CHINESE OPTICS LETTERS

Vol. xx, No. x | March 2025

Enhanced Profile Reconstruction of Small Angle X-ray Scattering Measurement via Correlation Learning

Hairui Yang (杨海瑞)[1,2,3], Zhaolong Wu(吴赵龙)[2],and Hong Yu (喻 虹)[1,2*]

[1] Key Laboratory of Quantum Optics, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China

[2] Zhangjiang Laboratory, Shanghai 201210, China

[3] University of Chinese Academy of Sciences, Beijing 100049, China

*Corresponding author: yuhong@zjlab.ac.cn

Received Month X, XXXX | Accepted Month X, XXXX | Posted Online Month X, XXXX

Small angle x-ray scattering (SAXS) is a promising metrology technology for complex nanostructures in semiconductor manufacturing. However, parameter reconstruction based on SAXS measurement often faces challenges in achieving high precision and repeatability, due to the increasing complexity of structures and the demands for precise measurement. To address these challenges, a correlation learning-based method is proposed to enhance the accuracy and reduce the uncertainty of the profile reconstruction in SAXS measurement. This method leverages the Long Short-Term Memory (LSTM) mechanism to capture and learn inherent parameter correlation effectively. The precision and reliability of the proposed method are demonstrated through the simulations of synthetic Si gratings. Our method exhibits remarkable measurement accuracy with an improvement of at least 13.9% and the measurement repeatability is nearly 1.4 times higher compared to the previous learning-based methods. We expect that our approach will provide a novel solution for SAXS measurement, enabling accurate and reliable profile reconstruction of nanostructures.

Keywords: X-ray optics, Small angle x-ray scattering, Profile metrology, Deep learning.
DOI: xxxxxxx/COLxxxxxxx.

1. Introduction

Small angle x-ray scattering (SAXS), as a non-contact, non-destructive, and high-resolution measurement scheme,[1,2] has gradually become one of the most important techniques for measuring diverse nanostructures, especially in materials science and biology.[3–5] It is also applied to inline critical dimension metrology in semiconductor manufacturing to ensure product quality, and optimize fabrication process.[6–8] However, SAXS measurement, a typical ill-posed inverse problem,[9] is not straightforward as it involves a complicated process of estimating structural parameters from the scattering patterns.[10,11] Low signal-to-noise ratio often leads to inaccurate or even erroneous reconstruction results, and high uncertainty.[12] And these drawbacks become more severe with the increasing complexity of the structures,[13] which will seriously interfere with the correct perception of nanostructures and fabrication process, thus it is crucial to employ certain methods to enhance measurement accuracy and reduce uncertainty.[12]

Incorporating certain priors into the reconstruction process can improve the measurement precision in various semiconductor measurement techniques.[11,12,14,15] This strategy is also beneficial to improve reconstruction accuracy and reduce uncertainty in the SAXS inverse problem, such as providing a fin-shaped mirrored prior.[16,17] However, conventional iteration-based methods,[3,7,11,18,19] despite incorporating certain prior knowledge, still suffer from inherent inefficiency, because they require calling the predefined forward physical model to calculate the scattering patterns and assess many candidate solutions.[13,20] Deep learning-based methods have gained significant attention in semiconductor metrology,[21–24] due to their exceptional capabilities in rapid inference and robust feature extraction.[25] Recently, researchers have extended its application to parameter reconstruction for nanostructured samples in SAXS measurement, including the tilt of memory hole structures, the width and height of simple trapezoidal gratings, and the profile of complex profile gratings.[22,26,27] These approaches can be

regarded as model-dependent neural networks,[28] providing prior knowledge from the perspective of the physical forward model. These successes provide valuable insights: by integrating reliable prior knowledge into the reconstruction algorithm or guiding the algorithm to learn such prior, we can greatly enhance the accuracy and reliability of the reconstruction process.[29] In semiconductor manufacturing, the etching process often leads to the aspect ratio-dependent etching (ARDE) effect[30] or significant tilt variation near the edge of the wafer.[31] These phenomena also result in strong nonlinear inter-layer correlation for the nanostructure.[32] Hence, guiding the reconstruction process to learn this correlation serves as a potential enhancement scheme.

In this work, we propose a correlation learning-based method for profile reconstruction in SAXS measurement. Utilizing the Long Short-Term Memory (LSTM) mechanism,[33] our approach can actively learn the inter-layer correlation of profile structural parameters. Based on this characteristic, we design and evaluate a novel network model to reconstruct the profile parameters of synthetic complex Si gratings. The analysis results demonstrate this learning-based approach significantly enhances accuracy and reduces uncertainty in nanostructure profile measurement.

2. Theory and method

2.1. X-ray scattering principle

SAXS measurement can well extract structural parameters of nanostructured samples, which the schematic diagram is shown in Fig. 1. Figure 1(a) illustrates the fundamental process of small angle x-ray scattering. The monochromatic x-ray beam impinges upon the sample at an angle of incidence (AOI) and the scattering pattern within a small angle region is recorded by a detector, subsequently. After exposures of multiple AOIs, we can integrate the recorded scattering patterns to get the scattering map of the sample,[16] which is shown in Fig. 1(b) and 1(c). According to the kinematical approximation, if we know the relative electron destiny $\rho(\mathbf{r})$ of a sample, its Fourier transform $F(\mathbf{q})$ is,

$$F(\mathbf{q}) = \int \rho(\mathbf{r}) \exp(-i\mathbf{q}\cdot\mathbf{r}) d\mathbf{r}, \tag{1}$$

where $\mathbf{r}$ is the spatial vector in real space, $\mathbf{q}$ is the transferred wave vector in the reciprocal space, and the scattering vector $\mathbf{q}$ is considered as the vector difference between the incidence wave vector and the scattering wave vector, $\mathbf{q} = \mathbf{k} - \mathbf{k}'$. According to the first-order Born approximation, the scattering map is expressed as $I(\mathbf{q}) \propto \langle F(\mathbf{q}), F^*(\mathbf{q})\rangle_\Omega = |F(\mathbf{q})|^2$. Thus, considering the periodicity of samples, surface roughness, and noise, the scattering map is further expressed as,[1,16]

$$I(\mathbf{q}) = |F(\mathbf{q})|^2 S(\mathbf{q}) \exp(-\mathbf{q}^2\sigma_{DW}^2) + I_{\text{noise}}(\mathbf{q}), \tag{2}$$

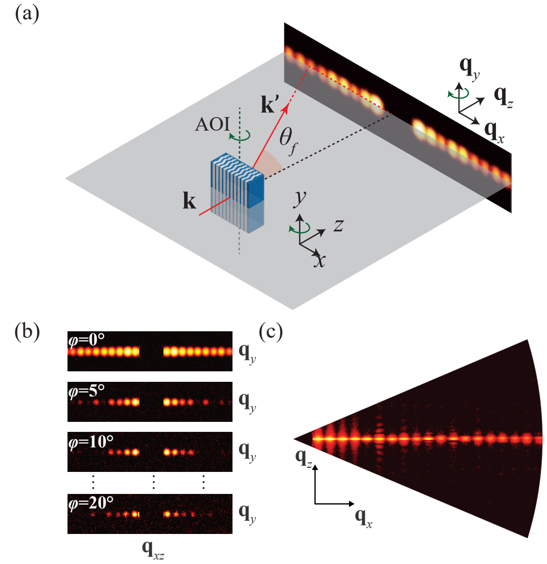

Fig. 1. (a) is the schematic diagram of SAXS measurement. (b) is the scattering patterns of multiple AOIs. (c) is the scattering map of the sample.

where $S(\mathbf{q})$ is the structure factor. $\exp(-\mathbf{q}^2\sigma_{DW}^2)$ is the Debye-Waller factor, and σ_{DW} is the root mean square of surface roughness amplitude. $I_{\text{noise}}(\mathbf{q})$ is the noise, including the background noise and Poisson noise.

Figure 2(a) is the top and side view of a complex profile grating sample. The sample is characterized by an inter-line distance L, which is called as the grating pitch. Hence, the structure factor $S(\mathbf{q})$ can be expressed as $\sum_{a=-\infty}^{\infty} \delta(|\mathbf{q}_x| - \frac{2\pi a}{L})$, where a is the diffraction order. In the side view, the sample can be modeled as a stack of rectangles. Its profile is demonstrated by a structural profile parameter set $\boldsymbol{t} = \{CD_n, CP_n\}_n^N$, where N is the structure's layer number and CD_n and CP_n are critical dimension and center position of the n-th layer, respectively.[34]

In an ideal condition, the CDs of each layer are equal, and the CPs of each layer are all zero. However in the fabrication processes (i.e. etching or multi-patterning processes),[35] nanostructures may demonstrate various types of phenomena,[7] including tilting, twisting, bowing, tapering, and potential combinations thereof, which are depicted in Fig.2(b). To investigate the impact of these phenomena on the profile parameters, we generated 100,000 samples. The samples consist of basic structures, including the above-mentioned basic profile phenomena.[7] These basic structures are randomly stacked to form different grating samples. Each grating sample is a stack of 40 rectangles with a thickness of 75 *nm* for a single rectangle. Through interpolation, the rectangle size of each layer varies smoothly with the height of a sample.[34] The mean value of CD at each height is about 75 *nm*, with a range of ± 20 *nm*, and the mean value of CP is 0 *nm*, with a range of ± 5 *nm*. The correlation ellipses for CPs and CDs are shown in Fig.2(d) and 2(e), which illustrate that the parameters of two adjacent layers are more correlated than

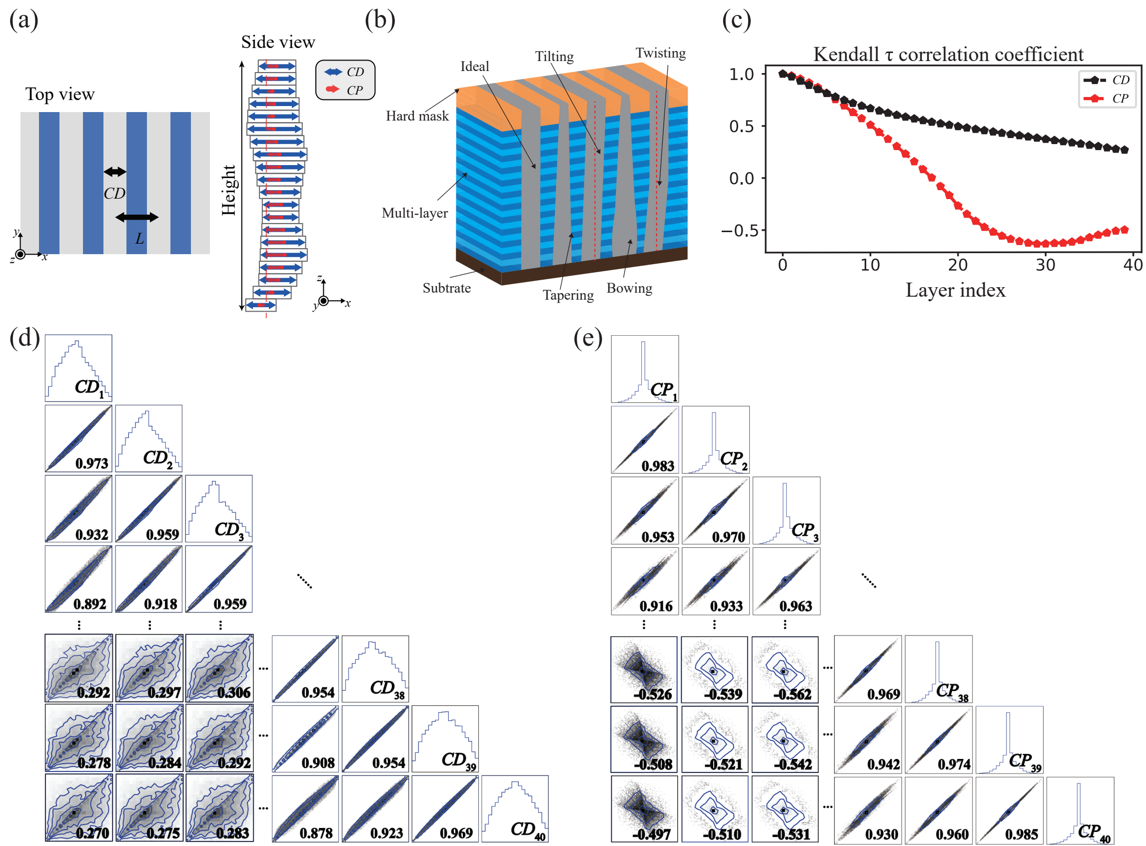

Fig. 2. (a) is the top and side view of a complex profile grating sample. The profile is demonstrated by a structural parameter set $t = \{CD_n, CP_n\}_n^N$. (b) is various types of phenomena, including tilting, twisting, bowing, tapering, and potential combinations thereof. (c) is the Kendall τ correlation coefficient for CDs and CPs. (d) and (e) are the correlation ellipses for CDs and CPs, which demonstrate that the parameters of two adjacent layers are more correlated than those of distant layers.

those of two distant layers, and this correlation may exhibit nonlinear characteristics. We calculate the Kendall τ correlation coefficient to quantify this kind of relationship in Fig.2(c). If two layers are further apart, the correlation coefficient between the corresponding parameters of these two layers will decrease. The correlation between CPs of two distant layers even reaches a nonlinear negative correlation. Similar characteristics can also be observed using Spearman correlation analysis. Utilizing this natural knowledge of the inherent inter-layer correlation may offer a potential way to improve SAXS measurement accuracy and uncertainty.

2.2. Correlation learning-based method

We propose a novel method that can learn the knowledge about the inter-layer correlation to enhance SAXS measurement accuracy and reduce uncertainty. For the SAXS inverse problem of reconstructing structural parameters from the scattering map, the scattering map (input) can be regarded as an angular series, while the structural parameters (output) can be seen as a layered series. The LSTM mechanism can effectively convey and express information across different series, and easily learn the inherent correlations within each series.[36] Hence, using the LSTM mechanism to solve this inverse problem has distinct advantages. Specifically, the LSTM achieves these through gate structures. The gate structures enable selective transmission of information through the Sigmoid function layer for point-by-point multiplication. The Sigmoid function regulates the weight of information exchange between neural layers. There are three types of gate structures in an LSTM cell, namely the forget gate, input gate, and output gate.

The forget gate determines which information from the last state needs to be forgotten. The gate gets the hidden state H_{t-1} from the last state and the input information x_t of the current state, and then outputs a weight f_t for the last cell state C_{t-1}. The Sigmoid function limits f_t between 0 to 1, which represents the forgetfulness extent of the hidden state H_{t-1}. In this way, LSTM can retain important information for a long period and the memory can be dynamically adjusted with the current input information x_t. The weight f_t, as well as its forget effect k_t on C_{t-1}, can

be written as follows:

$$\begin{cases} f_t = \sigma(w_f[H_{t-1}, x_t] + b_f) \\ k_t = C_{t-1} \otimes f_t, \end{cases} \tag{3}$$

where w_f and b_f are the weight and bias of the forget gate, respectively. $\otimes$ is the Hadamard product.

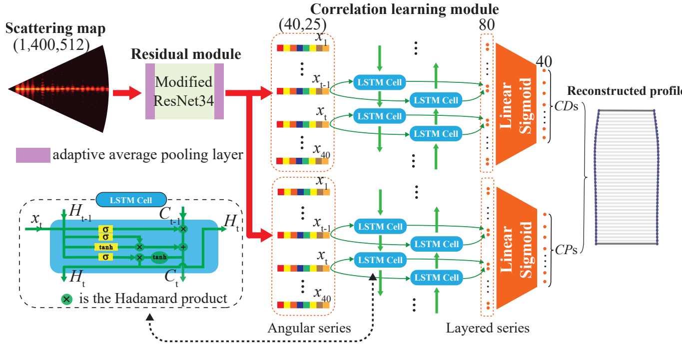

Fig. 3. Framework architecture of the correlation learning-based method. The framework, namely as CLNet, consists of a residual module and a correlation learning module. The residual module is a modified ResNet34, which has an adaptive average pooling layer at the beginning and another one at the end. The correlation learning module is a Bi-LSTM structure, which can directly reconstruct the profile parameters.

The input gate is the structure that determines how many components from the current input information x_t to add into the cell. The tanh function extracts the valid information from the x_t. Using the Sigmoid function, the input gate controls the extent i_t of valid information. In other words, the input gate filters the extracted valid information and scores each component from 0 to 1. The higher the score, the more component j_t will enter the current cell state C_t. The extent i_t of valid information and component j_t from the current input information x_t to add into the cell is demonstrated as,

$$\begin{cases} i_t = \sigma(w_i[H_{t-1}, x_t] + b_i) \\ j_t = \tanh(w_g[H_{t-1}, x_t] + b_g) \otimes i_t, \end{cases} \tag{4}$$

where w_i, w_g, b_i, and b_j are the weight and bias of the input gate, respectively.

The output gate is the neural gate that calculates the current state's output. Like the forget and input gate, the Sigmoid function also exists in the output gate, which helps to further extract the information from the current cell state C_t. Also, the current cell state C_t is mapped to the interval (-1, 1) as the current hidden state H_t by the tanh function. The current cell state C_t, output o_t and hidden state H_t are expressed as,

$$\begin{cases} C_t = k_t + j_t \\ o_t = \sigma(w_o[H_{t-1}, x_t] + b_o) \\ H_t = \tanh(C_t) \otimes o_t, \end{cases} \tag{5}$$

where w_o and b_o are the weight and bias of the output gate, respectively.

The overall framework (named as CLNet) has a residual module and a correlation learning module in Fig. 3. The residual module is used firstly to extract features from the scattering map and the correlation learning module is further used to reconstruct the profile parameters using the output of the residual module. The residual module is based on the backbone of ResNet34 and has a separate convolution layer at the front end, followed by four large blocks, which comprise three, four, six, and three groups of small blocks, respectively. Each of the small blocks contains two convolutional layers. An adaptive average pool layer then follows these and finally extracts a 40×25 angular series. The correlation learning module comprises two distinct channels dedicated to capturing the inherent parameter correlations of *CD*s and *CP*s, respectively. The channels directly reconstruct *CD*s and *CP*s respectively from the 40×25 series, which is shown in Fig.3. In each channel, the series needs the LSTM cell to take the progress of 40 time steps. The whole operations of the LSTM cell are also illustrated in the local enlargement of Fig.3, and the mechanism is described by Eq. (3) - (5). The bidirectional LSTM structure, which involves two sets of LSTM cells that process the series in the forward and backward directions is employed for accurate identification and reconstruction.[37] The bidirectional LSTM structure produces an output at each time step, and all the outputs form a new layered series to be fed into a fully connected layer with a Sigmoid function at the end of each channel. After linearly scaling the outputs of both channels to their respective reconstruction ranges, the reconstructed *CD*s and *CP*s are finally estimated. To suppress the overfitting caused by a large number of model weights, batch and layer normalization are performed after each convolution layer and LSTM layer.[38,39]

2.3. Data generation and training

We generate the scattering maps of grating samples mentioned in Section 2.1. The sample material is Si. The grating pitch L is 170 *nm*, and σ_{DW} is set to 2 *nm* for the surface roughness. The x-ray photon energy is 24 keV. The exposures of each sample are taken at 100 AOIs, uniformly spanning from -20° to 20°. The data recorded by the detector with 512×512 pixels only contains information in the middle few rows of pixels.[27] Hence, we take out 4×512 pixels from each AOI's scattering patterns. These data of multiple AOIs are concatenated together to 400×512 pixels. Poisson noise is added to the scattering data, where the noise level is related to the photon number. The photon number is 10 times the number of pixels in the scattering data, which is set according to actual experiment conditions.[40] The background noise level is a relative intensity

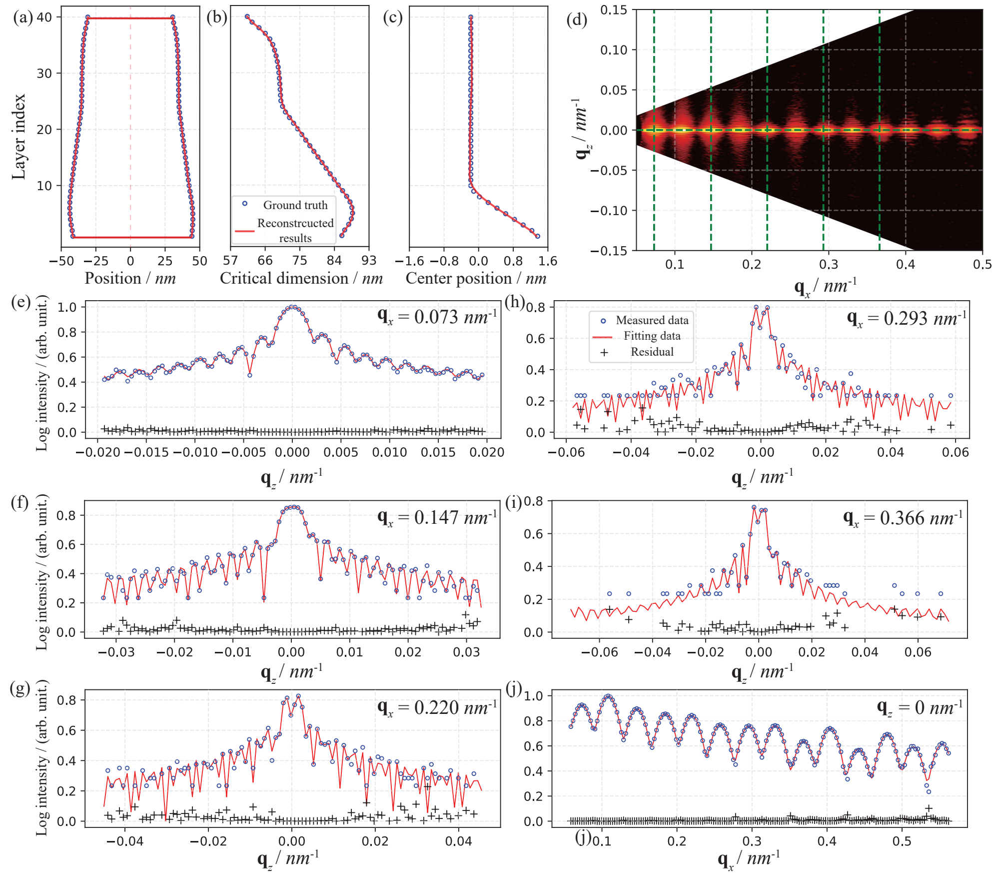

Fig. 4. Reconstructed results and corresponding fitting scattering data of a randomly selected sample. (a)-(c) are the reconstructed results for profile, *CD* and *CP*,respectively. (d) is the measured $\mathbf{q}_x$-$\mathbf{q}_z$ scattering map obtained from the scattering patterns at 100 AOIs, uniformly spanning from -20° to 20°. (e)-(j) are the measured data, fitting data, and the corresponding residuals at the $\mathbf{q}$-slices, which are indicated by dashed green lines in (d). In the high q regions, the residual increases due to the low signal-to-noise ratio of the measured data.

of 1×10^{-6}.[10] It should be noted that our method can apply to other X-ray photon energies theoretically.

The dataset includes the scattering maps and the corresponding structural parameters of the samples. The train, validation, and test sets are divided into 8:1:1. The model is trained on a workstation with two NVIDIA Quadro RTX 8000 GPUs and implemented using PyTorch. The Adam optimizer is used to update the network weights with the initial learning rate $lr = 1 \times 10^{-4}$ and $\beta_1 = 0.9, \beta_2 = 0.999$. The loss function is L1 loss with the physical symmetry prior.[27] The batch size is 16. These hyperparameters are chosen to ensure that the CLNet model can sufficiently understand and express the underlying information from the dataset. The training epoch is 200, taking about 25 hours in total.

3. Result and discussion

Figure 4 shows the reconstructed results and the corresponding fitting scattering data of one randomly selected instance in the test set. The reconstructed results for profile, *CD* and *CP* are shown in Fig. 4(a)-4(c), respectively. It shows that our method can precisely estimate the structural parameters. Figure 4(d) shows the measured scattering map of this sample recorded by the detector. To further evaluate the accuracy of our method, we recalculated the fitting data and chose six $\mathbf{q}$-slices to compare with the map in Fig. 4(d). Figure 4(e)-4(j) shows the measured data, fitting data, and corresponding residuals at $\mathbf{q}_x = 0.073, 0.147, 0.220, 0.293, 0.366\ nm^{-1}$ and $\mathbf{q}_z = 0\ nm^{-1}$. The fitting scattering data at these slices exhibits great consistency with the measured data.The proposed method achieves high fitting accuracy in low and medium q regions, where the measured scattering intensity is significant and the signal-to-noise ratio is high. However, in high q regions, the noise impact becomes more significant, leading to increased residuals and reduced accuracy.

Reconstruction can be achieved for diverse samples, including phenomena such as tilting, bowing, and tapering, as illustrated in Fig. 5. Figure 5(a) and 5(b) are the reconstructed *CD*s and *CP*s of three randomly selected

samples in the test set, respectively. In addition to CLNet, the results of well-trained ResNet34,[41] ResNet50,[27,41] AlexNet,[22,42] EffNet-b2,[43] and TinyViT-5m[44] are also shown in Fig. 5 for ablation and comparison experiments. The hyperparameters used by these models are consistent with those described in Section 2.3. It can be seen that our method performs more excellent in comparison. To quantitatively analyze and compare the performance of these six models, we use three kinds of key metrics to evaluate 10,000 instances in the test set: mean absolute error (MAE), R-squared value (R^2), and slope,

$$\begin{cases} \mathrm{MAE} = \frac{1}{M}\sum_{m=1}^{M}|U_m - V_m| \\ \mathrm{R}^2 = 1 - \frac{\sum_{m=1}^{M}(U_m - V_m)^2}{\sum_{m=1}^{M}(U_m - \bar{U})^2} \\ \mathrm{slope} = \frac{\sum_{m=1}^{M}(U_m V_m - M\bar{U}\bar{V})}{\sum_{m=1}^{M}U_m^2 - M\bar{U}^2} \end{cases}, \tag{6}$$

in which U is the ground truth, V is the reconstructed parameter, and M is the total number of samples.

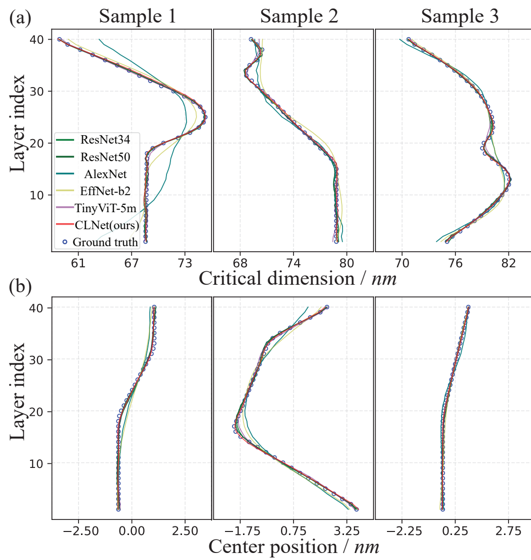

Fig. 5. Reconstructed results of three randomly selected samples in the test set. Each grating consists of 40 layers. (a) and (b) are the corresponding reconstructed *CD*s and *CP*s, respectively. Besides our method, the plots also include results from five other models, namely ResNet34, ResNet50, AlexNet, EffNet-b2, and TinyVit-5m. These are used for ablation and comparative experiments.

For clarity, it is essential to distinguish the three types of MAE since each grating sample has 40 layers, denoted as per-layer MAE, per-sample MAE and overall MAE.[27] The per-layer MAE refers to the MAE of each layer for all samples, and the per-sample MAE indicates the MAE of each sample for all 40 layers. The MAE of the entire test set is denoted as overall MAE. The results of per-layer MAE

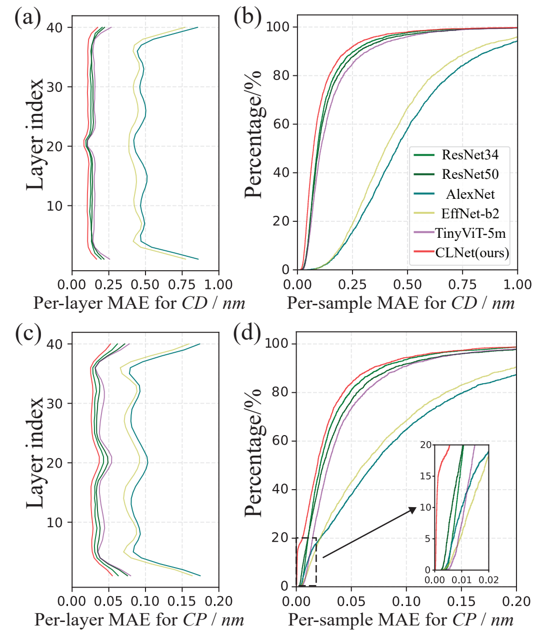

Fig. 6. (a) and (c) are the per-layer MAE for *CD* and *CP* in the test set, respectively. (b) and (d) are cumulative error distributions of per-sample MAE for *CD* and *CP*, respectively. Per-layer MAE represents the mean absolute errors for each layer of all samples, and per-sample MAE represents the mean absolute errors for individual samples.

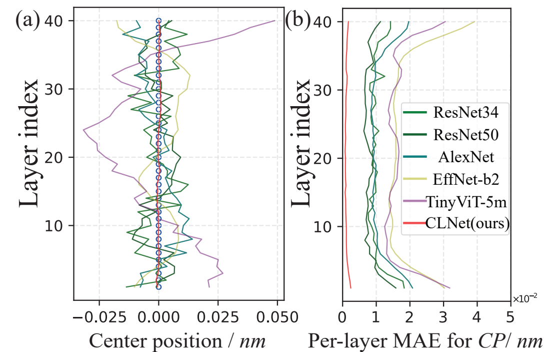

Fig. 7. (a) and (b) are the reconstructed *CP* and the corresponding per-layer MAE, respectively, under the ideal situation where ground truth *CP*s of each layer are equal to zero.

for ***CD*** and ***CP*** are shown in Fig. 6(a) and 6(c), respectively. Our method has the minimum error at each layer. The fitting results of CLNet for the edge layers demonstrate significant improvement compared to AlexNet and EffNet-b2. However, this phenomenon is insignificant for ResNet34, ResNet50, and TinyViT-5m. Figure 6(b) and 6(d) illustrate the cumulative error distribution of the per-sample MAE for ***CD*** and ***CP***, respectively. Compared with other models, the cumulative curves of CLNet reach 100 % most quickly, indicating that our method has achieved high accuracy in reconstructing structural parameters. The CLNet method has the overall MAE for ***CD*** of 0.110 *nm*,

Table 1. Evaluation comparison of six types of methods.

Method	Overall MAE / *nm*		Max per-sample MAE / *nm*		R^2 / (arb. units)		Slope / (arb. units)	
	CD	*CP*	*CD*	*CP*	*CD*	*CP*	*CD*	*CP*
CLNet (our method)	**0.110**	**0.031**	**1.982**	**0.631**	**0.9995**	**0.9953**	0.9993	**0.9942**
ResNet34	0.132	0.036	2.396	0.764	0.9993	0.9942	0.9978	0.9920
ResNet50	0.142	0.041	2.967	0.645	0.9992	0.9926	0.9995	0.9867
AlexNet	0.510	0.098	3.871	1.358	0.9933	0.9644	0.9934	0.9558
EffNet-b2	0.456	0.087	2.606	1.065	0.9948	0.9716	0.9968	0.9625
TinyViT-5m	0.156	0.045	2.264	0.742	0.9991	0.9921	**0.9997**	0.9897

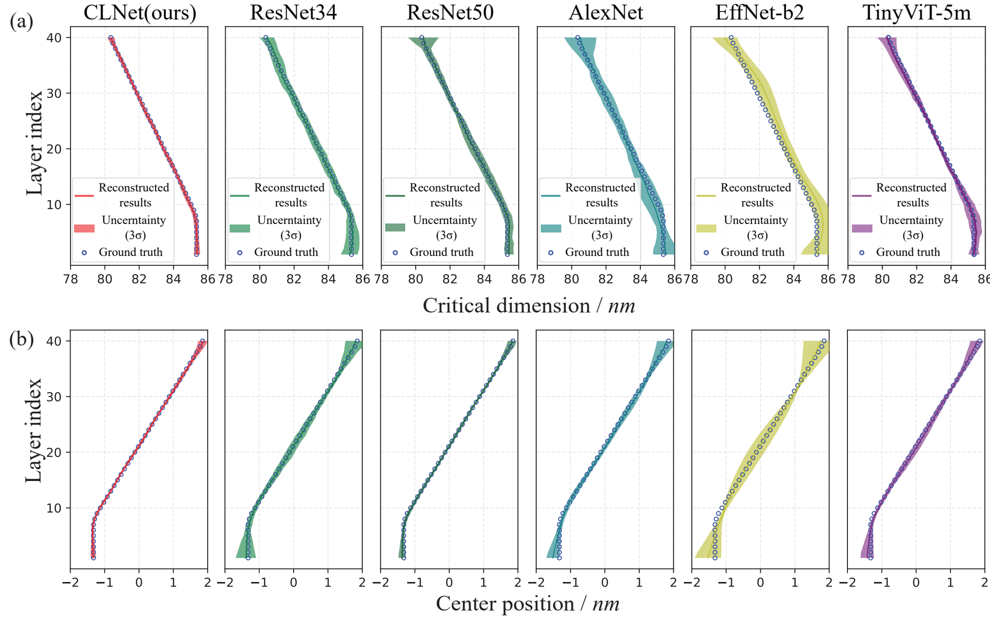

Fig. 8. (a) and (b) are model uncertainty quantification for *CD* and *CP*, respectively, based on six types of methods, including CLNet (our method), ResNet34, ResNet50, AlexNet, EffNet-b2, and TinyVit-5m. Our method performs best in terms of reconstruction accuracy and repeatability.

while the ResNet34, ResNet50, AlexNet, EffNet-b2, and TinyViT-5m have those of 0.132 *nm*, 0.142 *nm*, 0.510 *nm*, 0.456 *nm*, and 0.156 *nm*, respectively. The overall MAE for *CP* of our method is 0.031 *nm*, while the values of ResNet34, ResNet50, AlexNet, EffNet-b2, and TinyViT-5m are 0.036 *nm*, 0.041 *nm*, 0.098 *nm*, 0.087 *nm*, and 0.045 *nm*, respectively. The results indicate our method achieves a 20.0 % improvement in accuracy for reconstructing *CD* and a 13.9 % improvement for reconstructing *CP* than the second-highest achieved by ResNet34. The maximum per-sample MAE of CLNet is 1.981 *nm* for *CD* and 0.631 *nm* for *CP*, whereas ResNet-34 and TinyViT-5m yield the second-highest results of 2.264 *nm* for *CD*, and 0.764 *nm* for *CP*, respectively.

It should be noted that in Fig. 6(d), the cumulative error distribution curve rises quickly within 0.01 *nm*. Under the ideal sample situation where ground truth *CP*s of each layer are equal to zero, the CLNet's reconstructed results for *CP* are almost perfect, which is shown in Fig. 7. In other words, our method fully utilizes the LSTM mechanism to learn the prior knowledge about this kind of inter-layer correlation.

R^2 and slope are widely used to evaluate the fitting performance of methods. R^2 for *CD* of our proposed method is 0.9995, while those of the compared models are all less than 0.9993. For *CP*, R^2 of CLNet is 0.9953, while those of other methods are less than 0.9942. It indicates that our method has superior fitting ability. The slopes of CLNet are 0.9993 for *CD* and 0.9942 for *CP*, respectively. In addition to CLNet, the second-highest achieved results for *CD* and *CP* are 0.9997 (TinyViT-5m) and 0.9920 (ResNet34). The slope achieved by TinyVit-5m for *CD* is closer to 1 than our method, but the slope for *CP* achieved by TinyVit-5m is only 0.9897. The detailed evaluation results are shown in the Table 1. Notably, our method still demonstrates good fitting ability at the background level of 10^{-5}, with both the R^2 and the slope approaching 1.

Uncertainty quantification holds significant importance as it enables the estimation of method repeatability and facilitates comparability across diverse approaches.[45,46] In order to quantify the model uncertainty of our methods,

we sample the posterior distribution using stochastic gradient Langevin dynamics (SGLD)[47] and perform Bayesian inference averaging on the obtained samples. Typically, uncertainty is represented by calculating the three times standard deviation (3σ).[48] Figure 8 illustrates the uncertainty(3σ) quantifications for *CD* and *CP* of one randomly selected sample, using our method and the five methods of comparison. Besides accurately obtaining structural parameters, our method achieves confidence intervals better than the other methods. To further understand and evaluate the uncertainty of our method, we calculated the mean 3σ for the overall test set. The CLNet method has the mean 3σ for *CD* of 0.322 *nm*, while the ResNet34, ResNet50, AlexNet, EffNet-b2, and TinyViT-5m have those of 0.527 *nm*, 0.570 *nm*, 0.673 *nm*, 0.795 *nm*, and 0.424 *nm*, respectively. The mean 3σ for *CP* of our method is 0.116 *nm*, while the values of ResNet34, ResNet50, AlexNet, EffNet-b2, and TinyViT-5m are 0.204 *nm*, 0.155 *nm*, 0.216 *nm*, 0.231 *nm*, and 0.189 *nm*, respectively. The CLNet reduces the uncertainty for *CD* and *CP* by 24.1 % and 25.2 %, respectively. The results demonstrate that our method not only performs superior in terms of accuracy but also enhances the measurement repeatability by nearly 1.4 times.

4. Conclusion

In this work, we propose a correlation learning-based method for reconstructing complex profile nanostructures in the field of SAXS measurement. Considering the interlayer correlation of profile structural parameters and the nature of SAXS signals, we have designed and evaluated the CLNet framework which incorporates an LSTM mechanism. The CLNet actively learns inter-layer correlations of samples, significantly enhancing the accuracy and uncertainty of profile reconstruction. Notably, our novel approach improves reconstruction accuracy by at least 13.9 % and also exhibits a measurement repeatability that is nearly 1.4 times superior to the best among five other learning-based methods. In the future, we hope to extend our method to a broader range of nanostructures, such as 3D hole nanostructures. Additionally, this proposed approach may have potential applications in other profile metrology techniques, such as OCD, which need further exploration.

This work was supported by Zhangjiang Laboratory and the NSFC (Grant No. 11627811).

References

1. D. F. Sunday, S. List, J. S. Chawla, and R. Joseph Kline, "Evaluation of the effect of data quality on the profile uncertainty of critical dimension small angle x-ray scattering," Journal of Micro/Nanolithography, MEMS, and MOEMS **15**, 014001–014001 (2016).
2. Z. Jiang and B. Lee, "Recent advances in small angle x-ray scattering for superlattice study," Applied Physics Reviews **8** (2021).
3. D. F. Sunday, J. Ren, C. D. Liman, *et al.*, "Characterizing patterned block copolymer thin films with soft x-rays," ACS applied materials & interfaces **9**, 31325–31334 (2017).
4. B. A. Collins and E. Gann, "Resonant soft x-ray scattering in polymer science," Journal of Polymer Science **60**, 1199–1243 (2022).
5. H. He, C. Liu, and H. Liu, "Model reconstruction from small-angle x-ray scattering data using deep learning methods," Iscience **23** (2020).
6. A. F. Herrero, M. Pflüger, J. Probst, *et al.*, "Applicability of the debye-waller damping factor for the determination of the line-edge roughness of lamellar gratings," Optics Express **27**, 32490–32507 (2019).
7. M. Wormington, A. Ginsburg, I. Reichental, *et al.*, "X-ray critical dimension metrology solution for high aspect ratio semiconductor structures," in *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV,* vol. 11611 (SPIE, 2021), pp. 150–159.
8. J. Zhang, X. Chen, T. Yang, and S. Liu, "X-ray-based overlay metrology using reciprocal space slicing analysis," Optics Letters **48**, 6380–6383 (2023).
9. D. Calvetti and E. Somersalo, "Inverse problems: From regularization to bayesian inference," Wiley Interdisciplinary Reviews: Computational Statistics **10**, e1427 (2018).
10. A. F. Hannon, D. F. Sunday, D. Windover, and R. Joseph Kline, "Advancing x-ray scattering metrology using inverse genetic algorithms," Journal of Micro/Nanolithography, MEMS, and MOEMS **15**, 034001–034001 (2016).
11. M. Pflüger, R. J. Kline, A. F. Herrero, *et al.*, "Extracting dimensional parameters of gratings produced with self-aligned multiple patterning using grazing-incidence small-angle x-ray scattering," Journal of Micro/Nanolithography, MEMS, and MOEMS **19**, 014001–014001 (2020).
12. W.-l. Wu, R. Joseph Kline, R. L. Jones, *et al.*, "Review of the key milestones in the development of critical dimension small angle x-ray scattering at national institute of standards and technology," Journal of Micro/Nanopatterning, Materials, and Metrology **22**, 031206–031206 (2023).
13. J. Reche, Y. Blancquaert, G. Freychet, *et al.*, "Dimensional control of line gratings by small angle x-ray scattering: Shape and roughness extraction," in *2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC),* (IEEE, 2020), pp. 1–6.
14. R. Rohrich, G. Oliveri, S. Kovaios, *et al.*, "Uncertainty estimation and design optimization of 2d diffraction-based overlay metrology targets," ACS Photonics **7**, 2765–2777 (2020).
15. X. You, J. Liu, Y. Li, *et al.*, "3d microscopy in industrial measurements," Journal of Microscopy **289**, 137–156 (2023).
16. D. F. Sunday, S. List, J. S. Chawla, and R. J. Kline, "Determining the shape and periodicity of nanostructures using small-angle x-ray scattering," Journal of Applied Crystallography **48**, 1355–1363 (2015).
17. M. Pflüger, R. J. Kline, A. F. Herrero, *et al.*, "Extracting dimensional parameters of gratings produced with self-aligned multiple patterning using grazing-incidence small-angle x-ray scattering," Journal of Micro/Nanolithography, MEMS, and MOEMS **19**, 014001–014001 (2020).
18. R. Ciesielski, L. M. Lohr, H. Mertens, *et al.*, "Pushing the boundaries of euv scatterometry: reconstruction of complex nanostructures for next-generation transistor technology," in *Metrology, Inspection, and Process Control XXXVII,* vol. 12496 (SPIE, 2023), pp. 447–455.

19. A. F. Herrero, M. Pflüger, J. Puls, *et al.*, "Uncertainties in the reconstruction of nanostructures in euv scatterometry and grazing incidence small-angle x-ray scattering," Optics Express **29**, 35580–35591 (2021).
20. Z. Jiang, J. Wang, M. V. Tirrell, *et al.*, "Parameter estimation for x-ray scattering analysis with hamiltonian markov chain monte carlo," Journal of Synchrotron Radiation **29**, 721–731 (2022).
21. C. Rendón-Barraza, E. A. Chan, G. Yuan, *et al.*, "Deeply sub-wavelength non-contact optical metrology of sub-wavelength objects," Apl Photonics **6** (2021).
22. S. Liu, T. Yang, J. Zhang, *et al.*, "X-ray scatterometry using deep learning," in *Tenth International Symposium on Precision Mechanical Measurements,* vol. 12059 (SPIE, 2021), pp. 481–487.
23. T. Yang, X. Chen, S. Liu, *et al.*, "Condition-number-based measurement configuration optimization for nanostructure reconstruction by optical scatterometry," Measurement Science and Technology **34**, 125001 (2023).
24. B. Dey, D. Cerbu, K. Khalil, *et al.*, "Unsupervised machine learning based cd-sem image segregator for opc and process window estimation," in *Design-Process-Technology Co-optimization for Manufacturability XIV,* vol. 11328 (SPIE, 2020), pp. 317–327.
25. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature **521**, 436–444 (2015).
26. A. Baranovskiy, I. Grinberg, M. G. Greene, *et al.*, "Deep learning for the analysis of x-ray scattering data from high aspect ratio structures," in *Metrology, Inspection, and Process Control XXXVII,* vol. 12496 (SPIE, 2023), pp. 837–847.
27. D. Wang, H. Liang, H. Yang, and H. Yu, "Complex profile metrology via physical symmetry enhanced small angle x-ray scattering," Journal of Applied Physics **135** (2024).
28. S. Liu, X. Chen, T. Yang, *et al.*, "Inverse optical scatterometry using sketch-guided deep learning," Optics Express **32**, 20303–20315 (2024).
29. S. S. Ginosar, *Modeling Visual Minutiae: Gestures, Styles, and Temporal Patterns* (University of California, Berkeley, 2020).
30. J. Zhang, T. Lan, Y. Gao, *et al.*, "Cdsaxs study of 3d nand channel hole etch pattern edge effects and etched hole pattern variance," in *Metrology, Inspection, and Process Control XXXVIII,* vol. 12955 (SPIE, 2024), pp. 813–821.
31. O. Sorkhabi, J. Zhang, D. Hu, *et al.*, "Flash memory hole etch profile monitoring using x-ray and optical scatterometry," Journal of Micro/Nanopatterning, Materials, and Metrology **22**, 031207–031207 (2023).
32. P. Gin, M. Wormington, Y. Amasay, *et al.*, "Inline metrology of high aspect ratio hole tilt and center line shift using small-angle x-ray scattering," Journal of Micro/Nanopatterning, Materials, and Metrology **22**, 031205–031205 (2023).
33. Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," Neural computation **31**, 1235–1270 (2019).
34. W. Sun, Y. Goto, T. Yamamoto, and K. Hitomi, "Deep learning model for 3d profiling of high-aspect-ratio features using high-voltage cd-sem," Japanese Journal of Applied Physics **61**, SD1036 (2022).
35. C. Petti, "3d memory: etch is the new litho," in *Advanced Etch Technology for Nanopatterning VII,* vol. 10589 (SPIE, 2018), p. 1058904.
36. X. You, H. Chang, Q. Zhang, *et al.*, "Low-complexity characterized-long-short-term-memory-aided channel modeling for optical fiber communications," Applied Optics **62**, 8543–8551 (2023).
37. S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of lstm and bilstm in forecasting time series," in *2019 IEEE International conference on big data (Big Data),* (IEEE, 2019), pp. 3285–3292.
38. S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" Advances in neural information processing systems **31** (2018).
39. J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," arXiv preprint arXiv:1607.06450 (2016).
40. C. Settens, B. Bunday, B. Thiel, *et al.*, "Critical dimension small angle x-ray scattering measurements of finfet and 3d memory structures," in *Metrology, Inspection, and Process Control for Microlithography XXVII,* vol. 8681 (SPIE, 2013), pp. 200–208.
41. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition,* (2016), pp. 770–778.
42. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems **25** (2012).
43. M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning,* (PMLR, 2019), pp. 6105–6114.
44. K. Wu, J. Zhang, H. Peng, *et al.*, "Tinyvit: Fast pretraining distillation for small vision transformers," in *European conference on computer vision,* (Springer, 2022), pp. 68–85.
45. B. Bunday and G. Orji, "Metrology," in *2021 IEEE International Roadmap for Devices and Systems Outbriefs,* (2021), pp. 01–68.
46. T. Yang, X. Chen, S. Liu, *et al.*, "Bootstrap method for uncertainty evaluation in critical dimension small-angle x-ray scattering," IEEE Transactions on Instrumentation and Measurement (2024).
47. C. Li, C. Chen, D. Carlson, and L. Carin, "Preconditioned stochastic gradient langevin dynamics for deep neural networks," in *Proceedings of the AAAI conference on artificial intelligence,* vol. 30 (2016).
48. H. Yang, Z. Wu, K. Zhang, *et al.*, "Uncertainty quantification on small angle x-ray scattering measurement using bayesian deep learning," Journal of Applied Physics **136** (2024).