IOP Publishing Measurement Science and Technology
Meas. Sci. Technol. **34** (2023) 125001 (10pp) https://doi.org/10.1088/1361-6501/acec8d

Condition-number-based measurement configuration optimization for nanostructure reconstruction by optical scatterometry

Tianjuan Yang[1], **Xiuguo Chen**[1,*], **Shuo Liu**[1], **Jiahao Zhang**[1] and **Shiyuan Liu**[1,2]

[1] State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China
[2] Optics Valley Laboratory, Wuhan 430074, People's Republic of China

E-mail: xiuguochen@hust.edu.cn

Received 25 March 2023, revised 3 July 2023
Accepted for publication 2 August 2023
Published 10 August 2023

Abstract

The quality of the measured signature is influenced not only by the instrument's precision but also by the selected measurement configuration. In optical scatterometry, the purpose of measurement configuration optimization (MCO) is to select an optimal or suboptimal combination of measurement conditions, such as the angles of incidence, azimuth, polarization and wavelength, to achieve higher measurement accuracy. This analysis not only requires an effective optimization strategy but is also time-consuming. In this work, we propose a general MCO method that incorporates error propagation theory and condition-number-based error estimation technique, by which the MCO problem can be formulated as an optimization problem for the condition number of the coefficient matrix in the linear estimation of parameter deviations. The method is demonstrated on a multi-wavelength Mueller matrix scatterometry measuring a Si grating. With the help of the neural-network-based surrogate model, the feasibility of the method is verified by making a comparison with Latin hypercube sampling. Fitting results of the measured and calculated Mueller matrix spectra obtained at the selected optimal measurement configuration show a good agreement. The proposed method is promising to provide an alternate solution to globally evaluate the MCO problem in optical scatterometry and other measurement scenarios.

Keywords: optical scatterometry, measurement configuration, optimization, condition number, surrogate model

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

1. Introduction

Benefiting from non-contact, non-destructive, fast and low-cost characteristics, optical scatterometry has been used as a powerful tool for characterizing nanostructures in the semiconductor industry [1–4]. This technique involves measuring the corresponding scattering signatures induced by the sample and extracting the profiles by solving the inverse problem [5–7]. In scatterometry, the nanostructure's signature is measured under a proper measurement configuration. Here, the measurement configuration refers to the combination of measurement conditions such as wavelength, incidence angle,

* Author to whom any correspondence should be addressed.

azimuth angle, and polarization direction. Measurement configuration optimization (MCO) in optical scatterometry is highly desirable due to its dual benefits. On the one hand, the measurement precision or accuracy is affected by the chosen measurement configuration [8], the optimal measurement configuration mode can mitigate the effect of measurement noise on the extracted parameters. On the other hand, certain optimal measurement configurations can lead to the decorrelation of structural parameters [9], thus improving the robustness of the extracted parameters in the inverse problem, particularly when the complexity of nanoscale devices significantly increases.

Many works have been conducted for scatterometry optimization. Logofătu proposed an optimization approach by defining sensitivity as the estimated precision of the fitting parameters. The method involves scanning all possible measurement configurations and selecting the measurement conditions that provide the highest sensitivity [10]. Dong *et al* conducted a global sensitivity analysis to identify the optimal measurement configuration by assessing the individual influence of input parameters on the output signatures based on the analysis of noise level and the main effect defined in global sensitivity analysis [11]. Meng *et al* proposed a neural-network (NN)-based and density-based sensitivity analysis to identify the optimal measurement configuration that exhibits a significant change in optical responses with small variations in dimensions [12]. Foldyna *et al* proposed choosing optimal incidence and azimuth angles based on the standard evaluation of the parameter variances and parameter correlations in Mueller matrix scatterometry (MMS) [13]. Chen *et al* treated the MCO problem as a multi-objective optimization problem, incorporating both parameter variances and parameter correlations as optimization objectives, and they utilized a multi-objective genetic algorithm to tackle this problem [8]. Dong *et al* proposed a method to reduce the uncertainty of the parameters in Mueller matrix ellipsometer by removing the measurement configurations with redundant information based on the dependence-analysis theory [14]. Chen *et al* also introduced an MCO method based on the theoretical analysis of error propagation and treated the norm of the configuration error propagating matrix as a metric to evaluate the impact of configuration errors on measurement accuracy [15]. Similarly, Zhu *et al* proposed an MCO method utilizing error propagation theory and singular value decomposition, and the optimal set is determined through the minimization of the Frobenius norm of a pseudo-Jacobian matrix [16]. In summary, MCO methods can be roughly classified into two categories based on their optimization objectives. The first category is the sensitivity analysis-based MCO method, which focuses on evaluating the individual influence of an input profile parameter on the output signatures over all measurement configurations. The second category is the error analysis-based MCO method, which applies error propagation theory to the inverse problem and primarily focuses on optimizing parameters' uncertainties, correlations, or coefficient matrices. These methods effectively optimize measurement configurations based on various assumptions and merits. However, many of them lead to multiple optimization objectives. For instance, if there are N parameters to be measured, N objectives on parameter uncertainties or sensitivities, and $N(N-1)/2$ objectives on parameter correlations can be raised. Consequently, a decision has to be made among the selected measurement schemes, which can be inefficient for implementation. The development of a universal MCO method that can be applied not only to optical scatterometry but also to other relevant measurement scenarios effectively and efficiently remains a significant subject of ongoing exploration and discussion.

In this work, we proposed a condition-number-based MCO method, serving as a complement to existing MCO methods. Specifically, we utilize the first-order Taylor expansion of the least squares (LSQ) function to establish the foundation of our approach. Additionally, by employing the condition-number-based error estimation technique to assess the solution accuracy of the linear equation system, we illustrate that the MCO problem can be formulated as a max–min problem of the condition number of the coefficient matrix in the linear estimation of parameter deviations. We should emphasize that the proposed MCO method has applicability not only in OCD metrology but also in other related model-based metrologies such as small-angle x-ray scattering [17]. To address the computational burden in the MCO problem, we also adopt a NN-based surrogate model for replacing the time-consuming forward model. The utilization of this surrogate model not only accelerates the calculations considerably but also facilitates the adoption of more refined configuration combinations in the optimization process, thereby enhancing the accuracy of identifying the optimal configurations.

The remainder of this article is organized as follows. Section 2 briefly introduces the inverse problem in optical scatterometry and then introduces the condition-number-based MCO method and the NN-based surrogate model adopted in this work. Section 3 introduces the experimental apparatus, the profile detail of the investigated sample and the architecture of the neural network. Section 4 gives the experimental results to illustrate the validity of the proposed MCO method. Section 5 gives some conclusions.

2. Method

2.1. Condition-number-based MCO problem

The χ^2 function is usually adopted in the inverse problem of optical scatterometry to evaluate the fitting error between the measured and theoretical signatures, which is defined as [5]

$$\chi^2 = \sum_{i=1}^{M} w_i [y_i - f_i(\mathbf{x}, \mathbf{a})]^2 = [\mathbf{y} - \mathbf{f}(\mathbf{x}, \mathbf{a})]^{\mathrm{T}} \mathbf{W} [\mathbf{y} - \mathbf{f}(\mathbf{x}, \mathbf{a})], \tag{1}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_M)^{\mathrm{T}}$ represents the measured signature, $\mathbf{f}(\mathbf{x}, \mathbf{a}) = [f_1(\mathbf{x}, \mathbf{a}), f_2(\mathbf{x}, \mathbf{a}), \ldots, f_M(\mathbf{x}, \mathbf{a})]^{\mathrm{T}}$ represents the corresponding theoretical signature calculated at the measurement configuration $\mathbf{a} = [a_1, a_2, \ldots, a_S]^{\mathrm{T}}$ for structural parameter $\mathbf{x} = [x_1, x_2, \ldots, x_N]^{\mathrm{T}}$ under measurement. $\mathbf{W}$ is an $M \times M$

diagonal matrix with diagonal elements w_i, w_i is the weighting vector and usually chosen to be $w_i = 1/\sigma^2(y_i)$ ($i = 1, 2, \ldots, M$), $\sigma^2(y_i)$ is the variance of the measured signature. The inverse problem can be formulated as the least-square regression problem in optical scatterometry,

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

where $\hat{\mathbf{x}}$ denotes the extracted parameters, Ω is the structural parameter domain. The solution to the inverse problem can be achieved by some iterative algorithms such as the gradient-based trust region algorithm and the Levenberg–Marquardt algorithm. Note that this approach requires knowledge of the measurement noise and a proper choice of weighting factors could influence the reconstruction results.

In general, there always be some discrepancy between the extracted parameters and the true parameters due to the presence of random errors and systematic errors [18]. The random errors arise from some random noises in measurement, such as instrument and the environment, while the systematic errors arise from the deterministic offsets in measurement, such as instrument, measurement method and the adopted forward model. Supposing $\mathbf{x}_0$ be the true parameter value, $\Delta\mathbf{x}$ be the deviation between the theoretical and the extracted parameters which can be expressed as

$$\Delta\mathbf{x} = \mathbf{x}_0 - \hat{\mathbf{x}}. \tag{3}$$

Suppose that the function $\mathbf{f}(\mathbf{x}, \mathbf{a})$ is sufficiently smooth and can be expanded in a first-order Taylor expansion

$$\mathbf{f}(\mathbf{x},\ \mathbf{a}) = \mathbf{f}(\hat{\mathbf{x}},\ \mathbf{a}) + \mathbf{J}_{\mathbf{x}} \cdot (\mathbf{x} - \hat{\mathbf{x}}), \tag{4}$$

where $\mathbf{J}_{\mathrm{x}}$ is the $N \times M$ Jacobian matrices with respect to $\mathbf{x}$, whose elements are given by

$$[\mathbf{J}_{\mathbf{x}}]_{ij} = \left. \frac{\partial f_i(\mathbf{x},\ \mathbf{a})}{\partial x_j} \right|_{\mathbf{x}=\hat{\mathbf{x}}}. \tag{5}$$

Substituting $\mathbf{x} = \mathbf{x}_0$ into equation (4) gives

$$\mathbf{f}(\mathbf{x}_0,\ \mathbf{a}) = \mathbf{f}(\hat{\mathbf{x}},\ \mathbf{a}) + \mathbf{J}_{\mathbf{x}} \cdot (\mathbf{x}_0 - \hat{\mathbf{x}}) = \mathbf{f}(\hat{\mathbf{x}},\ \mathbf{a}) + \mathbf{J}_{\mathbf{x}} \Delta\mathbf{x}. \tag{6}$$

Substituting equation (6) into equation (1) yields

$$\chi^2 = [\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}, \mathbf{a}) - \mathbf{J}_{\mathbf{x}} \Delta\mathbf{x}]^{\mathrm{T}} \mathbf{W} [\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}, \mathbf{a}) - \mathbf{J}_{\mathbf{x}} \Delta\mathbf{x}]. \tag{7}$$

Taking the derivatives of each side in equation (7) with respect to each element of vector $\mathbf{x}$ yields the following equation

$$\left(\mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W}^{\frac{1}{2}} \right) \mathbf{W}^{\frac{1}{2}} [\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}, \mathbf{a}) - \mathbf{J}_{\mathbf{x}} \Delta\mathbf{x}] = 0. \tag{8}$$

After rearranging the terms in equation (8), we obtain the following equation

$$\mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W} \mathbf{J}_{\mathbf{x}} \Delta\mathbf{x} = \mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W} \Delta\mathbf{y}, \tag{9}$$

where $\Delta\mathbf{y} = \mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}, \mathbf{a})$. Some detailed derivations of equation (9) can be found in appendix. The above equation gives an estimate of the deviation $\Delta\mathbf{x}$ between $\mathbf{x}_0$ and $\hat{\mathbf{x}}$. We call equation (9) the linear estimator of $\Delta\mathbf{x}$ since it is based on the first-order Taylor expansion of the function $\mathbf{f}(\mathbf{x}, \mathbf{a})$. The final structure parameter $\mathbf{x}_{\mathrm{c}}$ can be corrected as the sum of the extracted parameter $\hat{\mathbf{x}}$ and the deviation $\Delta\mathbf{x}$: $\mathbf{x}_{\mathrm{c}} = \hat{\mathbf{x}} + \Delta\mathbf{x}$.

The uncertainties in the measured signatures will be propagated into the extracted parameters in the process of solving equation (9). The coefficient matrix, denoted as $\mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W} \mathbf{J}_{\mathbf{x}}$, is closely linked to the measurement configuration and offers valuable insights into the uncertainties associated with the extracted parameters. In fact, the covariance matrix of $\mathbf{x}$ can be estimated as the inverse of the coefficient matrix, represented by $\mathbf{C} = \left(\mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W} \mathbf{J}_{\mathbf{x}} \right)^{-1}$. Several MCO methods, as mentioned in [8, 13–16], employ this coefficient matrix as a foundation of the optimization objective. In this study, we treat the MCO problem based on the condition-number-based error estimation technique. The coefficient matrix $\mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W} \mathbf{J}_{\mathbf{x}}$, depicted in linear equation (9), often reveals the ill-conditioning of the equation, reflecting the solution's susceptibility to slight changes in the input variables. A small condition number of the coefficient matrix $\mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W} \mathbf{J}_{\mathbf{x}}$ suggests that the perturbations in the signatures at the right-hand side of equation (9) have a minimal impact on the deviation $\Delta\mathbf{x}$. Therefore, it is reasonable to focus on optimizing $\mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W} \mathbf{J}_{\mathbf{x}}$ for a more robust solution. Based on the above statement, the MCO problem can be described as a 'max–min' optimization in this work:

$$\mathbf{a}_{\mathrm{opt}} = \arg\min_{\mathbf{a}\in\Theta} \left[\max_{\mathbf{x}\in\Omega} \left(\mathrm{cond} \left(\mathbf{J}_{\mathbf{x}}^{\mathrm{T}} \mathbf{W} \mathbf{J}_{\mathbf{x}} \right) \right) \right], \tag{10}$$

where $\mathbf{a}_{\mathrm{opt}}$ represents the optimal measurement configuration. The inner operator 'max' calculates the maximum condition numbers of the coefficient matrix over all parameters in the parameter domain Ω under a certain measurement configuration. The outer operator 'min' calculates the condition numbers of the coefficient matrix over all measurement schemes in the domain Θ for the global minimum. The measurement configurations with smaller condition numbers can be regarded as optimal schemes. The 'max' operator ensures the robustness of the optimization for a batch of samples with parameter values fluctuating around their nominal values. The 'min' operator ensures the global optimization of the measurement configuration. The optimal measurement configuration can be determined using single-objective optimization methods or traversal methods.

2.2. NN-based surrogate model

It is necessary and critical to reduce the computational burden in the MCO problem. The complexity of the forward model, the alternative measurement schemes to be optimized and the chosen optimization strategy account for the computational burden in the MCO problem. First, the operation of the forward model is usually time-consuming in optical scatterometry.

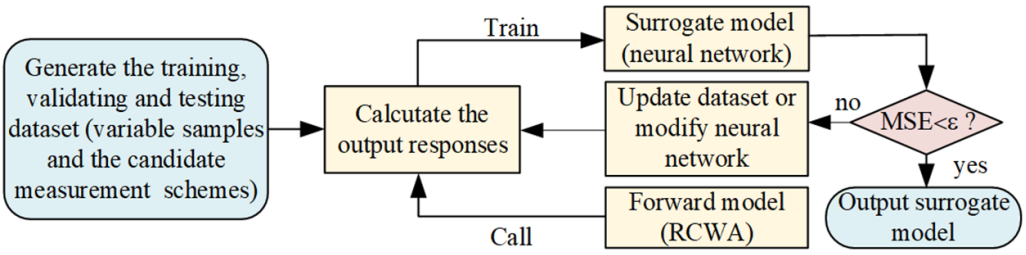

Figure 1. General procedure for generating the surrogate model.

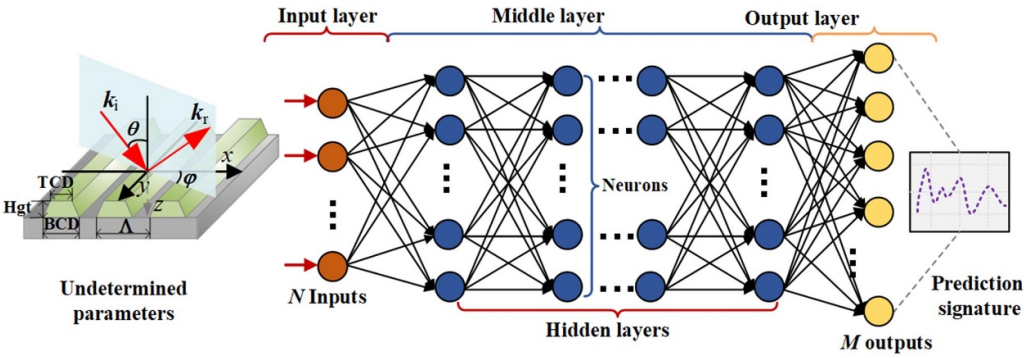

Figure 2. Schematic of the architecture of the neural-network-based surrogate model.

Rigorous coupled wave analysis (RCWA) is more effective in calculating periodic structures compared to other methods [19]. However, as the architectures become more complex, the computational efficiency significantly decreases due to the frequent invocation of RCWA with a greater number of slicing layers and higher truncated orders. Second, the MCO problem requires an adequate number of samples with parameter values fluctuating around their nominal values to derive an optimal measurement configuration that ensures robustness. Lastly, the computational time can be further burdened by the optimization strategy, particularly when time-consuming optimization algorithms and a substantial number of candidate schemes are being considered. The surrogate model offers an efficient solution for tackling such problems, particularly when establishing a complex or computationally expensive forward model is challenging. Some researchers have successfully utilized artificial neural networks in optical scatterometry [20–22]. The NN-based surrogate model has the potential for efficient and repeatable calculations, which is employed in this work.

Figure 1 depicts the general procedures for constructing such a surrogate model. The surrogate model is trained, validated or tested by the simulated dataset generated from the forward model, and the well-trained model replaces the forward model to accelerate the calculation process. A fully connected neural network is employed in this work to construct the mapping from parameters to signatures. The architecture of the NN-based surrogate model is presented in figure 2, The NN-based surrogate model establishes a mapping from N inputs to M outputs, comprising three essential components: an input layer for importing variable parameters, a middle layer for conducting non-linear computation, and an output layer for predicting feature signatures. The inputs contain both the structural parameters and the candidate measurement configurations, and the outputs are optical responses, e.g. reflectance, transmittance, and Mueller matrix. The middle layer consists of N_h hidden layers with N_n neurons in each layer. The back-propagation algorithm is employed in each iterative training process to constantly adjust the weights and biases of neurons. Additionally, the minimum square error is applied as the train loss function. The rectified linear unit activation function is selected for the hidden layers, while the output layer employs a linear function [22].

Please note that the computational performance of the NN-based surrogate model is contingent upon the configuration of its hidden layers. In practice, increasing the number of layers proves to be more advantageous than merely augmenting the number of neurons within a layer [12]. Hence, a series of preliminary numerical simulations are conducted to identify an optimal architecture for the hidden layers and the number of neurons. This step ensures that the surrogate model achieves excellent performance before replacing the forward model.

3. Experimental setup

The proposed method was tested on an industrial dual-rotating-compensator MMS (ME-L, Wuhan Eoptics Technology Co., China) for probing the optimal combination of incidence angle θ and azimuthal angle φ, namely $\mathbf{a} = [\theta, \varphi]^\mathrm{T}$. As schematically shown in figure 3, the system setting

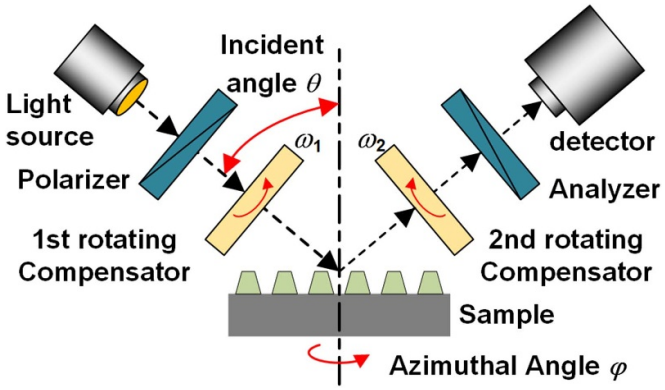

Figure 3. Principle of the dual-rotating compensator Mueller matrix scatterometry.

Table 1. The specifics of the training dataset.

Parameter	Nominal value	Lower bound	Upper bound	Training dataset size
TCD	37 nm	30 nm	40 nm	
Hgt	90 nm	85 nm	95 nm	
BCD	70 nm	60 nm	75 nm	500 00
θ	—	60°	70°	
φ	—	0°	90°	

of the spectroscopic MMS is PCr1SCr2A, where P and A stand for the polarizer and analyzer, and Cr1 and Cr2 stand for the 1st and 2nd rotating compensators. In MMS, it is common practice to fix the incidence angle θ and azimuth angle φ while varying the wavelength over a spectral range. In this work, the spectral range was varied from 200 to 800 nm with increments of 10 nm. The full 15 Mueller matrix elements can be obtained from this dual-rotating-compensator setting and are normalized by the first element. The range of the configuration domain Θ was defined as $\theta \in [60°, 70°]$ and $\varphi \in [0°, 90°]$, considering the geometric symmetry of the sample in this test. The calculation of $\mathbf{W}$ in equation (10) utilizes the variances of the measured signatures, which are provided by the MMS software.

A typical one-dimensional Si grating was tested in this work. The cross-section of the Si grating is described by a symmetric trapezoidal model, which includes the top critical dimension (TCD), grating height (Hgt), bottom critical dimension (BCD), and period Λ (as depicted in figure 2). The testing grating exhibits nominal values of TCD $=$ 37 nm, Hgt $=$ 90 nm, BCD $=$ 70 nm, and $\Lambda =$ 125 nm. In the MCO analysis, the grating pitch Λ remains fixed, while the parameters TCD, Hgt, and BCD vary within the parameter domain Ω, where TCD $\in$ [30, 40] nm, Hgt $\in$ [85, 95] nm, and BCD $\in$ [60, 75] nm. The parameter domain Ω typically depends on the variance in the fabrication process and can be empirically defined as a fluctuation of $\pm$10% around their nominal values in this paper [18]. The main reason for adopting Si grating in this work is due to its typical profile, higher refractive index contrast as well as long-term dimensional stability. It is noteworthy that while the investigated sample consists of only three parameters, the proposed MCO method is applicable to more complex nanostructures without a doubt. The simulated dataset was generated by RCWA and used for training, testing/validating. The specifics of the training dataset utilized in the experiments are presented in table 1, encompassing various structural parameters and the corresponding candidate measurement configurations. The training dataset is obtained by randomly selecting the parameters and measurement schemes within their ranges. The NN-based surrogate model was designed, trained, and implemented using PyTorch on a computer workstation (i7-9750H CPU @ 2.60 GHz).

4. Results and discussion

The NN architecture was identified by using an exhaustive grid search method on a sub-dataset of training data. Figure 4 presents the training losses of different NN architectures. The NN architecture with three or four hidden layers demonstrates superior approximation capability and exhibits lower loss compared to the architecture with only two hidden layers in this case. However, the NN architecture with four hidden layers does not exhibit significantly superior performance when compared to the architecture with three hidden layers. These findings suggest that a three-hidden-layer NN architecture is sufficient for this case, and further increasing the number of hidden layers is unlikely to result in substantial performance improvements. The loss of the NN architecture with three hidden layers decreases as the neuron size increases, reaching convergence when the neuron size exceeds 1200.

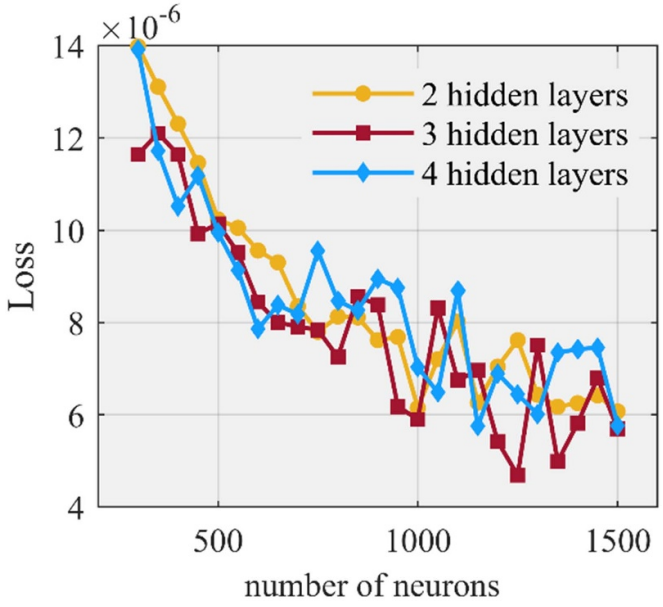

Figure 4. The training losses of different NN architectures.

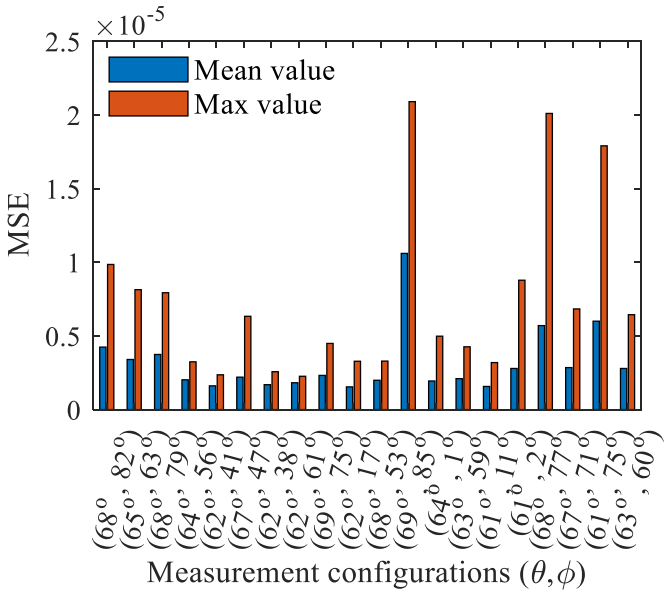

Figure 5. MSEs between the real and prediction signatures (Mueller matrix elements).

Consequently, a three-hidden-layer NN architecture with 1200 neurons per hidden layer is selected for this case.

To examine the performance of the surrogate model, we compare the mean square error (MSE) between the real signatures and predicted ones (Mueller matrix elements) under 20 groups of measurement configurations with 10 different choices of parameters randomly selected from the table 1 in each group of measurement configuration, MSE is given by

$$\mathrm{MSE} = \frac{1}{M}\sum_{m=1}^{M}(x_m^r - x_m^p)^2, \tag{11}$$

where x_m^r and x_m^p are the mth element of the real signatures and predicted signatures, respectively. M is the dimension of the signature. The real signatures were generated by RCWA. Figure 5 depicts the distributions of means and maximum MSEs of 10 groups of signatures under each configuration. Both the mean and maximum MSEs are on the order of 1.0×10^{-6}–1.0×10^{-5}, which are less than the noise level of Mueller matrix elements obtained from the MMS (typically $\sim 1.0 \times 10^{-3}$). The results indicate that the NN-based surrogate model can provide high performance to calculate the signatures in MCO problems.

Figure 6 presents the distributions of the maximum condition numbers of the coefficient matrix $\mathbf{J}_\mathbf{x}^\mathbf{T}\mathbf{W}\mathbf{J}_\mathbf{x}$ within the structural parameter domain Ω and the measurement configuration domain Θ. The incorporation of the NN-based surrogate model greatly facilitates the determination of the optimal measurement configurations. In particular, we explored the range of incidence angles θ from 60° to 70° with an increment of 1° increment, and azimuthal angles φ from 0° to 90° with an increment of 1°. Compared to previous works [8, 14–16], the implementation of this surrogate mode allows us to consider more refined angular step sizes during the optimization process, leading to improved accuracy in identifying optimal configurations. As can be seen from figure 6, the measurement configurations with smaller condition numbers (indicated by the purple region) are distributed within the ranges of $60° \leqslant \theta \leqslant 70°$ and $20° \leqslant \varphi \leqslant 80°$, as well as within the ranges of $64° \leqslant \theta \leqslant 70°$ and $80° \leqslant \varphi \leqslant 90°$, predicting that the accuracy of the extracted parameters under these measurement configurations is less susceptible to the errors in measured signatures. Based on this observation, we propose selecting optimal measurement configurations from these ranges to enhance the robustness of the extracted parameters.

To verify the validity of the proposed method indirectly, we examined the uncertainty distributions of three parameters

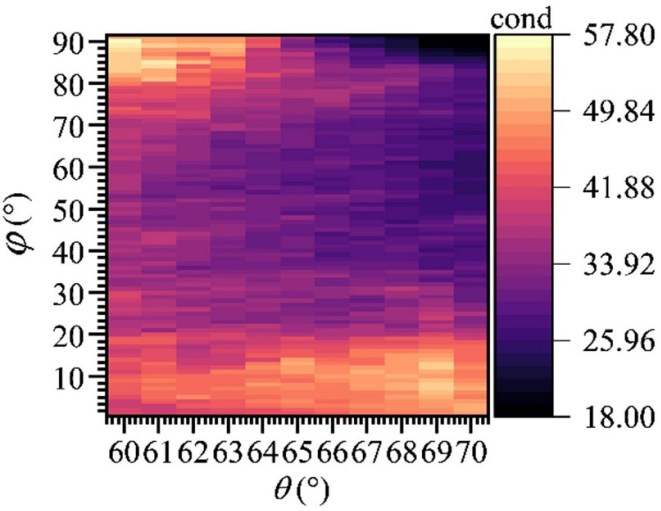

Figure 6. The distributions of the maximum condition numbers under different measurement configurations.

under different measurement configurations. In our research, we employed Latin hypercube sampling (LHS) as an efficient sampling method in optical scatterometry. LHS is particularly well-suited for scenarios with small sample sizes, enabling accurate estimation of geometry parameters, uncertainties, and repeatability [23]. To extract the structural parameters, we assumed Gaussian distributions for the Mueller matrices of the Si grating, with mean values derived from $\mathbf{x} = [37, 90, 70]^{\mathrm{T}}$ nm and variances obtained from MMS. The inverse problem was solved to determine the structural parameters, and the simulations were repeated under identical measurement configurations to assess the uncertainty associated with the extracted parameters

$$\mu(x_i) = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x_{i,n} - \bar{x}_i)^2}, \tag{12}$$

where N stands for the repeated number, $x_{i,n}$ represents the ith structural parameter extracted from the nth repeated simulation and $\bar{x}_i$ represents the mean of the ith structural parameter after N times of repeated simulations.

The repeated number N was tested under different measurement configurations to make sure of convergence when using LHS figure 7 illustrates the variations in estimated uncertainties for three extracted parameters as a function of the repeated number N at two randomly selected measurement configurations: $\theta = 60°$, $\varphi = 45°$, and $\theta = 65°$, $\varphi = 60°$. As expected, the estimated uncertainties of the three extracted parameters exhibit significant fluctuations when $N \leqslant 100$, but demonstrate relatively smaller variabilities when $N \geqslant 300$ for both measurement configurations. Notably, even within the range of N from 300 to 2500, the fluctuations remain relatively stable, indicating a convergence of the estimated uncertainties at $N = 300$.

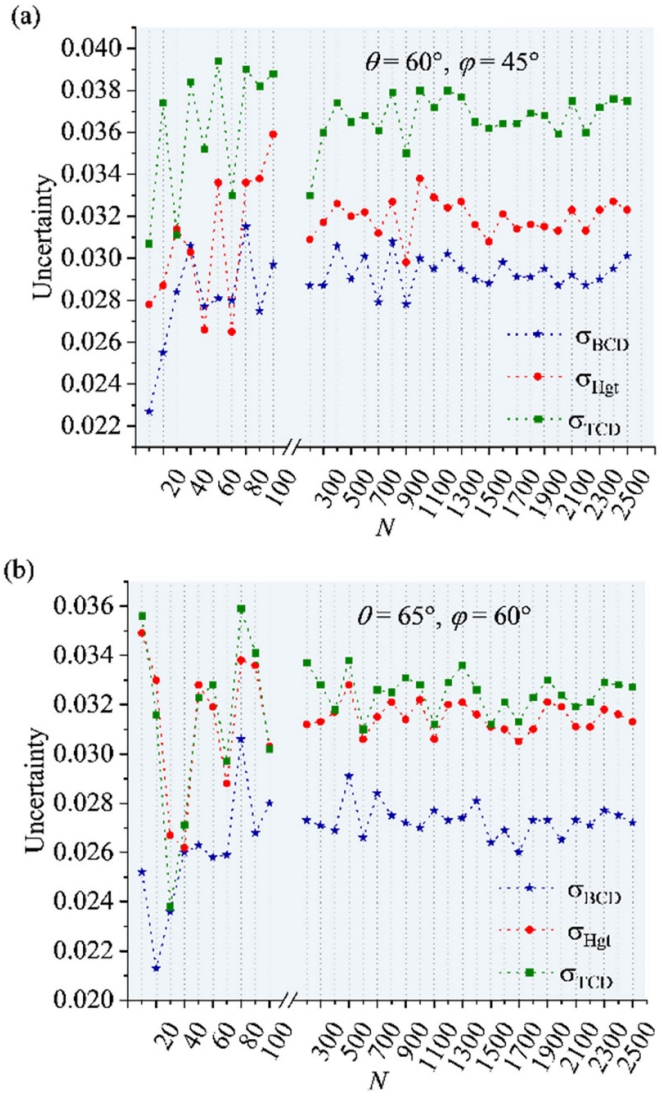

Figure 7. Variations in estimated uncertainties for three extracted parameters with respect to the repeated number N in LHS at two randomly selected measurement configurations: (a) $\theta = 60°$, $\varphi = 45°$, and (b) $\theta = 65°$, $\varphi = 60°$.

Figure 8 presents the uncertainty distributions of three structural parameters simulated by LHS at $N = 300$ under different measurement configurations. In the case of TCD (figure 8(a)), the measurement configurations with small uncertainties are located in the ranges of $60° \leqslant \theta \leqslant 70°$ and $20° \leqslant \varphi \leqslant 90°$. Similarly, for Hgt (figure 8(b)), the measurement configurations with small uncertainties are located in the ranges of $60° \leqslant \theta \leqslant 70°$ and $0° \leqslant \varphi \leqslant 80°$. In terms of BCD (figure 8(c)), the measurement configurations with small uncertainties are identified in two ranges: $60° \leqslant \theta \leqslant 64°$ and $20° \leqslant \varphi \leqslant 90°$, as well as $64° \leqslant \theta \leqslant 70°$ and $20° \leqslant \varphi \leqslant 85°$. The optimal configuration set, obtained by considering all three cases, corresponds to $60° \leqslant \theta \leqslant 64°$ and $20° \leqslant \varphi \leqslant 80°$.

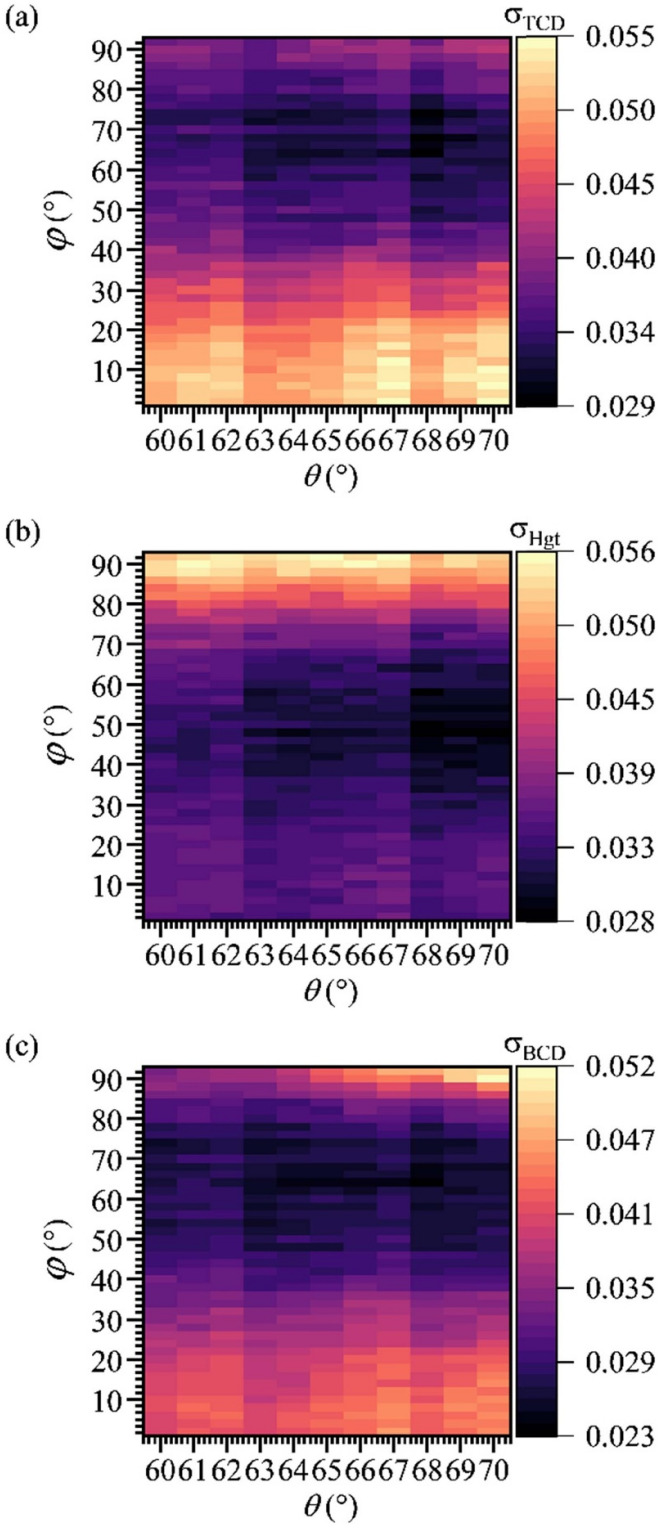

Figure 8. The distributions of the uncertainties of three extracted structural parameters at different measurement configurations simulated by LHS.

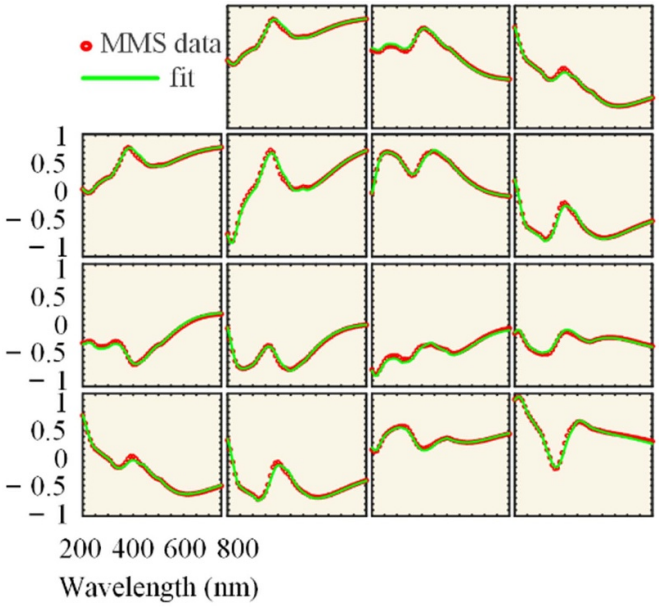

Figure 9. Fitting results of the measured and calculated Mueller matrix spectra of the Si grating at the selected measurement configuration of $\theta = 65^\circ$ and $\varphi = 60^\circ$.

This range aligns with the distribution of the optimal measurement configurations shown in figure 6, which exhibits smaller condition numbers. However, it should be noted that a different trend emerges in the ranges of $64^\circ \leqslant \theta \leqslant 70^\circ$ and $85^\circ \leqslant \varphi \leqslant 90^\circ$ (upper right corners of the graphs). Despite the small condition numbers in these ranges, the uncertainties for the BCT and Hgt parameters increase. The observed phenomenon is expected, as the uncertainties of the parameters are influenced by factors beyond the condition number of the coefficient matrix. As previously discussed, a smaller condition number of the coefficient matrix is associated with a smaller error estimate of the measured parameter and a more robust solution. However, the final extracted results may be affected by other various factors, including model errors and instrument errors. The presented results demonstrate that the proposed method effectively achieves a balanced trade-off among the three parameters, enabling a comprehensive evaluation. This differs from the commonly used optimization strategies that optimize multiple objectives individually, such as parameters' uncertainties, sensitivities and correlation coefficients, which are inefficient and a reasonable compromise has to be made ultimately among the selected measurement schemes. In contrast, our proposed one-time optimal strategy offers the ability to globally evaluate the MCO problem.

Figure 9 presents the fitting results of the measured Mueller matrix spectra and the corresponding calculated spectra for the Si grating at the measurement configuration ($\theta = 65^\circ$, $\varphi = 60^\circ$) selected from the optimal regions depicted in figure 6. This specific configuration was selected due to its ease of implementation on the measurement system (MMS). The fitting results show a good match, and the extracted parameters are TCD $= 37.3 \pm 0.56$ nm, Hgt $= 91.2 \pm 0.55$ nm, and BCD $= 72.3 \pm 0.72$ nm, which deviate slightly from the nominal values. Here, the uncertainties attached to each structural parameter were estimated at a confidence level of 95%. The chosen measurement configuration represents a near-optimal scheme, as indicated by the observations in figure 6. It should be noted that there are alternative optimal measurement

configurations available for experimentation. It is worth pointing out that sometimes it is not always feasible to achieve the globally optimal measurement configuration due to constraints imposed by specific experimental conditions or limited experimental sensitivity. In many cases, sub-optimal or near-optimal measurement configurations are adequate to meet measurement requirements while still offering improved parameter accuracy and practical feasibility. The optimal measurement configurations can be achieved by a simple traversal method under limited configuration conditions.

We should note that there is no one-fit-all optimal strategy that applies to all MCO problems. The application of the condition-number-based optimal method in this work suggests that it provides an alternative approach to the existing MCO methods, offering comprehensive and effective solutions. However, this does not imply its superiority over other methods. Each MCO approach possesses its own strengths based on various perspectives and optimization metrics. The selection of the final optimization strategy for an MCO problem depends on the specific application scenario, optimization index, and optimization method. It is recommended to conduct a comprehensive evaluation of the results obtained from various MCO methods and subsequently select the most appropriate configuration scheme. Furthermore, even within the proposed method, different combinations of measurement configurations, such as λ, θ, and φ, can yield different MCO outcomes. The preferences and decisions of the implementer also matter. In addition, more advanced intelligent optimization techniques, such as genetic algorithms, differential evolution, and simulated annealing, can be employed to further enhance performance.

5. Conclusions

The development of an effective and versatile MCO method is a meaningful research, deserving further exploration and innovation. This paper introduces a condition-number-based MCO method, offering a general and low-cost means of solving MCO problems in optical scatterometry. The full paper is summarized as follows:

(1) We utilized the first-order Taylor expansion of the LSQ function in the inverse problem considering the error propagation in the measurement. Consequently, we treated the MCO problem as an optimization problem for the condition number of the coefficient matrix in the linear estimation of the deviation of the parameters.
(2) Considering the computing burden in the MCO problem, we also adopted an NN-based surrogate model for replacing the time-consuming forward model.
(3) The method was validated on a commercial MMS, measuring a typical Si grating and probing the optimal combination of incidence and azimuth angles. The feasibility of the approach was verified by comparing it with the LHS results.

The proposed method exhibits several advantages, as substantiated by the experimental results. These advantages can be observed in the following aspects:

(1) The proposed method holds promise as a general method for solving MCO problems, not only in optical scatterometry but also in other related measurement instruments, providing a comprehensive evaluation capability.
(2) The incorporation of the NN-based surrogate model not only accelerates the calculations considerably but also facilitates employing more refined configuration combinations during the optimization process. As a result, it enhances the accuracy in identifying optimal configurations.

It should be noted that the condition-number-based MCO approach presented in this paper may not represent the most optimal choice of measurement configurations. However, it is expected to yield reconstructions that are more robust and accurate compared to randomly chosen configurations. Meanwhile, the neural network adopted in this work serves as a preliminary surrogate model for optimization. To achieve more accurate approximation, we recommend exploring alternative designs for the neural network architecture, such as incorporating transfer learning or utilizing other machine learning techniques like mainstream Bayesian methods in future research. Ultimately, this study is anticipated to provide a fresh perspective on nanostructure reconstruction in IC manufacturing, encompassing not only optical scatterometry but also other measurement domains.

Data availability statements

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This work was funded by the National Natural Science Foundation of China (Grant Nos. 52022034, 62175075, 52130504), National Key Research and Development Plan of China (Grant No. 2022YFF0709104), and Key Research and Development Plan of Hubei Province (Grant No. 2020BAA008).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Derivation of equation (9)

For the sake of brevity, a matrix $\mathbf{A}(\mathbf{x}) = \mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}, \mathbf{a}) - \mathbf{J}_{\mathbf{x}}\Delta\mathbf{x}$ is introduced in the following derivation. In this case, equation (7) can be rewritten as

$$\chi^2 = [\mathbf{A}(\mathbf{x})]^{\mathrm{T}}\mathbf{W}[\mathbf{A}(\mathbf{x})] = \left[\tilde{\mathbf{A}}(\mathbf{x})\right]^{\mathrm{T}}\left[\tilde{\mathbf{A}}(\mathbf{x})\right], \tag{A.1}$$

where $\tilde{\mathbf{A}}(\mathbf{x}) = \mathbf{W}^{\frac{1}{2}}\mathbf{A}(\mathbf{x})$. Taking the derivatives of each side in equation (A.1) with respect to each element of vector $\mathbf{x}$ yields the following equation

$$2\frac{\partial\left[\tilde{\mathbf{A}}(\mathbf{x})\right]^{\mathrm{T}}}{\partial\mathbf{x}}\tilde{\mathbf{A}}(\mathbf{x}) = 0, \tag{A.2}$$

$$\frac{\partial\left[\tilde{\mathbf{A}}(\mathbf{x})\right]^{\mathrm{T}}}{\partial\mathbf{x}} = -\left(\mathbf{W}^{\frac{1}{2}}\mathbf{J}_{\mathbf{X}}\right)^{T}. \tag{A.3}$$

Substituting equation (A.3) into equation (A.2) gives

$$-2\left(\mathbf{W}^{\frac{1}{2}}\mathbf{J}_{\mathbf{X}}\right)^{\mathrm{T}}\tilde{\mathbf{A}}(\mathbf{x}) = 0, \tag{A.4}$$

i.e.

$$\left(\mathbf{W}^{\frac{1}{2}}\mathbf{J}_{\mathbf{X}}\right)^{\mathrm{T}}\mathbf{W}^{\frac{1}{2}}[\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}, \mathbf{a}) - \mathbf{J}_{\mathbf{X}}\Delta\mathbf{x}] = 0. \tag{A.5}$$

After rearranging the terms in (A.5), we obtain the following equation

$$\mathbf{J}_{\mathbf{x}}^{\mathbf{T}}\mathbf{W}[\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}, \mathbf{a}) - \mathbf{J}_{\mathbf{X}}\Delta\mathbf{x}] = 0. \tag{A.6}$$

Supposing $\Delta\mathbf{y} = \mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}, \mathbf{a})$ and substituting it into equation (A.6) give

$$\mathbf{J}_{\mathbf{x}}^{\mathbf{T}}\mathbf{W}\mathbf{J}_{\mathbf{x}}\Delta\mathbf{x} = \mathbf{J}_{\mathbf{x}}^{\mathbf{T}}\mathbf{W}\Delta\mathbf{y}. \tag{A.7}$$

ORCID iDs

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

References

[1] Chen X, Gu H, Liu J, Chen C and Liu S 2022 Advanced Mueller matrix ellipsometry: instrumentation and emerging applications *Sci. China Technol. Sci.* **65** 2007–30
[2] Shimizu Y, Chen L C, Kim D W, Chen X, Li X and Matsukuma H 2021 An insight into optical metrology in manufacturing *Meas. Sci. Technol.* **32** 042003
[3] Guo C, Shi Y, Wu H, Li W, Zhang C, Jiang H and Liu S 2023 Integrated optical critical dimension metrology with Mueller matrix ellipsometry *Thin Solid Films* **768** 139695
[4] Sabbagh R, Stothert A, Sreenivasan S V and Djurdjanovic D 2023 Optical metrology of critical dimensions in large-area nanostructure arrays with complex patterns *J. Manuf. Sci. Eng.* **145** 061010
[5] Henn M-A, Gross H, Scholze F, Wurm M, Elster C and Bär M 2012 A maximum likelihood approach to the inverse problem of scatterometry *Opt. Express* **20** 12771–86
[6] Zhu J, Liu S, Chen X, Zhang C and Jiang H 2014 Robust solution to the inverse problem in optical scatterometry *Opt. Express* **22** 22031
[7] Guo C, Shi Y, Wu H, Xiang Y, Li W, Zhang C and Liu S 2023 A combination of library search and Levenberg-Marquardt algorithm in optical scatterometry *Thin Solid Films* **767** 139670
[8] Chen X, Gu H, Jiang H, Zhang C and Liu S 2018 Probing optimal measurement configuration for optical scatterometry by the multi-objective genetic algorithm *Meas. Sci. Technol.* **29** 045014
[9] 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
[10] Logofatu P C 2002 Sensitivity analysis of grating parameter estimation *Appl. Opt.* **41** 7179–86
[11] 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
[12] Meng K, Jiang B and Youcef-Toumi K 2021 Neural network assisted multi-parameter global sensitivity analysis for nanostructure scatterometry *Appl. Surf. Sci.* **570** 151219
[13] 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
[14] Dong Z, Chen X, Wang X, Shi Y, Jiang H and Liu S 2019 Dependence-analysis-based data-refinement in optical scatterometry for fast nanostructure reconstruction *Appl. Sci.* **9** 4091
[15] 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 MOMES* **12** 033013
[16] Zhu J, Shi Y, Goddard L and Liu S 2016 Application of measurement configuration optimization for accurate metrology of sub-wavelength dimensions in multilayer gratings using optical scatterometry *Appl. Opt.* **55** 6844–9
[17] Franke D, Jeffries C and Svergun D 2015 Correlation map, a goodness-of-fit test for one-dimensional x-ray scattering spectra *Nat. Methods* **12** 419–22
[18] Chen X, Liu S, Gu H and Zhang C 2014 Formulation of error propagation and estimation in grating reconstruction by a dual-rotating compensator Mueller matrix polarimeter *Thin Solid Films* **571** 653–9
[19] Pham H, Alcaire T, Soulan S, Le Cunff D and Tortai J 2022 Efficient rigorous coupled-wave analysis simulation of Mueller matrix ellipsometry of three-dimensional multilayer nanostructures *Nanomaterials* **12** 3951
[20] Liu J, Zhang D, Yu D, Ren M and Xu J 2021 Machine learning powered ellipsometry *Light Sci. Appl.* **10** 55
[21] Arunachalam A, Berriel S, Banerjee P and Basu K 2022 Machine learning-enhanced efficient spectroscopic ellipsometry modeling (arXiv:2201.04933 [cond-mat.mtrl-sci])
[22] Liu S, Chen X, Yang T, Guo C, Zhang J, Ma J, Chen C, Wang C, Zhang C and Liu S 2022 Machine learning aided solution to the inverse problem in optical scatterometry *Measurement* **191** 110811
[23] Heidenreich S, Gross H, Bär M and Wright L 2017 Uncertainty propagation in computationally expensive models: a survey of sampling methods and application to scatterometry *Measurement* **97** 79–87