



Machine learning aided solution to the inverse problem in optical scatterometry

Shuo Liu, Xiuguo Chen^{*}, Tianjuan Yang, Chunfu Guo, Jiahao Zhang, Jianyuan Ma, Chao Chen, Cai Wang, Chuanwei Zhang, Shiyuan Liu

State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Keywords:

Inverse problem
Machine learning
Optical scatterometry
Critical dimension
Mueller matrix ellipsometry

ABSTRACT

Optical scatterometry is the workhorse technique for in-line manufacturing process control in the semiconductor industry. However, as manufacturing processes develop, traditional methods for solving the inverse problem in optical scatterometry are struggling to continue improving productivity. To address this problem, machine learning can be a promising method, but it is a challenge to ensure robustness. In this paper, we propose a machine learning method to reconstruct the profile of nanostructures. The proposed method consists of three parts: compressing signature using a dimensionality reduction approach based on the principle component analysis, constructing a surrogate electromagnetic solver (SurEM) based on an artificial neural network mapping from parameters to signatures, and iteratively comparing the SurEM-predicted signatures with measured one to finally determine the results. Experiments have demonstrated that the proposed method can achieve fast and accurate measurement. This method is thus promising as an efficient in-line measurement method for nano- or micro-scale manufacturing.

1. Introduction

In high volume manufacturing of integrated circuits, the deviations in profile or position may affect the device performance, even lead to failure. Therefore, nanoscale metrology is a crucial part of manufacturing, by inspecting the patterns on the wafer before the wafer undergoes the next irreversible process. As an important technique for nanoscale metrology, ellipsometry is widely used in characterization of thicknesses of thin films and optical constants of both layered and bulk materials [1]. Since the year of around 2000, ellipsometry was introduced to measure critical dimensions of grating structures [2], which is also known as optical scatterometry in the community of semiconductor metrology [3]. Up to now, optical scatterometry has become an indispensable tool for the in-line monitoring of nanostructures profile in the semiconductor manufacturing [4–6].

Optical scatterometry is an indirect metrology technique. Through the analysis of the light intensity information measured by a detector, the optical signature of the sample can be obtained. And the parameters are extracted by comparing the theoretical optical signature with the measured one, which is known as an inverse scattering problem [7–9]. There are several methods to solve this problem, such as library search

[10] and non-linear regression [11]. The library search methods have faster solution speed relying on a pre-generated data library, but typically a large library is required to ensure the accuracy. The nonlinear regression methods are the most accurate, but usually time-consuming due to the demand of iterative computing the electromagnetic (EM) solvers, which must be established to calculate the theoretical signatures, such as the rigorous coupled-wave analysis (RCWA) [12], or the finite-difference time-domain method [13]. However, with the development of semiconductor industry, the dimensions of device continue to shrink and the architectures are becoming more complex [14,15]. The EM solver is becoming more difficult to establish and more time-consuming for calculation, due to the challenges arising from the development trend, such as, quantum confinement or high aspect ratio architectures.

Machine learning, as a data-driven algorithm, has offered a statistical perspective for constructing complex non-linear mapping relationships [16]. It is therefore effective in solving problems where physical models are computationally intensive or difficult to establish [17,18]. Machine learning has been successfully applied in many fields [19–22]. Recently, some researchers have been working for exploring the application of machine learning in optical scatterometry [23–28]. The main idea of

* Corresponding author.

E-mail address: xiuguo@hust.edu.cn (X. Chen).

those works is to map optical signatures to structural parameters using the artificial neural network (ANN). In order to meet the data needs of machine learning, the researchers usually utilize the EM solver to generate simulation data for training. However, due to an end-to-end computational process, the noise in data will be transmitted to the results. Robert et al. implemented ANN to characterize the wavelength-scale structure, and they pointed out that the simulation data added with Gaussian noise could improve the results and using principle component analysis (PCA) to process training data could decrease the training time [23]. Zhang et al. took the prediction structural parameters obtained by ANN as initial values, and then performed a non-linear algorithm to modify the bias in the results due to the presence of noise in the data, which speeded up the measurement process [24]. Godi et al. developed four ANN models for different basic profiles, and achieved reconstruction of more complex profile through a weighting method, but lacked experimental verification [26]. Thus, the simulation data added with noise or not, used for training would result in poor generalization performance of the trained ANN, because the noise varies with the instruments and measurements [29].

In this paper, we propose a machine-learning-enabled reconstruction method (MLER) for solving to the inverse problem in optical scatterometry. In the proposed MLER, we first present a signature dimensionality reduction approach using PCA to extract features from measured signatures. Then, we establish an ANN-based surrogate electromagnetic solver (SurEM) to learn the mapping from parameters about the samples to feature signatures. The results can be finally achieved by finding the best match between the measured and SurEM-predicted feature signatures using a nonlinear regression algorithm. The MLER method greatly reduce the parameter extraction time by utilizing a highly computationally efficient ANN and the proposed signature dimensionality reduction approach. Moreover, the MLER could effectively address the poor generalization performance of the trained ANN caused by measurement noise. To make a comparison, we also establish a direct extraction deep neural network (DEDNN) to learn the mapping from signatures to parameters (See the [Supplementary Material](#)). The experimental results have demonstrated that the proposed method could achieve accurate reconstruction comparable to the library search method with a much faster speed and is also more robust to measurement noise compared to DEDNN.

2. Method

2.1. Overview of MLER

Fig. 1 illustrates the summary block diagram of the proposed MLER method to reconstruct parameters of nanostructure from the optical signatures. Prior to the measurement, the geometric structure model of the testing sample is constructed. And we denote the undetermined parameters about the testing sample and measurement configuration as an m -dimensional vector $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$. Then, a SurEM based on ANN is developed and trained to calculate the signature for a given set of parameters. The dataset for training is generated by simulation or measurements, containing various sets of parameters and corresponding optical signatures, each of which is given by an n -dimensional vector $\mathbf{s} = [s_1, s_2, \dots, s_n]^T$. Moreover, PCA is employed on the dataset to find a transformation, which can compress the signatures into a lower feature space without losing too much information. The feature signature is noted as $\mathbf{z} = [z_1, z_2, \dots, z_R]^T$, where R is the reduced dimension of the signature ($R \ll n$). With the availability of the SurEM and the specification of the PCA transformation, the procedure of the MLER is described in details as follows.

Step 1: Data processing for the measured signature. The measured signature \mathbf{s}^{mea} is transformed into the measured feature signature \mathbf{z}^{mea} using PCA.

Step 2: Initialize the iterative solving configuration. The undetermined parameters are initially set to \mathbf{x}_0 , and the solving termination condition ε is set for nonlinear regression strategy.

Step 3: Calculate the feature signature $\tilde{\mathbf{z}}$ for the given parameters using the pre-trained SurEM.

Step 4: Calculate the mean square error (MSE) between the feature signature $\tilde{\mathbf{z}}$ obtained by SurEM and the measured feature signature \mathbf{z}^{mea} , given by

$$\text{MSE} = \frac{1}{R} \sum_{r=1}^R \left(\tilde{z}_r - z_r^{mea} \right)^2 \quad (1)$$

where \tilde{z}_r and z_r^{mea} are the r -th element of the calculated feature signature and measured feature signature, respectively.

Step 5: Determine whether the termination condition $\text{MSE} < \varepsilon$ is

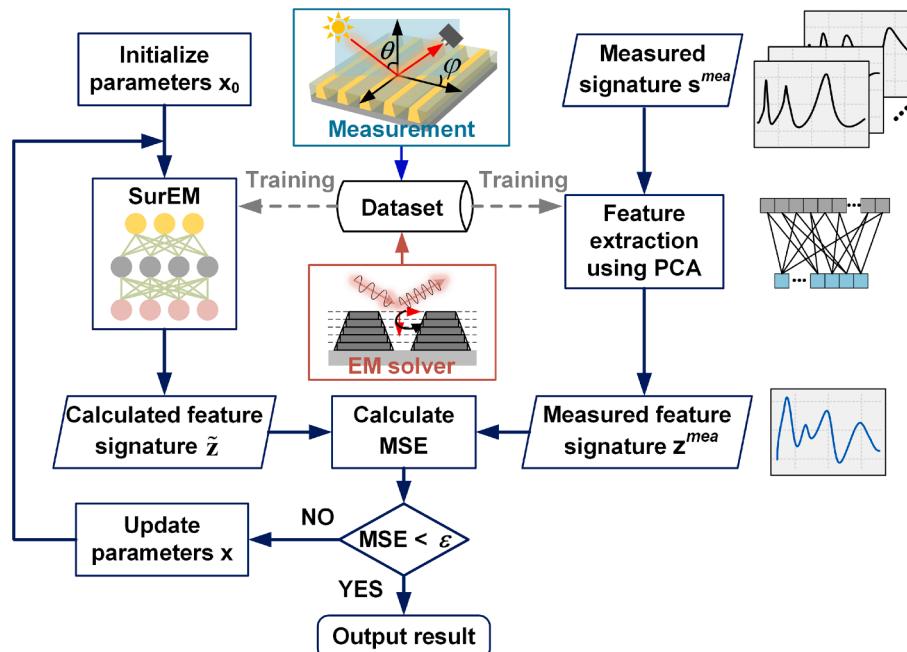


Fig. 1. Overview of the proposed MLER method.

reached. If yes, go to step 6. Otherwise, update the parameters using the nonlinear regression strategy and return to step 3 until the termination condition is met.

Step 6: Output measurement result \mathbf{x}^{mea} .

Through the above steps, the structural parameters of the sample can be determined, with the minimal difference between its corresponding calculated signature and the measured one. It is worth mentioning that the success of MLER method depends on the nonlinear regression strategy and the SurEM. Typically, Levenberg-Marquardt (LM) algorithm [30,31] is chosen in the proposed method, due to its fast convergence. The termination threshold could be set to a value equal to the minimum MSE between the feature signatures in the training dataset. The training dataset is generated by EM solvers, for the cases that the measurement process can be simulated. If there is not a physical model for measurement, the training dataset can be produced by measuring design of experiment wafers using other metrology techniques, such as scanning electron microscopy. In order to improve the performance of the SurEM, a signature dimensionality reduction approach is presented based on PCA and forward selection algorithm, and the architecture of the SurEM is meticulously designed. The signature dimensionality reduction approach and the design of SurEM will be discussed in detail in the following.

2.2. Signature dimensionality reduction using PCA

The goal of PCA is to find an optimal linear transformation, removing redundant information in signatures with lower dimensionality and minimal loss of information. Assuming there are N different signatures, an $N \times n$ signature matrix $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N]^T$ can be constructed, where the superscript “T” represents transpose. And the estimated mean vector \mathbf{Mu} can be calculated, defined by

$$\mathbf{Mu} = [\bar{s}_1, \bar{s}_2, \dots, \bar{s}_n] \quad (2)$$

where $\bar{s}_i = \frac{\sum_{j=1}^N s_{ji}}{N}$, ($i = 1, 2, \dots, n$). The estimated mean vector \mathbf{Mu} is used to centre the signature matrix \mathbf{S} and the centralized signature matrix $\tilde{\mathbf{S}}$ can be defined by

$$\tilde{\mathbf{S}} = \mathbf{S} - \mathbf{Mu} \otimes [1, 1, \dots, 1]^T \quad (3)$$

where “ \otimes ” represents the Kronecker product and $[1, 1, \dots, 1]^T$ is an N -dimensional vector. By calculating the covariance between the different elements of the signature, the covariance matrix \mathbf{C} is obtained, given by

$$\mathbf{C}_{k,l} = E[\mathbf{d}_k \cdot \mathbf{d}_l] - E[\mathbf{d}_k]E[\mathbf{d}_l], \quad (k, l = 1, 2, \dots, n) \quad (4)$$

where \mathbf{d} is the column vector of the centralized signature matrix $\tilde{\mathbf{S}}$, and $E[\cdot]$ denotes mathematical expectation. Then singular value decomposition is employed to calculate the eigenvalue λ_r and corresponding eigenvector (also known as principle component) ξ_r of the covariance matrix \mathbf{C} , which satisfy the following relationship:

$$\mathbf{C}\xi_r = \lambda_r \xi_r, \quad r \leq n \quad (5)$$

The larger the eigenvalue, the more information in the projection space corresponding to the eigenvector. Thus, the eigenvectors are sorted in descending order of corresponding eigenvalues. Once the value R of principal components is determined, the specific form of PCA transformation (PCAT) and PCA pseudo-inverse transformation (pi-PCAT) formula can be obtained, as follows:

$$\mathbf{F}(\mathbf{s}) = \mathbf{T}^T \cdot (\mathbf{s} - \mathbf{Mu}^T) \quad (6)$$

$$\tilde{\mathbf{F}}(\mathbf{z}) = \mathbf{T} \cdot \mathbf{z} + \mathbf{Mu}^T \quad (7)$$

where \mathbf{T} is an $n \times R$ transformation matrix, written as $\mathbf{T} = (\xi_1, \xi_2, \dots, \xi_R)$.

To determine the value of R , forward selection algorithm has been employed. In the process of forward selection shown in Fig. 2, the number of principal components starts from the initial setting value R_0 , and the minimum mean square error between the signatures in dataset is used as termination threshold ε_{PCA} , given by

$$\varepsilon_{PCA} \leq \min_{\mathbf{s}_i, \mathbf{s}_j \in S} \frac{1}{n} (\mathbf{s}_i - \mathbf{s}_j)^T \cdot (\mathbf{s}_i - \mathbf{s}_j) \quad (8)$$

According to initial value R_0 , the formula of PCAT and pi-PCAT can be specified. The maximum difference between the signatures before and after transformation within the dataset is noted as signatures loss $SLoss$. And, maximum mean square error is used to assess the signatures loss, given by

$$SLoss = \max_{\mathbf{s}_i \in S} \frac{1}{n} \left\{ \mathbf{s}_i - \tilde{\mathbf{F}}[\mathbf{F}(\mathbf{s}_i)] \right\}^T \left\{ \mathbf{s}_i - \tilde{\mathbf{F}}[\mathbf{F}(\mathbf{s}_i)] \right\} \quad (9)$$

And only single eigenvector is accessed in the next cycle in descending order of eigenvalues, until reaching the termination condition $SLoss < \varepsilon_{PCA}$. From the above, the reduced dimension of the signature can be determined. It is worth noting that in order to reduce processing time, the initial setting value and termination threshold can be estimated through performing the above procedure on a sub-dataset.

2.3. ANN-based SurEM solver

In order to address the time consuming of EM solvers during the process of solution to the inverse problem, a well-designed SurEM is constructed and trained. The SurEM with a fully connected neural network architecture will be built to construct the mapping from parameter space to high dimensional feature signature space. As shown in Fig. 3, the SurEM consists of three parts: input layer for importing the variable parameters \mathbf{x} , middle computational module (MCM) enabling the capacities of non-linear computation, and output layers for predicting feature signature $\tilde{\mathbf{z}}$. MCM is composed of n_h hidden layers, each of which has n_n neurons. Each neuron receives and processes the outputs of the neurons in the former layer, whose mathematical description is given as,

$$Output = f \left[\sum_i (I_i * w_i) + b \right] \quad (10)$$

where I_i and w_i are one of the inputs and corresponding weight, b represents bias term, and f is an activation function. Considering computational efficiency, the rectified linear unit (ReLU) is chosen as the activation function. For the sake of minimizing the errors between the

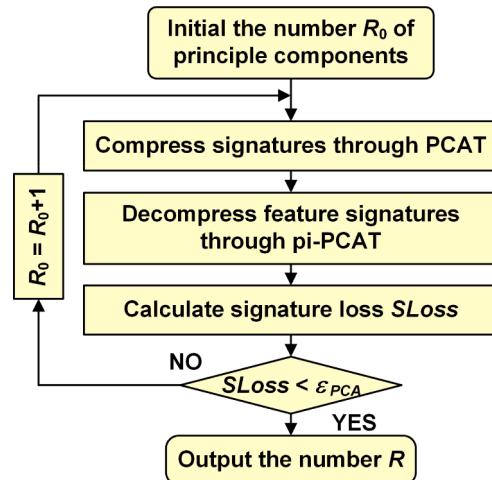


Fig. 2. Flowchart of determining the number of principle components.

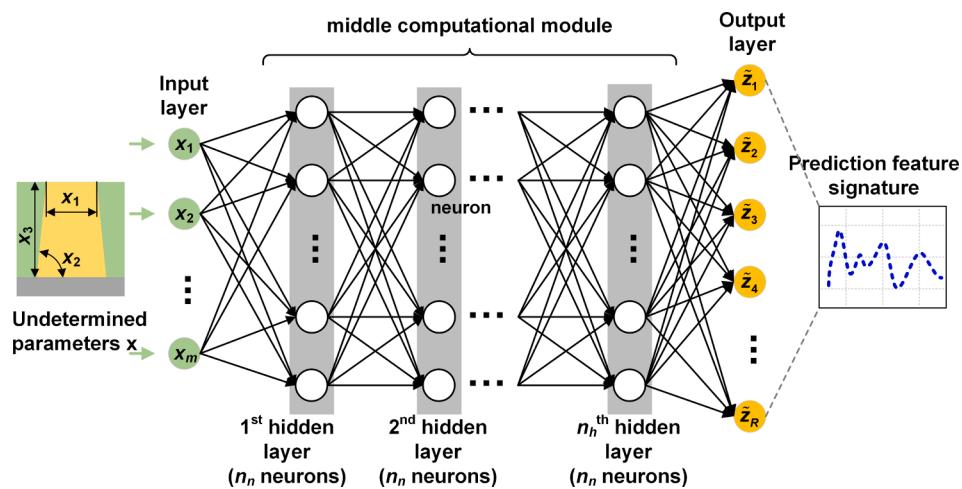


Fig. 3. Schematic of the architecture of SurEM.

prediction and the target feature signature, the back-propagation algorithm is performed to continuously tune the weights and biases of neurons during an iterative training process. And the MSE is used as the loss function.

Note that the calculation capacity of the SurEM depend on MCM. Thus, an exhaustive grid search method is employed to design the architecture of MCM. By comparing different structures, the optimal value of n_h and n_n can be determined, ensuring the SurEM with excellent performance.

3. Results

3.1. Training dataset generation

In order to assess the theoretic feasibility of the proposed MLER method, a typical one-dimensional periodic structure is investigated, as shown in Fig. 4. The sample consists of a tungsten (W) grating layer filled with silicon dioxide (SiO_2), two silicon dioxide layers and one silicon nitride (SiN) layer deposited on the silicon (Si) substrate. Optical properties of W, SiO_2 , SiN and Si were taken from Ref. [32]. To construct the geometric structure model, it is necessary to take the possible profiles into account as much as possible. However, considering the complexity of the model and the efficiency of parameter extraction, some structural parameters can be specified by the pre-measurement and are generally fixed, whose tiny errors have little effect on the signature. Therefore, the profile of the structure is characterized by middle critical dimension MCD , gating height Hg , sidewall angle SWA , top SiO_2 layer thickness Tt , SiN layer thickness Ts , bottom SiO_2 layer thickness Tb , and period $pitch$. Nominal dimensions of the testing sample are: $MCD = 34 \text{ nm}$, $Hg = 65 \text{ nm}$, $SWA = 90^\circ$, $Tt = 600 \text{ nm}$, $Ts = 27 \text{ nm}$, $Tb = 10.5 \text{ nm}$, and $pitch = 78 \text{ nm}$. In the experiments, structural parameters of the investigated sample that need to be extracted include

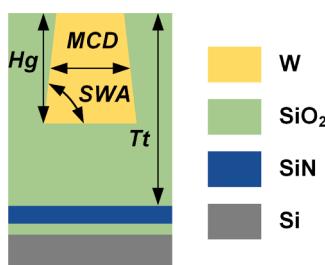


Fig. 4. (a) Schematic of the investigated periodic structure used in simulations and experiments.

MCD , Hg , SWA and Tt , while Ts , Tb and $pitch$ are fixed at their nominal dimensions. Since the signature varies with orientational placement of the testing sample, the azimuth angle φ is also taken into account, which describes the angle between the incident plane and the orientation of the grating. Therefore, the parameters to be measured consist of MCD , Hg , SWA , Tt and φ , called undetermined parameters.

An EM solver based on RCWA is developed to generate the dataset for training, including different sets of undetermined parameters and corresponding signatures. Mueller matrix spectra are used as optical signatures in this paper, since they have more plentiful information enabling sensitive nanostructure measurements [33–35]. The ranges of the structural parameters MCD , Hg , SWA , and Tt are 33–38 nm, 55–69 nm, 89–100°, and 576–630 nm. Considering the symmetry of the testing sample, the range of the azimuth angle φ is 0–90°. Details of the training dataset are listed in Table 1. The training dataset is generated by randomly selecting values in parameters ranges. The measurement conditions are set as follows: the incidence angle is fixed at 45°, spectral range is set to 320–800 nm with an interval of 5 nm.

3.2. Signature dimensionality reduction

To reduce the inevitable redundant information in signatures, the signature dimensionality reduction approach mentioned in Section 2.2 is implemented to extract the lower-dimensional feature signatures. As used in this paper, each Mueller matrix spectra $\mathbf{m}_{i,j}$ ($i, j = 1, 2, 3, 4$) is a 97-dimensional vector, and all the elements are concatenated in the order of rows except the \mathbf{m}_{11} . Therefore, each signature is a 1455-dimensional vector $\mathbf{s} = [\mathbf{m}_{12}, \mathbf{m}_{13}, \dots, \mathbf{m}_{44}]^T$. Then the specific formula of PCAT and pi-PCAT are obtained, which can be accomplished in two steps.

In the first step, a sub-dataset containing 10,000 signatures is generated by random selection from the training dataset. The minimum MSE between the signatures equals to 9.6365×10^{-6} . Considering the size of the training dataset, the value of termination threshold ε_{PCA} is taken as 1×10^{-6} , and the initial reduced dimension of the signature R_0 is set to 1. Then the data processing procedure described in Fig. 2 is

Table 1
Details of the training dataset used in simulations and experiments.

Parameter name	Nominal dimension	Lower bound	Upper bound	Training dataset size
MCD	34 nm	28 nm	40 nm	400,000
Hg	65 nm	51 nm	81 nm	
SWA	90°	85°	101°	
Tt	600 nm	565 nm	640 nm	
φ	–	0°	90°	

executed on the sub-dataset. When the number of principle components equals 152, the termination condition $S\text{Loss} < \varepsilon_{\text{PCA}}$ is reached.

Next, the data processing procedure described in Fig. 2 is executed on the training dataset. The $R_0 = 152$ is taken as the initial number of principal components. As shown in Fig. 5, the signatures loss $S\text{Loss}$ decreases as the number of principal components increases, and the termination condition is met as the number of principal components increasing to 204.

Through the above processing, the dimension of feature signature significantly drops down to 204, only one seventh of the original signature. Signature loss according to Eq. (9) before and after transformation is less than 1×10^{-6} , meaning that most of the information is retained. Then PCAT is performed on signatures in training dataset to extract the feature signatures for training the SurEM.

3.3. Training of the ANN-based SurEM solver

The neural network models are designed and trained by using Pytorch. All training processes are performed on a computer workstation with an Intel Xeon E5-2643v4 @3.4 GHz central processing unit (CPU), graphics processing unit (GPU) board NVIDIA GeForce RTX 2080 Ti, and 128 GB access memory.

For the sake of designing the SurEM, the training performances of different structures of SurEM are compared using exhaustive grid search method. In the design stage, the number n_h of hidden layers is chosen to be 2, 3, 4, and the number n_n of neurons per hidden layer is set to 200–1000 with an interval of 20. Then, these candidate models are constructed and trained on the sub-dataset. Fig. 6 presents that the training losses of different models after trained with a learning rate of 1×10^{-3} for 400 iterations. As can be observed from Fig. 6, the models with three and four hidden layers perform comparably, and better than those with two hidden layers. It means that increasing the number of hidden layers to more than three will not improve the performance of the model. The loss of models with 3 hidden layers tends to decrease as the number of neurons increases. And when the number of neurons is greater than 900, there is no significant improvement in performance. This suggests that the SurEM with 3 hidden layers and 900 neurons per hidden layer is the optimal model, considering the computational capability and complexity.

The training of SurEM is started with a learning rate of 1×10^{-3} and the learning rate reduce to 1×10^{-5} after 1000 iterations. The number of training iterations for SurEM is set to 2000. Adam [36] optimizer is chosen to tune the neural networks on account of its fast convergence. In addition, to demonstrate that signature dimensionality reduction is beneficial for performance improvement, a fully connected neural network with the same structure of MCM in SurEM is also constructed, named SNET, which could predict the signature for the given parameters. The training configuration of SNET is the same as that of SurEM, except that the number of training iterations is set to 3000, considering

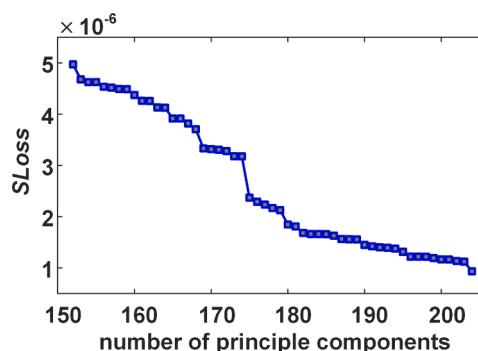


Fig. 5. Signature losses before and after transformation with different number of principle components.

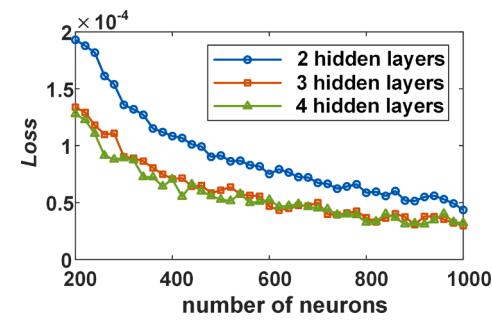


Fig. 6. The training losses of candidate models after trained.

the different number of hyperparameters.

3.4. Testing of the ANN-based SurEM solver

To examine the performance of these models, 100 groups of parameters are randomly generated within the range as listed in Table 1. The signature calculated by RCWA solver is treated as target signature. And prediction signatures are obtained by employing SurEM combined with pi-PCAT and SNET, respectively. Fig. 7 depicts the MSEs between the target and prediction signatures. In Fig. 7, the mean MSEs obtained by SurEM combined with pi-PCAT and SNET are 1.185×10^{-7} and 1.92×10^{-7} , respectively, which reveal that the loss of information caused by signature dimensionality reduction can be ignored. Fig. 8 presents the comparison between the target and prediction signature obtained by SurEM combined with pi-PCAT, whose MSE is the maximum in the testing dataset and equals to 5.41×10^{-7} . It is obvious that the prediction signature is in great agreement with the target.

The above results indicate that the proposed signature dimensionality reduction approach is instrumental in ensuring the accuracy of the surrogate model, simplifying the architecture and improving the computational efficiency of the SurEM, due to the effective reduction of the dimensionality and redundant information of origin signature. The well-designed SurEM is provided with capacity to calculate the feature signature for the given parameters.

3.5. Simulation results

Simulations are performed to demonstrate the validity of the proposed MLER method for the measurement of nanostructures. The simulated target signatures mentioned in Section 3.4 are treated as the measured signatures. The initial structural parameters are set to their nominal values, respectively, and the initial value of azimuth angle is set to the true value added by Gaussian random noise with standard deviation of 1° . LM algorithm in SciPy, an open-source software for mathematics, is utilized for the iterative solving, and the parameters used in the algorithm are the default values. The MLER method with the well-

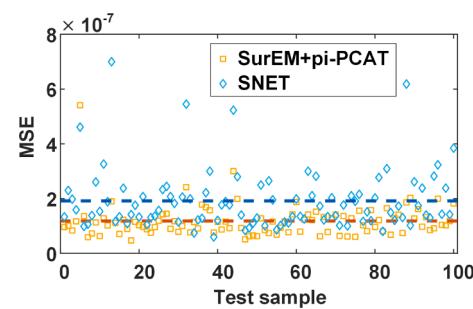


Fig. 7. The MSEs between the target and prediction signatures, with the dashed lines in red and blue denoting the mean MSEs of the SurEM combined with pi-PCAT and SNET, respectively.

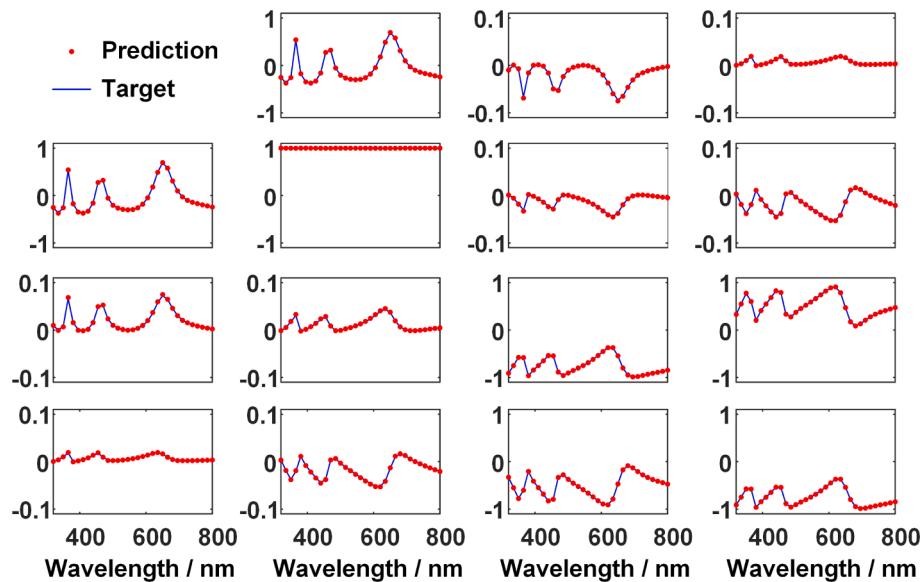


Fig. 8. Comparison between the target and prediction signature (Mueller Matrix spectrum) of SurEM combined with pi-PCAT. Parameters of the testing sample: $MCD = 33.403 \text{ nm}$, $Hg = 54.953 \text{ nm}$, $SWA = 85.570^\circ$, $Tt = 566.159 \text{ nm}$, $\varphi = 88.479^\circ$. We have lowered the sampling frequency of the prediction signature for a better demonstration.

trained SurEM is then implemented to extract parameters. The relative error is introduced to evaluate the differences between the extracted parameters by MLER and the true values, given by

$$\text{Relative error} = \frac{|p_{\text{MLER}} - p_{\text{true}}|}{p_{\text{true}}} \times 100\% \quad (11)$$

where p_{MLER} and p_{true} are the results of the MLER and true values, respectively. Fig. 9(a)-(e) depict the relative errors of the structural parameters for the 100 test samples with their mean relative errors. It is

observed that the relative errors of parameters MCD , Hg , SWA , Tt and φ are less than 0.15%, 0.30%, 0.09%, 0.012% and 0.15%, respectively. The mean relative errors of parameters MCD , Hg , SWA , Tt and φ are 0.028%, 0.062%, 0.020%, 0.003% and 0.027%, respectively. To make a comparison, we also test the performance of the DEDNN in parameter extraction. As indicated in Fig. S2, the mean relative errors of DEDNN are one order of magnitude larger than that of the MLER. From section 3.2, it is known that the minimum MSE between the signatures is less than 9.7×10^{-6} . The smaller the errors of the extracted parameters, the

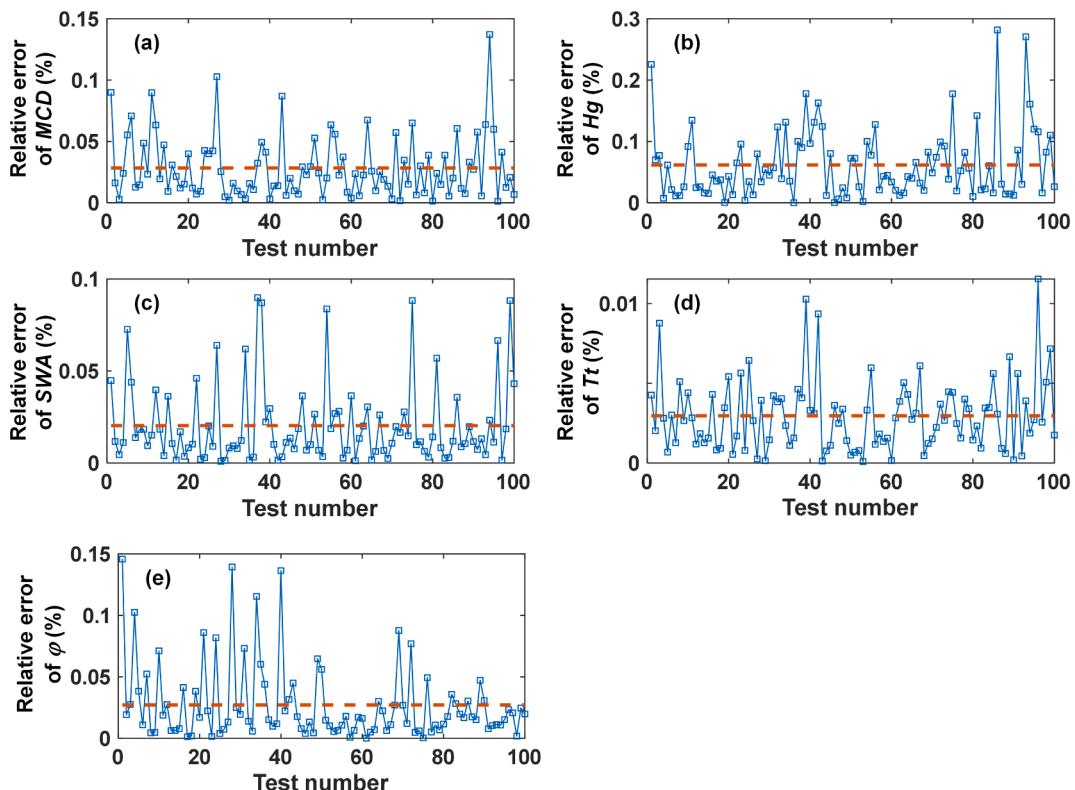


Fig. 9. (a)-(e) Relative errors of the extracted parameters MCD , Hg , SWA , Tt and φ for 100 tests, with the dashed lines denoting the mean relative errors.

higher the confidence of the MLER method. Consequently, these excellent results demonstrate that the MLER is sufficiently accurate to reconstruct the undetermined parameters of testing samples. In addition, in comparison with the DEDNN, the design and training of ANN in MLER are easier to implement for achieving the same accuracy.

3.6. Experimental results

In order to further confirm the robustness of MLER in practical measurements, a dual-rotating compensator Mueller matrix ellipsometer (ME-L, Wuhan Eoptics Technology Co.) is employed to measure the Mueller matrix spectra of the testing wafer. The wafer contains 91 dies, each of which has the same geometric structure shown in Fig. 4 but with different dimensions, including MCD , Hg , SWA and Tt . As schematically shown in Fig. 10, the system configuration of ME-L is $PC_r_1SC_r_2A$, where P and A stand for the fixed polarizer and analyzer, respectively, C_r_1 and C_r_2 are for the first and second rotating compensators, respectively, and S represents the sample. With the light source used in ME-L, the wavelengths available are in the 193–1000 nm range, including the spectral range used in this paper. The beam diameter used in the experiments is 200 μm . The two arms and the sample stage of the ME-L can be rotated to change the incidence and azimuthal angles in experiments. The incidence angle is set to 45°, and the zero-order diffraction beam is collected by the detector. Through analyzing the optical intensity signal, we can obtain the Mueller matrix spectra of the testing samples. The accuracy and repeatability of ME-L are 0.13 nm and 0.001 nm, respectively, tested on a standard SiO_2 film with a thickness of 96.75 nm. In addition, ME-L could give the estimated azimuth angle once calibrated before measurement.

Library search method is the standard solution for the inverse problem of the optical scatterometry during in-line process control. For illustrating the practicality and accuracy of the proposed MLER, the results of MLER method are compared with the state-of-the-art library search method. The library search requires a signature library prior to the measurement, in which a best match with the measured signature could be found. Therefore, a signature-parameters library is generated using RCWA solver. The ranges of the parameters in library are described in Table 1. Different from the training dataset used for SurEM, the selection of values within the parameters ranges is done by gridding. The grid intervals for the parameters MCD , Hg , SWA , Tt and φ are 2 nm, 2 nm, 2 nm and 2°, respectively. The library contains more than 1.76 million data pairs.

We then perform library search and MLER method to extract the parameters. In the MLER method, the initial values of structural parameters are set to their nominal values, and the initial azimuth angles are set to the output values of the ME-L. Fig. 11 depicts the results of extracted parameters MCD , Hg , SWA , Tt and φ by two different methods. As can be observed, the coefficients of determination R^2 between the results of parameters MCD , Hg , SWA , Tt and φ are 0.9944, 0.9899, 0.9948, 0.9999 and 1.0, respectively. The high consistency between the

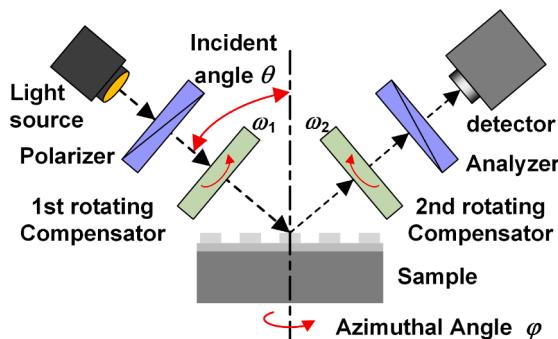


Fig. 10. Basic scheme of the dual-rotating compensator Mueller matrix ellipsometer used in the experiments.

results indicates that the proposed MLER method is very reliable and robust to experiment noise.

To quantify the results, the relative bias is introduced to evaluate the differences between extracted parameters by the two methods, given by Eq. (11), but p_{true} is replaced by the results of the library search. Fig. 12 (a)-(e) depict the relative biases of the extracted parameters of the two methods. The mean relative biases of parameters MCD , Hg , SWA , Tt and φ are 0.104%, 0.338%, 0.140%, 0.010% and 0.077%, respectively. There are slight differences between the two methods, due to the noise introduced by the instrument and measurement environment. We have conducted 30 repeatability measurement experiments on the wafer. The repeatability of the results for each die is within the same scale. For clarity, the results for only one randomly selected die are presented. And the mean values obtained by the two methods are listed in Table 2. The absolute bias of the mean values of parameters MCD , Hg , SWA , Tt and φ are 0.016 nm, 0.050 nm, 0.065°, 0.022 nm and 0.006°. As shown in Table 2, the two methods have the same level of repeatability. The theoretical signatures of the parameters extracted by the MLER and library search are calculated using RCWA solver, and the MSEs between the theoretical and measured signatures are shown in Fig. 12(f). The mean difference between the two MSEs equals to 3×10^{-6} . It illustrates that the two methods are quite equivalent. The biases could be reduced via precise calibration of instruments and measurement. Consequently, we can conclude that the proposed MLER method can achieve a high measurement accuracy. And it is worth mentioning that the amount of data required to implement MLER is only less than 1/4 of that to library search.

In comparison, Figs. S3(a)-(e) depict the biases between the results of the DEDNN and library search. The mean relative biases of parameters MCD , Hg , SWA , Tt and φ are 3.78%, 9.05%, 3.75%, 0.13% and 2.26%, respectively. The results indicate that the DEDNN is vulnerable to experimental noise. The MSEs between measured and theoretical signatures of DEDNN are slightly larger than that of library search and MLER, as shown in Fig. S3(f). It shows that even a little increase in the MSE between two signatures would lead to greater biases in parameters. The results demonstrate that the proposed MLER method is more feasible and robust than the method to map the signatures to profile parameters using the ANN.

The time cost of library search and MLER methods in the parameter extraction is evaluated. The calculation is carried out on a 2.60 GHz laptop with a GPU (NVIDIA GeForce GTX 1650). Fig. 13 shows the time cost for the 91 test samples. The mean time cost of the library search method is 1.2031 s, and that of the proposed MLER method is 0.0366 s. The measurement speed of MLER method is much faster than that of library search. The results demonstrate that the well-designed SurEM has the ability to significantly shorten the computation time in the MLER solution process. The reason is that the SurEM calculates the feature signatures with mathematical matrix operations, rather than from a physical point of view. What's more, the proposed signature dimensionality reduction approach can reduce the time cost of computing the error function.

4. Conclusions

In conclusion, a MLER method for accurate and fast reconstruction of nanostructures has been proposed and demonstrated. Firstly, the signature dimensionality reduction approach is presented to reduce the redundant information in signatures and extract corresponding feature signatures, based on the PCA and forward selection algorithm. Secondly, we establish and train an ANN-based SurEM, which is able to predict the feature signatures for given parameters. Finally, the results can be obtained by a comparison between the SurEM-predicted and measured feature signatures using a nonlinear regression algorithm. Simulations and experiments conducted on a periodic grating have confirmed that a fast measurement speed with mean time cost of 0.0366 s can be achieved by applying MLER method, which is only 1/30 of the library search. The

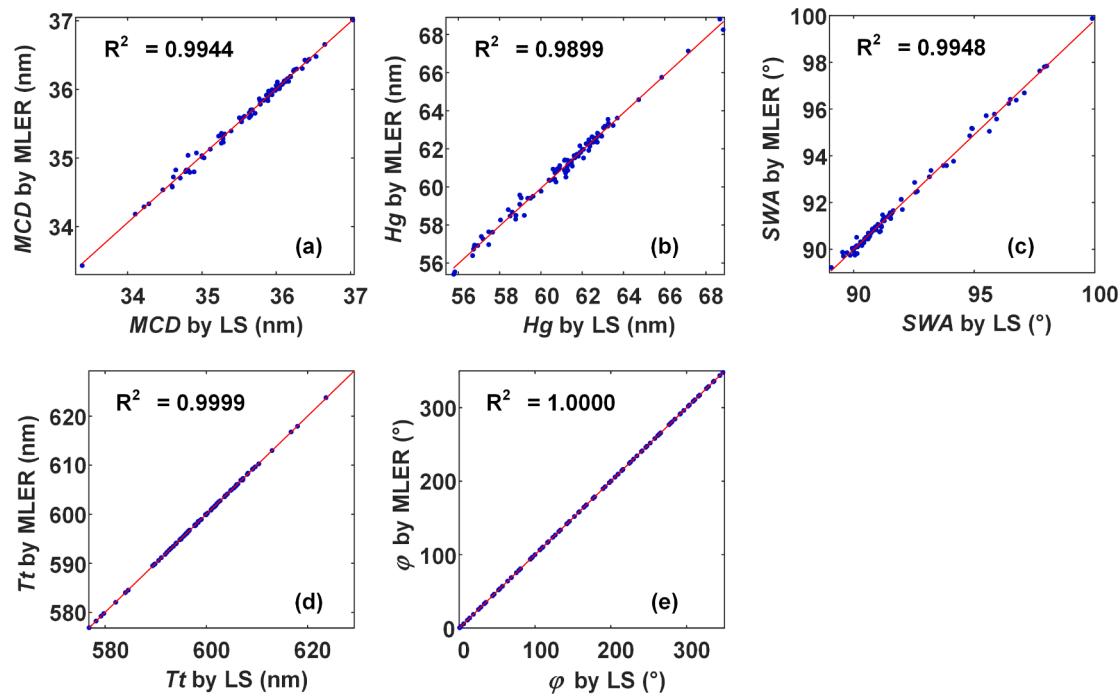


Fig. 11. Comparison between the parameters obtained by library search (LS) and MLER, respectively.

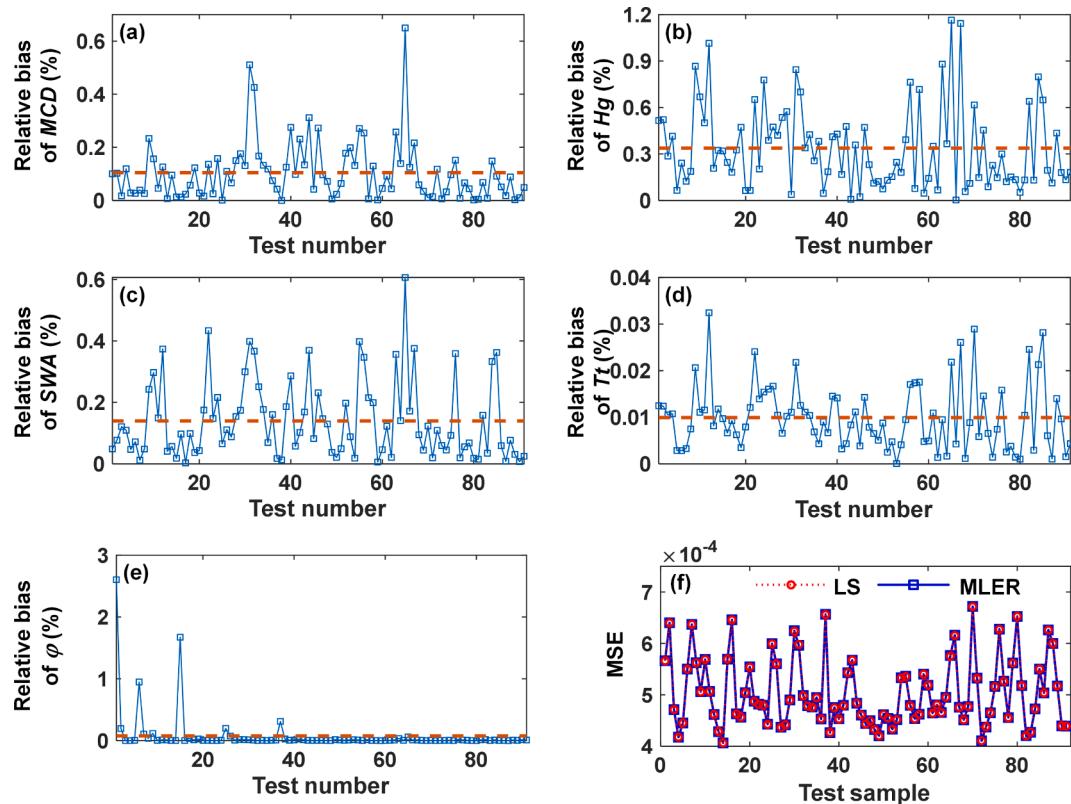
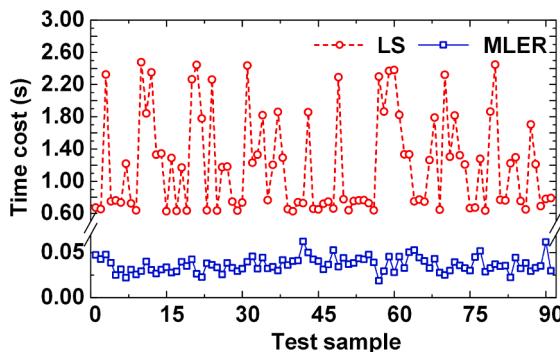


Fig. 12. (a)-(e) Relative biases between the extracted parameters MCD , Hg , SWA , Tt and φ extracted from practical measured signatures by library search (LS), and MLER, respectively, with the dashed lines denoting the mean relative biases. (f) MSEs between the measured and theoretical signatures for the parameters extracted by LS and MLER, respectively.

Table 2

The mean values and standard deviations of the results.

Parameter name	Mean values		Standard deviations	
	MLER	Library search	MLER	Library search
MCD	34.678 nm	34.662 nm	0.009 nm	0.009 nm
Hg	58.901 nm	58.951 nm	0.042 nm	0.041 nm
SWA	90.637°	90.702°	0.017°	0.016°
Tt	614.271 nm	614.249 nm	0.040 nm	0.039 nm
φ	2.025°	2.031°	0.013°	0.012°

**Fig. 13.** Time cost of the library search (LS) and MLER in the parameter extraction.

results also show that the accuracy of MLER is in line with library search, but using less pre-generated data. In contrast to the conventional ANN-based solution methods which directly map the signatures to profile parameters, the architecture of the neural network in the proposed MLER method is simpler and easier to establish. And the proposed method is robust in the presence of measurement noise. Although a simple nanostructure is tested in the experiments, the MLER could be expected to address the reconstruction of samples with more complex structures. We also believe that the proposed method promises to be an efficient in-line measurement method for nano manufacturing.

Acknowledgements

This work was partially funded by the National Key Research and Development Plan of China (2019YFB2005602), National Natural Science Foundation of China (Grant No. 52022034, 62175075, 51727809, 52130504), and Key Research and Development Plan of Hubei Province (Grant No. 2020BAA008).

Data availability

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

CRediT authorship contribution statement

Shuo Liu: Conceptualization, Methodology, Software, Writing – original draft. **Xiuguo Chen:** Conceptualization, Writing – review & editing, Funding acquisition, Supervision. **Tianjuan Yang:** Writing – review & editing. **Chunfu Guo:** Data curation, Resources. **Jiahao Zhang:** Writing – review & editing. **Jianyuan Ma:** Writing – review & editing. **Chao Chen:** Writing – review & editing. **Cai Wang:** Writing – review & editing. **Chuanwei Zhang:** Writing – review & editing. **Shiyuan Liu:** Funding acquisition, Supervision.

Declaration of Competing 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 A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.measurement.2022.110811>.

References

- [1] H. Fujiwara, Introduction to spectroscopic ellipsometry, in: H. Fujiwara (Ed.) Spectroscopic Ellipsometry Principles and Applications, 1–10. (John Wiley & Sons, England, 2007).
- [2] H.-T. Huang, W. Kong, F.L. Terry, Normal-incidence spectroscopic ellipsometry for critical dimension monitoring, *Appl. Phys. Lett.* 78 (25) (2001) 3983–3985.
- [3] C.J. Raymond, Scatterometry for semiconductor metrology, in: A.C. Diebold (Ed.) Handbook of silicon semiconductor metrology, 477–513. (CRC Press, Boca Raton, 2001).
- [4] M.H. Madsen, P. Hansen, Scatterometry-fast and robust measurements of nano-textured surfaces, *Surf. Topogr.: Metrol. Prop.* 4 (2016) 23003.
- [5] A.C. Diebold, A. Antonelli, N. Keller, Perspective: Optical measurement of feature dimensions and shapes by scatterometry, *APL Mater.* 6 (2018) 058201.
- [6] X. Chen, S. Liu, Scatterometry for semiconductor metrology, in: W. Gao (Ed.) Metrology. Precision Manufacturing, 477–513. (Springer, Singapore, 2019).
- [7] V.F. Paz, S. Peterhansel, K. Frenner, W. Osten, Solving the inverse grating problem by white light interference Fourier scatterometry, *Light Sci. Appl.* 1 (2012), e36.
- [8] P. Ansinielli, W.M.J. Coene, H.P. Urbach, Automatic feature selection in EUV scatterometry, *Appl. Opt.* 58 (22) (2019) 5916, <https://doi.org/10.1364/AO.58.005916>.
- [9] N. Farchmin, M. Hammerschmidt, P. Schneider, M. Wurm, B. Bodermann, et al., Efficient Bayesian inversion for shape reconstruction of lithography masks, *J. Micro/Nanolith. MEMS MOEMS* 19 (2020) 1.
- [10] X. Chen, S. Liu, C. Zhang, H. Jiang, Improved measurement accuracy in optical scatterometry using correction-based library search, *Appl. Opt.* 52 (2013) 6726–6734.
- [11] J. Zhu, S. Liu, C. Zhang, X. Chen, Robust solution to the inverse problem in optical scatterometry, *Opt. Express* 22 (2014) 22031.
- [12] M.G. Moharam, T.K. Gaylord, D.A. Pommet, E.B. Grann, Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach, *J. Opt. Soc. Am. A* 12 (5) (1995) 1077, <https://doi.org/10.1364/JOSAA.12.001077>.
- [13] H. Ichikawa, Electromagnetic analysis of diffraction gratings by the finite-difference time-domain method, *J. Opt. Soc. Am. A* 15 (1) (1998) 152, <https://doi.org/10.1364/JOSAA.15.000152>.
- [14] N.G. Orji, M. Badaroglu, B.M. Barnes, C. Beitia, B.D. Bunday, et al., Metrology for the next generation of semiconductor devices, *Nat. Electron.* 1 (2018) 532–547.
- [15] B. Bunday, A.F. Bello, E. Solecky, Vaid A. 7/5 nm Logic Manufacturing Capabilities and Requirements of Metrology, in: Proc. SPIE 105850I (SPIE: San Jose, California, United States, 2018).
- [16] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [17] K.A. Brown, S. Britzman, N. Maccaferri, D. Jariwala, U. Celano, Machine Learning in Nanoscience: Big Data at Small Scales, *Nano Lett.* 20 (1) (2020) 2–10.
- [18] J. Liu, D. Zhang, D. Yu, M. Ren, J. Xu, Machine learning powered ellipsometry, *Light Sci. & Appl.* 10 (2021) 55.
- [19] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, et al., Natural Language Processing (almost) from Scratch, *J. Mach. Learn. Res.* 12 (2011) 2493–2537.
- [20] W. Ma, F. Cheng, Y. Xu, Q. Wen, Y. Liu, Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy, *Adv. Mater.* 31 (2019) 1901111.
- [21] S. So, T. Badloe, J. Noh, J. Bravo-Abad, J. Rho, Deep learning enabled inverse design in nanophotonics, *Nanophotonics* 9 (2020) 1041–1057.
- [22] M. Liu, C. Fai Cheung, N. Senin, S. Wang, R. Su, R. Leach, On-machine surface defect detection using light scattering and deep learning, *J. Opt. Soc. Am. A* 37 (9) (2020) B53, <https://doi.org/10.1364/JOSAA.394102>.
- [23] S. Wei, L. Li, Measurement of photoresist grating profiles based on multiwavelength scatterometry and artificial neural network, *Appl. Opt.* 47 (13) (2008) 2524, <https://doi.org/10.1364/AO.47.002524>.
- [24] S. Robert, A. Mure-Ravaud, D. Lacour, Characterization of optical diffraction gratings by use of a neural method, *J. Opt. Soc. Am. A* 19 (1) (2002) 24, <https://doi.org/10.1364/JOSAA.19.000024>.
- [25] C. Zhang, S. Liu, T. Shi, Z. Tang, Improved model-based infrared reflectometry for measuring deep trench structures, *J. Opt. Soc. Am. A* 26 (2009) 2327.
- [26] J.S.M. Madsen, S.A. Jensen, J. Nygard, P.E. Hansen, Replacing libraries in scatterometry, *Opt. Express* 26 (2018) 34622.
- [27] M. Godi Tchéhé, S. Robert, Z.S. Fawzi, B. Bayard, Reconstruction of a complex profile shape by weighting basic characterization results for nanometrology, *Appl. Opt.* 58 (22) (2019) 6118, <https://doi.org/10.1364/AO.58.006118>.

- [28] I. Kim, Y. Bae, S. Gwak, E. Kum, T. Jo, et al., Machine learning aided profile measurement in high-aspect-ratio nanostructures, in: Proc. SPIE. 117830K (SPIE: Online Only, 2021).
- [29] T. Li, A. Chen, L. Fan, M. Zheng, J. Wang, et al., Photonic-dispersion neural networks for inverse scattering problems, *Light Sci. Appl.* 10 (2021) 154.
- [30] K. Levenberg, A method for the solution of certain non-linear problems in least squares, *Q. Appl. Math.* 2 (2) (1944) 164–168.
- [31] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.* 11 (2) (1963) 431–441.
- [32] H. Fujiwara, R.W. Collins, Spectroscopic Ellipsometry for Photovoltaics, in: W.T. Rhodes (Ed.) Applications and Optical Data of Solar Cell Materials, 322–573. (Springer, Switzerland, 2018).
- [33] T. Novikova, A. De Martino, S.B. Hatit, B. Dréville, Application of Mueller polarimetry in conical diffraction for critical dimension measurements in microelectronics, *Appl. Opt.* 45 (16) (2006) 3688, <https://doi.org/10.1364/AO.45.003688>.
- [34] S. Liu, X. Chen, C. Zhang, Development of a broadband Mueller matrix ellipsometer as a powerful tool for nanostructure metrology, *Thin Solid Films.* 584 (2015) 176–185.
- [35] M. Korde, S. Kal, C. Alix, N. Keller, G.A. Antonelli, et al., Nondestructive characterization of nanoscale subsurface features fabricated by selective etching of multilayered nanowire test structures using Mueller matrix spectroscopic ellipsometry based scatterometry, *J. Vac. Sci. Technol. B.* 38 (2020) 24007.
- [36] D.P. Kingma, B.J. Adam, A Method for Stochastic Optimization, CoRR. abs/1412.6980 (2015).