

Received 21 October 2022; revised 9 May 2023; accepted 16 May 2023. Date of publication 18 May 2023; date of current version 25 May 2023.  
The review of this article was arranged by Editor C. Bulucea.

Digital Object Identifier 10.1109/JEDS.2023.3277548

# Machine Learning-Assisted Device Modeling With Process Variations for Advanced Technology

YAOYANG LYU<sup>1,2</sup> (Student Member, IEEE), WANGYONG CHEN<sup>1,2</sup> (Member, IEEE), MINGYUE ZHENG,<sup>1,2</sup>  
BINYU YIN<sup>1,2</sup>, JINNING LI<sup>1,2</sup>, AND LINLIN CAI<sup>1,2</sup> (Member, IEEE)

<sup>1</sup> School of Microelectronics Science and Technology, Sun Yat-sen University, Guangzhou 510275, China

<sup>2</sup> Guangdong Provincial Key Laboratory of Optoelectronic Information Processing Chips and Systems, Sun Yat-sen University, Guangzhou 510275, China

CORRESPONDING AUTHORS: W. CHEN and L. CAI (e-mail: chenwangy@mail.sysu.edu.cn; caillin3@mail.sysu.edu.cn)

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515110358 and Grant 2023A1515011418, and in part by the National Natural Science Foundation of China under Grant 62204269.

**ABSTRACT** Process variations (PV), including global variation (GV) and local variation (LV), have become one of the major issues in advanced technologies, which is crucial for circuit performance and yield. However, developing a mature and physics-based model is challenging and time-consuming. Thus, in this work, we propose a machine learning (ML) based method for device modeling with PV and implement the corresponding circuit simulation, which is demonstrated on advanced Nanosheet FETs (NSFET). Verified by TCAD simulations, the artificial neural network (ANN)-based ML algorithm enables to capture PV, e.g., dimension and work function variations (WFV), with high accuracy and improved efficiency. For GV, the ANN surrogated NSFET-based ring oscillator (RO) simulation results show that the larger width (Wsh) or height (Hsh) of the Nanosheet leads to the higher RO frequency and lower circuit delay. For LV, the respective impacts of grain size and WF on circuit performance can be distinguished. The proposed workflow, from ANN model training to circuit simulation based on the generated Verilog-A model, is fully automatic, promising to shorten the procedure of device modeling and accelerate the development of advanced technologies.

**INDEX TERMS** Nanosheet FETs, device modeling, artificial neural network, machine learning, process variations.

## I. INTRODUCTION

Over the past decades, in the development of the integrated circuits (IC) industry, the feature size of transistors has been aggressively scaled, especially for gate-all-around Nanosheet FETs (NSFET). Although the scaled transistors lead to significant progress in high-density integration and performance, the process variations (PV) is challenging for performance and yield, including global variation (GV) and local variation (LV) [1], [2]. For the case of GV, the dimension variation is one of the prominent variability sources in determining the performance of devices. For the case of LV, work function variation (WFV) derived from the randomly distributed grain sizes and orientations plays a crucial role in statistical variability sources [3], [4], [5]. Moreover,

the interplay between GV and LV should be considered to capture the device performance accurately. Given that multiple variability sources are involved, obtaining a large amount of sample device data from simulations or experiments means a high computational or time cost. Hence, the device model accommodated the powerful capability of predicting PV more quickly and accurately is urgently needed [6], [7].

Generally, industrial compact models of devices consider the impacts of process variation by varying the model parameters, e.g., PHIG (gate work function), U0 (low field mobility), to reduce the model complexity. However, it is still quite challenging for equation-based models to fully automate the model parameter extraction process while achieving

a very high fitting accuracy [8]. Moreover, various emerging devices may exhibit a difference in electrical characteristics that the conventional compact models cannot sufficiently capture. It requires high expertise and a long period to develop the physics-based model equations for the new physical phenomena [9]. Therefore, with the more pronounced variability source effects in the scaled devices, shortening nanoscale device modeling with PV and achieving automatic and precise parameter extraction is highly desired.

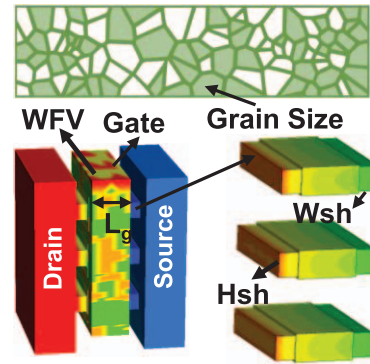
Recently, model creation based purely on data from experiments or simulations has attracted lots of interest, mainly due to their precision and ease of development [10], [11]. Data-driven models can be divided into parametric and nonparametric models. Parametric models, i.e., linear regression, have prior forms with fixed parameters obtained by physical knowledge. Nonparametric models, i.e., artificial neural network (ANN), do not have a specific format. The ANN model captures the inherent physical pattern from the training data, which are particularly suitable for emerging devices, and can be applied to different technology nodes [12], [13]. Moreover, the high-performance graphic processing unit (GPU) servers and efficient development frameworks for ANN training, i.e., TensorFlow [14] and PyTorch [15], provide conveniences for the implementation of ANN device modeling. For the reasons above, it is worthwhile to utilize the ANN-based methodology for emerging device modeling and apply it to circuit simulation [16], [17].

In this work, we propose a machine learning (ML) assisted device modeling method, where the ANN model is designed to capture variability sources and bias conditions effects efficiently. Verified by the calibrated TCAD simulation, the trained and verified model shows that superior accuracy can be achieved on the key figure of merits (FoMs) and their correlations. Furthermore, the model is employed to the circuit simulation with PV and demonstrates its capability on the NSFET-based ring oscillator (RO). The rest of this article is divided into the following sections.

Section II introduces the proposed workflow, including the design of experiments (DoE) for the device with PV impacts, the ANN model development, and the TCAD simulation for the current-voltage ( $I$ - $V$ ) and capacitance-voltage ( $C$ - $V$ ) characteristics of NSFET used to train and test the model. Section III presents the fitting results of  $I$ - $V$  and  $C$ - $V$ , the extraction results of FoMs and their variations with GV, and the simulation results of RO with devices affected by GV and LV. Our key conclusions are summarized in Section IV.

## II. DEVICE STRUCTURE AND METHODOLOGY

The schematic view of NSFET used in this work is illustrated in Fig. 1, indicating the random distributions of work function (WF) and metal grains for gate materials. Table 1 lists the parameter values of the process variability sources in this work. The variation of critical dimensions, including width (Wsh) and height (Hsh) of the Nanosheet, are used to implement GV. And the effects of WF acting as the major LV



**FIGURE 1.** The schematic of NSFET with global and local variations. The grain size (GS) and work function (WF) are the key parameters.

**TABLE 1.** Summary of process variability source parameters.

Local Variations (WF Variations)		
	Standard Deviation ( $3\sigma$ )	Average ( $\mu$ )
GS [nm] [18]	1.5	6.5
WF [eV] [19]	0.15	4.5
Global Variations (Dimension Variations)		
Parameter	DoE Values	
Hsh [nm] [2]	4-8	
Wsh [nm] [2]	15-25	

are considered. The relevant LV parameters are all assumed to be Gaussian distributions over the DoE space. It should be noted that WFV depends on the mean value of WF and grain size (GS). The variation range of GV parameters is selected based on the data in [2], while the values and variation range of GS and WF are chosen according to the data in [18] and [19], respectively.

Fig. 2 shows the flow chart for developing the ANN model from TCAD simulations and its implementation for circuit simulation. The proposed flow consists of initial data preparation over the DoE space, model development, model verification, and model application in the circuit. The first step is to prepare data at the DoE space, in which the device GV and LV sources are defined, DoE points based on GV values are designed, and data at the DoE points are generated from TCAD simulation. The second step is developing the model. To predict the  $I$ - $V$  and  $C$ - $V$  characteristics of the device, we design the model structure and use data in the first step to train the model. The third step is verifying the trained model, where the FoMs are extracted and the model is further optimized based on the error between the results of the model and the TCAD. Finally, apply the model in the circuit simulation, in which a Verilog-A model is generated for SPICE simulation, and the output waveforms can be further analyzed.

The structure of the ANN model is shown in Fig. 3. Based on the feed-forward neural network (FFNN), the model is composed of one input layer with seven neurons, one output

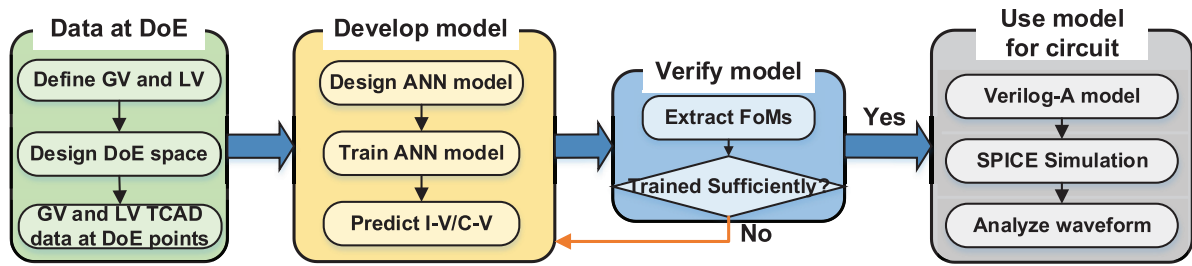


FIGURE 2. Flow chart for generating model using ML method and its application in circuit simulation.

layer with two neurons, and six hidden layers with 150, 150, 50, 30, 15, 30 neurons. The input features of the ANN model are bias conditions, dimensions, and work function parameters induced by GV and LV. By using these three types of input data, the network is trained to obtain the variability effects of NSFET on the ANN-based ML algorithm under different conditions. For hidden layers, the number of layers and neurons in each layer can be adjusted. Each hidden layer receives data from the previous layer and propagates a calculation result to the next layer. With the conversion function, multiple variables in the output layer of ANN can be converted to obtain  $I_{ds}$  and  $C_{gs}$ .

For a lower error rate, an activation function of rectified linear unit (RELU) is used to prevent the vanishing gradient problem that may appear due to the complexity of the algorithm. Besides, the adaptive moment estimation (ADAM) optimization function is employed for accurate error correction. The ADAM optimization function is an optimization function which is suitable for processing stochastic data, and it is applied to iteratively correct the learning process of the algorithm through the set error rate [15], [20]. The normalized root mean square error (NRMSE) is the indicator to quantify the accuracy of ANN [21]. The ANN-based ML algorithm is trained using the public release PyTorch 1.10.2 python library. The training process is repeated 46000 times or more. The total time for training the ANN model and saving the model results is 1 hour and 30 minutes. If the time of saving data is excluded, generating the model takes only 21 minutes.

Fig. 4 shows the  $I_{ds}-V_{gs}$  and  $C_{gs}-V_{gs}$  plots for 3024 NSFET device samples caused by PV and bias voltages at the device level, which are obtained through calibrated TCAD. All the bias voltages and PV parameters shown in Table 1 vary concurrently. The LV parameter values are randomly selected from the Gaussian distributions, while the GV parameters are assigned with the defined value. The metal gate in the device is titanium nitride (TiN). The physical models incorporated in the TCAD simulation includes the Shockley-Read-Hall (SRH) model to account for carriers generation and recombination, and the density gradient model to take the quantum confinement in nanoscale devices into account. Besides, the mobility models, including the high-field saturation model, the doping-dependent mobility model, and vertical field dependent mobility model, are

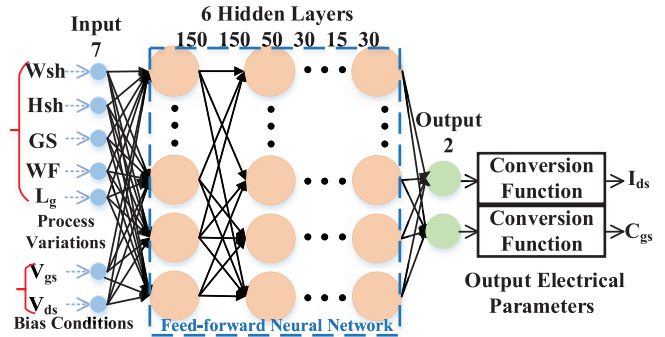


FIGURE 3. Artificial neural network (ANN) structure for the device modeling with variability, where bias conditions are also considered.

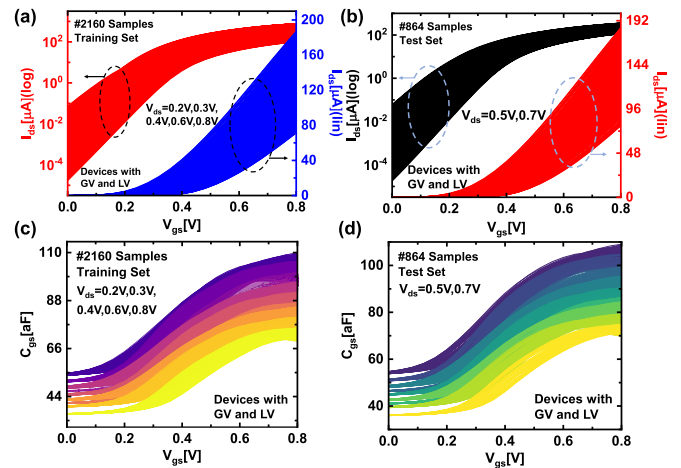


FIGURE 4.  $I_{ds}-V_{gs}$  and  $C_{gs}-V_{gs}$  plots of NSFET with different variability sources. (a)(c) for the training set and (b)(d) for the test set.

also applied in the TCAD simulations. The TCAD built-in functions are utilized to generate  $I_{ds}-V_{gs}$  and  $C_{gs}-V_{gs}$  plots. Notably, in Fig. 4(c)-(d), the value of  $C_{gs}$  is discrete at low  $V_{gs}$  because  $C_{gs}$  values increase with the enlarger of the nanosheet effective width, and the differences are prominent when  $C_{gs}$  values are small. After TCAD simulations, 70% of the TCAD data is used to train the ANN model, and 30% of the data is used to test it. Once the model development process finishes, the ANN model is automatically converted to a Verilog-A [22] model through a python script for circuit simulation at the SPICE level. To use a trained ANN

model in Verilog-A, several steps should be taken. First of all, the weights and biases of each layer in the ANN are extracted from the trained and saved model. These extracted values are then assigned to the corresponding variables in the Verilog-A model. Then, the input and output variables of the Verilog-A model are made consistent with the process variation sources and electrical characteristics being studied in this work. Next, the output value of each hidden layer is calculated according to equation (1), where  $A^i$  is the output of the current layer,  $A^{i-1}$  is the output of the previous layer, and  $W$  and  $B$  are the weight and bias connecting the current and previous layer. Before the output of the current layer is passed to its next layer, it is processed by the RELU activation function. Finally, the output variable values of the ANN-based Verilog-A model are converted to the corresponding electrical characteristics through equations (2) and (3).

$$A^i = W \times A^{i-1} + B \quad (1)$$

$$I_{ds} = V_{ds} \times 10^{y_1} \quad (2)$$

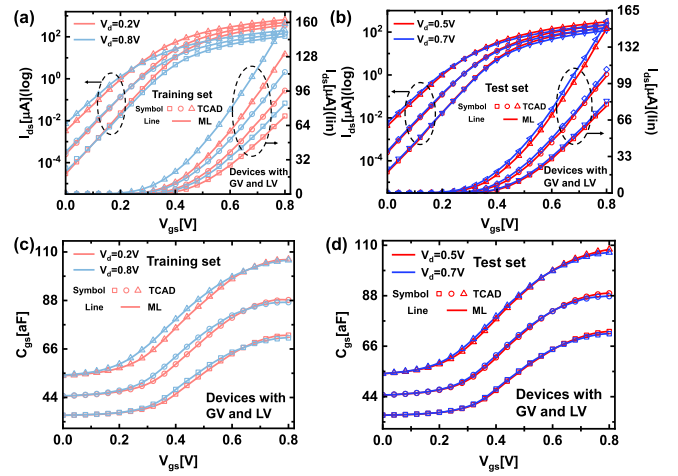
$$C_{gs} = C_0 \times y_2 \quad (3)$$

It is worth mentioning that the selection of the conversion function is critical for achieving high  $I-V$  and  $C-V$  model accuracy [8]. In this work, the conversion function (2) and (3) is applied for the  $I_{ds}$  and  $C_{gs}$ , where  $C_0$  is a normalization coefficient (i.e., 10aF), and  $y_1$  and  $y_2$  are the output variables from the relative neurons in the ANN output layer. The conversion function (2) guarantees the zero  $I_{ds}$  when  $V_{ds} = 0$  V to avoid violating the physical conservation law [8], [9], [10], and reduces the range of  $y_1$  even when  $I_{ds}$  vary by many orders of magnitude during TCAD simulations of NSFET. The variation range of  $C_{gs}$  is in the same order of magnitude, so  $y_2$  in (3) is not needed to be a logarithmic scale.

### III. RESULTS AND DISCUSSIONS

#### A. MODEL ACCURACY VERIFICATION AND STATISTICAL ANALYSIS OF FOMS WITH THE ANN MODEL

Fig. 5 illustrates the ANN model prediction results for NSFET under PV and bias voltage effects. The simulated and the predicted  $I_{ds}-V_{gs}$  plots for the training dataset are shown in Fig. 5(a)-(b), and the  $C_{gs}-V_{gs}$  curves are shown in Fig. 5(c)-(d). It can be observed that both  $I_{ds}-V_{gs}$  and  $C_{gs}-V_{gs}$  curves are in good agreement between simulation and prediction. In addition, the performance of a well-trained ANN model is also evaluated using a test dataset whose data differs from the train set. The ANN testing results of  $I_{ds}-V_{gs}$  and  $C_{gs}-V_{gs}$  characteristics fit the TCAD data with high accuracy, as shown in Fig. 5. The mean error of the model fitting of  $I_{ds}-V_{gs}$  /  $C_{gs}-V_{gs}$  characteristics is less than 0.5%. The FoMs scatter plots of model prediction results against TCAD simulations and the value of correlation coefficients between key FoMs over the target DoE space are shown in Fig. 6, respectively, where both GV and LV dominate the device performance. To extract the

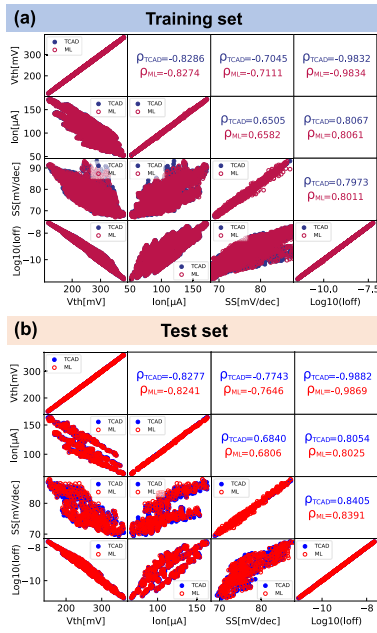


**FIGURE 5.** ANN model results (lines) versus targets (symbols) for NSFET under different bias conditions. (a)  $I-V$  and (c)  $C-V$  of the training set, (b)  $I-V$  and (d)  $C-V$  of the test set.

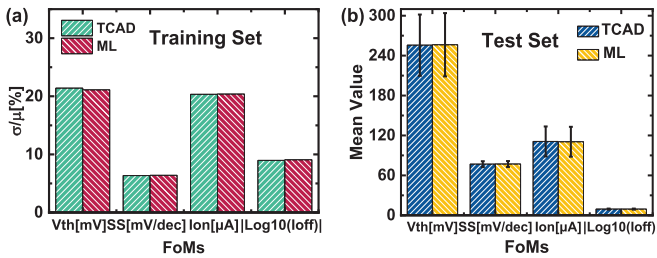
FoMs from the  $I_{ds}-V_{gs}$  curves, a python script based on the MLFoMPy library is utilized [23]. It should be noted that the design of DoE space not only covers the interested range of dimension variation but also suffer from LV effects at each node in the space. Both distributions and correlations of the training and test sets, including the effects of GV and LV, achieve a good match between the TCAD data and the extracted ANN data. The excellent agreement between the full  $I_{ds}-V_{gs}$  curves predicted by the ANN model and those generated by TCAD simulations can guarantee that all relevant FoMs are extracted with a high degree of accuracy. In addition, Fig. 7(a) demonstrates the comparison results of the standard deviation coefficient ( $\sigma/\mu$ ) for the training set. Meanwhile, the  $\mu$  and  $\sigma$  values of distributions for the test set are shown in Fig. 7(b). They all indicate that the train and test error rate is close to zero for all FoM distributions and their correlations. The mean error of FoM extraction results from the ANN model is less than 3%.

Fig. 8(a)-(b) shows the distribution variations of threshold voltage ( $V_{th}$ ) and on-state current ( $I_{on}$ ) with GV in the NSFET devices. The LV with different GS values are selected from the Gaussian distribution shown in Table 1. The work function of all the devices in Fig. 8 is 4.52eV. For  $V_{th}$ , as the value of  $W_{sh}$  or  $H_{sh}$  becomes larger, the mean and median value of  $V_{th}$  distributions decrease, and the variation range of  $V_{vth}$  is smaller. For  $I_{on}$ , as  $W_{sh}$  or  $H_{sh}$  increases, the mean and median value of  $I_{on}$  distributions increase, and the variation range of  $I_{on}$  is narrower. For subthreshold swing (SS) and off-state current ( $I_{off}$ ), the variation of the  $\sigma/\mu$  values for their distributions are shown in Fig. 8(c)-(d), indicating that the variation range of SS and  $I_{off}$  narrows down with the increment of  $W_{sh}$  or  $H_{sh}$ . It is clear that the key electrical characteristics improve and suffer less variation due to the increased effective channel width and cross-sectional area of NSFET.





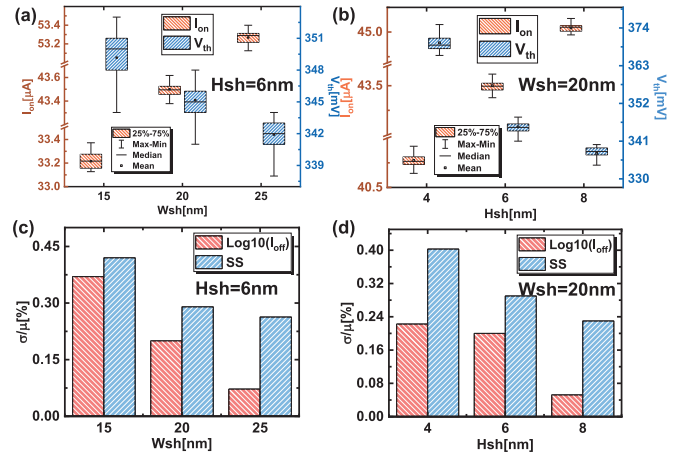
**FIGURE 6.** Comparison of scatter plots and correlations of the key electrical parameters between the TCAD data and the results from the ML approach. (a) training set (b) test set.



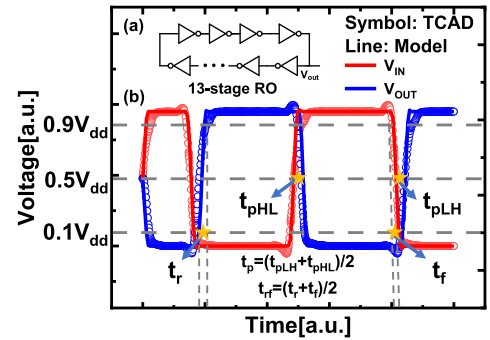
**FIGURE 7.** Results of TCAD simulation and ML-based prediction for the mean value of key electrical parameters and standard deviation coefficient for (a) training set and (b) test set.

## B. ANN MODEL APPLICATION IN RO SIMULATION

The comparison of RO simulation results obtained using the Verilog-A-based ANN model and TCAD mixed-mode simulations are demonstrated in Fig. 9, where the definitions of rising ( $t_r$ ), propagation ( $t_p$ ), and fall delay ( $t_f$ ) are marked. It shows that a good agreement is achieved between these two methods, and it suggests that the Verilog-A-based ANN device model can be utilized to analyze the impacts of GV and LV at the circuit level. The output waveforms are produced during the circuit simulation when using the ANN-based Verilog-A model with various GV and LV parameter values. Fig. 10(a)-(b) present some waveforms under GV effects, showing that waveforms shift right with the decrement of Wsh or Hsh. Meanwhile, Fig. 10(c)-(d) illustrate waveforms under LV effects, suggesting that waveforms shift right as GS decreases or WF increases. Overall, it means that the RO has a more extended period of oscillation with reduced NSFET size, decreased GS, and increased WF. For further analysis, the RO frequency and circuit delay are extracted from the simulated waveforms mentioned above.



**FIGURE 8.** Variation of the FoMs distributions from ANN model results with GV. (a)(b) on-state current ( $I_{on}$ ) and threshold voltage ( $V_{th}$ ) distributions. (c)(d) off-state current ( $I_{off}$ ) in the logarithmic scale  $\text{Log}_{10}(I_{off})$  and subthreshold swing (SS) distributions.



**FIGURE 9.** Schematic view of (a) 13-stage RO and (b) Comparison of its output voltage waveforms obtained from Verilog-A-based ANN model and TCAD mixed-mode simulations, where the definitions of delay parameters are also marked.

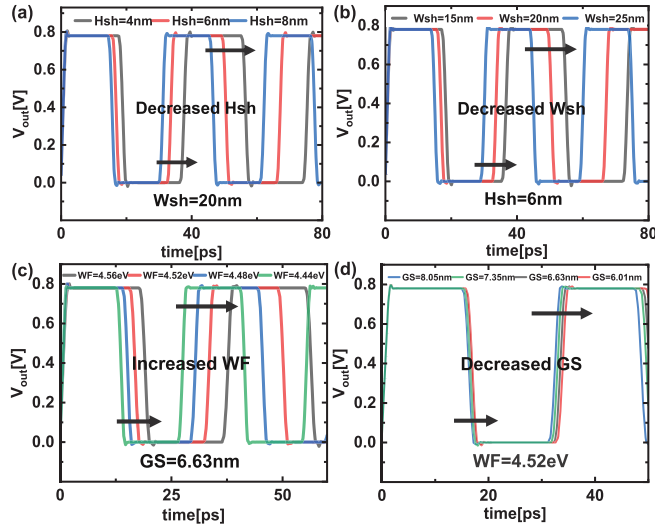
Fig. 11 presents the extracted distributions of RO frequency with PV. Fig. 11(a)-(b) are the results under GV effects, revealing that RO frequency is enhanced with increased Wsh or Hsh. Similarly, RO frequency distributions affected by LV are demonstrated in Fig. 11(c)-(d). It is evident that the distributions of RO frequency shift right when WF decreases or GS increases. Since the RO frequency  $f_{RO}$  is calculated with the following function:

$$f_{RO} = \frac{1}{2 \times n \times t_p} \quad (4)$$

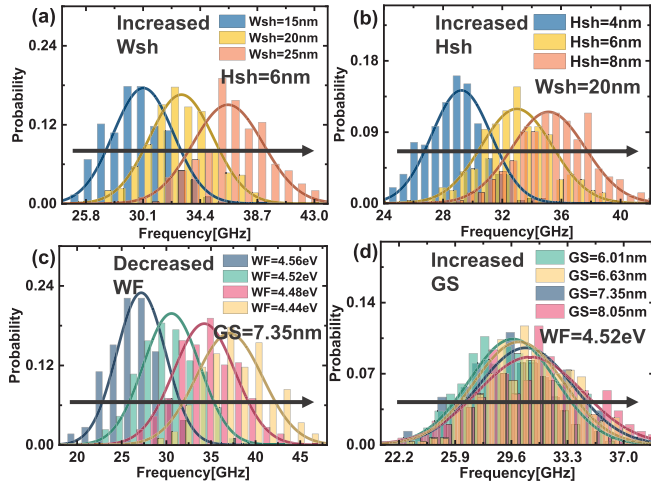
where  $n$  represents the number of inverters in the RO circuit. the shift of RO frequency is mainly attributed to the variation of  $t_p$  with GV and LV. The increased Wsh or Hsh, the decreased WF, and the increased GS decrease the value of  $t_p$ , resulting in higher RO frequency. The detailed GV and LV effects on  $t_p$  are discussed in Section III-C.

## C. GV AND LV EFFECTS ON CIRCUIT DELAY

Fig. 12 and Fig. 13 illustrate the change of circuit delay when devices are under GV and LV. The average value of

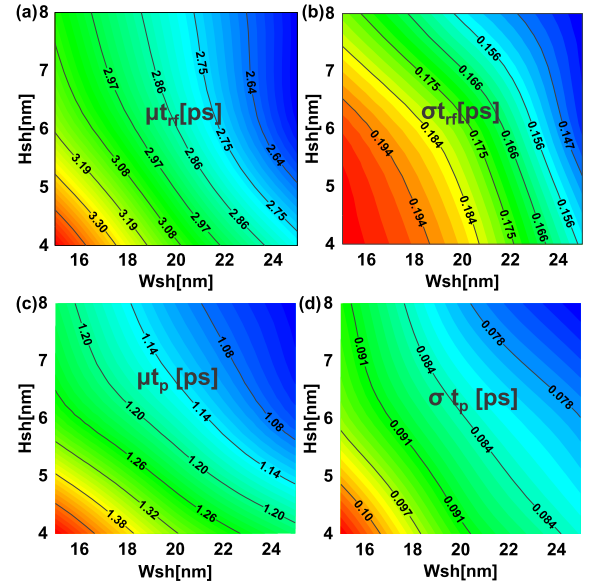


**FIGURE 10.** Output waveforms of 13-stage RO using ANN-based Verilog-A model with (a)(b) global variation (GV) and (c)(d) local variation (LV).

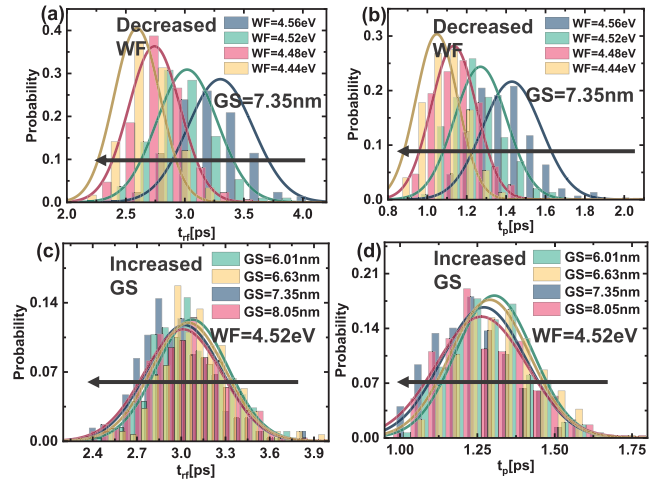


**FIGURE 11.** Statistical distributions of frequency of the 13-stage RO using ANN-based Verilog-A model with (a)(b) global variation (GV) and (c)(d) local variation (LV).

$t_r$  and  $t_f$  is defined as  $t_{rf}$ . For GV, the  $\mu$  and  $\sigma$  variation of  $t_{rf}$  and  $t_p$  distributions generated from devices with various Wsh and Hsh values are shown in Fig. 12. It can be seen that as Wsh or Hsh increases, the  $\mu$  and  $\sigma$  of the delay distributions become lower. It is because that increasing Wsh or Hsh leads to a higher drive current and less drive current variation of NSFET, according to Fig. 8, which brings a drop and minimizes variation in time delay. Fig. 13 presents the distributions of  $t_{rf}$  and  $t_p$  with the impacts of WF and GS. It is apparent that from Fig. 13(a)-(b) and Fig. 13(c)-(d) that delay distributions shift left as WF decreases or GS increases, respectively. It can be attributed to the reduction of the  $V_{th}$  of NSFET. When WF decreases or GS increases, the  $V_{th}$  of NSFET becomes smaller [24], [25], [26], which results in lower delay. Besides, the decreased WF leads to less  $V_{th}$  variation of NSFET [25], thus slighter variation in



**FIGURE 12.** The average value ( $\mu$ ) and standard deviation ( $\sigma$ ) of statistical delay distributions of the 13-stage RO using ANN-based Verilog-A model with global variation (GV). (a)(b) The average of rising and fall delay ( $t_{rf}$ ). (c)(d) Propagation delay ( $t_p$ ).



**FIGURE 13.** Statistical delay distributions of 13-stage RO using ANN-based Verilog-A model with local variation (LV). (a)(b) The mean value of work function (WF) changes. (c)(d) The average value of grain size (GS) changes.

circuit delay. On the contrary, increasing GS values leads to a broader  $V_{th}$  variation range [24], [26], thus severer variation in circuit delay.

#### D. COMPARISON WITH OTHER DEVICE MODELS

There are two alternative data-driven methods for the advanced technology and devices evaluation, including the lookup table (LUT)-based model [27], [28], [29] and the ANN-based model. The comparison of the key properties between the ANN-based model and the LUT-based model is as follows.

In general, to achieve the same level of modeling precision as the ANN model with the same input and output

**TABLE 2. Comparison with ANN models in reported works.**

Comparison with ANN models in reported works		
	This Work	Reported Works
Number of PVs	4, including GV and LV	2-4, only GV [6], [8], [12]
Analysis Level of PVEffects	device and circuit	device, with GV and LV [1], [7]
ANN Size	$N_{hid}$ : 6 $N_{neu}$ : 475	$N_{hid}$ : 1-3 $N_{neu}$ : 5-300 [7], [12]
ANN Training Time	1 hour and 30 minutes	>8 hours [1], [8]
$I_{ds}$ - $V_{gs}$ / $C_{gs}$ - $V_{gs}$ Fitting Error	<0.5%	<0.5% [8]
FoMs Extraction Error	<3%	<3% [1]

parameters, a LUT-based model may require a higher number of bias sampling points, which is due to that the interpolation relationships modeling capability of the LUT model may be weaker than that of the ANN model [30], [31], [32]. Because the advanced devices exhibit complex nonlinear relationships in their electrical characteristics, and the ANN model can automatically learn these complex characteristics from the given set of training data and maintain high interpolation accuracy [30].

Additionally, the variability evaluation capability of the ANN-based model may be better than the LUT-based model, because the advanced devices, such as NSFET, are susceptible to the coupling effect of multiple process variation sources, resulting in a nonlinear and complex interpolation relationship. The ANN-based model requires less human effort in model parameter extracting and fitting, as it can automatically learn from the given set of data [8], [33]. Furthermore, to model the coupling effect of multiple process variation sources and achieve the same level accuracy as the ANN model, the LUT-based model requires data with a fine grid-like structure, while the ANN-based model can be trained on data with a coarser grid-like structure or even scattered data [32]. Thus, modeling using ANN may require fewer data points.

Despite the advantages, there also exists challenges in ANN device modelling compared to conventional modeling techniques. The physical implications of ANN models are generally weaker compared to those of physics-based equation models, which can lead to the possibility of unphysical behaviors [9]. And the prediction accuracy of ANN models may become worse beyond the range of the training data [13], indicating that the DoE space is of great significance. In addition, the training and testing of ANN models may require lots of computational resources, imposing higher requirements for the computational hardware.

Table 2 shows the comparison between key properties of the ANN-based model in this work and some reported works [1], [6], [7], [8], [9], [10], [12], [32], in which the number of hidden layers in the ANN is defined as  $N_{hid}$ , the number of neurons is defined as  $N_{neu}$ . It can be noted that this work is at an advantage over the reported works in some aspects. First of all, although some reported works have studied ANN-based device models, to capture the device performance, the coupled effects of GV and LV are still lacking. In this work, 4 GV and LV sources are considered

coherently from the DoE space that covers the interested range of dimension variation, while other works only consider 2 to 4 GV sources [6], [8], [12]. Besides, in this work, the influence of GV and LV on device FoMs distributions and digital circuit performance parameters are presented, respectively. For the reported works considering GV and LV [1], [7], GV and LV effects are only studied at the device level. Notably, the  $I_{ds}$  and  $C_{gs}$  values of devices are directly treated as the ANN model output in this work, making it possible for the model to be converted to a Verilog-A-based FET device model in the SPICE. And the FoMs values can be automatically extracted from ANN or SPICE simulation results, reducing the time cost to verify the model at the device level and apply it at the circuit level.

Due to the increased number of PV kinds and PV sources, the ANN model in this work unavoidably needs more data samples to train. Thus, the ANN structure in this work is more complex, including six hidden layers and 475 neurons in total. Nevertheless, the ANN training and results-saving time is only 1 hour and 30 minutes, which is less than reported works which take more than 8 hours for ANN training. Actually, generating the ANN model used in SPICE takes only 21 minutes, with the data-saving process excluded. Moreover, the accuracy of the ANN model can be guaranteed with increased ANN complexity. The fitting error of  $I_{ds}$ - $V_{gs}$  and  $C_{gs}$ - $V_{gs}$  characteristics are within 0.5%, and the FoMs extraction error is within 3%.

#### IV. CONCLUSION

This work proposes a ML-assisted device modeling method to consider the multiple process variability source effects. Including GV and LV, the ANN-based nanoscale device modeling for circuit simulation with high efficiency and accuracy is implemented. The model provides excellent prediction capability in capturing variability sources of the device, verified by the conventional TCAD simulation. Besides, the Verilog-A model generated from trained ANN is further applied to analyze the effects of GV and LV, e.g., Wsh, Hsh, GS, and WF, on circuit performance via SPICE simulation, where the variation of NSFET-based RO frequency and delay distributions are investigated. Compared to other device models, the ANN-based model has a better capability of variability modeling which considers the coupled effects of GV and LV, reduces modeling and evaluation time for the advanced technology at the device and circuit level, and achieves high accuracy that is excellent among the ANN-based models in reported works. The proposed automatic flow of device modeling can be extended to other kinds of variability sources and emerging device technologies.

#### REFERENCES

- [1] K. Ko, J. K. Lee, M. Kang, J. Jeon, and H. Shin, "Prediction of process variation effect for ultrascaled GAA vertical FET devices using a machine learning approach," *IEEE Trans. Electron Devices*, vol. 66, no. 10, pp. 4474-4477, Oct. 2019, doi: [10.1109/TED.2019.2937786](https://doi.org/10.1109/TED.2019.2937786).

- [2] X. Wang et al., "FinFET centric variability-aware compact model extraction and generation technology supporting DTCO," *IEEE Trans. Electron Devices*, vol. 62, no. 10, pp. 3139–3146, Oct. 2015, doi: [10.1109/TED.2015.2463073](https://doi.org/10.1109/TED.2015.2463073).
- [3] X. Wang, A. Brown, B. Cheng, and A. Asenov, "Statistical variability and reliability in nanoscale FinFETs," in *Proc. IEDM*, Dec. 2011, pp. 5.4.1–5.4.4, doi: [10.1109/IEDM.2011.6131494](https://doi.org/10.1109/IEDM.2011.6131494).
- [4] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain orientation induced work function variation in nanoscale metal-gate transistors—Part I: Modeling, analysis, and experimental validation," *IEEE Trans. Electron Devices*, vol. 57, no. 10, pp. 2504–2514, Oct. 2010, doi: [10.1109/TED.2010.2063191](https://doi.org/10.1109/TED.2010.2063191).
- [5] K. Ko, M. Kang, J. Jeon, and H. Shin, "Compact model strategy of metal-gate work-function variation for ultrascaled FinFET and vertical GAA FETs," *IEEE Trans. Electron Devices*, vol. 66, no. 3, pp. 1613–1617, Mar. 2019, doi: [10.1109/TED.2019.2891677](https://doi.org/10.1109/TED.2019.2891677).
- [6] J. K. Lee, K. Ko, and H. Shin, "Analysis on process variation effect of 3D NAND flash memory cell through machine learning model," in *Proc. EDTM*, Penang, Malaysia, 2020, p. 4, doi: [10.1109/EDTM47692.2020.9117940](https://doi.org/10.1109/EDTM47692.2020.9117940).
- [7] K. Ko, J. K. Lee, and H. Shin, "Variability-aware machine learning strategy for 3-D NAND flash memories," *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1575–1580, Apr. 2020, doi: [10.1109/TED.2020.2971784](https://doi.org/10.1109/TED.2020.2971784).
- [8] J. Wang, Y. H. Kim, J. Ryu, C. Jeong, W. Choi, and D. Kim, "Artificial neural network-based compact modeling methodology for advanced transistors," *IEEE Trans. Electron Devices*, vol. 68, no. 3, pp. 1318–1325, Mar. 2021, doi: [10.1109/TED.2020.3048918](https://doi.org/10.1109/TED.2020.3048918).
- [9] Y. Lei, X. Huo, and B. Yan, "Deep neural network for device modeling," in *Proc. EDTM*, Kobe, Japan, Mar. 2018, pp. 154–156, doi: [10.1109/EDTM.2018.8421454](https://doi.org/10.1109/EDTM.2018.8421454).
- [10] L. Zhang and M. Chan, "Artificial neural network design for compact modeling of generic transistors," *J. Comput. Electron.*, vol. 16, no. 3, pp. 825–832, Apr. 2017, doi: [10.1007/s10825-017-0984-9](https://doi.org/10.1007/s10825-017-0984-9).
- [11] Y. Cao, X. Chen, and G. Wang, "Dynamic behavioral modeling of nonlinear microwave devices using real-time recurrent neural network," *IEEE Trans. Electron Devices*, vol. 56, no. 5, pp. 1020–1026, May 2009, doi: [10.1109/TED.2009.2016029](https://doi.org/10.1109/TED.2009.2016029).
- [12] Q. Chen and G. Chen, "Artificial neural network compact model for TFTs," in *Proc. CAD-TFT*, Beijing, China, Oct. 2016, p. 1, doi: [10.1109/CAD-TFT.2016.7785057](https://doi.org/10.1109/CAD-TFT.2016.7785057).
- [13] F. Klemme, J. Prinz, V. M. van Santen, J. Henkel, and H. Amrouch, "Modeling emerging technologies using machine learning: Challenges and opportunities," in *Proc. ICCAD*, San Diego, CA, USA, 2020, pp. 1–9, doi: [10.1145/3400302.3415770](https://doi.org/10.1145/3400302.3415770).
- [14] M. Abadi et al., "TensorFlow: A system for large scale machine learning," in *Proc. OSDI*, vol. 16, 2016, pp. 265–283, doi: [10.5555/3026877.3026899](https://doi.org/10.5555/3026877.3026899).
- [15] V. Subramanian, *Deep Learning With Pytorch*, 1st ed. New York, NY, USA: iG, 2018.
- [16] Y. Lyu, W. Chen, M. Zheng, B. Yin, J. Li, and L. Cai, "Machine learning assisted nanoscale device modeling for nanosheet FETs with process variation," in *Proc. SNW*, Honolulu, HI, USA, Jun. 2022, p. 2, doi: [10.1109/SNW56633.2022.9889011](https://doi.org/10.1109/SNW56633.2022.9889011).
- [17] L. Cai, W. Chen, G. Du, X. Zhang, and X. Liu, "Layout design correlated with self-heating effect in stacked nanosheet transistors," *IEEE Trans. Electron Devices*, vol. 65, no. 6, pp. 2647–2653, Jun. 2018, doi: [10.1109/ted.2018.2825498](https://doi.org/10.1109/ted.2018.2825498).
- [18] P. H. Vardhan, S. Ganguly, and U. Ganguly, "An accurate expression to estimate the metal gate granularity induced threshold voltage variability in NWFETs," *Solid-State Electron.*, vol. 152, pp. 65–71, Mar. 2019, doi: [10.1016/j.sse.2018.12.003](https://doi.org/10.1016/j.sse.2018.12.003).
- [19] H. Nam and C. Shin, "Study of high-*k*/metal-gate work-function variation using rayleigh distribution," *IEEE Electron Device Lett.*, vol. 34, no. 4, pp. 532–534, Apr. 2013, doi: [10.1109/LED.2013.2247376](https://doi.org/10.1109/LED.2013.2247376).
- [20] G. W. Burr et al., "Experimental demonstration and tolerancing of a large-scale neural network (165 000 synapses)," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3498–3507, Nov. 2015, doi: [10.1109/TED.2015.2439635](https://doi.org/10.1109/TED.2015.2439635).
- [21] J. Xiong, A. Yang, M. Raginsky, and E. Rosenbaum, "Neural networks for transient modeling of circuits," in *Proc. MLCAD*, Raleigh, NC, USA, Sep. 2021, p. 7, doi: [10.1109/MLCAD52597.2021.9531153](https://doi.org/10.1109/MLCAD52597.2021.9531153).
- [22] *The Designer's Guide to Verilog-AMS*. Boston, MA, USA: Kluwer, 2004.
- [23] "MLFoMpy documentation, version 0.0.0." 2021. [Online]. Available: <https://gitlab.citius.usc.es/modev/mlfompy>
- [24] P. H. Vardhan, S. Mitta, S. Ganguly, and U. Ganguly, "Analytical estimation of threshold voltage variability by metal gate granularity in FinFet," *IEEE Trans. Electron Devices*, vol. 64, no. 8, pp. 3071–3076, Aug. 2017, doi: [10.1109/TED.2017.2712763](https://doi.org/10.1109/TED.2017.2712763).
- [25] N. A. Kumari and P. Prithvi, "Performance evaluation of GAA nanosheet FET with varied geometrical and process parameters," *Silicon*, vol. 14, pp. 9821–9831, Feb. 2022, doi: [10.1007/s12633-022-01695-7](https://doi.org/10.1007/s12633-022-01695-7).
- [26] E. Mohapatra, T. P. Dash, J. Jena, S. Das, and C. K. Maiti, "Design study of gate-all-around vertically stacked nanosheet FETs for sub-7nm nodes," *SN Appl. Sci.*, vol. 3, no. 5, p. 540, May 2021, doi: [10.1007/s42452-021-04539-y](https://doi.org/10.1007/s42452-021-04539-y).
- [27] J. Wang, N. Xu, W. Choi, K.-H. Lee, and Y. Park, "A generic approach for capturing process variations in lookup-table-based FET models," in *Proc. SISPAD*, Washington, DC, USA, Sep. 2015, pp. 309–312, doi: [10.1109/SISPAD.2015.7292321](https://doi.org/10.1109/SISPAD.2015.7292321).
- [28] R. A. Thakker, C. Sathe, A. B. Sachid, M. S. Baghini, V. R. Rao, and M. B. Patil, "A novel table-based approach for design of FinFET circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 7, pp. 1061–1070, Jul. 2009, doi: [10.1109/TCAD.2009.2017431](https://doi.org/10.1109/TCAD.2009.2017431).
- [29] V. Bourenkov, K. G. McCarthy, and A. Mathewson, "MOS table models for circuit simulation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 3, pp. 352–362, Mar. 2005, doi: [10.1109/TCAD.2004.842818](https://doi.org/10.1109/TCAD.2004.842818).
- [30] Q. Yang et al., "Transistor compact model based on multigradient neural network and its application in SPICE circuit simulations for gate-all-around Si cold source FETs," *IEEE Trans. Electron Devices*, vol. 68, no. 9, pp. 4181–4188, Sep. 2021, doi: [10.1109/TED.2021.3093376](https://doi.org/10.1109/TED.2021.3093376).
- [31] H. Bai, C. Liu, E. Breaz, and F. Gao, "Artificial neural network aided real-time simulation of electric traction system," *Energy AI*, vol. 1, Aug. 2020, Art. no. 100010, doi: <https://doi.org/10.1016/j.egyai.2020.100010>.
- [32] D. E. Root, "Future device modeling trends," *IEEE Microw. Mag.*, vol. 13, no. 7, pp. 45–59, Nov./Dec. 2012, doi: [10.1109/MMM.2012.2216095](https://doi.org/10.1109/MMM.2012.2216095).
- [33] Y. Zhang, G. He, G. Wang, and Y. Li, "XBarNet: Computationally-efficient memristor crossbar model using convolutional autoencoder," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 41, no. 12, pp. 5489–5500, Dec. 2022, doi: [10.1109/TCAD.2022.3163895](https://doi.org/10.1109/TCAD.2022.3163895).