Microelectronic Engineering 216 (2019) 111051

Contents lists available at ScienceDirect

Microelectronic Engineering

journal homepage: www.elsevier.com/locate/mee

Research paper

Deep learning denoising of SEM images towards noise-reduced LER measurements

E. Giannatou[a,c,*], G. Papavieros[b,c], V. Constantoudis[b,c], H. Papageorgiou[a], E. Gogolides[b,c]

[a] Institute for Language and Speech Processing (ILSP), Athena R.C., Maroussi, Greece
[b] Institute of Nanoscience and Nanotechnology (INN), N.C.S.R. Demokritos, Agia Paraskevi 15341, Greece
[c] Nanometrisis p.c, Greece

ARTICLE INFO

Keywords:
Neural networks
Deep learning
Line edge roughness
Noise effects
SEM images

ABSTRACT

As chip sizes decrease and node dimensions break the sub-10 nm barrier, Line Edge Roughness (LER) metrology becomes a critical issue for the semiconductor research and industry. Scanning Electron Microscopy (SEM) imaging being the widely used tool for LER metrology suffers from the presence of noise that degrades measurement accuracy. To solve this issue without damaging the measured pattern, the applicability of deep Convolutional Neural Networks (CNNs) is explored, tackling the problem at the image level. The SEM image Denoising model (SEMD) is trained on synthesized image data to detect the variations of noise and provides as an output a denoised image of the pattern. The results are presented and compared with the state-of-art predictions, showing the effectiveness and enhanced performance of the SEMD method.

1. Introduction

Chip sizes rescale transistors to sub-10 nm size and a major issue that arises, following this rescale, is the impact of stochastic phenomena in photon irradiation and material composition. These phenomena cause random deviations in the shape and size of lithographic patterns. The most popular metric to measure these deviations is the Line Edge Roughness (LER), i.e., the sidewall roughness of lithographic line patterns as it is depicted with bright line edges on top-down Scanning Electron Microscopy (SEM) images [2]. According to [3], LER is fully described by three parameters: (i) root-mean-square (σ_{LER} or rms) deviation, (ii) correlation length (ξ or ksi), and (iii) roughness exponent (α or alpha). Rms quantifies the standard deviation of the bright line edge pixels from their best linear fit. Correlation length ksi describes how closely correlated are the edge points, i.e. the range of correlations of edge points along edge direction. The roughness exponent alpha quantifies the relative contribution of high frequencies on the total LER value. The rms^2 value equals to the area under the Power Spectral Density (PSD) curve, ksi is defined as the distance r at which the normalized autocorrelation function is 1/e and alpha can be extracted by exponent of the power law drop of the PSD at high frequencies. LER leads to local variations of line widths, which is usually called Line Width Roughness (LWR). When the rms of LWR (i.e. σ_{LWR}) is beyond a certain point, circuit performance degrades and production yield drops [4]. According to ITRS 2013, accepted $3\sigma_{LWR}$ values should be lower than the 8% of the Critical Dimension (CD) (i.e. width) of the lines. This threshold for sub-10 nm nodes is located at sub-nm levels, which makes LER metrology a difficult endeavor. One of the main reasons is the presence of noise in SEM images that distorts high level information and undermines the accuracy of LER measurement. Noise affects the accurate detection of line edges (Fig. 1), increasing the measured edge roughness of the structures [9].

Because of the physics in SEM imaging, microscope users struggle as they find themselves constrained between two cliff edges. On one hand, they can choose to use a large number of frames per scan, thus reducing the presence of noise in the acquired image (blue curve in Fig. 2) but at the same time increasing sample damage caused by the high energy electron beam (red curve in Fig. 2). On the other hand, they have the option to reduce the number of frames. However, though the damage is partially avoided, noise starts to dominate and distort the edges of image pattern as we can see in the detected edges of the images shown in Fig. 1b.

To overcome this blurring net, several computational methods for noise free LER metrology have been proposed. Some implement image filtering techniques that treat the problem on an image level or use physical models that simulate SEM operation and image acquisition [5–9]. State of the art techniques do not apply image filters nor physical modeling, but rather process the acquired noisy edge data using additional information gained from analysis, based on advanced metrics, such as the PSD or the Height to Height Correlation Function (HHCF) [10,11].

* Corresponding author at: Nanometrisis p.c, Greece.
E-mail address: e.giannatou@nanometrisis.com (E. Giannatou).

https://doi.org/10.1016/j.mee.2019.111051
Received 30 November 2018; Received in revised form 10 June 2019; Accepted 24 June 2019
Available online 27 June 2019
0167-9317/

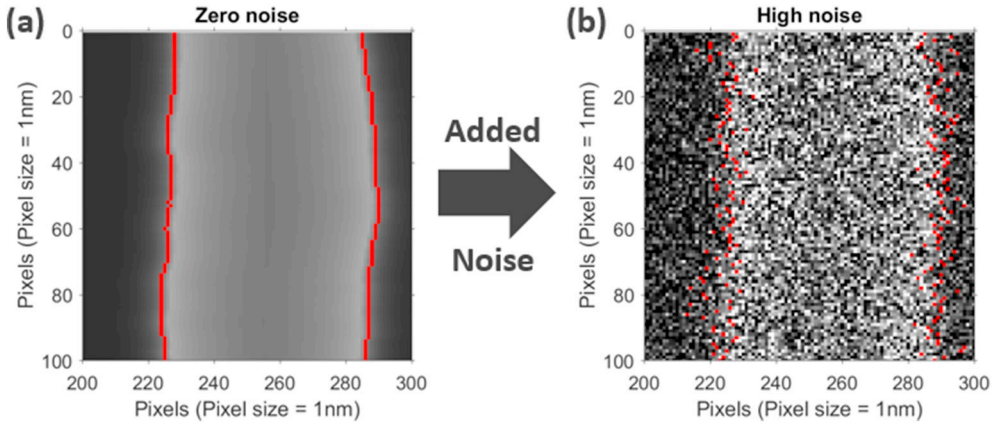

Fig. 1. a) Edge detection in Zero Noise (ZN) synthesized image b) Edge detection in High Noise (HN) synthesized image, consisting of ZN after noise degradation.

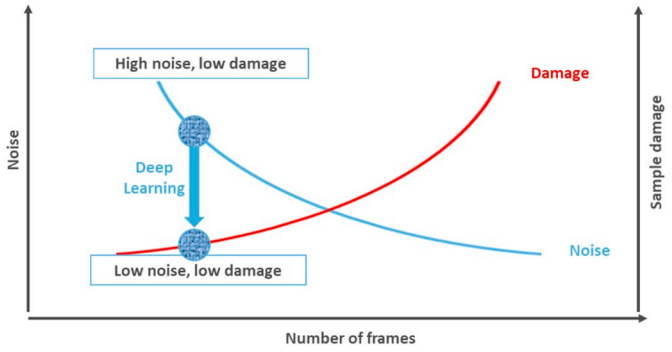

Fig. 2. The effect of SEM electron beam energy on noise distortion of image and pattern damage.

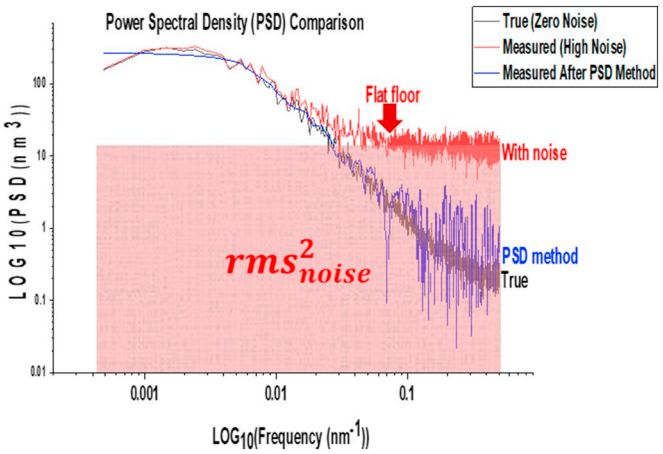

Fig. 3. PSD curves estimated from the noisy image directly (red line) and after the application of the PSD-based method (blue line). The ground truth for the PSD curve is shown with the black line and is calculated from the zero noise image. This analysis was performed on a synthesized image of size 2048nm × 2048nm depicting 10 lines with LER parameter values rms, ksi and alpha equal to 2.4 nm, 25.5 nm and 0.5 accordingly. The noisy image is degraded by additive Gaussian and Poisson noise, $Q_g = 2$ and $Q_p = 5$ respectively (see Section 2.1). According to [11] $rms_{measured}^2 = rms_{true}^2 + rms_{noise}^2$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The state of the art Power Spectral Density (PSD) method for unbiased (noise free) LER measurements [10,11], reduces the noise contamination effects by applying Fourier analysis on the detected edges of the patterns. Fourier transform assumes that the frequency components of the signal lie predominantly at low frequencies and those of the noise dominate predominantly at high frequencies. The removal of noise is achieved by subtracting the observed PSD flat floor due to noise domination at high frequencies (Fig. 3). By employing the knowledge that the area under the PSD curve is the square of the rms (Parseval's theorem), one can calculate the rms_{noise}^2 that corresponds to the noise and subtract it from the measured-biased $rms_{measured}^2$. In the same practice for the HHCF, noise dominates at short distances, therefore one can calculate the unbiased HHCF by subtracting the flat floor and use it to calculate the unbiased LER parameters.

The above state of the art method, though successful in many cases, may contain flaws and meet limitations as it cannot overcome the following issues:

1. The noise domination at high frequencies makes it hard for someone to tell whether the removed volume doesn't contain real high frequency signal.
2. The method cannot deal with cases were no clear flat floor is observed.

In order to overcome the above issues, we propose an orthogonal approach in LER metrology based on the use of deep learning techniques in an effort to denoise the image before the edge detection algorithm is applied. Deep learning (DL) is an emerging approach for directly mapping low quality images to desirable high-quality images. It has achieved impressive results to many practical problems, where the goal is to use big data to train a model. Deep Learning Neural Network architectures (DLNN) are robust, scalable and generalizable. They automatically learn the random variations of real-life data while their performance improves with more training data. Furthermore, the same network can be used for many different applications and data types. Even though DLNNs are computationally complex, parallelization and efficient contemporary hardware permits taking advantage of their statistical strength for multi-task learning. However, the optimal choice of model architecture and parameters, depends on the task at hand and remains a challenge for achieving further improvements. Convolutional neural networks (CNNs) are similar in architecture to conventional DLNNs but they explicitly assume that the input is an image.

In this paper, we explore the applicability of a DL-based method, in an effort to overcome the need for a large number of frames (i.e. high-energy electron beam) in order to reduce the presence of noise in SEM images. We take advantage of recent DL success in image denoising, thus being able to develop a post processing step for denoising SEM images while preserving the structural information of the specimen unaltered.

The structure of the paper is as follows: The next section describes the DL method including the training data and the model structure and training. The evaluation and results of the DL method are presented in Section 3 along with comparison with state-of-art predictions. The

paper closes with the main conclusions and future steps in the last Section 4.

2. Deep learning denoising

CNN models aim to learn the underlying relationship between noisy and noise free images from a training set of degraded and ground-truth image pairs. These data can be real SEM images, physical modeling and simulation of SEM image acquisition and synthesized computer-generated SEM images. We choose to work with synthesized SEM images, because it is: a) less time consuming to generate large amount of data and b) there is total control or roughness parameters and noise level.

2.1. Training data and processing

The methodology of producing top down grayscale (single channel) SEM images of predefined roughness parameters and controlled noise magnitude consists of 3 steps [12,13]. In the first step, the line edge skeleton of the image is constructed. This step is implemented with the assistance of an inverse Fourier technique that uses as input the following autocorrelation model function [12]:

$$C(r) = rms^2 e^{-\left(\frac{r}{\xi}\right)^{2\alpha}} \quad (1)$$

where rms, ξ and α are the LER parameters.

The next step involves the building process of the image where an exponentially decaying pixel intensity profile is applied around the edges, following the method described in [13]. The output of this process is a synthesized image that corresponds to an ideal experimental noise free SEM image. In the third step, noise is added to the output of step 2 considering both noise types met in SEM images (Gaussian and Poisson) [12]. Poisson noise, originates from the varying number of electrons that hit the specimen at each measurement spot (shot noise) while Gaussian noise, is the result of microscope electronics (electronics noise).

Noise is added to synthesized images using the following formula:

$$C_n(i,j) = C_s(i,j) + (Q_g + Q_p\sqrt{C_s})R(i,j) \quad (2)$$

where $C_s(i,j)$ is the synthesized image intensity of the pixel (i,j), $C_n(i,j)$ is the final intensity of the pixel (i,j) after the addition of noise, Q_g, Q_p are the coefficients of the Gaussian and Poisson noise magnitudes respectively and $R(i,j)$ is a random number with Gaussian distribution. The Poisson noise is simulated as an intensity-dependent Gaussian and the skewness of the distribution is neglected. For the purpose of this paper, zero mean Gaussian noise with standard deviation $Q_g = 2$ and Poisson noise with mean $Q_p = 5$ has been added. Following noise addition, image pixels are scaled to the range [0,255].

Examples of images generated by the above method can be seen in Fig. 4.

The dataset used for training the CNN model consists of 1000 pairs of high noise-zero noise synthesized SEM images of size 2048 × 2048. Each synthesized image contains on average 10 lines. During the evaluation of trained models, we do not focus on the fitting success of the training data but on how well the model will perform on unseen data after training. Therefore, before training the model we split the dataset into two sets, the training and the test set. We keep 920 image pairs for training set and the remaining 80 image pairs for test set. The training set is used for training models with different architectures and hyper parameters and then the test set is used to evaluate the performance of the trained models. The test set provides us with an estimate of how well the model performs on unseen data. Finally, the hyper parameters that performed best on the test data are selected for training the final deep learning denoising model.

During the preprocessing step we split the training set images of size 2048 × 2048 into smaller patches of size 200 × 200. For each train set image we slide a 200 × 200 window by 100 pixels at a time and we split each image into 361 overlapping patches. Thus, we have 920 high noise images of 361 patches each which lead to a total of 332.120 patches. We repeat the same process for the 920 zero noise images.

One training epoch is completed when the entire dataset completes one forward and backward pass through the convolutional neural network. Since, one epoch is too big to be fed to the computer at once it is a common practice to divide the dataset into smaller batches. Therefore, we group the patches into 10.379 mini-batches of size 32. Using mini batches for training gives a significant computational speed up and it can be a solution for running out of memory issues.

2.2. Model training

The goal of image denoising is to recover a clean image x from a noisy observation y which follows an image degradation model y = x + v, where v represents additive noise [14].

Convolutional neural networks can be used to estimate $\bar{x}$ from the noisy measurement y. They are being trained on a dataset containing noisy-clean image pairs. CNNs aim to learn the tunable parameters Θ which will minimize a given loss function. CNN models achieve state of the art performance and fast testing speeds [15].

Convolutional neural network layers transform the original image pixel values to the final predicted pixel values. The local connectivity property of convolutional layers gives CNNs the ability to understand visual data. Convolutional layers differing on the numeric values of the filter matrix, can detect different features from an image, for example edges, curves etc. High number of filters lead to higher number of extracted features which results to better performing models when it comes to recognizing patterns in unseen images. In the case of image denoising, filter values which consist of tunable weights and biases are not known in advance but they are learned during the training process in a way such that they minimize a given cost function. Weight initialization plays a decisive role in the loss minimization process of the model while it determines the minimum that the optimization method will aim to approach.

For the task of SEM image denoising, we trained models based on the DnCNN [1] architecture. For ease of reference we refer to the SEM denoising model as SEMD. SEMD is a residual learning convolutional neural network for SEM image denoising which can also be used for blind denoising images with unknown noise levels. Instead of directly predicting the zero noise image, it can be trained to estimate the noise v_i at each pixel of the high noise image. Then, the estimated zero noise image can be computed by subtracting the estimated residual mapping from the high noise image (see Fig. 5).

SEMD is trained on synthesized SEM images to learn the noise distribution through a predefined nonlinear function $F(y;\Theta)$, where Θ denotes trainable weights assigned to the pixel values of the (i,j) position. SEMD non-blind denoising model is trained on synthesized SEM images with specific Gaussian and Poisson noise levels, $Q_g = 2$ and $Q_p = 5$ accordingly. However, SEMD can also be trained for blind denoising images with varying noise levels.

The input block of SEMD consists of a convolutional layer mapping the single channel grayscale image, into n feature maps, followed by a Rectified Linear Unit (ReLU) activation function. Rectified Linear Unit ReLU(x) = max(x,0) [16], is commonly used in deep learning networks to introduce non-linearity. The convolutional layers of the middle blocks are followed by a Batch Normalization (BN) step and a ReLU activation function. BN [17] is used for keeping the activations in a specified Gaussian range. It improves gradient flow through the network and it allows faster learning rates. BN also reduces the strong dependence of initialization and acts as a form of regularization, balancing the weights from being too large or too small. Finally the output block consists of a convolution with one filter which is used to reconstruct the output. SEMD has a total of 149.185 tunable parameters.

The SEMD output R(y;Θ) can be expressed as

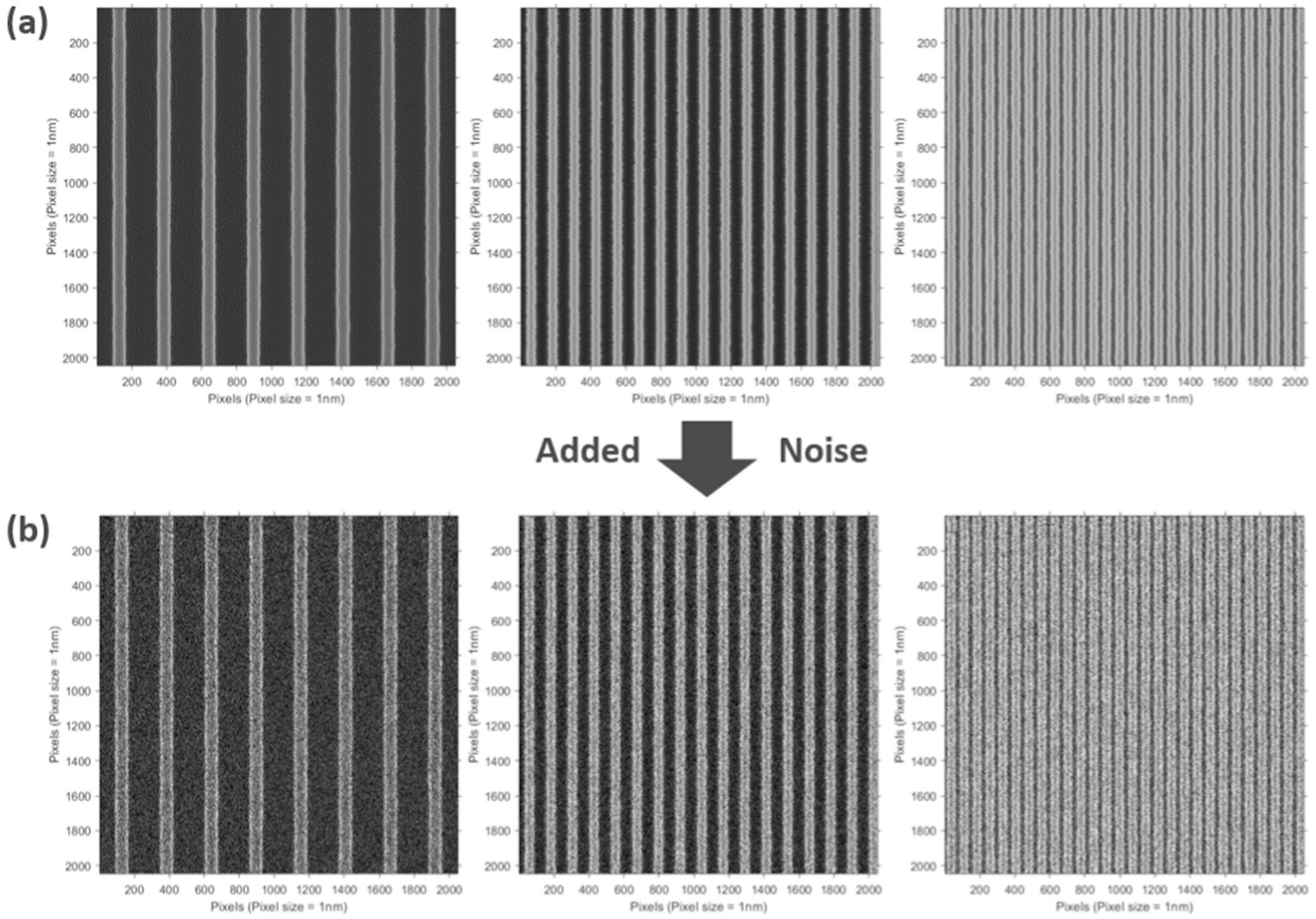

Fig. 4. Examples of a) Zero Noise (ZN) and b) High Noise (HN) synthesized top-down SEM images with various pattern characteristics demonstrating the potential of generation algorithm to produce any type of line pattern.

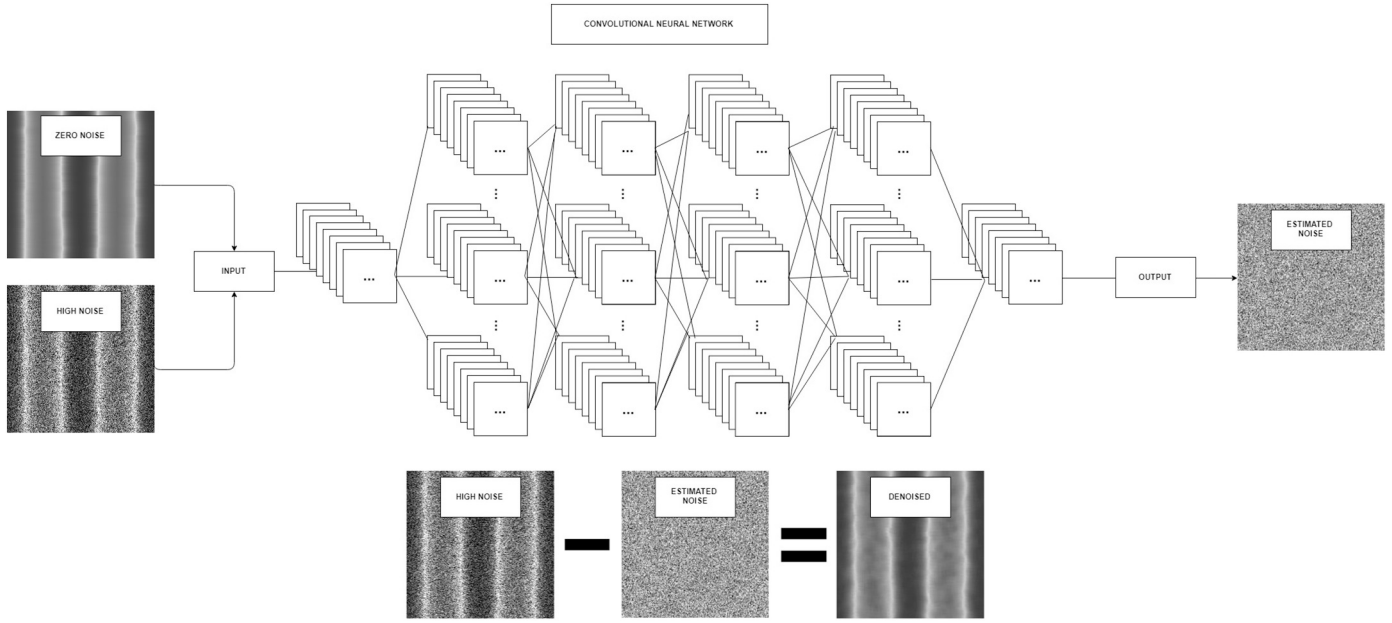

Fig. 5. DLNN model scheme for SEM image denoising (SEMD model).

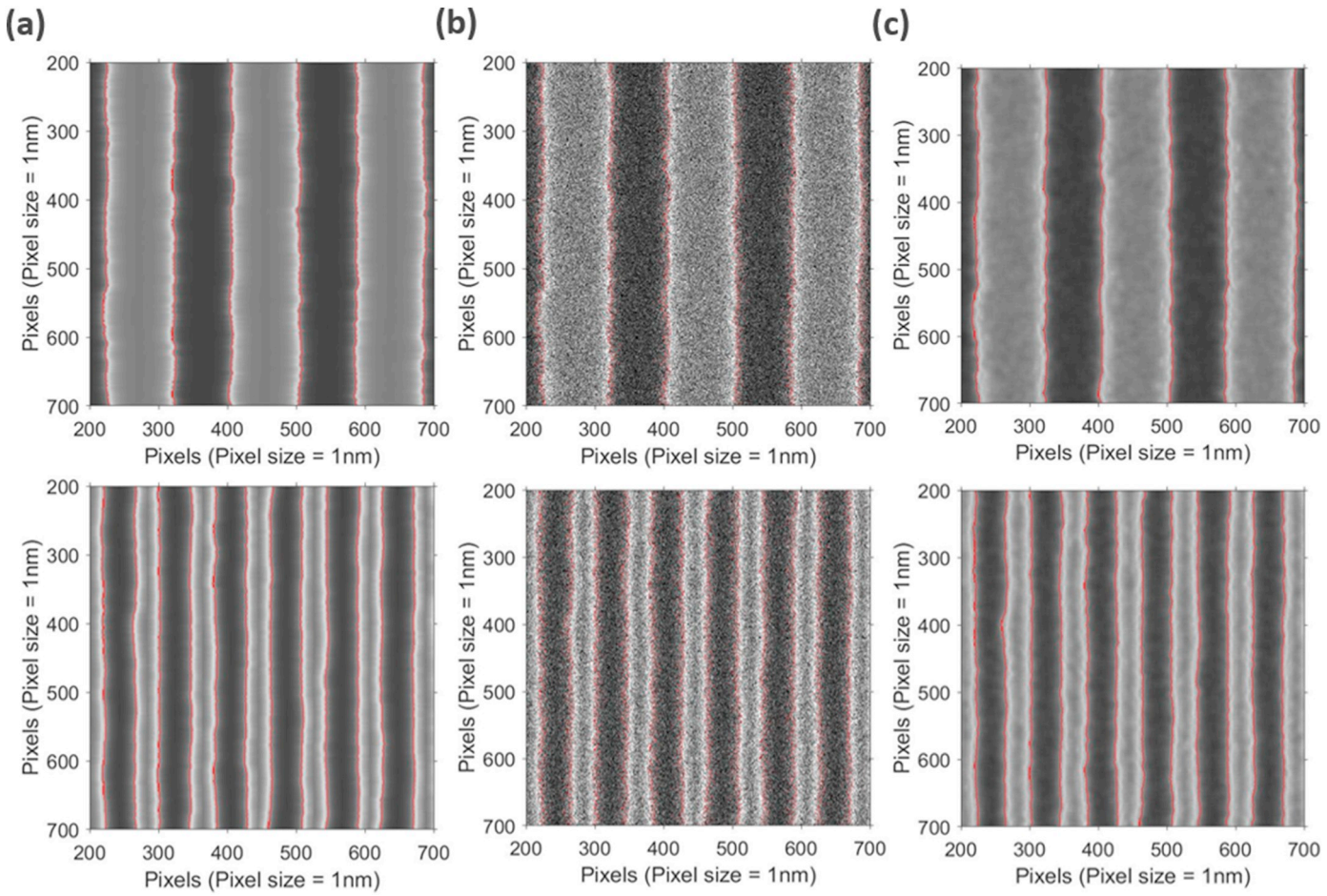

Fig. 6. Cropped image parts of size 500 × 500 pixels representing 2 examples of a) Zero Noise synthesized test SEM images, b) High Noise synthesized test SEM images and c) result of SEMD model when applied in images of column (b). Detected edges are represented with red lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\mathrm{R}(y;\Theta) = (W_n(...(RELU\,(BN_{\gamma_2\beta_2}(W_2 RELU\,(W_1 y + b_1))))...) + b_n) \tag{3}$$

where y is the input high noise image, $\mathrm{R}(y;\Theta)$ is the model output or estimated residual image, W_i represents the trainable weights of the i_{th} layer, b_i is the trainable bias term of the i_{th} convolution layer, RELU is a nonlinear function, BN represents batch normalization, and Θ is the set of all trainable parameters including W_i and b_i.

The learning process requires the definition of a loss function that quantifies the difference between the estimated residual image $R(y_k;\Theta) = y_k - \bar{x}_k$ and the actual residual image $\nu_k = y_k - x_k$. SEMD aims to find an optimal parameter set Θ to minimize a loss function L,

$$\sum_{k=1}^{K} L(\nu_k, R(y_k;\Theta)) \tag{4}$$

Where y_k and x_k denote the k_{th} input and output training image patch pair respectively [18].

For the task of denoising SEM images, L is defined as the Mean Squared Error (MSE) between the actual and the predicted residual mapping $\nu_k$and $R(y_k;\Theta)$ accordingly. MSE is calculated over a set of multiple input-output training patch pairs (y_k, x_k) and it is used to train the tunable parameters Θ.

The mean squared error between $R(y_k;\Theta)$ and ν_k can be is defined as

$$L(\Theta) = \frac{1}{2K}\sum_{k=1}^{K} \|R(y_k;\Theta) - (y_k - x_k)\|_F^2 \tag{5}$$

Here (y_k, x_k) represents K noisy-clean training image patch pairs and Θ represents the trainable parameters. $R(y_k;\Theta)$ represents the estimated residual mapping and $\nu_k = y_k - x_k$ represents the actual residual mapping.

Assuming that loss functions are differentiable, the minimization problem or network training in Eq. (5) can be solved by the back-propagation method [19]. Back-propagation is an algorithm used for calculating the gradients. The negative gradient is a vector of partial derivatives which point to the direction of the greatest decrease of the loss function [20]. Gradients are used for determining how a single training example would successfully change the weights and biases with respect to the most rapid decrease to the cost function. SEMD uses the Adaptive Moment Estimation (Adam) [21] optimization algorithm. Adam calculates a moving estimate of the first and second moment as a weighted sum of the gradients and the squared gradients accordingly. During the update step, it steps towards the minimum using both the first moment and the second moment. The step size or learning rate is a hyper parameter controlling how far the algorithm steps at each iteration towards the direction of the minimum.

It takes 33 h to train SEMD. After being trained, SEMD can denoise a synthesized SEM image of size 2048 × 2048 and 1024 × 1024 in approximately 0.43 and 0.11 s accordingly.

3. Results

Fig. 6 represents two examples of Zero Noise-High Noise image pairs and the reconstruction of the High Noise images after the application of the SEMD denoising model. Rms, ksi and alpha values (referred to as "Estimated") calculated from SEMD denoised images should be as close

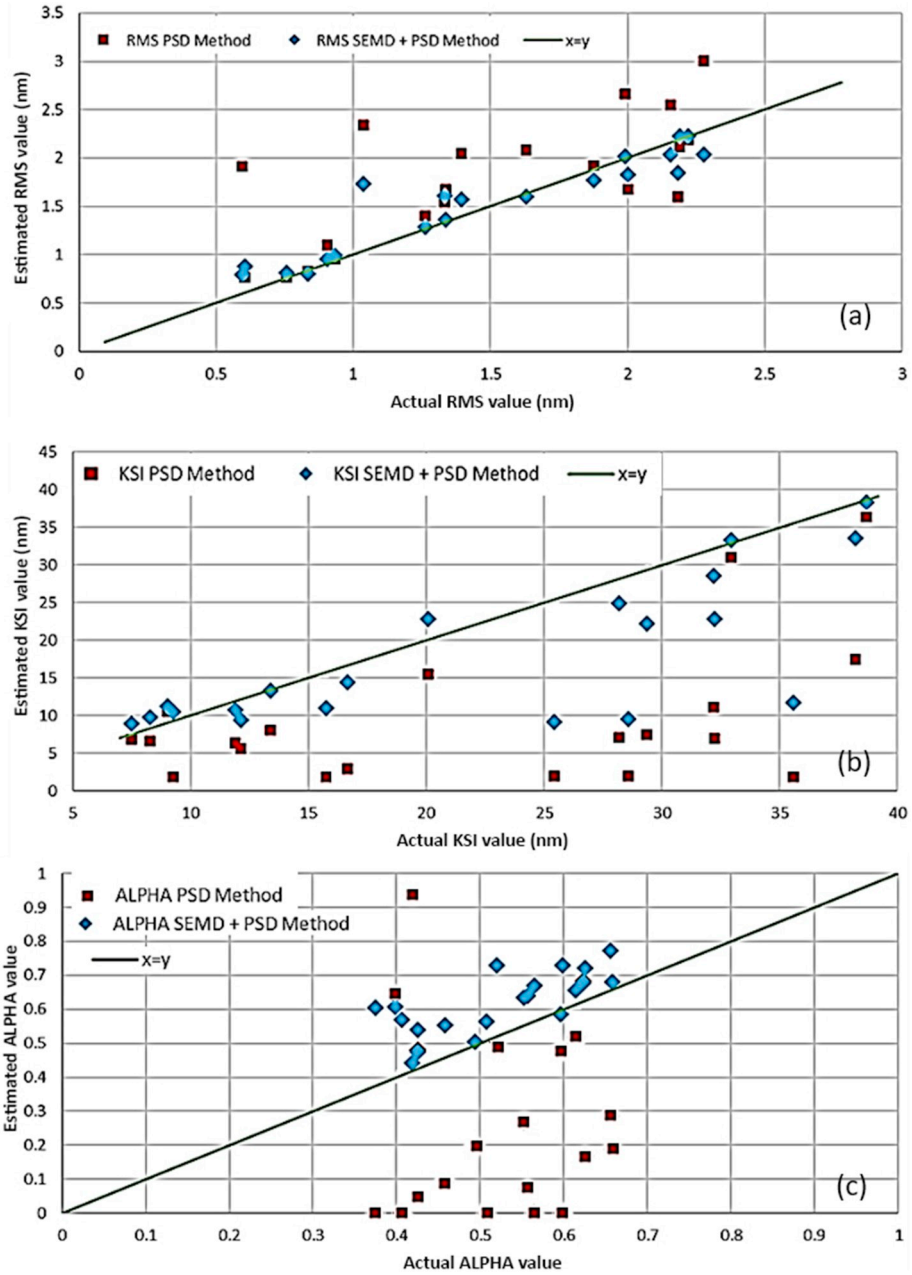

Fig. 7. Estimated LER parameters with PSD-based method (red squares) and with SEMD + PSD based method (blue diamonds) versus the actual mean value of parameters: a) Rms value, b) correlation length ksi and c) roughness exponent alpha. One can notice the improved accuracy when SEMD method is combined with PSD-based one. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

as possible to the rms, ksi and alpha values (referred to as "True") calculated from zero noise images. Therefore, we use rms, ksi and alpha metrics for the objective assessment of the SEMD model.

We test the trained SEMD model on 20 synthesized High Noise (HN) – Zero Noise (ZN) image pairs with known noise levels, producing 20 extra Denoised (DN) images. We calculate the biased rms of both HN, ZN and DN images. Then, we apply the current state of the art PSD [10,11] method for calculating the unbiased rms of HN and DN images. Thus, we have the rms values estimated straight from the high noise images (RMS_HN_{biased}), the rms estimated from the high noise images using the PSD method ($RMS\_HN_{unbiased}$), the rms estimated from the denoised images (RMS_DN_{biased}) and finally the rms estimated from the denoised images using the PSD method ($RMS\_DN_{unbiased}$). $RMS\_DN_{unbiased}$ is the result of a method combination between the trained SEMD model and the state of the art method described in Section 1. We repeat the exact same process for calculating the ksi and alpha values of both ZN, HN and DN images.

In Fig. 7, we compare the performance of the state of the art method (PSD-based method) applied on high noise images and on SEMD denoised images (SEMD + PSD method). Each point in Fig. 7 represents a single test image of size 2048nm × 2048nm consisting of an average of 10 lines. The results of the state of the art method (PSD method) are represented with red squares while the results of the combined method (SEMD + PSD) are represented with blue diamonds. The green diagonal lines represent the actual rms, ksi and alpha as calculated from the zero noise images. Thus, the closer the points are to the diagonal line the better. We observe that the state of the art method on average results to higher rms and lower ksi and alpha value estimations. However, the SEMD model can be used to significantly reduce the rms, ksi and alpha calculation error. Moreover, the PSD method fails in 5 images,

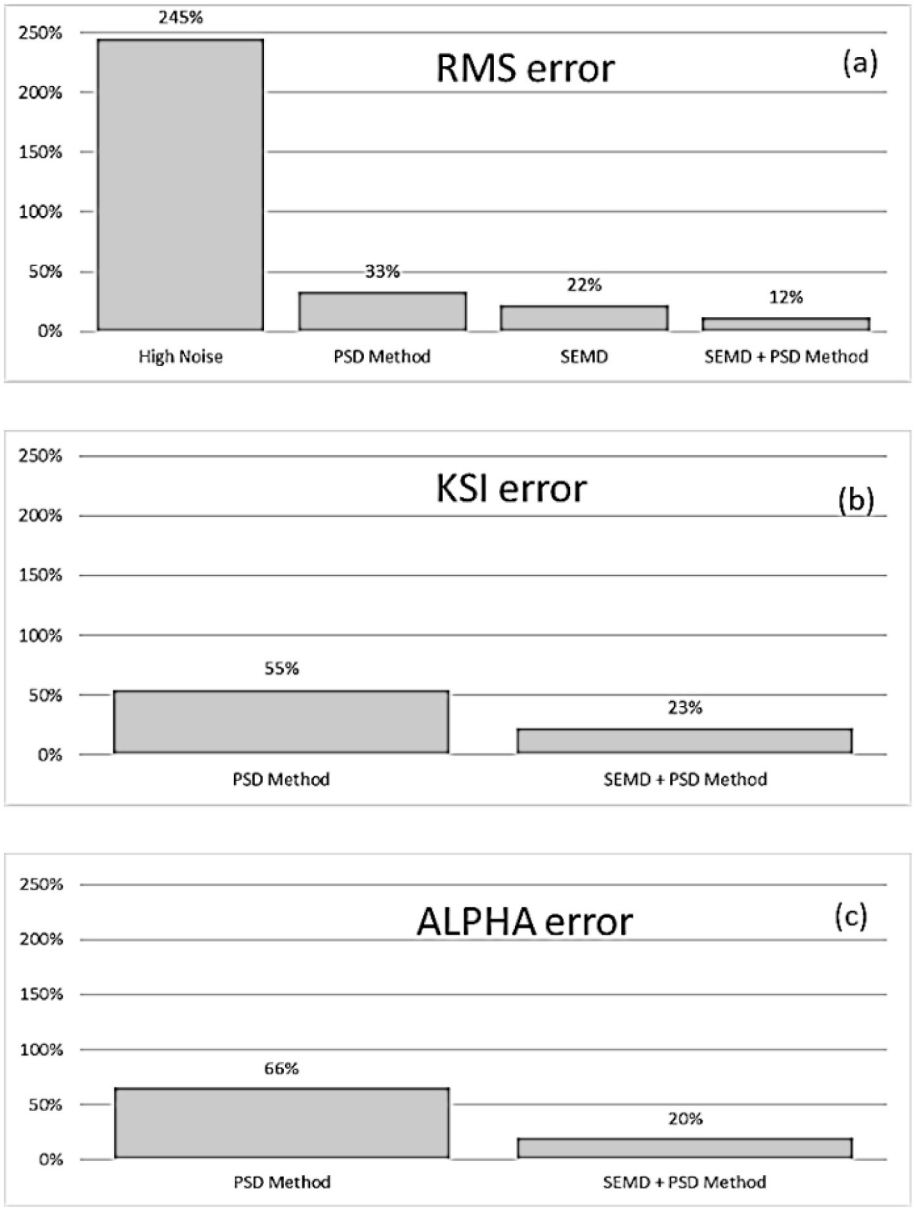

Fig. 8. Mean absolute LER parameter percentage errors of LER parameter predictions (RMS (a), KSI (b), APLHA (c)) with the PSD method, SEMD and their sequential combination SEMD + PSD method. In a) we also show the large error of RMS value when no noise reduced method is applied for the sake of comparison. In all cases, one can see the improved accuracy gained when SEMD method is involved to denoise the highly noisy SEM images.

calculating zero alpha values while in these cases the SEMD model leads to valid estimations.

We also calculate the percentage rms, ksi and alpha error for each test set image as

$$RMS_{error}\% = \left(\frac{|RMS_{estimated} - RMS_{actual}|}{RMS_{actual}} \right) * 100\% \quad (6)$$

$$KSI_{error}\% = \left(\frac{|KSI_{estimated} - KSI_{actual}|}{KSI_{actual}} \right) * 100\% \quad (7)$$

$$ALPHA_{error}\% = \left(\frac{|ALPHA_{estimated} - ALPHA_{actual}|}{ALPHA_{actual}} \right) * 100\% \quad (8)$$

Where RMS_{actual}, KSI_{actual}, $ALPHA_{actual}$ denote the LER parameters calculated from the zero noise images, while $RMS_{estimated}$, $KSI_{estimated}$, $ALPHA_{estimated}$ denote the LER parameters estimated from noise degraded or denoised images. Fig. 8 shows the mean absolute RMS_{error}%, KSI_{error}% and $ALPHA_{error}$% for all test images. When no method is applied the mean absolute $RMS_{HNerror}$% is as high as 245%, the current state of the art PSD method significantly reduces the error to $RMS_{PSDerror} = 33\%$, the SEMD model alone reduces even more the error to $RMS_{SEMDerror} = 22\%$ and finally the combined method (SEMD + PSD) leads to the lowest $RMS_{SEMD+PSDerror}$% of 12%. Similarly, $KSI_{PSDerror}$% is equal to 55% and $KSI_{SEMD+PSDerror}$% is equal to 23%. Finally, $ALPHA_{PSDerror}$% is equal to 66% and $ALPHA_{SEMD+PSDerror}$% is equal to 20%.

4. Conclusions

SEM is nowadays the driving workhorse in the metrology of nanostructures providing high resolution images in a wide variety of materials and applications. However, when it is used in the deep nanoscale (< 10 nm), it suffers from the competition of noise blurring and specimen damage: at short dwell times of electron beam (few number of frames), where the specimen remains intacted by e-beam energy, the image is highly noisy whereas when we increase the number of frames

to smooth out the noise the measured pattern is damaged especially in the case of soft materials.

In order to solve the competition and enable accurate measurements with no noise effects and damage of the pattern, in this paper we elaborated a Deep Learning technique to denoise noisy SEM images. The key idea is to train a DL model using synthesized SEM images to extract the noise from a noisy SEM image and receive as output the denoised image where noise-free measurements of LER parameters can be easily estimated.

We applied the DL approach to address one of the most difficult metrology problems in semiconductor devices: the measurement of Line Edge Roughness of lithography patterns through the analysis of SEM images. To this end, a CNN model based on the DnCNN [1] architecture was trained using synthesized top-down SEM images with rough line patterns and tested on new data to remove the noise, making possible the estimation of noise-free LER metrics. Due to its image-based nature, this DL approach can be combined sequentially with the PSD-based method. It has been shown that the synergy of both methods provides more accurate predictions of LER parameters (rms, correlation length, roughness exponent) in synthesized data from a large variety of lithographic patterns. Future steps of this work may concern the model and/or the training set. Regarding the model, we can increase the network depth and the number of tunable parameters or incorporate the LER evaluation in its loss function. Also, we can enhance the training set to incorporate synthesized SEM images with variable noise level to come closer to the experimental conditions.

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.

References

[1] K. Zhang, et al., Beyond a Gaussian Denoiser: residual learning of deep CNN for image denoising, IEEE Trans. Image Process. 26 (2017) 3142–3155.
[2] V. Constantoudis, et al., Sidewall roughness in nanolithography: Origins, metrology and device effects, Nanolithography: The Art of Fabricating Nanoelectronic and Nanophotonic Devices and Systems, 2013, pp. 503–537.
[3] V. Constantoudis, et al., Noise-free estimation of spatial line edge/width roughness parameters, Proc. SPIE 7272, Metrology, Inspection, and Process Control for Microlithography XXIII, 72724B, 2009.
[4] Y. Ban, et al., Electrical impact of line-edge roughness on sub-45-nm node standard cells, J. Micro/Nanolith. MEMS MOEMS 9 (4) (2010) 041206.
[5] G. Patsis, et al., Quantification of line-edge roughness of photoresists. I. A comparison between off-line and on-line analysis of top-down scanning electron microscopy images, J. Vac. Sci. Technol. B 21 (3) (2003).
[6] D. Li, et al., Noise filtering for accurate measurement of line edge roughness and critical dimension from SEM images, J. Vac. Sci. Technol. B 34 (6) (2016).
[7] T. Kameda, et al., SEM image prediction based on modeling of electron-solid interaction, Proc. SPIE 10145, Metrology, Inspection, and Process Control for Microlithography XXXI, 101451I, 2017.
[8] A. Yamaguchi, et al., Bias-free measurement of LER/LWR with low damage by CD-SEM, Proc. SPIE 6152, Metrology, Inspection, and Process Control for Microlithography XX, 61522D, 2006.
[9] J.S. Villarrubia, B.D. Bunday, Unbiased estimation of linewidth roughness, Proc. SPIE 5752, Metrology, Inspection, and Process Control for Microlithography XIX, 2005.
[10] G.F. Lorusso, et al., Unbiased roughness measurements: subtracting out SEM effects, Microelectron. Eng. 190 (2018) 33–37.
[11] V. Constantoudis, et al., Line edge roughness metrology: recent challenges and advances toward more complete and accurate measurements, J. Micro/Nanolith. MEMS MOEMS 17 (4) (2018) 041014.
[12] G. Papavieros, et al., Line edge roughness measurement through SEM images: effects of image digitization and their mitigation, Proc. SPIE 10446, 33rd European Mask and Lithography Conference, 104460K, 2017.
[13] P. Cizmar, et al., Simulated SEM images for resolution measurement, Scanning 30 (2008) 381–391.
[14] W. Dong, et al., Nonlocally centralized sparse representation for image restoration, IEEE Trans. Image Process. 22 (4) (2013) 1620–1630.
[15] K. Zhang, et al., Learning deep CNN denoiser prior for image restoration, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2808–2817.
[16] V. Nair, G. Hinton, Rectified linear units improve restricted boltzmann machines, Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 807–814.
[17] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, International Conference on Machine Learning, 2015, pp. 448–456.
[18] E. Kang, et al., A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction, Med. Phys. 44 (10) (2017).
[19] A. Cotter, et al., Better mini-batch algorithms via accelerated gradient methods, Adv. Neural Inf. Proces. Syst. 24 (2011) 1647–1655 (NIPS 2011).
[20] CS231n: Convolutional neural networks for visual recognition, Stanford University CS231n: Convolutional Neural Networks for Visual Recognition, 2018 (cs231n.stanford.edu/).
[21] D.P. Kingma, et al., Adam: A Method for Stochastic Optimization, CoRR abs/1412.6980 (2015).