JVST B
Journal of Vacuum Science & Technology B

ARTICLE

avs.scitation.org/journal/jvb

Estimation of critical dimension and line edge roughness using a neural network

Cite as: J. Vac. Sci. Technol. B **39**, 032602 (2021); doi: 10.1116/6.0000806
Submitted: 21 November 2020 · Accepted: 5 April 2021 ·
Published Online: 4 May 2021

View Online

Export Citation

CrossMark

Dehua Li,[1] Soo-Young Lee,[1,a)] Jin Choi,[2] Seom-Beom Kim,[2] and Chan-Uk Jeon[2]

AFFILIATIONS

[1]Department of Electrical and Computer Engineering, Auburn University, Auburn, Alabama 36849
[2]Samsung Electronics, Mask Development Team, 16 Banwol-Dong, Hwasung, Kyunggi-Do 445-701, South Korea

[a)]**Electronic mail:** leesooy@eng.auburn.edu

ABSTRACT

While electron-beam (e-beam) lithography is widely employed in the pattern transfer, the proximity effect makes features blurred, and the stochastic nature of the exposure and development processes causes the roughness in the feature boundaries. In an effort to reduce the proximity effect and line edge roughness (LER), it is often necessary to estimate the critical dimension (CD) and LER. In our previous study, the e-beam lithographic process was modeled using the information extracted from SEM images for the estimation of CD and LER. This modeling involves several parameters to be determined and tends to require a long computation time. In this study, the possibility of improving the accuracy of the CD and LER estimation using a neural network (NN) is investigated. In the NN-based estimation, the explicit modeling of the e-beam lithographic process can be avoided. This paper describes the method of estimating the CD and LER using a NN, including the issues of training, tuning, and sample reduction and presents results obtained through an extensive simulation.

Published under an exclusive license by the AVS. https://doi.org/10.1116/6.0000806

I. INTRODUCTION

Electron beam (e-beam) lithography is often employed in the transfer of patterns with high spatial resolution onto the resist,[1–4] e.g., photo-masks, experimental circuits, etc. However, the proximity effect due to electron scattering in the resist results in a deviation from the target critical dimension (CD).[5] Also, the stochastic e-beam exposure and resist development processes cause the line edge roughness (LER), which does not scale with the feature size. As the feature size goes well below 100 nm, the effect of LER becomes significant and limits the minimum feature size as well as the maximum circuit density achievable in practice. Many empirical and experimental efforts to reduce the LER and correct the proximity effect have been made.[6–8]

In this paper, the term "dose" refers to the amount of incident electron charge on the unit area of resist and the term "exposure" the energy deposited in the unit volume of resist.

In our previous study, a 2D modeling approach was proposed to model the e-beam lithographic process with three components, i.e., the line spread function (LSF), conversion formula (which determines the developing rate for a given exposure),[9] and noise process (exposure fluctuation).[10] The CD and LER measured from SEM images are utilized in this approach, minimizing the modeling error through iterations. Then, the process model is employed in estimating the CD and LER for a given pattern. However, this modeling approach involves several parameters to derive (model) and tends to require a long computation time. Also, the estimation errors were substantial in some cases, e.g., thick resist and low beam energy.

In this study, the possibility of improving the estimation accuracy of CD and LER and reducing the computational requirement by using a neural network (NN) is investigated. NNs are widely employed in a variety of applications, such as the machine vision, speech recognition, signal processing, robotics, etc.[11–14] The use of a NN makes unnecessary an explicit modeling of the e-beam lithographic process and thus does not involve the three components. A NN is trained by a number of samples obtained from the e-beam lithographic process, and the trained NN is used in estimating the CD and LER, given a set of the e-beam lithographic parameters, i.e., the resist thickness, beam energy, beam diameter, linewidth (design linewidth), and normalized dose. In this paper, the procedures of training and tuning the NN are described, and the estimation results obtained from the trained NN through an extensive simulation are discussed. Also, the possibility of reducing the training samples to lower the overhead of sample preparation is investigated.

Published under an exclusive license by the AVS.

The rest of the paper is organized as follows. The e-beam lithographic process is introduced, and the behaviors of the CD and LER are briefly reviewed in Sec. II. The neural network and its training for the estimation of CD and LER are described in Sec. III. The results are provided with a discussion in Sec. IV, and the issue of reducing training samples is discussed in Sec. V, followed by a summary in Sec. VI.

II. E-BEAM LITHOGRAPHIC PROCESS

A typical substrate system consists of a certain type of resist on top of the substrate as shown in Fig. 1. The resist layer is exposed according to the dose required for a circuit pattern to be transferred. The exposure distribution, $e(x, y, z)$, is related to the dose, $d(x, y, 0)$, and the point spread function, $psf(x, y, z)$, through the convolution. That is,

$$e(x, y, z) = \int_{y'}\int_{x'} d(x - x', y - y', 0)psf(x', y', z)\,dx'\,dy'.$$

Then, the substrate system goes through a resist-development process and the transferred pattern is obtained as the remaining resist profile. A typical SEM image of a line/space pattern and the detected edges (boundaries) are shown in Fig. 2. The resist development is governed by the developing rate, which is derived from the exposure through the conversion formula.[9] Also, the simulated contours (feature boundaries) of a remaining resist profile at the top and bottom layers are provided in Fig. 3 for a substrate system. From reference resist profiles, the feature boundaries at the top and bottom layers are detected from which the top and bottom CDs/LERs are determined, respectively. The CD is computed as the average of top and bottom CDs. The standard deviation of the edge location is calculated along each boundary (at the top and bottom layers) of a line. Then, the LER is computed as the average of the standard deviations of all lines, i.e., 1-σ LER.[10]

In a previous study,[10] this e-beam lithographic process was modeled explicitly by the LSF, conversion formula, and exposure fluctuation to estimate the CD and LER in a given pattern. An alternative approach, i.e., employing a NN, that avoids the explicit modeling and, therefore, reduces the computational requirement is taken in this study. The weights of all edges (links between neurons or nodes) in the NN may be considered to collectively model the e-beam lithographic process and are determined to realize the behaviors of CD and LER.

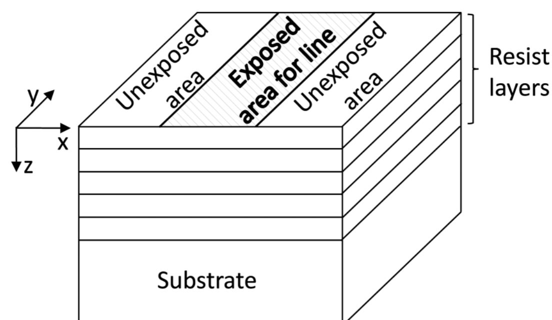

FIG. 1. In this study, the resist is modeled by five layers, and the line feature is assumed to be infinitely long along the Y-dimension.

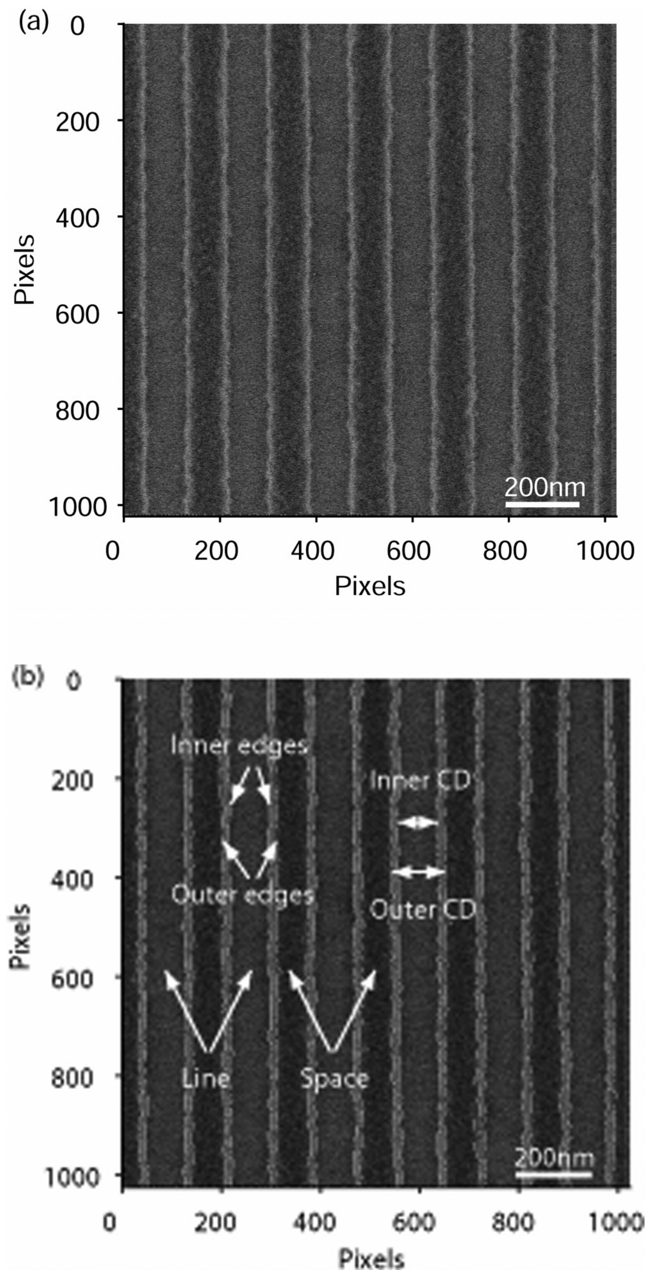

FIG. 2. (a) SEM image of a line/space pattern and (b) the detected inner and outer edges overlaid on the SEM image: a substrate system of PMMA on Si, a resist thickness of 100 nm, a beam energy of 50 keV, a normalized dose of 1.082, and a linewidth of 120 nm.

Published under an exclusive license by the AVS.

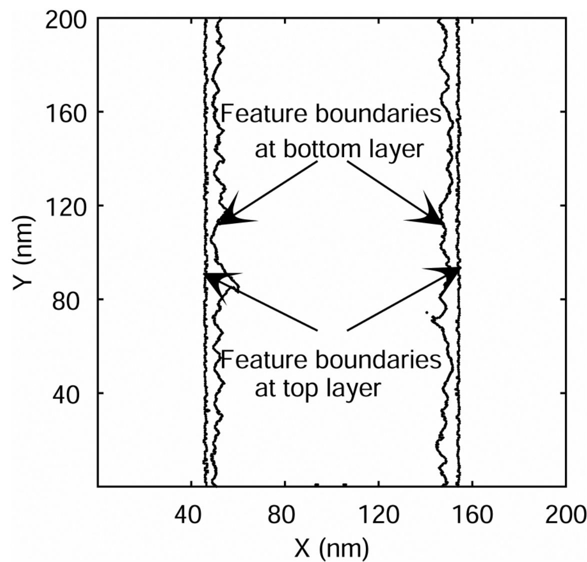

FIG. 3. Simulated contours of the remaining resist profiles at the top and bottom layers: a substrate system of PMMA on Si, a resist thickness of 100 nm, a beam energy of 10 keV, a beam diameter of 10 nm, and a normalized dose of 0.335.

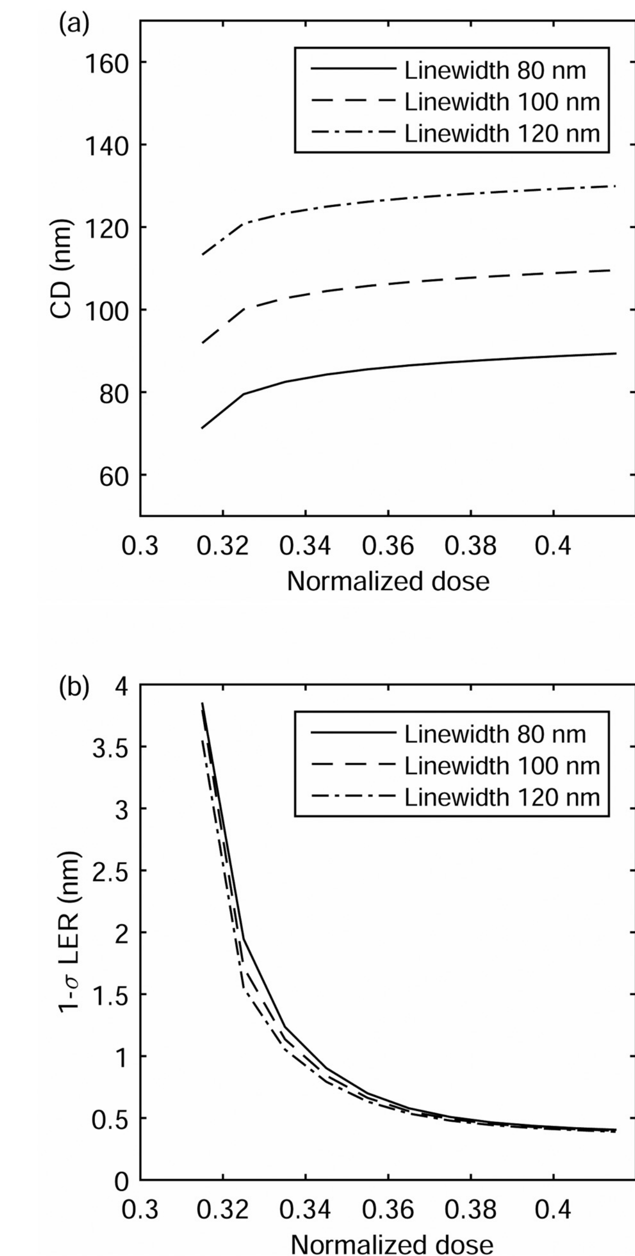

FIG. 4. Dependency of (a) CD and (b) 1-σ LER on the dose when the development time is fixed: a substrate system of PMMA on Si, a resist thickness of 100 nm, a beam energy of 10 keV, a beam diameter of 10 nm, a normalized dose of 0.335, an exposure interval of 1 nm, and a development time of 0.5 min.

Both CD and LER vary with e-beam lithographic parameters in general. For example, the CD monotonically increases with the increment of dose as shown in Fig. 4(a). When the dose is relatively low, the CD increases fast with the dose. However, once the edge location moves out of the exposed area (i.e., feature), the rate of CD increase significantly drops as the dose continues to increase. This is due to the fact that the exposure level decreases rapidly over the edge of the exposed area. A similar dependency of LER on the dose can be seen in Fig. 4(b) except that the LER decreases with the dose. As the dose increases, the relative fluctuation of exposure decreases and the edge location tends to move out of the exposed area, both making the LER smaller. Note that the absolute fluctuation of exposure is much smaller in the outside of a feature than in the inside.

III. ESTIMATION OF CD AND LER

The neural network employed for the estimation of the CD and LER in this study is a multilayer NN, which is trained by the error back-propagation (EBP) algorithm with a set of training samples. Each training sample is a known CD or LER for a combination of e-beam lithographic parameters.

A. Structure of NN

The NN consists of an input layer, an output layer, and hidden layers between the input and output layers as illustrated in Fig. 5. There are multiple nodes (neurons) in each layer, and every node in a layer is connected to all the nodes in the following layer. The output from node j in a layer is expressed as $O_j = f(\sum_{i=1}^{N} w_{ij} I_i)$, where $f()$ is the activation function, N is the number of inputs from the previous layer to node j, I_i is the ith input, i.e., the output from node i in the previous layer, and w_{ij} is the weight of the link (edge) between node i and node j. In this study, the sigmoid function $f(x) = \frac{1}{1+e^{-x}}$ is employed as the activation function.

Each training sample, which is an input vector of the NN, consists of e-beam lithographic parameters, i.e., resist thickness,

Published under an exclusive license by the AVS.

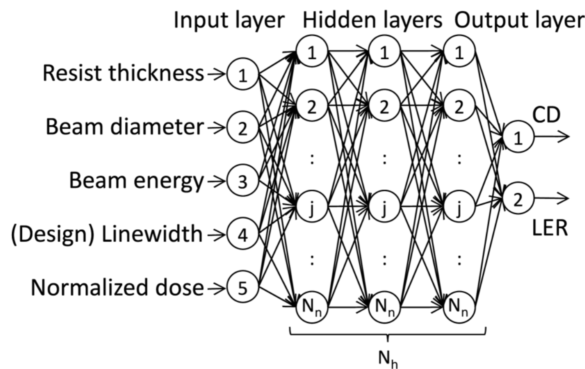

FIG. 5. Illustration of a neural network with an input layer of five inputs (resist thickness, beam energy, beam diameter, linewidth, and normalized dose), N_h hidden layers with N_n neurons each, and an output layer of two outputs (CD and 1-σ LER).

beam energy, beam diameter, linewidth, and normalized dose, and the output vector of the NN contains the CD and LER. For each training sample, the CD and LER are known. From this point on, each element (i.e., e-beam lithographic parameter) of the input vector will be referred to as an input and, similarly, each element of the output vector as an output.

B. Training with e-beam lithographic samples

Through the input layer, each training sample is fed into the NN. After the forward calculation, the corresponding CD and LER become available at the output layer. Then, the training errors (CD and LER) are computed, which are propagated backward through the layers to determine the amount of weight (w_{ij}) adjustment for every edge. This process of training is repeated until the NN converges; e.g., the training errors no longer decrease. The edge weights may be updated after each training sample or a batch of training samples. The overall flow of training is shown in Fig. 6. The specific issues involved in the training are discussed below.

1. Normalization of inputs

The range of an input may vary significantly with input, e.g., the resist thickness (nm) from 100 to 500, the beam energy (keV) from 10 to 50, and the normalized dose from 0.3 to 3.5. Therefore, in order to have the same or a similar range for all inputs in the NN, each input is normalized as follows:

$$x'_k = \frac{x_k - E[x_k]}{\sigma_k}, \tag{1}$$

where x_k is an input where k is the input index, x'_k is the normalized input, and $E[x_k]$ and σ_k are the mean and standard deviation computed for the distribution of x_k, respectively.

2. Initialization of weights

The initial weights of edges may have a substantial effect on the convergence and error of training. In order to avoid the

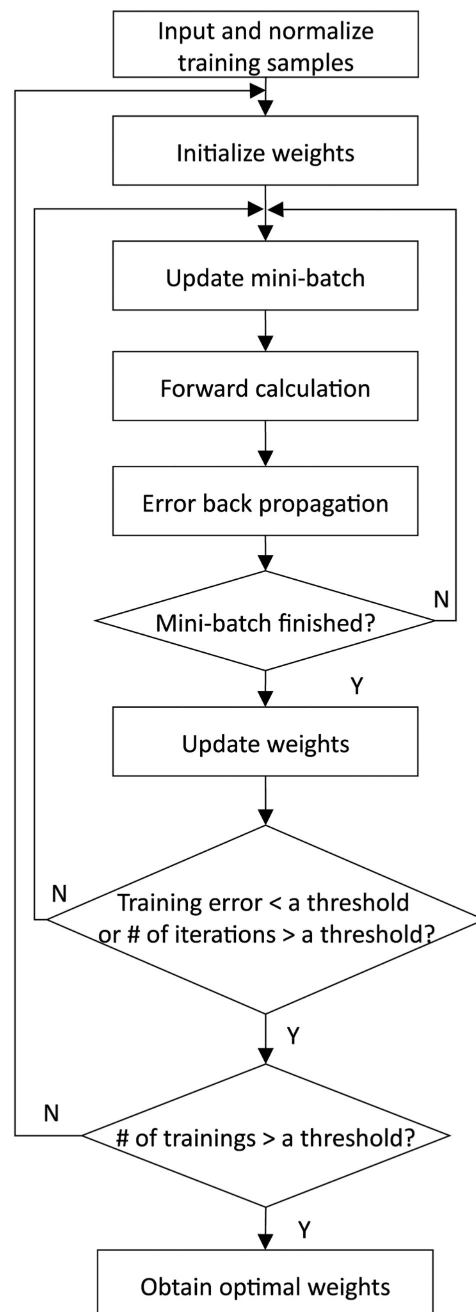

FIG. 6. NN training process employs the EBP algorithm using a minibatch. Each minibatch of samples is selected randomly in an iteration. The weights of nodes are updated after the training on a minibatch of samples is finished.

saturation region (i.e., too small a derivative) of the activation function, the combined input ($\sum_{i=1}^{N} w_{ij} I_j$) to a node should not be too large. A reasonable choice is to maintain the range of the normalized input. Note that the input normalized according to Eq. (1) has

Published under an exclusive license by the AVS.

the mean of 0 and the standard deviation of 1. The standard deviation of the sum of N uncorrelated samples of a random variable is $\sqrt{N}$ times larger than the standard deviation of the random variable. Hence, each weight may be initialized to a value selected from the range of $[-1/\sqrt{N} \quad 1/\sqrt{N}]$, where N is the number of inputs to a node.[15]

3. Gradient-descent training

In the error back-propagation, the gradient-descent training algorithm is employed.[16] The loss function consisting of the CD and LER training errors, i.e., half of the mean square error, is employed, and the weights are adjusted such that the cost function decreases most.

The amount of weight adjustment for the link between node i in a layer and node j in the next layer is computed as follows:

$$\Delta w_{ij} = -\eta \cdot \delta_j \cdot I_i, \tag{2}$$

where η is the learning rate, δ_j is the error at node j, and I_i is the input to node j from node i.

For an output node j (a node in the output layer), δ_j is computed from the difference between the output O_j and the known (target) output T_j (CD or LER);[16] i.e.,

$$\delta_j = (O_j - T_j) \cdot O_j \cdot (1 - O_j). \tag{3}$$

For a hidden layer node j (a node in a hidden layer) connected to N nodes in the next layer, δ_j is computed as follows:

$$\delta_j = O_j \cdot (1 - O_j) \cdot \sum_{k=1}^{N} w_{jk}\delta_k. \tag{4}$$

That is, the error at node j is derived from the errors at the nodes in the next layer.

4. Learning rate

The learning rate sets the relative amount of weight adjustment that is allowed in each weight update. A too large learning rate makes a weight change faster but possibly oscillates without converging. On the other hand, with a smaller learning rate, a weight tends to converge to a stable value, but the training takes longer and may be possible to be trapped locally. In the NN that generates normalized outputs, i.e., their absolute values <1, as in this study, the weight updates through the back-propagation can become so small that an effective training is not possible, especially when the number of layers is large.[17,18] Hence, a proper selection of the learning rate is necessary in such cases. In this implementation, the learning rate is determined empirically.

5. Minibatch

Updating weights for every training sample may cause them to oscillate, hampering their convergence and leading to a relatively large training error. On the other hand, accumulating weight changes for all training samples and reflecting the accumulated changes of weights in one step tend to make the behaviors of weights more stable but the convergence slower. A reasonable approach to achieving an acceptable training error and speed would be to adopt the concept of minibatch; i.e., weights are updated after weight changes are accumulated for a subset of training samples (a minibatch).

The main purpose of employing a minibatch in this application is to reduce the training time without sacrificing the accuracy substantially.[19] As the minibatch size increases, the accuracy tends to improve, but the training time increases. Also, once the minibatch size reaches a certain level, the accuracy does not improve significantly. That is, there exists a trade-off between the training accuracy and time.

Training samples may be stored in a specific order; e.g., one of the lithographic parameters is varied with the others fixed. Taking minibatches of training samples in such an order is likely to follow a biased training path by the gradient-descent method, deviated from the path leading to the minimal training error. To enhance the possibility of achieving the minimal error, randomly selected training samples are included in each minibatch.

C. Estimation

Once the NN is trained, it is straightforward to use it for the estimation of CD and LER. Given a specification of the lithographic parameters, the forward computation through the NN is carried out, and the estimated CD and LER become available at the output of NN without any iteration as shown in the flowchart in Fig. 7.

IV. RESULTS AND DISCUSSION

For training and testing the NN, remaining resist profiles for which the CD and LER are known are employed. A number of such remaining resist profiles are generated through simulation for various combinations of the lithographic parameters. The CD and

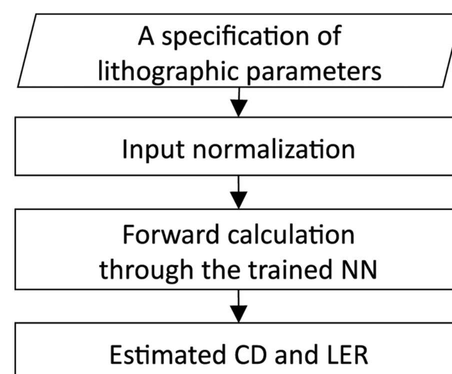

FIG. 7. Process for estimating CD and 1-σ LER with one forward calculation using the trained NN.

Published under an exclusive license by the AVS.

LER measured from the reference remaining resist profiles are used as the training and test samples.

In this study, the simulated resist profiles are used in the training and estimation in order to obtain a large number of samples encompassing a wide range of each parameter. However, the same NN (structure and training method) as described in the paper can be employed for experimental results.

A. Reference remaining resist profiles

In creating reference remaining resist profiles, each of the lithographic parameters is varied: the resist thickness from 100 to 500 nm with an interval of 100 nm, the beam energy from 10 to 50 keV with an interval of 10 keV, the beam diameter of 2, 5, 10, 15, and 20 nm. The exposure distribution in the resist is computed using the point spread function generated by the software CASINO.[20] The substrate system is composed of PMMA on Si. For each combination of the lithographic parameters, 50 instances of the 5-layer stochastic PSF are generated. Then, a fast path-based method[9] for simulation of resist development is used to obtain the remaining resist profile from which the CD and LER are computed.

A line feature is considered where the linewidth is varied from 50 to 150 nm with an interval of 10 nm. The normalized dose (reference dose: $320\,\mu\mathrm{C/cm^2}$) is varied such that both underdeveloped and overdeveloped remaining resist profiles are obtained in each case. From each reference remaining resist profile, the CD and LER are measured according to the edge location using the method defined in Sec. II from one line in one image and the length of the line feature is 2000 nm, where the pixel size is 1.0 nm. The total number of samples used for the training is 4350 and that for the estimation is 2875.

B. Training and estimation errors

The CD training error is defined as the average difference between the trained CD (i.e., CD obtained after the training) and the known CD and the CD estimation error as the average difference between the estimated CD and known CD as follows:

$$\epsilon_{CD} = \frac{1}{N}\sum_{i=1}^{N} |CD_i - CD_i'|, \tag{5}$$

where N is the number of samples used in the training or estimation, CD_i is the trained or estimated CD and CD_i' is the known CD.

The LER training and estimation errors are defined similarly with

$$\epsilon_{LER} = \frac{1}{N}\sum_{i=1}^{N} |LER_i - LER_i'|, \tag{6}$$

where LER_i is the trained or estimated LER, and LER_i' is the known LER.

The percent training and estimation errors for the CD and LER are computed as follows:

$$\epsilon_{CD.p} = \frac{\frac{1}{N}\sum_{i=1}^{N} |CD_i - CD_i'|}{\frac{1}{N}\sum_{i=1}^{N} CD_i'} \times 100\%, \tag{7}$$

$$\epsilon_{LER.p} = \frac{\frac{1}{N}\sum_{i=1}^{N} |LER_i - LER_i'|}{\frac{1}{N}\sum_{i=1}^{N} LER_i'} \times 100\%. \tag{8}$$

A small value of CD (or LER) tends to make the percent CD error larger, which affects the overall average percent error of CD much disproportionately. To avoid this, the average percent CD error is computed as the average CD error normalized to the average CD rather than the average of the CD errors normalized to the respective CDs.

C. Tuning of NN

To achieve the best performance in both accuracy and computation time, the structure and parameters of NN need to be optimized. In the training of NN, the loss function is defined as follows:[21]

$$loss = \frac{1}{2}\left(\sum_{i=1}^{N} (CD_i - CD_i')^2 + \sum_{i=1}^{N} (LER_i - LER_i')^2\right). \tag{9}$$

However, in the analysis of performance, only the LER error is considered since the CD error is much smaller (than the LER error).

The size of minibatch affects the performance of NN substantially. As shown in Fig. 8(a), the LER estimation error is around 10% for the minibatch size of 5. For the minibatch size of 20 or larger, the LER estimation error is around 5% or less. On the other hand, the computation time is longer for a larger size of minibatch. For this application, it turns out that the minibatch size of 20 is a reasonable choice. The training takes about 50 h on the PC with Intel Core i5-4210U and 12 GB RAM running at 2.4 GHz when 4096 samples are used as a batch but about 500 s with the minibatch of size 20. That is, using a minibatch reduces the training time greatly without sacrificing the accuracy substantially. For comparison to the direct modeling, it takes about 120 h on the same PC to model the e-beam lithographic process (for one lithographic configuration) with the line spread function and noise, considering a feature of $120 \times 1000\,\mathrm{nm}^2$ and 5 layers of the resist. The effect of the learning rate on the LER estimation error is shown in Fig. 8(b). It is seen that the LER estimation error is larger than 15% for the small learning rate of 0.025 and decreases close to 5% as the learning rate increases to 1.0. A larger learning rate tends to make the convergence faster; however, too large a learning rate such as 2.5 in Fig. 8(b) leads to the divergence of the NN. Hence, the learning rate of 1.0 is used. Though there are fluctuations of the training error in the training process, this learning rate makes a fast training and achieves smaller training/estimation errors for both CD and LER.

Published under an exclusive license by the AVS.

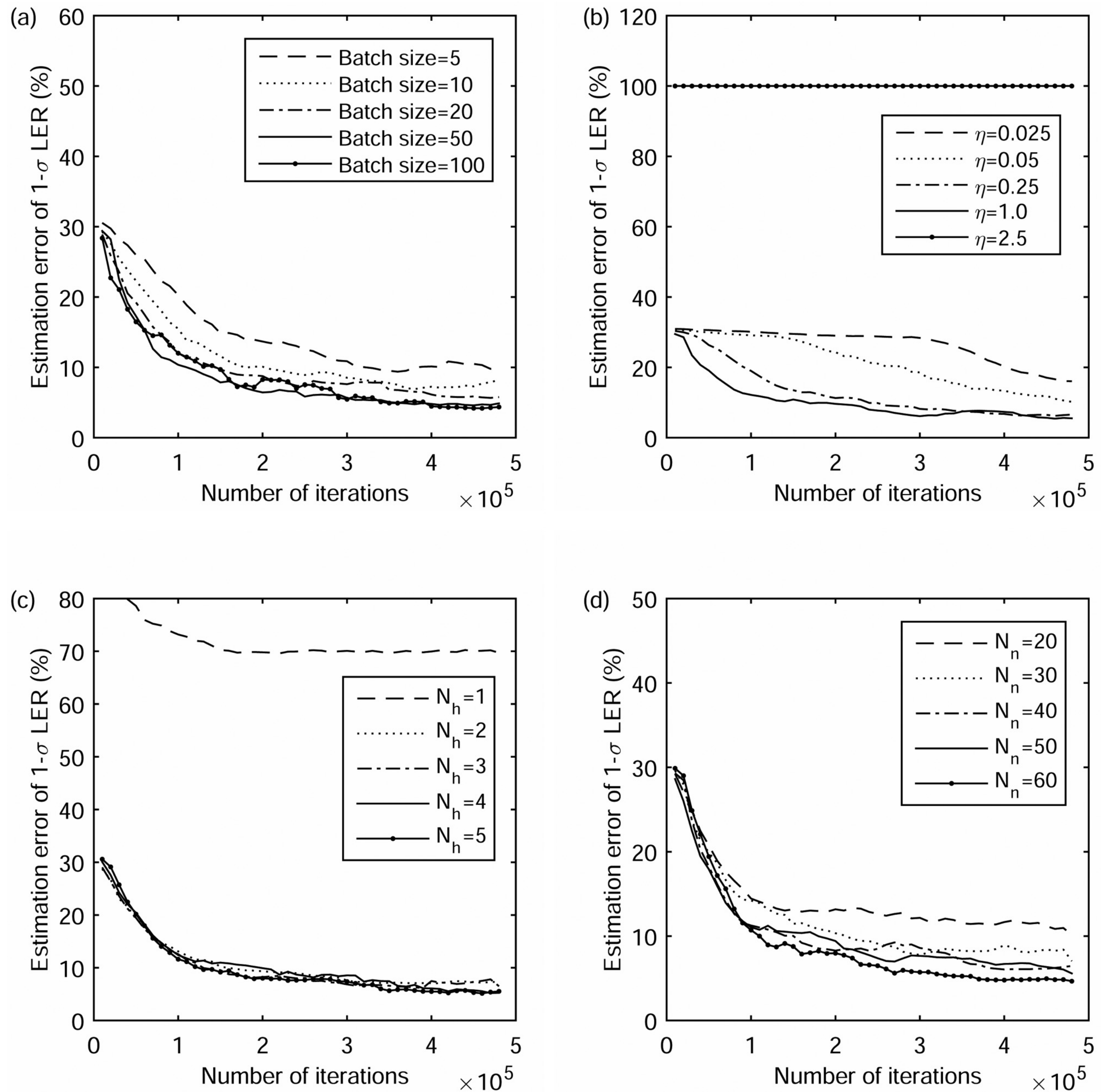

FIG. 8. 1-σ LER estimation errors in the training process with varying (a) minibatch size, (b) learning rate, (c) the number of hidden layers, and (d) the number of nodes in a hidden layer.

The structure of NN may be characterized by the number of hidden layers (N_h) and the number of nodes (neurons) per hidden layer (N_n). In Fig. 8(c), the dependency of the LER estimation error on N_h is shown. The LER estimation error is as high as 70% when only a single hidden layer is employed and decreases down to around 5% as N_h increases to 4. Adding more hidden layers beyond 4 does not decrease the LER estimation error further. More nodes in each hidden layer result in a smaller LER estimation error as can be seen in Fig. 8(d) but require a longer computation time. The LER estimation error is around 13% for $N_n = 20$ and decreases down to 5% as N_n increases to 50.

Published under an exclusive license by the AVS.

D. Estimation of CD and LER

The optimized NN consists of five layers in total, i.e., three hidden layers, an input layer, and an output layer. There are five nodes in the input layer for the five lithographic parameters: resist thickness, beam diameter, beam energy, linewidth, and dose. Each hidden layer is composed of 50 nodes. The output layer has two nodes for the CD and LER. In the optimized NN, the minibatch size is set to 20 and the learning rate is set to 1.0. Each lithographic parameter is sampled for the training and estimation as described in Sec. IV A. In the optimized NN, the linewidths and doses for the estimation are selected to be different from those used in the training. The linewidth is varied from 50 to 150 nm with an interval of 20 nm for the training and from 60 to 140 nm with an interval of 20 nm for the estimation.

The optimized NN is trained with a subset of samples such that the training errors are minimized, and the CD and LER are estimated for various combinations of the lithographic parameters using the rest of the samples. The estimation errors are provided for various combinations of resist thickness and beam energy in Table I where each error is the error averaged over the other three lithographic parameters, i.e., beam diameter, linewidth, and dose.

TABLE I. CD and 1-σ LER estimation errors in nm and percentage for varying resist thickness and beam energy for the substrate system of PMMA on Si: Errors are averaged over the beam diameters of 2, 5, 10, 15, and 20 nm; linewidths of 60, 80, 100, 120, and 140 nm; and normalized doses.

Resist thickness (nm)	Beam energy (keV)	ϵ_{CD} (nm)	$\epsilon_{CD.p}$ (%)	ϵ_{LER} (nm)	$\epsilon_{LER.p}$ (%)
100	10	0.41	0.40	0.05	4.39
	20	0.58	0.57	0.07	5.87
	30	0.45	0.44	0.05	4.16
	40	0.39	0.39	0.05	4.09
	50	0.58	0.58	0.07	4.73
200	10	0.65	0.65	0.07	4.18
	20	0.50	0.50	0.04	2.78
	30	0.45	0.45	0.07	4.16
	40	0.69	0.67	0.08	5.11
	50	0.62	0.61	0.05	3.27
300	10	0.62	0.64	0.06	3.35
	20	0.43	0.42	0.07	3.65
	30	0.69	0.68	0.10	4.86
	40	0.40	0.39	0.10	5.08
	50	0.61	0.60	0.09	5.02
400	10	1.08	1.15	0.07	4.07
	20	0.87	0.92	0.10	4.30
	30	0.64	0.65	0.08	3.37
	40	0.55	0.55	0.07	3.19
	50	0.54	0.54	0.09	4.50
500	10	1.79	2.68	0.10	5.87
	20	1.09	1.58	0.08	3.24
	30	0.78	1.07	0.07	2.74
	40	1.17	1.55	0.10	4.30
	50	0.89	1.16	0.10	3.98

The average CD and LER estimation errors for all the cases are 0.68% and 4.12%, respectively. It may be said that the accuracy of estimation is high enough for the practical use. One observation is that the largest estimation errors are obtained for the resist thickness of 500 nm and a beam energy of 10 keV. The thicker the resist is and the lower the beam energy is, the broader (less sharp) the point spread function is. Hence, the exposure contrast over the feature boundary is smaller, which makes the location of the feature edge less stable. That is, the samples for the estimation are likely less consistent with those for the training. This results in larger estimation errors.

V. REDUCTION OF TRAINING SAMPLES

In general, using a larger number of training samples results in a smaller estimation error, however, requires more experiments and a longer time to prepare samples. Also, adding more training samples beyond some point may not improve the estimation accuracy. Therefore, it is desirable to minimize the number of training samples without increasing the estimation error significantly. For each lithographic parameter, the effect of reducing training samples (from the training samples used to obtain the results in Sec. IV D) on the estimation errors is analyzed. The results are provided in Figs. 9(a)–9(e).

In Fig. 9(a), it is observed that when the number of (different) linewidths is reduced from 6 to 3, the estimation errors, especially the CD estimation error, are increased significantly. However, when four different linewidths are kept including the two linewidths at or close to the two ends of the original range, the increase of the estimation errors is not significant. Hence, the number of linewidths may be reduced to four at the expense of a minor increase in the estimation errors. Both CD and LER vary with the dose rapidly when the dose level is relatively low; however, changes vary slowly once the dose exceeds a certain level. Therefore, as long as the dose is sampled sufficiently in the fast-varying region, a significant increase of the estimation errors can be avoided as indicated in Fig. 9(b). Depending on the acceptable accuracy of estimation, one may readily reduce the number of doses in the dose range down to four.

The effect of the change in the beam diameter on the sharpness of PSF and in turn on the CD and LER is smaller when the beam diameter is relatively large. Hence, removing one of the large beam diameters does not increase the estimation errors much as shown in Fig. 9(c).

It is seen in Figs. 9(d) and 9(e) that the removal of all training samples for a resist thickness or beam energy increases the estimation errors, the LER estimation error, in particular, significantly. Therefore, it is advisable to maintain the sampling intervals of 100 nm and 10 keV for the resist thickness and beam energy, respectively, which are already sufficiently coarse. It would be still worthwhile to investigate what the sampling interval is optimal for each input considering the accuracy and sample size.

The training time depends on several factors, e.g., the number of samples, acceptable accuracy, the number of times each sample is used, etc. Hence, a reduced set of training samples does not necessarily lead to a shorter training time.

Published under an exclusive license by the AVS.

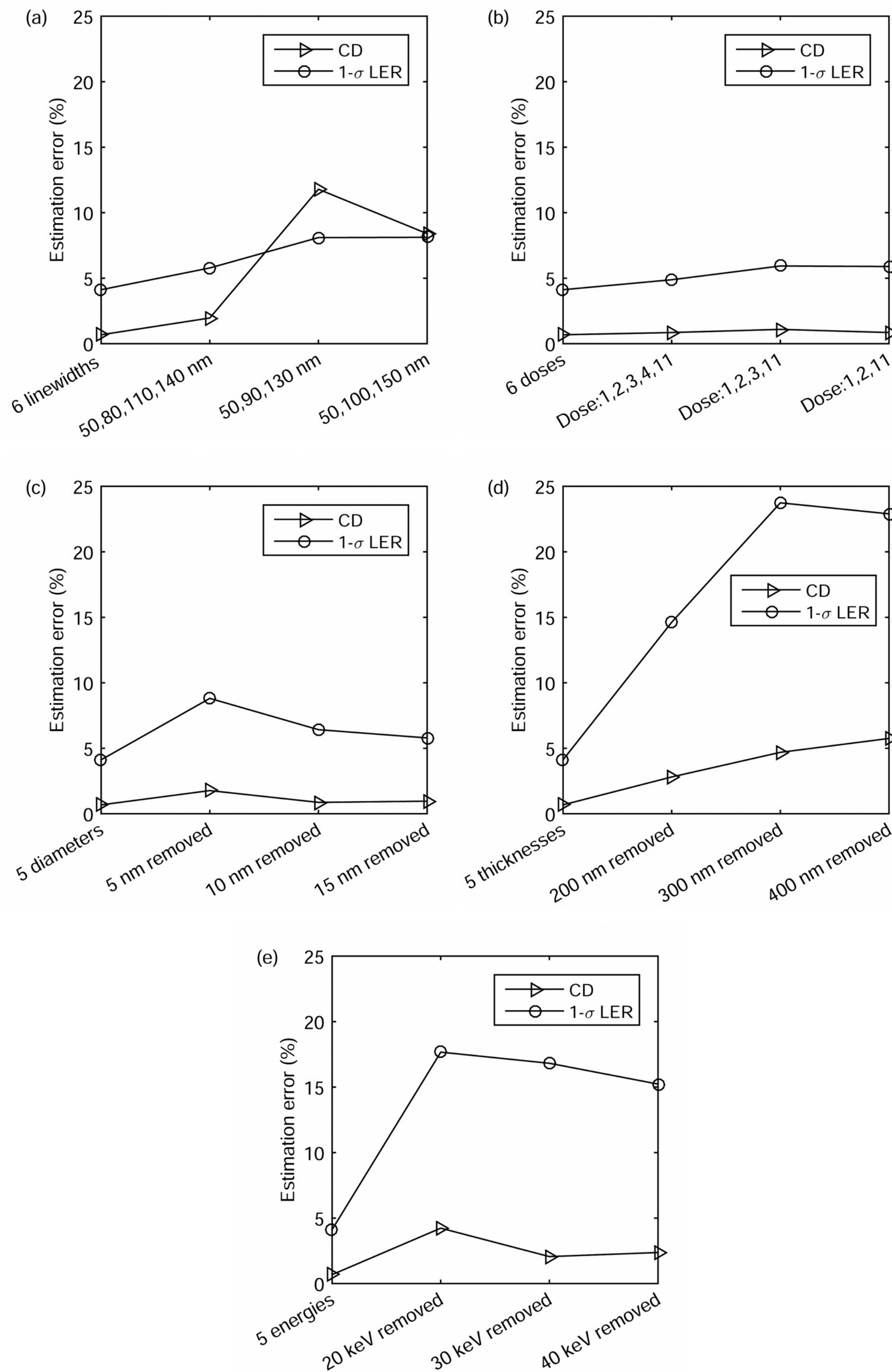

FIG. 9. CD and 1-σ LER estimation errors with a reduced number of training samples for (a) linewidth, (b) normalized dose, (c) beam diameter, (d) resist thickness, and (e) beam energy.

Published under an exclusive license by the AVS.

VI. SUMMARY

The line edge roughness becomes a more critical issue as the CD decreases down to nanometers. An accurate estimation of CD and LER is an important task by itself and also a necessary step toward the minimization of the proximity effect and LER in the e-beam lithography. The estimation of the CD and LER for the e-beam lithographic process using an NN is investigated.

A number of remaining resist profiles are generated using the exposure and development simulation programs, and from each profile, the CD and LER are measured. The NN is trained and tuned by the measured CD and LER. Given a line feature with the specification of e-beam lithographic parameters, the CD and LER are estimated by the NN. Through an extensive performance analysis, it has been shown that, with proper network structure and training, the NN can accurately estimate the CD and LER for wide ranges of the lithographic parameters. Using an NN does not require a direct modeling of the e-beam lithographic process, which is computationally very intensive, and, therefore, avoids the involvement of the process components. Also, an addition or deletion of certain lithographic parameter can be readily done in the training and estimation.

A large number of training samples is required for an acceptable estimation accuracy. However, it has been shown that, for this application to the e-beam lithography, it is possible to reduce the sample size substantially without degrading the accuracy much. As training samples from experiments are accumulated, the accuracy of CD and LER estimation by an NN would be improved.

When the value of a parameter is outside the range of respective training samples, the estimation error is expected to increase but still acceptable unless it is far away from the range. The NN model and the training method described in this paper are not dependent on the substrate system or circuit pattern. Hence, it is expected that the NN is able to handle other (substrate) materials and patterns with a proper set of samples prepared.

ACKNOWLEDGMENTS

This work was supported by a research grant from Samsung Electronics Co., Ltd.

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1] F. Hu and S.-Y. Lee, J. Vac. Sci. Technol. B **21**, 2672 (2003).
[2] X. Zhao, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, and C.-U. Jeon, J. Vac. Sci. Technol. B **32**, 06F505 (2014).
[3] Q. Dai, S.-Y. Lee, S.-H. Lee, B.-G. Kim, and H.-K. Choi, J. Vac. Sci. Technol. B **29**, 06F314 (2011).
[4] G. P. Watson, L. A. Fetter, and J. A. Liddle, J. Vac. Sci. Technol. B **15**, 2309 (1997).
[5] S.-Y. Lee and B. D. Cook, IEEE Trans. Semicond. Manuf. **11**, 108 (1998).
[6] G. Owen and P. Rissman, J. Appl. Phys. **54**, 3573 (1983).
[7] E. Seo, B. Kyung Choi, and O. Kim, Microelectron. Eng. **53**, 305 (2000).
[8] A. Schleunitz and H. Schift, J. Micromech. Microeng. **20**, 095002 (2010).
[9] Q. Dai, R. Guo, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, C.-U. Jeon, B.-G. Kim, and H.-K. Cho, Microelectron. Eng. **127**, 86 (2014).
[10] R. Guo, S.-Y. Lee, J. Choi, S.-H. Park, I.-K. Shin, and C.-U. Jeon, J. Vac. Sci. Technol. B **34**, 011601 (2016).
[11] M. Riedmiller and H. Braun, *Proceedings of IEEE International Conference on Neural Networks*, *San Francisco, CA, March 28–April 1, 1993* (IEEE, New York, 1993), Vol. 1, p. 586
[12] H. A. Rowley, S. Baluja, and T. Kanade, IEEE Trans. Pattern Anal. Mach. Intell. **20**, 23 (1998).
[13] G. Hinton *et al.*, Signal Process. **29**, 82 (2012).
[14] R. Koker, Inf. Sci. **222**, 528 (2013).
[15] X. Glorot and Y. Bengio, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, *Sardinia, Italy, May 13–15, 2010* (Microtome Publishing, Brookline, MA, 2010), Vol. 9, p. 249
[16] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, Biol. Cybern. **59**, 257 (1988).
[17] B. M. Wilamowski and J. Korniak, *19th 2015 IEEE International Conference on Intelligent Engineering Systems (INES)*, *Bratislava, Slovakia, September 3–5, 2015* (IEEE, New York, 2015), pp. 21–29.
[18] S. Hochreiter, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. **6**, 107 (1998).
[19] A. Cotter, O. Shamir, N. Srebro, and K. Sridharan, *Advances in Neural Information Processing Systems*, *Granada, Spain, December 12–17, 2011* (Curran Associates Inc., Red Hook, NY, 2011), Vol. 24.
[20] D. Drouin, A. Couture, D. Joly, X. Tastet, V. Aimez, and R. Gauvin, Scanning **29**, 92 (2007).
[21] W. Kristjanpoller and E. Hernandez, Expert Syst. Appl. **84**, 290 (2017).

Published under an exclusive license by the AVS.