PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Unsupervised machine learning based CD-SEM image segregator for OPC and process window estimation

Dey, Bappaditya, Cerbu, Dorin, Khalil, Kasem, Halder, Sandip, Leray, Philippe, et al.

Bappaditya Dey, Dorin Cerbu, Kasem Khalil, Sandip Halder, Philippe Leray, Sayantan Das, Yasser Sherazi, Magdy A. Bayoumi, Ryoung Han Kim, "Unsupervised machine learning based CD-SEM image segregator for OPC and process window estimation," Proc. SPIE 11328, Design-Process-Technology Co-optimization for Manufacturability XIV, 113281G (23 March 2020); doi: 10.1117/12.2552055

SPIE.

Event: SPIE Advanced Lithography, 2020, San Jose, California, United States

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Unsupervised Machine Learning based CD-SEM image segregator for OPC and Process Window Estimation

Bappaditya Dey[a], Dorin Cerbu[a], Kasem Khalil[b], Sandip Halder[a], Philippe Leray[a], Sayantan Das[a], Yasser Sherazi[a], Magdy A. Bayoumi[b,c], and Ryoung Han Kim[a]

[a]imec, Kapeldreef 75, 3001 Leuven, Belgium
[b]The Center for Advanced Computer Studies, University of Louisiana at Lafayette, Louisiana, USA
[c]Department of Electrical and Computer Engineering, University of Louisiana at Lafayette, Louisiana, USA

ABSTRACT

As we are stepping towards sub-10 nm nodes, process window monitoring for systematic defects is becoming more and more critical. In traditional process window excursion and control (PWEC) methods often optical defect inspection is done on a focus and dose modulated wafer first. Once the different systematic defects are detected in a particular focus/energy die, we flag the repeating defect locations as potential hotspots and rank them based on how early/late they fail in a focus/energy modulated columns. So, during this first pass we get a rough idea of which locations are failing. However, due to limited resolution of optical tools, the true process window can only be gathered during a second pass with an e-beam tool. The key idea to define a true process window demands a detailed analysis of CD and other underlying features. We have proposed a new method of analyzing the process window with an unsupervised machine learning approach. Our proposed algorithm will extract the underlying key features and encode these to latent feature vectors or latent vector space instead of the conventional CD, given a dataset of thousands of CD-SEM images, and then rank the images based on a similarity index and then to automatically determine the process window. This work addresses the following problems (1) with a defect inspection tool this task seems tedious and time consuming and often require human intervention to analyze a large number of features, (2) a CD-SEM based process window analysis might not always match with a defect inspection-based process window. Our generalized variational auto-encoder based approach does this automatically. Also, we have analyzed and validated our result against conventional approach.

Keywords: machine learning, autoencoder, process window, defect, unsupervised, deep learning, similarity index, generative-adversarial-network, OPC, CD-SEM.

1. INTRODUCTION

A semiconductor device such as an integrated chip is produced by a sequence of process steps. Many of them require a lithographic mask through which 1 or 2-D pattern is printed on a photoresist layer, followed by one or more etch processes for transferring the pattern to the underlying layers. The design of such lithographic masks nowadays involves the prediction of the printed pattern considering optical, resist and etch phenomena, which occur at the scale of the mask below. These effects lead to unavoidable deviations on the printed pattern with respect to the design intent. Dedicated EDA tools, as an example OPC (Optical Proximity Correction) software are used for determining and making corrections to the intended initial mask design to provide best possible approximation of the design intent on the printed wafer. OPC optimization is based on the fragmentation of the initial design edges to compensate for the given phenomena and relies on an accurate modeling to predict simulated contours of the printed features. The window defines the printability performance limits in terms of the focus and exposure dose settings of a lithographic tool, within which limits a reliable print of the pattern is obtainable. Some patterns can have bigger printable windows than other patterns, due to various reasons such as design geometry, accuracy of the OPC modelling etc. The simulated windows are subsequently verified experimentally by manufacturing a test mask and using it to print various pattern features on a plurality of die areas of a photosensitive

Further author information: (Send correspondence to Bappaditya Dey)
Bappaditya Dey.: E-mail: Bappaditya.Dey@imec.be
Kasem Khalil: E-mail: kasem.khalil1@louisiana.edu

Design-Process-Technology Co-optimization for Manufacturability XIV, edited by Chi-Min Yuan, Proc. of SPIE Vol. 11328, 113281G · © 2020 SPIE
CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2552055

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

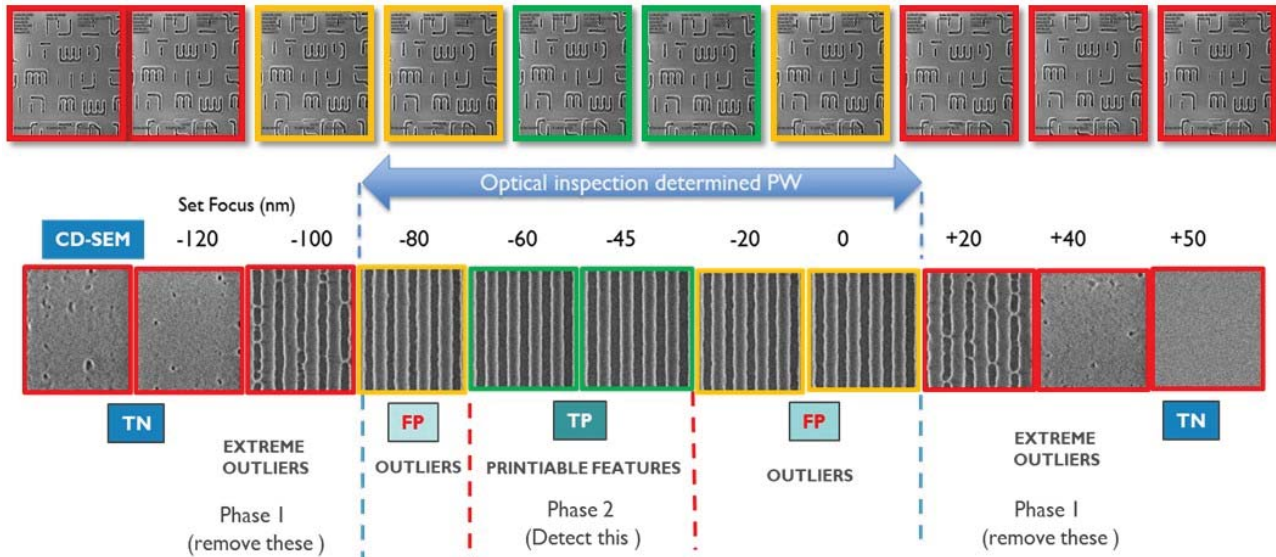


Figure 1: Optical inspection determined Process Window.

resist layer. Each die area is being printed with varying values of focus and exposure conditions. The process windows are determined by measuring features/CD on the plurality of die areas. Differences between the experimental and simulated windows are evaluated and may necessitate further OPC-based optimization of the mask design.

Different measurement techniques are known for analyzing the printed dies. Optical inspection may be used but as the feature dimensions shrink to critical dimensions under 32nm, these techniques are not sufficiently accurate because optical tools cannot resolve the structures. Therefore e-beam tools such as CD-SEM (Critical Dimension-Scanning Electron Microscope) are used to obtain detailed images and the printable window is derived from a manual inspection of these images. This is however a time-consuming effort and it represents a bottleneck that slows down the mask development process. Within the printable window, a narrower process window may be determined, related for example to specific defects, i.e. defects which occur when the dose and focus are outside the limits of the process window. The determination of this narrow process window today takes place primarily through the manual comparison of CD-SEM images and is therefore equally a labor intensive and time-consuming effort. The printable window and the defect-related process window are examples of process limits needed for evaluating an OPC based mask design and/or for correctly printing a given set of features on a semiconductor wafer. Process limits are determined not only in terms of dose and focus applied in lithography processes, but also as a function of various other process parameters, such as etch parameters or parameters related to deposition or planarization processes. Today, the determination of these process limit is based only on CD measurement and not completely on pattern fidelity.

Our proposed approach provides an alternative automated method for determining process limits that does not suffer from the above described problems. We have divided the entire problem statement into 2 phases. In the first phase, our aim is to remove the extreme outliers (TNs, True Negatives) or non-printable features as depicted in Fig. 1 for pre-processing OPC data (data clean up). In the second phase, our aim is to exploit the same approach to automatically define defect process windows by providing a significant decision boundary between TPs (True Positives) and most importantly FPs (False Positives), where TPs are the printable features and TNs are non-printable features or erroneous features.

We have proposed a new method of analyzing the process window with an unsupervised machine learning approach. Our proposed algorithm will extract the underlying key features and encode these to latent feature vectors or latent vector space instead of the conventional CD, given a dataset of thousands of CD-SEM images, and then rank the images based on a similarity index and then to automatically determine the process window. The novelty of this idea lies in the fact that (1) with a defect inspection tool this task seems tedious and time consuming and often require human intervention to analyze a large number of features, (2) a CD-SEM based process window analysis might always not match with a defect inspection-based process window. Our generalized variational auto-encoder based approach does this automatically. Also, we have analyzed and validated our result against conventional tool approach, which is only CD-based. For 2D patterns

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

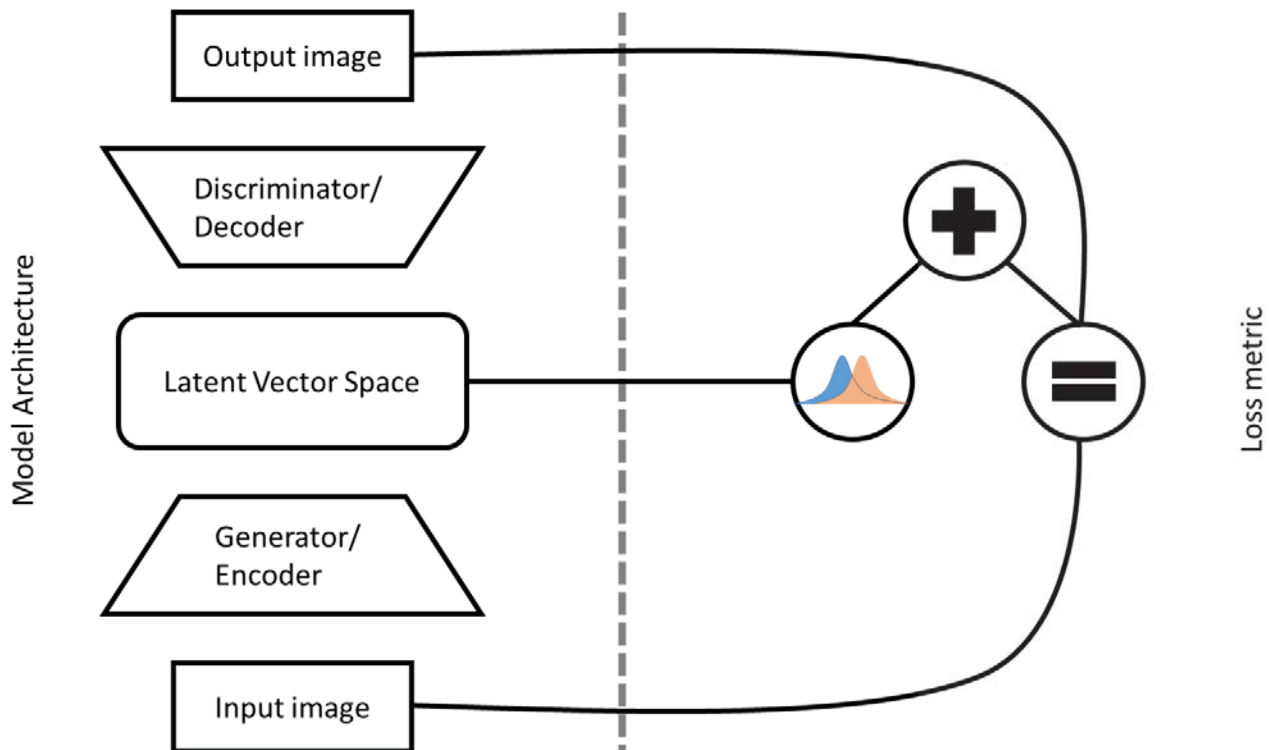


Figure 2: Generative Var-AE Model architecture.

this method becomes even more powerful.

The remainder of the paper is organized as follows. In section 2, we have summarized some related work and context. In section 3, We have discussed deep Generative Adversarial Networks (GAN) and variational autoencoder architecture. In section 4, we have shown the experimental results and analyzed and validated our proposed method against conventional techniques. In section 5, we have concluded the paper.

2. RELATED WORK

In this section, we briefly discuss existing research approaches and methodologies in the context of defect process window monitoring. In this paper, the authors have discussed and analyzed the context of the Process Window Discovery methodology and demonstrated the impact of EUV stochastics on the overall process window Ref. 1. Halder et. al. had carefully investigated the scope of performing future analysis with utilization of existing enormous review-SEM images for quantification with CCA and K-means algorithms Refs. 2, 3. In this research work, the authors have compared the full map of SEM overlay with a full map of optical overlay for high order correction in the context that CD-SEM could estimate asymmetry property of target from an image of pattern edge. They also evaluated the difference metrics between the via-pattern and relaxed pitch gratings Ref. 4. Bisschop et. al. in his research work Refs. 5, 6 had presented a quantification-metric for stochastic failure and investigated experimental parameters dependency as well as presented a reasonable analysis for prediction of stochastic CD-variability in the context of OPC-verification. Halder et. al. had proposed an innovative methodology to improve accuracy for evaluating process window on wafer. They have adopted "sectorization" technique to differentiate different AOI for loss-less analysis in the context of CD-SEM images Ref. 7. In this paper Ref. 8, the authors have discussed the context of higher defect detection sensitivity to identify precise process window and capabilities for automatic sub-categorization of defects. This research paper Ref. 9 demonstrated techniques to improve OPC metrology flow to optimize statistical process variation present in modern fabrication process flow.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

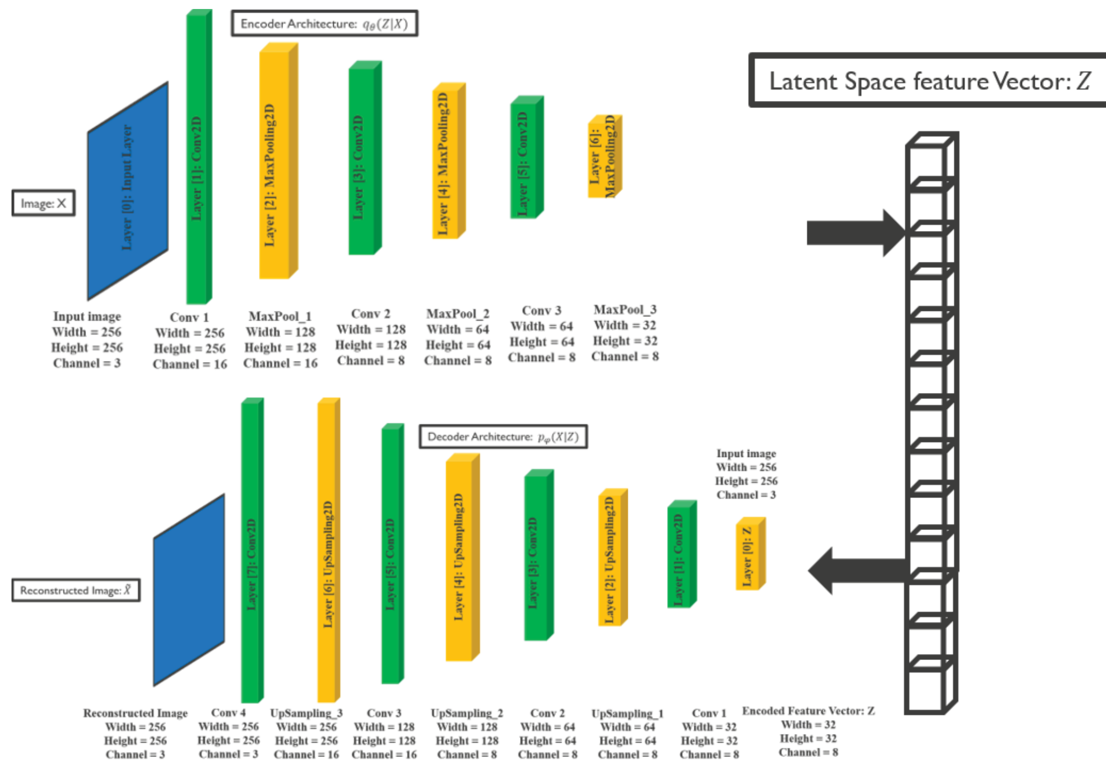


Figure 3: The Encoder model compresses image data into a latent vector space (Z). The Decoder model reconstructs the image ($\hat{X}$) given the learned hidden representations.

3. THE PROPOSED METHOD

Generative adversarial networks (GAN) are unsupervised or semi-supervised deep neural net architectures comprised of two architectures as shown in Fig. 2, competing one against the other Ref. 1, 10–13. A Variational-Autoencoder Refs. 1, 14, 15 in generative-adversarial network architecture perspective, consists of 2 different sub-architectures as (1) a generator/encoder and (2) a discriminator/decoder and a trade-off metric between these 2 as a loss-function. The encoder model is termed as a "bottleneck" as it must learn an efficient and significant compression of the data into it's corresponding lower dimensional stochastic space. The VAR-autoencoder architecture as shown in Fig. 3 can be summarized as follows:

1. An input image is passed through an encoder network.
2. The encoder outputs parameters of a distribution $q_\theta(Z|X)$, a Gaussian probability distribution
3. A latent vector Z is sampled from $q_\theta(Z|X)$. If the encoder learned to do its job well, most chances are Z will contain the information describing X.
4. The decoder decodes Z into an reconstructed image $\hat{X}$. On the R.H.S we have the loss:
   (a) Reconstruction error: the output should be similar to the input. Difference between $q_\theta(Z|X)$ and $p_\phi(X|Z)$.
   (b) $q_\theta(Z|X)$ should be similar to the prior p(Z) (multivariate standard Gaussian).
5. Loss function term with Kuulback-Leibler divergence as:
   $L(\theta, \phi) = -E_{(Z\sim q_\theta)(Z|X)}[log\, p_\phi(X|Z)] + KL(q_\theta(Z|X)||p(Z)$
   We train the variational autoencoder using gradient descent to optimize the loss w.r.t parameters θ and ϕ.

In order to generate new images, we can directly sample a latent vector from the prior distribution, and decode it into an image. Our VAR-autoencoder architecture has an Encoder/Generator model with 3 pairs of convolutional-maxpool layers.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

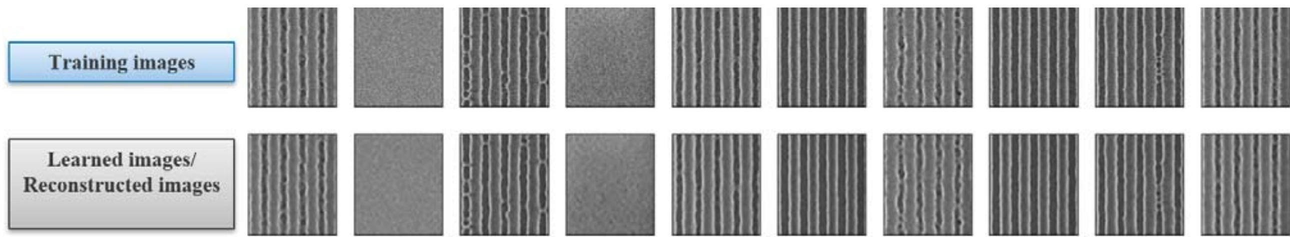
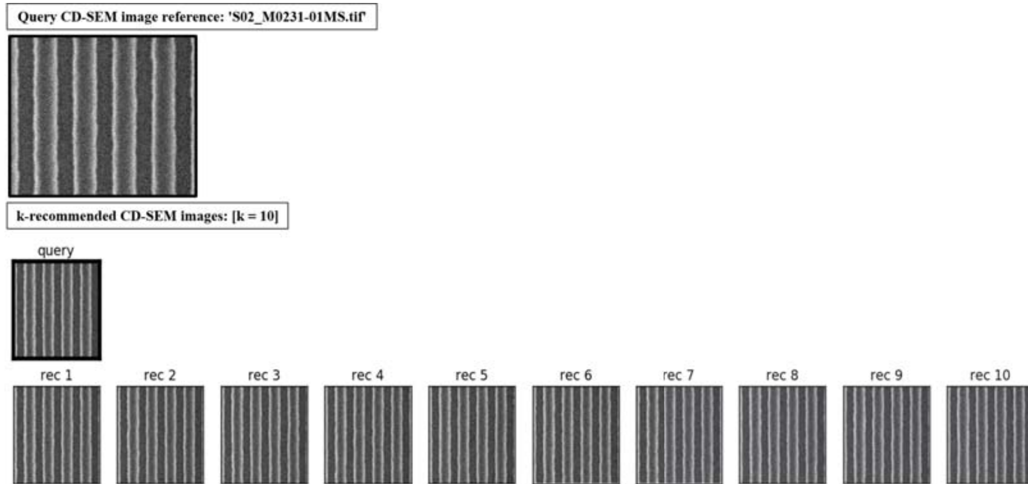
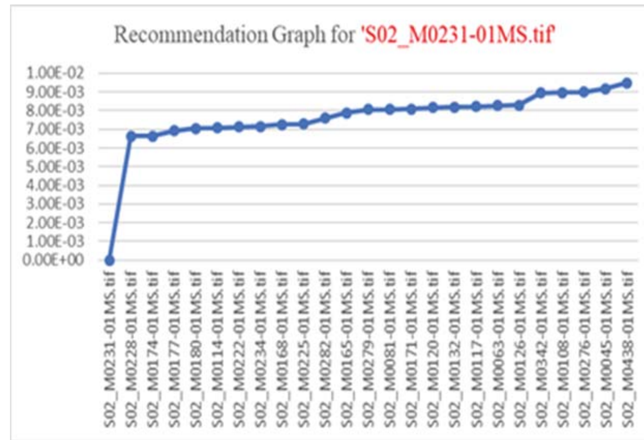


Figure 4: (a) Training images, (b) Learned/Reconstructed images by the model.



(a)

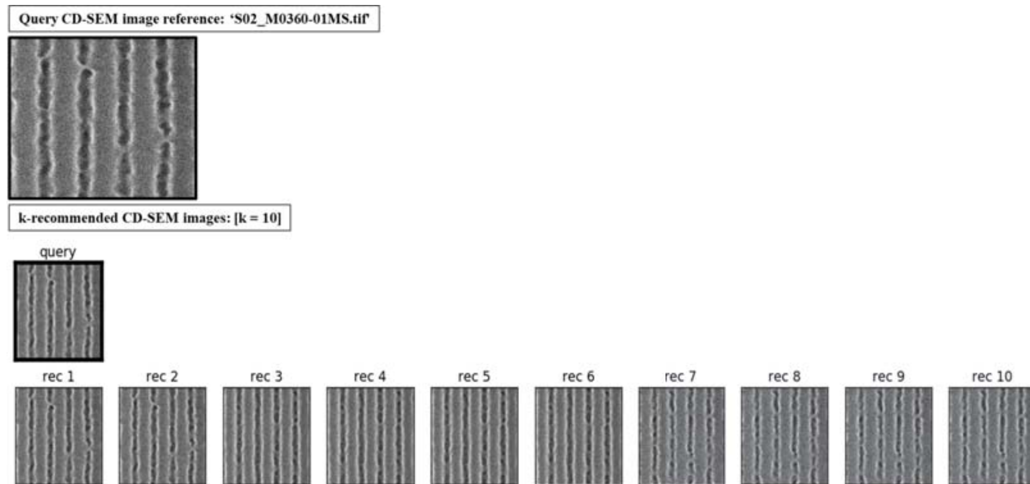


(b)

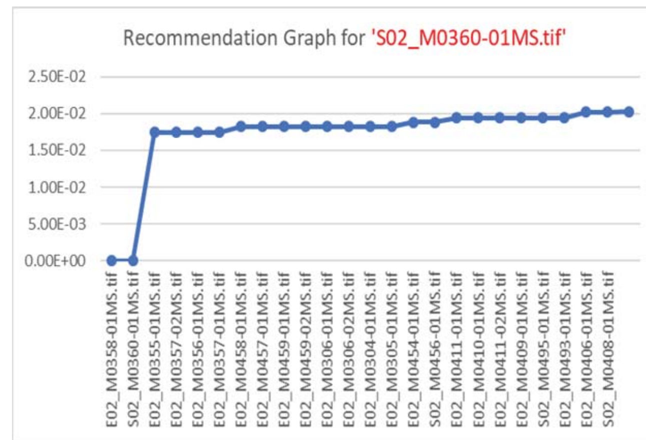
Figure 5: (a) Classification and Recommendation for CD (line-space) image pattern at dose: 42.0, focus: -0.09. The first 10 best recommended similar printable CD-SEM patterns(images) has been shown out of 230 diverse datasets of CD-SEM images. (b) Recommendation graph-plot for "query-pattern" index S02-M0231-01MS .

The Decoder model has 4 convolutional layers and 3 up- sampling layers. We have used a convolutional kernel and a pool kernel. The encoder encoded each input image of size into a latent space vector of size . The decoder took the learned internal representations as latent space vector of that image using the learned weights and reconstructed the image back as . We have used "ReLU" as activation function and kept the padding as "same". We have a total of 4,963 trainable parameters.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use



(a)



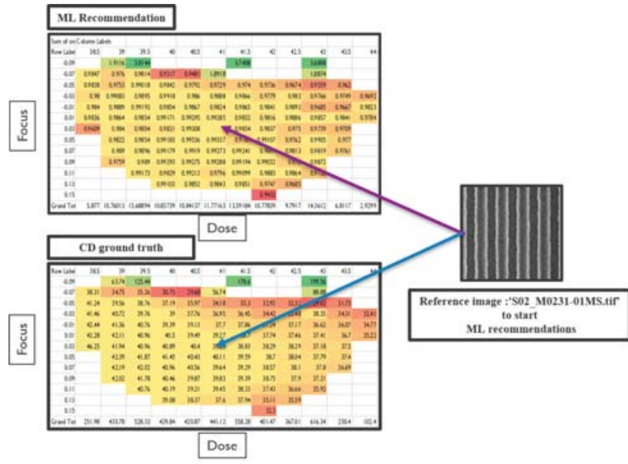
(b)

Figure 6: (a) Classification and Recommendation for CD (line-space) image pattern at dose: 42.0, focus: -0.09. The first 10 best recommended similar non-printable CD-SEM patterns(images) has been shown out of 230 diverse datasets of CD-SEM images. (b) Recommendation graph-plot for "query-pattern" index S02-M0360-01MS.

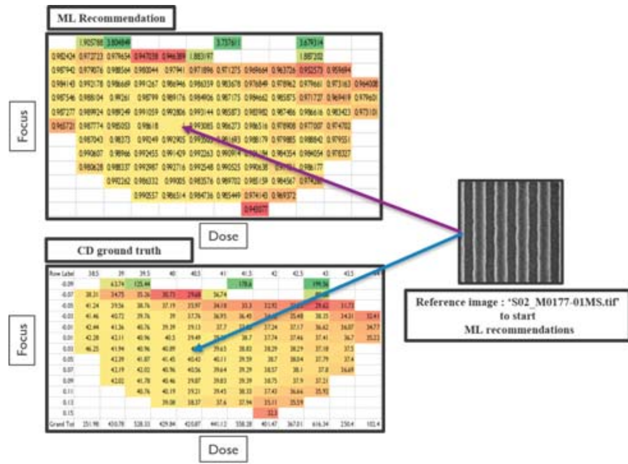
4. THE IMPLEMENTATION AND EXPERIMENTAL RESULT

Our proposed approach is based on 2 major steps. In the 1st step, we have trained a prototype of Conventional Variational-AE Refs. 14, 15 with a set of 230 FEM wafer images for training and testing purposes. The dataset was divided into a training set, a validation set and a test set with a percentage ratio of [60:20:20]. The training set consisted of 187 CD-SEM images and the test set consisted of 43 CD-SEM images for different dose and focus. We took the validation set by random shuffling the CD-images both from test and train set for each epoch. We also implemented data-augmentation techniques (rotation at 90 degrees, rotation at finer angels etc.) to balance/increase the diversity of training dataset image patterns. The model was trained with Intel Core i5-8350U CPU 1.70GHz node. The VAR-AE model was trained using Keras library Ref. 16 and the Tensorflow library Ref. 17 backend in the python programming environment. The Anaconda version was 4.6.8. We reiterated the training process with variable batch-size ranging from 8 to 256 Ref. 18. All CD-SEM images were normalized to have values in the range [0, 1]. We choose Adam optimization technique as stochastic gradient decent algorithm and binary-cross-entropy as loss-metric Refs. 19, 20. The total number of trainable parameters were 4,963. Once training step has been completed with minimum reconstruction loss, we saved the model and learned weights for later use. We have extracted the saved learned encoder part of the model to extract/generate the representative features or embedded characteristics of the CD-SEM images (latent vector space encoding) during inference as shown in Fig. 4.

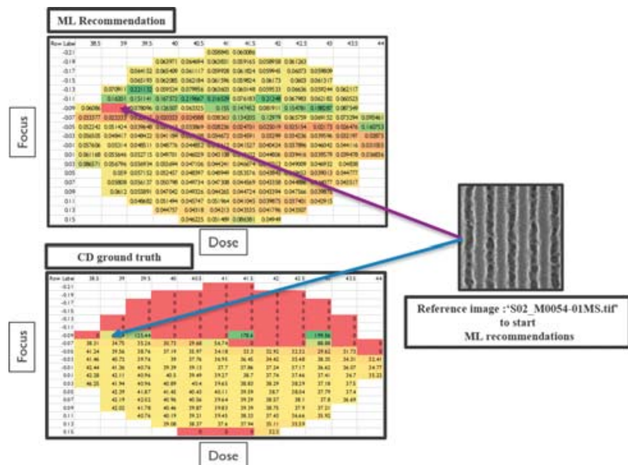
Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use



(a)



(b)



(c)

Figure 7: Wafer-map representation for Conventional technique metric vs. proposed ML recommendation for (a) a very good qualitative Process-Window (best printable feature CD-SEM image). (b) for a near good qualitative Process-Window (a random near best printable feature CD-SEM image).(c) for a worst qualitative Process-Window (random non-printable feature CD-SEM image).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

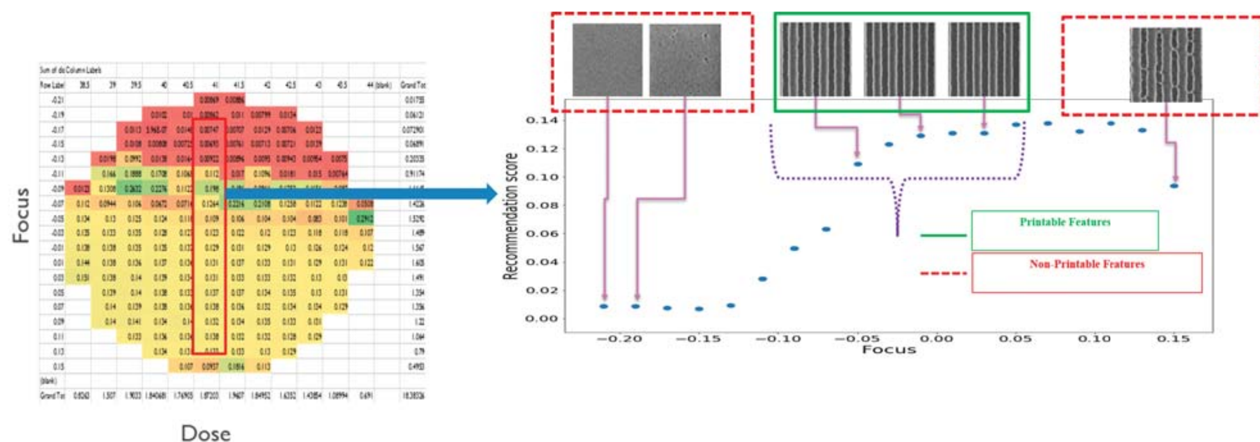


Figure 8: Qualitative Process-Window analysis for a reference Dose-Column: 41.0 with varying focus.

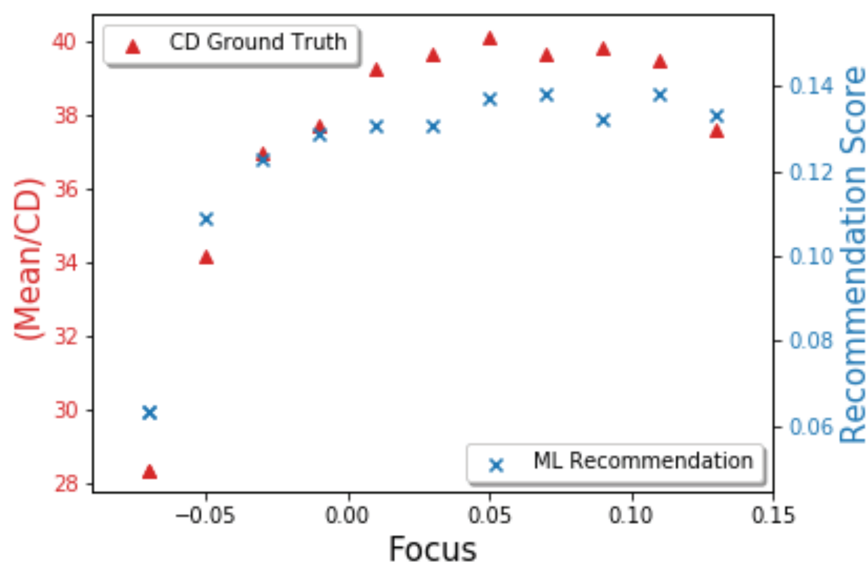


Figure 9: Plot for Normalized Mean/CD-ground truth value against proposed ML-Recommendation score for the selected Dose column: 41.0 mJ/cm2 against varying focus.

In the second step, we have performed similar pattern retrieval or similar CD-SEM image retrieval on given test database of images by implementing well-known KNN algorithm Ref. 21 on the above extracted encoded feature space. We have experimented with 5 popular different distance metrics as Cosine-Similarity, Minkowski Distance, Manhattan Distance, Euclidean distance and Jaccard Distance and found the context of choosing the Cosine-Similarity as distance metric for KNN learning fits the best as shown in Table. 1.

We believe the training has been performed with limited resources and with very limited scope, so further careful analysis and improvement of hyperparameters (both the model architecture and area-of-subject-research) may significantly improve the scope of training process. We have termed any test-case "Unlabeled" CD-SEM image patterns as a "Query-image" or a "Query-pattern" against rest of unseen diverse dataset of other CD-SEM images/patterns and our proposed approach has ranked them based on most "closer" or "similar" learned encoded-feature metric in an unsupervised way of training. In other words, our proposed approach allows us to extract the process limits directly from the score/rank values rather than from a manual comparison of the images. Fig. 5 and Fig. 6 have shown 2 different sets of classification-[1: (N-1)] recommendation pair for respective printable "query" CD-SEM image pattern (indexed here as S02-M0231-01MS.tif) and non-printable "query" CD-SEM image pattern (indexed here as S02-M0360-01MS). The respective recommendation

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Table 1: NORMALIZED DISTANCE METRIC FOR 5 DIFFERENT DISTANCE FUNCTION FOR RANDOMLY SELECTED 10 CD-SEM IMAGES

Image name	**Euclid dist norm**	**Manhattan dist norm**	**Jaccard dist norm**	**Cos sim dist norm**	**Minkow dist norm**
E02-M0022-01MS.tif	1.014483	1.074927	1.000136	1.008405	0.980607
E02-M0052-01MS.tif	1.046246	1.060798	1.000136	1.00414	1.01445
E02-M0055-01MS.tif	1.047701	1.086538	1.000136	1.013524	1.004681
E02-M0056-01MS.tif	1.047701	1.086538	1.000136	1.013524	1.004681
E02-M0057-01MS.tif	1.047701	1.086538	1.000136	1.013524	1.004681
E02-M0057-02MS.tif	1.047701	1.086538	1.000136	1.013524	1.004681
E02-M0058-01MS.tif	1.008539	1.053082	1.000136	1.013524	0.986324
E02-M0091-01MS.tif	1.036588	1.045348	1.000136	1.005846	1.012918
E02-M0092-01MS.tif	1.036588	1.045348	1.000136	1.005846	1.012918

Table 2: NORMALIZED GROUND-TRUTH MEAN/CD OBTAINED FROM CONVENTIONAL TECHNIQUES

Dose	**Focus**	**Mean-CD-norm**	**Delta-CD**
39	-0.09	0.830597	-0.20276
39.5	-0.09	0.817305	-0.21605
39.5	-0.09	0.817305	-0.21605
39.5	-0.09	0.817305	-0.21605
39.5	-0.09	0.817305	-0.21605
41	-0.07	0.73938	-0.29398
41.5	-0.09	1.16367	0.13031
41.5	-0.09	1.16367	0.13031
41.5	-0.09	1.16367	0.13031

variations have been also depicted in Fig. 5 (b) and Fig. 6 (b).

Fig. 7 shows the wafer-map representation for 3 different case-scenario, where we have extracted the process-limits directly from the obtained score-values/ranking metric and plotted against conventional CD-ground truth (obtained from conventional tool approach).Obtaining a score value representative of a given image characteristics is enabled by a targeted choice of one or more of the query images and object images, and/or by determining the score as an aggregate score value, calculated from the scores obtained with respect to a set of different "query" images. We have shown the ground-truth Mean/CD values obtained from the conventional tool approach for random 10 CD-SEM images in Table. 2. The aggregate score may be the maximum or the average of the scores of the object images, obtained with respect to a set of "query" images. From Fig. 8 and Fig. 9, supports our claim that our proposed unsupervised machine learning approach can classify and segregate printable CD-SEM images/patterns from the non-printable CD-SEM images/patterns and extract the probable process limits directly from the score values, rather than from a manual comparison of the images. Fig. 9 also depicts that the scores furthermore closely follow the variations of the CD beyond sharp rise. This result shows that it is possible to determine a CD-related process-window (i.e. a narrower window within the limits of the printable window) based on the score values alone, provided a preliminary calibration is performed.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

5. CONCLUSION AND FUTURE WORK

The proposed method provides a way of automatically detecting the probable printable window. We have merged concepts of Unsupervised Learning and Deep-Generative-Network to first learn the characteristic features for a given dataset of numerous different CD-SEM images and then ranked them based on a numeric score-metric, derived from comparison of the images in the latent feature space(1:N-1). When the query image is chosen judiciously, the score of the object images exhibits a monotonic relation to CD of the images. The method becomes more relevant for complex 2D shapes/patterns where only CD-measurement is not enough.Hence, our proposed deep unsupervised learning approach offers an effective method for deriving the printable window from the score values, without necessitating the actual CD-measurements, by detecting the presence of a step in the relation between the score and the focus and/or dose. Our future goal is to improve the model accuracy and performance by fine-tuning different hyper-parameters to clearly distinguish small variations by sensing context-pixel differences only and to look for stochastic defects during review of defect locations.

6. ACKNOWLEDGEMENTS

The authors used the imec Leuven research computing facility to conduct part of the research. We thank our colleagues who provided insight and expertise that helped us to improve the manuscript.

REFERENCES

[1] Goodfellow, I., Bengio, Y., and Courville, A., [*Deep Learning*], The MIT Press (2016).

[2] Halder, S., Leray, P., Sah, K., Cross, A., and Parisi, P., "Connected component analysis of review-SEM images for sub-10nm node process verification," in [*Metrology, Inspection, and Process Control for Microlithography XXXI*], Sanchez, M. I., ed., **10145**, 604 – 610, International Society for Optics and Photonics, SPIE (2017).

[3] Halder, S., Cerbu, D., Saib, M., and Leray, P., "Review-sem image analysis with k-means algorithm: Am: Advanced metrology/di: Defect inspection," in [*2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*], 255–258 (April 2018).

[4] Leray, P., Halder, S., Lorusso, G., Baudemprez, B., Inoue, O., and Okagawa, Y., "Hybrid overlay metrology for high order correction by using CDSEM," in [*Metrology, Inspection, and Process Control for Microlithography XXX*], Sanchez, M. I., ed., **9778**, 691 – 698, International Society for Optics and Photonics, SPIE (2016).

[5] Bisschop, P. D. and Hendrickx, E., "Stochastic printing failures in EUV lithography," in [*Extreme Ultraviolet (EUV) Lithography X*], Goldberg, K. A., ed., **10957**, 37 – 56, International Society for Optics and Photonics, SPIE (2019).

[6] Bisschop, P. D., "Stochastic effects in EUV lithography: random, local CD variability, and printing failures," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **16**(4), 1 – 17 (2017).

[7] Halder, S., Mailfert, J., Leray, P., Rio, D., Peng, Y.-H., and Laenens, B., "Design-based metrology: beyond CD/EPE metrics to evaluate printability performance," in [*Metrology, Inspection, and Process Control for Microlithography XXX*], Sanchez, M. I., ed., **9778**, 290 – 300, International Society for Optics and Photonics, SPIE (2016).

[8] Leray, P., Halder, S., Lorenzo, P. D., Wang, F., Zhang, P., Fang, W., Liu, K., and Jau, J., "Study of design-based e-beam defect inspection for hotspot detection and process window characterization on 10nm logic device," in [*Metrology, Inspection, and Process Control for Microlithography XXX*], Sanchez, M. I., ed., **9778**, 234 – 240, International Society for Optics and Photonics, SPIE (2016).

[9] Ward, B. S., Berthiaume, S., Brist, T., and Brooker, P., "Metrology data cleaning and statistical assessment flow for modeling applications," in [*Metrology, Inspection, and Process Control for Microlithography XXIV*], Raymond, C. J., ed., **7638**, 646 – 656, International Society for Optics and Photonics, SPIE (2010).

[10] LeCun, Y., Bengio, Y., et al., "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks* **3361**(10), 1995 (1995).

[11] Karpathy, A., "Stanford university cs231n: Convolutional neural networks for visual recognition," *URL: http://cs231n. stanford. edu/syllabus. html* (2018).

[12] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *nature* **521**(7553), 436 (2015).

[13] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).

[14] Tschannen, M., Bachem, O., and Lucic, M., "Recent advances in autoencoder-based representation learning," (2018).

[15] Kingma, D. P. and Welling, M., "An introduction to variational autoencoders," (2019).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

[16] Chollet, F. et al., "Keras: Deep learning library for theano and tensorflow," *URL: https://keras. io/k* **7**(8), T1 (2015).

[17] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al., "Tensorflow: A system for large-scale machine learning," in [*12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*], 265–283 (2016).

[18] Ioffe, S. and Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167* (2015).

[19] Bottou, L., "Large-scale machine learning with stochastic gradient descent," in [*Proceedings of COMPSTAT'2010*], 177–186, Springer (2010).

[20] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).

[21] Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K., "Knn model-based approach in classification," in [*OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*], 986–996, Springer (2003).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 26 Mar 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use