IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Probabilistic artificial neural network for line-edge-roughness-induced random variation in FinFET

JAEHYUK LIM, JINWOONG LEE, AND CHANGHWAN SHIN

Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea

Corresponding author: Changhwan Shin (cshin@skku.edu)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2020R1A2C1009063). And, this work was supported by the Future Semiconductor Device Technology Development Program (20003551) funded by the Ministry of Trade, Industry & Energy (MOTIE) and the Korea Semiconductor Research Consortium (KSRC). The EDA Tool was supported by the IC Design Education Center (IDEC), Korea.

ABSTRACT Line-edge-roughness (LER) is one of undesirable process-induced random variation sources. LER is mostly occurred in the process of photo-lithography and etching, and it provokes random variation in performance of transistors such as metal oxide semiconductor field effect transistor (MOSFET), fin-shaped field effect transistor (FinFET), and gate-all-around field effect transistor (GAAFET). LER was analyzed/characterized with technology computer-aided design (TCAD), but it is fundamentally very time consuming. To tackle this issue, machine learning (ML)-based method is proposed in this work. LER parameters (i.e., amplitude, and correlation length X, Y) are provided as inputs. Then, artificial neural network (ANN) predicts 7-parameters [i.e., off-state leakage current (I_{off}), saturation drain current (I_{dsat}), linear drain current (I_{dlin}), low drain current (I_{dlo}), high drain current (I_{dhi}), saturation threshold voltage (V_{tsat}), and linear threshold voltage (V_{tlin})] which are usually used to evaluate the performance of FinFET. First, how datasets for training process of ANN were generated is explained. Next, the evaluation method for probabilistic problem is introduced. Finally, the architecture of ANN, training process and our new proposition is presented. It turned out that the prediction results (i.e., non-Gaussian distribution of device performance metrics) obtained from the ANN were very similar to that from TCAD in the respect of both qualitative and quantitative comparison.

INDEX TERMS Line edge roughness, process-induced random variation, FinFET, machine learning, artificial neural network.

I. INTRODUCTION

Over the past a few decades, complementary metal oxide semiconductor (CMOS) technology has been evolved with advanced techniques such as stress engineering in 90 nm technology node [1], high-k/metal-gate (HK/MG) in 45 nm technology node [2], and three-dimensional advanced device structure in 22 nm technology node [3]. Those new techniques have enabled the physical dimension of metal oxide semiconductor field effect transistor (MOSFET) to be successfully scaled down, resulting in the improved functions of integrated circuit (IC) per cost. However, there still exists secondary effects in aggressively scaled MOSFETs, and they should be overcome. Especially, one of those challenges, i.e., process-induced random variation, which randomly cause variation in transistor performance

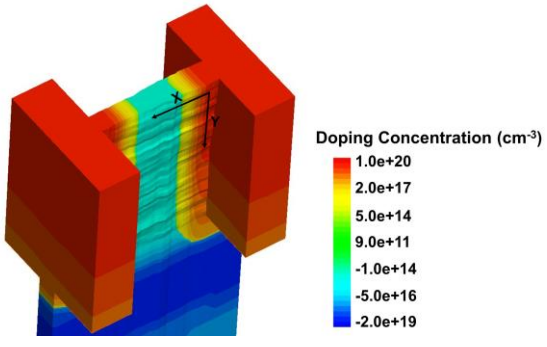

FIGURE 1. A bird's-eye view of FinFET with a three-dimensional line-edge-roughness (LER) on its sidewall. LER parameters used in this example are as follows: Δ = 0.5 nm, Λ_x = 20 nm, Λ_y = 50 nm, α = 1, and Θ = 0.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3088461, IEEE Access

metrics such as threshold voltage, on-state drive current, and off-state leakage current, become significant as CMOS technology is evolved [4]. The primary causes of process-induced random variation can be classified as (i) line edge roughness (LER), (ii) random dopant fluctuation (RDF), and (iii) work function variation (WFV) [5]. Among them, because LER can affect the other random variation sources (i.e., RDF and WFV) by inducing the deformation of device structure [6], it would degrade the device performance more severely. With the most radical shift in device architecture at 22 nm node, i.e., from planar bulk MOSFET to three-dimensional fin-shaped field effect transistor (FinFET), the process-induced random variation becomes much more severe [7]. As expected that a more complicated device structure such as multiple bridge channel field effect transistor (MBCFET), stacked nano-wire FET, complementary FET (CFET) would be adopted at 3 nm node [8], understanding and analyzing the impact of LER on device performance is essential for designing variation-robust silicon device [9].

To understand and quantify the impact of LER on device performance, TCAD has been used so far. TCAD simulation is, however, fundamentally very time-consuming. As another approach, compact model [10] has been used to overcome the time-inefficiency of using TCAD tools. However, the compact model for LER is still based on 2-D analysis, despite that 3-D analysis for LER is necessary [7]. As an alternative way to avoid those obstacles, we have focused on machine learning (ML)-based artificial neural network (ANN). Machine Learning technology has already been in spotlight in various fields such as geology, and biology [11-13]. Especially in semiconductor technologies, a number of studies have been reported in many branches such as fabrication [14, 15], optimization [16], and modeling [17]. Following this trend, we suggested a ML-based ANN model in our previous work [18]. However, it turned out that the ML-based ANN had limits, in that target performance metrics were assumed to follow multi-variate Gaussian distribution only. This assumption would cause some inevitable errors. Moreover, the ANN model has its own intrinsic limit in estimating the other performance metrics such as I_{dlo}, and I_{dhi} (which, in real, do not follow Gaussian distribution). Therefore, we would like to develop and propose an upgraded ANN model with enhanced accuracy (note that this new ANN model can overcome the limit mentioned above).

In this work, we show the way how FinFETs with LER are simulated as well as how those data are preprocessed for training process of ANN. Afterwards, the evaluation method used to assess the results of proposed work is introduced. Finally, it is shown how ANN was composed and built, including its geometrical structure, hyper parameters, and the process of grafting probability.

TABLE I. NOMINAL DEVICE PARAMETERS OF FINFET [8].

Device design parameters			
Symbol	Description	Unit	Value
L_g	Gate length	nm	20
T_{ox}	Equivalent oxide thickness	nm	0.3
W_{fin}	Fin width	nm	7
H_{fin}	Fin height	nm	50
V_{DD}	Power supply voltage	V	0.7
$N_{S/D}$	Source/drain doping concentration	cm^{-3}	10^{20}
N_{RCD}	Peak concentration of Punch-through Stopper	cm^{-3}	2×10^{19}

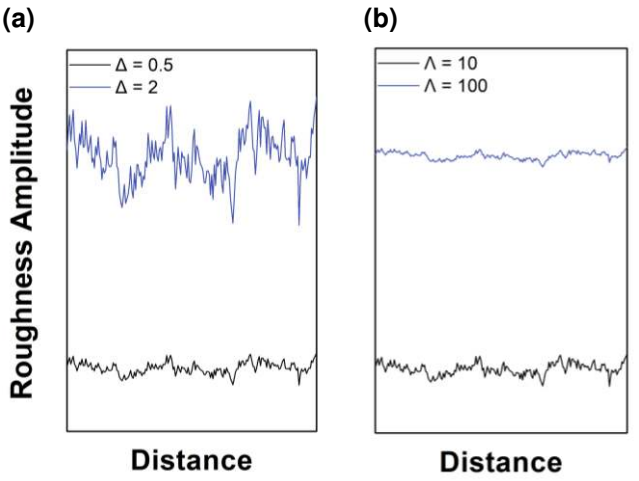

FIGURE 2. Examples of roughness amplitude when (a) Λ = 10, and (b) Δ = 0.5.

II. DATA GENERATION

A. TCAD SET-UP WITH LINE EDGE ROUGHNESS

As done in the previous study [18], 3-D quasi atomistic model for line edge roughness was used [15]. Three parameters (i.e., Δ, Λ_x, and Λ_y) are used to describe and reconfigure LER profile. The physical meaning of those parameters are as follows [see Fig 2]:

(i) Amplitude (Δ): it indicates the rms (root-mean-squared) value of surface roughness.

(ii) Correlation length (Λ): it means how line edge is closely correlated with its neighboring edge. A larger Λ indicates a smoother line.

$$ACVF(x,y) = \Delta^2 \exp\left[-\left\{\frac{(x\cos\Theta + y\sin\Theta)^2}{\Lambda_x^2} + \frac{(-x\cos\Theta + y\sin\Theta)^2}{\Lambda_y^2}\right\}^{\frac{1}{2}}\right] \quad (1)$$

In Eq. (1), Λ_x and Λ_y are the correlation length along x-direction and y-direction, respectively [see Fig. 1]. Θ indicates the relation between x-direction and y-direction.

Using parameters mentioned above and two-dimensional auto covariance function (ACVF), we have simulated FinFET with MATLAB and TCAD Sentaurus Structure Editor and Device [19, 20]. The detailed steps how to create a rough sidewall surface of FinFET are provided in [18, 19].

Nominal device parameters of FinFET is summarized in Table I. Drift-diffusion simulation for the FinFETs with surface roughness are executed, using various models such as doping-dependent mobility model, thin-layer mobility model for carrier transport, the Shockley-Read-Hall (SRH) model for generation and recombination, high field

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3088461, IEEE Access

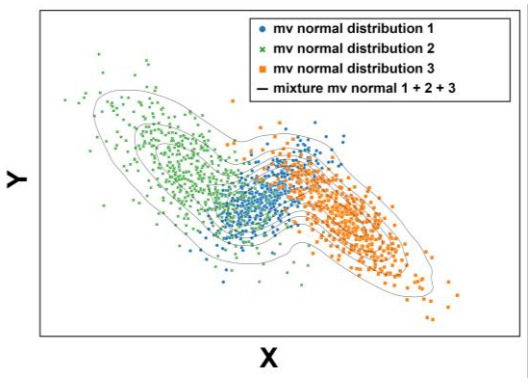

FIGURE 3. Conceptual diagram for showing the mixture of multivariate normal distributions.

TABLE II. PERFORMANCE METRICS.

Symbol	Description
I_{off}	I_d at V_g = 0 [V], V_d = 0.7 [V], V_s, V_b = 0 [V]
V_{tsat}	V_g at I_d = 535 [nA], V_d = 0.7 [V], V_s, V_b = 0 [V]
V_{tlin}	V_g at I_d = 535 [nA], V_d = 0.05 [V], V_s, V_b = 0 [V]
I_{dsat}	I_d at V_g = 0.7 [V], V_d = 0.7 [V], V_s, V_b = 0 [V]
I_{dlin}	I_d at V_g = 0.7 [V], V_d = 0.05 [V], V_s, V_b = 0 [V]
I_{dlo}	I_d at V_g = 0.35 [V], V_d = 0.7 [V], V_s, V_b = 0 [V]
I_{dhi}	I_d at V_g = 0.7 [V], V_d = 0.35 [V], V_s, V_b = 0 [V]

Note that I_d is the drain current, V_d is the drain voltage, V_s is the source voltage, V_b is the body voltage, and V_g is the gate voltage.

saturation model for velocity saturation, and a density-gradient quantization model for quantum-mechanical effects.

B. DATA COMPOSITION

To train, validate, and test the ANN model that we have newly built up, two kinds of datasets were separately generated: (1) The first kind of dataset has 130 datasets. Each dataset contains the performance metrics of 50 different FinFETs, so that total 6,500 FinFETs are used in 130 datasets. 70% of them are used as the training datasets, and the others are used as the validation datasets. (2) The second kind of dataset has 10 different datasets. Each dataset contains the performance metrics of 250 different FinFETs, so that total 2,500 FinFETs are used in 10 datasets.

The LER parameters are chosen in the range below:

Amplitude (Δ): 0.1 nm ~ 0.8 nm
Correlation length X (Λ_x): 10 nm ~ 100 nm
Correlation length Y (Λ_y): 20 nm ~ 200 nm

The reference parameter set (Δ = 0.5 nm, Λ_x = 20 nm, Λ_y = 50 nm) is obtained, based on experimental results [21-24]. Then, the range is selected on the basis of the reference parameter set. Note that the performance metrics of transistor were extracted from simulated drain current versus gate voltage (I_d-vs.-V_g) characteristic. The details on the device performance metrics are summarized in Table II. These metrics are extracted using the Sentaurus TCAD inspect [20].

III. EVALUATION METHOD

To quantitatively verify the distribution of values obtained from the new ANN model, we used earth-mover's distance (EMD) score (or be referred to as Wasserstein metric in mathematics). The EMD score is used to measure

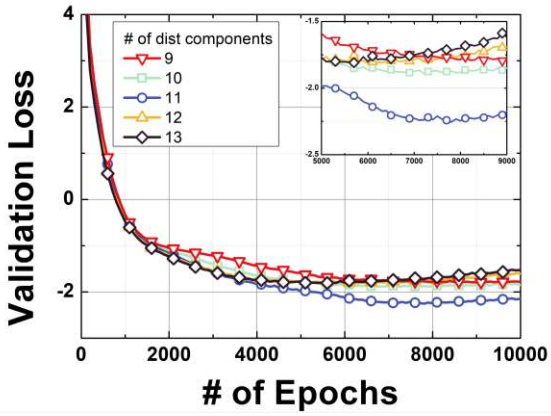

FIGURE 4. Validation loss vs. the number of epochs, when the number of distribution used to mixture is varied from 9 to 13.

how two probability distributions are different from each other [25]. The definition of the score is "The minimal amount of work needed to transform one distribution to another distribution". The EMD score can be calculated following the steps below:

Step I: Calculate the difference of cumulative distribution function (CDF) of TCAD datasets and ANN prediction datasets.

Step II: Normalize the calculated value in Step I.

In this work, we used Gaussian kernel density estimation (KDE) to estimate CDF of datasets. The EMD score is "0" when two distributions are exactly identical.

IV. MODEL GENERATION

A. MIXTURE-MULTIVARIATE NORMAL DISTRIBUTION

Different from the previous study [10], the mixture of multivariate normal distributions (MVN) is used in this work. By using the mixture of MVNs, we can respond to many other unknown distribution shapes. It is trivial that performance metrics of transistor approximately follow Gaussian distribution [26], [27], [28]. However, triggered by many non-ideal effects in transistors (i.e., short-channel effects [29]), there are some skewness, kurtosis, and/or non-linear correlation in-between the performance metrics. Moreover, distribution shapes are also quite different, for each LER parameter. For those reasons, the mixture of MVNs was used to deal with the non-ideal cases. A conceptual diagram for the mixture of MVNs is shown in Fig. 3.

To determine the number of components (i.e., MVN distributions) used in generating the mixture of MVNs, an optimization was first done. Validation datasets were used in the training process, to find the best working model. In Fig. 4, it can be noted that the validation loss was minimized with the number of components of 11 at 7,800 epochs. This means that the ANN model with 11 MVNs works best to describe the distribution of performance metrics. The optimized ANN with the mixture of MVNs has 3 neurons for the input layer, 81 neurons for the first hidden layer, 162 neurons for the second hidden layer, 324 for the third hidden layer, and 324 neurons for the output layer. This output neurons are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3088461, IEEE Access

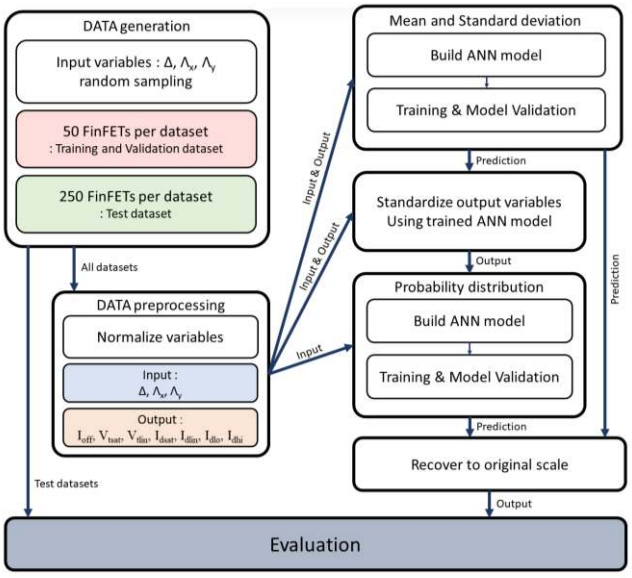

FIGURE 5. **The flow chart how to build/train/test the ANN model.**

TABLE III. TIME SPENT TO TRAIN ANN MODEL.

Model	Time
Previous study [10]	776 sec (for I_{off}, I_{dsat}, V_{tsat}, and SS) 1191 sec (for I_{off}, I_{dlin}, I_{dsat}, I_{dlo}, I_{dhi}, V_{tlin} and V_{tsat})
Non-separated ANN	1412 sec
This work	185 sec (84 sec for mean and standard deviation + 101 sec for mixture of MVN)

connected to probabilistic layer with the mixture of multivariate normal distribution, for the generation of power density function (PDF) of output variables (performance metrics).

B. ANN FOR MEAN AND STANDARD DEVIATION

Unlike the proposed ANN in the previous section, another simple ANN was built only for estimating mean and standard deviation. With this simple ANN, the performance metrics were standardized in the training process of ANN with the mixture of MVNs [see Fig. 5]. By using additional simple ANN, we can limit the role of ANN with the mixture of MVNs, to just estimate the shape of probability distribution. Thus, the time spent to train the ANN model can be largely reduced (see Table III). The simple ANN has 3 neurons for the input layer, 3 neurons for the first hidden layer, 7 neurons for the second hidden layer, 14 neurons for the third hidden layer, and 14 neurons for the output layer. To compare this work against the non-separated ANN (note that the ANN model can predict the mean, standard deviation, and the shape of distribution, at once), the same work mentioned in Part A was repeated to find the optimized ANN (i.e., the number of MVNs, training epochs, etc.).

C. TRAINING METHODOLOGY

The weight matrices and bias vectors of ANN model are updated for the given/specified number of iterations. These matrices and vectors determine the output of ANN model. The probabilistic layer attached to output neurons returns the PDF of variables while training process. Thus, conventional mean-squared error cannot be used as loss function. Instead, "Negative log likelihood" (Negloglik) was used as a loss function [see Eq. (2)]. The training process is executed to minimize this loss function. That is, training ANN becomes the process of Maximum Likelihood Estimation (MLE) [30].

$$Negloglik(P,Q) = -\sum_x P(x) log Q(x) \qquad (2)$$

In Eq. (2), P(x) and Q(x) denotes the PDF of observation and hypothesis, respectively.

Using Adam Optimizer [31], the training process was executed for 14,880 epochs (84 sec) for mean and standard deviation of ANN model and 7,800 epochs (101 sec) for the mixture of MVNs ANN model. Both models are trained with the learning rate of 10^{-4}, and then optimized to prevent the occurrence of over-fitting by using validation datasets. ReLU was used as the activation function for both models. Note that the ANN model was built using the Tensorflow 2.0 and Tensorflow-probability python library [32, 33].

V. RESULTS AND EVALUATION

Based on the PDF determined by the mixture of MVNs, the standardized prediction data was randomly extracted. Then, these standardized values are recovered with the predicted mean and standard deviation to the original scale (see Fig. 5).

Fig. 6 shows the comparison between the previous work, non-separated ANN, and this work, for a given LER of Δ = 0.505 nm, Λ_x = 48.62 nm, Λ_y = 67.99 nm. As shown in Fig. 5(a, b), previous work with plain MVN cannot predict metrics with non-linear correlation, skewness and kurtosis. On the other hand, ANN with mixture of MVN [see Fig. 6(c-f)] successfully predicts skewness, kurtosis, and non-linear correlation, which is distinctly different from plain MVN. EMD score also proves that prediction accuracy is highly improved (0.0170 vs 0.00928). As shown in Fig. 6(c-f), there is no significant performance degradation between non-separated ANN model and this work, in spite of huge time saving (1412 sec to 185 sec) [see Table III].

The pair plots for Δ = 0.505 nm, Λ_x = 48.62 nm, Λ_y = 67.99 nm, and Δ = 0.690 nm, Λ_x = 85.09 nm, Λ_y = 80.98 nm are shown in [See Fig. 7 and Fig. 8]. We can notify that distribution for each parameter (a diagonal line) and correlation between parameters (the rest except for a diagonal line) are well predicted.

Table IV shows the EMD score comparison of performance metric. In the respect of quantitative analysis, it is shown that this work shows quite similar performance with non-separated ANN, when the amplitude of LER profile is around 0.5 or 0.6. However, it shows enhanced performance in a wider range than the non-separated ANN model. This work not only predicts the mean and standard deviation at a high level with the additional simple ANN, but also shows higher consistency for specific points such as the tail of distribution by using the mixture of MVNs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3088461, IEEE Access

VI. CONCLUSION

We have proposed newly developed ANN models with enhanced accuracy. A ML-based model [14] was first suggested to estimate LER-induced random variation, and its simulation time was shorter than using compact models. Herein, compared against the previous ML-based model [14], the newly proposed ANN models have shortened the simulation time by ~ 6 times (from 1,191 seconds to 185 seconds). Especially, non-Gaussian features of device performance metrics' distribution (i.e., skewness, kurtosis, and non-linear correlation) are successfully predicted while

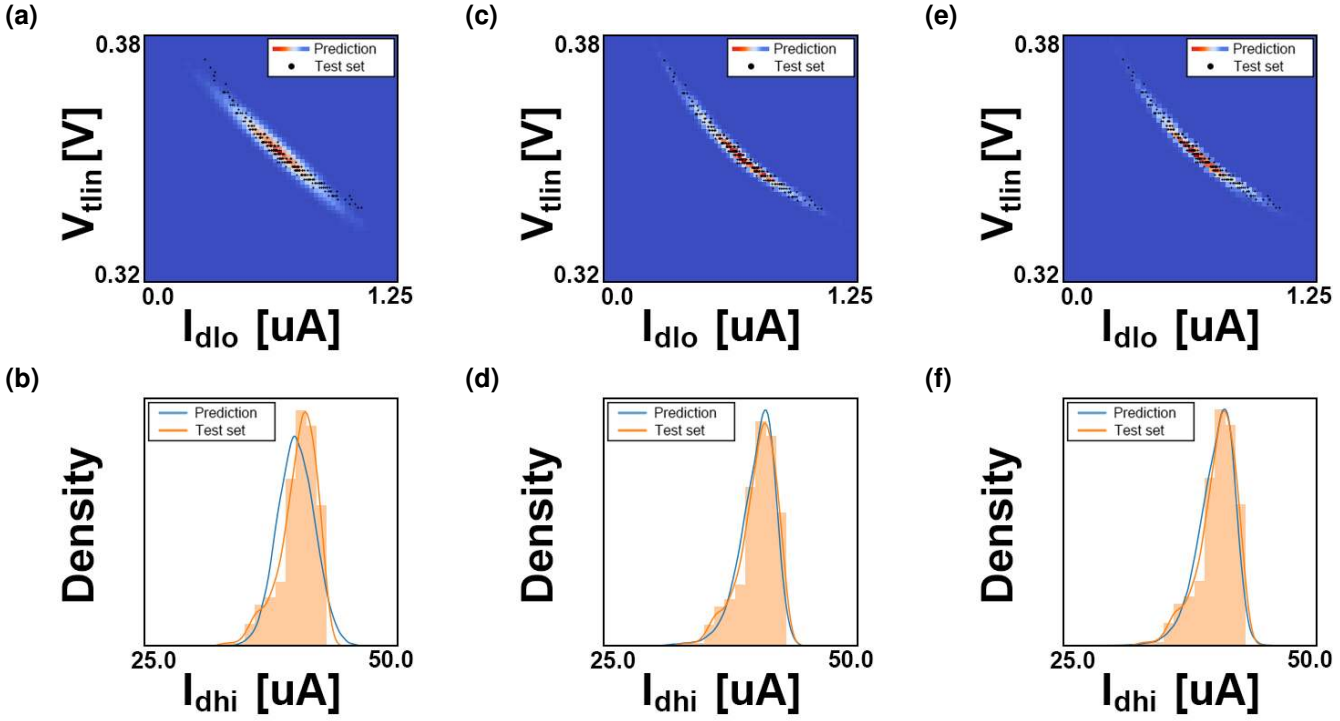

FIGURE 6. Histograms of (a, b) previous work, (c, d) non-separated model, and (e, f) this work. EMD score of (b), (d), and (f) is 0.0170, 0.00785, and 0.00928, respectively.

TABLE IV. COMPARISON OF EMD SCORE BETWEEN THIS WORK AND NON-SEPARATED ANN MODEL.

Test dataset	LER Parameters			Model	EMD Score							
	amp(Δ)	corx(ξ_x)	cory(ξ_y)		I_{off}	I_{dsat}	I_{dlin}	I_{dlo}	I_{dhi}	V_{tsat}	V_{tlin}	Total
#1	0.625	17.01	173.86	This work	0.0218	0.0129	0.0174	0.0257	0.0131	0.0188	0.0194	0.129
				Non-separated	0.0205	0.0194	0.0207	0.0248	0.0185	0.0194	0.0186	0.142
#2	0.557	80.92	80.57	This work	0.00645	0.0114	0.0102	0.00755	0.0120	0.00576	0.00551	0.0590
				Non-separated	0.00765	0.00692	0.0124	0.00826	0.00740	0.00712	0.00631	0.0561
#3	0.505	48.62	67.99	This work	0.00501	0.00857	0.0111	0.00712	0.00928	0.00601	0.00496	0.0520
				Non-separated	0.00527	0.00695	0.0123	0.00621	0.00785	0.00526	0.00516	0.0491
#4	0.560	25.43	69.40	This work	0.0132	0.0113	0.00684	0.0163	0.0107	0.0141	0.0127	0.0853
				Non-separated	0.00871	0.00511	0.00563	0.0126	0.00507	0.00911	0.00800	0.0543
#5	0.637	69.149	65.33	This work	0.00830	0.0128	0.0102	0.0103	0.0136	0.00808	0.00789	0.0714
				Non-separated	0.00789	0.0111	0.0115	0.0101	0.0113	0.00791	0.00731	0.0672
#6	0.730	89.39	195.62	This work	0.00805	0.00546	0.00772	0.0103	0.00524	0.00789	0.00671	0.0514
				Non-separated	0.0157	0.0129	0.0146	0.0170	0.0120	0.0125	0.0108	0.0960
#7	0.153	23.32	51.25	This work	0.00287	0.00136	0.00348	0.00339	0.00139	0.00284	0.00274	0.0181
				Non-separated	0.00288	0.00597	0.00713	0.00407	0.00596	0.00267	0.00244	0.0311
#8	0.690	85.09	80.98	This work	0.00785	0.0102	0.00740	0.00757	0.0108	0.00838	0.00805	0.0603
				Non-separated	0.0105	0.0109	0.0133	0.0108	0.0107	0.00957	0.00880	0.0749
#9	0.259	78.12	28.08	This work	0.00859	0.00582	0.00552	0.00992	0.00497	0.00677	0.00603	0.0476
				Non-separated	0.00461	0.00318	0.00513	0.00504	0.00327	0.00375	0.00317	0.0281
#10	0.141	96.71	186.68	This work	0.00255	0.00558	0.00526	0.00342	0.00524	0.00231	0.00161	0.0260
				Non-separated	0.00274	0.00438	0.00331	0.00388	0.00422	0.00246	0.00278	0.0237

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3088461, IEEE Access

the previous ML model only did with a shape of Gaussian distribution and linear correlation. Thus, the accuracy of the ANN model is significantly improved in the respect of both quantitative and qualitative comparisons. Especially, we extend the prediction target from 4 parameters (I_{off}, I_{dsat}, V_{tsat}, and SS) to 7 parameters (I_{off}, I_{dsat}, I_{dlin}, I_{dlo}, I_{dhi}, V_{tsat}, and V_{tlin}). This enables simulating electrical behavior of transistor as well as DC behavior of digital circuit blocks such as SRAM bit cell [34]. This work can pave a new road to analyzing the impact of LER, and thereby, to timely design the process integration for integrated circuits.

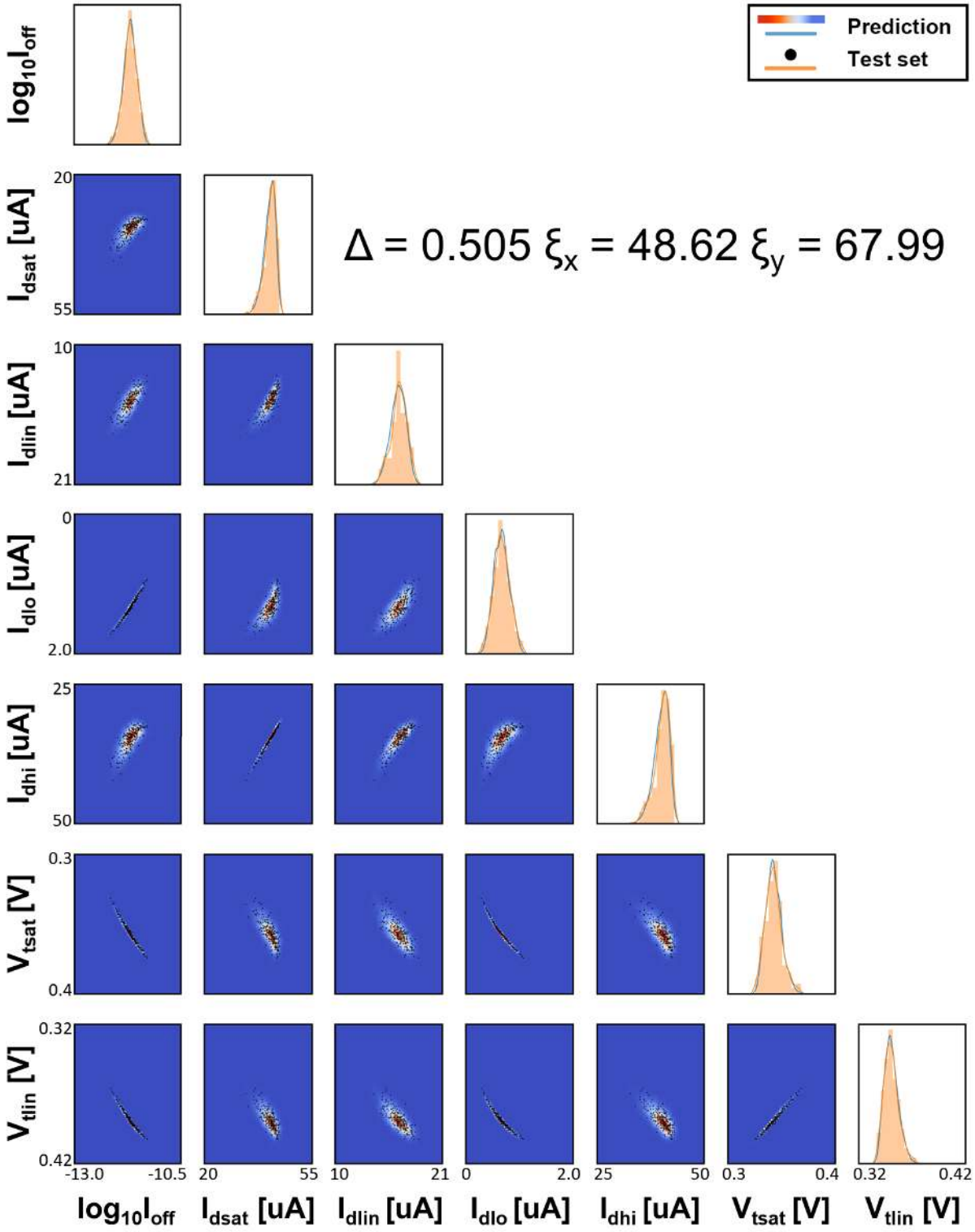

FIGURE 7. Pair plot of performance metrics for (a) Δ = 0.505 nm, Λ_x = 48.62 nm, Λ_y = 67.99 nm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3088461, IEEE Access

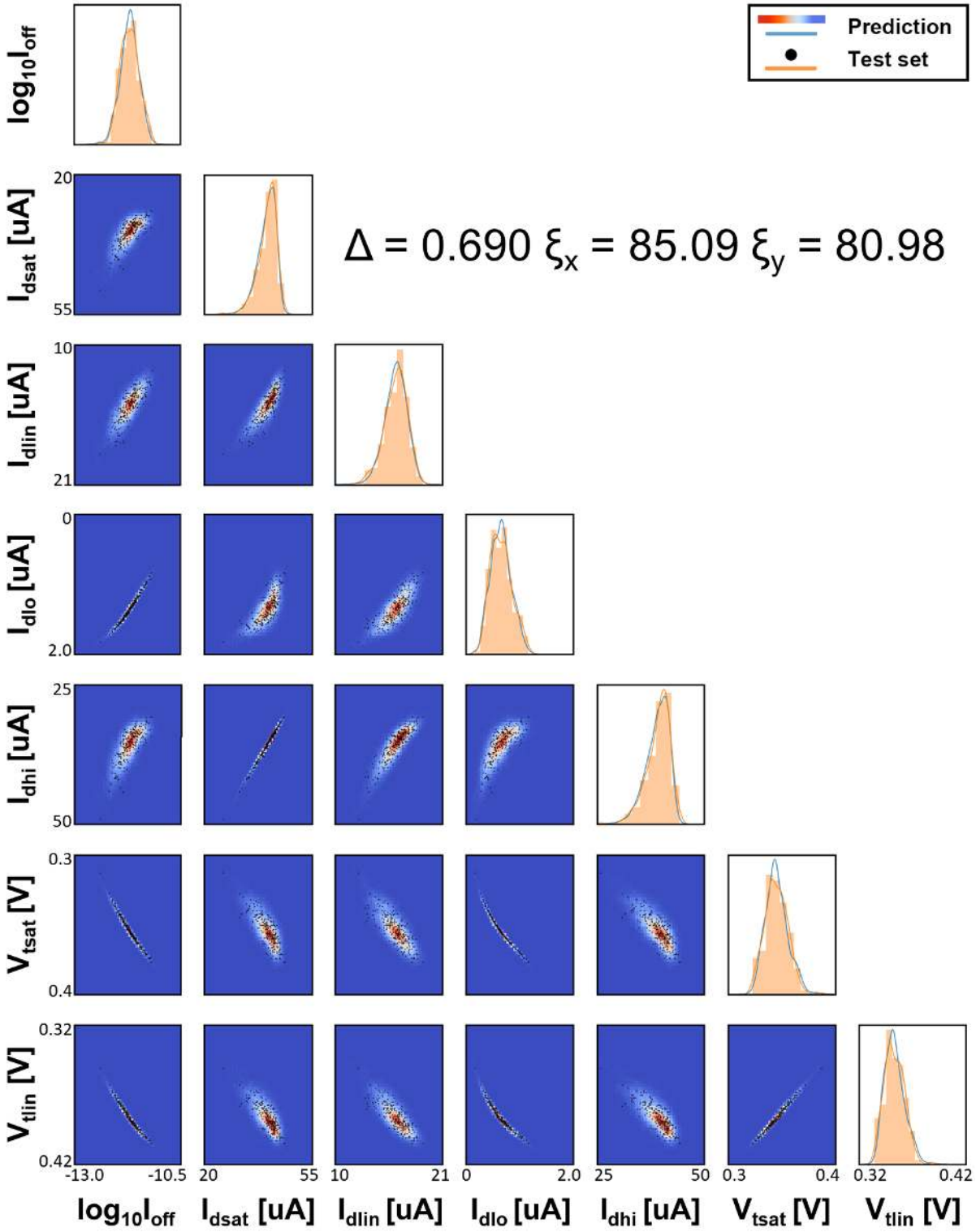

FIGURE 8. Pair plot of performance metrics Δ = 0.690 nm, Λ_x = 85.09 nm, Λ_y = 80.98 nm.

REFERENCES

[1] K. Mistry *et al.*, "Delaying forever: Uniaxial strained silicon transistors in a 90nm CMOS technology," in *Digest of Technical Papers. 2004 Symposium on VLSI Technology, 2004.*, 2004: IEEE, pp. 50-51.

[2] C. Auth *et al.*, "45nm high-k+ metal gate strain-enhanced transistors," in *2008 Symposium on VLSI Technology*, 2008: IEEE, pp. 128-129.

[3] C. Auth *et al.*, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *2012 symposium on VLSI technology (VLSIT)*, 2012: IEEE, pp. 131-132.

[4] K. Agarwal and S. Nassif, "The impact of random device variation on SRAM cell stability in sub-90-nm CMOS technologies," *IEEE Transactions on Very Large Scale Integration Systems,* vol. 16, no. 1, pp. 86-97, 2007.

[5] S. Markov, A. S. M. Zain, B. Cheng, and A. Asenov, "Statistical variability in scaled generations of n-channel UTB-FD-SOI MOSFETs under the influence of RDF, LER, OTF and MGG," in *2012 IEEE International SOI Conference (SOI)*, 2012: IEEE, pp. 1-2.

[6] G. Leung and C. O. Chui, "Interactions between line edge roughness and random dopant fluctuation in nonplanar field-effect transistor variability," *IEEE transactions on electron devices,* vol. 60, no. 10, pp. 3277-3284, 2013.

[7] W.-T. Huang and Y. Li, "The impact of fin/sidewall/gate line edge roughness on trapezoidal bulk FinFET devices," in *2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2014: IEEE, pp. 281-284.

[8] D. S. Holmes *et al.*, "IRDS 2018," 2018.

[9] G. Bae *et al.*, "3nm GAA technology featuring multi-bridge-channel FET for low power and high performance applications," in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018: IEEE, pp. 28.7. 1-28.7. 4.

[10] X. Jiang, X. Wang, R. Wang, B. Cheng, A. Asenov, and R. Huang, "Predictive compact modeling of random variations in FinFET technology for 16/14nm node and beyond," in *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015: IEEE, pp. 28.3. 1-28.3. 4.

[11] M. H. Al Banna *et al.*, "Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges," vol. 8, pp. 192880-192923, 2020.

[12] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, and M. J. B. i. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia," vol. 7, no. 1, pp. 1-21, 2020.

[13] A. Whata and C. Chimedza, "Deep Learning for SARS COV-2 Genome Sequences," *IEEE Access,* 2021.

[14] H.-C. Choi, H. Yun, J.-S. Yoon, and R.-H. Baek, "Neural Approach for Modeling and Optimizing Si-MOSFET Manufacturing," *IEEE Access,* vol. 8, pp. 159351-159370, 2020.

[15] D. Jiang, W. Lin, and N. Raghavan, "A Gaussian Mixture Model Clustering Ensemble Regressor for Semiconductor Manufacturing Final Test Yield Prediction," *IEEE Access,* vol. 9, pp. 22253-22263, 2021.

[16] J.-S. Yoon, S. Lee, H. Yun, and R.-H. Baek, "Digital/Analog Performance Optimization of Vertical Nanowire FETs Using Machine Learning," *IEEE Access,* vol. 9, pp. 29071-29077, 2021.

[17] K. Ko, J. K. Lee, and H. Shin, "Variability-aware machine learning strategy for 3-D NAND flash memories," *IEEE Transactions on Electron Devices,* vol. 67, no. 4, pp. 1575-1580, 2020.

[18] J. Lim and C. Shin, "Machine Learning (ML)-Based Model to Characterize the Line Edge Roughness (LER)-Induced Random Variation in FinFET," *IEEE Access,* vol. 8, pp. 158237-158242, 2020.

[19] S. Oh and C. Shin, "3-D quasi-atomistic model for line edge roughness in nonplanar MOSFETs," *IEEE Transactions on Electron Devices,* vol. 63, no. 12, pp. 4617-4623, 2016.

[20] *TCAD Sentaurus™ User Guide Version P-2019.03*. Synopsys, Inc., Mountain View CA, USA, , 2019.

[21] C. Shin, *Variation-aware advanced CMOS devices and SRAM*. Springer, 2016.

[22] E. Dornel *et al.*, "Hydrogen annealing of arrays of planar and vertically stacked Si nanowires," vol. 91, no. 23, p. 233502, 2007.

[23] T. Tezuka *et al.*, "{110}-facets formation by hydrogen thermal etching on sidewalls of Si and strained-Si fin structures," vol. 92, no. 19, p. 191903, 2008.

[24] Y. Ma, H. J. Levinson, and T. Wallow, "Line edge roughness impact on critical dimension variation," in *Metrology, Inspection, and Process Control for Microlithography XXI*, 2007, vol. 6518: International Society for Optics and Photonics, p. 651824.

[25] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, 1998: IEEE, pp. 59-66.

[26] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness," *IEEE Transactions on Electron Devices,* vol. 50, no. 5, pp. 1254-1260, 2003.

[27] J. Min and C. Shin, "Study of line edge roughness on various types of gate-all-around field effect transistor," *Semiconductor Science Technology,* vol. 35, no. 1, p. 015004, 2019.

[28] Y.-N. Chen *et al.*, "Impacts of work function variation and line-edge roughness on TFET and FinFET devices and 32-bit CLA circuits," vol. 5, no. 2, pp. 101-115, 2015.

[29] S. Kaya, A. Brown, A. Asenov, D. Magot, and T. LintonI, "Analysis of statistical fluctuations due to line edge roughness in sub-0.1 μm MOSFETs," in *Simulation of Semiconductor Processes and Devices 2001*: Springer, 2001, pp. 78-81.

[30] J.-X. Pan and K.-T. Fang, "Maximum likelihood estimation," in *Growth curve models and statistical diagnostics*: Springer, 2002, pp. 77-158.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980,* 2014, [online] Available: http://arxiv.org/abs/1412.6980.

[32] M. Abadi *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," 2016.

[33] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265-283.

[34] A. E. Carlson, *Device and circuit techniques for reducing variation in nanoscale SRAM*. University of California, Berkeley, 2008.