Bayesian dropout approximation in deep learning neural networks: analysis of self-aligned quadruple patterning

Halle, Scott, Bloomfield, Max, Gabor, Allen, Dunn, Derren, Shephard, Mark

Scott D. Halle, Max O. Bloomfield, Allen H. Gabor, Derren Dunn, Mark Shephard, "Bayesian dropout approximation in deep learning neural networks: analysis of self-aligned quadruple patterning," Proc. SPIE 11329, Advanced Etch Technology for Nanopatterning IX, 113290B (6 May 2020); doi: 10.1117/12.2551649

SPIE.

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

Bayesian Dropout Approximation in Deep Learning Neural Networks: Analysis of Self-Aligned Quadruple Patterning

Scott D. Halle[a], Max O. Bloomfield[b], Allen H. Gabor[a], Derren Dunn[a], and Mark Shephard[b]

[a]IBM TJ Watson Research Center, 257 Fuller Road, Albany, NY,12203 USA;
[b]Rensselaer Polytechnic Institute, 110 8th St, Troy, NY 12180, USA

ABSTRACT

Predictive modeling of the pitch-walk variance from multistep coupled processes, such as SAQP using experimental metrology observables, has the potential to give both deep understanding and a control mechanism for pitch-walk variance. In this study, with the Bayesian dropout approximation, a methodology using Bayesian inference via use of stochastic neural networks was employed to both model and predict the SAQP pitch-walk variance distribution. Bayesian neural networks were implemented as variational ensembles of networks with hidden layers, where the neural net training uses conventional dropout, while the forward solves employ a dropout Bayesian vector methodology previously developed by Gal and Ghahramani.[1,2] An important distinction here is that the forward propagations effectively sample the network to make a prediction, resulting in a distribution of outputs achieving the best model, not just a single expectation value. A complete dataset of fin module OCD metrology measurements per chip at top mandrel, bottom mandrel, and final fin reveal were used. Since the measured dataset was limited to small number chip locations, data augmentation with the highly efficient method of the volume of simplex was used to generate 30K samples. The synthetic data and the experimental data were used for neural network calibration and validation, respectively.

Keywords: pitch walk, self-aligned quadruple patterning (SAQP), deep learning neural nets, bayesian dropout approximation

1. INTRODUCTION

Self-aligned Quadruple Patterning (SAQP) is a a method for enabling sub-lithographic patterning that has been extensively discussed in the literature[3] and is actively employed in semiconductor manufacturing of finfet devices.[4] This process employs multiple non-lithographic sidewall spacer image transfers to reduce the structure to a quarter of the original lithographic pitch. The numerous process steps are required for SAQP making it susceptible to process errors which can propagate through each deposition and etch step, which can result in unwanted variations in the final step structure. The difficulty of controlling the complicated process sequence of SAQP has been noted by many authors[3,5,6] . Given the complexity of this coupled multistep process simply relying on metrology specifications may be insufficient to guard against unacceptable process variations of neighboring fin space widths, commonly referred to as pitch walk. The complexitity of pitch walk in SAQP is extensively discussed in Section 3. This work presents a model that predicts the fin space widths variations at defined stages in the process flow. The application of deep learning neural networks (DNN) is the predictive modeling methodology of choice in this study. One advantage of a predictive model for the downstream fin pitch-walk outcome is the ability to flag wafers/lots early in the process flow. If the predicted pitch-walk value exceeds a targeted threshold, there is potential for reworking the wafers. The ability to early identify individual chips or regions of a wafer with fin pitch-walk excursions would be invaluable for either pre-sorting the chips in wafer or preventing unnecessary costly processing in the fin process module. A model that only gives a singular value for the predicted pitch-walk has usefulness; however, its inability to give the error (*i.e.*or precision) of the pitch-walk prediction is a major limitation. The ability to early predict not just a single value fin pitch-walk from a model, but also to give the uncertainty of pitch walk prediction at a given site on the wafer, would give a very large boost to the confidence in the predictive outcome. This power of prediction uncertainty would make the deployment of this model much more feasible. In this work, we show such a methodology which allows for the prediction

Send correspondence to halle@us.ibm.com

Advanced Etch Technology for Nanopatterning IX, edited by Richard S. Wise, Catherine B. Labelle, Proc. of SPIE Vol. 11329, 113290B · © 2020 SPIE · CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2551649

of both the mean and uncertainty of the pitch walk using a novel approach within the inference model of a deep learning neural network. This feedfoward-solve DNN inference methodology, called the Bayesian Dropout Approximation, developed by Gal and Ghahramani,[1] will be explored in Section 2.1. While this methodology is applied here to early predict pitch walk, we note that this approach has a broad applicability to model many complex problems related to devices and semiconductor processes.

2. METHODOLOGY FOR SAQP DEEP LEARNING NEURAL NETWORKS

Chao *et al.*[4] used optical critical dimension (OCD)/optical scatterometry data to create a calibrated SAQP model, using a data feed-forward approach and verification by reference metrology. That work enables the extraction of precise 2D measurements of the stack geometric parameters used in this study. The OCD measurements were obtained at five discrete process steps in the SAQP process (as described in Section 3). The OCD dataset of approximately 30 wafers was collected periodically from a stable process route under active process control without any experimental splits. For each of the 30 wafers undergoing metrology, 20 sites were measured per wafer. At each process step, a number of geometric stack parameters were extracted. For modeling purposes we considered 16 parameters, collected from the five measurement steps in the OCD dataset. The complete dataset included only wafers in which all sites were available for all five OCD measurements (*i.e.*scrapped wafers were not included). Additional culling of the data included filtering with a goodness-of-fit parameter threshold. We shall discuss in Section 3 more about parameters included in the model for each of the five measurement steps in the process.

The framework for the modeling of the SAQP process is a fully connected deep learning neural network (DNN). The inputs to the DNN correspond to different geometric parameters extracted from the dataset of OCD measurements. Two pitch-walk metrics (OCD parameters of the fin space-width differences) define the outputs of the network. The network topology for modeling SAQP will be described later in Section 4. For the SAQP DNN, hyperparmeters such as the learning rate are optimized and a dropout probability of 0.15 is used in the training of the DNN. Originally introduced by Hinton *et al.*,[7] dropout has enjoyed success as a algorithm [8,9] for regularizing DNN models. Dropout is typically applied during training, with the output of a randomly selected set of nodes being set to zero or "dropped out", and with a new set of node outputs being selected after each backpropogation step. This zeroing out has the effect of changing the topology of the network connections with each update to the network. Dropout is implemented on a layer-by-layer basis and can be applied to the outputs of any of the hidden layers. While dropout may slow down the training, its advantage is to reduce overfitting and to help improve model accuracy on validation and test datasets. Next we discuss the methodology employed to augment the training data of the network from the OCD dataset.

Data augmentation is a well-known strategy to increase the size of the training dataset of the neural net where a larger data set is not available, without collecting new data. Parameter optimization of the neural network may require larger amounts of data with an increasing number of model parameters to prevent overfitting in model training. For example, data augmentation of images with convolutional neural nets typically employ techniques that make small changes to the existing data set by symmetry operations such as translation, rotation, or flipping about an axis. For application domains with limited experimental data and a large number of parameters, augmentation techniques may be required. For this limited measurement data, we used the simplex method to generate 30K interpolated data points from the original 560 experimental results. To enable a linear interpolation within a unique higher-dimensional parameter space, the simplex method allows interpolation on non-gridded data points. A simplex is defined as a *n+1* vertex polytope embedded in n dimensions. Simplices in the n input dimensions use *n+1* data points. A set of barycentric coordinates are chosen and the outputs are interpolated for each vertex.

2.1 Bayesian Dropout Approximation

Standard deep learning neural networks do not capture model uncertainty. However; the lure of Bayesian neural networks to propogate a distribution of weights through the network with the potential to give model uncertainty, has attracted many researchers over the last 30 years. Neal[10] and Williams[11,12] both explored the relationship between DNN and the Gaussian process, showing that as a Gaussian probability distribution is placed over each weight in a network with an infinite number of weights, a Gaussian process is recovered. MacKaye[13] and

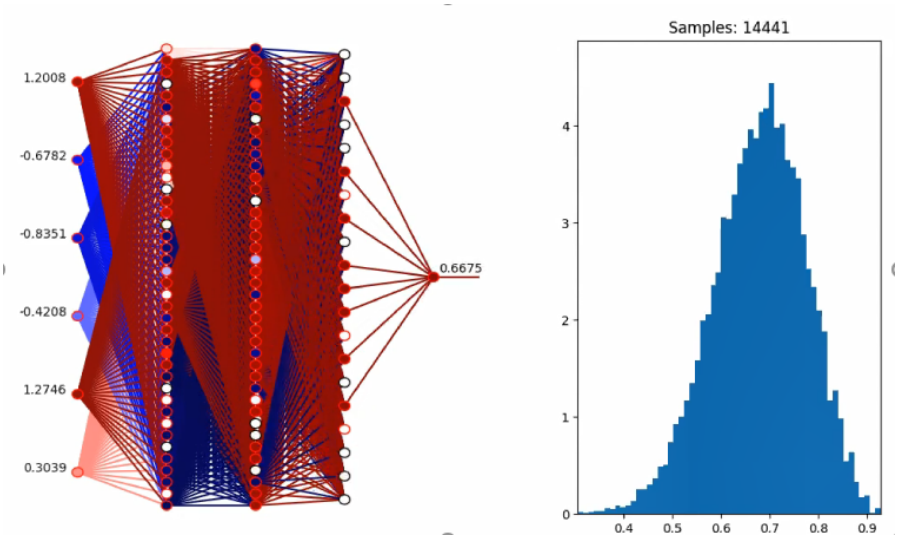

Figure 1. Shown is a DNN that illustrates the feedforward-solve inference with the Bayesian dropout approximation and its associated outcome histogram of the probability density. This example DNN has the following network topology: 6 input layers, 3 hidden layers, 1 output layer The dropouts of the nodes are colored white, while the fully connected active layers are colored red. In this example, an arbitrary number of repititions of the feedfoward solve is 14441.

others[14] have explored a modified Gaussian process with a finite number of weights, known as a Bayesian neural network, which yields the uncertainty by placing a distribution over the weights. To address the mathematical intractability of implementing a Bayesian neural network, Graves[15] and Hinton *et al.*[16] applied variational inference to Bayesian neural networks with limited success. Advances in variational inference techniques[17] have also been implemented on Bayesian neural networks; however, these approaches suffer from prohibitive computational costs to represent uncertainty.

This study exploits a new theoretical framework for interpreting dropout, developed by Gal and Ghahramani[1,18,19] . The mathematically grounded insight of the Bayesian dropout approximation shows that dropout applied before every weight layer in DNNs is equivalent to approximate Bayesian inference in deep Gaussian processes. Some concerns about BDA have been addressed by Hron *et al.*? This insight of BDA leads to a methodology for computing model uncertainty using dropouts in the DNN that is computationally fast and does not degrade test accuracy. For a network trained with dropout, this approach yields the predictive mean and the confidence (ie. the predictive variance) in the model predictions. Since the dropout network is simply a Gaussian process approximation, the model precision can be calculated as the Gaussian process precision, according to the prescriptive of Gal.[19] The network (inference) output, under BDA, is generated by a feedforward-solve while dropping out random nodes from the trained network. Forward passes are solved over a number of repetitions, with different nodes dropped each time to obtain a probability density distribution. A visualization of the Bayesian dropout approximation, equivalent to a Bayesian Neural Network, with the outcome of the probability density distribution is shown in Figure 1. The nodes in Figure 1 are fill colored to show the following different behavior at the nodes: probablistically dropouts (white), active non-zero weights (negative and postive weights for blue and red, respectively), and dropped nodes due effectively zero weight (black). Before presenting the results of the application of this feedfoward-solve inference model to the DNN, it is necessary to gain insight into the self-aligned quadruple patterning.

3. SELF-ALIGNED QUADRUPLE PATTERNING

The output layers of the SAQP DNN are the pitch-walk parameters defined by the space-width differences of neighboring fins. These pitch-walk space-width differences can be defined by a geometrical SAQP model as shown in Figure 2. A geometrical SAQP model both simplifies the description of the process steps and describes how the parameters such as mandrel widths and the space-width of the mandrels can influence the space-width differences

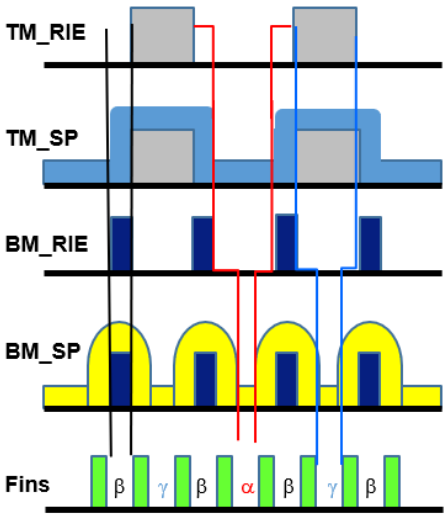

Figure 2. This shows the process steps for SAQP

in the fin structures. The elucidation of the geometrical SAQP model is important not only to provide a definition of the fin space-width differences, but is required for the interpretation of the different network topologies of DNN based on different process steps parameters. In addition, as previously mentioned, the input dataset for the SAQP DNN is based on measurements of the geometric parameters from optical scatterometry. However, it should be emphasized that because the SAQP DNN is not a physics-based model, geometrical rules and relationships are not built into the DNN.

Although the modern FIN process module has a multitude of steps, we simplify the module description into five distinct steps that correspond to the OCD measurement steps of this dataset. These five steps, shown in Figure 2, include the following: 1) the top mandrel after lithography and RIE etching (TM_{RIE}), 2) the top mandrel post spacing deposition (TM_{SP}), 3) the bottom mandrel, post mandrel pull and RIE etching (BM_{RIE}), 4) the bottom mandrel post spacer deposition (BM_{SP}), 5) the final fin formation at fin reveal (FIN). Chou *et al.*[4] and Kagalwala *et al.*[20] have previously described the nearest neighbor spacing output from the fins with three spacing parameters. For the spacing parameters denoted as α, β, and γ, the unique dependencies on the discrete processing steps are now elaborated. The definition of pitch walk used in the output layer of the DNN is the space-width differences $\alpha - \beta > \epsilon$ and $\alpha - \gamma > \epsilon$, where ϵ is a technology-defined non-zero threshold value. At each process step, increasing or decreasing the mandrel width or spacer width may have an impact on the fin space-widths $\alpha - \beta$ and $\alpha - \gamma$. From Figure 2, we visually see that the variation of geometrical parameters in both the top and bottom mandrels have a significant effect on the pitch walk defined by $\alpha - \beta$. We shall next examine the role of top and bottom mandrel parameters on $\alpha - \gamma$.

From the geometrical SAQP model shown in Figure 2, we can define the relationship of increasing or decreasing the mandrel width and mandrel space-width at each process step on the fin space-widths α, β and γ. Table 1 explicitly shows that an inverse relationship with α is seen for the following parameters; the top mandrel width (TM) in TM_{RIE}, the width between the spacers (SP_{TM}) in TM_{SP}, the bottom mandrel width (BM) in BM_{RIE}, and the width between the spacers (SP_{BM}) in BM_{SP}. Since the spacer deposition process in both top and bottom mandrels has extremely tight process control, the impacts on the α space are expected to have less impact for all fin space-widths. Table 1 shows that the fin space-width γ increases as the top mandrel width (TM) in TM_{RIE}increases. However, an inverse relationship is shown in Table 1 for the fin space-width γ with the bottom mandrel width (BM) in BM_{RIE}, and the width between the spacers (BM_{SP}) in BM_{SP}. The width between the spacers in (TM_{SP}) does not have an impact on γ. Finally, the relationship between fin space-width β is examined in Table 1. As both the width between the spacers (SP_{TM}) in $TM_{SP}$and the bottom mandrel width (BM) in BM_{RIE}increase, the value of β should increase. In contrast, an increase in both the top mandrel width (TM) in TM_{RIE}, and the width between the spacers (SP_{BM}) in BM_{SP}should have no impact. Increasing either the

Table 1. The table shows the relationships between the process steps TM_{RIE}, TM_{SP}, BM_{RIE}, and $BM_{SP}$and the SAQP spacing parameters of α, β, γ to the TopMandrel width (TM), top mandrel spacer width (SP_{TM}), the bottom mandrel width (BM), and the bottom mandrel spacer width (SP_{BM}). The upward and downward directions of the vertical arrows indicates increasing and decreasing values, respectively of the geometrical parameters. A horizontal arrow indicates no dependency.

Process Step	α space	γ space	β space
TM_RIE	TM↑ → α↓	TM↑ → γ↑	TM↑ → β↔
TM_SP	SP_{TM}↑ → α↓	SP_{TM}↑ → γ↔	SP_{TM}↑ → β↑
BM_RIE	BM↑ → α↓	BM↑ → γ↓	BM↑ → β↑
BM_SP	SP_{BM}↑ → α↓	SP_{BM}↑ → γ↓	SP_{BM}↑ → β↔

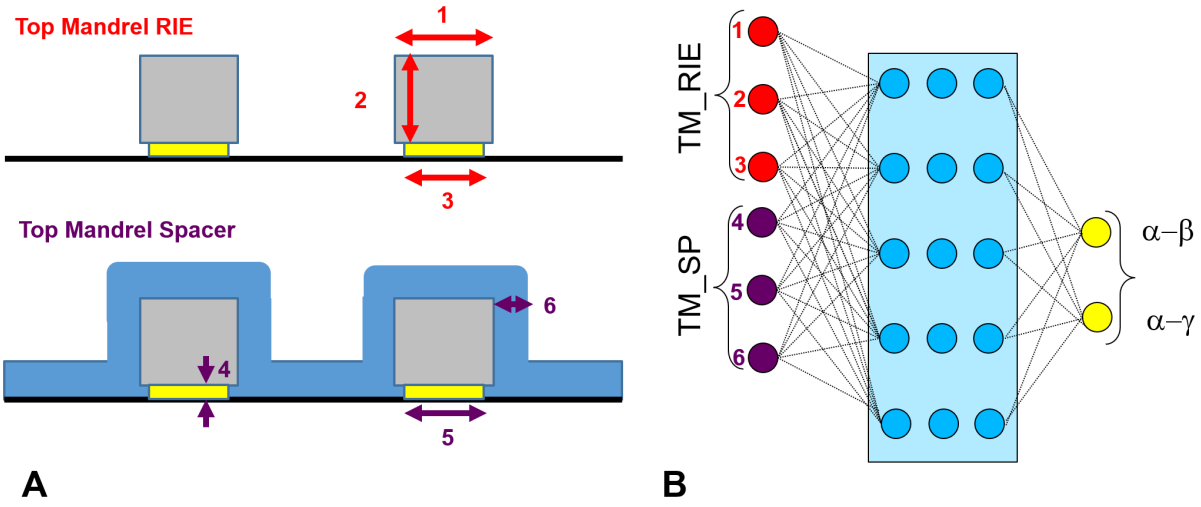

Figure 3. This shows the stack geometry parameter definitions for the top mandrel for the six-parameter network

bottom mandrel width (BM) in BM_{RIE}, or the width between the spacers (SP_{BM}) in BM_{SP}, is found to both decrease α and increase γ. Since α and γ areanti-correlated in the bottom mandrel, the bottom mandrels have a marginal impact on this fin space-width difference whereas the top mandrel features have a direct influence on $\alpha - \gamma$. This is in contrast to the previously stated pitch-walk parameter $\alpha - \beta$, where the variation of geometrical parameters in both the top and bottom mandrels can have a significant effect.

4. NETWORK TOPOLOGY FOR SAQP

Previous models of SAQP[4,21] have employed feed-forward process-sequence ordered inputs. In contrast, in this network topology, any number of geometric parameter input layer nodes from any OCD measurement step, independent of process-sequence order, can be constructed. However, for assessing the contributions of parameters from the top and bottom mandrels, a few networks are constructed. Three different n-parameter networks are constructed with the following number of input layer nodes: 6, 10, and 14, from top and bottom mandrel parameters. Using python-based TensorFlow,[9] these different networks can be fully trained with at least 3 hidden layers. The output layers for all these networks are the pitch-walk metrics $\alpha - \beta$ and $\alpha - \gamma$. For all three SAQP networks in this study, the number of nodes in the 3 hidden layers are 100, 100, and 50, respectively. Next, the descriptions for the three networks are given.

The six-parameter network contains inputs only from the top mandrel, with 6 geometric parameters from $TM_{RIE}$and TM_{SP}. Figure 3A shows the top mandrel stack at TM_{RIE}, where the geometric stack parameters 1,

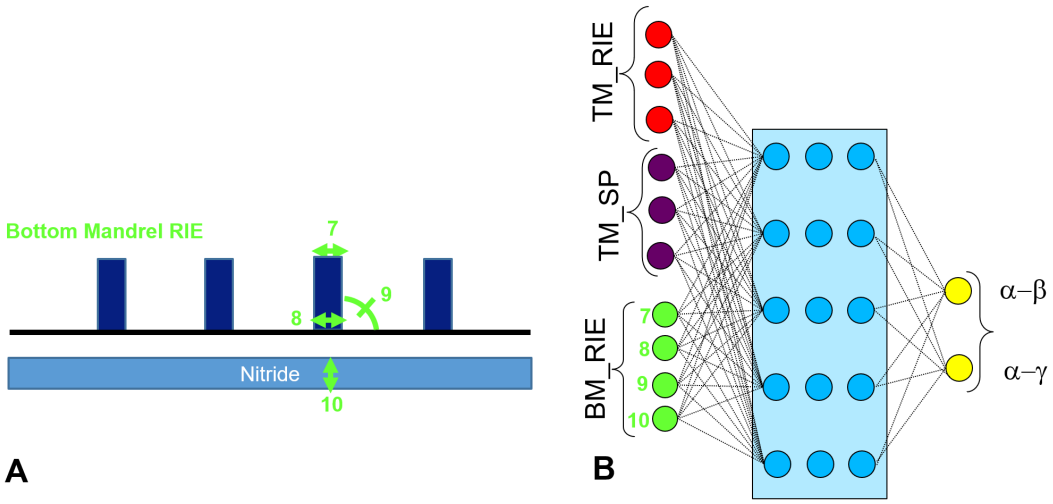

Figure 4.
This shows the stack geometry parameter definitons for the bottom mandrel $BM_{RIE}$for the ten-parameter network

2, and 3 correspond to top mandrel top width, top mandrel height and undercut hardmask layer bottom width, respectively. Figure 3A also shows the top mandrel stack at TM_{SP}, where the geometric stack parameters 4, 5, and 6 correspond to undercut hardmask layer height, undercut hardmask layer bottom width, and the sidewall spacer width along the top mandrel. The illustration of the six-parameter network is shown in Figure 3B with the labeled same input layers definition of Figure 3A. The ten-parameter network contains all input layer nodes from the six-parameter network plus 4 additional geometric parameters from bottom mandrel step at BM_{RIE}. Figure 4A shows the bottom mandrel stack at BM_{RIE}, where the geometric stack parameters 7, 8, 9, and 10 correspond to bottom mandrel top width, bottom mandrel bottom width, bottom mandrel sidewall angle and stack nitride thickness, respectively. Likewise, the fourteen-parameter network contains all input layer nodes from the ten-parameter network plus 4 additional geometric parameters from bottom mandrel step at BM_{SP}. Figure 5A shows the at bottom mandrel stack at BM_{SP}, where the geometric stack parameters 11, 12, 13, and 14 correspond to bottom mandrel top width, bottom mandrel sidewall angle, sidewall spacer width, and bottom mandrel height, respectively.

5. RESULTS AND DISCUSSION

In this section we examine the output of a fully trained n-parameter SAQP DNN network using the methodology previous described. Examination of the output of the forward-solve inference of a given n-parameter network using the Bayesian dropout approximation prescription allows a comparison of the predicted uncertainty of the distribution for the pitch-walk output layers $\alpha - \beta$ and $\alpha - \gamma$ from the probability density output. The forward-solve inference can be calculated on either a given experimental wafer/chip location or an artificial site which is reflective of all parameter space. To gauge distribution uncertainty of the latter, the mean (or centroid) of each of the parameters is determined. This centroid determination is readily achieved, since neural network activations require normalized inputs to ensure meaningful gradient behavior. This histogram of the probability density for the six-parameter network is shown in Figure 6A for $\alpha - \beta$ and Figure 6B for $\alpha - \gamma$. The mean value of these "Gaussian" probability density distributions can be readily determined. While these mean values are not meaningful *per se* in this case, since they are determined from an "artificial" mean parameter site, it is noted that for this dataset the $\alpha - \gamma$ has a negative value ($i.e. \gamma > \alpha$) , while the $\alpha - \beta$ is positive ($i.e. \alpha > \beta$) and is self-consistent with the SAQP model. The approximate range of the predicted uncertainty for the $\alpha - \gamma$ is roughly double that of $\alpha - \beta$. The predicted uncertainty determined as a Gaussian 1σ value for $\alpha - \beta$, and $\alpha - \gamma$ is 0.12 and 0.24 nm, respectively. The prediction uncertainty reflects the underlying distribution of the experimentally measured pitch-walk. The validity of this prediction uncertainty can be accessed from the correlation of the

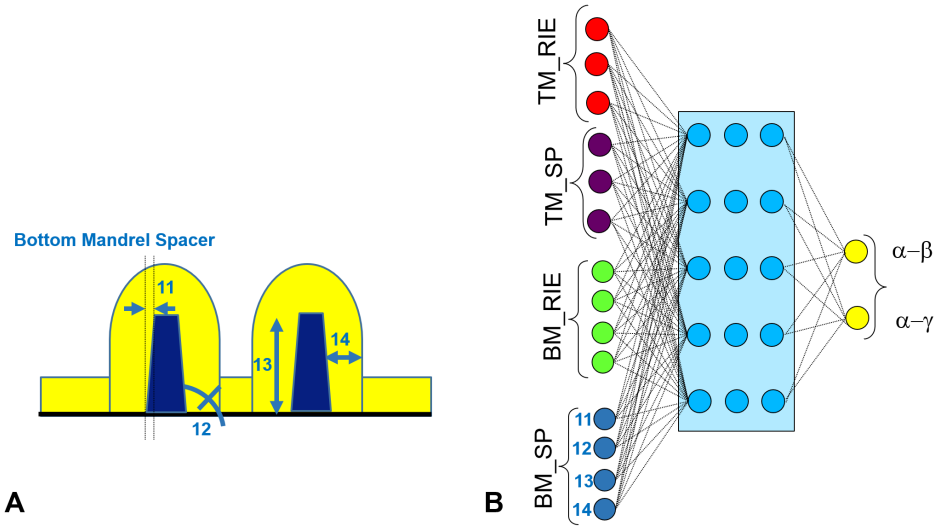

Figure 5. This shows the stack geometry parameter definitions for the bottom mandrel BM_{SP} for the ten-parameter network with BM_{SP}

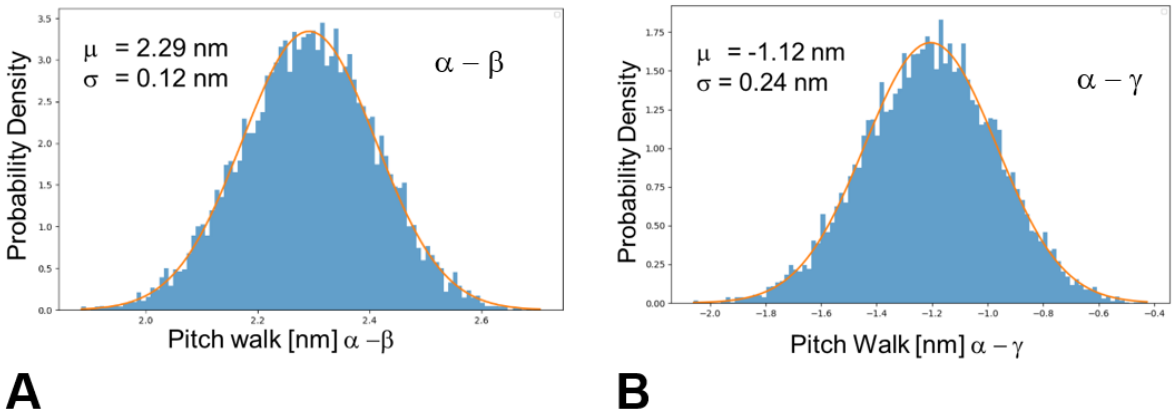

Figure 6. This shows the probability density for the six-parameter network for a) $\alpha - \beta$ and b) $\alpha - \gamma$ for the inference using the centroid of the parameters distributions

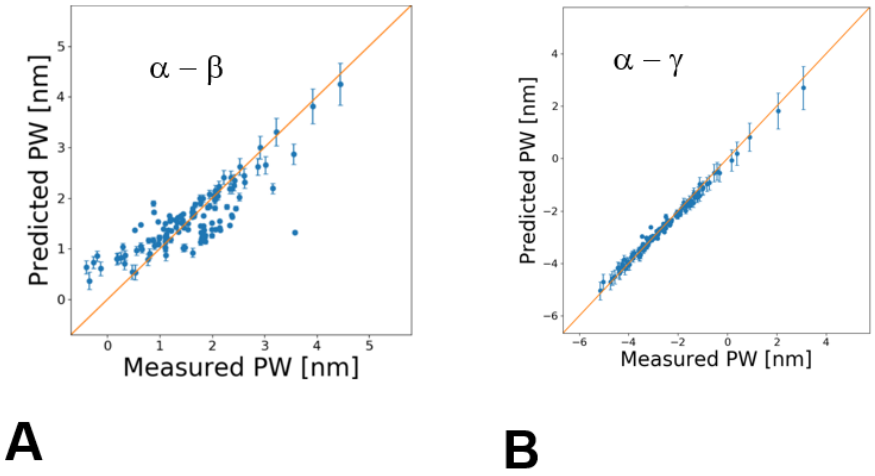

Figure 7. This shows the correlation of the measured pitch walk (PW) as a function of the predicited output of the six-parameter network for both a) $\alpha - \beta$ and b) $\alpha - \gamma$ for the inference using the centroid of the parameters distributions

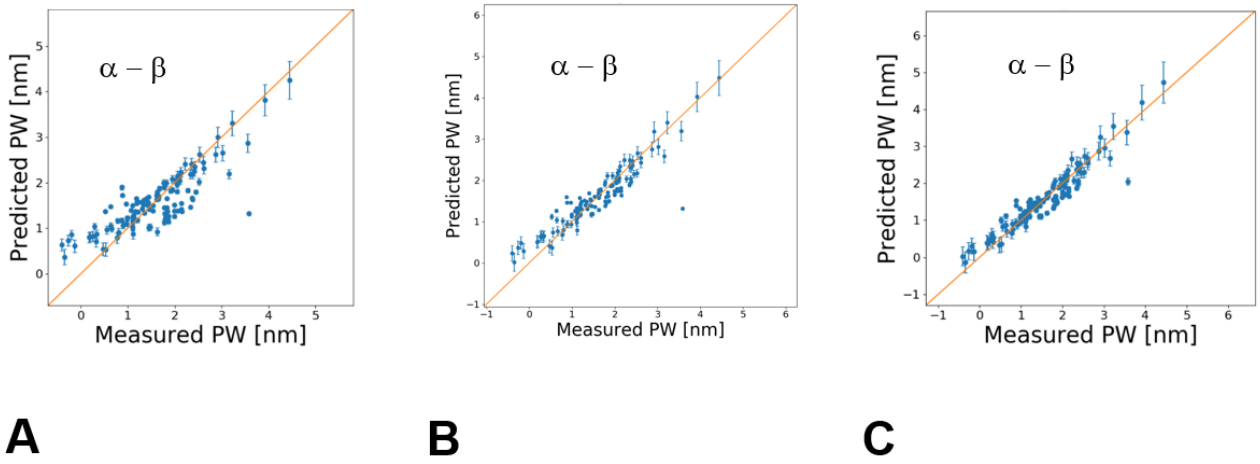

Figure 8. This shows the correlation of the measured pitch walk (PW) as a function of the predicted pitch walk for the output layer $\alpha - \beta$ in the following networks a) the six-parameter b) the ten-parameter , and c) the fourteen parameter networks)

experimentally measured pitch-walk metrics as a function of the predicted pitch walk. This correlation scatter plot is shown in Figure 7A for $\alpha - \beta$ and 7B for $\alpha - \gamma$. The calculated 1σ values of the predicted pitch walk are shown. Note that the scale of the plots are displayed in nanometers and are not normalized for comparison. The uncertainty (*i.e.*1σ values) at both the lower and upper values of the correlation scatterplot are significantly larger than those values which are clustered at the center of the plots. The data at the center of the plots, with a larger number of experimental values, are better determined. It is observed that the mean values of the predicted $\alpha - \beta$ are not as precisely predicted at the $\alpha - \gamma$. This observation from the six-parameter network, reflecting only parameters from the top mandrel, is consistent with the previous discussion of SAQP. The $\alpha - \gamma$ pitch walk is determined primarily from the top mandrel, whereas the $\alpha - \beta$ pitch-walk value is inherently determined from both top and bottom mandrel parameters, where the later parameter is not defined for this network.

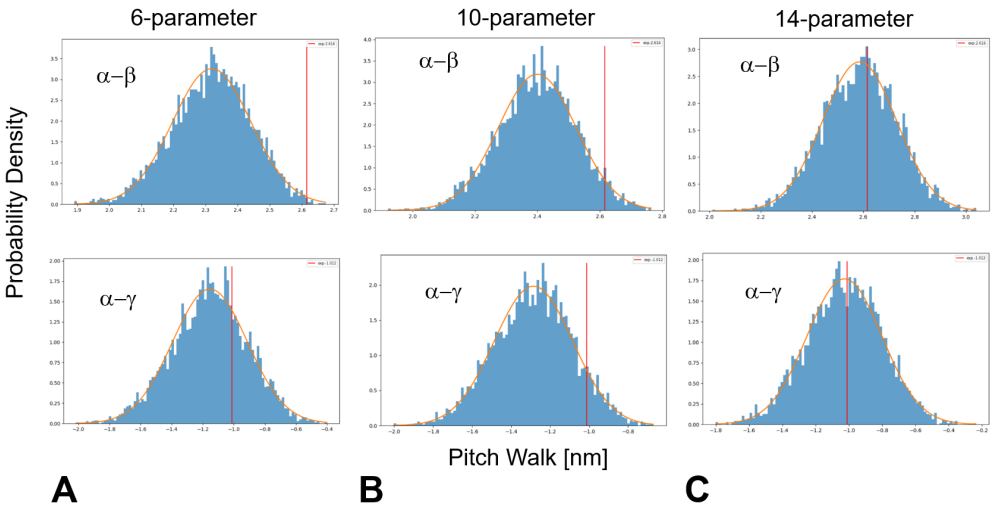

Figure 9. The outcome of the probability density of the $\alpha - \beta$ and $\alpha - \gamma$ pitch walk for an arbitrary wafer site location is shown for the following networks a) the six-parameter b) the ten-parameter , and c) the fourteen parameter networks. The experimental mean of the chip-site data is indicated by a red-colored vertical bar.)

In the previous discussion, we examined results from the forward-solve inference with different n-parameter networks for an artificial site defined by the mean (or centroid) of the geometric parameters. Using this approach, we gain an insight to the overall behavior of the SAQP networks. Alternatively, applying the forward-solve inference to individual experimental wafer/chip-sites is a more realistic use-case for applying the pitch-walk prediction. The forward-solve inference is applied to an arbitrary single chip location of a wafer with different n–parameter networks, allowing a comparison of pitch-walk predictions at different process steps of the SAQP process. This histograms of the probability density for both the $\alpha - \beta$ and $\alpha - \gamma$ pitch walks are shown in Figure 9 for each n-parameter network. Figure 9 shows a comparison of the predicted distributions to the experimental mean of the chip-site data, indicated by a red colored vertical bar. By visually comparing the means of the predicted distributions to the experimental means, we can easily see that for both $\alpha - \beta$ and $\alpha - \gamma$ pitch-walks, increasing the number of input layer/nodes in the network improves the agreement between the predicted and experimental means. From the six-parameter, ten-parameter, to fourteen parameter networks, the improvement between the predicted and observed value for $\alpha - \beta$ pitch walk is substantial, with a reduction of 2σ. This result is consistent with our previous discussion, in which the scatterplots of the predicted to measured $\alpha - \beta$ pitch walk improve with increasing the number of parameters. The pitch-walk uncertainty for both the $\alpha - \beta$ and $\alpha - \gamma$ does not significantly improve or degrade with increasing number of input layers. This will be discussed in Section 6. The application of pitch-walk predictions over a number of sites on a wafer, relative to a defined threshold value, can be applied to a predicative disposition process.

5.1 Methodology: Sensitivity to Input Parameters

A more nuanced understanding of the SAQP DDN is gained by exploiting the feedforward-solve inference under the Bayesian dropout approximation, giving insight into the parameter sensitivity of the network. A methodology for gauging the sensitivity of input layer parameters to DNN, the subject of a future investigation,[22] is briefly explored here.

The predicted distribution of an output parameter is generated by systematically varying a chosen input layer parameter (η) over a small range, where the impact of η on the output. Figure 10 shows two 2D histograms, generated using the fourteen parameter network, of the probability density distribution for $\alpha - \gamma$ pitch walk as a function of two different η, where η projects into the page and is allowed to vary over a range of 0.25 standard deviations. In Figure 10A, η is chosen to be one of the $\mathrm{BM_{RIE}}$parameters, and the distribution of predicted $\alpha - \gamma$ ptich walk is shown to be rather insensitive to variations in η. However, when η is chosen to be a $\mathrm{TM_{RIE}}$

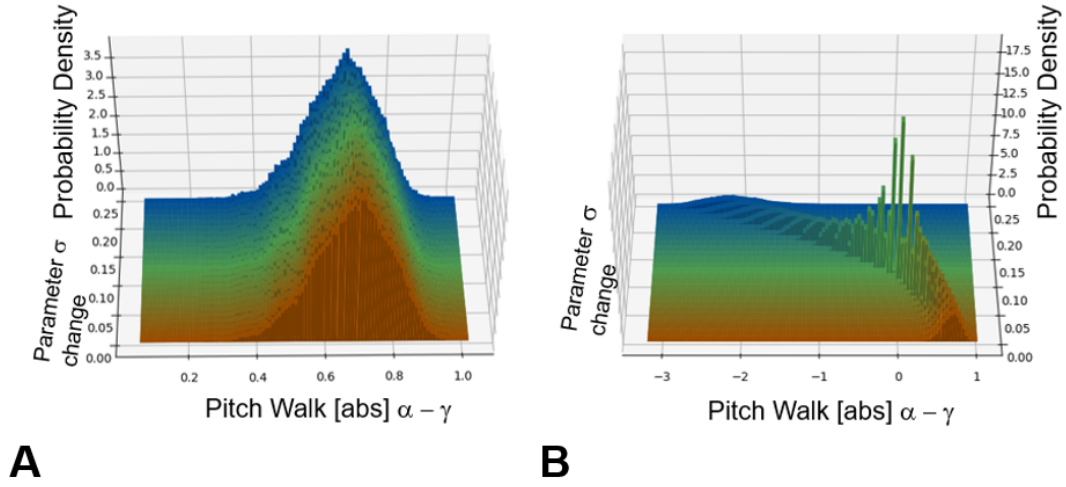

Figure 10. Figure 10 shows two 2D histograms, generated using the fourteen parameter network, of the probability density distribution for $\alpha - \gamma$ pitch walk as a function of two different η, where η projects into the page and is allowed to vary over a range of 0.25 standard deviations

parameter in Figure 10B, we see that the predicted pitch walk distribution is quite sensitive and the histogram responds strongly.

The area of high probability density presumably corresponds to the tight distribution of the parameter where the network was trained. These pitch-walk sensitivities shown here are consistent with our understanding of SAQP. The pitch walk $\alpha - \gamma$ should be sensitive to certain top mandrel parameters, and not sensitive to bottom mandrel parameters. In principal, this analysis can be extended to n-dimensional input parameters. n-dimensional sensitivity is graphically complex and understanding parameter interactions is beyond the scope of this work. In summary, the ablity to interrogate a DNN with this methodology is a computationally fast and powerful tool for understanding process sensitivities.

6. CONCLUSIONS

In this work we have described in detail the complexity of the self-aligned quadruple patterning and the need to have a predictive model for both the mean and the uncertainty of pitch walk. Elaboration of the network input layers, the top and bottom mandrel geometric stack parameters, and their contribution to three types of space-widths: α, β, and γ in the final pattern is given. The relationship of the stack parameters to the output layers of the network, pitch walk $\alpha - \beta$ and $\alpha - \gamma$, is demonstrated. We defined the relevant network topologies and the input layers for modeling SAQP: six-parameter top mandrel only network, and ten and fourteen parameter networks including both top and bottom mandrels. Modeling the complex SAQP process with a stochastic DNN achieved a very good correlation of measured to predicted pitch-walk values despite the fact that the DNN model contains no knowledge of the physics of SAQP. The use of the Bayesian dropout approximation to perform Bayesian inference is an effective, easy to implement, and computational fast prescription for making sophisticated predictions about the pitch walk observed in SAQP. These predictions provide quantitative uncertainties and can be used in further business-relevant calculations for process outcomes.

The predicted pitch-walk for the n-parameter network gives a range for uncertainty from the probability density that is not found to significantly change by increasing the size of the input layer nodes. However, increasing the number of input layer nodes/parameters does improve the overall goodness-of-fit of the model predictions to the measured values for pitch walk $\alpha - \beta$. Thus, increasing the number of input parameters does improve the pitch-walk prediction, (*ie.* the predicted to experimental mean). For the available dataset, it is not unreasonable that the magnitude of the pitch-walk uncertainty is dominated by the top mandrel process and that

further downstream processes do not increase or decrease the uncertainty. Thus, the ability to predict both the mean and uncertainty for the pitch-walk outcome early in SAQP process flow is a powerful methdology, which could be deployed for reliable manufacturing process disposition.

While the network input nodes in this SAQP study was limited by design intent, the methodology with the BDA approximation can easily be scaled to a much larger set in input parameters and network sizes. We have demonstrated that DNNs can be effectively used to model the complexity of SAQP and with the use of the BDA approximation provide high confidence actionable results. This approach can be readily extended to modeling other complex patterning processes, such as self-aligned double patterning and self-aligned octuple patterning. In addition, there are a host of opportunities to deploy the methodology demonstrated here to other device-yield scenarios.

ACKNOWLEDGMENTS

The authors would like to acknowledge both Drs. Geng Han and Robert Baseman for their useful insights and IBM for funding this project with the CCI (Center for Computational Innovations) at RPI.

REFERENCES

[1] Gal, Y. and Ghahramani, Z., "Dropout as a Bayesian approximation: Insights and applications," in [*Deep Learning Workshop, ICML*], (2015).

[2] Gal, Y. and Ghahramani, Z., "Bayesian convolutional neural networks with Bernoulli approximate variational inference," in [*4th International Conference on Learning Representations (ICLR) workshop track*], (2016).

[3] Standaert, T., Beique, G., Chen, H.-C., Chen, S.-T., Hamieh, B., Lee, J., McLaughlin, P., McMahon, J., Mignot, Y., Mont, F., et al., "Beol process integration for the 7 nm technology node," in [*2016 IEEE International Interconnect Technology Conference/Advanced Metallization Conference (IITC/AMC)*], 2–4, IEEE (2016).

[4] Chao, R., Breton, M., L'herron, B., Mendoza, B., Muthinti, R., Nelson, F., Pena, A. D. L., li Le, F., Miller, E., Sieg, S., Demarest, J., Gin, P., Wormington, M., Cepler, A., Bozdog, C., Sendelbach, M., Wolfling, S., Cardinal, T., Kanakasabapathy, S., Gaudiello, J., and Felix, N., "Advanced in-line metrology strategy for self-aligned quadruple patterning," in [*Metrology, Inspection, and Process Control for Microlithography XXX*], Sanchez, M. I., ed., **9778**, 380 – 389, International Society for Optics and Photonics, SPIE (2016).

[5] Raley, A., Mohanty, N., Sun, X., Farrell, R. A., Smith, J. T., Ko, A., Metz, A. W., Biolsi, P., and Devilliers, A., "Self-aligned blocking integration demonstration for critical sub-40nm pitch mx level patterning," in [*Advanced Etch Technology for Nanopatterning VI*], **10149**, 101490O, International Society for Optics and Photonics (2017).

[6] Felix, N., Corliss, D., Petrillo, K., Saulnier, N., Xu, Y., Meli, L., Tang, H., De Silva, A., Hamieh, B., Burkhardt, M., et al., "Euv patterning successes and frontiers," in [*Extreme Ultraviolet (EUV) Lithography VII*], **9776**, 97761O, International Society for Optics and Photonics (2016).

[7] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R., "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580* (2012).

[8] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research* **15**(1), 1929–1958 (2014).

[9] 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).

[10] Neal, R., "Bayesian learning for neural networks [phd thesis]," *Toronto, Ontario, Canada: Department of Computer Science, University of Toronto* (1995).

[11] Williams, C. K., "Computing with infinite networks," in [*Advances in neural information processing systems*], 295–301 (1997).

[12] Williams, C. K., "Computation with infinite neural networks," *Neural Computation* **10**(5), 1203–1216 (1998).
[13] MacKay, D. J., "A practical bayesian framework for backpropagation networks," *Neural computation* **4**(3), 448–472 (1992).
[14] Tishby, N., Levin, E., and Solla, S. A., "Consistent inference of probabilities in layered networks: Predictions and generalization," in [*International Joint Conference on Neural Networks*], **2**, 403–409 (1989).
[15] Graves, A., "Practical variational inference for neural networks," in [*Advances in neural information processing systems*], 2348–2356 (2011).
[16] Hinton, G. E. and Van Camp, D., "Keeping the neural networks simple by minimizing the description length of the weights," in [*Proceedings of the sixth annual conference on Computational learning theory*], 5–13 (1993).
[17] Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D., "Weight uncertainty in neural network," in [*Proceedings of the 32nd International Conference on Machine Learning*], Bach, F. and Blei, D., eds., *Proceedings of Machine Learning Research* **37**, 1613–1622, PMLR, Lille, France (07–09 Jul 2015).
[18] Gal, Y. and Ghahramani, Z., "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," *arXiv:1506.02142* (2015).
[19] Gal, Y., *Uncertainty in Deep Learning*, PhD thesis, University of Cambridge (2016).
[20] Kagalwala, T., Vaid, A., Mahendrakar, S., Lenahan, M., Fang, F., Isbester, P., Shifrin, M., Etzioni, Y., Cepler, A., Yellai, N., Dasari, P., and Bozdog, C., "Scatterometry-based metrology for SAQP pitch walking using virtual reference," in [*Metrology, Inspection, and Process Control for Microlithography XXX*], Sanchez, M. I., ed., **9778**, 627 – 633, International Society for Optics and Photonics, SPIE (2016).
[21] Halle, S. D., Yeo, K. m., and Derren, D., "unpublished," (2016).
[22] Bloomfield, M. O. and Halle, S. D., "to be published,".