IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 June 2022, accepted 28 June 2022, date of publication 5 July 2022, date of current version 11 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3188690

RESEARCH ARTICLE

A Machine Learning Approach to Modeling Intrinsic Parameter Fluctuation of Gate-All-Around Si Nanosheet MOSFETs

RAJAT BUTOLA[1,2], YIMING LI[1,2,3,4,5,6], (Member, IEEE), AND SEKHAR REDDY KOLA[1,2], (Graduate Student Member, IEEE)

[1]Parallel and Scientific Computing Laboratory, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan
[2]Electrical Engineering and Computer Science International Graduate Program, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan
[3]Institute of Communications Engineering, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan
[4]Institute of Biomedical Engineering, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan
[5]Department of Electrical Engineering and Computer Engineering, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan
[6]Center for mmWave Smart Radar System and Technologies, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan

Corresponding author: Yiming Li (ymli@nycu.edu.tw)

This work was supported in part by the Ministry of Science and Technology (MOST), Taiwan, under Grant MOST 110-2221-E-A49-139, Grant MOST 110-2218-E-492-003-MBK, and Grant MOST 109-2221-E-009-033; and in part by the "Center for mmWave Smart Radar Systems and Technologies" under Grant MOST 109-2634-F-009-030 and Grant MOST 110-2634-F-009-028.

ABSTRACT The sensitivity of semiconductor devices to any microscopic perturbation is increasing with the continuous shrinking of device technology. Even the small fluctuations have become more acute for highly scaled nano-devices. Therefore, these fluctuations need to be addressed extensively in order to continue further device scaling. In this paper, we mainly focus on three sources of intrinsic parameter fluctuation including the work function fluctuation (WKF), the random dopant fluctuation (RDF), and the interface trap fluctuation (ITF) for gate-all-around (GAA) silicon (Si) nanosheet (NS) MOSFETs. Generally, the effect of these fluctuations is analyzed using a time-consuming device simulation process. A machine learning (ML) based powerful and efficient artificial neural network (ANN) model is used to accelerate this process. Firstly, the effects of fluctuation sources are analyzed individually by using the ANN model and results have been presented that show the WKF variations dominate the variation of threshold voltage, off-state current, and on-state current among other fluctuation sources. Next, we examine the combined effect of three fluctuation sources. It is crucial because considering only one fluctuation can result in unexpected variations due to other fluctuations appearing in the device. Consequently, the ANN model is used to estimate the combined effects as well. The results show that the proposed model predicts the outputs with an R^2-score of 99% and an error rate of less than 1%. In addition, the ML is also utilized to calculate the permutation importance of input variables as a measure to investigate the contribution of each fluctuation source.

INDEX TERMS Artificial neural network, machine learning, GAA Si NS MOSFETs, intrinsic parameter fluctuation, WKF, RDF, ITF, characteristic fluctuation, threshold voltage, off-state current, on-state current.

I. INTRODUCTION

The threat of semiconductor device variability grows as the transistor size shrinks. The aggressive scaling of transistors is subject to different variations [1]. However, continuous reduction in technology nodes is the demand for advanced integrated circuits (ICs). The miniaturized nanoscale devices provide a remarkable improvement in power consumption [2] and allow more transistors to be fabricated on a given area of IC [3]. In order to function properly, these highly scaled ICs are required to fabricate all transistors such that their characteristics should be identical to each other. However, the semiconductor industry is facing a major challenge owing to the fact that each transistor in an IC has different electrical

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

characteristics due to a variety of causes. Among them, the random fluctuations are the major causes of variation in device characteristics, such as threshold voltage (V_{TH}), off-state current (I_{OFF}), on-state current (I_{ON}), drain-induced barrier lowering (DIBL), and subthreshold slope (SS), etc. [4]. These variations in advanced nano-devices are becoming the most severe challenge for further device scaling [5]. No matter how, it is imperative to conduct an extensive analysis of these fluctuation sources to keep producing better, cheaper, and faster nano-devices.

The gate-all-around (GAA) is a widely accepted device structure for technology nodes beyond 3 nm. There are two types of GAA structures: nanowire (NW) and nanosheet (NS). Due to a high surface-to-volume ratio, NS provides more drive current as compared to NW. Therefore, the GAA silicon (Si) NS metal-oxide-semiconductor field-effect transistor (MOSFET) has gained enormous attention as a promising candidate for a high-degree-scaling technology node [6], [7]. It exhibits excellent short-channel control and improves the performance of transistors at reduced gate length. It also provides more conducting channels, improved power performance, and better area scaling compared to other transistors, such as FinFETs and nanowire FETs [8]–[10]. However, nanoscale devices are suffered from different kinds of fluctuations that cause variability in their electrical characteristics [11]–[15] and the miniaturization increases the device sensitivity to microscopic perturbation. Thus, even a small fluctuation may become a critical issue for device characteristics. These fluctuation sources are the key obstructions for future potential applications of GAA Si NS MOSFETs. Accordingly, their analysis is of utmost interest and needs further investigation.

The fluctuation is referred to as a type of variability. Among several fluctuation sources, the intrinsic parameter fluctuations are less controllable due to their microscopic nature and are the most serious challenges to overcome. We consider in this work three major sources of intrinsic parameter fluctuation, the work function fluctuation (WKF), the random dopant fluctuation (RDF), and the interface trap fluctuation (ITF). In general, fluctuation analyses are mainly accomplished by using device simulation techniques [16]. Effects of fluctuation on GAA Si NS MOSFETs have been studied [17], [18]; however, despite the high accuracy of device simulation, it is a very time-consuming process and desired alternative time-efficient methods [19].

A. PRELIMINARIES AND RELATED WORKS

Various machine learning (ML) approaches have recently been proposed for device simulation to shorten the turn-around time. ML has emerged as an efficient method that is having applications in different fields such as healthcare [20], agriculture [21], cybersecurity [22], education [23], data governance [24], finance [25], marketing [26], etc. In the past few years, it has also been used significantly in the semiconductor industry to solve modeling and optimization problems [27]–[33]. It is successfully employed as a strategy to speed up semiconductor device development and reduce the computational resources [34]–[39]. Device simulation assisted ML frameworks have been reported to solve different problems such as device variation and operating temperature analysis [40], point defect prediction [41], hotspot detection [42], anomaly detection [43], device structural variation identification, and inverse design [44], etc. These works have focused to speed up and automate the device modeling task with high accuracy. Some of recent researches applied ML-based frameworks to predict next-generation device characteristics [45]–[48]. They employed the ML models as an alternative to the compact model for advanced devices to predict their electrical characteristics. Notably, these ML frameworks are entirely based on device data and do not include device physics. In addition to this, the ML techniques have been further used to predict and model the characteristics variations that occurred in different semiconductor devices due to WKF [49], random dopant distribution [50], line-edge-roughness [51], [52], process variation effect [53], etc. In some studies, ML was also applied for the estimation of threshold voltage variability induced by random telegraph noise fluctuation [54] and by device parameter variability [55], etc. In our recent work, the effect of WKF was modeled using ML for NS devices such as, in [56], the ML technique was proposed to suppress the WKF on device characteristics of a three-channel NS device by using a random forest regressor-based ML model. In [57], ML was utilized to identify the WKF patterns on the metal gate to reduce its impact on DC characteristics of GAA Si NS MOSFETs. However, all previous ML works have considered only one specific type of fluctuation and to the best of our knowledge, this is the first time when ML is used to model more than one fluctuation sources simultaneously.

The goal of the current work is to avoid the intensively performing 3-dimensional (3D) device simulation of GAA Si NS MOSFETs by determining the effect of the source of variation through the ML-based ANN model. In addition to this, we also estimate the variability in important parameters such as V_{TH}, I_{OFF}, and I_{ON} by considering the effects of random numbers of different fluctuations. In this way, we also accelerate the overall 3D statistical device simulation by using the ML model. Furthermore, the fair prediction of the effect of all types of fluctuation sources and their combined effects demonstrate the generality of the proposed strategy to construct accurate ML models for advanced nano-devices.

B. NOVEL CONTRIBUTIONS IN THIS WORK

1. This is the first time the ML approach is proposed to capture multi-fluctuation variability in nano-devices. Their individual as well as combined effects are analyzed comprehensively.
2. We, for the first time, investigate the denser ANN-based ML algorithm to analyze the complex physical behavior of multi fluctuation sources. The ANN model proved to be a general and more robust model in contrast to the if/else

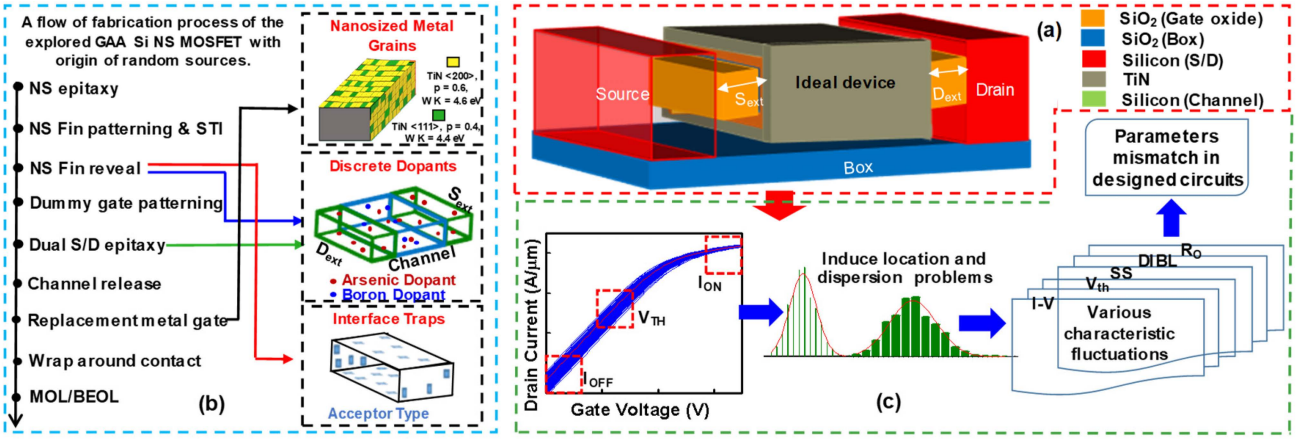

FIGURE 1. A flow of fabrication process of the explored GAA Si NS MOSFET with origin of random sources. Random nano-sized metal grains, random discrete dopants, and random interface traps are raised during the complicated process. (a) An illustration of the explored GAA Si NS MOSFET. (b) A flow of fabrication process with introduced random factors of the explored GAA Si NS MOSFET. (c) The fluctuated I_D-V_G curves and the associated device parameters: V_{TH}, I_{OFF}, and I_{ON}, as well as their impact on device variability and circuit mismatch.

statistical function based model (such as in the random forest regressor model).

3. We derive the benefit from ML to investigate the contribution of each fluctuation source on electrical characteristics when their combined effect is analyzed. The individual permutation scores for all input features are also calculated that revealed the contribution of each fluctuation source on the device characteristics.

The rest of the paper is organized as follows. Section II presents the structure of GAA Si NS MOSFET with device simulation settings, physical and electrical parameters, and an overview of fluctuation source. Section III discusses the machine learning modeling approach and integration with device simulation. In Section IV, the predicted results for different fluctuation sources are compared with the simulation data and discussed in detail. Finally, Section V concludes the main work of this study.

II. DEVICE SIMULATION, FLUCTUATION SOURCES, AND DATA GENERATION

The 3D statistical device simulations are centered around the GAA Si NS MOSFET, as shown in Fig. 1(a). A flow of device fabrication for the explored GAA Si NS MOSFET with the origin of various fluctuation sources including random nano-sized metal grains, random discrete dopants, and random interface traps is shown in Fig. 1(b) [7]. Starting with the epitaxial growth, the fin with a height of 5 nm and a width of 25 nm is fabricated. To meet the requirement of advanced technological nodes, the sidewall image transfer technique is used. Dummy gates are defined before the anisotropic etching and the source/drain regions are grown epitaxially. After the removal of dummy gates, the channel release is performed and Si wires are released completely during the replacement metal gate. The contacts are wrapped around to maintain good electrostatic integrity. Fig. 1(c) shows the fluctuated drain current-gate voltage (I_D-V_G) curves and device parameters, and their impacts on circuit mismatch.

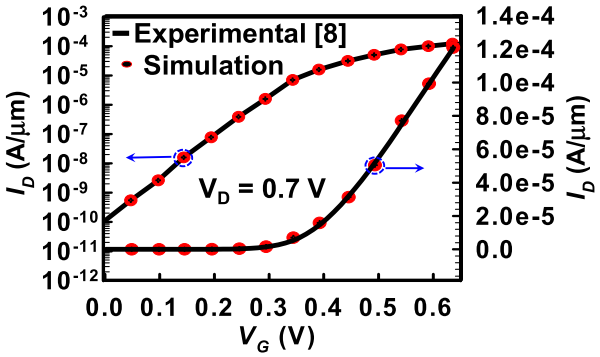

FIGURE 2. A comparison between the numerically simulated (line) and experimentally measured (symbol) [8] I_D-V_G curves that are used to calibrate the nominal device simulation of GAA Si NS MOSFET.

The ideal device without any fluctuation is referred to as a "nominal device". The adopted device and material parameters, dimension, doping concentration, and the achieved nominal values of electrical characteristics and short-channel effect parameters are summarized in Table 1. As shown in Fig. 2, before performing the 3D statistical device simulation on the explored devices, a nominal device simulation is first performed and calibrated with the experimentally measured I_D-V_G curves [8] of a stacked GAA Si NS MOSFET by tuning the device parameters, physical models, and doping profile to reproduce the best accuracy of the transfer characteristics. We then apply the similar calibration methodology to statistically simulate a single-channel GAA Si NS MOSFET for various fluctuations. Notably, the intrinsic parameter fluctuation of a single-channel GAA Si NS MOSFET is studied because it is more critical than that of devices with multi stacked NS channels.

The GAA Si NS MOSFET with sources of intrinsic parameter fluctuation is shown in Fig. 3. Total number of simulated devices with different fluctuations is listed in Table 2. Statistically, three fluctuation sources are generated by following the Gaussian distribution for the explored GAA Si NS MOSFET.

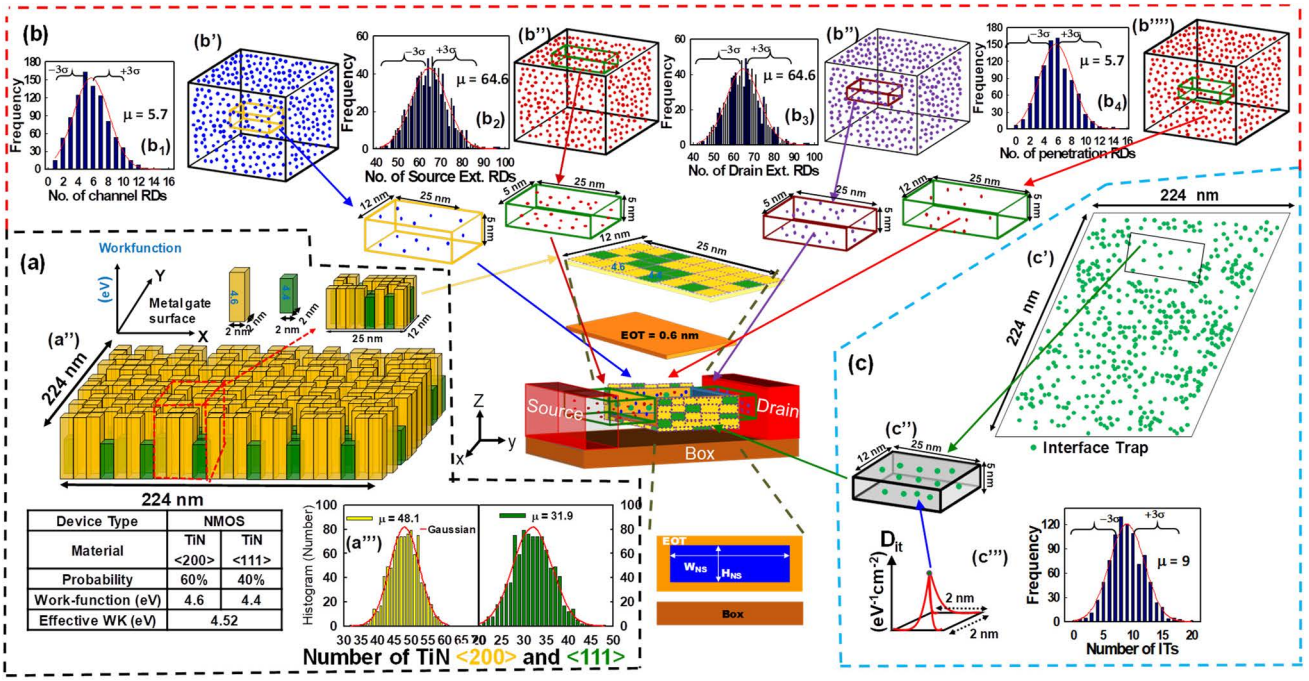

FIGURE 3. An illustration of the 3D architecture of nominal GAA Si NS MOSFET including various fluctuation sources i.e. (a) WKF, (b) RDF, and (c) ITF. (a') random pattern of metal grains of TiN on gate area. (a'') High and low work functions with orientation <111> for WK_L and <200> for WK_H. (a''') Histograms of sub-planes following the Gaussian distribution. (b')-(b'''') Random dopants in channel, source/drain extensions, and penetration from the S/D extensions into the channel regions the average dopants in each square is 5.7, 64.6, 64.6, and 5.7, respectively. (b_1)-(b_4) Histograms showing the Gaussian distribution of random dopants. (c') Interface traps randomly generate in a large plane of 224 nm^2. (c'') Sub-plane of small dimensions. (c''') The size of each trap is 2 nm^2 with density of 1.5-7.6$\times 10^{13}$ cm^{-2} eV^{-1} and histogram shows that the interface traps follow Gaussian distribution.

TABLE 1. List of the adopted device parameters and achieved device characteristics of the GAA Si NS MOSFETs corresponding to sub-3-nm technological nodes.

Parameters	Value
Gate Length (L_G) (nm)	12
Channel Doping (cm^{-3})	5×10^{17}
Nanosheet Width (W_{NS}) (nm)	25
Nanosheet Height (H_{NS}) (nm)	5
Sext/Dext Length (nm)	5
S/D Doping (cm^{-3})	1×10^{20}
Sext/Dext Doping (cm^{-3})	4.8×10^{18}
Effective Oxide Thickness (*EOT*) (nm)	0.6
Density of Interface Traps ($cm^{-2}eV^{-1}$)	$1.5\text{-}7.6 \times 10^{13}$
S/D$_{ext}$ Dopants Doping (cm^{-3})	1.45×10^{19}
Channel Penetration Dopants Doping (cm^{-3})	1.1×10^{19}
Channel Dopants Doping (cm^{-3})	1.63×10^{19}
TiN <200> Work Function (eV)	4.6
TiN <111> Work Function (eV)	4.4
V_{TH} (mV)	260
I_{OFF} (A/μm)	4.48×10^{-12}
I_{ON} (A/μm)	2.29×10^{-5}
SS (*mV/dec*)	66.2
DIBL (*mV/V*)	31.8

TABLE 2. List of different sources of fluctuation and their corresponding total number of fluctuated devices.

Type of fluctuation	No. of devices
Work Function Fluctuation (WKF)	1000
Random Dopant Fluctuation (RDF)	1000
Interface Trap Fluctuation (ITF)	1000
Combined Fluctuation (among WKF, RDF, and ITF)	1000

A. GENERATION OF WKF SAMPLES

The titanium nitride (TiN) layer is used as the metal gate for the device and it is composed of metal grains (MG) of sizes 3.92 nm × 3 nm. For all WKF fluctuated devices, the random patterns of low/high work function (WK) are generated by using the Monte Carlo method [14]. The low WK is referred to as TiN<111> and high WK is referred to as TiN<200> with the probability of occurrence as 40% and 60% with WK = 4.4 eV and WK = 4.6 eV, respectively as shown in Fig. 3(a). The effective work function is 4.52 eV. Total number of MG is fixed (MGN = 80) for all devices. More details about the generation of WKF are given in Appendix A.

B. GENERATION OF RDF SAMPLES

To simulate RDF fluctuated devices, 1725, 19396, and 1718 random dopants are generated in a large cuboid with the doping concentration of 5×10^{17}, 1.1×10^{19}, and 3.36×10^{17} cm^{-3} for the channel (CH), source-drain extension (S/D_{EXT}), and channel penetration (C_{PE}), respectively, as shown in Fig. 3(b). The large cuboid is then partitioned

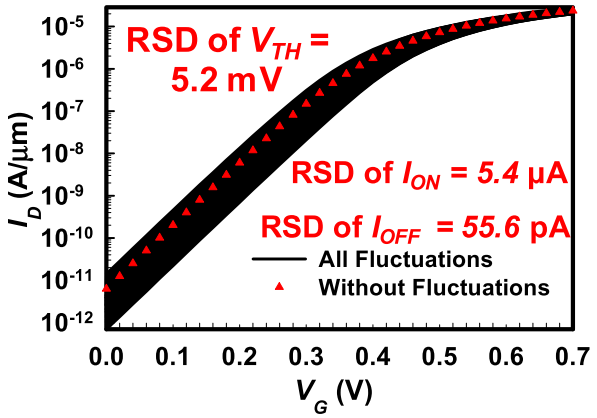

FIGURE 4. An illustration of 1000 I_D-V_G curves (line) of the GAA Si NS MOSFET induced by intrinsic parameter fluctuation, i.e., work function fluctuation, random dopant fluctuation and interface trap fluctuation. Symbols are of the nominal characteristic of I_D-V_G (i.e., the GAA Si NS MOSFET device without any fluctuation.

into 1000 sub-cuboids which are mapped into CH, S/D_{EXT}, and C_{PE} regions for each fluctuated device [5], [16]. The average number of dopants in CH, S/D_{EXT}, and C_{PE} regions are 5.7, 64.6, and 5.7, respectively. See Appendix B for more discussion about the RDF.

C. GENERATION OF ITF SAMPLES

Similarly, for ITF fluctuated devices, 2700 random interface traps (RITs) are generated in a large square plane of size 224 nm × 224 nm, as shown in Fig. 3(c). The RITs are generated arbitrarily at the interface between the gate stack and Si channel using the statistical generation simulator [5]. The dimension of each RIT is 2 nm^2 with a density of around 1.5-7.6 × 10^{13} cm^{-2} eV^{-1}. From this plane, 1000 sub-planes are pulled out randomly. The average number of RITs in-plane is 9 and plane density is kept as 3 × 10^{13} cm^{-2}. The ITF is further discussed in detail in Appendix C.

D. GENERATION OF COMBINED FLUCTUATION SAMPLES

Finally, to generate the devices with combined fluctuations (CF), the distribution of all sources (WKF, RDF, and ITF) remains the same as mentioned above. Moreover, 1000 such fluctuated devices have been randomly generated and simulated for the CF case as well [5].

Eventually, through comprehensive simulations, the I_D-V_G characteristics of all above 4000 fluctuated devices for four different cases are collected which are later used for ML analysis as the training and the test datasets. The simulated electrical parameters: V_{TH}, I_{OFF}, and I_{ON} are also determined which are later compared with the ML predicted values. Fig. 4 shows the nominal device curve (red triangular) and 1000 simulation curves for devices with the combined fluctuations (black lines) with relative standard deviation (RSD) (explained in detail in Section IV) of V_{TH}, I_{OFF}, and I_{ON}.

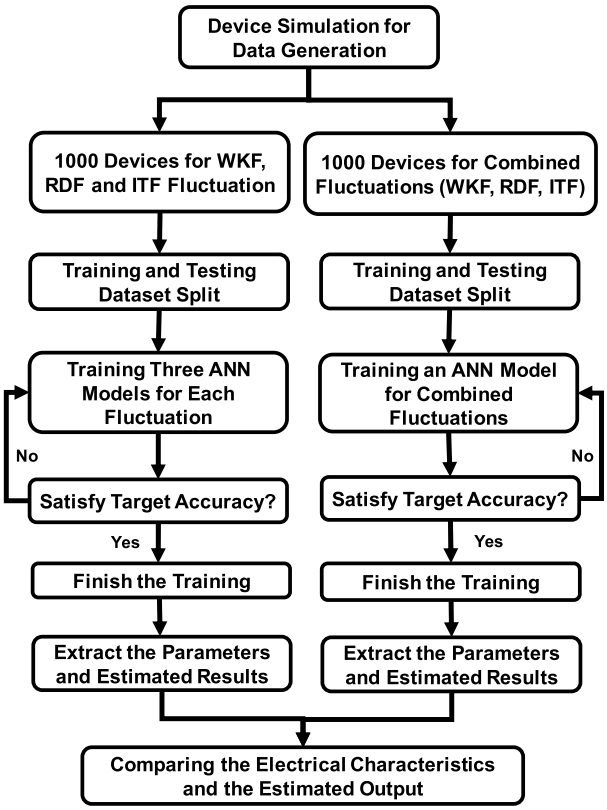

FIGURE 5. A comprehensive flow chart showing the proposed methodology with data generation, data collection, splitting training and testing processes, extraction of the electrical parameters and then comparison of the electrical characteristics.

III. MACHINE LEARNING APPROACH

The overall purpose of the NN-based framework is to develop the ML technique that can accelerate the process of modeling the characteristic variations imposed by various fluctuations. The NN-based framework used in this work is outlined in Fig. 5, which consists of device simulation steps and machine learning process flow. The simulations are performed for four fluctuation sources to obtain the training and test datasets as described in Section II. Figs. 6(a) and (b) show the ANN architectures for the different sources of variations. The ANN models are implemented using Python's Keras library with Tensorflow working in its backend [58]. All architectures have three layers referred to as input layer, hidden layer, and output layer. The input layer consists of fluctuation sources governing parameters as inputs such as for WKF count of high and low WK [WK_H, WK_L], for RDF, count of random dopants in the channel, source-drain extensions and channel penetration [CH, S/D_{EXT1}, S/D_{EXT2}, PE], for ITF, count of interface traps at the top, bottom, side-1 and side-2 [IT_T, IT_B, IT_{S1}, IT_{S2}], and at last, for CF, the input pattern consist of all fluctuation counts [WK_H, WK_L, CH, S/D_{EXT1}, S/D_{EXT2}, PE, I_T, I_B, I_{S1}, I_{S2}]. Variation in V_{TH}, I_{OFF}, and I_{ON} is predicted by considering the distribution of fluctuation sources using the output layer. The ANN models for all cases have

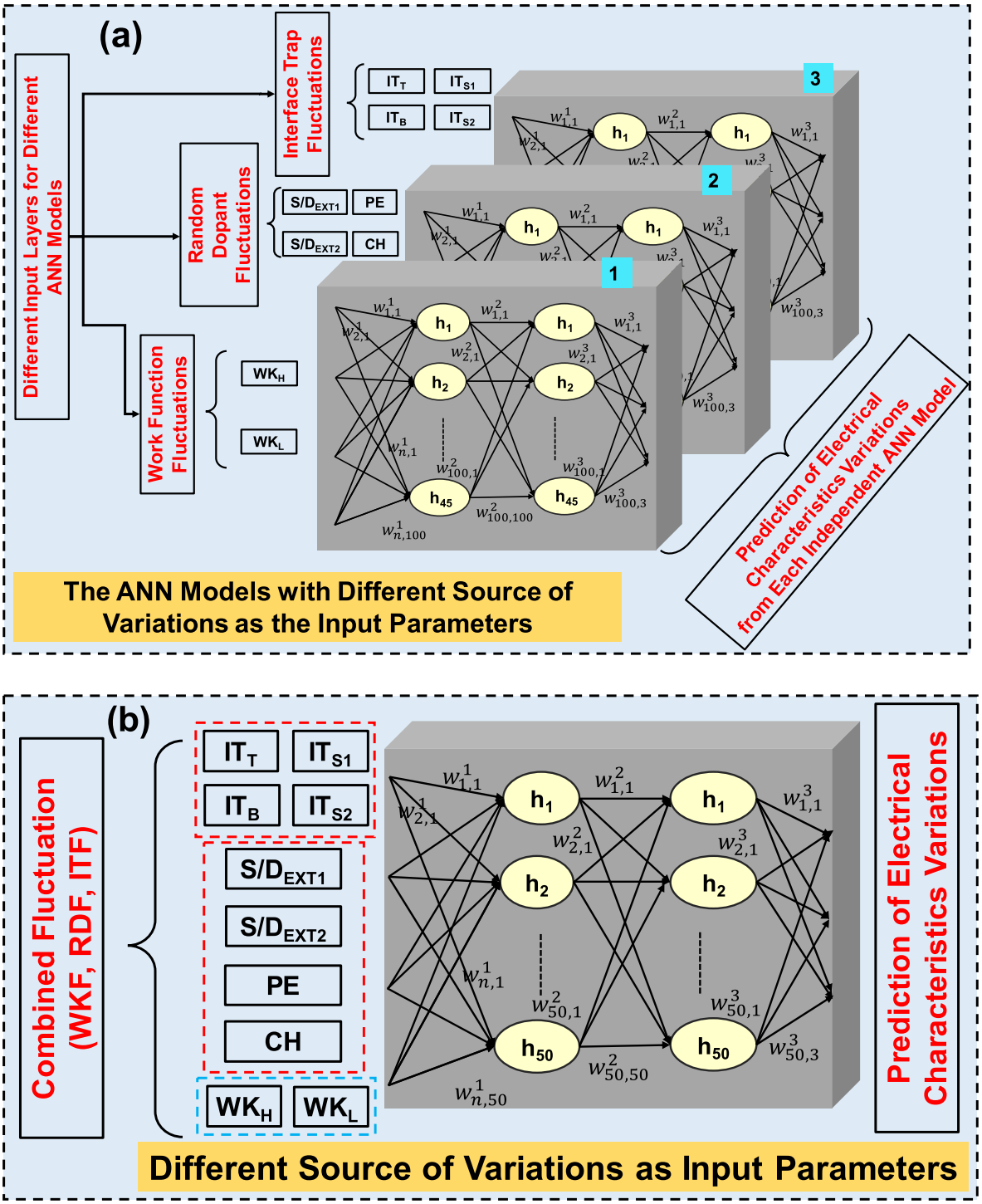

FIGURE 6. Modeling of ML algorithm using data obtained from the GAA Si NS MOSFET. (a) Three ANN models trained and evaluated separately using three different intrinsic parameter fluctuation, i.e., WKF, RDF and ITF and (b) the ANN model trained and evaluated using combined fluctuation, i.e., all variations: WKF, RDF, and ITF, simultaneously.

two hidden layers. The list of various hyperparameters such as activation function, type of optimizer, number of neurons in each hidden layer, learning rate, and batch size are also summarized in Table 3 and explained in detail as follows.

The input (I/P) fed into the ML model has distinct dimensions depending on the fluctuation source, such as for WKF I/P dimension is 2, i.e., WK_H, and WK_L and for RDF the I/P dimension is 4 i.e. CH, S/D_{EXT1}, S/D_{EXT2}, PE, etc. However, each model has an output (O/P) dimension in

TABLE 3. List of model parameters add hyperparameters for various ANN models with inputs: WKF, RDF, ITF, and CF, respectively.

Hyperparameters	ANN_{WKF}	ANN_{RDF}	ANN_{ITF}	ANN_{CF}
(I/P, O/P) Dimension	(2, 36)	(4, 36)	(4, 36)	(10, 36)
Number of Hidden Layers	2	2	2	2
Number of Neurons in each layer	45	45	45	50
Activation Function	Tanh	Tanh	Tanh	Tanh
Learning Rate	0.001	0.001	0.001	0.001
Optimizer	RMSprop	RMSprop	RMSprop	RMSprop
Batch Size	5	5	5	5
Epochs (Fixed)	5000	5000	5000	5000
Early stopping Epoch	1257	1140	1381	1284

a monochromatic fashion, i.e., 36 (due to the number of I_D values being 36). The activation function governs the threshold at which a neuron is activated. The hyperbolic tangent activation function is investigated for nonlinear mapping, due to its property to deal with -1 to $+1$ range of the dataset. Furthermore, the optimization algorithm widely affects the accuracy of the ML models. it is utilized to tune the weights and biases through an appropriate learning rate to reduce the overall loss and also improve the training speed of the model. For this work, RMSprop is used to minimize the loss while tuning the entire hyperparameters conjointly. The other crucial hyperparameters are epoch and batch size. Epoch represents the number of times the ML model is trained on the training dataset and batch size is the number of samples utilized in one iteration. We use the batch size as 5 and the number of epochs is set to 5000. Thus, 5 samples are used at a time before the model's loss function is updated and 5000 epochs are utilized to complete the training process. However, to deal with the overfitting and the underfitting issues of the ML models, the Keras early stopping function is used. It is a regularization technique that enables the training to be stopped automatically when the accuracy has stopped improving. Under this function the input argument, ''patience value'', is set to 20. It is the number of epochs without improvement after which the model training is stopped. It not only avoids overfitting and underfitting but also reduces the training time. Therefore, even though the epoch values are fixed to 5000 epochs but the training stopped in a smaller number of epochs than this value for different ANN models is listed in Table 3. More discussion of underfitting and overfitting can be found in Appendices D and E. Hence, all hyperparameters are set manually which provides the minimum error and accurate prediction.

The biasing condition examined for generating the device electrical characteristics are as follows: drain voltage is set at $V_D = 0.7$ V, the gate voltage V_G is swept from 0 to 0.7 V with a step size (Δ) of 0.02 V i.e. V_G is discretized into 36 points. The V_{TH} is defined at V_G where drain current $I_D = 3.2 \times 10^{-8}$A/μm, I_{OFF} is I_D at $V_G = 0$ V and I_{ON} is I_D at $V_G = V_D = 0.7$ V. The dataset of 1000 samples is divided into training, testing, and validation set, where 80% of data is used for training the ANN model 10% is used for validation, and 10% data is considered for testing the accuracy of the model. The inputs are first transformed into a suitable form by Min-Max Scaler from scikit-learn [59] to fit them within a specific range before feeding into the ANN model. A hypothesis, also called a fitness function, is developed for each case according to the relationship between their inputs and corresponding outputs. During the training phase, the ANN model optimizes this hypothesis by constantly modifying weights and biases to minimize the loss function. The model is trained until the loss function gets as low as possible and is stabilized. The loss function figures out how far we are from our desired solution or final convergence. The loss function is defined in eq. 1.

IV. RESULTS AND DISCUSSION

The ability and the efficiency of the ML-based approach are discussed in detail. The modeling results with respect to three types of fluctuation sources and their effects are studied comprehensively to understand the relation between the device electrical characteristics and various fluctuations. The variability caused by these fluctuation sources in important key parameters such as V_{TH}, I_{OFF}, and I_{ON} is also estimated by using the proposed ANN model. Therefore, the performance of the model is evaluated using rmse [60] and R^2-score [61] after the successful training. The rmse is the most common regression loss function and inquires how close the predicted value is to the real value. The R^2-score (coefficient of determination) represents the goodness of the fitting of explored ML model and is calculated as an evaluation technique. It defines the proportion of the variance in the dependent variable that is predictable from the independent variable. It varies between zero to 1. The higher value of the R^2-score shows that the input variables are perfectly correlated, whereas, a value closer to 0 shows that the ML model is not valid and suffering from many problems related to train/test data split, noise in the data, unavailability of tuned hyperparameters of the ML model, over fitting and so on.

The rmse and R^2 score are defined as:

$$rmse = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Simulated_i - Predicted_i)^2}, = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}, \quad (1)$$

and

$$R^2 = 1 - \frac{Residual\ sum\ of\ squares}{A\ total\ sum\ of\ squares} = 1 - \frac{\sum_i^n \left(Y_i - \hat{Y}\right)^2}{\sum_i^n (Y_i - Y_m)^2}, \quad (2)$$

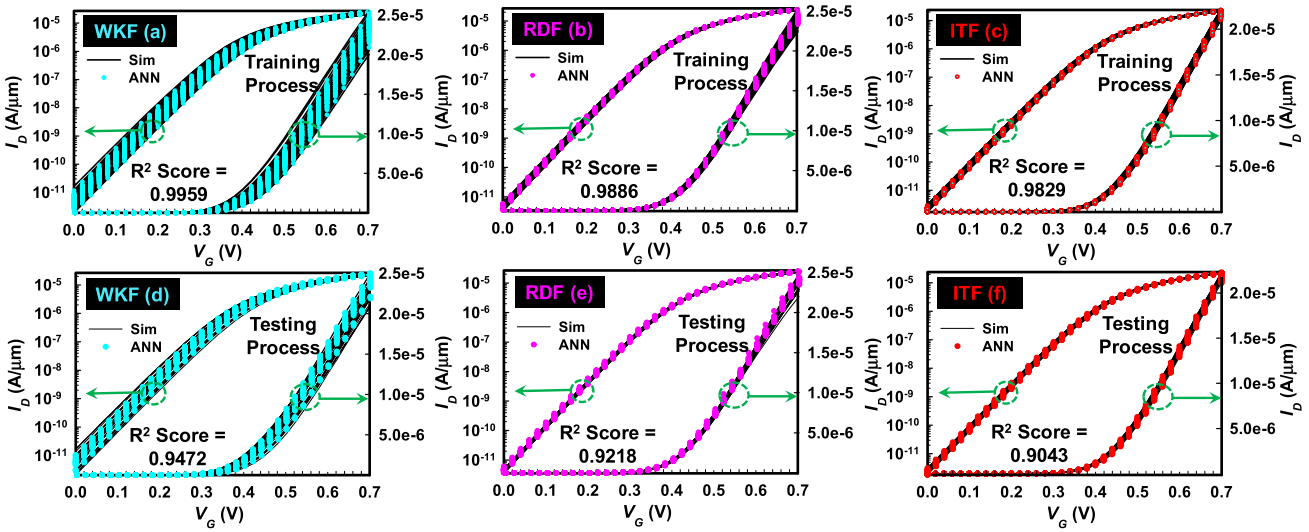

FIGURE 7. A comparison of I_D-V_G curves generated through device simulation and their respective predicted electrical characteristics through the ML-based ANN models. (a), (b) and (c) represent the I_D-V_G curves utilized by training of the three independent ANN models while considering the source of variations, i.e., WKF, RDF and ITF, respectively. Similarly, (e), (f) and (g) show the simulated and predicted I_D-V_G curves by testing the three independent ANN models for WKF, RDF and ITF, respectively.

where Y_i is the simulated value, $\hat{Y}$ is the ANN predicted value and Y_m is the mean of the simulated values. In the same way, we also calculate the error rate [62] in terms of the variance of the simulated dataset and the ANN predicted values to evaluate the performance of the model.

The error rate is defined as follows:

$$Error\ Rate = \frac{\sigma_{Sim} - \sigma_{Pred}}{\sigma_{Sim}} \times 100, \tag{3}$$

where, σ_{Sim} and σ_{Pred} represent the standard deviation of the simulated and the ANN predicted test dataset, respectively.

The training set of simulated and predicted *I-V* curves for various fluctuation sources (WKF, RDF, and ITF) are shown in Figs. 7(a)-(c), in both linear and logarithmic scales. The predicted training I-V curves for each fluctuation source show that the ANN models learn the variations accurately and there is a good agreement between ANN models predictions and simulated data. To further ensure this, the statistical measures such as rmse and R^2-score are also calculated from eq. 1 and eq. 2, respectively, and summarized in Table 4. It can be observed from the table that, for all ANN models, the training fitting parameter R^2 score is more than 98% which ensures that the models get trained efficiently. The minimum rmse is also achieved for the train as well as the test dataset. In addition to this, to determine the true performance of the models, the validation is done on the unseen test dataset for unbiased evaluation. Figs. 7(d)-(f) show the simulated and predicted test I_D-V_G curves. It can be observed from the test results that the predicted values are closely flowing with simulated values. The models predict the test data with R^2-score more than 90% and error rates are less than 3% for every case. The test error is likely to be higher than the training error for obvious reasons. Similarly, Figs. 8(a) and (b) show the training and testing results for CF with R^2-score of 0.9941 and 0.9138 for train and test results, respectively.

TABLE 4. R^2-score and RMSE values of the training and testing of various ann models with respect to their input source of variations such as WKF, RDF, ITF and CF, respectively.

Model	R^2-score (%) (Train)	R^2-score (%) (Test)	rmse (Train)	rmse (Test)
ANN_{WKF}	99.59	94.72	1.68e-07	2.66e-07
ANN_{RDF}	98.86	92.18	4.44e-07	4.79e-07
ANN_{ITF}	98.29	90.43	1.98e-07	2.55e-07
ANN_{CF}	99.41	91.38	3.77e-07	5.29e-07

From the analysis of the electrical characteristics (in Figs. 7 and 8), special attention is drawn to the fact that the effect of every fluctuation varies differently. The ITF induced very less variations on the device characteristics. The fluctuation induced by random dopants is moderate but has a low impact as compared to WKF. However, the device characteristics are majorly affected due to WKF. The variation, on electrical parameters, V_{TH}, I_{OFF}, and I_{ON}, caused by WKF is approximately five times more than ITF and RDF. The impact of CF on device characteristics has smaller variations than WKF because of the counterbalancing effect among fluctuation sources. Here at this point, it is interesting to note that despite the differences in the nature of fluctuation sources and their effects, the ANN model is efficient in modeling these effects on electrical characteristics of GAA Si NS MOSFETs. This shows the generality, robustness and the prediction capability of the proposed ANN model.

To further inspect the model accuracy, we extract the predicted parameters V_{TH}, I_{OFF}, and I_{ON} and compare them

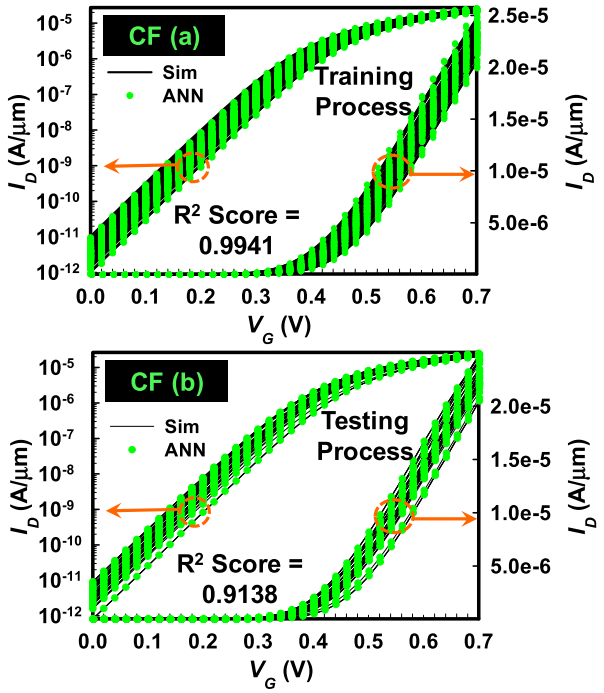

FIGURE 8. (a) and (b) show the comparison of training and testing results for simulated and predicted I_D-V_G curves of the ANN model using the combined fluctuations among WKF, RDF, and ITF of GAA Si NS MOSFETs.

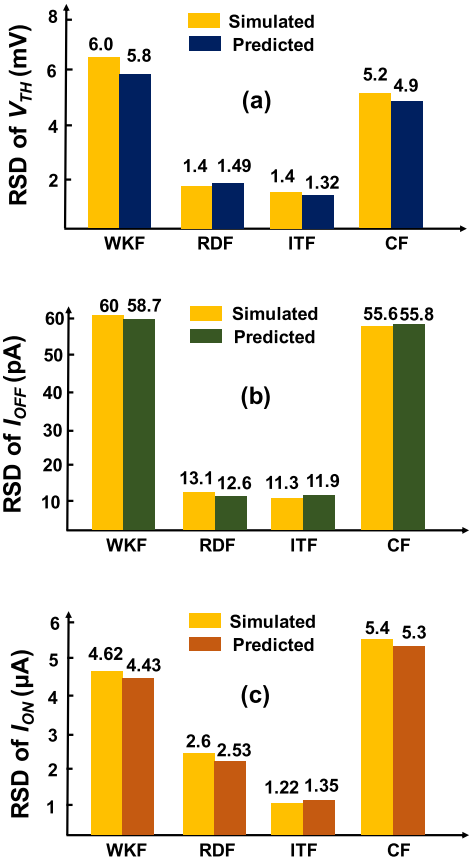

FIGURE 9. A RSD comparison between the device simulation values and the predicted values of ML-based ANN model, i.e., σ/μ of (a) V_{TH}, (b) I_{OFF}, (c) I_{ON}, using the testing dataset.

with the true simulated values by calculating the relative standard deviation which is useful when comparing the predicted results with the actual simulated data. The small differences between the simulated and predicted RSD represent, the ANN predicted the data more precisely. In Fig. 9, the RSD between actual and predicted V_{TH}, I_{OFF}, and I_{ON}, is shown using a bar graph for all fluctuation sources. The RSD, also known as the coefficient of variance, is a measure of deviation of simulated and predicted values from the mean values. The smaller is the deviation the closer will be the predicted values to the simulated values. The yellow bar shows the actual simulated RSD values represented as ''true'' and other colors i.e. green, orange, and blue bars show the predicted RSD value for V_{TH}, I_{OFF}, and I_{ON}, respectively. It can be observed from the bar chart that the RSD of predicted values is very close to the actual simulated values.

For better visual analysis of simulated and ANN predicted values, the violin plots for the three electrical characteristics are shown in Fig. 10. These plots are used to analyze the distribution of regression data. They are especially useful when the objective is to compare multiple outputs from multiple models. The plots have peaks, tails, and, valleys for every predicted and simulated value and show how close the values are. The x-axis represents parameters V_{TH}, I_{OFF}, and I_{ON} values for simulated and ANN predicted data whereas the y-axis represents observed output values. The plots clearly show that the peaks, tails, and valleys of the outputs predicted by the ANN models have almost identical position and shape as for simulation data. The width of each curve correlates with the approximate frequency of the data points in each region. These plots show one more way to confirm that the ANN models predicted the output efficiently.

The selection of independent variables plays a crucial role in the accuracy of the ML model. Fig. 11 shows the importance of each input parameter in predicting the impact of fluctuation sources on device characteristics. The feature importance is the measure of the relative importance of input features. A feature is important if a change in its value escalates the model error rapidly. The ML model firmly depends on such input features for high predictive accuracy. Likewise, a feature will be less important if the change in its value does not have much impact on the model output. The inputs can be classified into three separate categories: WKF, RDF, and ITF as marked on the y-axis. Features with higher values for this score are more important than features that have small values. We can see that the WKF category inputs are more important than the others. This is an expected outcome as the WKF has a predominant effect on the transfer characteristics. The second important category is the RDF, where the random dopant in the channel, source/drain extensions, and channel penetration affect the output significantly. The last category is the ITF inputs with minimum score and therefore, have less effect in predicting the fluctuation variations.

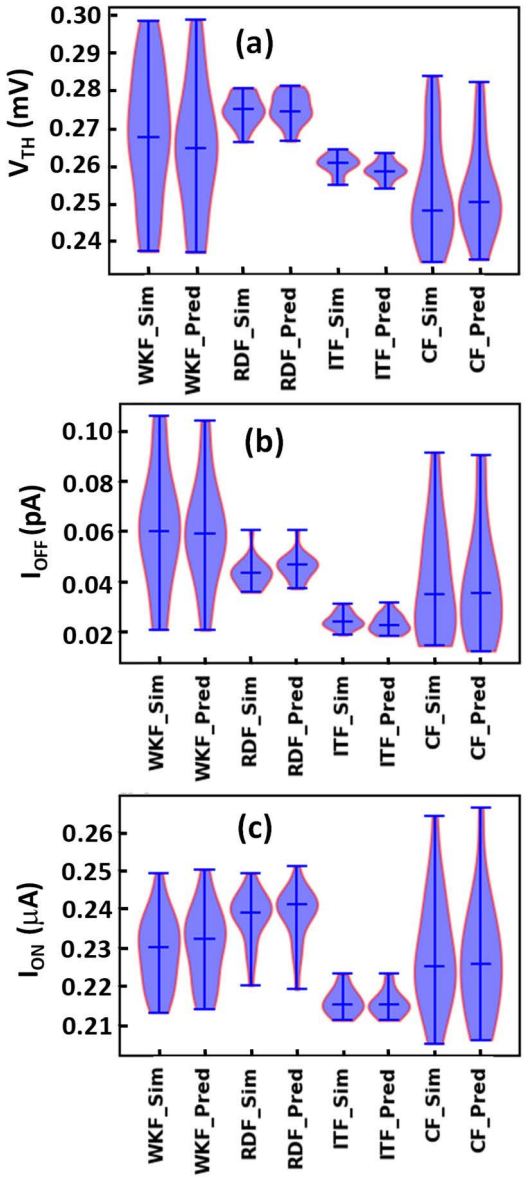

FIGURE 10. Illustrations of the violin plots for the simulated and predicted values for (a) V_{TH}, (b) I_{OFF}, and (c) I_{ON}, using testing dataset.

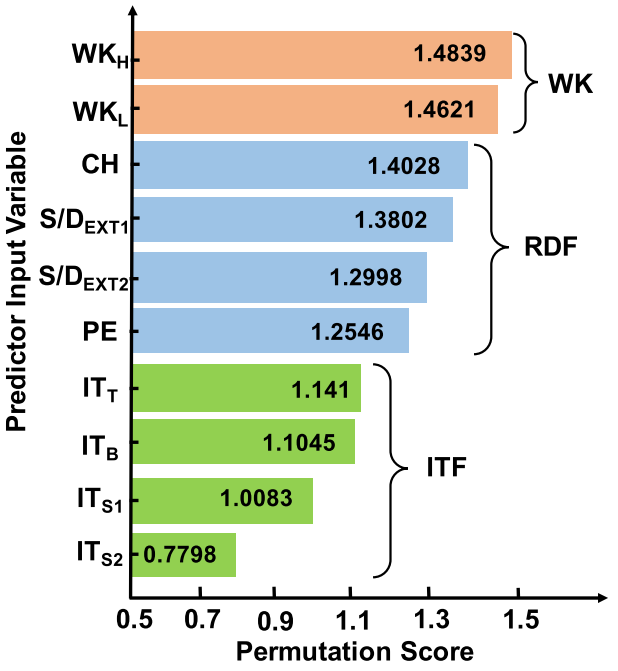

FIGURE 11. Comparing all sources of variation, permutation score determines the most important parameter contributes the modeling and prediction of the ANN model. The higher value of WK_H and WK_L show that WKF has a greater impact on the electrical characteristics of GAA Si NS MOSFET.

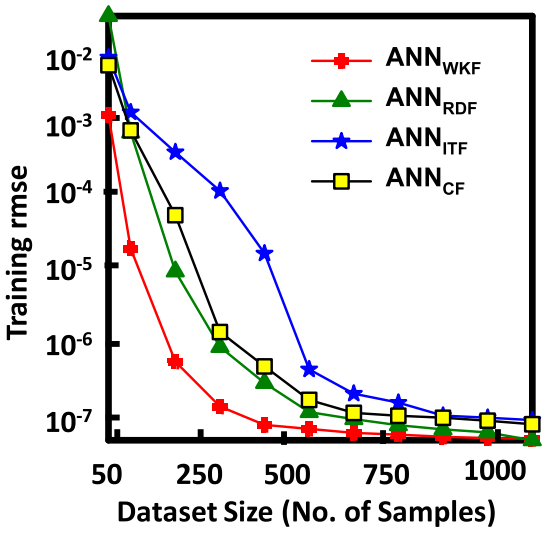

FIGURE 12. Comparison of the rmse values versus the number of required training size for all ANN models with respect to different sources of variation.

To ensure the dataset requirement in training the ANN models, the training rmse is calculated against the different dataset sizes (number of samples). The model is trained with different sample sizes and the corresponding rmse is recorded. The plot of rmse versus different dataset sizes is shown in Fig. 12. The rmse value decreases abruptly as the number of samples is increases. The error reached saturation at around 800 input samples. Results indicate that the ANN-based ML models learned the trends successfully even from the different fluctuation datasets with high predictive accuracy and have a good performance on both training and testing datasets. Moreover, the ML model can be a time-efficient method for fluctuation analysis. In this work, simulations take approximately 30 mins/sample. In order to generate 4000 samples, approximately 83 days have been taken. However, with the same computing resources, the ANN model is trained in only 180 seconds. This shows that the ML is much faster than simulation and the best alternative to accelerate the device modeling process.

Here an important point to note is that to train the ML model we need the simulated data one time. Once the model is

well trained then we can predict as many devices as we want and can save time by avoiding further device simulations. In addition to this, transfer learning (TL) is one more promising way to avoid the need to train a model from scratch. The TL uses the knowledge of the trained ML model from the dataset of one specific device and applies them to evaluate other devices with similar characteristics. This way the ML development reduces the resources and required time to train new models. We can share the generalized knowledge of one ML model for different devices.

Furthermore, the leverage of the ML model can be extended to the circuit simulation as with every new advanced device, it takes just too long to extend the compact models or come up with a completely new model for a specific device. Therefore, in the future, our focus will be on the application of the ML model in the circuit simulation to minimize the huge computational power and time. Our purpose is to integrate the ML model into circuit simulation to accelerate the optimization and evaluation of the semiconductor manufacturing industry.

V. CONCLUSION

In summary, this study has comprehensively analyzed the potential of the ANN-based ML strategy in modeling the effect of fluctuation sources on electrical characteristics of GAA Si NS MOSFETs. A total of 4000 fluctuated devices are simulated, with intrinsic parameter fluctuation sources (WKF, RDF, ITF, and their combination), to collect a complete dataset for the training and testing of ANN models. Their independent as well as combined effects have been analyzed successfully by modeling the variations of threshold voltage, on-state current, and off-state current. The developed ANN models show good agreement for training and testing data with small rmse and high R^2-score for every case. The ANN models are effectively able to predict the variations from a microscopic change (caused by ITF) to a comparatively larger change (caused by WKF). This demonstrates the generality and the powerful prediction capability of the ANN model.

In addition, we applied ML to measure the influence of each input feature (of intrinsic parameter fluctuation sources) on predicting the transfer characteristics while analyzing their combined effect. Based on these scores, we inferred that the WKF inputs have a greater impact on device characteristics than other fluctuation inputs. Furthermore, the ANN model speeds up the process considerably by taking only a few minutes to train and effectively minimizes the time and complexity challenges associated with full numerical device simulation. As a result, we believe that ML modeling technique has tremendous potential to enhance the design and fabrication of semiconductor nano-devices in the near future.

APPENDIX

A. DETAILS OF GENERATION OF WKF SAMPLES

The device we explored is the GAA Si NS MOSFET with TiN/HfO_2 gate stack which has an EOT of 0.6 nm. To generate the WKF-simulated devices, random high and low WKs on the gate area of the explored device are statistically assigned according to the material property of metal gate. First, we partitioned the gate area of the NS device into several small sub-regions, as shown in Fig. 3(a'). The device consists of TiN as a material for metal gate and we set the grain orientations of TiN as TiN<111> for high WK and TiN<200> for low WK, as shown in Fig. 3(a''). Different probabilities of occurrence are assigned to high and low WK. High WK has a probability of 60% and low WK has a probability of 40%. The work function of high WK has 4.6eV and low WK has 4.4 eV and the effective work function is 4.52 eV. The metal grains are fixed to 80 grains for all WK-fluctuated devices. The grain size is calculated as follows: (1) Total gate area $= 2[W_{NS} + t_{SIO2} + t_{HfO2}] + [H_{NS} + t_{SIO2} + t_{HfO2}] \times L_G = 2(29.6 + 9.6)\ ] \times 12 = 940.8\ nm^2$; and (2) Number of grain = gate area / grain area = $940.8/3.92 \times 3 = 80$ MG. This way a total of 1000 randomly generated samples are picked up by our Monte Carlo (MC) program and devices are simulated to examine the WKF. Furthermore, in Fig. 3(a''') the histograms show that the sub-planes are following the Gauss distribution.

B. DETAILS OF GENERATION OF RDF SAMPLES

We illustrate in Fig. 3(b) various regions of the GAA Si NS MOSFET with different types of random dopants appearing at channel (*CH*), source extension (S_{EXT}), drain extension (D_{EXT}), and penetration from the S/D extensions into the channel (*PE*) where random dopants are generated. First, 1725, 19396, and 1718 dopants were randomly generated in a large cuboid consisting of 1000 small cuboids, where the equivalent doping concentration are 5×10^{17} cm^{-3}, 1.1×10^{19} cm^{-3}, and 3.36×10^{17} cm^{-3} for the *CH*, S/D_{ext}, and *PE*, as shown in Figs. 3(b')-(b''''), respectively. The Gaussian distribution is involved to generate the dopants, as shown in the histograms in Figs. 3(b_1)-(b_4). These cuboids are partitioned into 1000 sub-cuboids for RDF devices.

The exact number of dopants is calculated as Number of Dopants = (Length of large cuboid × width of large cuboid × height of large cuboid × density of plain) / (dopant length × dopant width × dopant height × dopant concentration). A step function N_A is used to define the concentration and position of dopants [16].

$$N_A = \sum_{i=0}^{k} N_A{}^{Dopant} [H\ (x - x_1, y - y_1, z - z_1) - H\ (x - x_u, y - y_u, z - z_u) \quad \text{(B1)}$$

and

$$H\ (x,y,z) = \left\{ \begin{array}{l} 1,\ x \geq 0, y \geq 0, z \geq 0 \\ 0,\ otherwise \end{array} \right\}, \quad \text{(B2)}$$

where, (x_1, y_1, z_1) and (x_u, y_u, z_u) are the lower and upper coordinates of a random dopant, respectively; k is the number of dopants in the device region. N_A^{dopant} is the associated doping concentration for a doping concentration for dopant within a cuboid. In the generation of random dopants, we mainly follow our earlier work [5].

C. DETAILS OF GENERATION OF ITF SAMPLES

For the device simulation of the ITF source of variation, 2700 acceptor-type are generated randomly in a considerably larger 2D plane. The plane size is 224 $\times$ 224 nm^2 and its concentration is 3 $\times$ 10^{13} cm^{-2}, as shown in Fig. 3(c'). Furthermore, the entire plane is segregated into many sub-planes and each sub-planes are consisting of a random number of ITF. Predominantly, the number of RITs is uncertainly ranging from 1 to 14, and, the average number of RITs is 9. Whereas, the sub-planes having no RIT are counted as the nominal cases inducing no threshold variability in the explored device. Moreover, Fig. 3(c'') illustrates the dimension of the sub-plane segment having the size of 25 $\times$ 12 nm^2 of the top as well as the bottom walls. The size of the side walls is 12 $\times$ 5 nm^2. Therefore, the dimension of each RIT is 2 $\times$ 2 nm^2 and the density concentration is varying in the range of 1.5 $\times$ 10^{13} cm^{-2}eV^{-1} to 7.6 $\times$ 10^{13} cm^{-2} eV^{-1}. In order to determine the density of RIT (D_{it}) of the segregated plane, it can be calculated by the relation of the trap's density and its corresponding trap energy, as depicted in Fig. 3(c'''). Additionally, we have considered the same trap's energy in the entire 3D device simulation process for a rational outcome. Thus, using this simulation procedure, 1000 randomly statistical 3D device simulations can be performed to scrutinize the electrical characteristics of the advanced nanodevices.

D. COMMENTS ON UNDERFITTING

The machine learning model underfits when it is not able to capture the input-output relationship of the data. In this condition, the model has high bias and low variance. In the case of underfitting, the ML model does not train efficiently and produces a high error and low R^2-score on training data. The underfitting is usually occurred due to less training dataset. Therefore, to avoid the underfitting, we generated enough data samples for training the ML model.

E. COMMENTS ON OVERFITTING

Overfitting occurs when the ML model performs well on training data but has poor performance on test data. In this condition, the model has low bias and high variance. These models are too closely fit on the training data with very less error and very high R^2-score but on the contrary, they have very high error and low R^2-score on unseen test data. There are a number of techniques to deal with the overfitting of the ML model. Since our model is ANN which is trained by an iterative process so we use the ''Early Stopping'' regularization technique from the sklearn machine learning library. We want to train the ANN model accurately with an appropriate number of epochs to let the model learn the relationship between the inputs and outputs. If the number of epochs is too high the model will get trained excessively on training data and overfit, and if the model is trained with less number of epochs then it will not learn the relationship properly and perform poorly. Therefore, we use ''Early Stopping'' which enables the training to be automatically stopped when a chosen metric has stopped improving. One more advantage of early stopping is that it also saves the training time and makes ML model faster and more robust.

REFERENCES

[1] J.-S. Yoon, J. Jeong, S. Lee, J. Lee, S. Lee, J. Lim, and R.-H. Baek, ''DC performance variations by grain boundary in source/drain epitaxy of sub-3-nm nanosheet field-effect transistors,'' *IEEE Access*, vol. 10, pp. 22032–22037, 2022, doi: 10.1109/ACCESS.2022.3154049.

[2] C. Enz, ''An MOS transistor model for RF IC design valid in all regions of operation,'' *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 1, pp. 342–359, Jan. 2002, doi: 10.1109/22.981286.

[3] M. Haselman and S. Hauck, ''The future of integrated circuits: A survey of nanoelectronics,'' *Proc. IEEE*, vol. 98, no. 1, pp. 11–38, Jan. 2010, doi: 10.1109/JPROC.2009.2032356.

[4] K. Takeuchi, A. Nishida, and T. Hiramoto, ''Random fluctuations in scaled MOS devices,'' in *Proc. Int. Conf. Simulation Semiconductor Processes Devices*, Sep. 2009, pp. 1–7, doi: 10.1109/SISPAD.2009.5290243.

[5] Y. Li, H.-W. Cheng, Y.-Y. Chiu, C.-Y. Yiu, and H.-W. Su, ''A unified 3D device simulation of random dopant, interface trap and work function fluctuations on high-*k*/metal gate device,'' in *IEDM Tech. Dig.*, Dec. 2011, pp. 107–110, doi: 10.1109/IEDM.2011.6131495.

[6] D. Jang, D. Yakimets, G. Eneman, P. Schuddinck, M. G. Bardon, P. Raghavan, A. Spessot, D. Verkest, and A. Mocuta, ''Device exploration of nanosheet transistors for sub-7-nm technology node,'' *IEEE Trans. Electron Devices*, vol. 64, no. 6, pp. 2707–2713, Jun. 2017, doi: 10.1109/TED.2017.2695455.

[7] S. R. Kola, Y. Li, C.-Y. Chen, and M.-H. Chuang, ''Statistical 3D device simulation of full fluctuations of gate-all-around silicon nanosheet MOSFETs at sub-3-nm technology nodes,'' in *Proc. Int. Symp. VLSI Technol., Syst. Appl. (VLSI-TSA)*, Apr. 2022, pp. 1–2, doi: 10.1109/VLSI-TSA54299.2022.9771002.

[8] N. Loubet, T. Hook, P. Montanini, C. W. Yeung, S. Kanakasabapathy, M. Guillom, T. Yamashita, J. Zhang, X. Miao, J. Wang, and A. Young, ''Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET,'' in *Proc. Symp. VLSI Technol.*, Jun. 2017, pp. 230–231, doi: 10.23919/VLSIT.2017.7998183.

[9] J. Ajayan, D. Nirmal, S. Tayal, S. Bhattacharya, L. Arivazhagan, A. S. A. Fletcher, P. Murugapandiyan, and D. Ajitha, ''Nanosheet field effect transistors—A next generation device to keep Moore's law alive: An intensive study,'' *Microelectron. J.*, vol. 114, Aug. 2021, Art. no. 105141, doi: 10.1016/j.mejo.2021.105141.

[10] S. R. Kola, Y. Li, and N. Thoti, ''Characteristics of gate-all-around silicon nanowire and nanosheet MOSFETs with various spacers,'' in *Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD)*, Sep. 2020, pp. 1–4, doi: 10.23919/SISPAD49475.2020.9241603.

[11] Y. Li, C.-H. Hwang, and M.-H. Han, ''Simulation of characteristic variation in 16 nm gate FinFET devices due to intrinsic parameter fluctuations,'' *Nanotechnology*, vol. 21, no. 9, Feb. 2010, Art. no. 095203, doi: 10.1088/0957-4484/21/9/095203.

[12] X. Li, X. Yang, Z. Zhang, T. Wang, Y. Sun, Z. Liu, X. Li, Y. Shi, and J. Xu, ''Impact of process fluctuations on reconfigurable silicon nanowire transistor,'' *IEEE Trans. Electron Devices*, vol. 68, no. 2, pp. 885–891, Feb. 2021, doi: 10.1109/TED.2020.3045689.

[13] F.-L. Yang, J.-R. Hwang, and Y. Li, ''Electrical characteristic fluctuations in sub-45 nm CMOS devices,'' in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2006, pp. 691–694, doi: 10.1109/CICC.2006.320881.

[14] W.-L. Sung, Y.-S. Yang, and Y. Li, ''Work-function fluctuation of gate-all-around silicon nanowire n-MOSFETs: A unified comparison between cuboid and Voronoi methods,'' *IEEE J. Electron Devices Soc.*, vol. 9, pp. 151–159, 2021, doi: 10.1109/JEDS.2020.3046608.

[15] L. Sponton, L. Bomholt, D. Pramanik, and W. Fichtner, ''A full 3D TCAD simulation study of line-width roughness effects in 65 nm technology,'' in *Proc. Int. Conf. Simulation Semiconductor Processes Devices*, Sep. 2006, pp. 377–380, doi: 10.1109/SISPAD.2006.282913.

[16] Y. Li, C. H. Hwang, and T. Y. Li, ''Random-dopant-induced variability in nano-CMOS devices and digital circuits,'' *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1588–1597, Aug. 2009, doi: 10.1109/TED.2009.2022692.

[17] Amita, A. Gorad, and U. Ganguly, "Analytical estimation of LER-like variability in GAA nano-sheet transistors," in *Proc. Int. Symp. VLSI Technol., Syst. Appl. (VLSI-TSA)*, Apr. 2019, pp. 1–2, doi: 10.1109/VLSI-TSA.2019.8804637.

[18] N. Seoane, J. G. Fernandez, K. Kalna, E. Comesana, and A. Garcia-Loureiro, "Simulations of statistical variability in *n*-type FinFET, nanowire, and nanosheet FETs," *IEEE Electron Device Lett.*, vol. 42, no. 10, pp. 1416–1419, Oct. 2021, doi: 10.1109/LED.2021.3109586.

[19] C. Jeong, S. Myung, I. Huh, B. Choi, J. Kim, H. Jang, H. Lee, D. Park, K. Lee, W. Jang, J. Ryu, M.-H. Cha, J. M. Choe, M. Shim, and D. S. Kim, "Bridging TCAD and AI: Its application to semiconductor design," *IEEE Trans. Electron Devices*, vol. 68, no. 11, pp. 5364–5371, Nov. 2021, doi: 10.1109/TED.2021.3093844.

[20] N. Cummins, A. Baird, and B. W. Schuller, "Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning," *Methods*, vol. 151, pp. 41–54, Dec. 2018, doi: 10.1016/j.ymeth.2018.07.007.

[21] F. Balducci, D. Impedovo, and G. Pirlo, "Machine learning applications on agricultural datasets for smart farm enhancement," *Machines*, vol. 6, pp. 1–22, Sep. 2018, doi: 10.3390/machines6030038.

[22] O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 2186–2193, doi: 10.1109/BigData.2017.8258167.

[23] R. R. Halde, "Application of machine learning algorithms for betterment in education system," in *Proc. Int. Conf. Autom. Control Dyn. Optim. Techn. (ICACDOT)*, Sep. 2016, pp. 1110–1114, doi: 10.1109/ICACDOT.2016.7877759.

[24] M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy artificial intelligence," *Government Inf. Quart.*, vol. 37, Jul. 2020, Art. no. 101493, doi: 10.1016/j.giq.2020.101493.

[25] H. Ghoddusi, G. G. Creamer, and N. Rafizadeh, "Machine learning in energy economics and finance: A review," *Energy Econ.*, vol. 81, pp. 709–727, Jun. 2019, doi: 10.1016/j.eneco.2019.05.006.

[26] S. M. Carta, S. Consoli, L. Piras, A. S. Podda, and D. R. Recupero, "Explainable machine learning exploiting news and domain-specific lexicon for stock market forecasting," *IEEE Access*, vol. 9, pp. 30193–30205, 2021, doi: 10.1109/ACCESS.2021.3059960.

[27] M. Li, O. Irsoy, C. Cardie, and H. G. Xing, "Physics-inspired neural networks for efficient device compact modeling," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 2, pp. 44–49, 2016, doi: 10.1109/JXCDC.2016.2636161.

[28] N. Hari, M. Ahsan, S. Ramasamy, P. Sanjeevikumar, A. Albarbar, and F. Blaabjerg, "Gallium nitride power electronic devices modeling using machine learning," *IEEE Access*, vol. 8, pp. 119654–119667, 2020, doi: 10.1109/ACCESS.2020.3005457.

[29] 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.

[30] 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, doi: 10.1109/ACCESS.2020.3019933.

[31] Y. Lei, X. Huo, and B. Yan, "Deep neural network for device modeling," in *Proc. IEEE 2nd Electron Devices Technol. Manuf. Conf. (EDTM)*, Mar. 2018, pp. 154–156, doi: 10.1109/EDTM.2018.8421454.

[32] R. Butola, Y. Li, and S. R. Kola, "Deep learning approach to modeling and exploring random sources of gate-all-around silicon nanosheet MOSFETs," in *Proc. Int. Symp. VLSI Technol., Syst. Appl. (VLSI-TSA)*, Apr. 2022, pp. 1–2, doi: 10.1109/VLSI-TSA54299.2022.9771019.

[33] 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, doi: 10.1109/ACCESS.2021.3059475.

[34] T. Wu and J. Guo, "Speed up quantum transport device simulation on ferroelectric tunnel junction with machine learning methods," *IEEE Trans. Electron Devices*, vol. 67, no. 11, pp. 5229–5235, Nov. 2020, doi: 10.1109/TED.2020.3025982.

[35] S.-C. Han, J. Choi, and S.-M. Hong, "Acceleration of semiconductor device simulation with approximate solutions predicted by trained neural networks," *IEEE Trans. Electron Devices*, vol. 68, no. 11, pp. 5483–5489, Nov. 2021, doi: 10.1109/TED.2021.3075192.

[36] C.-W. Teo, K. L. Low, V. Narang, and A. V.-Y. Thean, "TCAD-enabled machine learning defect prediction to accelerate advanced semiconductor device failure analysis," in *Proc. Int. Conf. Simulation Semiconductors Processes Devices (SISPAD)*, Sep. 2019, pp. 1–4, doi: 10.1109/SISPAD.2019.8870440.

[37] N. Chatterjee, J. Ortega, I. Meric, P. Xiao, and I. Tsameret, "Machine learning on transistor aging data: Test time reduction and modeling for novel devices," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Mar. 2021, pp. 1–9, doi: 10.1109/IRPS46558.2021.9405188.

[38] Y. S. Bankapalli and H. Y. Wong, "TCAD augmented machine learning for semiconductor device failure troubleshooting and reverse engineering," in *Proc. Int. Conf. Simulation Semiconductors Processes Devices (SISPAD)*, Sep. 2019, pp. 1–4, doi: 10.1109/SISPAD.2019.8870467.

[39] J. Chen, M. B. Alawieh, Y. Lin, M. Zhang, J. Zhang, Y. Guo, and D. Z. Pan, "PowerNet: SOI lateral power device breakdown prediction with deep neural networks," *IEEE Access*, vol. 8, pp. 25372–25382, 2020, doi: 10.1109/ACCESS.2020.2970966.

[40] H. Y. Wong, M. Xiao, B. Wang, Y. K. Chiu, X. Yan, J. Ma, K. Sasaki, H. Wang, and Y. Zhang, "TCAD-machine learning framework for device variation and operating temperature analysis with experimental demonstration," *IEEE J. Electron Devices Soc.*, vol. 8, pp. 992–1000, 2020, doi: 10.1109/JEDS.2020.3024669.

[41] J. Kim, S. J. Kim, J.-W. Han, and M. Meyyappan, "Machine learning approach for prediction of point defect effect in FinFET," *IEEE Trans. Device Mater. Rel.*, vol. 21, no. 2, pp. 252–257, Jun. 2021, doi: 10.1109/TDMR.2021.3069720.

[42] D. Ding, X. Wu, J. Ghosh, and D. Z. Pan, "Machine learning based lithographic hotspot detection with critical-feature extraction and classification," in *Proc. IEEE Int. Conf. IC Des. Technol.*, May 2009, pp. 219–222, doi: 10.1109/ICICDT.2009.5166300.

[43] G. A. Susto, M. Terzi, and A. Beghi, "Anomaly detection approaches for semiconductor manufacturing," *Proc. Manuf.*, vol. 11, pp. 2018–2024, Jun. 2017, doi: 10.1016/j.promfg.2017.07.353.

[44] K. Mehta, S. S. Raju, M. Xiao, B. Wang, Y. Zhang, and H. Y. Wong, "Improvement of TCAD augmented machine learning using autoencoder for semiconductor variation identification and inverse design," *IEEE Access*, vol. 8, pp. 143519–143529, 2020, doi: 10.1109/ACCESS.2020.3014470.

[45] Z. Zhang, R. Wang, C. Chen, Q. Huang, Y. Wang, C. Hu, D. Wu, J. Wang, and R. Huang, "New-generation design-technology co-optimization (DTCO): Machine-learning assisted modeling framework," in *Proc. Silicon Nanoelectron. Workshop (SNW)*, Jun. 2019, pp. 1–2, doi: 10.23919/SNW.2019.8782897.

[46] G. Gil, S. Park, and S. Woo, "Solution for model parameter optimization and prediction of next-generation device DC characteristics," *IEEE Trans. Electron Devices*, vol. 69, no. 4, pp. 1781–1785, Apr. 2022, doi: 10.1109/TED.2022.3154326.

[47] J. Wei, W. Mao, H. Fang, Z. Zhang, J. Zhang, B. Lan, and J. Wan, "Advanced MOSFET model based on artificial neural network," in *Proc. China Semicond. Technol. Int. Conf. (CSTIC)*, Jun. 2020, pp. 1–3, doi: 10.1109/CSTIC49141.2020.9282457.

[48] 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.

[49] C. Akbar, Y. Li, and W.-L. Sung, "Deep learning algorithms for the work function fluctuation of random nanosized metal grains on gate-all-around silicon nanowire MOSFETs," *IEEE Access*, vol. 9, pp. 73467–73481, 2021, doi: 10.1109/ACCESS.2021.3079981.

[50] H. Carrillo-Nuñez, N. Dimitrova, A. Asenov, and V. Georgiev, "Machine learning approach for predicting the effect of statistical variability in Si junctionless nanowire transistors," *IEEE Electron Device Lett.*, vol. 40, no. 9, pp. 1366–1369, Sep. 2019, doi: 10.1109/LED.2019.2931839.

[51] 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, doi: 10.1109/ACCESS.2020.3020066.

[52] J. Lim, J. Lee, and C. Shin, "Probabilistic artificial neural network for line-edge-roughness-induced random variation in FinFET," *IEEE Access*, vol. 9, pp. 86581–86589, 2021, doi: 10.1109/ACCESS.2021.3088461.

[53] 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.

[54] E. Oh, J. K. Lee, Y. Seo, and H. Shin, "Methodology to predict random telegraph noise induced threshold voltage shift using machine learning," in *Proc. 4th IEEE Electron Devices Technol. Manuf. Conf. (EDTM)*, Apr. 2020, pp. 1–4, doi: 10.1109/EDTM47692.2020.9117805.
[55] S. Moparthi, C. Yadav, G. K. Saramekala, and P. K. Tiwari, "Machine learning based device simulation using multi-variable non-linear regression to assess the impact of device parameter variability on threshold voltage of double gate-all-around (DGAA) MOSFET," in *Proc. IEEE 2nd Int. Conf. Circuits Syst. (ICCS)*, Dec. 2020, pp. 64–67, doi: 10.1109/ICCS51219.2020.9336608.
[56] C. Akbar, Y. Li, and W. L. Sung, "Machine learning aided device simulation of work function fluctuation for multichannel gate-all-around silicon nanosheet MOSFETs," *IEEE Trans. Electron Devices*, vol. 68, no. 11, pp. 5490–5497, Nov. 2021, doi: 10.1109/TED.2021.3084910.
[57] C. Akbar, Y. Li, and W.-L. Sung, "Deep learning approach to inverse grain pattern of nanosized metal gate for multichannel gate-all-around silicon nanosheet MOSFETs," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 4, pp. 513–520, Nov. 2021, doi: 10.1109/TSM.2021.3116250.
[58] A. Gulli and S. Pal, "Neural networks foundations," in *Deep Learning With Keras*, 1st ed. Birmingham, U.K.: Packt, 2017, pp. 9–43.
[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[60] S. Bregni, M. Carbonelli, D. De Seta, and D. Perucchini, "Impact of slave clock internal noise on Allan variance and root mean square time interval error measurements," in *Proc. IEEE Instrum. Meas. Technol. Conf. (IMTC)*, May 1994, pp. 1411–1414, doi: 10.1109/IMTC.1994.352160.
[61] A. C. Cameron and F. A. Windmeijer, "An R-squared measure of goodness of fit for some common nonlinear regression models," *J. Econ.*, vol. 77, no. 2, pp. 329–342, 1997, doi: 10.1016/S0304-4076(96)01818-0.
[62] C. Tofallis, "Least squares percentage regression," *J. Modern Appl. Stat. Methods*, vol. 7, no. 2, pp. 526–534, Nov. 2008, doi: 10.22237/jmasm/1225513020.

RAJAT BUTOLA received the master's degree in electronics and communication engineering from Jaypee University, Noida, Uttar Pradesh, India, in 2012. He is currently pursuing the Ph.D. degree with the Parallel and Scientific Computing Laboratory, Electrical Engineering and Computer Science International Graduate Program, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. His research interests include application of machine learning and deep learning algorithms in advance semiconductor device modeling.

YIMING LI (Member, IEEE) has been a Visiting Professor with Stanford University, Grenoble INP, and Tohoku University, since 2011. He is currently a Full Professor of electrical and computer engineering with the National Yang Ming Chiao Tung University, Hsinchu, Taiwan. He is the author or coauthor of more than 400 technical articles, including the book *Physics of Semiconductor Devices* (4th Edition, Wiley, 2021). His current research interests include computational electronics, device physics, semiconductor nanostructures, modeling and parameter extraction, biomedical and energy harvesting devices, and optimization techniques. He has been a Program Committee of IEDM, since 2011.

SEKHAR REDDY KOLA (Graduate Student Member, IEEE) received the B.S. degree in mathematics, electronics and computer science and the M.S. degree in electronics science from Sri Venkateswara University, Tirupati, Andhra Pradesh, India, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Parallel and Scientific Computing Laboratory, Electrical Engineering and Computer Science International Graduate Program, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. His research interests include device simulation of FinFETs, GAA nanowire MOSFETs, and nanosheet devices along with circuit design optimization in the semiconductor industry. He achieved the Best Paper Award from IEDMS 2019.