is converted to a single objective optimization problem. The main difficulty with this approach is selecting a proper weight vector for each run. Typically, different weight vectors will lead to different optimal results. For this reason, we resort to the multi-objective optimization algorithm, by which multiple objectives could be optimized simultaneously. The genetic algorithm (GA) is a metaheuristic inspired by the mechanism of natural selection and natural genetics [20]. In the GA, a population of candidate solutions to the optimization problem is evolved towards better solutions through successive iterations called generations. A new generation is formed using the bio-inspired operators, such as mutation, crossover and selection, according to the fitness value of each candidate solution. After several generations, the algorithm is expected to converge to the optimum or sub-optimal solution of the problem. Due to the population-based nature of the GA, it is an ideal candidate for solving multi-objective optimization problems [21, 22], and is thus chosen to probe the optimal measurement configuration for optical scatterometry.

The remainder of this paper is organized as follows. In section 2, the inverse problem in optical scatterometry is firstly revisited. The estimation of the uncertainties in the extracted structural parameters as well as parameter correlations is also briefly introduced. Then, we reformulate the MCO problem in optical scatterometry as a multi-objective optimization problem. The solution of the MCO problem by MOGA is briefly introduced with an emphasis on important concepts related with MOGA. In section 3, the MMS is exemplified to illustrate the implementation of MOGA for optimizing the combination of incidence and azimuthal angles in measuring a Si grating. Monte Carlo simulations are performed to demonstrate the feasibility of MOGA in handling the MCO problem. Some conclusions are finally drawn in section 4.

2. Method

2.1. Inverse problem in optical scatterometry

Without loss of generality, we denote the structural parameters under measurement as a M-dimensional vector $\mathbf{p} = [p_1, p_2, \ldots, p_M]^T$, where the superscript 'T' represents the transpose. The signature measured at a measurement configuration $\mathbf{a} = [a_1, a_2, \ldots, a_L]^T$ is denoted as an N-dimensional vector

$\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$, where $a_1, a_2, \ldots, a_L$ represent the measurement conditions such as wavelength, incidence angle and azimuthal angle in the experiment, and $y_1, y_2, \ldots, y_N$ represent the measured data in terms of reflectance, ellipsometric angles, Stokes vector elements, or Mueller matrix elements etc. The corresponding theoretical signature calculated at the same measurement configuration for any vector $\mathbf{p}$ is denoted as $\mathbf{f}(\mathbf{p}, \mathbf{a}) = [f_1(\mathbf{p}, \mathbf{a}), f_2(\mathbf{p}, \mathbf{a}), \ldots, f_N(\mathbf{p}, \mathbf{a})]^T$. The χ^2 function is usually adopted to estimate the fitting error between the measured and theoretical signatures, which is defined as

$$\chi^2 = \sum_{k=1}^{N} w_k [y_k - f_k(\mathbf{p}, \mathbf{a})]^2 = [\mathbf{y} - \mathbf{f}(\mathbf{p}, \mathbf{a})]^T \mathbf{W} [\mathbf{y} - \mathbf{f}(\mathbf{p}, \mathbf{a})], \tag{1}$$

where w_k ($k = 1, 2, \ldots, N$) are the weighting factors, which are usually chosen to be $w_k = 1/\sigma^2(y_k)$ with $\sigma^2(y_k)$ being the variances of the measured signature, and $\mathbf{W}$ is an $N \times N$ diagonal matrix with diagonal elements being w_k. The inverse problem in optical scatterometry is typically formulated as a least-square regression problem such that

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p} \in \Omega} [\mathbf{y} - \mathbf{f}(\mathbf{p}, \mathbf{a})]^T \mathbf{W} [\mathbf{y} - \mathbf{f}(\mathbf{p}, \mathbf{a})], \tag{2}$$

where $\hat{\mathbf{p}}$ denotes the solution of the inverse problem consisting of the extracted structural parameters, and Ω is the associated domain of the structural parameters. In the solution of equation (2), typically only $\mathbf{p}$ is floated while both $\mathbf{a}$ and $\mathbf{W}$ are fixed.

The uncertainty in the measured signature will be propagated into the extracted structural parameters in the solution of equation (2). The uncertainty in the extracted structural parameters can be estimated according to the covariance matrix of $\mathbf{p}$ defined as

$$\mathbf{C} = \left(\tilde{\mathbf{J}}^T \tilde{\mathbf{J}}\right)^{-1}, \tag{3}$$

where $\tilde{\mathbf{J}} = \mathbf{W}^{1/2} \mathbf{J}$ with $\mathbf{J}$ being the Jacobian matrix with respect to $\mathbf{p}$, whose (k, i)th element is given by

$$J_{ki} = \left. \frac{\partial f_k(\mathbf{p}, \mathbf{a})}{\partial p_i} \right|_{\mathbf{p} = \hat{\mathbf{p}}}. \tag{4}$$

With equation (3), the standard deviation of the extracted structural parameter p_i ($i = 1, 2, \ldots, M$) can be estimated from the ith diagonal element of $\mathbf{C}$ by

$$\sigma(p_i) = \sqrt{C_{ii}}. \tag{5}$$

We can further estimate the uncertainty in parameter p_i by

$$u(p_i) = \kappa \sigma(p_i), \tag{6}$$

where κ is the coverage factor associated with the prescribed confidence level. Since the uncertainty of Type B in optical scatterometry is typically constant at different measurement configurations, we only take the error sources of Type A into account in the uncertainty estimation in equation (6) [23]. According to equation (3), we can also estimate the correlation coefficient between parameters p_i and p_j by

$$\rho(p_i, p_j) = \frac{\sqrt{C_{ij}}}{\sqrt{C_{ii}} \sqrt{C_{jj}}}, \tag{7}$$

which takes values between -1 and 1 with 0 meaning no (linear) correlation and -1 or 1 meaning perfect (linear) correlation, and moreover, $\rho(p_i, p_j) = \rho(p_j, p_i)$.

2.2. Measurement configuration optimization by multi-objective genetic algorithm

As revealed in the previous work [12, 13], different measurement configurations will lead to extracted structural parameters with different precisions. It is highly desirable that if we could pick out the configurations that lead to a higher parameter precision, namely smaller uncertainties in the extracted structural parameters. In addition, some measurement configurations could yield weaker parameter correlations [6], which are beneficial to the solution of the inverse problem in optical scatterometry, especially for complex nanostructures with a large number of structural parameters under measurement. Since smaller uncertainties in the extracted structural parameters do not necessarily indicate weaker parameter correlations [17], it is also preferred to choose the configurations with smaller correlation coefficients between the structural parameters. Based on the above description, the MCO problem in optical scatterometry can be formulated as

$$\min_{\mathbf{a} \in \Xi} \left\{ \begin{array}{c} \max\limits_{\mathbf{p} \in \Omega} [u(p_1)], \max\limits_{\mathbf{p} \in \Omega} [u(p_2)], \ldots, \max\limits_{\mathbf{p} \in \Omega} [u(p_i)], \ldots, \max\limits_{\mathbf{p} \in \Omega} [u(p_M)]; \\ \max\limits_{\mathbf{p} \in \Omega} [|\rho(p_1, p_2)|], \max\limits_{\mathbf{p} \in \Omega} [|\rho(p_1, p_3)|], \ldots, \max\limits_{\mathbf{p} \in \Omega} [|\rho(p_i, p_j)|], \ldots, \max\limits_{\mathbf{p} \in \Omega} [|\rho(p_{M-1}, p_M)|] \end{array} \right\}, \tag{8}$$

where $u(p_i)$ and $|\rho(p_i, p_j)|$ ($i, j = 1, 2, \ldots, M$ and $i < j$) are estimated according to equations (3)–(7) by replacing the value of $\mathbf{p}$ ($\mathbf{p} = \hat{\mathbf{p}}$) in equation (4) with a given central value from a parameter domain Ω. It is noted that, here we take the absolute value of $\rho(p_i, p_j)$, $|\rho(p_i, p_j)|$, as the objective, since the parameter correlation coefficient $\rho(p_i, p_j)$ takes a value between -1 and 1. Considering the local property of the partial derivative in equation (4) for the definition of the Jacobian matrix $\mathbf{J}$, which typically varies with the given central value of $\mathbf{p}$ because of the nonlinearity of $\mathbf{f}(\mathbf{p}, \mathbf{a})$, we first scan the values of the objectives $u(p_i)$ and $|\rho(p_i, p_j)|$ by changing the central value of $\mathbf{p}$ from the domain Ω for their maxima. Then we scan the maxima of $u(p_i)$ and $|\rho(p_i, p_j)|$ obtained at different measurement configurations $\mathbf{a}$ from the domain Ξ for the minimum of each objective. The configuration that corresponds to the minima of all the objectives will be the final optimal configuration. The rationale behind the 'max-min' optimization given in equation (8) is that, if the achieved optimal measurement configuration is workable even in the worst case of $u(p_i)$ and $|\rho(p_i, p_j)|$ in a parameter domain, we have reasons to believe that it should be applicable for other cases of $u(p_i)$ and $|\rho(p_i, p_j)|$ in the same parameter domain. The

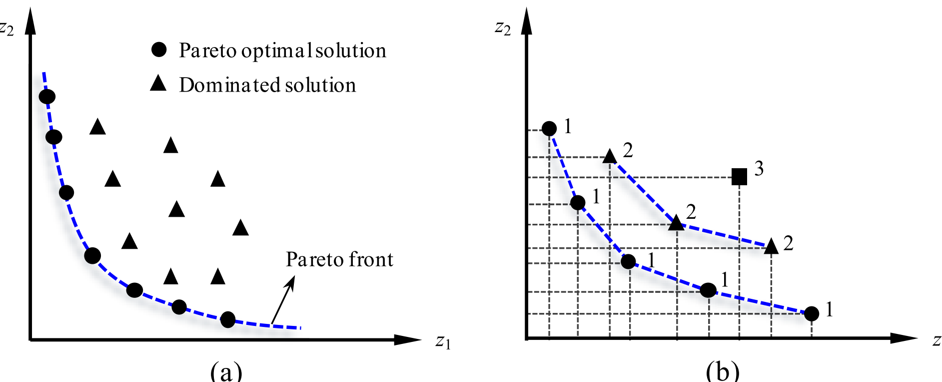

Figure 1. (a) Schematic of the Pareto optimal solution and the dominated solution for a multi-objective minimization problem, (b) schematic of the Pareto ranking based fitness assignment, where the number is the assigned rank of the corresponding solution.

former (internal) scan ensures the robustness of the optimization process for the changes of central values. The latter (external) scan ensures the optimization of the overall parameter precision and correlation. Note also that in equation (8) the external 'min' operator is not to find the minimum among all the objectives, but to try to find the minimum of each objective. Obviously, equation (8) describes a multi-objective minimization problem that consists of $M(M + 1)/2$ objectives for a M-dimensional vector $\mathbf{p}$.

For a single-objective optimization problem, we attempt to find the best solution, which is typically absolutely superior to all other alternatives. However, in a multi-objective scenario, such as the MCO problem described in equation (8), it is almost impossible for us to find a solution that could simultaneous optimize all the objectives. A possible approach to solve the multi-objective problem is to find a set of solutions, each of which could satisfy the objectives at an acceptable level without being dominated by other alternatives. Mathematically, for a K-objective minimization problem with K-objective being $\{z_1 = g_1(\mathbf{x}),\ z_2 = g_2(\mathbf{x}),\ \ldots,\ z_K = g_K(\mathbf{x})\}$, a solution $\mathbf{x}$ is said to *dominate* anther feasible solution $\mathbf{y}$ ($\mathbf{x} \succ \mathbf{y}$), if and only if, $g_i(\mathbf{x}) \leqslant g_i(\mathbf{y})$ for all indices $i \in \{1, 2, \ldots, K\}$ and $g_j(\mathbf{x}) < g_j(\mathbf{y})$ for at least on index $j \in \{1, 2, \ldots, K\}$ [21, 22]. A solution is said to be *Pareto optimal* if it not dominated by any other solution in the solution space. The concept of the Pareto optimal solution and the dominated solution is schematically shown in figure 1(a). Obviously, for a Pareto optimal solution, an improvement in any objective is impossible without sacrificing at least one of the other objectives. The set of all the feasible nondominated solutions in the solution space is referred to as the *Pareto optimal set*, and the set of the objective function values corresponding to a Pareto optimal set in the objective space is called the *Pareto front*. The aim of MOGA is to identify the solutions in the Pareto optimal set, namely the Pareto optimal solutions, for a multi-objective optimization problem. To do this, MOGA adopts a mechanism called Pareto ranking to assign a proper fitness value to each candidate solution [20]. As schematically shown in figure 1(b), the Pareto ranking procedure is illustrated as follows: assigning rank 1 to the nondominated solutions and removing them from contention, then finding the next set of nondominated solutions and assigning rank 2 to them, and repeating this process until the entire population is ranked.

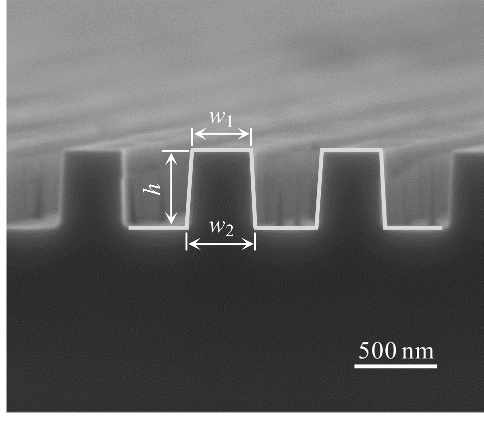

Figure 2. SEM cross-section image of the investigated Si grating.

For a minimization problem, a lower rank corresponds to a better solution. Many other Pareto ranking approaches have also been proposed over the past decades. One can consult [21, 22] for more details about MOGA. The specific implementation of MOGA in this paper was carried out by using the 'gamultiobj' function in MATLAB® (version R2014a, The MathWorks, Inc., Natick, MA, USA) running on a workstation equipped with double 2.0 GHz Intel Xeon CPUs.

3. Results and discussion

To demonstrate the feasibility of MOGA in optimizing measurement configurations for optical scatterometry, we take the spectroscopic MMS as an example. Compared with other scatterometry techniques, MMS can provide much more useful information about the sample by exploring the collected 4×4 Mueller matrix and has demonstrated a great potential in semiconductor manufacturing [6–9]. In spectroscopic MMS, it is the common practice to vary the wavelength λ in a spectral range with the incidence angle θ and azimuthal angle ϕ fixed at specified values [6–9], where ϕ is defined as the angle between the plane of incidence and the periodic direction of the nanostructure under test. The MCO problem for spectroscopic MMS is to find the optimal combination of incidence and azimuthal angles, namely the vector $\mathbf{a} = [\theta, \phi]^{\mathrm{T}}$, with which a higher measurement precision can be achieved. The nanostructure under test in this section is a

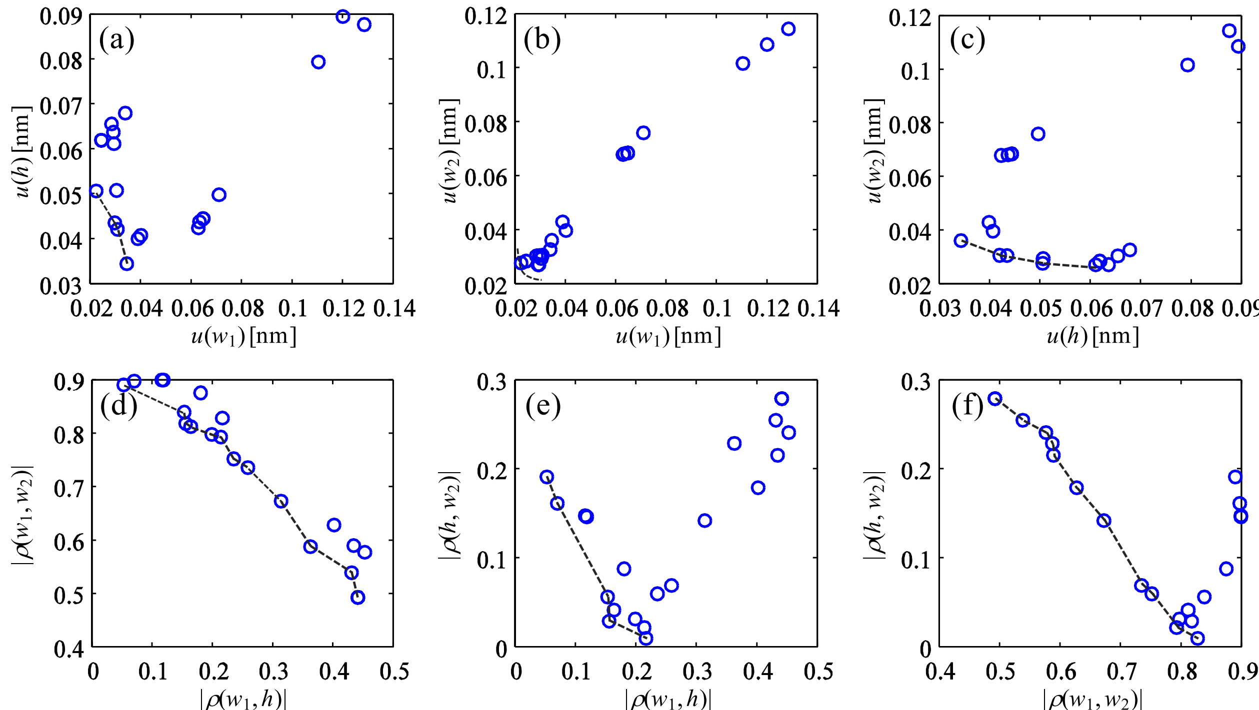

Figure 3. 2D cross-sections of the Pareto front achieved by MOGA for (a) $u(w_1)$ and $u(h)$, (b) $u(w_1)$ and $u(w_2)$, (c) $u(h)$ and $u(w_2)$, (d) $|\rho(w_1, h)|$ and $|\rho(w_1, w_2)|$, (e) $|\rho(w_1, h)|$ and $|\rho(h, w_2)|$, and (f) $|\rho(w_1, w_2)|$ and $|\rho(h, w_2)|$.

1D Si grating [18, 19], whose SEM cross-section image is shown in figure 2. As can be observed, the cross-section of the Si grating is characterized by a symmetrical trapezoidal model with three structural parameters under measurement, including the top CD w_1, the grating height h, and the bottom CD w_2, namely the vector $\mathbf{p} = [w_1, h, w_2]^T$. Dimensions of the above three structural parameters obtained from figure 2 are about $w_1 = 350$ nm, $h = 472$ nm, and $w_2 = 383$ nm, respectively. The nominal period of the Si grating is 800 nm, which is fixed in the solution of the inverse problem. The Si grating is chosen for this study mainly due to its typical geometrical profile in optical scatterometry as well as the limited computation resources in our lab. Although the investigated Si grating only has three structural parameters with relatively large dimensions under measurement, we should note from equation (8) that the proposed MCO method is not specific for the investigated sample but can be readily used for complex nanostructures with more and much smaller structural parameters under measurement.

In the optimization, the incidence angle θ was varied from 45° to 65°. Thanks to the geometrical symmetry of the investigated Si grating, we could restrict the range of azimuthal angles ϕ from 0° to 90°. The Mueller matrices of the Si grating were calculated by the rigorous coupled-wave analysis (RCWA) [24–26] for each $\mathbf{p}$ ($\mathbf{p} \in \Omega$) and $\mathbf{a}$ in the spectral range of 200–800 nm with an increment of 10 nm. In the calculation of the covariance matrix $\mathbf{C}$ by equation (3), the variances of the Mueller matrix elements were obtained from a noise model given in our previous work [27], which were functions of the specific $\mathbf{p}$ and $\mathbf{a}$. According to equation (8), we first scanned the values of the objectives $u(w_1)$, $u(h)$, $u(w_2)$, $|\rho(w_1, h)|$, $|\rho(w_1, w_2)|$ and $|\rho(h, w_2)|$ for their maxima within a parameter domain Ω of $w_1 \in [345, 355]$ nm, $h \in [465, 475]$ nm, and $w_2 \in [378, 388]$ nm. Here, the parameter domain Ω was determined according to the structural parameter values measured by SEM. In the general case, the ranges of structural parameters, typically depending on the process tolerances, could be specified at $\pm 10\%$ of their nominal dimensions as a rule of thumb [10]. The achieved maxima of the above objectives were then scanned within the prescribed ranges of incidence and azimuthal angles for the minimum of each objective. Here, the estimated uncertainties in the structural parameters all had a 95% confidence level ($\kappa = 2$). Since there were six objectives in the MCO problem of the investigated Si grating, the Pareto front achieved by MOGA would be six-dimensional. It was thus impossible to plot the whole Pareto front intuitively. Instead of showing the whole Pareto front, we exemplified several two-dimensional (2D) cross-sections of the achieved Pareto front for illustration.

Figure 3 presents some 2D cross-sections of the achieved Pareto front by MOGA, as indicated by the dotted lines. As can be noted from this figure, the presented 2D cross-sections of the achieved Pareto front show a large difference in comparison with the Pareto front schematically shown in figure 1(a). In addition, there are also many dominated solutions in figure 3. This is because of the insufficient convergence of MOGA, and the algorithm was terminated due to the excess of the prescribed number of generations in implementation (the default number of generations was 100 times of the number of variables to be optimized in the implementation of MOGA by MATLAB, and here was 200 for the MCO problem of the investigated Si grating). Strictly speaking, the achieved Pareto front shown in figure 3 is not the true but a quasi or approximate Pareto front for the MCO problem of the investigated

Table 1. The Pareto optimal set and Pareto front achieved by MOGA.

No.	Pareto optimal set θ (deg)	ϕ (deg)	Pareto front $u(w_1)$ (nm)	$u(h)$ (nm)	$u(w_2)$ (nm)	$\lvert\rho(w_1, h)\rvert$	$\lvert\rho(w_1, w_2)\rvert$	$\lvert\rho(h, w_2)\rvert$
1	55.3	22.9	0.0630	0.0424	0.0678	0.0531	0.8900	0.1907
2	62.5	82.5	0.0225	0.0506	0.0276	0.3630	0.5872	0.2284
3	52.5	83.2	0.0346	0.0344	0.0360	0.1993	0.7977	0.0311
4	50.9	79.8	0.0402	0.0407	0.0396	0.2169	0.8279	0.0094
5	64.1	77.5	0.0296	0.0611	0.0270	0.4348	0.5894	0.2150
6	47.0	67.8	0.1202	0.0894	0.1085	0.1562	0.8180	0.0289
7	56.5	32.0	0.0648	0.0444	0.0683	0.1163	0.8993	0.1472
8	64.9	82.5	0.0246	0.0619	0.0283	0.4416	0.4924	0.2788
9	56.3	31.1	0.0634	0.0437	0.0680	0.0706	0.8974	0.1611
10	52.1	29.5	0.0711	0.0497	0.0758	0.1809	0.8752	0.0875
11	64.9	80.2	0.0286	0.0655	0.0303	0.4314	0.5386	0.2544
12	47.0	67.0	0.1286	0.0876	0.1143	0.1536	0.8390	0.0558
13	64.9	77.5	0.0293	0.0636	0.0270	0.4531	0.5769	0.2406
14	55.8	80.4	0.0298	0.0435	0.0304	0.2592	0.7351	0.0688
15	47.4	70.8	0.1105	0.0793	0.1015	0.1643	0.8122	0.0413
16	51.9	83.1	0.0390	0.0399	0.0428	0.2142	0.7929	0.0217
17	55.3	79.5	0.0310	0.0420	0.0305	0.2358	0.7518	0.0595
18	60.2	77.0	0.0305	0.0507	0.0293	0.3142	0.6727	0.1415
19	64.1	77.6	0.0340	0.0679	0.0325	0.4025	0.6278	0.1785
20	56.5	32.1	0.0649	0.0445	0.0683	0.1188	0.8994	0.1460

Si grating. Although we could increase the number of generations for the implementation of MOGA to approach to the true Pareto front, it would take more time to run the algorithm. On the other hand, it is also unnecessary to achieve a globally optimal measurement configuration, suboptimal measurement configurations that could yield higher parameter precisions and lower parameter correlations are enough in most cases. As will be illustrated below, the achieved quasi Pareto front shown in figure 3 can indeed deliver some useful suboptimal measurement configurations. It is worth pointing out that the suboptimal measurement configurations would be more beneficial to a complex nanostructure with more structural parameters under measurement, since a globally optimal measurement configuration for the complex nanostructure will take even more time to run the algorithm due to the more optimization objectives.

For the convenience of illustration, table 1 presents the Pareto optimal set and the Pareto front achieved by MOGA for the MCO problem of the investigated Si grating. Note that the shown 20 groups of data in table 1 are not artificially selected from the many different possible measurement configurations but are all of the solutions output by MOGA after its termination. The number of the solutions achieved by MOGA depends on the complexity of the optimization problem as well as the termination conditions assigned to MOGA, such as the number of generations, the limit of the average relative change in the best fitness function value, and the prescribed running time. The default setting for the termination conditions was adopted in the implementation of MOGA by MATLAB. Due to the insufficient convergence of MOGA, there are still many dominated solutions in the achieved Pareto optimal set. First, we try to distinguish these dominated solutions. As can be observed from table 1, the structural parameters w_1 and w_2 have relatively stronger parameter correlation. As a rule of thumb, we remove the groups of data with $\lvert\rho(w_1, w_2)\rvert$ larger than 0.7, which were shown with a gray background in table 1, since strong parameter correlation is apt to make the solution of the inverse problem in optical scatterometry fall into local optima. The remaining groups of data (Nos. 2, 5, 8, 11, 13, 18 and 19) now all have small parameter correlation coefficients and also small parameter uncertainties, and the corresponding Pareto optimal solutions can be treated as the suboptimal measurement configurations. In addition, it can also be noted from table 1 that the measurement configuration of $\theta = 62.5°$ and $\phi = 82.5°$ in Group No. 2 shows a better performance than other remaining groups of data in terms of parameter uncertainties and also parameter correlation coefficients and can be chosen as the final optimal measurement configuration.

To verify the measurement configurations achieved by MOGA, Monte Carlo simulations were performed for the investigated Si grating. In the simulations, the incidence angle θ was varied from 45° to 65° with an increment of 2° and the azimuthal angle ϕ was varied from 0° to 90° also with an increment of 2°. The Mueller matrices of the Si grating were first calculated by RCWA for $\mathbf{p} = [350, 472, 383]^{\mathrm{T}}$ nm and each $\mathbf{a}$ in the spectral range of 200–800 nm with an increment of 10 nm. Random noise generated by the same noise model [27] as in the optimization was then added to the calculated Mueller matrices. Next, we extracted the structural parameters of the Si grating from the simulated Mueller matrix spectra by solving the inverse problem described by equation (2). Repeated simulations were performed at the same measurement configuration to estimate the uncertainties in the extracted parameters by

$$u_{\mathrm{MC}}(p_i) = \kappa \sqrt{\frac{1}{N_{\mathrm{MC}} - 1} \sum_{k=1}^{N_{\mathrm{MC}}} (\hat{p}_{i,k} - \bar{p}_i)^2}, \tag{9}$$

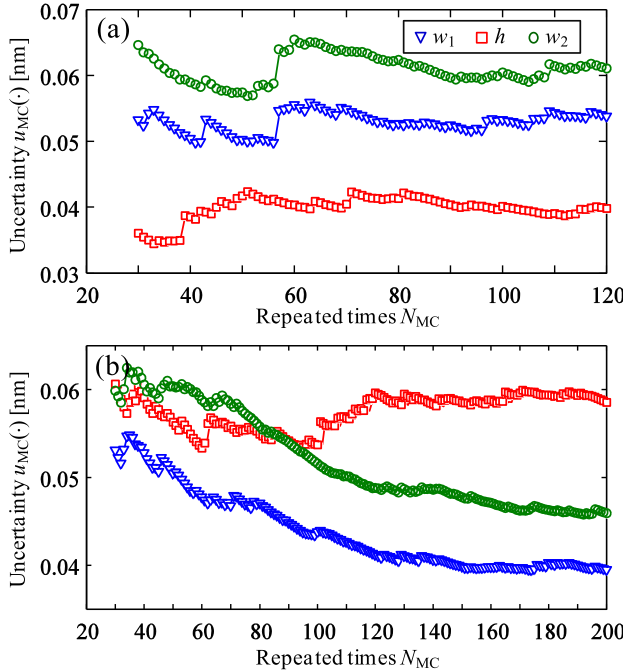

Figure 4. Variation of the estimated uncertainties in the extracted structural parameters with respect to the repeated times in Monte Carlo simulations at two randomly selected measurement configurations of (a) $\theta = 53°$, $\phi = 18°$ and (b) $\theta = 63°$, $\phi = 88°$.

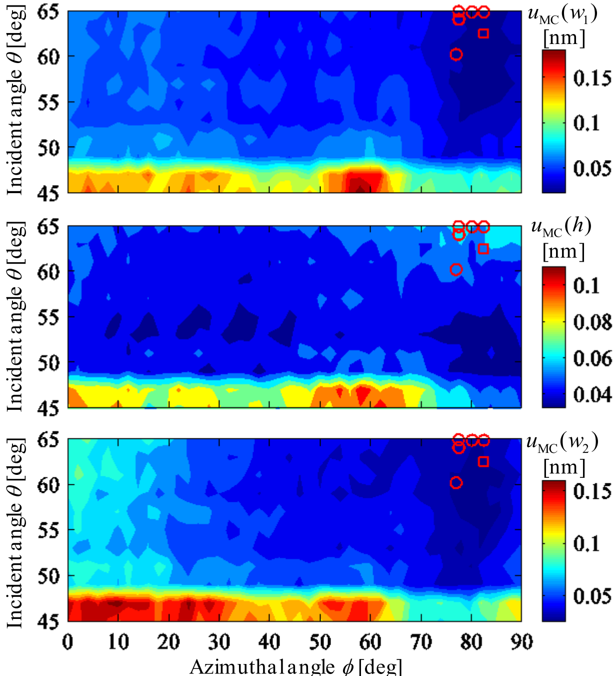

Figure 5. Monte Carlo simulated variation of the uncertainties in the extracted structural parameters at different measurement configurations, where the data points marked with red circles and squares correspond to the measurement configurations in table 1 shown with a white background. The data point marked with a red square corresponds to the optimal measurement configuration (Group No. 2 in table 1).

where N_{MC} represents the repeated times, $\hat{p}_{i,k}$ represents the ith structural parameter extracted in the kth repeated simulation, $\bar{p}_i = \sum_{k=1}^{N_{\mathrm{MC}}} \hat{p}_{i,k}/N_{\mathrm{MC}}$ represents the mean of the ith structural parameter after N_{MC} times of repeated simulations, and κ holds the same meaning as in equation (6) and also takes the same value as in the optimization. Note that here we use $u_{\mathrm{MC}}(\cdot)$ to denote the uncertainties in the extracted structural parameters estimated by Monte Carlo simulations to make a distinction with the approach given in equations (3)–(6). It was found that the estimated uncertainties in the extracted structural parameters $u_{\mathrm{MC}}(\cdot)$ tended to convergence when $N_{\mathrm{MC}} = 120$ for most of the measurement configurations. For some other measurement configurations where the estimated uncertainties were not fully convergent, much larger values of N_{MC} were adopted to make sure the convergence. As an example, figures 4(a) and (b) present the variation of the estimated uncertainties in the extracted structural parameters $u_{\mathrm{MC}}(\cdot)$ with respect to the repeated times N_{MC} at two randomly selected measurement configurations of $\theta = 53°$, $\phi = 18°$ and $\theta = 63°$, $\phi = 88°$, respectively.

Figure 5 presents the Monte Carlo simulated variation of uncertainties in the extracted structural parameters at different measurement configurations. As can be observed, the measurement configurations with larger incidence angles have much smaller parameter uncertainties. According to table 1, we know that all the measurement configurations achieved by MOGA locate in a region with $\theta \geqslant 47.0°$ and $\phi \geqslant 22.9°$, which is in good accordance with the corresponding regions shown in figure 5 that have smaller parameter uncertainties. As for the selected suboptimal measurement configurations from table 1, which were marked with red circles and squares in figure 5, we can observe that they are all located near the regions in figure 5 that have the smallest parameter uncertainties, especially for parameters w_1 and w_2. Although the measurement configuration in Group No. 2 was chosen as the final optimal measurement configuration, we should note from figure 5 that the measurement configurations in other groups (Nos. 5, 8, 11, 13, 18 and 19) actually could also be adopted in the experiment. The result shown in figure 5 therefore clearly demonstrates the feasibility of MOGA in handling the MCO problem in optical scatterometry.

Figure 6 present the fitting result of the measured and calculated best-fit Mueller matrix spectra of the Si grating at the selected optimal measurement configuration of $\theta = 62.5°$ and $\phi = 82.5°$ achieved by MOGA (Group No. 2 in table 1). Here, the measured Mueller matrix spectra were collected by using a ME-L ellipsometer (Wuhan Eoptics Technology Co., China), which employed a dual rotating-compensator system layout [8] and could provide full 4×4 Mueller matrices in the spectral range of 200~1000 nm with a high precision. As can be observed from figure 6, the calculated best-fit Mueller matrix spectra exhibit good agreement with the measured spectra, and the fitting result yields extracted structural parameters being $w_1 = 351.4 \pm 0.69$ nm, $h = 474.6 \pm 1.23$ nm, and $w_2 = 390.8 \pm 0.72$ nm, respectively. Here, the appended uncertainties (with a 95% confidence level) to the extracted structural parameters were estimated according to equation (6) with only the error sources of Type A being taken into account. The difference between the extracted structural parameters and the SEM-measured values given at the beginning of this section might be induced by the native oxide

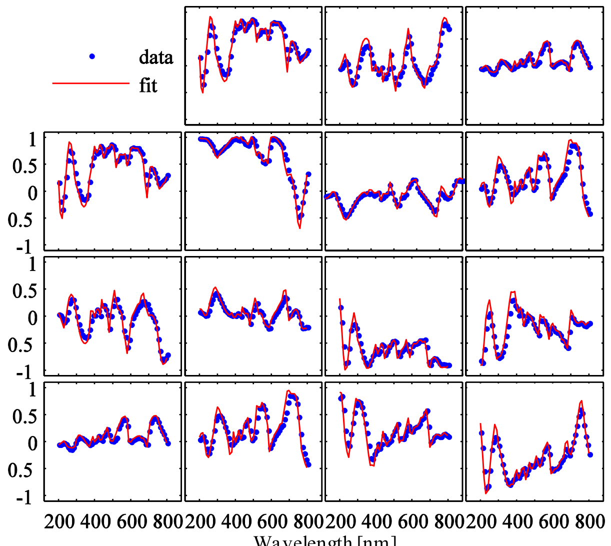

Figure 6. Fitting result of the measured and calculated best-fit Mueller matrix spectra of the Si grating at the optimal measurement configuration of $\theta = 62.5°$ and $\phi = 82.5°$.

layers coated on the Si grating, which typically have a thickness of 1–3 nm.

We can note the difference between the optimal measurement configuration achieved in this paper and that in our previous work in [18]. This is because of the different optimization goals between them. In this paper, we want to probe the optimal measurement configuration with which a higher measurement precision can be achieved, while in [18] the goal is to achieve a higher measurement accuracy. Nevertheless, it should be noted that the proposed MCO method based on MOGA in this paper can be readily extended by adding the objective function in [18] to equation (8) so as to probe a measurement configuration with which higher measurement precision and accuracy can be achieved. In addition, it took about twenty days for the employed workstation in our lab to run the MOGA to obtain the result shown in table 1. However, the Monte Carlo simulations took more than four months at the same workstation for the result shown in figure 5. The more calculation time of the Monte Carlo simulations is because that all the measurement configurations need to be traversed, and moreover, a large amount of repeated calculations are required to be performed at each measurement configuration to ensure the convergence of the estimated uncertainties in the extracted structural parameters. In comparison, in the proposed MCO method based on MOGA, the measurement configurations are efficiently evolved towards the optimum or suboptimal results based on some bio-inspired operators, such as mutation, crossover and selection. In addition, since MOGA is well-suited for parallel implementation, the parallel computation strategy can be adopted to further improve the computational efficiency of MOGA.

4. Conclusions

In this paper, the MCO problem in optical scatterometry was first formulated as a multi-objective optimization problem and the MOGA was introduced to probe the optimal measurement configuration with which a higher measurement precision could be achieved. The MMS was exemplified to show the implementation of MOGA for optimizing the combination of incidence and azimuthal angles in measuring an Si grating. The comparison with the Monte Carlo simulation results clearly demonstrated the feasibility of MOGA in handling the MCO problem. Although the investigated Si grating only had a simple geometrical profile with three structural parameters under measurement, it should be noted that the proposed MCO method based on MOGA could be readily used for nanostructures with more complex geometrical profiles and more structural parameters under measurement in advanced technology nodes of semiconductor manufacturing. Moreover, the proposed MCO method could also easily be extended by adding the deigned objectives to achieve the desired optimization goal. Therefore, the proposed MCO method is expected to provide a more general and practical means to solve the MCO problem in state-of-the-art optical scatterometry. It should also be pointed out that the main purpose of MCO in optical scatterometry is to provide a low-cost approach to achieve a higher measurement precision/accuracy for a given measuring instrument. However, if the measuring instrument inherently does not have enough sensitivity to a sample under test, it is difficult and even impossible to apply MCO to realize the measurement of the sample.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (Grant Nos. 51775217, 51475191, 51525502, and 51727809) and the National Science and Technology Major Project of China (Grant No. 2017ZX02101006-004).

ORCID iDs

Xiuguo Chen https://orcid.org/0000-0002-7067-5084
Shiyuan Liu https://orcid.org/0000-0002-0756-1439

References

[1] Minhas B K, Coulombe S A, Sohail S, Naqvi H and McNeil J R 1998 Ellipsometric scatterometry for metrology of sub-0.1 μm linewidth structures *Appl. Opt.* **37** 5112–5
[2] Huang H T, Kong W and Terry F L 2001 Normal-incidence spectroscopic ellipsometry for critical dimension monitoring *Appl. Phys. Lett.* **78** 3983–5
[3] Ko C H and Ku Y S 2006 Overlay measurement using angular scatterometer for the capability of integrated metrology *Opt. Express* **14** 6001–10
[4] Wurm M, Endres J, Probst J, Schoengen M, Diener A and Bodermann B 2017 Metrology of nanoscale grating structures by UV scatterometry *Opt. Express* **25** 2460–8
[5] Qin J, Silver R M, Barnes B M, Zhou H, Dixson R G and Henn M A 2016 Deep subwavelength nanometric image reconstruction using Fourier domain optical normalization *Light Sci. Appl.* **5** e16038
[6] Novikova T, De Martino A, Hatit S B and Drévillon B 2006 Application of Mueller polarimetry in conical diffraction

for critical dimension measurements in microelectronics *Appl. Opt.* **45** 3688–97
[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** 21336–43
[8] Liu S, Chen X and Zhang C 2015 Development of a broadband Mueller matrix ellipsometer as a powerful tool for nanostructure metrology *Thin Solid Films* **584** 176–85
[9] Dixit D *et al* 2016 Optical critical dimension metrology for directed self-assembly assisted contact hole shrink *J. Micro/Nanolith. MEMS MOEMS* **15** 014004
[10] Raymond C J 2001 Scatterometry for semiconductor metrology *Handbook of Silicon Semiconductor Metrology* ed A Diebold (New York: Dekker)
[11] Paz V F, Peterhänsel S, Frenner K and Osten W 2012 Solving the inverse grating problem by white light interference Fourier scatterometry *Light Sci. Appl.* **1** e36
[12] Logofătu P C 2002 Sensitivity analysis of grating parameter estimation *Appl. Opt.* **41** 7179–86
[13] Logofătu P C 2002 Phase-modulation scatterometry *Appl. Opt.* **41** 7187–92
[14] Littau M, Forman D, Bruce J, Raymond C J and Hummel S G 2006 Diffraction signature analysis methods for improving scatterometry precision *Proc. SPIE* **6152** 615236
[15] Gross H and Rathsfeld A 2008 Sensitivity analysis for indirect measurement in scatterometry and the reconstruction of periodic grating structures *Wave Random Complex* **18** 129–49
[16] Vagos P, Hu J, Liu Z and Rabello S 2009 Uncertainty and sensitivity analysis and its applications in OCD measurement *Proc. SPIE* **7272** 72721N
[17] Foldyna M, De Martino A, Garcia-Caurel E, Ossikovski R, Licitra C, Bertin F, Postava K and Drévillon B 2008 Critical dimension of biperiodic gratings determined by spectral ellipsometry and Mueller matrix polarimetry *Eur. Phys. J. Appl. Phys.* **42** 351–9
[18] Chen X, Liu S, Zhang C and Jiang H 2013 Measurement configuration optimization for accurate grating reconstruction by Mueller matrix polarimetry *J. Micro/Nanolith. MEMS MOEMS* **12** 033013
[19] Dong Z, Liu S, Chen X and Zhang C 2014 Determination of an optimal measurement configuration in optical scatterometry using global sensitivity analysis *Thin Solid Films* **562** 16–23
[20] Goldberg D E 1989 *Genetic Algorithms in Search, Optimization and Machine Learning* (Boston, MA: Addison-Wesley)
[21] Konak A, Coit D W and Smith A E 2006 Multi-objective optimization using genetic algorithms: a tutorial *Reliab. Eng. Syst. Safety* **91** 992–1007
[22] Deb K 2001 *Multi-Objective Optimization using Evolutionary Algorithms* (Chichester: Wiley)
[23] JCGM 100:2008 2008 Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (GUM)
[24] Moharam M G, Grann E B, Pommet D A and Gaylord T K 1995 Formulation of stable and efficient implementation of the rigorous coupled wave analysis of binary gratings *J. Opt. Soc. Am.* A **12** 1068–76
[25] Li L 1996 Use of Fourier series in the analysis of discontinuous periodic structures *J. Opt. Soc. Am.* A **13** 1870–6
[26] Li L 1996 Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings *J. Opt. Soc. Am.* A **13** 1024–35
[27] Chen X G, Liu S Y, Gu H G and Zhang C W 2014 Formulation of error propagation and estimation in grating reconstruction by a dual-rotating compensator Mueller matrix polarimeter *Thin Solid Films* **571** 653–9