Computers and Electrical Engineering 105 (2023) 108554

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Application of long short-term memory modeling technique to predict process variation effects of stacked gate-all-around Si nanosheet complementary-field effect transistors

Rajat Butola [a,b], Yiming Li [a,b,c,d,e,f,*], Sekhar Reddy Kola [a,b], Chandni Akbar [a,c], Min-Hui Chuang [a,c]

[a] Parallel and Scientific Computing Laboratory, National Yang-Ming Chiao Tung University, Hsinchu, 300093, Taiwan
[b] EECS International Graduate Program, National Yang-Ming Chiao Tung University, Hsinchu, 300093, Taiwan
[c] Institute of Communications Engineering, National Yang-Ming Chiao Tung University, Hsinchu, 300093, Taiwan
[d] Institute of Biomedical Engineering, National Yang-Ming Chiao Tung University, Hsinchu, 300093, Taiwan
[e] Department of Electronics and Electrical Engineering, National Yang-Ming Chiao Tung University, Hsinchu, 300093, Taiwan
[f] Center for mmWave Smart Radar System and Technologies, National Yang-Ming Chiao Tung University, Hsinchu, 300093, Taiwan

ARTICLE INFO

Keywords:
Complementary-field effect transistors
Statistical device simulation
Long short-term memory
Machine learning
Multi-layer perceptrons
Process variation effects

ABSTRACT

Emerging machine-learning (ML) methodology has been overcoming the challenging task of analyzing the process variation effect of nanoscale devices using 3-D stochastic device simulation. In this study, the effects of process variations on the electrical characteristics of the stacked gate all around (GAA) silicon nanosheet (NS) complementary field effect transistors (CFETs) are predicted using long short-term memory (LSTM) based ML model. The LSTM algorithm is less time-consuming and computationally effective as compared to conventional device modeling tools. For this work, a two-channel stacked CFET, formed by folding n-FET on top of p-FET, is considered. We utilize six important process parameters as input features to the LSTM model and estimate the variations in key electrical characteristics, such as threshold voltage (V_{TH}), off-current (I_{OFF}), on-current (I_{ON}), subthreshold slope (SS), and drain induced barrier lowering (DIBL). To avoid overfitting in the training phase, an early stopping regularization technique is applied to construct a high-performance LSTM model. In addition, to compare the predictive performance of our proposed ML-based LSTM model, the baseline MLP (multi-layer perceptrons) model is implemented. The results of this study show that the LSTM model with an r^2-score $>$ 95% outperforms the baseline model which has r^2-score $<$ 75%. Furthermore, the estimated rmse of LSTM in an order of 10^{-7} is 10 times lower than that of the baseline model whose rmse is in an order of 10^{-6}.

1. Introduction

The evolution of semiconductor technology nodes to achieve Moore's law brings device scaling to the nanoscale regime. Beyond 22 nm nodes, the planar MOSFETs suffered from excessive leakage current due to short channel effects (SCEs) [1]. As the conventional planar devices reached their limits, 3D structures, such as FinFETs, gate-all-around (GAA) nanowire (NW) FETs, and GAA nanosheet

* Corresponding author.
E-mail address: ymli@nycu.edu.tw (Y. Li).

https://doi.org/10.1016/j.compeleceng.2022.108554
Received 30 May 2022; Received in revised form 12 December 2022; Accepted 14 December 2022
Available online 24 December 2022

(NS) FETs have been explored as possible alternatives [2,3,4,5]. These are the multigate devices that provide more control on the channel and reduce the leakage current. In FinFETs, the channel is surrounded by the gate from the three sides whereas in GAA NW and NS FETs the gate completely covered the channel region. The GAA NW and NS FETs provide additional control on the gate makes them more suitable for low-power applications. However, beyond the 3-nm technological nodes, the FinFET and GAA FETs also encounter the similar problems as planar transistors did. Therefore, in continuation of this advancement, the complementary field effect transistor (CFET) has emerged as a further evolution of nanodevice structure for sub-3-nm nodes. It outperforms the other devices and meets 3 nm node requirements for performance, power, area, and cost (PPAC). With CFET, the Moore's Law can continue to fuel its life beyond sub-3-nm nodes [6,7]. The structure of CFET is a more complex form of GAA, where the p-FET is folded on top of n-FET and controlled by a common gate. This structure eliminates the bottleneck of n-to-p separation and consequently offers a potential area scaling in a design technology co-optimization [8]. To construct CFET the two obvious GAA choices were NS and NW. The NW transistor has a limitation of lower drive current as compared to NS one due to narrower effective channel widths [9,10]. On the other hand, NS transistor provides a higher I_{ON} and superior sub-threshold region characteristics for similar fabrication process [11]. The GAA NS FET also has the advantage that they can provide more current per layout footprint because of their larger effective widths and also deliver improved electrostatics control [12]. Furthermore, the CFET structure with the GAA NS FETs provides excellent electrostatic integrity. Apart from this, due to design flexibility, the stacked GAA NS structure also provides more device optimization options. However, there are several challenges associated with such highly scaled devices that affect the device performance.

The nano-devices are suffered from many problems that are sensitive to process variations and become increasingly challenging to sustain [13,14]. A slight variation can affect the characteristics and reliability of nano-devices. This can be a barrier to further scaling down the technology nodes. There are many variations that have a significant effect on device performance. Among them, the process variation effect (PVE) is one of the challenges that affect device-to-device performance variations. The PVE has always been a critical aspect of semiconductor fabrication. In past, the analysis of these variations is performed using conventional device simulators such as technology computer-aided design (TCAD) [15,16,17]. To investigate the effects of PV, a large number of devices must be simulated. These simulators are highly accurate, but they take much time to predict the effect of these variations. For the rapid growth of technology, however, device modeling is required as soon as possible. Therefore, an alternative to conventional device modeling is needed which can be fast and accurate. Recently, ML is successfully used as an alternative for device modeling [18,19,20]. ML is a data-driven technique that develops relationships between inputs and outputs and then predicts outputs for unseen data. The ML techniques are computationally less expensive and relatively very fast. Many ML techniques have been used in past for predicting the effects of different variations on the electrical properties of different nanodevices. In [21], an analytical method based on an ANN-based ML algorithm is proposed to assess the degree of process variability in ultra-scaled devices. Similarly, a TCAD-ML framework is proposed to assist the analysis of device-to-device variation and operating temperature without the need for physical quantities extraction [22]. The authors used principal component analysis followed by a third-order polynomial regression ML algorithm. Their framework showed great promise to accelerate the development of new device technologies. In [23], the authors investigated the process variation effect for 3D NAND flash memory cells using an ANN-based ML approach. Their model is proved to be a good fit for TCAD simulation with an error rate of less than 5%.

This is, however, the first attempt to the best of our knowledge, where ML is used to predict the effect of PV for CFET devices. Here, we proposed a reliable, generic, and data-driven LSTM-based ML model to predict the variations in the device characteristics and extracted key electrical properties such as V_{TH}, I_{OFF}, I_{ON}, SS, and DIBL accurately. We focus on the six prominent process variation factors that vary the device characteristics. These consists of gate length (L_G), NS width (W_{NS}), NS height (H_{NS}), oxide thickness (T_{OX}), and the spacing between n-FET and p-FET (D_{SEP}). Since we use a stacked NS CFET structure with 2 channels, therefore, we include one more important parameter which is the spacing between these two channels (L_{SEP}). These six process parameters are used as input features to the LSTM model and the corresponding change in IV characteristics is estimated as output. The LSTM is a powerful network used to handle sequence dependence. A long sequence of drain current (I_D) is predicted using the LSTM networks because they can maintain the state (memory) across very long sequences. The LSTM model accurately captured the relationship between the process parameters and electrical characteristics of stacked GAA Si NS CFET. The results of the LSTM model are compared with the baseline model for which we use the MLP model that can handle the nonlinear effects effectively. The results demonstrate that the proposed LSTM model clearly outperforms the MLP model and shows more accurate results.

1.1. Contribution

In this work, we investigate the effects of PV and predict the variations of transfer characteristics of the GAA Si NS CFET by using the LSTM model.

- The aim of this work is to develop an ML-based model as an alternative to conventional device modeling and provide a great opportunity to speed-up technology development.
- Once the ML model is trained rigorously, it can be used to extract the important figure of merits, such as threshold voltage, on current, off current, etc. for further analyses of device characteristics.

1.2. Organization of paper

The rest of the paper is organized as follows. Section 2 presents the structure of CFET device in detail with simulation methodology and process parameters. Section 3 demonstrates the LSTM and the baseline model. The data collection and data pre-processing

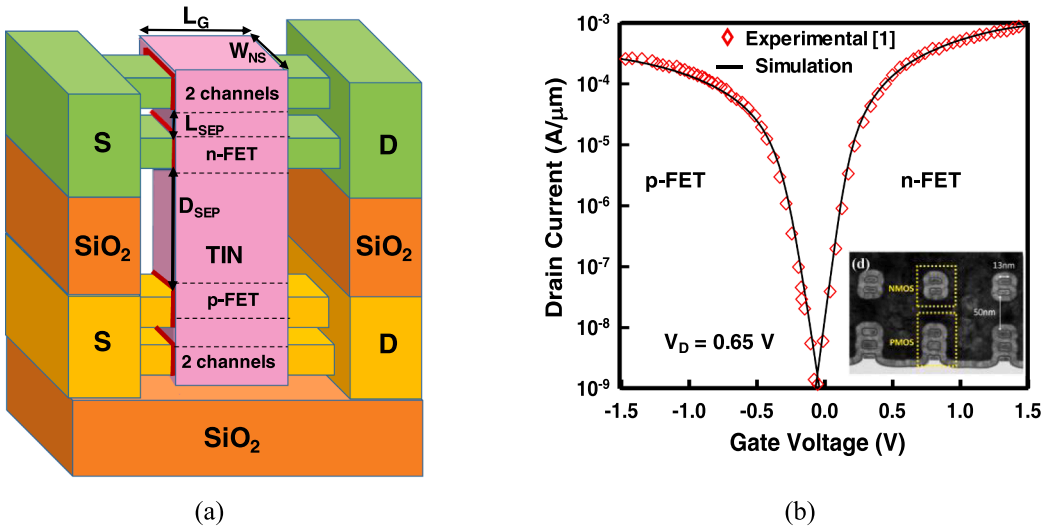

Fig. 1. (a) A 3D schematic illustration of the vertically stacked GAA Si NS CFET. (b) Plot of the experimental calibration curves, where the solid line is simulation and the symbol is measured data from [1].

Table I
Stacked GAA Si NS CFET Device Parameters for the sub-3-nm Technology Node.

Sr. No.	Parameters	Values
1	Gate Length (L_G) (nm)	20
2	Channel Doping (cm^{-3})	5×10^{17}
3	Nanosheet Width (W_{NS})(nm)	25
4	Nanosheet Height (H_{NS}) (nm)	5
5	Spacing between nanosheet (L_{SEP}) (nm)	10
6	Spacing between N- and P-FET (D_{SEP}) (nm)	20
7	S_{ext}/D_{ext} Length (nm)	5
8	S/D Doping (cm^{-3})	1×10^{20}
9	S_{ext}/D_{ext} Doping (cm^{-3})	4.8×10^{18}
10	Number of Channels	2
11	EOT (nm)	0.6

technique is explained in Section 4. In Section 5, results obtained for the LSTM and baseline models are discussed in detail and finally, Section 6 concludes the outcomes of this work.

2. Device structure and process parameters

The 3-nm node GAA Si NS CFET and standard CMOS with separate N-/P-FETs are designed based on the IRDS 2021 high-performance specifications [24]. Fig. 1(a) depicts a 3D schematics of two-channel vertically stacked GAA Si NS CFET, with a gate length (L_G) of 20 nm and NS width, W_{NS}, of 25 nm. The nominal NS device has a height, H_{NS}, of 5 nm, the distance between adjacent channels, L_{SEP}, is 10 nm, the distance between n-and p-devices, D_{SEP}, is 20 nm, and the source and drain extensions (S_{ext} and D_{ext}) are 5 nm. In addition, the doping concentration of the channel is 5×10^{17} cm^{-3} and the source/drain is 10^{20} cm^{-3}. In the explored device, the gate stack consists of $SiO_2/HfO_2/TiN$ with an effective oxide thickness (EOT), T_{ox}, of 0.6 nm [25]. As stated in Table I, the parameters of the nominal device with respect to process parameters and doping concentration are listed. The device simulation was calibrated using Intel® measurement data for CFETs to achieve the highest level of precision, as depicted in Fig. 1(b) [1]. In the validated calibration, the W_{NS}/H_{NS} of the channel, the S/D including S_{ext}/D_{ext}, channel doping concentrations, the work function of the metal gate, T_{ox}, and the mobility models are adjusted to obtain the best accuracy of the device simulation. In particular, the device simulation is achieved by solving the 3D quantum-mechanically corrected transport model validated by the nonequilibrium Green's function approach and tuning the electron effective mass to take into account physically reasonable parameters within the transport model [26,27]. Similarly, CFET PVEs device simulations are carried out using similar calibration methodologies, models, and doping levels. We randomly vary six important parameters, L_G, T_{ox}, H_{NS}, W_{NS}, L_{SEP}, and D_{SEP} as part of the PVE simulations. Moreover, the PVE method is demonstrated in [28]; and, 1000 samples are randomly generated by varying different process variations parameters. Not shown here, we have also statistically generated the process variations for characteristic fluctuations by using our own simulator. The varied parameters and their mean and standard deviation parameters are: L_G is 20 nm and $\pm$ 2 nm, L_{SEP} is 10 nm and $\pm$ 1.0 nm, W_{NS} is 25 nm and $\pm$ 2.5 nm, H_{NS} is 5 nm and $\pm$ 0.5 nm, S_{ext}/D_{ext} is 5 nm and $\pm$ 0.5 nm, and T_{ox} is 0.6 nm and $\pm$ 0.06 nm. By varying these

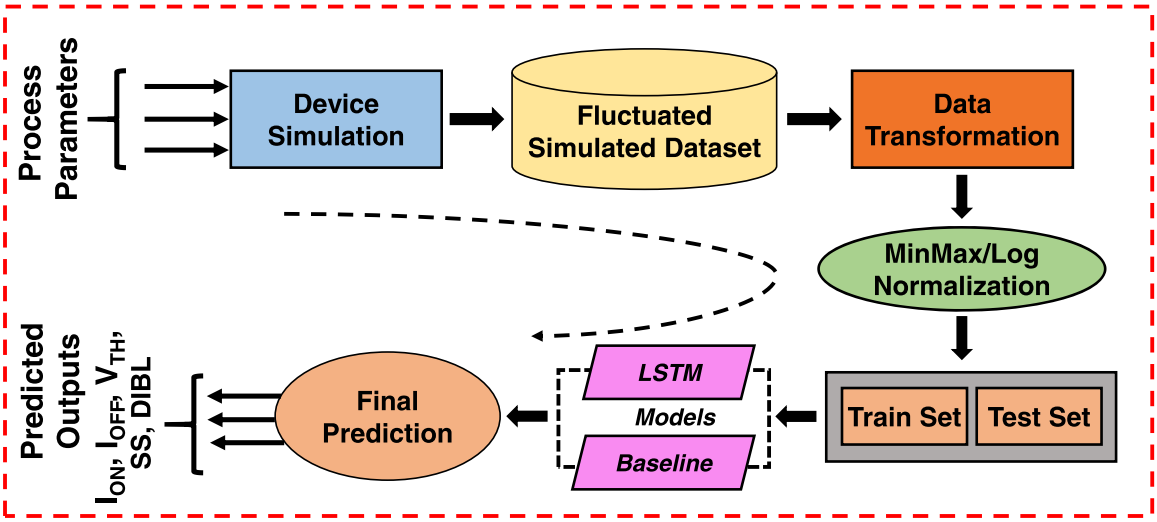

Fig. 2. A procedure for the ML-based device modeling.

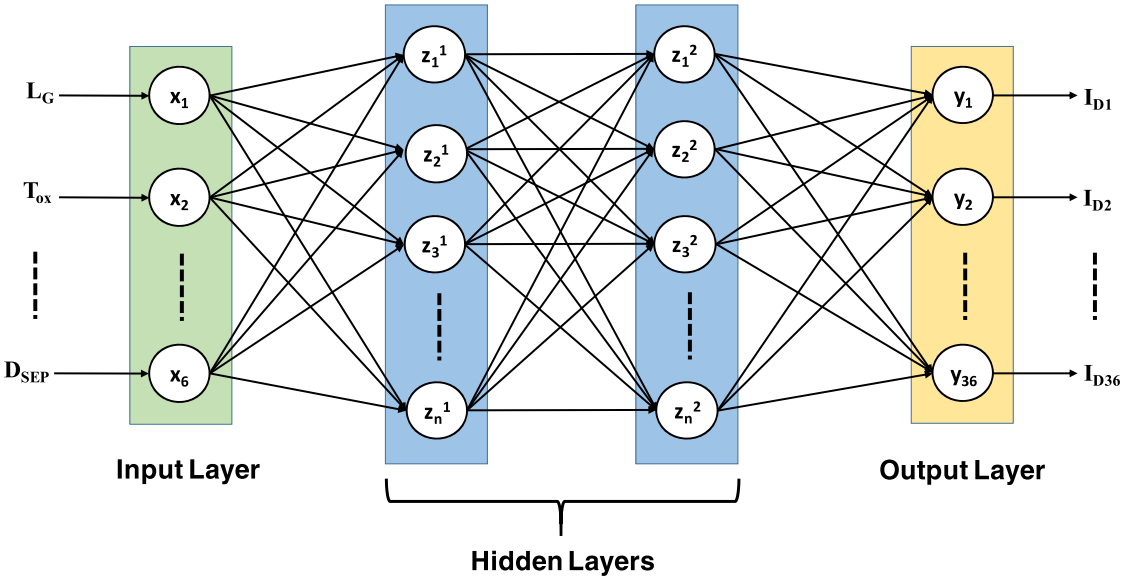

Fig. 3. An illustration of the MLP model structure.

parameters, we simulated 1000 devices that have fluctuated characteristics. The V_{TH} is extracted from I_D-V_G curves at $V_D = 0.05$ and 0.7 V, where the V_{TH} is estimated by the constant current method via the current criterion 10^{-7} A.

3. Machine learning modeling

Semiconductor technology and ML are the two fastest-growing areas in today's life. Therefore, ML presents new opportunities with a wide range of applications in semiconductor manufacturing. This encouraged us to exploit the capability of ML in modeling the effects of PV in CFET devices. ML has the ability to learn the hidden relationships in large datasets. We illustrate here if the right ML model is selected and its hyperparameters are tuned and optimized precisely, the effects of PV can be estimated with high accuracy and in less time as compared to physical device modeling. The ML models explored in this work include the MLP model which is used as the baseline model and the LSTM model. The procedure of the ML-based approach is shown in Fig. 2.

3.1. Baseline model

The MLP is selected as the baseline model as it provides a lot of flexibility and has proven itself to be useful and reliable over

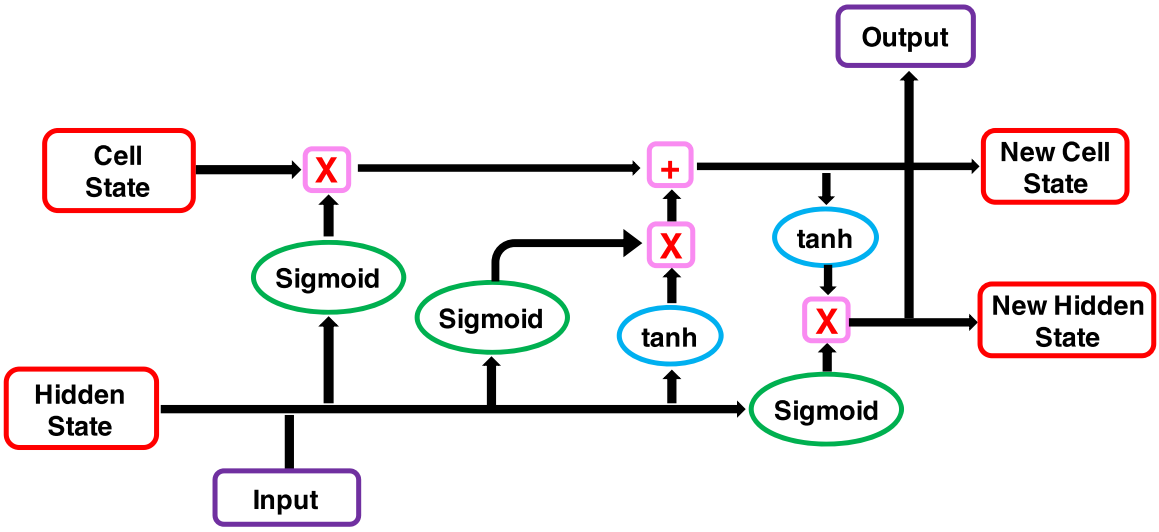

Fig. 4. An illustration of the implemented architecture of LSTM network.

decades for regression problems. Without a "reference model", it would be difficult to determine whether the proposed LSTM model achieves its purpose satisfactorily or not. The baseline model serves as a "reference model" in ML tasks. The MLP is a classical type of fully connected feed-forward neural network (NN). It is used to investigate how an artificial biological brain-type network can be used to solve complex tasks such as predictive modeling. MLP is used to develop robust ML techniques that can model difficult problems. It has the ability to learn the hidden relationship in the training data and mapped it into the output data that is to be predicted. Therefore, MLP is suitable for regression analysis where the task is to estimate the effect of some explanatory variable on the dependent variable. It is flexible and generally used to learn a mapping from inputs to outputs. The predictive capability of MLP comes from its multi-layer architecture. The inputs and outputs are connected with each other through one or more dense hidden layers. Each layer consists of a number of neurons, which are simple computational units with weighted input signals and activation functions. The activation function also called a transfer function, that governs the threshold at which a neuron is activated. For our work, we use the non-linear hyperbolic tangent (tanh) as an activation function. The MLP network has three types of layers: input layer, hidden layer, and output layer. The input layer takes the inputs from the dataset. The hidden layer followed the input layer and there can be more than one hidden layer in an MLP structure. These hidden layers operate on the bases of the activation function. The final layer is called the output layer and it produces the outputs. The architecture of the baseline MLP model used in this work is shown in Fig. 3. The equation holding the input and output relationship is straightforward as:

$$y = a\left(\sum x \times w + b\right), \tag{1}$$

where x is the input, w is the weight, b is the bias, and a is the activation function of the output y.

3.2. Long short-term memory (LSTM)

The LSTM is a type of recurrent neural network (RNN) that is trained using a backpropagation algorithm and overcomes the vanishing gradient problem. It is used to address complex sequence problems in ML and achieve state-of-the-art outcomes. In contrast to the baseline MLP model which has neurons, the LSTM model consists of memory blocks also called memory cells that are connected through layers [29]. The cell is more advanced than a neuron and has the memory to remember recent sequences. The cell contains different gates that control the cell state and output. The cell operates upon a sequence of inputs and the gates inside the cell use sigmoid as an activation function to manage the triggering for making a change of state [30].

The architecture of the LSTM network is shown in Fig. 4. It consists of three types of gates: input gate, output gate, and forget gate. The input gate selects the values from the input to update the memory state, the output gate decides what output to produce based on the input and the memory of the cell, and finally, the forget gate decides whether the information to retain or throw away from the cell. Therefore, the LSTM has the ability to memorize the input and output values for a short time interval. Generally, sigmoid and tanh activation functions are utilized in the LSTM model. The range of sigmoid and tanh is 0 to 1 and -1 to $+1$, respectively. The mathematical notation of forget gate is given below:

$$f_t = \sigma\left(x_t \times w_f^x + h_{t-1} \times w_f^x + b_f\right), \tag{2}$$

where σ is the activation function that behaves as a forget gate to remove the state value if the input signal is close to 0 and retain the state if closer to 1. Similarly, the mathematical notation of the input gate is given as:

Table II
Hyperparameters for LSTM and Baseline Model.

Model	Parameters	Value
LSTM Model	LSTM Cells	15
	Return Sequence	True
	Dense Layer	1
	Number of Neurons	10
	Activation Function	Tanh & Sigmoid
	Epochs	2000
	Optimizer	Adam
	Learning Rate	0.001
	Input Dimension	(1000, 1, 6)
	Output Dimension	(1000, 36)
	Loss Function	MSE
Baseline Model (MLP)	Hidden layers	2
	Number of Neurons	(10,10)
	Activation Function	Tanh
	Epochs	2000
	Learning Rate	0.001
	Loss Function	MSE
	Input Dimension	(1000, 6)
	Output Dimension	(1000, 36)
	Optimizer	Adam

$$i_t = \sigma\left(x_t \times w_i^x + h_{t-1} \times w_i^x + b_i\right), \tag{3}$$

where $\boldsymbol{x_t}$, $\boldsymbol{w_i^x}$, $\boldsymbol{h_{t-1}}$ and $\boldsymbol{b_i}$ are the input, weight, previous hidden state, and bias of the input gate, respectively. Furthermore, the cell state is intrigued by the tanh activation function. The mathematical notation of the cell state or the memory of the LSTM model is given as:

$$i_t = f_t \times c_{t-1} + i_t \times tanh\left(x_t \times w_c^x + h_{t-1} \times w_c^x + b_c\right). \tag{4}$$

Lastly, the hidden state is the repository of the previous input and the updated state is formulated as given below:

$$o_t = \sigma\left(x_t \times w_o^x + h_{t-1} \times w_o^x + b_o\right), \tag{5}$$

and

$$h_t = o_t \times tanh(c_t). \tag{6}$$

In this work, we are building the LSTM model with 15 LSTM cells connected recurrently with each other. After flattening the output from the LSTM model, the output sequence, as well as the hidden sequence, are fed into the dense layer. The dense layer having 10 neurons optimized the LSTM model to avoid the vanishing gradient problem. The list of hyperparameters is summarized in Table II. These hyperparameters are tuned in such a way that the overall loss function optimizes its value and high prediction accuracy can be achieved. Due to its recurrence property, the sequence of electrical characteristics is analyzed according to the position of I_D with its corresponding values.

4. Data collection and pre-processing

The ML models are data-driven models, which means they learn the relationship between the input and output itself during the training phase. Therefore, for training and validation of the model, we need the required dataset. To collect the dataset for this work, we performed the 3D device simulation for the GAA Si NS CFET by varying the process parameters for linear and saturation regimes. A total of 1000 samples of fluctuated devices are generated and their transfer characteristics (I_D-V_G) are collected and used to train and test the ML models. The simulated data is referred to as raw data since it needs to be transformed and normalized before feeding into the ML model. The raw data has inconsistent formating and a high range of inputs and outputs which makes the dataset more complicated and challenging for the ML model and hence resulting in poor performance. Data pre-processing is required to resolve these issues before feeding the data into the ML model. The process parameters that are used as input features for our problem are first normalized using the min-max normalization to confine them into the range between 0 and 1. Similarly, the I_D values are the predicted outputs of the ML model. The range of I_D is very small at the subthreshold voltages as compared to those which are over the threshold voltage. Therefore, log normalization is used to scale the I_D values to bring them into the same range and resolve the unbalanced distribution. Once the pre-processing is done the data is ready for the ML model.

5. Results and discussion

After the pre-processing of data, it is split into the training (80%), testing (10%), and validation (10%) sets. The training dataset is

Table III
Comparison of r^2-score, rmse, Training Time and Epochs for LSTM and Baseline Model.

Model	Type	Biasing	r^2_score train data	r^2_score test data	rmse train data	rmse test data	Time (*sec*)	Early stopping epoch
LSTM	N-Type	Linear	0.9942	0.9735	1.4914e-07	1.5938e-07	410	585
		Saturation	0.9976	0.9869	4.0467e-07	6.3596e-07	574	816
	P-Type	Linear	0.9827	0.9558	1.0565e-07	1.2782e-07	346	384
		Saturation	0.9926	0.9531	2.8799e-07	3.7902e-07	214	232
Baseline	N-Type	Linear	0.9923	0.7038	5.4362e-07	8.7063e-06	63	89
		Saturation	0.9882	0.7187	5.9878e-07	2.3332e-06	98	103
	P-Type	Linear	0.9854	0.6893	2.4265e-07	6.5207e-06	783	947
		Saturation	0.9582	0.7435	7.2470e-07	7.7101e-06	564	678

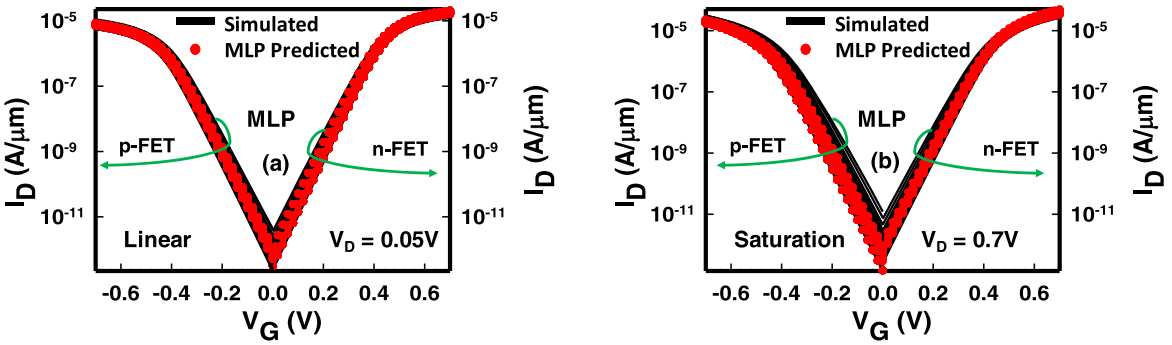

Fig. 5. Prediction of I_D-V_G curves using the MLP model for the (a) linear (b) saturation regions.

used to build up the ML algorithm. The model evaluates the data repeatedly to learn the hidden relationship and then adjusts the parameters accordingly to provide accurate results. The validation dataset, on the other hand, is a sample of data that is kept aside from the training of the model. The model is frequently evaluated on this dataset during training time to give an unbiased estimation of model fit on the training data while tuning the model hyperparameters. Finally, the test dataset is the unseen data that is used for the final evaluation of the model fit. The input and target values of the LSTM and baseline MLP model are process parameters and I_D-V_G curves, respectively. Each I_D-V_G curve is discretized into 36 I_D values corresponding to 36 gate voltages V_G from 0 to 0.7 V with a step size of 0.02 V for n-FET, and V_G from −0.7 to 0 V with a step size of −0.02 V for p-FET, at a specific biasing condition (V_D = 0.05 V and 0.7 V).

5.1. Performance evaluation metrics

Two different ML algorithms, MLP and LSTM, have been implemented in python Tennsorflow 2.0 with Keras and Scikit-Learn library. The performance of models is evaluated using two evaluation metrics: the root mean squared error (rmse) and r^2-score. In the regression problem, rmse is expressed as the square root of the sum of the square of the difference between the predicted values and the simulated values divided by the total number of data points. The rmse is calculated by

$$rmse = \sqrt{\frac{\sum_{i=0}^{n}\left(y_{sim,i} - y_{pred,i}\right)^2}{n}}, \tag{7}$$

where $y_{sim,i}$, and $y_{pred,i}$ are the drain current obtained through the simulation and prediction of the MLP/LSTM model, respectively. The r^2-score, on the other hand, represents the goodness of the fitting of the explored ML models and is calculated as an evaluation technique.

$$r^2 = 1 - \frac{\sum_i \left(y_{sim,i} - y_{pred,i}\right)^2}{\sum_i \left(y_{sim,i} - y_{mean}\right)^2} \tag{8}$$

where y_{mean} is the average value of the simulated output.

5.2. Dealing with overfitting: early stopping

With ML techniques overfitting is a very common and serious issue. The ML model with more input parameters is generally more prone to overfitting. Overfitting occurs when the model is trained excessively on training data. This happened when the number of epoch is too high. This makes the model overfit on training data to such an extent that it lost its generalization capability. It means the

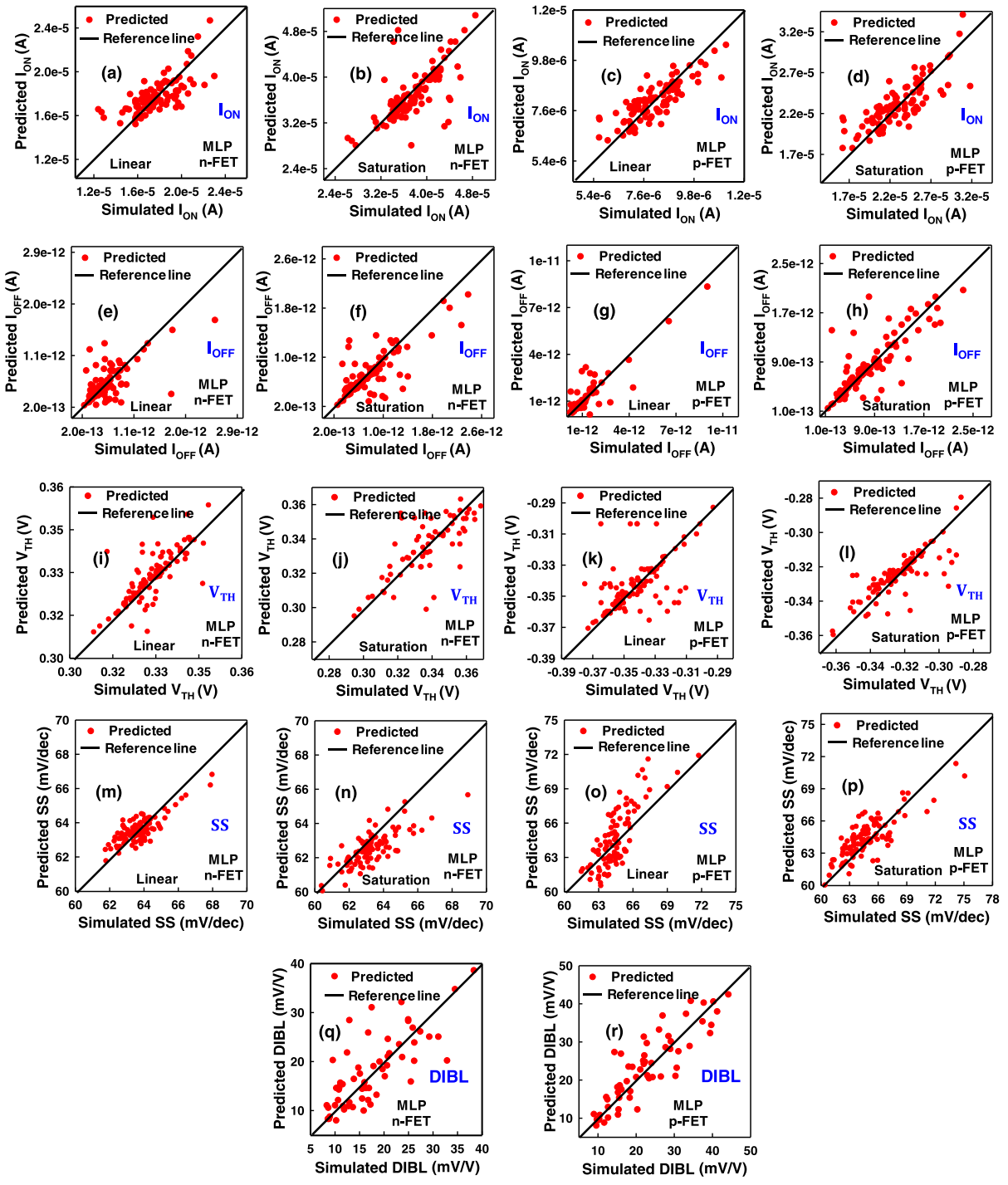

Fig. 6. Scatter plots for (a)-(d) I_{ON}, (e)-(h) I_{OFF}, (i)-(l) V_{TH}, (m)-(p) SS, and (q)-(r) DIBL extracted from the MLP predicted I_D-V_G curves for the n-FET and p-FET at the linear and saturation regions.

model will underperform on unseen test data. To avoid overfitting regularization is used and early stopping is one such technique. We use early stopping from the Keras library. In early stopping, a criterion is set in terms of the "Patience" parameter. This is the number of epochs without improvement in which the training will be stopped. We set the patience value as 50 and validation loss is set as the criteria. So the training will be stopped if there will be no improvement in the validation loss for 50 epochs in succession. Early stopping not only avoids overfitting but also saves training time and makes the overall model faster and more robust.

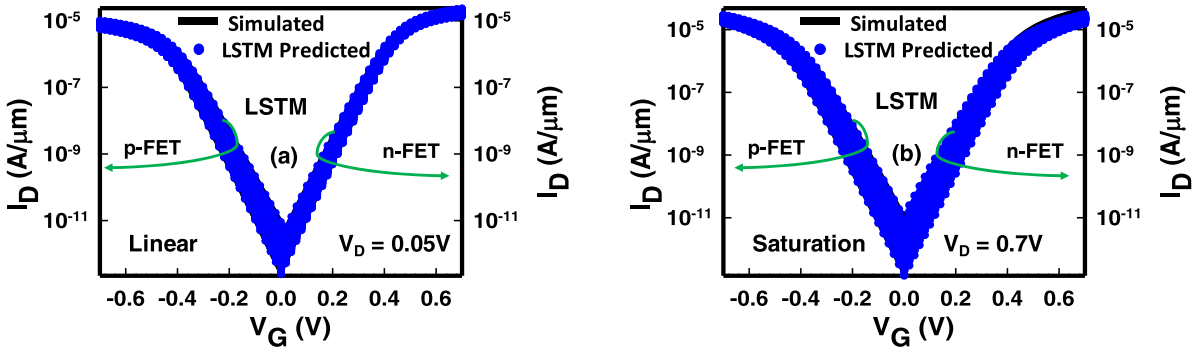

Fig. 7. Prediction of I_D-V_G curves using the LSTM model for the (a) linear (b) saturation regions.

5.3. Performance of baseline model

Firstly, the baseline MLP model is trained to understand the conventional relation of CFET structure and its process parameters. The MLP model is constructed using two hidden layers with 10 neurons in each. The tanh is used as the activation function and the Adam algorithm is used as an optimizer. The model is primarily trained for 2000 epoch cycles with a batch size of 5. The model training is stopped way earlier than the 2000 epochs because of the early stopping regularization technique. The early stopped epochs for different models are listed in Table III. After the successful training, the MLP model is exposed to the unseen test dataset, and prediction capability is evaluated. The test I_D-V_G curves predicted by the MLP model are compared with the simulated curves and the result is shown in Fig. 5. It can be observed that the MLP model is unable to predict the I_D-V_G curves of GAA SI NS CFET accurately. Furthermore, the rmse and r^2-score of the model for the n-FET and p-FET predicted values are summarized in Table III. The training r^2-scores of the baseline model for the n-FET at linear and saturation regions are 0.9923 and 0.9882, respectively and similarly for the p-FET is 0.9854 and 0.9582, respectively. The test r^2-scores of the baseline model for the n-FET at linear and saturation regions are 0.7038 and 0.7187, respectively. Similarly, for the p-FET it is 0.6893 and 0.7435, respectively. The high r^2-score for training and low r^2-score for test dataset show that the MLP model cannot perform well for the unseen dataset. This is the condition of high variance and it occurred either due to overfitting or due to fewer training samples. Since the overfitting is already taken care of by using the regularization technique, therefore, it can be inferred that the MLP model requires more training data for better prediction. With the limited dataset, the MLP performance is highly unsatisfactory. The same can be observed from the high rmse values for test data. We further extract the I_{ON}, I_{OFF}, V_{TH}, SS, and DIBL from the predicted dataset and compare them with the simulated values. The scatter plots between the simulated and predicted values are represented in Fig. 6. The *x*-axis depicts the simulated values and *y*-axis depicts the predicted values. The red dots in the figure represent the output values predicted from the baseline model and the black line shows the reference line. The reference line here indicates a line where the predicted values will be exactly equal to the simulated values. This is the ideal situation when the machine learning model will predict the output values with 100% accuracy. Therefore, the scatter plots show how close the predicted values are to the simulated values. If the values are near the reference lines, then the model predicts the output values with good accuracy and if the values are far away from the reference line, then the prediction accuracy of the machine learning model is not satisfactory. Therefore, the reference line is also called the ideal line. Here, plots 6 (a)-(d) show linear and saturation scatter plots of I_{ON} for the n-FET and p-FET devices. Here, very few values on the reference lines illustrate that the baseline model could not estimate the I_{ON} values accurately. A similar conclusion can be drawn from Figs. 6(e)-(h), Figs. 6(i)-(l), Figs. 6(m)-(p), and Figs. (q)-(r) for I_{OFF}, V_{TH}, SS, and DIBL, respectively. The predicted values are clearly dispersing instead of concentrating on the reference line. We can also deduce from Fig. 6 that due to PVE the values of these extracted parameters are varying. The values of V_{TH}, SS, and DIBL are varying between 0.03 to 0.1 V, 6 to 12 mV/ dec., and 30 to 40 (mV/V) respectively for the n-FET and p-FET under different biases. From these results, it is evident that for the complex CFET device structure, the baseline MLP model has very poor performance.

5.4. Performance of LSTM model

Therefore, it is necessary to use a more advanced ML model that can learn the complex relationship between process parameters and transfer characteristics. Hence, the LSTM model has been implemented to understand the sequence of electrical characteristics of the n-FET and p-FET devices. The LSTM model has the ability to memorize the past value in the sequence. Additionally, it can handle a sequence of data and its predicted output depends on the input as well as the previous output. Therefore, the LSTM model is created with a layered array that contains a sequence input layer, one LSTM layer, and an output layer that produces the predicted outputs. The size of the sequence input layer is fixed and equal to the number of input features in the training data. The size of the output layer is fixed according to the number of outputs to be predicted at a time. To prevent the gradient from exploding the gradient threshold is set to 1. The learning rate, a hyperparameter that governs the pace at which the model updates and learns the parameters, is set to 0.001. The LSTM model is trained using the same number of epochs, i.e., 2000 cycles, and the batch size of 5 samples. However, the model training is stopped early because of the regularization technique. The transfer characteristics of the n-FET and p-FET for the LSTM

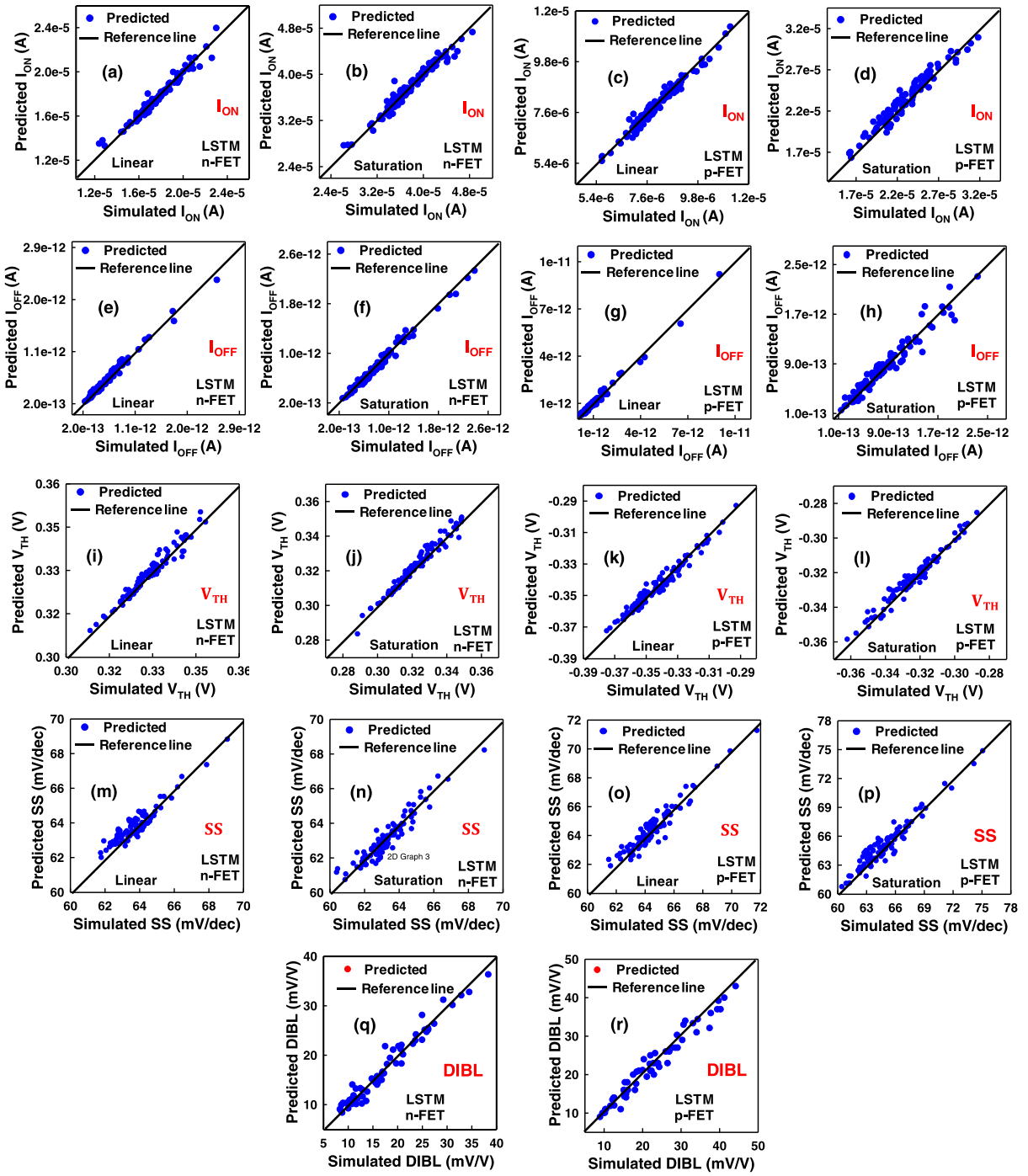

Fig. 8. Scatter plots for (a)-(d) I_{ON}, (e)-(h) I_{OFF}, (i)-(l) V_{TH}, (m)-(p) SS, and (q)-(r) DIBL extracted from the LSTM predicted I_D-V_G curves for the n-FET and p-FET under the linear and saturation regions, respectively.

prediction and simulation are shown in Fig. 7. The training r^2-scores of the LSTM model for n-FET at linear and saturation regions are 0.9942 and 0.9976, respectively and similarly for the p-FET is 0.9827 and 0.9926, respectively. The test r^2-scores of the baseline model for the n-FET at linear and saturation regions are 0.9735 and 0.9869, respectively. Similarly, for the p-FET, it is 0.9558 and 0.9531, respectively. The performance metrics summarized in Table III show that the proposed LSTM model outperforms the baseline MLP model. Therefore, the proposed LSTM model has a better prediction performance than the baseline model.

In order to further examine the performance of the LSTM model, the key electrical parameters I_{ON}, I_{OFF}, V_{TH}, SS, and DIBL are

extracted from the predicted values and a comparison is established against the simulation values. These key electrical parameters are important indicators for device performance. The scatter plots in Fig. 8 show the linear relationship between the LSTM predicted values (blue dots) and simulated values of I_{ON}, I_{OFF}, V_{TH}, SS, and DIBL. The training and testing performance matrices suggest that the proposed model is able to capture the underlying hidden relationship between process parameters and device characteristics. The key electrical parameters extracted with high accuracy from the predicted output I-V curves confirm this. Figs. 8(a)-(d) are the scatter plots for I_{ON} for the n-FET and p-FET devices at linear and saturation regions. As compared to the baseline model results, it can be observed from the plots that the I_{ON} predicted values are falling on the reference line except for a few outliers. It represents that the proposed model is able to transfer the training knowledge to validate the test data and has an excellent prediction capability. Similarly, the predicted values of I_{OFF}, V_{TH}, SS, and DIBL in Figs. 8(e)-(h), Figs. 8(i)-(l), Figs. 8(m)-(p), and Figs. 8(q)-(r), respectively, are very close to the reference line and clearly show that the proposed model is giving much more reasonable predictions. Therefore, the LSTM model predicts the values that are closely following the simulated values and are concentrated along the reference line, unlike the baseline model where the predicted values are dispersed. It can be inferred that the LSTM model is a generic ML model that can capture the underlying physics of transfer characteristics and its significant electrical parameters.

In particular, the LSTM shows a significantly lower error and high prediction accuracy as compared to the baseline model. In addition to this, Table III also summarizes the training time taken by the baseline and the LSTM model. It can be analyzed that the ML-based device modeling is fast and is able to predict the process variations effects in the electrical characteristics of the CFET devices accurately with the limited dataset. The ML algorithms can significantly speed up the modeling of electrical characteristics of advanced semiconductor devices.

6. Conclusions

In summary, the LSTM-based ML model was proposed to estimate the effects of PV in the GAA Si NS CFET and predict the variations in transfer characteristics. The rmse value and r^2-score depict that the single LSTM model can handle the sequence of I_D-V_G values with an accuracy $> 95\%$. The key parameters I_{ON}, I_{OFF}, V_{TH}, SS, and DIBL were also extracted from the predicted values and showed a good agreement with the simulated values. To further check the effectiveness of the proposed LSTM model, its relative performance was compared with the baseline MLP model. However, the baseline model shows that the prediction performance to capture the variability in electrical characteristics is not accurate. The results exhibit that the LSTM model outperformed the baseline model in terms of a good r^2-score and lower rmse values. We also applied the early stopping criteria as a regularization technique to avoid the model overfitting and save more time by preventing the models from unnecessary training for more epochs. In the future, our focus is on the use of ML models in CFET-based circuit simulation to minimize the huge computational power and time.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgement

This work was supported in part by the National Science and Technology Council, Taiwan, under Grant MOST 111–2221-EA49–181, MOST 111–2634-F-A49–008, and NSCT 111–2218-E-492–009-MBK; in part by the “2022 Qualcomm Taiwan Research Program (NYCU)” under Grant NAT-487835 SOW; and in part by the “Center for mm Wave Smart Radar Systems and Technologies” under the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan.

References

[1] Huang CY, et al. 3-D self-aligned stacked NMOS-on-PMOS nanoribbon transistors for continued moore’s law scaling. In: IEEE International Electron Devices Meeting; 2020. 20.6.1-4.
[2] Mangal G, et al. Numerical investigation and temperature-based analysis of the analog performance of fully gate-covered junctionless FinFET. Comput Electr Eng 2022;101.
[3] Kola SR, et al. Characteristics of Gate-All-Around Silicon Nanowire and Nanosheet MOSFETs with Various Spacers. Simulation of semiconductor processes and devices. 2020. p. 1–4.
[4] Chuang M-H, et al. DC Characteristics and Dynamic Properties of Multi-Channel Nanosheet MOSFETs with and without Tungsten Metal Sidewall for Sub-3-nm Technological Nodes. ECS Journal of Solid State Science and Technology 2022;11:065001.
[5] Sung W-L, et al. Characteristics of stacked gate-all-around Si nanosheet MOSFETs with metal sidewall source/drain and their impacts on CMOS circuit properties. IEEE Trans Electron Devices 2021;68:3124–8.
[6] Cheng CK, et al. Multirow Complementary-FET (CFET) Standard Cell Synthesis Framework Using Satisfiability Modulo Theories (SMTs). Journal on Exploratory Solid-State Computational Devices and Circuits 2021;7(1):43–51.

[7] Jung SG, et al. Performance analysis on complementary FET (CFET) relative to standard CMOS with nanosheet FET. IEEE Journal of the Electron Devices Society 2021;10:78–82.
[8] Jiang L, et al. Complementary FET device and circuit level evaluation using fin-based and sheet-based configurations targeting 3 nm node and beyond. Simulation semiconductor processes devices. 2022. p. 323–6.
[9] Lu W, et al. Nanowire Transistor Performance Limits and Applications. IEEE Trans Electron Devices 2008;55(11):2859–76.
[10] Li Y, et al. The effect of the geometry aspect ratio on the silicon ellipse-shaped surrounding-gate field-effect transistor and circuit. Semicond Sci Technol 2009;24 (9):095018.
[11] Nagy D, et al. Benchmarking of FinFET, nanosheet, and nanowire FET architectures for future technology nodes. IEEE Access 2020;8:53196–202.
[12] Mohapatra E, et al. Design study of gate-all-around vertically stacked nanosheet FETs for sub-7 nm nodes. SN Applied Sciences 2021;540:1–13.
[13] Li Y, et al. Random Work-Function-Induced Threshold Voltage Fluctuation in Metal-Gate MOS Devices by Monte Carlo Simulation. IEEE Trans Semicond Manuf 2012;25(2):266–71.
[14] Li Y, et al. Process-variation effect, metal-gate work-function fluctuation, and random-dopant fluctuation in emerging CMOS technologies. IEEE Trans Electron Devices 2009;57(2):437–47.
[15] Paul BC, et al. Impact of a Process Variation on Nanowire and Nanotube Device Performance. IEEE Trans Electron Devices 2007;54(9):2369–76.
[16] Li B, et al. Process variation dependence of total ionizing dose effects in bulk nFinFETs. Microelectron Reliab 2018;88(90):946–51.
[17] Mohapatra E, et al. Strain induced variability study in Gate-All-Around vertically-stacked horizontal nanosheet transistors. Physics Scripta 2020;95:1–10.
[18] Oh MH, et al. Sensitivity Analysis Based on Neural Network for Optimizing Device Characteristics. IEEE Electron Device Lett 2020;41(10):1548–51.
[19] Butola R, et al. A Machine Learning Approach to Modeling Intrinsic Parameter Fluctuation of Gate-All-Around Si Nanosheet MOSFETs. IEEE Access 2022;10: 71356–69.
[20] Akbar C, et al. Deep Learning Approach to Inverse Grain Pattern of Nanosized Metal Gate for Multichannel Gate-All-Around Silicon Nanosheet MOSFETs. IEEE Trans Semicond Manuf 2021;34(4):513–20.
[21] Ko K, et al. Prediction of Process Variation Effect for Ultrascaled GAA Vertical FET Devices Using a Machine Learning Approach. IEEE Trans Electron Devices 2019;66(10):4474–7.
[22] Wong HY, et al. TCAD-Machine Learning Framework for Device Variation and Operating Temperature Analysis With Experimental Demonstration. IEEE Journal of the Electron Devices Society 2020;8:992–1000.
[23] Lee K, et al. Analysis on Process Variation Effect of 3D NAND Flash Memory Cell through Machine Learning Model. Electron Devices Technology & Manufacturing Conference 2020:1–4.
[24] International Roadmap for Devices and Systems (IRDS™) 2021 Edition. [Online]. Available: https://irds.ieee.org/editions/2021.
[25] Franco J, et al. Novel low thermal budget gate stack solutions for BTI reliability in future Logic Device technologies. In: International Conference on IC Design and Technology; 2021. p. 1–4.
[26] Kola SR, et al. Random telegraph noise in gate-all-around silicon nanowire MOSFETs induced by a single charge trap or random interface traps. J Comput Electron 2020;19:253–62.
[27] Sung W-L, et al. DC/AC/RF characteristic fluctuations induced by various random discrete dopants of gate-all-around silicon nanowire n-MOSFETs. IEEE Trans Electron Devices 2018;65(6):2638–46.
[28] Li Y, et al. Process variation effect, metal-gate work-function fluctuation and random dopant fluctuation of 10-nm gate-all-around silicon nanowire MOSFET devices. In: IEEE International Electron Devices Meeting; 2015. p. 887–90.
[29] Kang J, et al. Long short-term memory-based Malware classification method for information security. Computers & Electrical Engineering 2019;77:366–75.
[30] Wang H, et al. IncLSTM: incremental Ensemble LSTM Model towards Time Series Data. Computers & Electrical Engineering 2021;77:107156.

Rajat Butola is currently pursuing his Ph.D. degree with the Parallel and Scientific Computing Laboratory, the EECS International Graduate Program, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. His research interests include machine learning and deep learning algorithms, and their applications in semiconductor device simulation..

Yiming Li is currently a Full Professor of electrical and computer engineering, 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 (Wiley, 2021). His current research interests include computational electronics, device modeling, numerical simulation, parameter extraction, and optimization techniques.

Sekhar Reddy Kola is currently pursuing his Ph.D. degree with the Parallel and Scientific Computing Laboratory, the EECS International Graduate Program, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. His research interests include simulation of nano-CMOS devices using technology computer-aided design tools.

Chandni Akbar is currently a Postdoctoral Researcher of the Parallel and Scientific Computing Laboratory, the Institute of Communications Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. Her research interests include machine learning and deep learning algorithms, and their applications in semiconductor device simulation.

Min-Hui Chuang is currently pursuing her Ph.D. degree with the Parallel and Scientific Computing Laboratory, the Institute of Communications Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. Her research interests include numerical simulation of emerging semiconductor devices, device model parameter extraction and computational intelligence.