

# Deep Learning Approach to Modeling and Exploring Random Sources of Gate-All-Around Silicon Nanosheet MOSFETs

Rajat Butola<sup>1,2</sup>, Yiming Li<sup>1-6,\*</sup>, and Sekhar Reddy Kola<sup>1,2</sup>

<sup>1</sup>Parallel and Scientific Computing Laboratory, <sup>2</sup>Electrical Engineering and Computer Science International Graduate Program, <sup>3</sup>Institute of Communications Engineering, <sup>4</sup>Department of Electrical Engineering and Computer Engineering, <sup>5</sup>Institute of Biomedical Engineering, and

<sup>6</sup>Center for mmWave Smart Radar System and Technologies, National Yang-Ming Chiao Tung University, Hsinchu 300, Taiwan, 1001 Ta-Hsueh Rd., Hsinchu 300, Taiwan; \*Tel: +886-3-5712121 ext. 52974; Fax: +886-3-5726639; Email: yml@nycu.edu.tw

**Abstract**—In this paper, for the first time, deep learning (DL) based artificial neural network (ANN) is applied to model the effects of various random variations: work function fluctuation, random dopant fluctuation, and interface trap fluctuation, on gate-all-around silicon nanosheet MOSFETs. The number of fluctuations for each source variation is used as input features and their effects on devices of interest are studied qualitatively and quantitatively. The key figures of merit (FoM) are also extracted accurately from the transfer characteristics, which shows the competency of the ANN model in the domain of device modeling.

## I. INTRODUCTION

The challenges of transistor scaling are becoming increasingly vital with the advancement of device technology nodes. It has turned out to be more difficult for device manufacturers to deliver better performance and reduced power at low device area and cost. The transistor structures are modified to achieve continuous scaling. Gate all around (GAA) transistor is one such modified structure in which the gate wraps the channel from all the sides and provides continuous scaling [1]. Among the different variants of GAA, silicon (Si) nanosheet (NS) is appeared to be the best structure, to construct future transistors [2]. But the high-scale devices like NS are generally suffered from various fluctuations such as work function fluctuation (WKF), random dopant fluctuation (RDF), interface trap fluctuation (ITF), line edge roughness (LER) [3]-[5], etc. It is very important in nanodevices to consider and study these effects. Therefore, in this paper, to analyze the impacts of these fluctuations on newly discovered nanodevices, we incorporate DL, precisely the ANN model. We consider three different fluctuations WKF, RDF and ITF simultaneously and collectively investigate their combined effect on electrical characteristics for 3 nm node technology.

## II. DEVICE STRUCTURE AND THE ANN APPROACH

The 3D structure of explored GAA Si NS MOSFET device is shown in Fig. 1(a) with various random fluctuation sources utilized in this work. The 2D cross-section view at the channel of the device structure is also shown in Fig. 1(a). The NS device structures are simulated after validating with the experimental calibration. A total of 1000 GAA Si NS fluctuated devices are simulated with statistically generated random sources (WKF, RDF, and ITF), having a different number of counts for each device. The metal grains of high work function (WK<sub>H</sub>) and low work function (WK<sub>L</sub>) are introduced with 60% and 40% probability, respectively, to generate WKF. For RDF, the channel (CH), source extension (SEXT), drain extensions (D<sub>EXT</sub>) and channel penetration (PE) concentrations are added with  $5 \times 10^{17} \text{ cm}^{-3}$ ,  $1.1 \times 10^{19} \text{ cm}^{-3}$ ,  $1.1 \times 10^{19} \text{ cm}^{-3}$  and  $3.36 \times 10^{17} \text{ cm}^{-3}$ , respectively. Similarly, for ITF, traps of size  $2 \text{ nm}^2$  with density ranging from  $1.5 \times 10^{13}$  to  $7.6 \times 10^{13} \text{ cm}^{-2} \text{ eV}^{-1}$  are used. The physical device parameters and doping profiles of all the fluctuated NS devices are kept the same and listed in Fig. 1(b). The overall flow chart of integration of device simulation and DL algorithm is shown in Fig. 2. The ANN architecture, illustrated in Fig. 3(a), has one input layer (7 neurons), two hidden layers (60 neurons each), and one output layer (36 neurons). The list of hyperparameters used in the ANN model is summarized in Fig. 3(b). The biasing condition for generating the characteristics are as follows:  $V_D = 0.7 \text{ V}$ ,  $V_G$  is swept from  $0 \text{ V} - 0.7 \text{ V}$  with a step size ( $\Delta$ ) of  $0.02 \text{ V}$ . For the ANN model, the input features consist of the count of fluctuations [WK<sub>H</sub>, WK<sub>L</sub>, ITF, CH, SEXT, D<sub>EXT</sub>, PE] and the output data structure consists of 36 [ $I_{D1}, I_{D2}, \dots, I_{D36}$ ] drain current values corresponding to each  $V_G$  [ $0 \text{ V}, 0.02 \text{ V}, \dots, 0.7 \text{ V}$ ] value.

## III. RESULTS AND DISCUSSION

As we know, training (Train) the DL model is referring to feeding the model to learn its hidden parameters, whereas testing (Test) refers to prediction on new data from the trained DL model. The dataset preprocessing techniques, such as standardization is performed before feeding the data into the DL model. In this work, the standardization is only performed on the target ( $I_D$ ) values. After preprocessing, the number of fluctuations of WK, RD, and IT is fed into our explored DL model. While training the ANN model, the loss values are calculated in each epoch to optimize the weights and biases of the hidden layers, are shown in Fig. 3(c). For the training dataset, the simulated and the predicted  $I_D$ - $V_G$  curves are shown in Fig. 4 (a). It can be observed that both  $I_D$ - $V_G$  curves are in good agreement with each other. The performance of a well-trained ANN model is evaluated using an unknown dataset. Similarly, for the testing dataset, the simulated and the predicted transfer characteristics are overlapping to each other, as shown in Fig. 4(b). For better visualization, randomly two fluctuated devices are selected from the testing dataset and the performance of the well-trained ANN model is evaluated, as shown in Fig. 4(c). Moreover, Fig. 4(c') shows the number of fluctuations of WK, RD and IT, for both fluctuated devices. Furthermore, the performance of the ANN model is evaluated using root mean squared error (RMSE). The smaller the error, the better will be the performance of the DL model and vice versa. The mathematical notation as well as the comparison between train and test RMSE values are shown in Fig. 4(d). The key figure of merit (FoM) such as transconductance ( $g_m$ ), on-current ( $I_{ON}$ ), and off-current ( $I_{OFF}$ ), and characteristic fluctuations are also determined from the simulated and the predicted  $I_D$ - $V_G$  curves. In Figs. 5(a)-(c), along with linear ideal curve, true versus predicted  $I_{ON}$ ,  $I_{OFF}$  and  $g_m$  are shown, respectively. Similarly, Figs. 5(a')-(c') show the fluctuation of true and predicted  $I_{ON}$ ,  $I_{OFF}$ , and  $g_m$ , respectively. Results show the potential of the ANN model in advanced nanodevice modeling due to its excellent fitting capability.

## CONCLUSIONS

In summary, to the best of our knowledge, the DL method is utilized, for the first time, to determine the electrical characteristics of GAA Si NS MOSFET by analyzing the effect of the number of fluctuations of WK, RD, and IT. We established that without compromising the accuracy, the DL modeling is proved to be an inexpensive way to explore the relationship among different source variations and the transfer characteristics. We are currently exploring more advanced emerging nanodevices using several DL techniques.

## ACKNOWLEDGMENT

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 110-2221-E-A49-139, Grant MOST 109-2221-E-009-033 and Grant MOST 109-2634-F-009-030, and in part by the “Center for mm-Wave Smart Radar Systems and Technologies” under the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan.

## REFERENCES

- [1] J. Ajayan, *et al.*, *Microelectron. Eng.*, vol. 114, pp. 1-12, Aug. 2021.
- [2] N. Loubet, *et al.*, *Symposium on VLSI Technology*, pp. 230-231, Aug. 2017.
- [3] S. R. Kola, *et al.*, *SISPAD*, pp. 1-4, Nov. 2020.
- [4] H. Carrillo-Nuñez, *et al.*, *IEEE Electron Device Letter*, vol. 40, no. 9, pp. 1366-1369, July 2019.
- [5] C. Akbar, *et al.*, *IEEE Transactions on Semiconductor Manufacturing*, vol. 34, no.4, pp. 513-520, Nov. 2021.

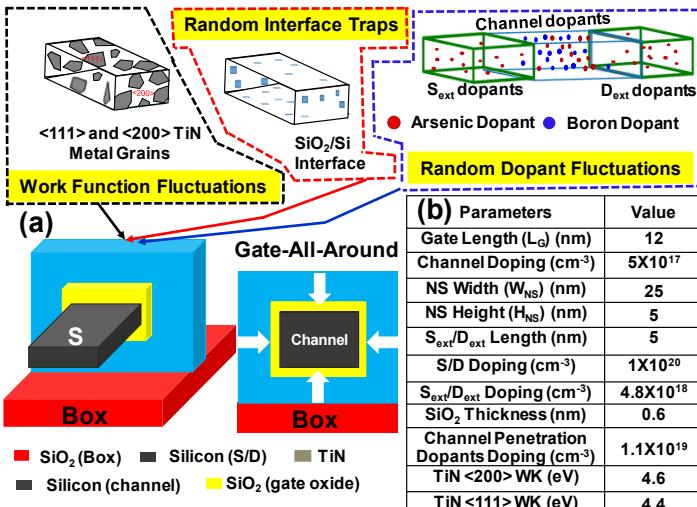


Fig. 1 (a) An illustration of the explored GAA Si NS MOSFET device with three uncertain random factors, i.e., the random dopant fluctuation (RDF), the work function fluctuation (WKF) and the random interface traps (ITF), respectively. We note that the total numbers of WKF, RDF and ITF, and device characteristics are generated for 1,000 fluctuated devices by using the validated 3D device simulation. (b) A list of the adopted device parameters in this study.

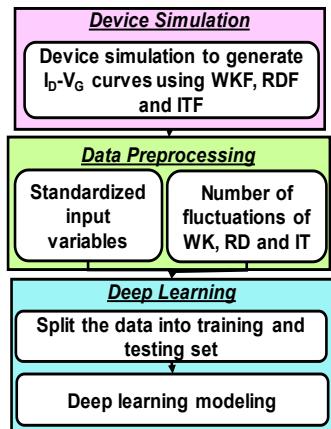


Fig. 2 A flow of the 3D device simulation combined with the DL methodology, where characteristics of GAA Si NS MOSFETs are simulated.

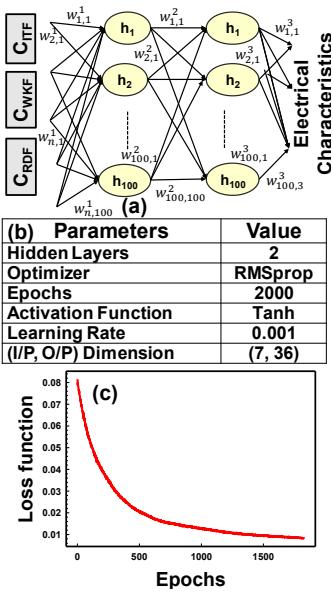


Fig. 3 (a) The adopted architecture of explored ANN model with input parameters (number of RDF, WKF and ITF) and output ( $I_D$  values), (b) the tuned hyperparameters and their corresponding values and (c) the decay of loss function with number of epochs.

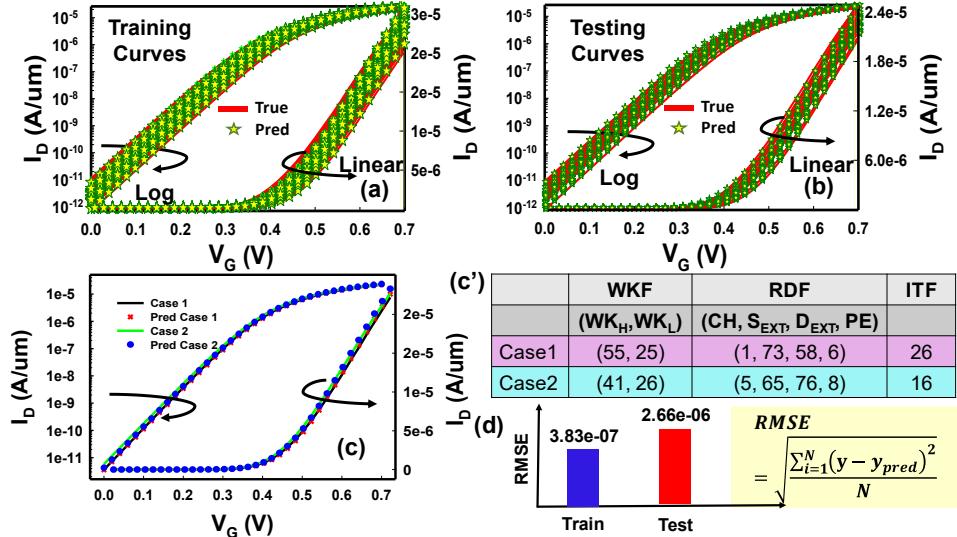


Fig. 4 The transfer characteristics are predicted by the ANN model using the device simulation of GAA Si NS MOSFET in linear and logarithmic scale. (a) While training the ANN model,  $I_D$ - $V_G$  curves are predicted by considering the simulated curves as target values. (b) After training the model, the transfer characteristics are predicted using the testing dataset. The bunch of  $I_D$ - $V_G$  curves for device simulation as well as the prediction results of the ANN model shows a good agreement. (c) To visualize the accuracy of the output from the ANN model, two fluctuated devices are selected from the testing dataset. The simulated and the predicted  $I_D$ - $V_G$  curves show the outperformance of the ANN model. The detailed device configuration for both cases are shown in (c'). (d) To validate the performance of ANN model, the RMSE value for the training and the testing datasets are calculated. The small train and the test RMSE values shows the well-trained performance of the ANN model.

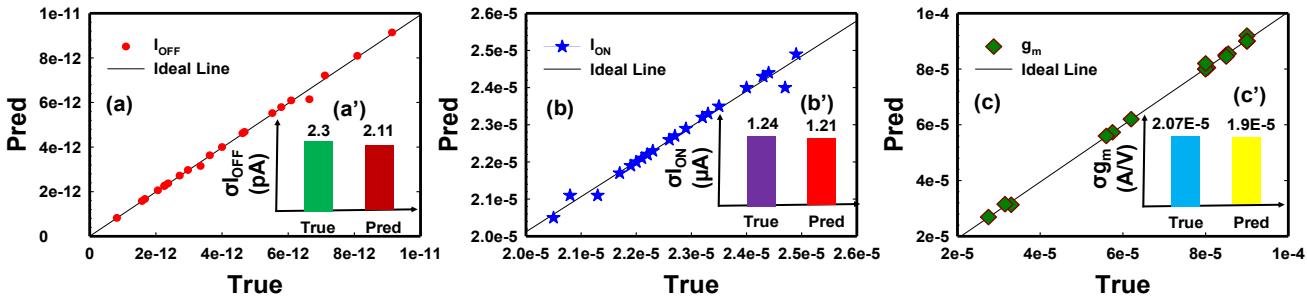


Fig. 5 The key figure of merit ( $I_{\text{ON}}$ ,  $I_{\text{OFF}}$  and  $g_m$ ) is extracted from the simulated and the predicted  $I_D$ - $V_G$  curves. (a) It shows the comparison between the true and the predicted  $I_{\text{OFF}}$  values using the testing transfer characteristics and (a') illustrates the fluctuation of  $I_{\text{OFF}}$  for true and predicted transfer characteristics. (b) and (c) present the comparison of true versus predicted  $I_{\text{ON}}$  and  $g_m$  with the ideal line, respectively. (b') and (c') show the fluctuation of  $I_{\text{ON}}$  and  $g_m$  for true and predicted  $I_D$ - $V_G$  curves (for testing dataset), respectively. The slight difference between them shows the accuracy of the ANN model.