Solid-State Electronics 207 (2023) 108710

Contents lists available at ScienceDirect

Solid State Electronics

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

An accurate machine learning model to study the impact of realistic metal grain granularity on Nanosheet FETs☆,☆☆

Julian G. Fernandez [a,*], Natalia Seoane [a], Enrique Comesaña [b], Juan C. Pichel [a], Antonio Garcia-Loureiro [a]

[a] Centro Singular de Investigación en Tecnoloxías Intelixentes, Universidade de Santiago de Compostela, Spain
[b] Escola Politécnica Superior de Enxeñaría, Campus Terra, Universidade de Santiago de Compostela, Spain

ARTICLE INFO

Keywords:
Machine learning
TCAD
Nanosheet FET
Metal grain granularity
Variability

ABSTRACT

In this work, we present a machine learning neural network model to predict the impact of realistic metal grain granularity (MGG) variability on the threshold voltage V_{Th} and on the $I_D - V_G$ characteristics of a silicon-based 12 nm gate length nanosheet FET. This model is based on the multi-layer perceptron (MLP) machine learning architecture. As realistic MGG maps consist of the distribution of grains on the gate with different work-function values, it is relevant to apply algorithms such as the principal component analysis to reduce these features to the most representative ones. Once the realistic MGG features are correctly reduced without losing information, we train two different neural networks with the neurons in the output layer as the only difference, to predict the V_{Th} and the $I_D - V_G$ characteristics, respectively. The comparison between TCAD results and the model, shows excellent agreement for the mean and standard deviation of V_{Th} distributions for different average grain sizes values (from 3 nm to 10 nm) demonstrating the accuracy of the machine learning model. Also, we study the amount of data needed to accurately train the MLPs, leading to results that allow us to drastically reduce the computational time required to perform variability studies for state-of-art nano FET devices.

1. Motivation

Nanosheet FETs are currently considered one of the preferred architectures for the next technology nodes [1]. Due to the expensive manufacture of new devices, other solutions, such as technology-aided computer design (TCAD), are needed to evaluate the reliability of future transistors, being the metal grain granularity (MGG) one of the most harmful sources of variability [2]. As the realistic simulation of these variability studies is computationally demanding, it is essential to explore new techniques such as the Pelgrom-based predictive model [3] or the application of machine learning methodologies [4]. Previous works, combined machine learning models with the application of synthetic MGG profiles with fixed-size rectangular grains [5] or square grains [6]. However, since the average metallic grain size (GS) depends on the annealing temperature and duration of the gate deposition process [7], it is crucial to evaluate a variety of GSs while capturing the realistic random shapes of the grains deposited on the gate.

In this context, we present a multi-layer perceptron (MLP) neural network to estimate the impact of MGG on a 12 nm gate length nanosheet (NS) FET. Also, by modifying the output layer of the same MLP, we can accurately predict the threshold voltage (V_{Th}) or the $I_D - V_G$ characteristics of the studied transistors. In order to correctly describe the impact of MGG on nanosized transistor, we use random realistic MGG profiles based on Poisson–Voronoi diagrams for several GSs (from 3 nm to 10 nm) to feed the MLP model.

The contents of this paper are distributed as follows. Section 2 shows the methodology, from the description of the simulation process to the MLP structure. Then, Section 3 presents the MLP performance, and its comparison with TCAD data. Finally, Section 4 summarizes the main conclusions of this work.

2. Methodology: From TCAD to MLP

This study is based on a Si NSFET (see Fig. 1) previously calibrated in [8] against the experimental device reported in [9] at high drain

☆ Work supported by the Spanish MICINN, Xunta de Galicia, Spain and FEDER, Spain funds (RYC-2017-23312, PID2019-104834GB-I00, ED431F 2020/008, PLEC2021-007662, and ED431C 2022/16).
☆☆ The review of this paper was arranged by Sorin Cristoloveanu.
* Corresponding author.
E-mail address: julian.garcia.fernandez2@usc.es (J.G. Fernandez).

https://doi.org/10.1016/j.sse.2023.108710

Available online 22 July 2023

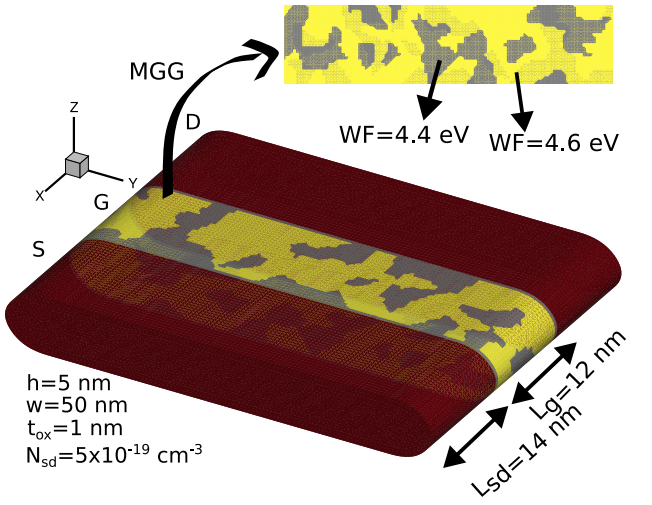

Fig. 1. 12 nm gate length (Lg) nanosheet FET affected by MGG. Regions: source (S), gate (G), and drain (D). Lsd and Nsd are the length and doping of S and D. w and h are the channel width and height. *tox* is the effective oxide thickness. The TiN work functions (WF) are 4.4 eV (40%) and 4.6 eV (60%).

bias ($V_D = 0.7V$). The simulations were carried out using the in-house built three-dimensional semiconductor device simulator toolbox VENDES [10]. This software allows us to reproduce the physics inside the device by applying the drift-diffusion (DD) transport method, coupled with density-gradient quantum corrections. The chosen criteria to extract the V_{Th} is the linear extrapolation (LE) method [11].

MGG consists of the appearance of different metallic grain orientations with different work functions (WF) during the gate deposition process. To implement this source of variability in a realistic way, we have generated random MGG profiles where the grains are created with Poisson–Voronoi diagrams, depending on the GS [12] (see an example of a realistic MGG profile in Fig. 1). The TiN metal gate has two grain orientations with WF of 4.4/4.6 eV, and occurrence probabilities of 40/60%, respectively. To have statistical significance, we generate around 900 profiles for each of the GSs studied in this work ($3, 5, 7, 10$ nm).

The MLP was developed using Python 3.9, the Scikit-learn 1.0.2 [13], and the PyTorch Lightning 1.9.0 library. Several hyperparameters, such as the batch size ($bs = 64$), the initial learning rate ($lr = 0.1$), or the number of neurons and hidden layers, were calibrated to optimize the MLP performance using the Ray Tune 2.2.0 library [14]. Fig. 2 shows the structure of the MLP, with an input layer corresponding to the number of features of the realistic MGG profile (N_{x_i}), two hidden layers with 234 and 44 neurons, and an output layer with 1 neuron corresponding to the V_{Th}. Furthermore, an identical structure with 21 number of neurons in the output layer is used to predict the $I_D - V_G$ characteristics. ReLU is used as the activation function, the mean square error (MSE) as the loss function, and an adaptative lr scheduler to avoid divergence in the MSE minimization. The stochastic gradient descent (SGD) optimization algorithm with a momentum = 0.9 is implemented. Also, the initialization of the neural network layers and nodes weights is implemented with a normal distribution with mean 0 and standard deviation 0.01, which were also optimized with Ray Tune.

The main issue of using a realistic MGG profile (368 × 41 discretization points) is its huge amount of features in each MGG profile. The number of features ($N_{x_i} = 15088$), which is greater than the sample size ($N_{sample} = 3604$), can produce problems of overfitting of the neural network, limiting its generalization and prediction capacity [15]. To deal with this issue, the principal component analysis technique (PCA) is applied with a 95% threshold cut-off of the cumulative variance [16], to determine the representative N_{x_i} value. With this methodology (see Fig. 3) the train dataset features are reduced from $N_{x_i} = 15088$ to

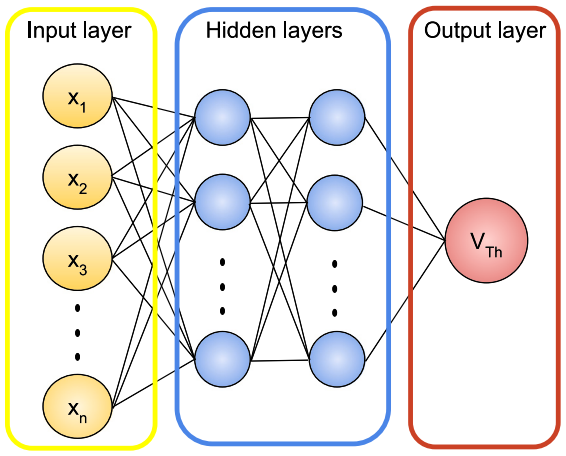

Fig. 2. A multi-layer perceptron neural network with an input layer, two hidden layers and an output layer. x_1 to x_n (input) are the MGG features. V_{Th} (output) is the threshold voltage.

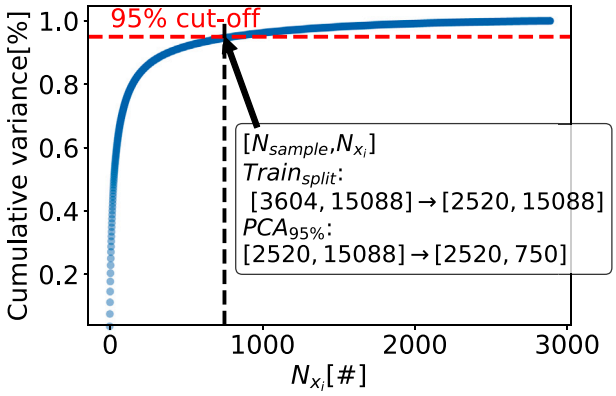

Fig. 3. Cumulative variance against the number of features (N_{x_i}). The data reduction process is explained in the box, first the split of the dataset for the training process and after the PCA reduction for the 95% cut-off of the cumulative variance.

$N_{x_i} = 750$ making $N_{x_i} < N_{train}$. The sample is split into three subsets (train, validation, test), being their size $N_{train} = 2520$, $N_{test} = 540$, and $N_{val} = 544$. Once this procedure is applied, the MGG data is ready to feed the MLP.

3. Numerical results: Performance and prediction

With the conditions previously mentioned, the MLP was trained to predict the V_{TH} MGG-induced variability on a NSFET. The computational time (t_{comp}) to reduce the features with PCA and train the MLP network is 6 min, with the advantage of being usable for future predictions without any extra computational cost. Fig. 4 shows the comparison between the MLP predicted and the VENDES simulated V_{Th} values for the test dataset. The metrics used to evaluate the training process are the coefficient of determination (R^2) and the mean absolute percentage error ($MAPE$), obtaining for the test values a $R^2 = 0.975$ and a $MAPE = 1.5\%$.

Based on an identical MLP structure used for V_{Th}, the I_D–V_G characteristics of each device impacted by realistic MGG can be predicted only by varying the number of neurons in the output layer. In this case, to train the MLP network the output neurons were modified from 1 (V_{Th}) to 21 (number of points of the simulated $I_D - V_G$ characteristics). The main advantage of predicting the $I_D - V_G$ characteristics is to have all the information about the performance of the devices, being able to extract not only the V_{Th} but also other relevant figures of merit as the

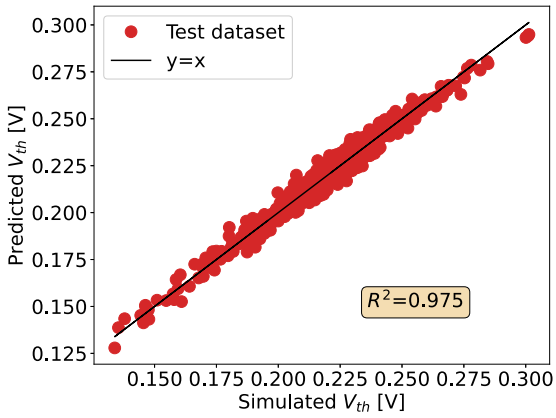

Fig. 4. Comparison between simulated TCAD threshold voltage (V_{Th}) and MLP predictions for the test dataset. The coefficient of determination (R^2) is also shown.

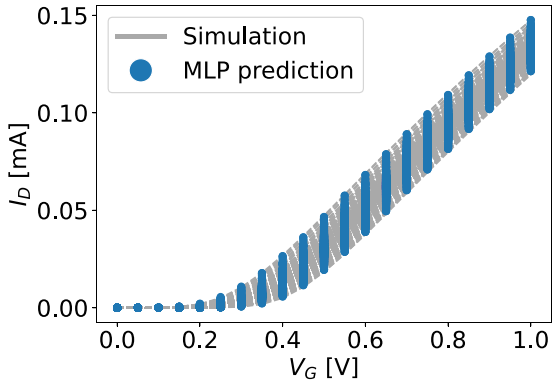

Fig. 5. Comparison of $I_D - V_G$ characteristics simulated using VENDES versus predicted values using the MLP for the test dataset.

off current, the subthreshold slope, or the on current. Fig. 5 displays the comparison between the MLP $I_D - V_G$ predictions (blue points) and the TCAD $I_D - V_G$ values (dashed gray lines). The metrics for the test of the $I_D - V_G$ characteristics are $R^2 = 0.967$ and $MAPE = 1.8\%$. In this case, the performance metrics for the test are slightly lower than those of the MLP V_{Th} but for this second application of the model, the hyperparameters have not been optimized.

To compare the predictive power for the two MLPs and their accuracy with the TCAD simulation results, we evaluate the statistics for the V_{Th} distributions as a function of the GS. For this purpose, the same training dataset is used to feed both MLPs, and also the comparison is made with the same testing dataset. For consistency, the criteria chosen to extract the V_{Th} from the predicted $I_D - V_G$ values is the same as in the TCAD, the LE method [11]. Fig. 6 shows the mean threshold voltage (μV_{Th}, top figure) and the standard deviation threshold voltage (σV_{TH}, bottom figure) of the V_{Th} distributions for each GS due to the three compared methodologies (TCAD, MLP V_{Th}, and MLP $I_D - V_G$). It can be seen that the deviations in the μV_{Th} values is lower than 3.5 mV, together with a perfect matching of the σV_{Th} values, demonstrating the accuracy of the model presented in this work.

Considering that in an Intel Core i9-10850K CPU 3.60 GHz processor each quantum corrected DD simulation takes 7.5 h, decreasing N_{train} will lead to a massive reduction in t_{comp}. Fig. 7 shows the effect after PCA features reduction, of decreasing the fraction of the training dataset from 1 ($N_{train} = 2520$) to 0.2 ($N_{train} = 504$). The performance metrics for the lower fraction of the training dataset (0.2) for MLP V_{Th} are $R^2 = 0.942$, $MAPE = 2.05\%$, and for MLP $I_D - V_G$ are $R^2 = 0.940$, $MAPE = 2.10\%$. These results demonstrate the feasibility of reducing an 80% the number of simulations required to train the MLPs,

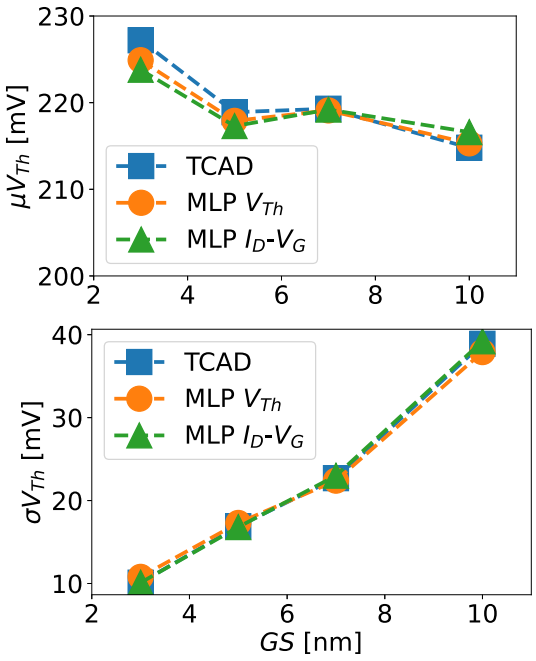

Fig. 6. Impact of the average metal grain size (GS) on the V_{Th} statistics for the three methodologies presented (TCAD, MLP V_{Th}, and MLP $I_D - V_G$). The top/bottom figure shows the mean threshold voltage/threshold voltage standard deviation ($\mu V_{Th}/\sigma V_{Th}$).

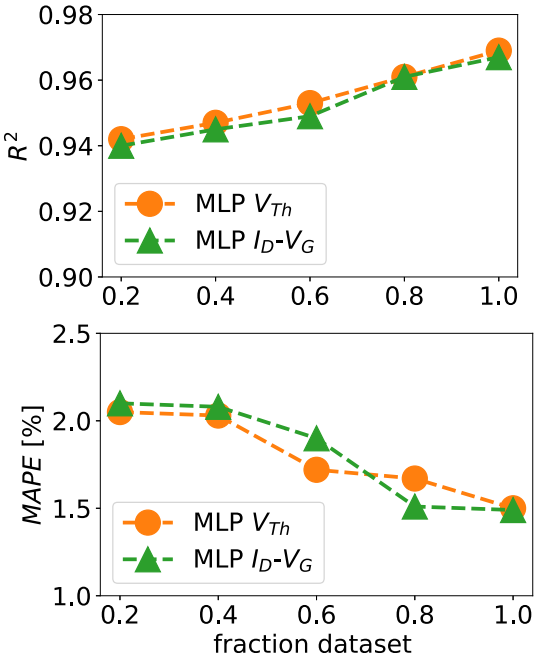

Fig. 7. Coefficient of determination (R^2, top) and mean absolute percentage error ($MAPE$, bottom) versus the fraction of dataset used to train the two different MLPs of this work.

lowering 5× the t_{comp}, without a significant loss of accuracy. The metric variations between training with the whole dataset and its 20% are $\Delta R^2 \sim 0.02$, and $\Delta MAPE \sim 0.5\%$.

4. Conclusions

We have presented two strategies based on a machine learning model to accurately estimate the MGG-induced variability on a 12 nm

NSFET. The first strategy consists of directly predicting the V_{Th} from realistic MGG maps, whereas the second strategy predicts the complete I_D–V_G characteristics of the devices impacted by MGG. The comparison between TCAD simulations and the MLP neural networks shows an accurate prediction of the V_{Th} distributions as their deflections for the mean and standard deviation are almost negligible.

We demonstrated that the presented model could obtain coefficients of determination of $R^2 \sim 0.94$ and mean absolute percentage errors of $MAPE \sim 2\%$ when using only the 20% of the training dataset, reducing 5.0× the computational time with respect to the total train dataset. Moreover, once the MLP model is trained, it can accurately predict the impact of realistic MGG variability on V_{Th} and on the I_D–V_G characteristics, with no further simulations.

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.

References

[1] More moore. In: IEEE IRDS. 2022.
[2] Seoane N, et al. 2021. http://dx.doi.org/10.1109/LED.2021.3109586.
[3] Fernandez JG, et al. 2022. http://dx.doi.org/10.1109/JEDS.2022.3214928.
[4] Carrillo-Nuñez, et al. 2019. http://dx.doi.org/10.1109/LED.2019.2931839.
[5] Butola R, et al. 2022. http://dx.doi.org/10.1109/TMTT.2022.3198659.
[6] Akbar C, et al. 2022. http://dx.doi.org/10.1016/j.compeleceng.2022.108392.
[7] Dadgour H, et al. 2008. http://dx.doi.org/10.1109/IEDM.2008.4796792.
[8] Nagy D, et al. 2020. http://dx.doi.org/10.1109/ACCESS.2020.2980925.
[9] Loubet N, et al. 2017. http://dx.doi.org/10.23919/VLSIT.2017.7998183.
[10] Seoane N, et al. 2019. http://dx.doi.org/10.3390/ma12152391.
[11] Conde AO, et al. 2013. http://dx.doi.org/10.1016/j.microrel.2012.09.015.
[12] Indalecio G, et al. 2016. http://dx.doi.org/10.1109/TED.2016.2556749.
[13] Pedregosa F, et al. 2011. http://dx.doi.org/10.48550/arXiv.1201.0490.
[14] Liaw R, et al. 2018. http://dx.doi.org/10.48550/arXiv.1807.05118.
[15] Hastie T, et al. 2009. http://dx.doi.org/10.1007/978-0-387-21606-5.
[16] Jolliffe IT. 2002. http://dx.doi.org/10.1002/9781118445112.stat06472.