

UC Berkeley

UC Berkeley Previously Published Works

Title

Neural Network-Based BSIM Transistor Model Framework: Currents, Charges, Variability, and Circuit Simulation

Permalink

<https://escholarship.org/uc/item/9mx680pk>

Journal

IEEE Transactions on Electron Devices, 70(4)

ISSN

0018-9383

Authors

Tung, Chien-Ting

Hu, Chenming

Publication Date

2023-04-01

DOI

10.1109/ted.2023.3244901

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

Neural Network-Based BSIM Transistor Model Framework: Currents, Charges, Variability, and Circuit Simulation

Chien-Ting Tung, *Graduate Student Member, IEEE*, and Chenming Hu, *Life Fellow, IEEE*

Abstract—We present a neural network (NN)-based transistor modeling framework which includes drain, source, and gate currents and charges, and their variabilities. The training data is generated by a Berkeley Short-channel IGFET Model (BSIM) with ranges of channel lengths, widths, and oxide thicknesses. The NNs are trained to learn the geometry dependence. The drain, source, and gate currents are modeled with one NN, and the charges by another NN. The NNs are trained to produce accurate variability prediction and derivatives of currents and charges. Quality and robustness tests such as Gummel symmetry, harmonic balance, and ring oscillator are performed and show excellent results.

Index Terms— Compact model, machine learning, neural network (NN), variability modeling, field-effect transistor (FET).

I. INTRODUCTION

Transistor models are important to the semiconductor industry which needs fast and accurate models for circuit simulation and design optimization. Industry-standard compact models, such as the Berkeley Short-channel IGFET Model (BSIM) series of models [1, 2], use physics-based equations. Developing accurate and computationally efficient analytic equations for each and every complex transistor behavior, such as short channel effects and quantum effects in gate-all-around transistors [3], can be time-consuming.

Neural network (NN)-based compact models [4-6] hold the potential of reducing the time of developing models of future new devices. The matrix multiplication nature and the ease of GPU acceleration endow NN-based compact models with the potential of reducing the time needed for model calculation during circuit simulation. Several previous works have studied using NN to model the variation in transistors. Ref. [7, 8] uses NN to predict some key merits in process variation such as I_{ON} , I_{OFF} , and V_{TH} without modeling the entire IV characteristics. Ref. [9] uses process variation parameters as inputs to train a NN that can reproduce IV characteristics line tunnel FETs. Still,

This work was supported by the Berkeley Device Modeling Center, University of California at Berkeley, Berkeley, CA, USA. The review of this article was arranged by Editor XXX.

(Corresponding author: Chien-Ting Tung.)

The authors are with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: cttung@berkeley.edu).

much more investigation is needed to determine whether NN-based models meet all the requirements of a practical compact model.

In this work, we present a NN-based compact model of source, drain, gate currents and charges, with variability modeling; and demonstrate its robustness for circuit simulation. Different from our previous work [10] which only models drain current and gate charge, gate and drain leakage currents are included in the neural networks. Those leakages are important in evaluating the circuit performance. Furthermore, source and drain charges are also included which are essential and contribute to the transient currents at source/drain terminals, which is also what BSIM does to ensure charge conservation. All currents are included in one network and all charges are included in the other network. The improved loss functions are developed to accurately train the networks with these additional outputs by considering higher order derivatives. We focus on four process variation parameters: gate length (L), fin height (H_{FIN}), equivalent oxide thickness (EOT) and work function difference ($\Delta\phi$) of FinFETs. It is shown that the demonstrated NN model can fit the IV and CV characteristics in the presence of process variations to the higher-order derivatives of current and charge. This model can also predict the statistical distribution of the device merits when used in Monte Carlo simulations.

II. MODELING FRAMEWORK

A. Currents

To have a NN trained with the process variations of L , H_{FIN} , EOT, and $\Delta\phi$, we include these parameters as inputs in NN together with V_{GS} and V_{DS} . Fortunately, for $\Delta\phi$, we know from physics that the effect of $\Delta\phi$ is equivalent to a gate voltage shift. Therefore, $\Delta\phi$ does not need to be included among the inputs, rather it is treated as a gate voltage shift during training and inferencing as shown in (1). In this way, we can reduce the complexity of the NN and training time. Other variations such as fin thickness (T_{FIN}) and temperature (T) are not included in this work for simplicity and will be added in the future follow-up study. The IV NN is trained to model drain current (I_D) and gate current (I_G). Again, for simplicity in this study, we assume a floating-body device so the $I_B=0$. There are 3 outputs for the IV NN. The first output y_1 is the transform of I_D as shown in (2) [10]. For I_G , we cannot easily determine its sign as I_D . To make I_G scaled by \ln function, we separate it into positive

and negative parts using the smoothing functions as (3) and transform them into y_{2p} and y_{2n} where Δ is the smoothing factor and I_0 is used to prevent 0. Thus, there are three outputs of this NN. The loss function is improved from [10] and shown in (4), where RMS is the root-mean-square error function, g_m is the transconductance, g_m' is its derivative, g_{ds} is the output conductance, g_{ds}' is its derivative, and a to f are the coefficients for each loss. We include up to second-order derivatives to obtain the desired accuracy. The process of determining the coefficients in (4) is similar to [10].

$$x = [V_{GS} - \Delta\phi, V_{DS}, H_{FIN}, L, EOT, (T_{FIN}, T...)], \quad (1)$$

$$I_D = V_{DS} e^{y_1}, \quad y_1 = \ln\left(\frac{I_D}{V_{DS}}\right), \quad (2)$$

$$y_{2p} = \ln\left(\frac{I_G}{2} + \frac{\sqrt{I_G^2 + \Delta^2}}{2} + I_0\right), \quad (3)$$

$$y_{2n} = \ln\left(\frac{-I_G}{2} + \frac{\sqrt{I_G^2 + \Delta^2}}{2} + I_0\right), \quad (3)$$

$$\begin{aligned} loss = & a \cdot RMS(y_1) + b \cdot RMS(g_m) + c \cdot RMS(g_{ds}) + d \cdot RMS(g_m') \\ & + e \cdot RMS(g_{ds}') + f \cdot RMS(y_{2p}) + f \cdot RMS(y_{2n}), \end{aligned} \quad (4)$$

B. Charges

For QV NN, different from our previous work [10] where we only trained Q_G , here we train Q_G , Q_S , and Q_D all in one network and Q_B is $-(Q_G + Q_S + Q_D)$. Inputs are the same as (1) and outputs are shown in (5). In the loss function (6), we also include up to second-order derivatives to obtain good accuracy where a' to f' are the coefficients. The y in (6) represents $y_{1,2,3}$ in (5).

$$y_{1,2,3} = Q_{G,S,D}, \quad (5)$$

$$\begin{aligned} loss = & a' \cdot RMS(y) + b' \cdot RMS\left(\frac{\partial y}{\partial V_{GS}}\right) + c' \cdot RMS\left(\frac{\partial y}{\partial V_{DS}}\right) \\ & + e' \cdot RMS\left(\frac{\partial^2 y}{\partial V_{GS}^2}\right) + f' \cdot RMS\left(\frac{\partial^2 y}{\partial V_{DS}^2}\right), \end{aligned} \quad (6)$$

III. IMPLEMENTATION & RESULTS

The NN is implemented with the Tensorflow package in Python and uses tanh as the activation function. The training data is generated using a BSIM-CMG [1] model that is calibrated to the Intel 10nm-node FinFET [11] with 10 fins and L is 18nm and H_{FIN} is 46nm. We use that scaling capability of BSIM-CMG to generate data for $L=[14, 16, 18, 20, 22, 24]$ nm, $H_{FIN}=[38, 42, 46, 50, 54]$ nm, and $EOT=[0.68, 0.73, 0.78, 0.83, 0.88]$ nm for training the NN to cover the range of possible device variations. Training uses 150 devices. Each device has a full I_D & I_G characteristic with V_{GS} and V_{DS} varying from -0.8V to 0.8V. The importance of training the full bias spectrum is discussed in [10].

The training results are shown in Fig. 1 & 2 where we show the I_D fitting for several L , H_{FIN} , and EOT combinations.

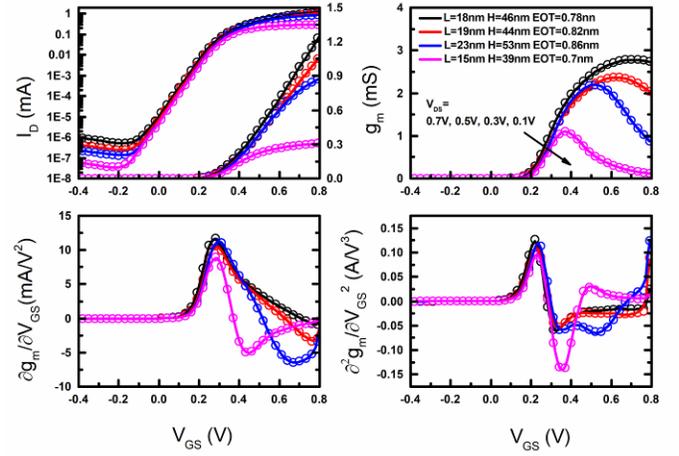


Fig. 1. The fitted $I_D V_G$ curves at different L , H_{FIN} , and EOT where the lines are the NN and symbols are the BSIM-CMG data.

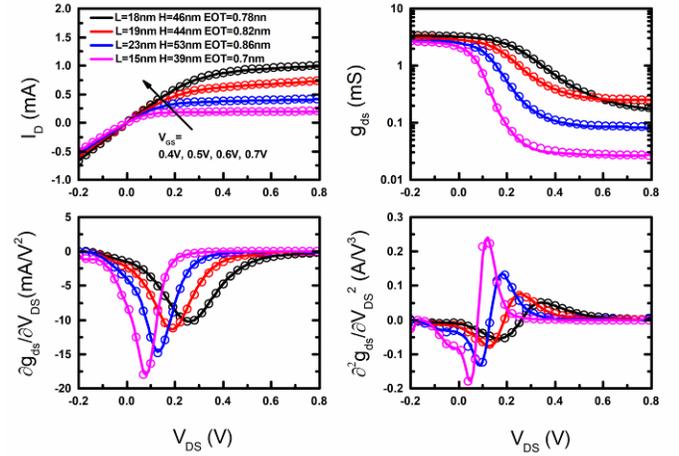


Fig. 2. The fitted $I_D V_D$ curves at different L , H_{FIN} , and EOT where the lines are the NN and symbols are the BSIM-CMG data.

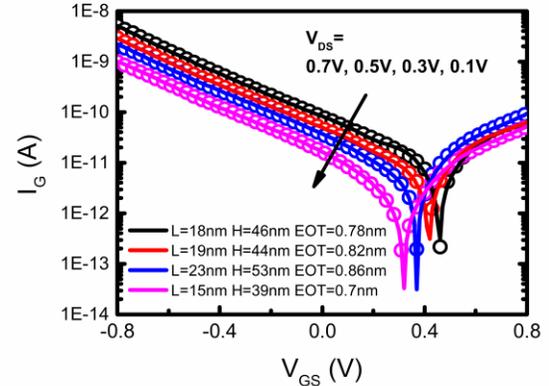


Fig. 3. The fitted I_G characteristics for different structures where the lines are the NN and symbols are the BSIM-CMG data.

including some that are not in the training set. Both $I_D V_G$ and $I_D V_D$ are accurate to higher-order derivatives especially for g_{ds} in the saturation region which is known to be difficult to model. Fig. 3 shows the NN modeling results of I_G for several structures and bias conditions. We can see that this modeling framework models all currents well by just using one network.

For QV NN, same inputs are fed in with three outputs Q_G , Q_S , and Q_D . In Fig. 4 & 5, we show the CV fitting accuracy

concluding data that are not in the training data set. The NN model can fit the capacitances well for varying geometries by using just one network.

To show the model's capability of variability modeling, we use Monte Carlo simulation to generate 2000 devices having certain variations (σ) in L , H_{FIN} , EOT and $\Delta\phi$ with BSIM-CMG. Then, we extract the I_{ON} of these devices and compare them with the prediction of the NN model. Fig. 6 shows the I_{ON} distributions relative to the mean value. We can see that, the mean (μ) and standard deviation (σ) of I_{ON} predicted by the NN model are in excellent agreement with those predicted by BSIM.

We also performed quality tests on the NN models. Fig. 7a shows the Gummel symmetry plot at 4th derivative is continuous and smooth. Because our model framework directly trains the NN using data from negative V_{DS} to positive V_{DS} , it can easily pass the Gummel test without applying smoothing functions like [4]. Therefore, this framework is applicable to devices that are inherently unsymmetric such as a MOSFET with different source and drain doping profiles. Fig. 7b shows the harmonic balance test result. The NN model produces the correct slope of each harmonic component. Finally, we show the transient simulation of a 17-stage ring oscillator and SRAM with the NN model in Fig. 8 & 9 by hard-coding the weights, biases, and matrix multiplications into Verilog-A. The result matches the BSIM-CMG perfectly with no convergence issue. We cannot compare the circuit simulation speeds for NN versus BSIM-CMG models at this time because Verilog-A has no efficient matrix calculation [4, 10]. However, we can compare the DC model evaluation speeds of NN and equation-based models using Python [10]. In Fig. 10, we show the speed of inferring a NN-IV versus calculating a simplified BSIM-CMG core IV model in Python. We code the core quasi-static IV calculation of BSIM-CMG in Python. We use NumPy and test them on Intel Xeon Platinum 8260 versus bias points. NN model holds about 13 times speed advantage. In the case of 10 million DC points, NN takes 59.6s while BSIM-CMG takes 806s. If we can further optimize the network and use hardware acceleration such as GPU, the NN speed advantage may be even more.

VI. CONCLUSION AND DISCUSSION

We present an NN-based model framework. It contains two parts: IV and QV network. For IV network, we use one NN to model I_{D} and I_{G} including leakage current. One QV NN is used to model all charges (Q_{G} , Q_{S} , Q_{D}). We demonstrate that the proposed model can accurately predict the variability of the device and give smooth and correct high order derivatives.

BSIM-CMG is used to generate data in this paper which has several benefits. Equation-based models such as BSIM-CMG can serve as a "noise filter" for the measured device data and training the NN with BSIM model data is an excellent approach for obtaining accurate high-order derivatives and variability capabilities as well as the charges which are difficult to get from measurements. The possibly faster inferring of NN would replace the evaluation of model equations. The benefit would be faster circuit simulations.

In our future work, we will keep completing the NN-based model by including parasitics, temperature effects, self-heating,

and so on.

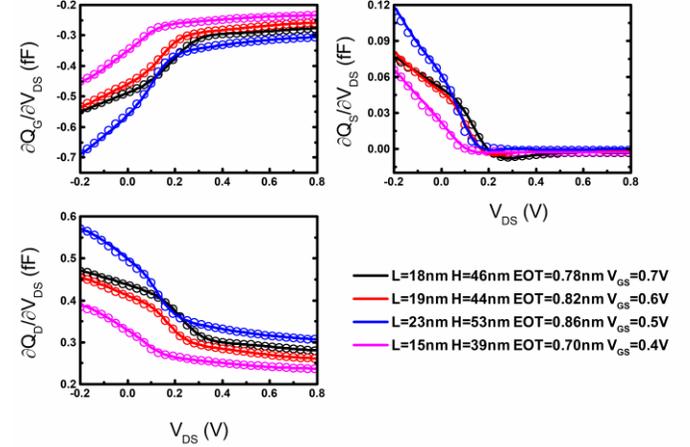


Fig. 4. CV fitting of Q_{G} , Q_{S} and Q_{D} versus V_{DS} for different L , H_{FIN} and EOT .

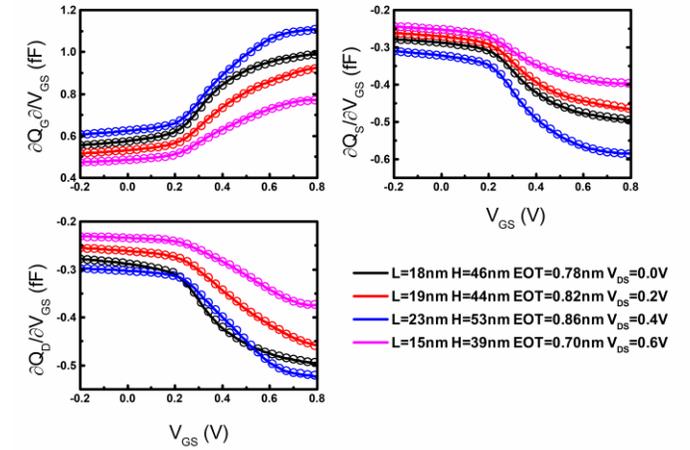


Fig. 5. CV fitting of Q_{G} , Q_{S} and Q_{D} versus V_{GS} for different L , H_{FIN} and EOT .

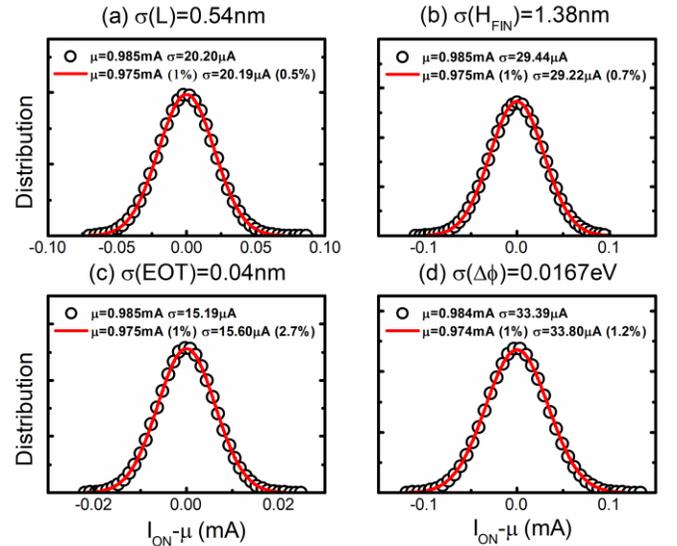


Fig. 6. I_{ON} distribution relative to mean by Monte Carlo simulation for (a) σ of $L = 0.54\text{nm}$, (b) σ of $H_{\text{FIN}} = 1.38\text{nm}$, (c) σ of $EOT = 0.04\text{nm}$, and (d) σ of $\Delta\phi = 0.0167\text{eV}$. The symbols are the data generated with BSIM-CMG and the lines are the predictions of the NN model. In the parentheses, we show the error rate of μ and σ between NN and BSIM-CMG.

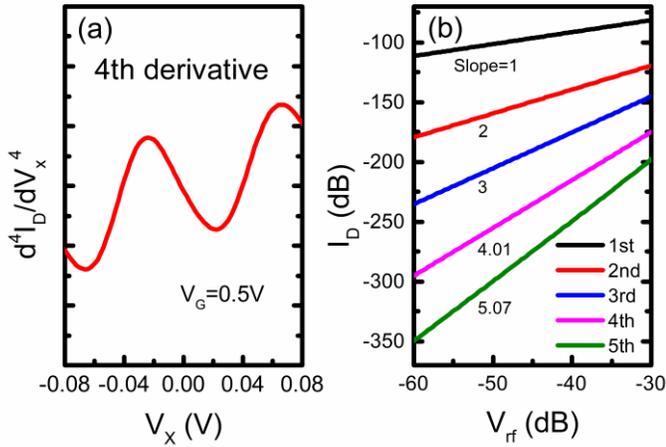


Fig. 7. (a) Gummel test at 4th derivative. (b) Harmonic balance test. The slope of each line meets the theoretical prediction.

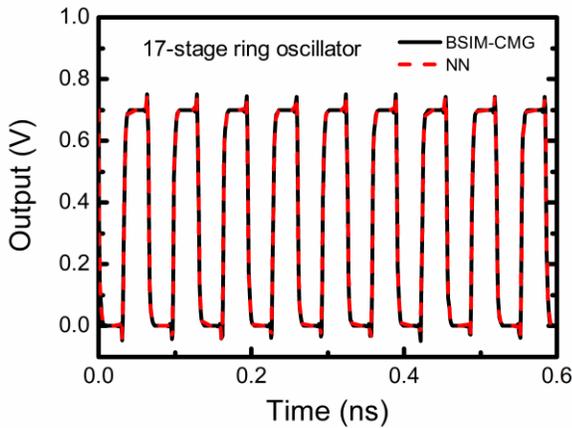


Fig. 8. The 17-stage ring oscillator simulations of BSIM-CMG and the NN model.

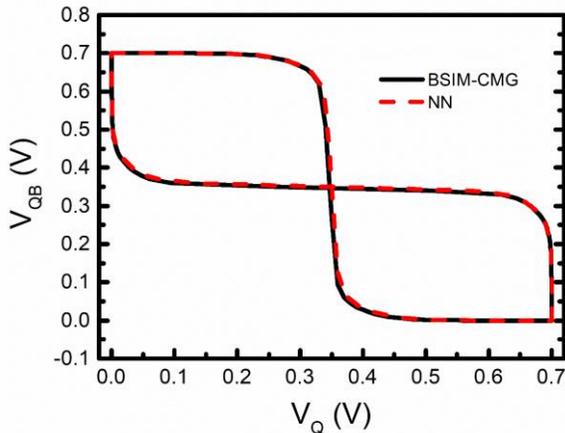


Fig. 9. The SRAM SNM (signal noise margin) simulations of BSIM-CMG and the NN model.

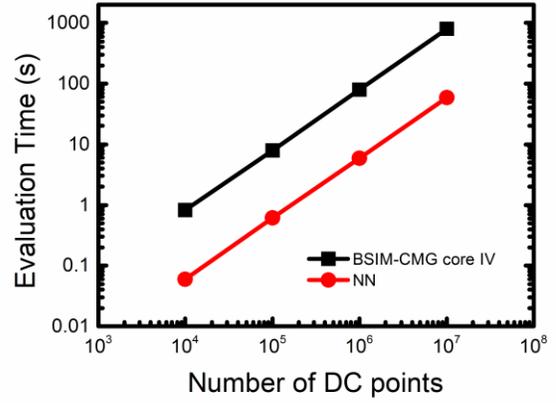


Fig. 10. Evaluation time comparison between NN and BSIM-CMG IV model.

REFERENCES

- [1] J. P. Duarte *et al.*, "BSIM-CMG: Standard FinFET compact model for advanced circuit design," in *ESSCIRC Conference 2015 - 41st European Solid-State Circuits Conference (ESSCIRC)*, 14-18 Sept. 2015 2015, pp. 196-201, doi: 10.1109/ESSCIRC.2015.7313862.
- [2] S. Khandelwal *et al.*, "BSIM-IMG: A Compact Model for Ultrathin-Body SOI MOSFETs With Back-Gate Control," *IEEE Transactions on Electron Devices*, vol. 59, no. 8, pp. 2019-2026, 2012, doi: 10.1109/TED.2012.2198065.
- [3] A. Dasgupta *et al.*, "BSIM Compact Model of Quantum Confinement in Advanced Nanosheet FETs," *IEEE Transactions on Electron Devices*, vol. 67, no. 2, pp. 730-737, 2020, doi: 10.1109/TED.2019.2960269.
- [4] 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 Transactions on Electron Devices*, vol. 68, no. 3, pp. 1318-1325, 2021, doi: 10.1109/TED.2020.3048918.
- [5] M. Li, O. İrsoy, C. Cardie, and H. G. Xing, "Physics-Inspired Neural Networks for Efficient Device Compact Modeling," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 2, pp. 44-49, 2016, doi: 10.1109/JXCDC.2016.2636161.
- [6] M. Y. Kao, H. Kam, and C. Hu, "Deep-Learning-Assisted Physics-Driven MOSFET Current-Voltage Modeling," *IEEE Electron Device Letters*, pp. 1-1, 2022, doi: 10.1109/LED.2022.3168243.
- [7] 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 Transactions on Electron Devices*, vol. 66, no. 10, pp. 4474-4477, 2019, doi: 10.1109/ted.2019.2937786.
- [8] H. Carrillo-Nunez, 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 Letters*, vol. 40, no. 9, pp. 1366-1369, 2019, doi: 10.1109/led.2019.2931839.
- [9] C. Akbar, Y. Li, and N. Thoti, "Device-Simulation-Based Machine Learning Technique for the Characteristic of Line Tunnel Field-Effect Transistors," *IEEE Access*, vol. 10, pp. 53098-53107, 2022, doi: 10.1109/access.2022.3174685.
- [10] C. T. Tung, M. Y. Kao, and C. Hu, "Neural Network-Based I – V and C – V Modeling With High Accuracy and Potential Model Speed," *IEEE Transactions on Electron Devices*, pp. 1-4, 2022, doi: 10.1109/TED.2022.3208514.
- [11] C. Auth *et al.*, "A 10nm high performance and low-power CMOS technology featuring 3rd generation FinFET transistors, Self-Aligned Quad Patterning, contact over active gate and cobalt local interconnects," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017 2017, pp. 29.1.1-29.1.4, doi: 10.1109/IEDM.2017.8268472.