



# Compact Model Build Upon Dynamic Weighting Artificial Neural Network Approach for Complementary Field Effect Transistors

Rajat Butola<sup>ID</sup>, Yiming Li<sup>ID</sup>, Member, IEEE, and Sekhar Reddy Kola<sup>ID</sup>, Member, IEEE

**Abstract**—In this work, a dynamic weighting-artificial neural network (DW-ANN) methodology is presented for quick and automated compact model (CM) generation. It takes advantage of both TCAD simulations for high accuracy and SPICE simulations for cost-effectiveness. This methodology is developed for gate-all-around (GAA) silicon (Si) nanosheet (NS) complementary field effect transistor (CFET), a potential candidate for future CMOS technology due to its innate properties. The critical process variation (PV) sources that severely degrade the CFET performance are estimated using DW-ANN. It reduces the computation cost and predicts the effects of PV sources with less than 1% error. Furthermore, a compact DW-ANN-based Verilog-A model has been developed that captures the dc as well as transient behavior accurately for circuit-level analysis. CFET-based circuits such as inverter, 6T-static random access memory (SRAM), and ring oscillator (RO) are characterized and implemented using DW-ANN-based CM. The overall average error of the model is reported as less than 2%. Therefore, the proposed device and circuit modeling approach provides a feasible solution for the rapid compact modeling of new emerging devices with good convergence and high accuracy.

**Index Terms**—Circuit simulation, compact modeling, complementary FETs, dynamic weights, machine learning, nanosheet (NS), process variation (PV).

## I. INTRODUCTION

THE growth in the semiconductor market is driven by the continuous increase in transistor performance. However,

Manuscript received 6 July 2023; accepted 28 August 2023. Date of publication 22 September 2023; date of current version 2 January 2024. This work was supported in part by the National Science and Technology Council (NSTC), Taiwan, under Grant NSTC 112-2221-E-A49-171 and Grant NSCT 112-2218-E-006-009-MBK; and in part by the “2022 Qualcomm Taiwan Research Program, National Yang Ming Chiao Tung University (NYCU),” under Grant NAT-487835 SOW. The review of this article was arranged by Editor H. Agarwal. (Corresponding author: Yiming Li.)

Rajat Butola and Sekhar Reddy Kola are with the Parallel and Scientific Computing Laboratory, Electrical Engineering and Computer Science International Graduate Program, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan.

Yiming Li is with the Parallel and Scientific Computing Laboratory, Electrical Engineering and Computer Science International Graduate Program, the Institute of Communications Engineering, the Institute of Biomedical Engineering, the Department of Electronics and Electrical Engineering, the Institute of Pioneer Semiconductor Innovation, and the Institute of Artificial Intelligence Innovation, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan (e-mail: yml@nycu.edu.tw).

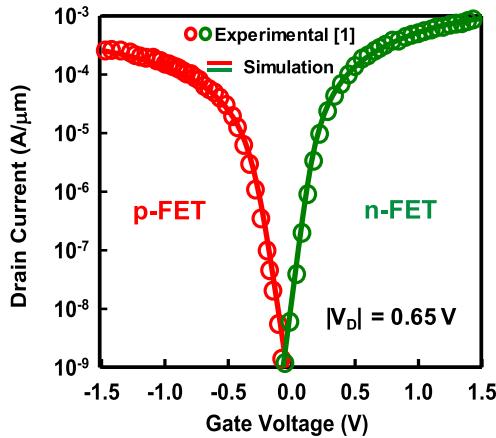
Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TED.2023.3312634>.

Digital Object Identifier 10.1109/TED.2023.3312634

recently CMOS technology has been scaled down rapidly to achieve the high packing density of integrated circuits (ICs) [1], [2], [3], [4]. This creates the need for a specific device structure that should consume less cell area. One such device that recently drew the attention of researchers is the gate-all-around (GAA) silicon (Si) nanosheet (NS) complementary field effect transistor (CFET) [5], [6]. CFET is a promising technique that can effectively reduce a device area by vertically stacking nFET and pFET on top of each other into the same structure. CFETs not only provide better area scaling benefits but also better power and performance over other emerging devices such as FinFETs and GAA FETs [7]. However, the aggressive scaling is causing unexpected issues in devices such as the process variation effect (PVE). Global variability (GV) sources (such as critical dimensions) are the most predominant variability sources and have always been a vital aspect of semiconductor fabrication. Hence, accurate modeling with respect to process variation (PV) is critical for the circuit performance benchmark [8]. Therefore, there exists a major requirement of a compact model (CM) for CFET devices to analyze and characterize the performance of ICs.

Compact modeling plays an imperative role in the description of emerging device behavior [9]. Before the fabrication of a new circuit design, CM enables the assimilation of those devices into the circuit simulation. However, developing a new CM is a very complicated task that takes a long time and requires special expertise in device physics [10]. Moreover, for such nanoscale devices, PV becomes more significant [11] and CM lacks robustness in terms of the parametric sweeps of circuits [12]. Thus, they are fast but unable to capture variation effects efficiently.

To overcome these challenges, machine learning (ML)-based attempts have been made in the recent past using artificial neural networks (ANNs) with and without device physics inclusion which provide a comprehensive platform to build CMs [13], [14], [15], [16]. The physics-based ANN-CMs that developed either lacked the computation speed [13] or the model shows non-convergence in certain voltages [15]. Notably, a recent work [17] addresses these problems successfully, however, these models still require skill and proficiency in device physics. Besides, pure ANN-based CMs are also implemented for new emerging devices [18], [19], [20], [21]. These ANN-based models drastically bring down the CM development time period and achieved optimized circuit performance. However, ANN is a fully data-driven approach, and the flexible architecture of ANN models makes them



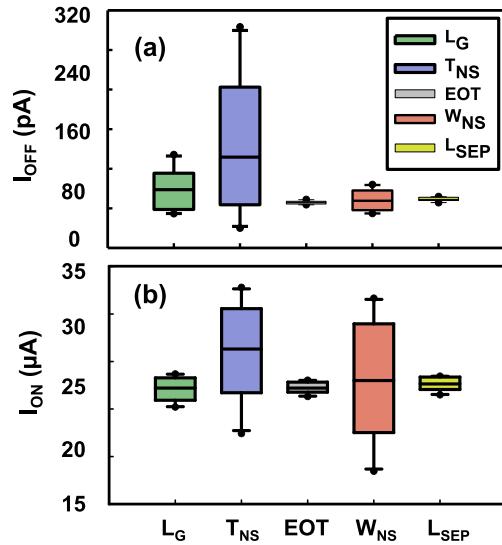
**Fig. 1.** Simulated (line) and experimental [1] (symbol) transfer characteristics for the accuracy calibration for nFET and pFET are shown.

suitable for PV data evaluation and provides them ability to accurately capture the nonlinear multivariate relationships in the device characteristics. However, they require massive datasets to train on. As the need for ANN accuracy grows, so does its data requirement. Likewise, the quality of the data is also significant, a lack of which can reduce the model's validity.

Therefore, in this article, we propose a novel dynamic weighting-artificial neural network (DW-ANN) strategy that uses a hybrid dataset of TCAD and SPICE simulations for training and takes both accuracy and time factor into account. Our goal is to develop a fast DW-ANN-based CM that can effectively project device-level performance to circuit level for both dc as well as transient simulations with minimum required training data. The rest of the article is organized as follows. Section II presents device parameters and an overview of PV. Section III discusses DW-ANN modeling strategy and integration with device simulation. In Section IV, predicted results for device and circuit simulation are discussed in detail. Section V consists of highlights of our contribution and a comparison with other related works. Finally, Section VI concludes the work of this article.

## II. DEVICE PARAMETERS AND DATA GENERATION

The CFET device for this study is simulated using the GAA NS structure. The simulated curves for nFET and pFET calibrated against the experimental data [1] are shown in Fig. 1. For accurate calibration we include the quantum correction model, quantum transport model such as the hydrodynamic model, density gradient quantization model, etc. Additional effects such as velocity saturation and impact ionization are also given due consideration and accurately modeled to ensure their incorporation into the TCAD simulations. In our previous study, we investigated different variability sources: NS gate length ( $L_G$ ), NS width ( $W_{NS}$ ), NS thickness ( $T_{NS}$ ), effective oxide thickness (EOT), and spacing between NS ( $L_{SEP}$ ). We analyze that  $L_G$ ,  $W_{NS}$ , and  $T_{NS}$  have major impacts on CFET electrical characteristics [22]. It can be deduced from Fig. 2 that for ON-current ( $I_{ON}$ ) and OFF-current ( $I_{OFF}$ ), the variation is most prominent for these three sources. Therefore, for this work, we consider these variability sources, and the corresponding current-voltage ( $I-V$ ) and capacitance-voltage ( $C-V$ ) characteristics are generated for nFET and



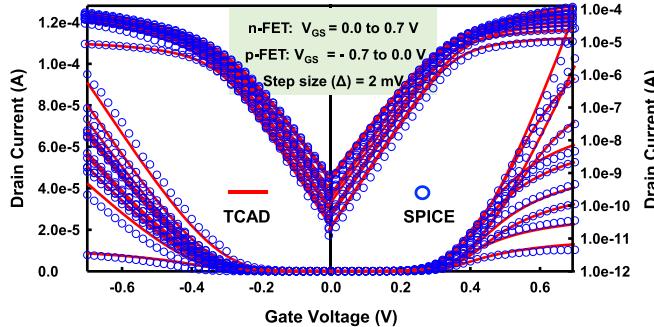
**Fig. 2.** Illustration of variation of process parameters ( $L_G$ ,  $T_{NS}$ , EOT,  $W_{NS}$ ,  $L_{SEP}$ ) on electrical properties, i.e., (a)  $I_{OFF}$  and (b)  $I_{ON}$  [22].

**TABLE I**  
LIST OF CFET NOMINAL DEVICE PARAMETERS CORRESPONDING  
TO SUB 3-nm TECHNOLOGY NODE AND SUMMARY  
OF PROCESS VARIABILITY SOURCES

Parameters	Nominal Value	Range
Gate Length ( $L_G$ ) (nm)	16	13 – 19 step size = 1
Width ( $W_{NS}$ ) (nm)	25	22 – 28 step size = 1
Thickness ( $T_{NS}$ ) (nm)	5	4 – 8 step size = 0.5

pFET of CFET. The nominal values and the variation of the device parameters are listed in Table I. Besides variability sources, the values for terminal voltages are set in the range of  $V_G$  [0: 0.7 V step = 0.02 V] and  $V_D$  [50 mV and 0.3:0.7 V step = 0.1 V] for  $I_D-V_G$  and  $V_D$  [0:0.7 V step = 0.02 V] and  $V_G$  [50 mV and 0.3 : 0.7 V step = 0.1 V] for  $I_D-V_D$ . Since accurate digital circuit simulation is possible for specific terminal voltages, appropriate  $V_D$  and  $V_G$  values are selected for transfer and output characteristics, respectively. This accumulates a total dataset of 2646 samples ( $7_{L_G} \times 7_{W_{NS}} \times 9_{T_{NS}} \times 6_{V_D}$ ) each for  $I_D-V_G$ ,  $I_D-V_D$ , and  $C-V$ . To generate this dataset, we use both TCAD simulation tool (through Sentaurus workbench) [23] as well as SPICE simulation tool [24] that used BSIM-CMG model [9] (using a transistor-level netlist by including same device geometries as TCAD), shown in Fig. 3.

Thus, the device characteristics are the function of critical dimensions and input bias voltages f ( $L_G$ ,  $W_{NS}$ ,  $T_{NS}$ ,  $V_G$ ,  $V_D$ ), which are modeled by the DW-ANN algorithm. The dataset is randomly divided by devices into three categories: a training set for hyperparameter fitting (70% TCAD and SPICE samples), a validation set for cross-validation during intermediate training of model (10% TCAD samples), and a test set for final model evaluation (20% TCAD samples). This dataset is collectively used by the DW-ANN model.



**Fig. 3.** It shows the deviation of the SPICE-generated transfer characteristics ( $I_D - V_G$ ) (for both nFET and pFET) with respect to TCAD-based  $I_D - V_G$  curves by utilizing identical PVs, device parameters, temperature, etc.

### III. DW-ANN COMPACT MODELING FRAMEWORK FOR CFET

DW-ANN is derived from the concept of dynamic neural network (DNN) which is emerging rapidly in the field of deep learning (DL) [25], [26]. Traditionally, static ANNs have been trained with fixed input parameters and fixed problem-solving skills which may limit their representation power, interpretability, and efficiency. Furthermore, the attributes of input and their dependencies keep changing the scenarios. Dynamic networks, on the other hand, can modify their parameters (such as weights) according to the inputs and offer additional advantages over static models in terms of efficiency, interpretability, adaptiveness, and generality. The DW-ANN algorithm can dynamically adjust the weights of its network according to the input attributes. Additionally, considering different inputs in PV data can have diverse computational demands, therefore, it is important to have dynamic parameters conditioned on each sample. Thus, the weights are adjusted to give more accurate inference and representation power to the model. The concept of the proposed DW-ANN is illustrated in Fig. 4. To begin with, the process parameters range is determined and the transistor level TCAD ( $X_{\text{INPUTS}}$ ,  $I_{D,\text{TCAD}}$ ) and SPICE ( $X_{\text{INPUTS}}$ ,  $I_{D,\text{SPICE}}$ ) samples are generated for nFET and pFET of CFET. The input vector ( $X_{\text{INPUTS}}$ ) includes critical PV sources  $L_G$ ,  $W_{\text{NS}}$ , and  $T_{\text{NS}}$  as well as bias inputs  $V_G$  and  $V_D$ . The high computational cost of the TCAD tool can be overcome by SPICE simulation with a slight compromise in accuracy. Therefore, the accuracy of the TCAD tool, acceleration of SPICE simulation, and compatibility of ML modeling are integrated into the proposed DW-ANN framework. This framework basically consists of three ANNs. Foremost, the SPICE samples are used to train the first  $\text{ANN}_1$  model. The predicted response of  $\text{ANN}_1$  is labeled as  $I_{D,\text{ANN}}$ , and weights of the network are also evaluated. These weights have less importance since the accuracy of the SPICE data is insufficient. Therefore, in the next stride, the TCAD data are incorporated to ensure overall accuracy. Thus, in the second step, based on three different current samples (i.e.,  $I_{D,\text{SPICE}}$ ,  $I_{D,\text{TCAD}}$ , and predicted  $I_{D,\text{ANN}}$ ), two sets of weights  $W_{\text{Target}}^{\text{ANN}}$  and  $W_{\text{Target}}^{\text{TCAD}}$  are generated by using the dynamic weighting function (DWF).

The DWF performs the calculations and estimates the weight matrix as represented in Table II. The outcomes of DWF are new weights ( $W_{\text{Target}}^{\text{ANN}}$ ,  $W_{\text{Target}}^{\text{TCAD}}$ ) associated with the target values. These weights have been defined with the

**TABLE II**  
WEIGHTS MATRIX FORMATION USING DWF

S.No.	Condition (if:)	Weights $W_{\text{Target}}^{\text{ANN}}$	Weights $W_{\text{Target}}^{\text{TCAD}}$
1	$I_{D,\text{SPICE}} > I_{D,\text{TCAD}} > I_{D,\text{ANN}}$ $I_{D,\text{SPICE}} < I_{D,\text{TCAD}} < I_{D,\text{ANN}}$	0	1
2	$I_{D,\text{SPICE}} > I_{D,\text{ANN}} > I_{D,\text{TCAD}}$ $I_{D,\text{SPICE}} < I_{D,\text{ANN}} < I_{D,\text{TCAD}}$	1	0
3	$I_{D,\text{TCAD}} > I_{D,\text{SPICE}} > I_{D,\text{ANN}}$ $I_{D,\text{TCAD}} < I_{D,\text{SPICE}} < I_{D,\text{ANN}}$	$(I_{D,\text{ANN}} - I_{D,\text{TCAD}}) / (I_{D,\text{SPICE}} - I_{D,\text{TCAD}})$	$(I_{D,\text{ANN}} - I_{D,\text{TCAD}}) / (I_{D,\text{SPICE}} - I_{D,\text{TCAD}})$

purpose that the TCAD values ( $I_{D,\text{TCAD}}$ ) should augment the accuracy in the final output. Thus, if the  $I_{D,\text{SPICE}}$  is too large and too low, i.e., out of the interval of  $I_{D,\text{TCAD}}$ , and  $I_{D,\text{ANN}}$ , then the values are transformed into (0, 1) and (1, 0) for  $W_{\text{Target}}^{\text{ANN}}$  and  $W_{\text{Target}}^{\text{TCAD}}$  (as shown in rows 1 and 2 of Table II). On the other hand, if  $I_{D,\text{SPICE}}$  value is in between  $I_{D,\text{TCAD}}$ , and  $I_{D,\text{ANN}}$  then the values for  $W_{\text{Target}}^{\text{ANN}}$  and  $W_{\text{Target}}^{\text{TCAD}}$  is calculated by a ratio (shown in row 3). These ratios are based on the values of  $I_{D,\text{SPICE}}$ ,  $I_{D,\text{TCAD}}$ , and  $I_{D,\text{ANN}}$  currents and assign appropriate values to weights  $W_{\text{Target}}^{\text{ANN}}$  and  $W_{\text{Target}}^{\text{TCAD}}$ . These updated weights are constructed in a way that the  $I_{D,\text{TCAD}}$  contribute to enhancing the overall accuracy of the model. The weights are modeled with the PV inputs using two other ANNs ( $\text{ANN}_2$  and  $\text{ANN}_3$ ) that are used to achieve adaptiveness and equilibrium in the DW-ANN. Each ANN has one hidden layer with five neurons and is trained with datasets ( $X_{\text{INPUTS}}$ ,  $W_{\text{Target}}^{\text{ANN}}$ ) and ( $X_{\text{INPUTS}}$ ,  $W_{\text{Target}}^{\text{TCAD}}$ ). The values of weights are updated and improved at each iteration until they converge. The predicted outputs of  $\text{ANN}_2$  and  $\text{ANN}_3$  are represented as  $W_{\text{Pred}}^{\text{TCAD}}$  and  $W_{\text{Pred}}^{\text{ANN}}$ , respectively. These predicted dynamic weights are utilized to determine the PV effects in nFET and pFET device characteristics. The outcome of DW-ANN model is weighted summation calculated below

$$I_{\text{DW-ANN}} = I_{D,\text{ANN}} \times W_{\text{Pred}}^{\text{ANN}} + I_{D,\text{TCAD}} \times W_{\text{Pred}}^{\text{TCAD}}.$$

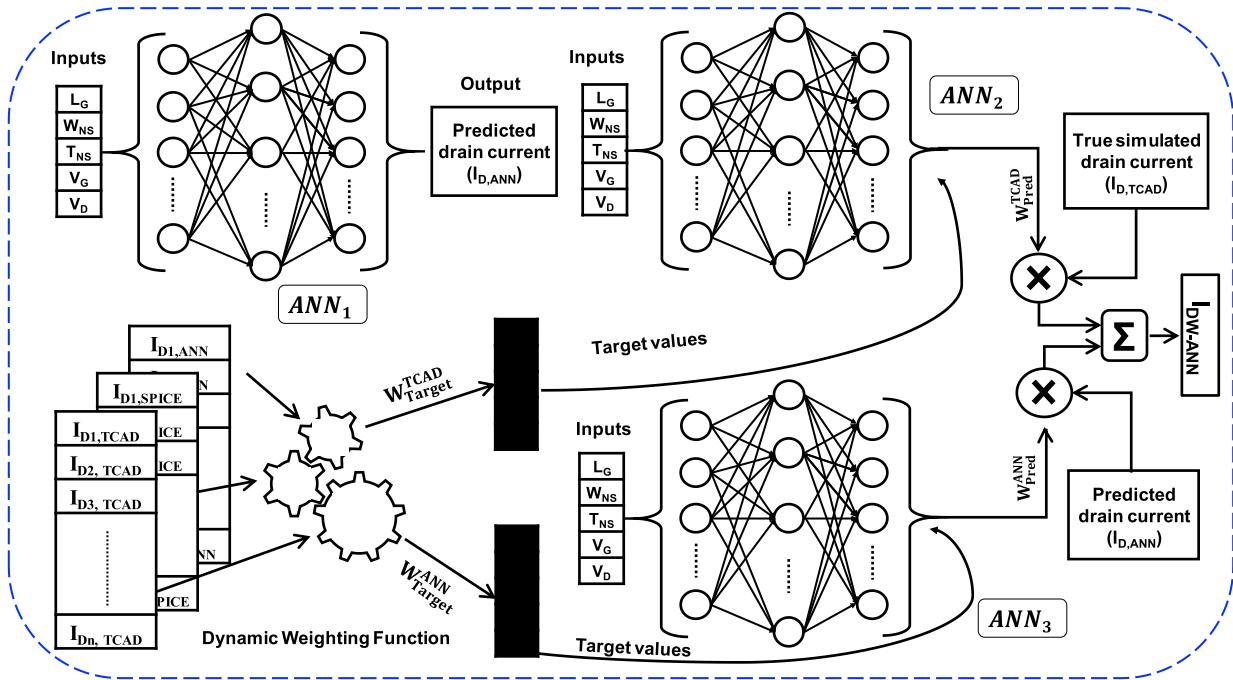
For circuit verification, the convergence issues of the model can affect the simulation results. Therefore, a conversion function is essential to ensure good model convergence. It is explicitly defined as

$$I_{D,\text{DW-ANN}} = V_D \times 10^{I_{\text{DW-ANN}}}.$$

Additionally, this function confirms no current at zero drain bias [18]. Loss function is an essential factor in determining the accuracy of a neural network since it directs the optimization process. The loss function for DW-ANN is defined as

$$\begin{aligned} \text{Loss} = & \text{RMSE}(I_{D,\text{true}} - I_{D,\text{pred}}) + a_1 \cdot \text{RMSE}(e^{I_{D,\text{true}}} - e^{I_{D,\text{pred}}}) \\ & + a_2 \cdot \text{RMSE}\left(\frac{\partial X}{\partial V_G}\right) + a_3 \cdot \text{RMSE}\left(\frac{\partial X}{\partial V_D}\right) \end{aligned}$$

where  $a_1$ ,  $a_2$ , and  $a_3$  are the weight coefficients that control the weighting of each component in the loss function and  $X$  stands for drain current/terminal charge for smoothing the higher-order derivatives with respect to  $V_G$  and  $V_D$ . The algorithm of DW-ANN is explained in Fig. 5. To implement the DW-ANN, we use Python's Keras library with Tensorflow working in its backend. Furthermore, instead of a fixed number of



**Fig. 4.** Implementation of the proposed DW-ANN model is illustrated. The dynamic network performs adaptive computation with respect to different input process variability sources. The model consists of three ANN models. The DWF provides the updated weights which are further trained with two other ANNs. The outcome of the DW-ANN model is the weighted summation.

#### Algorithm: Dynamic Weighting Artificial Neural Network

1. Procedure TRAIN
2.  $X_1(L_G, W_{NS}, T_{NS}, V_G, V_D) \leftarrow$  Training Dataset of size  $m \times n$
3.  $y_1(I_{D,SPICE}) \leftarrow$  Label of samples in  $X$
4.  $(X_1, y_1) \rightarrow ANN_1 \rightarrow I_{D,ANN}$  Input and output of the ANN<sub>1</sub>
5. For  $i = 1$  to  $m$ 
  - If (Condition 1 ( $I_{D,TCAD}, I_{D,SPICE}, I_{D,ANN}$ )):
  $w_{Target}^{ANN} = 0$   
 $w_{Target}^{TCAD} = 1$
  - Else if (Condition 2 ( $I_{D,TCAD}, I_{D,SPICE}, I_{D,ANN}$ )):
  $w_{Target}^{ANN} = 1$   
 $w_{Target}^{TCAD} = 0$
  - Else:
 DWF ( $(I_{D,TCAD}, I_{D,SPICE}, I_{D,ANN})$ )
6.  $X_2(L_G, W_{NS}, T_{NS}, V_G, V_D) \leftarrow$  Training Dataset for ANN<sub>2</sub>
7.  $y_2(w_{Target}^{ANN}) \leftarrow$  Dynamic Weights as a Label for ANN<sub>2</sub>
8.  $(X_2, y_2) \rightarrow ANN_2 \rightarrow w_{Pred}^{ANN}$  Input and output of ANN<sub>2</sub>
9.  $X_3(L_G, W_{NS}, T_{NS}, V_G, V_D) \leftarrow$  Training Dataset for ANN<sub>3</sub>
10.  $y_3(w_{Target}^{TCAD}) \leftarrow$  Dynamic Weights as a Label for ANN<sub>3</sub>
11.  $(X_3, y_3) \rightarrow ANN_3 \rightarrow w_{Pred}^{TCAD}$  Input and output of ANN<sub>3</sub>
12.  $I_{D,ANN} = I_{D,ANN} \times w_{Pred}^{ANN} + I_{TCAD} \times w_{Pred}^{TCAD} \leftarrow$  Final Output of DW-ANN model

**Fig. 5.** Step-by-step illustration of the dynamic weighting algorithm.

epochs, we use a regularization technique called early stopping to combat the overfitting issue and minimize the training time. It observes the performance of the model on a held-out validation set during the training phase and breaks off the training as soon as the validation error reaches the minimum. We use validation loss as a performance measure to stop the training and the patience value (number of epochs with no improvement) is set to 20. So, the training stops if there is no improvement in validation loss for continuous 20 epochs and avoids overfitting. In addition to this, to compare the results of the DW-ANN, we also train a conventional ANN model with three hidden layers and 18 neurons each as a baseline

**TABLE III**  
HYPERPARAMETERS OF THE DW-ANN AND BASELINE MODELS

Hyperparameters	DW-ANN	Baseline
(I/P, O/P) Dimension	(5, 1)	(5, 1)
No. of Hidden Layers	1	3
No. of Neurons in each layer	ANN <sub>1</sub> = 25, ANN <sub>w1</sub> = 5 ANN <sub>w2</sub> = 5	(18, 18, 18)
Activation Function	Tanh	Tanh
Learning Rate	0.001	0.001
Optimizer	Adam	Adam
Batch Size	25	25
Epochs (Fixed)	3000	3000
Early stopping Epoch	784	1090

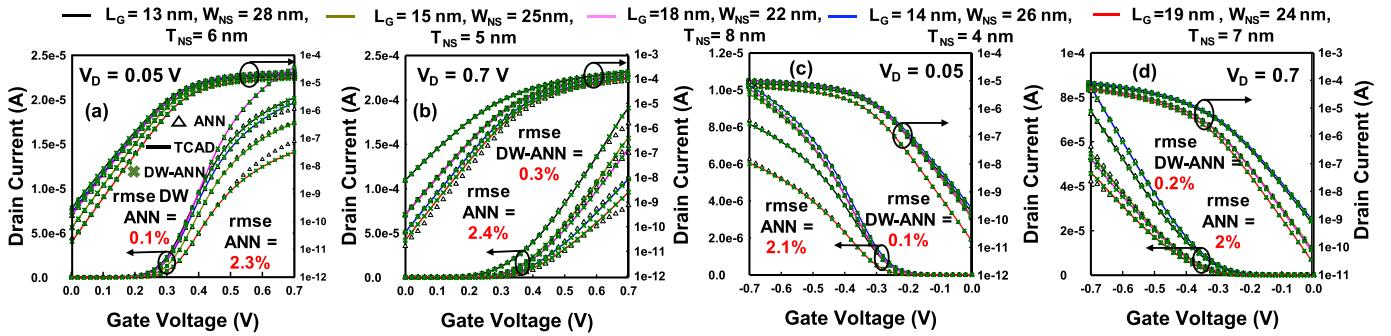
model. The other important hyperparameters of the DW-ANN model and the baseline model are summarized in Table III. The results are explained in detail in the next section.

## IV. RESULTS AND DISCUSSION

This section shows the contribution of the proposed DW-ANN model that performs at the device level for both nFET and pFET and at the circuit level [for CFET-based inverter, static random access memory (SRAM), and ring oscillator (RO)] using emerging GAA Si NS CFETs structure and demonstrates that the effects of PV distribution of CFET device can be captured accurately in an optimal time as compared to traditional simulation tools, i.e., TCAD and SPICE.

### A. DW-ANN Predictive Capability for CFET Modeling

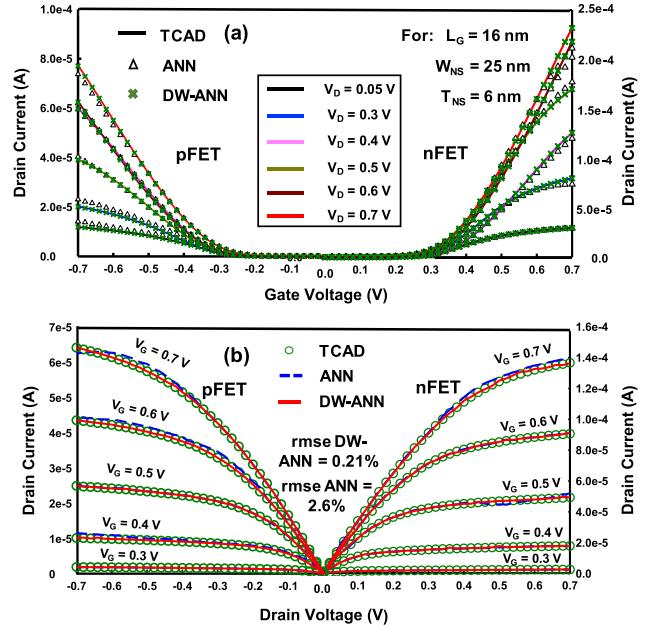
The training of DW-ANN and baseline ANN algorithm depends on the hybrid dataset gathered through two different



**Fig. 6.** Comparison of simulated (line) versus predicted (symbol) transfer characteristics ( $I_D - V_G$ ) curves obtained through baseline ANN, TCAD, DW-ANN for varying input process parameters. (a) and (b) Represent the  $I_D - V_G$  curves for nFET of CFET for different terminal biases  $V_D$  values, i.e., 0.05 V (linear) and 0.7 V (saturation), respectively, in linear and in logarithmic scale. Similarly, (c) and (d) show the  $I_D - V_G$  curves for pFET of CFET for  $V_D$  = 0.05 and 0.7 V, respectively.

simulation tools. After training the proposed DW-ANN and baseline ANN model, the performance of both models is evaluated through the testing dataset obtained through the TCAD simulation tool, as shown in Fig. 6. The  $I_D - V_G$  plots of nFET and pFET devices for different PV sources such as  $L_G$ ,  $W_{NS}$ , and  $T_{NS}$ , for linear bias ( $V_D = 0.05$  V) are illustrated in Fig. 6(a) and (c), respectively, in both linear and logarithmic scales. In the plots, the solid line represents the true TCAD values, and markers ( $\Delta$  and  $\times$ ) represent the ANN and DW-ANN predicted values, respectively. The predicted  $I_D - V_G$  curve of GAA Si NS CFET using different PV sources illustrates that the DW-ANN model fits the data accurately due to its capability to take benefit of both simulation tools, whereas the baseline could not perform well because of its lack of generalization. For example, for  $L_G = 13$  and 19 nm, the fitting of the baseline model failed to capture the variations in drain current, whereas the DW-ANN clearly outperforms. Our model successfully estimates the non-linearity of the device variability parameters by using the adaptive dynamic weight. The RMSE value recorded for DW-ANN is 0.1% for both linear and saturation biasing whereas, for the baseline model it is 2.3% ( $V_D = 0.05$  V) and 2.1% ( $V_D = 0.7$  V). Moreover, the coefficient of determination,  $R^2$ -score for DW-ANN is 99.98% which is very close to 1 as opposed to 98.35% for the baseline model. Similar results can be observed for  $V_D = 0.7$  V (saturation bias) in Fig. 6(b) and (d) for nFET and pFET, respectively.

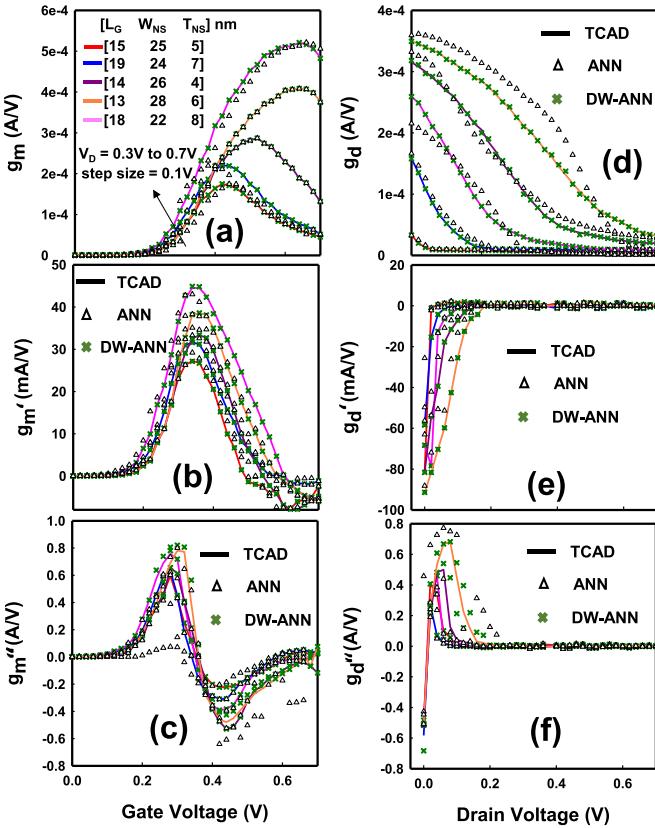
Fig. 7(a) shows the performance comparison of two methods with another set of results for a single nFET and pFET device at various bias conditions such as  $V_D = 50$  mV, 0.3 V, 0.4 V, 0.5 V, 0.6 V, and 0.7 V. The predicted  $I_D$  values obtained through the proposed model are perfectly matched with the true TCAD simulated values, however, the baseline model shows lower performance as compared to DW-ANN. Similarly, in Fig. 7(b), the performance of GAA Si NS CFET is evaluated by predicting the output characteristics ( $I_D - V_D$  curves) for various  $V_G$  values (0.3, 0.4, 0.5, 0.6, and 0.7 V) for nFET and pFET using DW-ANN and baseline models. The  $R^2$ -score of the baseline model for nFET and pFET devices measured with respect to the TCAD-generated  $I_D - V_D$  curves is 98.13% and 98.25%, respectively. On the other hand, an improved  $R^2$ -score of 99.87% and 99.93% is achieved for the proposed DW-ANN model for nFET and pFET, respectively. Each case study shows that the DW-ANN



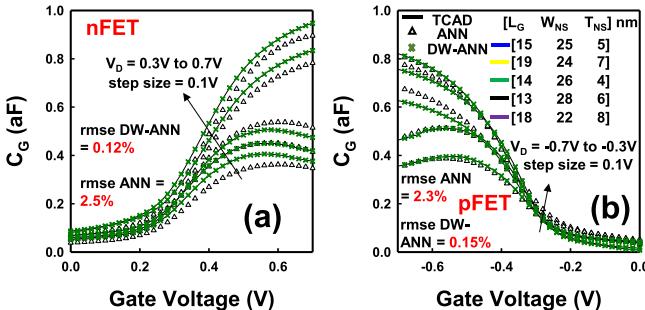
**Fig. 7.** Comparison of (a) transfer characteristics curves for different drain biases  $V_D = 50$  mV and 0.3–0.7 V with step size of 0.1 V and (b) output characteristics curves for different gate bias  $V_G$  = 0.3–0.7 V step size of 0.1 V for n and p-FET devices with  $L_G = 16$  nm,  $W_{NS} = 25$  nm and  $T_{NS} = 6$  nm.

model performs better as compared to the baseline ANN model in terms of  $R^2$ -score as well as RMSE value.

In order to implement a robust DW-ANN-based CM, it is important to satisfy high-order derivatives of transfer and output characteristics for circuit simulation applications. Consequently, the accurate modeling up to third order derivative of  $I_D - V_G$  and  $I_D - V_D$  characteristics including transconductance ( $g_m$ ), output conductance ( $g_d$ ),  $\partial^2 I_D / \partial V_G^2$  ( $g'_m$ ),  $\partial^3 I_D / \partial V_G^3$  ( $g''_m$ ),  $\partial^2 I_D / \partial V_D^2$  ( $g'_d$ ), and  $\partial^3 I_D / \partial V_D^3$  ( $g''_d$ ) are evaluated that is critical for circuit analysis as well as for iterative solving algorithms in SPICE circuit simulation. The performance of DW-ANN prediction for  $g_m$ ,  $g_d$  and higher order derivatives is shown in Fig. 8(a)–(f), respectively, under different bias conditions, i.e.,  $V_D = 0.3$ , 0.4, 0.6, and 0.7 V. The proposed model shows an excellent agreement with TCAD values for up to second order derivatives whereas the baseline clearly lacks the ability to replicate the target values. Even for third order derivative the DW-ANN has satisfactory performance,

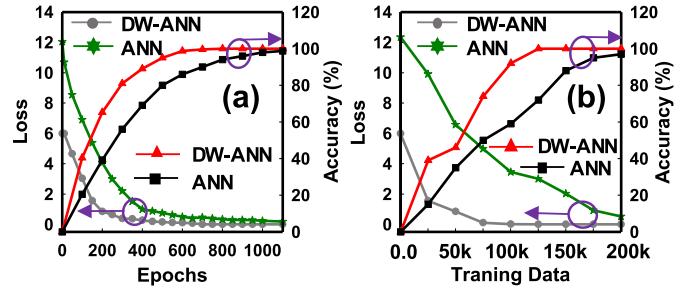


**Fig. 8.** Comparison of TCAD versus baseline and DW-ANN predicted results for higher order derivatives of transconductance and output conductance. (a)  $g_m$  versus  $V_G$ , (b)  $g'_m$  versus  $V_G$ , (c)  $g''_m$  versus  $V_G$ , (d)  $g_d$  versus  $V_D$ , (e)  $g'_d$  versus  $V_D$ , and (f)  $g''_d$  versus  $V_D$  for different terminal biases  $V_D = 0.3\text{--}0.7$  V, with step size of 0.1 V.



**Fig. 9.** (a) and (b) Show capacitive behavior for nFET and pFET, respectively with an average RMSE of 0.13% for the DW-ANN and 2.4% for the baseline.

however, at high-order derivatives, the TCAD itself suffers due to the numerical errors. Additionally, the assessment of the capacitive behavior of CFET is also important for the transient analysis of circuits. Therefore, the DW-ANN model is also trained for capacitance values with respect to terminal voltage for both nFET and pFET for different PVs, as shown in Fig. 9(a) and (b). The performance metric (RMSE value) of DW-ANN model is always less than 1% and more accurate as compared to the ANN model. Thus, the  $C_G$ - $V_G$  characteristic is significant for the evaluation of circuit performance. Moreover, the results are evident for the compatibility of studying the effect of PV of GAA Si NS CFET using the DW-ANN model. Therefore, it can be utilized as an alternative to a conventional CM for general-purpose circuit simulation.



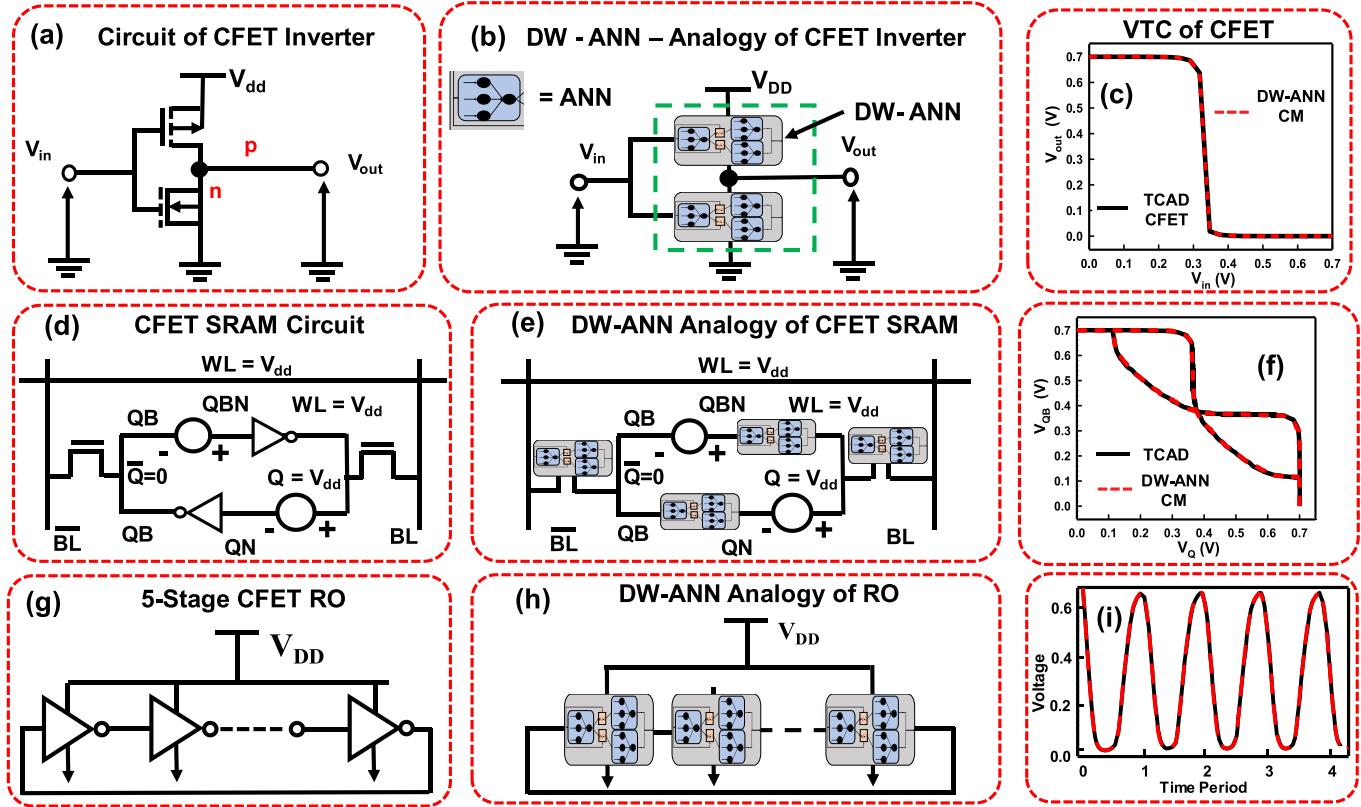
**Fig. 10.** Comparison of loss (in terms of RMSE) and accuracy (in terms of  $R^2$ -score) with respect to (a) number of epochs and (b) training data samples for DW-ANN and baseline models.

To compare the performance of the proposed DW-ANN model and the baseline model, the correlation between accuracy, epoch, loss, and acquired training data is calculated, as shown in Fig. 10. Here, accuracy and loss are examined in terms of  $R^2$ -score and RMSE value, respectively. In Fig. 10(a), it can be observed that DW-ANN converges rapidly and depicts minimum RMSE value at an early stage, i.e., at 784 epochs, as compared to the baseline model which takes 1090 epochs to get trained. Hence, it can be stated that the DW-ANN model is computationally efficient and can achieve more than 99% accuracy with minimum loss in a short timeframe. Similarly, in Fig. 10(b), the DW-ANN outclasses the baseline model in terms of loss and accuracy by acquiring 50% smaller number of training samples. Therefore, DW-ANN is not only cost-efficient but also requires less amount of data. The DW-ANN model has the proficiency to predict diverse devices for different dimensions and biases. It can be an efficient method for device engineering and variation analysis. For this study, TCAD simulation takes 30 min/sample and SPICE takes 0.6 s/sample approximately. To generate the whole dataset, 15 days are consumed by TCAD. In contrast, the DW-ANN training is accomplished within 45 min and then the testing takes even less than 5 ms, which is very fast compared to the TCAD simulation. In other words, the DW-ANN model prediction is approximately a million times faster than TCAD simulations if the ML model is properly designed and tested. The DW-ANN is even more time efficient than baseline ANN because all three neural networks used in DW-ANN are shallow networks with only one hidden layer and fewer neurons and hence take fewer epochs to get trained. In short, by taking advantage of both TCAD and SPICE techniques, our optimal ML structure of the proposed model also helps to accelerate the time and reduce computing resources required for accurate circuit simulation analysis.

### B. Circuit Verification of DW-ANN-Based CM

To achieve the projection of device-level performance to the circuit level, the CM is developed using DW-ANN. The proposed framework is applied to the CFET-based circuit by characterizing and implementing the simulations of logic circuits for dc as well as transient simulations. Results for circuit simulations based on the DW-ANN CM are discussed in this section.

**1) DC Simulation Performance:** The weights and biases of the trained model are extracted to integrate the DW-ANN into the circuit simulation. These weights and biases are then incorporated into the Verilog-A program to capture the



**Fig. 11.** Illustration of circuit simulation using DW-ANN-based ML methodology. DC and transient analyses of logical circuits are shown. (a), (d), and (g) Illustrate the circuits of inverter, SRAM, and five-stage RO, respectively. Similarly, (b), (e), and (h) represent the equivalent analogy of these circuits. (c) Shows the comparison of VTCs between TCAD and DW-ANN-based inverter circuit. (f) Depicts the comparison between 6T SRAM butterfly curve using TCAD simulation and DW-ANN-based CM circuit simulation. Similarly, (i) represent the verification of the five-stage CFET-based RO.

device behavior accurately for circuit-level analysis. To verify the dc simulation performance of DW-ANN-based CM, we explore the behavior of inverter and 6T-SRAM digital circuits based on CFET structure. Due to the unavailability of the commercial CM of CFET, circuit simulation of CFET structure relies on the computationally expensive TCAD simulation. Therefore, we compare the results of DW-ANN CM with TCAD-simulated circuits. The conventional CFET-based inverter and its equivalent DW-ANN analogy are shown in Fig. 11(a) and (b).

The result validates that the DW-ANN-based CM is closely following the TCAD waveform with an error of 0.5%. The agreement between TCAD-based and DW-ANN-based CM circuit simulation shows that DW-ANN-based CM inverter has the capability to produce an agreeable voltage transfer characteristic (VTC) waveform with a minimum computational cost without compromising accuracy. Similarly, Fig. 11(d) and (e) show the schematic of CFET-based and DW-ANN-based SRAM circuits, respectively. In addition, Fig. 11(f) depicts the simulation results of the SRAM butterfly curves for DW-ANN against TCAD model with 0.3% error. The excellent matching of waveforms shows that DW-ANN CM has the ability to accurately reproduce SRAM circuit.

**2) Transient Simulation Performance:** For transient analysis using DW-ANN CM, the standard five-stage RO circuit based on CFET is chosen for the experiment as it is the most popular circuit topologies in the digital world. In this work, RO consists of five CFETs connected in series that are composed of an odd number of inverters in a ring where the response of the last inverter is fed back to the first inverter. Its schematic

is shown in Fig. 11(g). The DW-ANN analogy of RO is illustrated in Fig. 11(h). As we know, the C–V characteristics are significant and highly desirable for the transient analysis of logic circuits, therefore their modeling is important. The output of RO oscillates between two voltage levels. This can be verified from Fig. 11(i), where the transient simulation result of the RO circuit for the DW-ANN model is oscillating between high and low voltages along with the TCAD result. The error is recorded as less than 1.5%. Thus, the proposed approach allows precise dc as well as transient analysis of complex circuits without any convergence issues.

## V. HIGHLIGHTS OF OUR CONTRIBUTIONS

The contribution of our work is discussed in this section by drawing a comparison with other related works. As listed in Table IV, we have closely examined the limitations of prior papers [10], [18], [19], [20], and [21], and have manifested the methodology that contributes in the following ways.

- 1) For the first time, we have presented the DW-ANN model that is trained using the accurate TCAD simulation and by minimizing the computational cost through SPICE simulation.
- 2) To the best of our knowledge, this is the first attempt to construct neural network (NN)-based CM for CFET devices.
- 3) This work predicted the testing devices instead of just modeling the untrained portion of training devices.
- 4) The proposed DW-ANN model requires the lowest data acquisition for training and to implement the CM for

**TABLE IV**  
COMPARISON OF DW-ANN METHODOLOGY WITH THE PAST WORKS THAT UTILIZED ANN MODELS FOR CIRCUIT SIMULATION

Examined features	[10]	[18]	[19]	[20]	[21]	This work
Explored device	TFT	GAA FETs	MOSFET	GAA FETs	GAA CSFET	GAA Si NS CFET
ML Novelty	C-ANN	C-ANN	C-ANN	C-ANN	MNN	DW-ANN
Device parameters	L, W	L, W, Temp.	L, W, Temp.	-	L, W	L <sub>G</sub> , W <sub>NS</sub> , T <sub>NS</sub>
Training data required	High	High	Very High	High	High	Low
Circuit for DC analysis	No	No	Yes	Yes	No	Yes
Circuit for transient analysis	No	Yes	No	No	Yes	Yes
Circuit implemented	No	RO	LDO	Inverter	RO, Op-Amp	Inverter, SRAM, RO

\*C-ANN = Conventional Artificial Neural Network,

MNN = Multigradient Neural Network

LDO = Low-Dropout Regulator

multiple logic circuits (inverter, SRAM and RO) to perform dc and transient analysis.

## VI. CONCLUSION

This study comprehensively analyzed the potential application of ML in the device and circuit modeling to assess the process variability in CFETs. A novel DW-ANN algorithm has been proposed that can take advantage of both highly accurate TCAD simulation and cost-effective SPICE simulation to develop a fast and accurate CM. We showed that the analysis of variability sources using our proposed approach can considerably improve the calculation efficiency besides maintaining the high accuracy of TCAD simulation. It also overcomes the limitation of classical static ANNs by dynamically adjusting the weights of its network according to the input attributes. DW-ANN has been evaluated in terms of its scalability and fitting accuracy for dc as well as transient analysis on CFET-based inverter, 6T-SRAM, and RO circuits with less than 2% error and without convergence issues. Our future research focuses on reducing the TCAD data samples to further minimize the training data generation time while maintaining the same high accuracy of the model.

## REFERENCES

- [1] C.-Y. Huang et al., “3-D self-aligned stacked NMOS-on-PMOS nanoribbon transistors for continued Moore’s law scaling,” in *IEDM Tech. Dig.*, Dec. 2020, pp. 20.6.1–20.6.4.
- [2] J. Kedzierski et al., “Extension and source/drain design for high-performance finFET devices,” *IEEE Trans. Electron Devices*, vol. 50, no. 4, pp. 952–958, Apr. 2003.
- [3] A. Kar, S. Sarker, A. Dasgupta, and Y. S. Chauhan, “Impact of corner rounding on quantum confinement in GAA nanosheet FETs for advanced technology nodes,” in *Proc. Device Res. Conf. (DRC)*, Columbus, OH, USA, Jun. 2022, pp. 1–2.
- [4] D. Jang et al., “Device exploration of NanoSheet transistors for sub-7-nm technology node,” *IEEE Trans. Electron Devices*, vol. 64, no. 6, pp. 2707–2713, Jun. 2017.
- [5] S.-G. Jung, D. Jang, S.-J. Min, E. Park, and H.-Y. Yu, “Performance analysis on complementary FET (CFET) relative to standard CMOS with nanosheet FET,” *IEEE J. Electron Devices Soc.*, vol. 10, pp. 78–82, 2022.
- [6] X.-R. Yu et al., “Integration design and process of 3-D heterogeneous 6T SRAM with double layer transferred Ge/2Si CFET and IGZO pass gates for 42% reduced cell size,” in *IEDM Tech. Dig.*, Jan. 2022, pp. 20.5.1–20.5.4.
- [7] J. Ryckaert et al., “The complementary FET (CFET) for CMOS scaling beyond N3,” in *Proc. IEEE Symp. VLSI Technol.*, Jun. 2018, pp. 141–142.
- [8] T. Dutta, G. Pahwa, A. Agarwal, and Y. S. Chauhan, “Impact of process variations on negative capacitance FinFET devices and circuits,” *IEEE Electron Device Lett.*, vol. 39, no. 1, pp. 147–150, Jan. 2018.
- [9] J. P. Duarte et al., “BSIM-CMG: Standard FinFET compact model for advanced circuit design,” in *Proc. 41st Eur. Solid-State Circuits Conf. (ESSCIRC)*, Graz, Austria, Sep. 2015, pp. 196–201.
- [10] Q. Chen and G. Chen, “Artificial neural network compact model for TFTs,” in *Proc. 7th Int. Conf. Comput. Aided Design Thin-Film Transistor Technol. (CAD-TFT)*, Oct. 2016, p. 1.
- [11] X. Yang et al., “Impact of process variation on nanosheet gate-all-around complementary FET (CFET),” *IEEE Trans. Electron Devices*, vol. 69, no. 7, pp. 4029–4036, Jul. 2022.
- [12] Z. Zhang, X. Jiang, R. Wang, S. Guo, Y. Wang, and R. Huang, “Extraction of process variation parameters in FinFET technology based on compact modeling and characterization,” *IEEE Trans. Electron Devices*, vol. 65, no. 3, pp. 847–854, Mar. 2018.
- [13] M.-Y. Kao, H. Kam, and C. Hu, “Deep-learning-assisted physics-driven MOSFET current-voltage modeling,” *IEEE Electron Device Lett.*, vol. 43, no. 6, pp. 974–977, Jun. 2022.
- [14] X. Gao, A. Huang, N. Trask, and S. Reza, “Physics-informed graph neural network for circuit compact model development,” in *Proc. Int. Conf. Simul. Semiconductor Processes Devices (SISPAD)*, Sep. 2020, pp. 359–362.
- [15] G. Qi et al., “The device compact model based on multi-gradient neural network and its application on MoS<sub>2</sub> field effect transistors,” in *Proc. 6th IEEE Electron Devices Technol. Manuf. Conf. (EDTM)*, Oita, Japan, Mar. 2022, pp. 88–90.
- [16] Y.-S. Yang, Y. Li, and S. R. Kola, “A physical-based artificial neural networks compact modeling framework for emerging FETs,” *IEEE Trans. Electron Devices*, early access, May 5, 2023, doi: [10.1109/TED.2023.3269410](https://doi.org/10.1109/TED.2023.3269410).
- [17] K. Sheelvardhan, S. Guglani, M. Ehteshamuddin, S. Roy, and A. Dasgupta, “Machine learning augmented compact modeling for simultaneous improvement in computational speed and accuracy,” *IEEE Trans. Electron Devices*, pp. 1–7, 2023, doi: [10.1109/TED.2023.3251296](https://doi.org/10.1109/TED.2023.3251296).
- [18] 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.
- [19] J. Wei, H. Wang, T. Zhao, Y.-L. Jiang, and J. Wan, “A new compact MOSFET model based on artificial neural network with unique data preprocessing and sampling techniques,” *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 42, no. 4, pp. 1250–1254, Apr. 2023.
- [20] C. T. Tung, M. Y. Kao, and C. Hu, “Neural network-based and modeling with high accuracy and potential model speed,” *IEEE Trans. Electron Devices*, vol. 69, no. 11, pp. 6476–6479, Nov. 2022.
- [21] 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.
- [22] S. R. Kola, Y. Li, and M.-H. Chuang, “Intrinsic parameter fluctuation and process variation effect of vertically stacked silicon nanosheet complementary field-effect transistors,” in *Proc. 24th Int. Symp. Quality Electron. Design (ISQED)*, San Francisco, CA, USA, Apr. 2023, pp. 1–8.
- [23] *Sentaurus Device User Guide, Version O-2018.06*, Synopsys, Mountain View, CA, USA, 2018.
- [24] *HSPIRE Reference Manual*, Synopsys, Mountain View, CA, USA, 2016.
- [25] Y. Han, G. Huang, S. Song, L. Yang, H. Wang, and Y. Wang, “Dynamic neural networks: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7436–7456, Nov. 2022.
- [26] Z.-Q. Shen and F.-S. Kong, “Dynamically weighted ensemble neural networks for regression problems,” in *Proc. Int. Conf. Mach. Learn. Cyberv.*, vol. 6, Aug. 2004, pp. 3492–3496.