Received August 13, 2020, accepted August 19, 2020, date of publication August 28, 2020, date of current version September 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3020066

Machine Learning (ML)-Based Model to Characterize the Line Edge Roughness (LER)-Induced Random Variation in FinFET

JAEHYUK LIM AND CHANGHWAN SHIN, (Senior Member, IEEE)
Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South 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 Ministry of Science and ICT (MSIT) (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).

ABSTRACT ML (Machine Learning)-based artificial neural network (ANN) model is proposed to estimate the LER (line edge roughness)-induced performance variation in Fin-shaped Field Effect Transistor (FinFET). For a given LER features such as rms amplitude(Δ), correlation length along x-direction (Λ_X), and correlation length along y-direction (Λ_Y), the metrics for device performance such as on-state drive current, off-state leakage current, threshold voltage, and subthreshold swing can be computing-efficiently estimated with the ANN model.

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

I. INTRODUCTION

For the last a few decades, complementary metal oxide semiconductor (CMOS) technology has been successfully evolved with the adoption of new techniques such as stress engineering in 90 nm technology node and beyond [1], high-k/metal-gate in 45 nm technology node and beyond [2], and 3-D advanced device structure in 22 nm technology node and beyond [3]. In every new CMOS technology platform, the physical dimension of metal oxide semiconductor field effect transistor (MOSFET) has been scaled down not only to increase the density of devices in integrated circuit (IC) but also to improve the functions of IC per cost. However, process-induced random variations (i.e., transistors' electrical characteristics such as threshold voltage, on-state drive current, and off-state leakage current, are randomly fluctuated/affected while fabricating transistors in FAB), have negatively affected the manufacturability of CMOS devices, and thereby, it would significantly hinder the evolution of CMOS technology [4]. The root-causes of process-induced random variation are classified as (i) line edge roughness (LER), (ii) random dopant fluctuation (RDF), and (iii) work function variation (WFV) [5]. Especially, LER would degrade the device performance but also indirectly affect the other random variation sources (i.e., RDF and WFV) because it induces structural variations in device [6]. With the most radical shift in device structure in the year of 2011, i.e., from planar bulk MOSFET to 3-D MOSFET (i.e., FinFET), the process-induced technical issues become much more severe [7]. Therefore, as the device architecture becomes more complicated (in reality, multiple bridge channel field effect transistor (MBCFET), stacked nano-wire FET, stacked nano-slab FET, etc. for 3 nm CMOS technology node [8] and beyond), understanding the impact of LER on device performance is desperately required in developing variation-robust silicon device at 3 nm technology node and beyond [9].

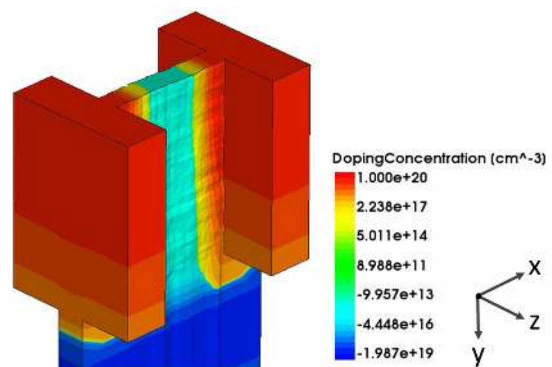

FIGURE 1. A bird's-eye view of FinFET with a 3-D LER on its sidewall. LER parameters used in this example are as follows: $\Delta = 0.5$ nm, $\Lambda_X = 20$ nm, $\Lambda_Y = 50$ nm, $\alpha = 1$, $\Theta = 0$.

The associate editor coordinating the review of this manuscript and approving it for publication was Sneh Saurabh.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

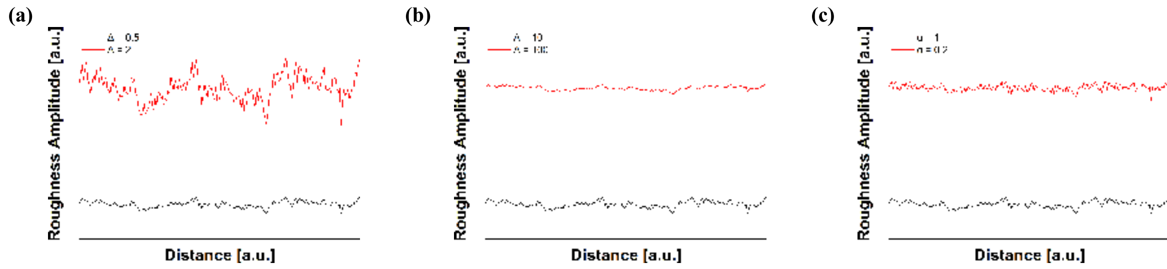

FIGURE 2. Examples of roughness amplitude when (a) $\Lambda = 10$, $\alpha = 1$, (b) $\Delta = 0.5$, $\alpha = 1$, and (c) $\Delta = 0.5$, $\Lambda = 10$.

TABLE 1. Device parameters of The FinFET used in simulation [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_{RC}	Peak concentration of retrograde channel doping profile	cm^{-3}	2×10^{19}

TABLE 2. Performance metrics.

Symbol	Description	Extraction Criteria
I_{off}	Off-state leakage current	I_d at $V_g = 0$
V_t	Threshold voltage	V_g at $I_d = 535$ [nA]
I_{on}	On-state drive current	I_d at $V_g = 0.7$ [V]
SS	Subthreshold swing	$\dfrac{0.01[V]}{log_{10}Id_{V_g=0.01[V]} - log_{10}Ioff}$

Note that I_d is the drain current, and V_g is the gate voltage.

A few studies have reported to understand, quantify, and analyze the impacts of LER on device characteristics [10]–[12]. TCAD (Technology Computer Aided Design)-based method has been adopted to propose model for finely and accurately predicting the impact of LER [13]. However, the TCAD-based approach is fundamentally very time-consuming and computationally-inefficient when predicting thousands of LER-induced input transfer characteristics of MOSFETs in integrated circuit. Thus, a few studies [14], [15] have tried to compactly model the impact of LER on the device performance. Nevertheless, due to many technical barriers in developing a new compact model, the compact model for analyzing the impact of LER [14], [15] would not be timely developed, even though the LER on the fin sidewall of FinFET should be modeled for two-dimensionally characterizing/understanding the sidewall surface [7], [13]. Therefore, using Machine Learning (ML) technique, simple but eye-catching novel approach with reasonable accuracy is proposed in this work, to provide an alternative device solution for predicting the process-induced variation.

II. DEVICE DESIGN AND DATA GENERATION

A. LINE EDGE ROUGHNESS PARAMETERS

Generally, 2 or 3 parameters (e.g., Δ, Λ, and α) are used to describe the LER profile in planar MOSFETs, and 3 or 4 parameters (e.g., Δ, Λ_x, Λ_y, and α) are used in 3-D MOSFETs. The impact of each parameter in LER profile is comparatively described in Fig. 2. The details of each parameter used in Fig. 2 are as below [16]:

(i) Amplitude (Δ): the root-mean-square(rms) value of roughness amplitude. The smaller Δ is, the smoother the surface is.

(ii) Correlation length (Λ): how closely the correlated edge is associated to its neighboring edge. The larger Λ is, the smoother the surface is.

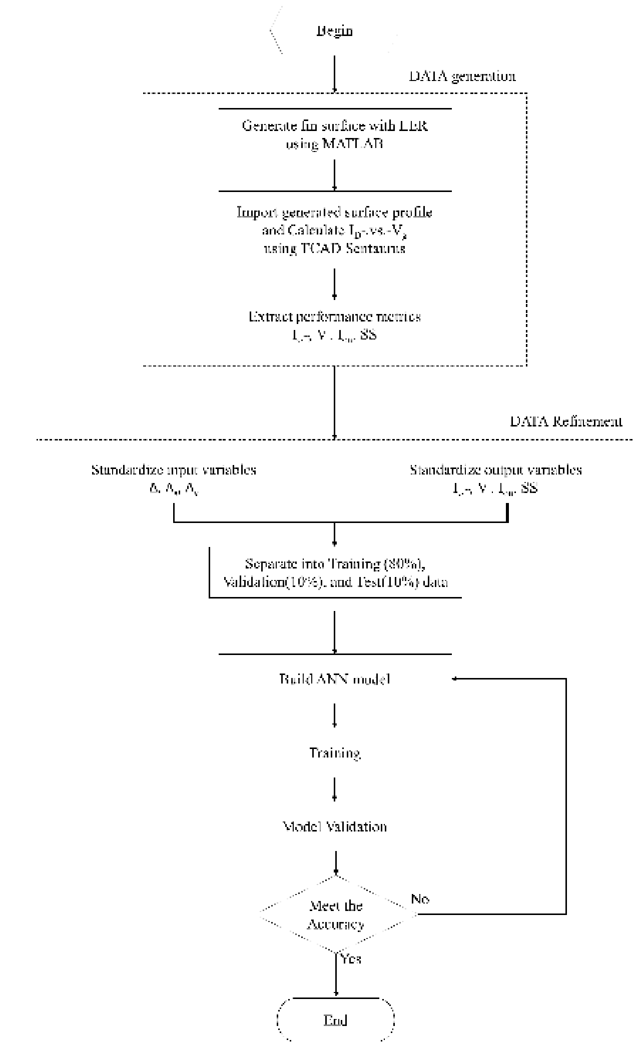

FIGURE 3. The flow chart how to build/train/test the ANN model.

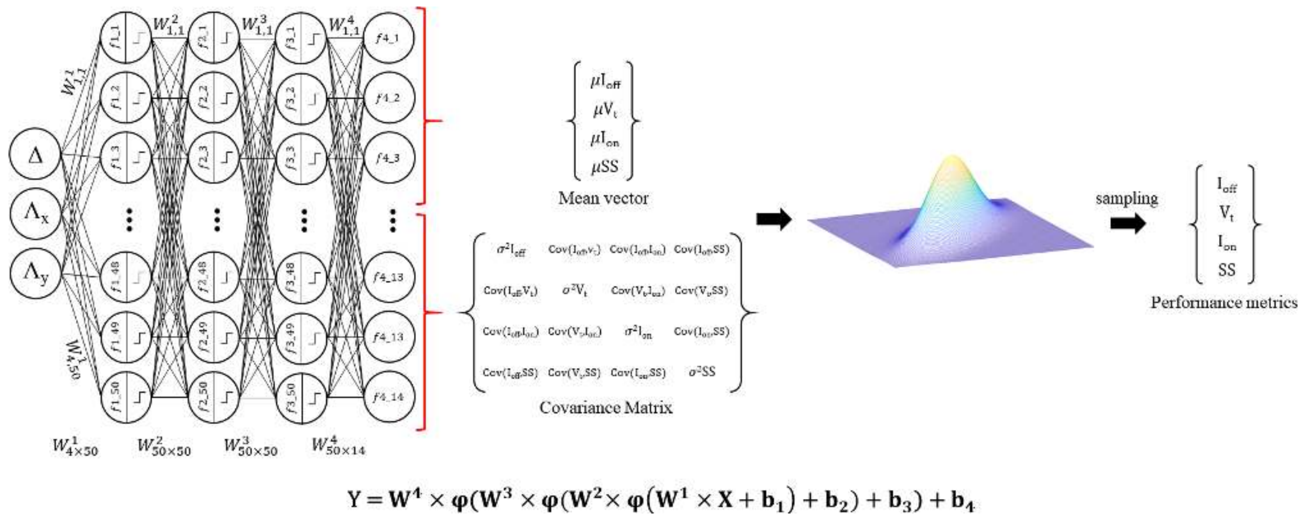

FIGURE 4. Artificial neural network (ANN) structure: The model has 3 input variables and 4 output variables which are randomly sampled from probability distribution function. The probability distribution function is determined by the mean vector and covariance matrix.

(iii) Roughness exponent (α): the high frequency component of roughness. The larger α is, the smoother the surface is.

B. DEVICE DESIGN WITH LINE EDGE ROUGHNESS

A three-dimensional (3-D) bird's-eye view of FinFET with a 3-D LER on its sidewall fin is shown in Fig.1. The device design parameters of nominal FinFET device are summarized in Table 1. To reconfigure the surface roughness on the sidewall fin of FinFET, the quasi-atomistic model [13] was used. The steps to generate a rough surface are as below:

Step I: Define key parameters such as Δ, Λ_x, Λ_y, α, Θ of 2-D ACVF [see (1) below].

Step II: Obtain the 2-D power spectrum by taking the fast Fourier transformation (FFT).

Step III: Obtain the amplitude spectrum by taking the square root of the result in Step II.

Step IV: Obtain the 2-D impulse response by taking the inverse fast Fourier transformation (IFFT) on the result in Step III.

Step V: Generate the white Gaussian noise (wgn) and take the 2-D convolution of the result in Step IV and wgn.

Step VI: Once the steps above are done, import the generated surface coordinates to TCAD with Sentaurus Structure Editor.

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

In (1), Λ_x and Λ_y are the correlation length along x-direction and y-direction of surface, respectively. Θ determines the relation between x and y direction.

C. DATA GENERATION

To build and verify the Artificial Neural Network (ANN) model, 100 different data sets (note that each data set consists of 50 different FinFETs with identical LER parameters) were first created. To generate those data sets of FinFET device, three LER parameters (i.e., Δ, Λ_x, Λ_y) should be determined. Herein, based on the previous experimental data [16]–[19], a reference LER parameter set was first chosen; $\Delta = 0.5$ nm, $\Lambda_x = 20$ nm, $\Lambda_y = 50$ nm, $\alpha = 1$, $\Theta = 0$. Afterwards, the value of three LER parameters (Δ, Λ_x, Λ_y) are randomly chosen from a given range for each LER parameter, as follows: Δ from 0.2 nm to 0.8 nm, Λ_x from 10 nm to 100 nm, and Λ_y from 20 nm to 200 nm. The distribution of each LER parameter in the limited range follows the uniform distribution. Note that α is set to 1, and Θ is set to 0. In fact, in order to take account into the impact of α on a LER profile, a very small sampling distance is necessary.

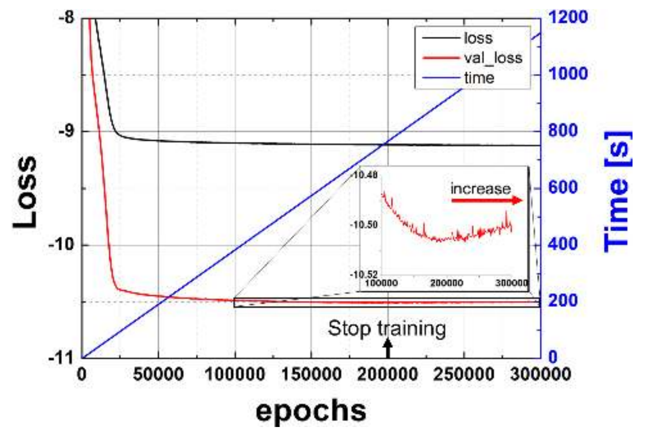

FIGURE 5. Training, validation loss, and time vs. training epochs.

TABLE 3. Time comparison between TCAD and ANN.

No. of Data	Time [hr]		Ratio
	TCAD	ANN	
1,000	62.5	312.8	0.2
5,000	312.5	312.8	1
10,000	625	312.8	2
30,000	1875	312.8	6
50,000	3125	312.9	10
100,000	6250	313	20

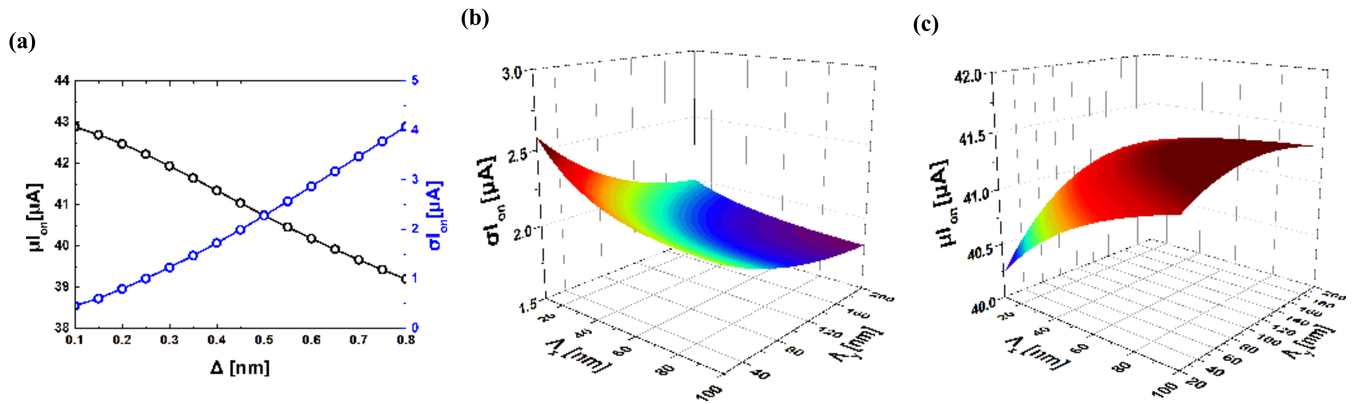

FIGURE 6. (a) Mean and standard deviation of Ion by amplitude(Δ) when $\Lambda_x = 20$ and $\Lambda_y = 50$, (b) standard deviation and (c) mean of Ion by correlation length X(Λ_x) and Y(Λ_y) when $\Delta = 0.5$.

TABLE 4. Mean and standard deviation of the performance metrics.

Data set		V_t [mV] μ	V_t [mV] σ	$\log_{10}(I_{off})$ μ	$\log_{10}(I_{off})$ σ	I_{on} [uA] μ	I_{on} [uA] σ	SS [mV/dec] μ	SS [mV/dec] σ
Feature #1: Δ = 0.246 nm, Λ_x = 79.43 nm, Λ_y = 159.69 nm									
TCAD		342.3	3.809	-11.677	0.0817	42.50	0.8547	61.61	0.1805
ML generated sample #1		341.3	2.755	-11.656	0.0621	42.74	0.7650	61.66	0.1563
ML generated sample #2		340.7	3.620	-11.642	0.0770	42.72	0.8609	61.69	0.1717
Population prediction		340.7	3.415	-11.666	0.0727	42.71	0.7765	61.64	0.1647
Confidence interval	Max	342.8	4.282	-11.635	0.0914	42.99	0.9757	61.70	0.2070
	Min	340.3	2.533	-11.688	0.0541	42.43	0.5772	61.58	0.1224
Feature #2: Δ = 0.715 nm, Λ_x = 30.47 nm, Λ_y = 164.03 nm									
TCAD		345.3	11.145	-11.70	0.2221	39.57	3.0965	61.74	0.4789
ML generated sample #1		345.8	12.378	-11.71	0.2551	39.61	3.0842	61.73	0.5814
ML generated sample #2		346.0	10.538	-11.72	0.2141	39.71	2.8570	61.69	0.4738
Population prediction		345.2	12.105	-11.71	0.2456	40.08	3.1170	61.72	0.5439
Confidence interval.	Max	349.6	15.211	-11.79	0.3086	41.20	3.9167	61.91	0.6834
	Min	340.1	8.975	-11.61	0.1820	38.96	2.3110	61.52	0.4032

Confidence interval of 99% is used to evaluate the predictions of ANN model.

However, the small sampling distance should cause the tremendous amount of computational works in TCAD simulation runs. In real, α is usually out of sight in many other studies on LER [11], [14], [15], [20]. Regarding Θ, we set Θ as 0, for simplicity. This means that the roughness along x-direction is independent of that along y-direction. Then, I_d-vs.-V_g characteristic of all FinFETs in 100 different data sets were simulated using the TCAD, and thereafter, the performance metrics (e.g., I_{off}, V_t, I_{on}, SS) were extracted out [see Table 2].

Those data sets were separated into three groups: training data sets, validation data sets, and test data sets. The training data sets are used to update the ANN model components such as weight matrices and bias vectors. The validation data sets are used to monitor if the ANN model is well trained or over-fitted in the training process. After the training process is finished, the test data sets are used to verify if the ANN model is well trained or not [see Fig. 3].

III. ARTIFICIAL NEURAL NETWORK MODELING

A. FULLY CONNECTED LAYERS

This ANN model has 1 input layer, 1 output layer and 3 hidden layers with 3 activation functions (φ), [see Fig. 4]. The hyperbolic tangent (tanh) is used for activation functions. It is mathematically defined as follows:

$$tanh = \frac{e^{2x} - 1}{e^{2x} + 1} \tag{2}$$

Weight matrices ($W^1, W^2, \ldots, W^4$) and bias vectors ($b^1, b^2, \ldots, b^4$) of ANN model can determine outputs. When training the ANN model, those matrices and vectors are

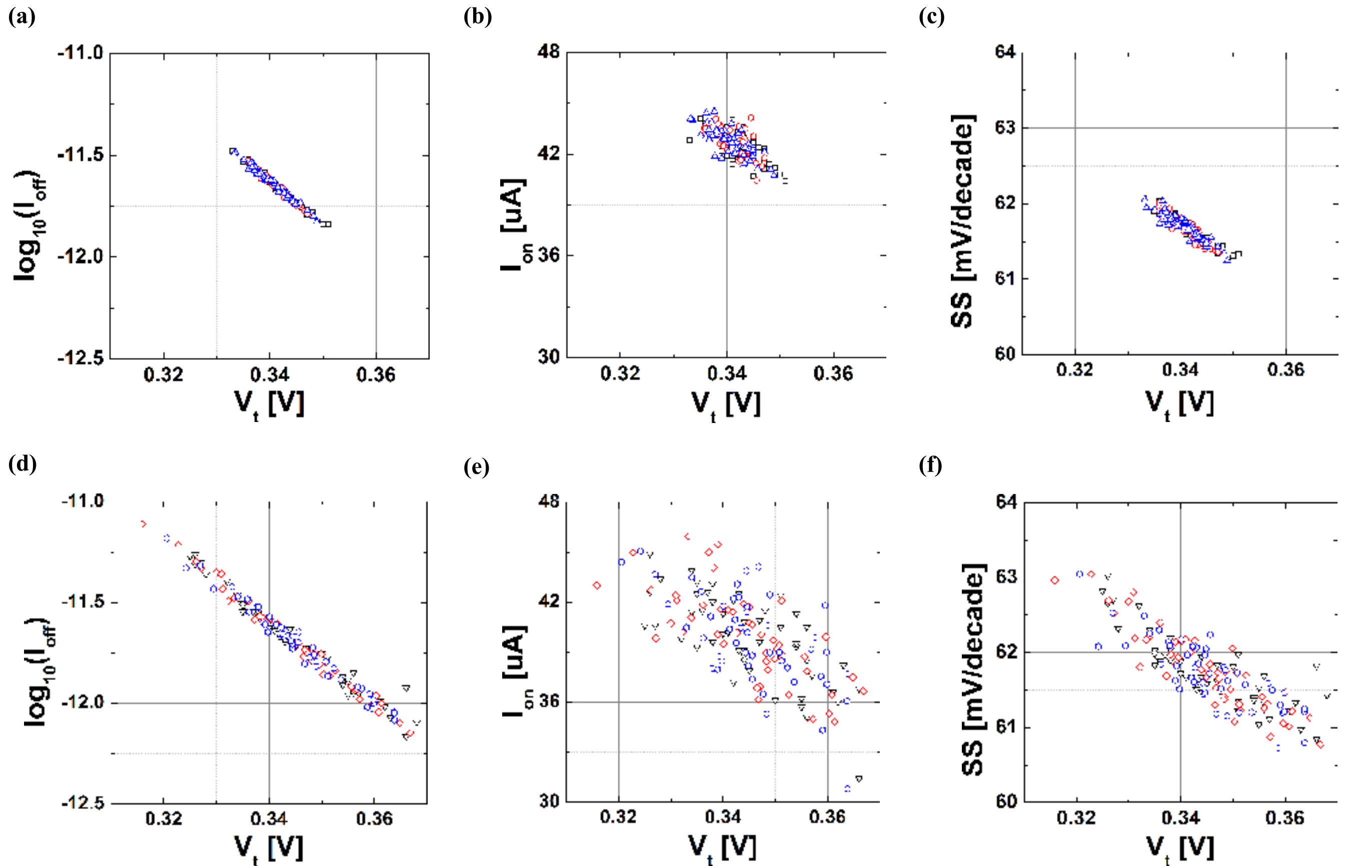

FIGURE 7. Test data set from TCAD vs. data set predicted by ML model. For the given LER profile (i.e., Δ = 0.246 nm, Λ_x = 79.43 nm, Λ_y = 159.69 nm), (a) $\log_{10}(I_{off})$-vs.-V_t, (b) I_{on}-vs.-V_t, (c) SS-vs.-V_t. For the other LER profile (i.e., Δ = 0.715 nm, Λ_x = 30.47 nm, Λ_y = 164.03 nm), (d) $\log_{10}(I_{off})$-vs.-V_t, (e) I_{on}-vs.-V_t, (f) SS-vs.-V_t. Note that black-colored square, red-colored circle, and blue-colored triangle indicates TCAD sample, ML sample 1, and ML sample 2, respectively.

updated in order to be fit to the training data sets for specified number of iterations.

B. GRAFTING PROBABILITY DISTRIBUTION

In this study, we assumed that the distribution of performance metrics follows the multi-variate Gaussian distribution to securely build the model for estimating the LER-induced performance variation of device. It is known that the LER-induced variation of V_t, I_{on}, SS, and $\log_{10}I_{off}$ approximately follows the Gaussian distribution in various devices [11], [21], [22].

To train the ANN model with probabilistic layer, we used Maximum likelihood estimation (MLE) method. Based on the observation (e.g., Y), the MLE method is a technique for estimating parameter θ, when there is the input X. In other words, the final goal in this method is to find θ that maximizes $\boldsymbol{P(Y|X;\theta)}$ or can be mathematically rewritten as in (3):

$$\theta_{ML} = arg\max_{\theta} P\left(Y\,|\,X;\theta\right) \tag{3}$$

The parameters such as X, Y, and θ can be redefined in our model as follows:

X: Δ, Λ_x, and Λ_y (LER parameters)

Y: $\{y_1, y_2, \ldots y_{50}\}$, y_i: observed I_{off}, V_t, I_{on}, and SS

θ: mean vector and covariance matrix

To train the probability-grafted ANN, we used "Negative log likelihood" (negloglik) as a loss function. Negloglik notifies how much two other distributions are different from each other.

Using Adam Optimizer [23], the training process was executed for 200,000 epochs (776 sec) with learning rate of 10^{-5}. The model was trained without overfitting [see Fig. 5].

IV. RESULTS AND EVALUATION

Fig. 6 shows how I_{on} is varied with modifying the LER parameters. Table 4 and Fig. 7 show the comparison between the TCAD data (=test data set) and the prediction data by the ANN model. Based on the probability density function determined by the mean vector and covariance matrix, the prediction data was "randomly" extracted. Hence, they are slightly different from TCAD samples, but they can never be identical to TCAD samples. Thus, the accuracy of prediction data was evaluated using the confidence interval calculated by the standard error of mean and standard deviation [24]. Herein, the predicted values of population mean and standard deviation by the ANN model are considered as the true population mean and standard deviation.

$$\text{Standard error of mean} = \frac{\sigma}{\sqrt{n}} \tag{4}$$

n : number of samples in 1 set of data.

$$\text{Standard error of standard deviation} \approx \frac{\sigma}{\sqrt{2\,(n-1)}} \tag{5}$$

Table 3 shows the comparison of simulation time of TCAD vs. ANN. It is noteworthy that the advantage of using the ANN model becomes conspicuous when the number of data (or the size of data) is bigger than 10,000 or more. Note that the ANN model was built using the Tensorflow 2.0 and Tensorflow-probability python library [25], [26].

V. CONCLUSION

Line edge roughness (LER) is one of key sources inducing undesirable variation in transistor performance. These undesirable fluctuations affect the operation of circuit, and thereby, they can cause unexpected errors. Therefore, it is important to understand the factors causing the random variation in an accurate manner within reasonable time. In FinFET, the structural deformation by LER appears not as a shape of line but plane. Thus, the compact modeling method would not be the right option for solving a problem with increased complexity. To avoid these difficulties, we used the ANN model and suggest alternatives to predict the process-induced random fluctuations. With accurate predictions (which meets the confidence interval of 99%), our method is expected to help analyze the effects of LER in fabrication process and to evaluate yield of integrated circuit (IC).

REFERENCES

[1] K. Mistry, M. Armstrong, C. Auth, S. Cea, T. Coan, T. Ghani, T. Hoffmann, A. Murthy, J. Sandford, R. Shaheed, K. Zawadzki, K. Zhang, S. Thompson, and M. Bohr, "Delaying forever: Uniaxial strained silicon transistors in a 90 nm CMOS technology," in *Symp. VLSI Technol. Dig. Tech. Papers*, Honolulu, HI, USA, 2004, pp. 50–51.

[2] C. Auth, "45 nm high-k + metal gate strain-enhanced CMOS transistors," in *Proc. IEEE Custom Integr. Circuits Conf.*, San Jose, CA, USA, Sep. 2008, pp. 379–386.

[3] C. Auth *et al.*, "A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *Proc. Symp. VLSI Technol. (VLSIT)*, Honolulu, HI, USA, Jun. 2012, 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 Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 1, pp. 86–97, Jan. 2008.

[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 *Proc. IEEE Int. SOI Conf. (SOI)*, Napa, CA, USA, Oct. 2012, 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 Trans. Electron Devices*, vol. 60, no. 10, pp. 3277–3284, Oct. 2013.

[7] W.-T. Huang and Y. Li, "The impact of fin/sidewall/gate line edge roughness on trapezoidal bulk FinFET devices," in *Proc. Int. Conf. Simulation Semiconductor Process. Devices (SISPAD)*, Sep. 2014, pp. 281–284.

[8] M. Badaroglu. (2018). *International Roadmap for Device and Systems (IRDS)*. [Online]. Available: http://irds.ieee.org

[9] G. Bae *et al.*, "3 nm GAA technology featuring multi-bridge-channel FET for low power and high performance applications," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2018, pp. 28.7.1–28.7.4.

[10] F. Adamu-Lema, X. Wang, S. M. Amoroso, C. Riddet, B. Cheng, L. Shifren, R. Aitken, S. Sinha, G. Yeric, and A. Asenov, "Performance and variability of doped multithreshold FinFETs for 10-nm CMOS," *IEEE Trans. Electron Devices*, vol. 61, no. 10, pp. 3372–3378, Oct. 2014.

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

[12] E. Baravelli, A. Dixit, R. Rooyackers, M. Jurczak, N. Speciale, and K. De Meyer, "Impact of line-edge roughness on FinFET matching performance," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2466–2474, Sep. 2007.

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

[14] Amita, S. Mittal, and U. Ganguly, "The first compact model to determine V_T distribution for DG-FinFET due to LER," *IEEE Trans. Electron Devices*, vol. 65, no. 11, pp. 4772–4779, Nov. 2018.

[15] X. Jiang, X. Wang, R. Wang, B. Cheng, A. Asenov, and R. Huang, "Predictive compact modeling of random variations in FinFET technology for 16/14 nm node and beyond," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2015, pp. 28.3.1–28.3.4.

[16] C. Shin, *Variation-Aware Advanced CMOS Devices and SRAM*, vol. 56. Dordrecht, The Netherlands: Springer, 2016.

[17] E. Dornel, T. Ernst, J. C. Barbé, J. M. Hartmann, V. Delaye, F. Aussenac, C. Vizioz, S. Borel, V. Maffini-Alvaro, C. Isheden, and J. Foucher, "Hydrogen annealing of arrays of planar and vertically stacked Si nanowires," *Appl. Phys. Lett.*, vol. 91, no. 23, Dec. 2007, Art. no. 233502.

[18] T. Tezuka, N. Hirashita, Y. Moriyama, N. Sugiyama, K. Usuda, E. Toyoda, K. Murayama, and S.-I. Takagi, "110-facets formation by hydrogen thermal etching on sidewalls of Si and strained-Si fin structures," *App. Phys. Lett.*, vol. 92, no. 19, 2008, Art. no. 191903.

[19] Y. Ma, H. J. Levinson, and T. Wallow, "Line edge roughness impact on critical dimension variation," *Proc. SPIE*, vol. 6518, Apr. 2007, Art. no. 651824.

[20] M. Martin and G. Cunge, "Surface roughness generated by plasma etching process of silicon," *J. Vac. Sci. Technol. B, Microelectron. Nanometer Struct. Process., Meas., Phenomena*, vol. 26, no. 4, pp. 1281–1288, 2008.

[21] J. Min and C. Shin, "Study of line edge roughness on various types of gate-all-around field effect transistor," *Semicond. Sci. Technol.*, vol. 35, no. 1, Jan. 2020, Art. no. 015004.

[22] Y.-N. Chen, C.-J. Chen, M.-L. Fan, V. Hu, P. Su, and C.-T. Chuang, "Impacts of work function variation and line-edge roughness on TFET and FinFET devices and 32-bit CLA circuits," *J. Low Power Electron. Appl.*, vol. 5, no. 2, pp. 101–115, May 2015, doi: 10.3390/jlpea5020101.

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

[24] S. Ahn and J. A. Fessler, "Standard errors of mean, variance, and standard deviation estimators," Dept. EECS, Univ. Michigan, Ann Arbor, MI, USA, Tech. Rep. 413, Jul. 2003.

[25] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467

[26] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," 2016, *arXiv:1605.08695*. [Online]. Available: http://arxiv.org/abs/1605.08695