Machine Learning Approach for Predicting the Effect of Statistical Variability in Si Junctionless Nanowire Transistors

Hamilton Carrillo-Nuñez, Nadezhda Dimitrova, Asen Asenov *Fellow, IEEE*, Vihar Georgiev, *Member, IEEE*

***Abstract*— This work investigates the possibility to replace numerical TCAD device simulations with a multi-layer neural network (NN). We explore if it is possible to train the NN with the required accuracy in order to predict device characteristics of thousands of transistors without executing TCAD simulations. In order to answer this question, here we present a hierarchical multi-scale simulation study of a silicon junctionless nanowire field-effect transistor (JL-NWT) with a gate length of 150 nm and diameter of an Si channel of 8 nm. All device simulations are based on the Drift-Diffusion (DD) formalism with activated density gradient (DG) quantum corrections. For the purpose of this work, we perform statistical numerical experiments of a set of 1380 automictically different JL-NWTs. Each device has a unique random distribution of discrete dopants (RDD) within the silicon body. From those statistical simulations, we extract important figures of merit (FoM), such as OFF-current (I_{OFF}) and ON-current (I_{ON}), subthreshold slope (SS) and voltage threshold (V_{TH}). Based on those statistical simulations, we train a multi-layer NN and we compare the obtained results with a general linear model (GLM). Our work shows the potential of using NN in the field of device modelling and simulation with a potential application to significantly reduce the computational cost.**

***Index Terms*—nanowires, TCAD simulations, statistical variability, deep learning, neural networks.**

I. Introduction

SILICON nanowires have a wide spectrum of promising applications, such as current field-effect transistors [1] [2], photovoltaics [3], energy conversion and storage [4] and qubits [5]. In our previous work, we have shown an extensive comparison between simulations and experimental results for JL-NWTs with Ω-gated region and with a channel length of 150 nm [6]. In another study, we have discussed statistical simulation results based on an ensemble of 500 JL-NWTs, where each device is atomistically unique with random distribution of discreet dopants in the channels [7]. Results obtained from those previously reported works have allowed us to suggest an improvement of the device design, predict the device performance and to extract important Figures of Merit

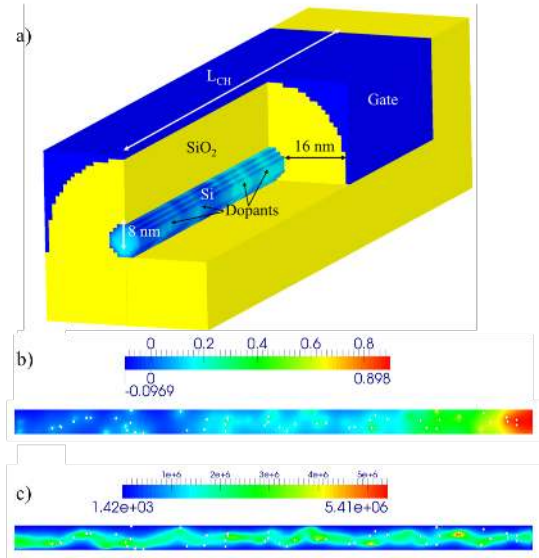

Fig. 1 a) Sketch of the JL-NWT, with cuts along and perpendicular to the transport direction. The Si channel shows the potential where fingerprints from the randomly distributed dopants (RDD) are visible. b) shows the 2D potential and c) the current density in the Si channel along the transport direction of wire. The white dots are the discrete random dopants.

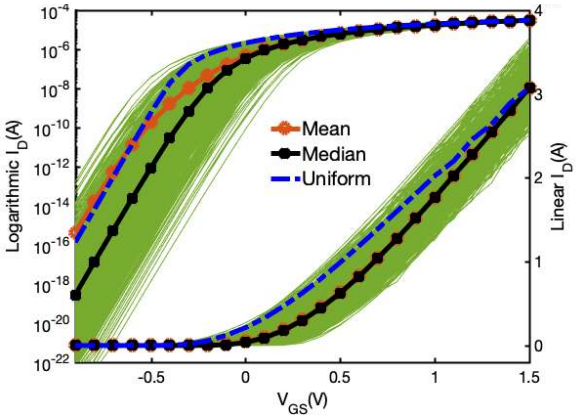

Fig. 2 I_D-V_G characteristics of 1380 JL-NWTs devices used in this work considering unique random dopant distribution for each device (green curves). The uniform device (blue curve) is continuously doped with a doping concertation of $N_D = 10^{19}$ cm^{-3}.

This work is supported by the UKRI under Projects EP/P009972/1 and EP/S001131/1.

H. Carrillo-Nunez, N. Dimitrova, A. Asenov and V. P. Georgiev are with the School of Engineering, University of Glasgow, Glasgow G12 8LT, UK (vihar.georgiev@glasgow.ac.uk).

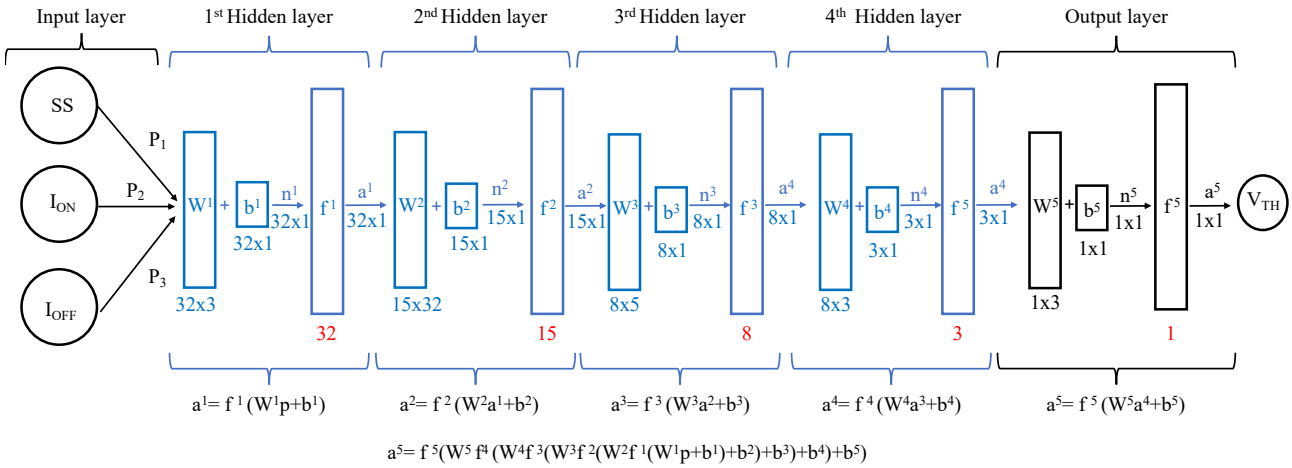

Fig. 3 Schematic representation of the multi-layer neural network (NN) used in this work. Weight, bias and logistic function are depicted by W^i, b^i and f^i respectively. A liner function is annotated by f^4 and the output of each neuron is illustrated by a^i.

(FoM), such as OFF-current (I_{OFF}) and ON-current (I_{ON}), subthreshold slope (SS) and voltage threshold (V_{TH}).

In this letter, we would like to extend our previous works [6] [7] further and explore the opportunity to include multi-layer neural network (NN) in our analysis in order to predict device characteristics without running numerical TCAD device simulations. To the best of our knowledge this is the first paper that establishes a direct link between NN and TCAD simulations.

This paper is organised as follows. The simulation methodology is presented in Section II. Then we report the main findings obtained from the numerical TCAD simulations and NN results in Section III. The final Section IV describes the conclusions.

II. Simulation Methodology and Results

Fig. 1 shows the device geometry. In all devices, the channel length is kept constant at L_{CH} = 150 nm, the Si cross-section is 8 nm and the oxide thickness is fixed at 16 nm. The only parameter that differs for each device is the position and the number of RDD in the Si region. Performing statistical numerical simulations of the current-voltage (I_D-V_G) characteristics (see Fig. 2) allows us to extract FoM (I_{OFF}, I_{ON}, SS, V_{TH}).

In this work, JL-NWTs are simulated using the commercial TCAD simulator—GARAND, which is now part of the Synopsys TCAD Sentaurus Simulator. All device simulations are based on the Drift-Diffusion (DD) formalism with activated Density Gradient (DG) quantum correction [8]. The DD+DG methodology is necessary in order to take into account quantum confinement effects in ultra-scaled JLNWs as shown in Fig. 1. Also, the inverse sombrero shape of the Coulomb potentials, associated with individual discrete charges created by the randomly distributed dopants, is accurately captured by the DG correction [8], [16]. In our simulations the DD+DG formalism is coupled with the Poisson equation in a self-consistent loop [9] that gives us the opportunity to calculate current-voltage characteristics of the devices (see Fig. 2).

Fig. 2 shows the impact of the random dopant distribution (RDD) on current–voltage ($I_D - V_G$) characteristics. For the purpose of this analysis, 1380 JL-NWTs with different RDDs are simulated. Based on those $I_D - V_G$ curves we extract FoM, such as I_{OFF}, I_{ON}, SS and V_{TH}.

Using three of those FoM (I_{OFF}, I_{ON}, SS) as an input to a multi-layer NN, we train the NN by applying a back-propagation algorithm [10] to predict the values for the fourth FoM – V_{TH}. Fig. 3 shows the NN produced as a result of testing various configurations of NN by changing the number of hidden layers. Single input layer (a row vector of dimension of 3 × 1), four hidden layers (with 32, 15, 8, 3 neurons respectively) and an output layer form the final structure. Fig. 3 also shows the W^i and b^i for a given layer i. The NN is implemented in Python making use of the TensorFlow and Keraus package [11]. The gradient descent technique [12] which minimises the MSE of the final result is used to determine the values of W^i and b^i. In the NN, each function (f) except f^5 corresponds to a logistic function used at each layer.

More specifically in this work, we use the data obtained from the device simulations to train the NN. The aim is to use three of the FoMs: SS, I_{OFF} and I_{ON}, to predict the fourth one – V_{TH}. In order to achieve that, the data for all 1380 devices is split into two sets. The first set has 700 devices that are split 70% to 30% for training and testing sets of the NN, correspondingly. Once the NN is trained, we use the second (validation) set of 680 devices in order to evaluate the accuracy of the trained NN.

The main descriptors that are used to validate the accuracy of the trained NN are the mean square error (MSE) and the mean absolute error (MAE). Fig. 4 shows the MSE and MAE for the training and validation sets as a function of the training steps (Epoch). As expected, the values of both errors quickly decrease and reach a constant value before the defined accuracy of the NN is reached.

To help to analyse the results, two plots are presented in Fig. 5 and Fig. 6. Fig. 5 shows the predicted V_{TH} from the NN and general linear model (GLM) versus the true value of the V_{TH} taken from the numerical device simulations. GLMs are powerful statistical tools as they accommodate response variables of non-normal distributions thus allowing for multiple statistical models such as Poisson regression, Linear Regression and Logistic regression to come together [14] [15].

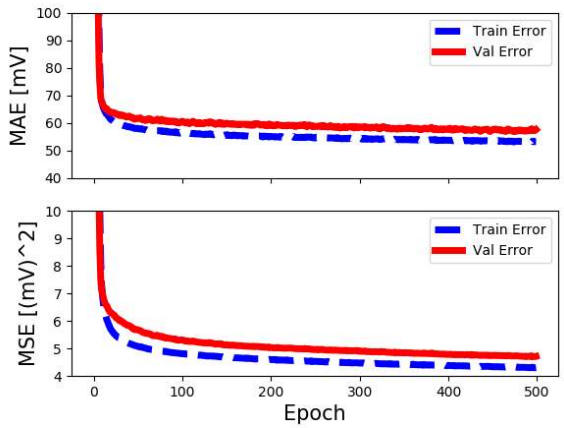

Fig. 4 Mean square error (MSE) and mean absolute error (MAE) function for the neural network (NN). The blue line represents the error on the training set as a function of each epoch and the red line corresponds to the error on the validation set as a function of each epoch.

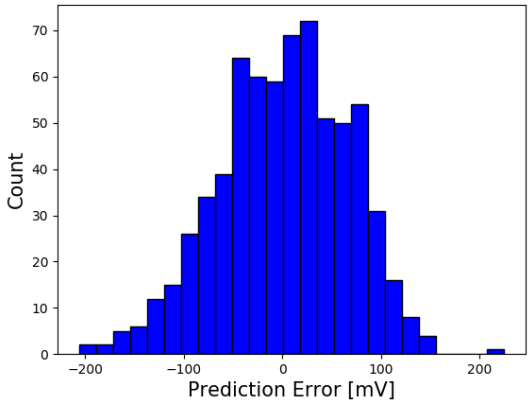

Fig. 7 Error histogram distribution plot for the predicted value of the V_{TH}. The error is evaluated on the second data set containing 680 devices which are used only for testing the NN.

The GLM is set up to train and test on the exact same data as the NN, ensuring that their results can be compared effectively. In general, if the predicted value of V_{TH} is a 100% match with the true value of V_{TH}, each blue or grey dot should be on the top of the red line. In our case, both NN and GLM are in agreement with the low values of the V_{TH} and larger variability for devices with V_{TH} below the value of -300 mV.

However, in order to establish if the NN performs better than the GLM, in Fig. 6 we plot the MAE and MSE for both methods. Fig. 6 shows that the MSE obtained by the NN equals 4.32 mV^2 and GLM gives a higher value of 10.33 mV^2. MAE is also smaller for the NN in comparison to the GLM, 53.34 mV and 64.06 mV, correspondingly. This is somehow surprising because the NN requires more parameters (782 parameters in this case) if compared to the GLM and it still gives the lowest MSE and MAE values. Those results show that our NN is trained well and performs better than the GLM.

Another useful method for displaying the results is by using a bar chart diagram that shows the error for the predicted values of V_{TH} for the testing data (see Fig. 7). The error histogram follows closely a Gaussian distribution which is another indicator of the success of the training procedure in the NN.

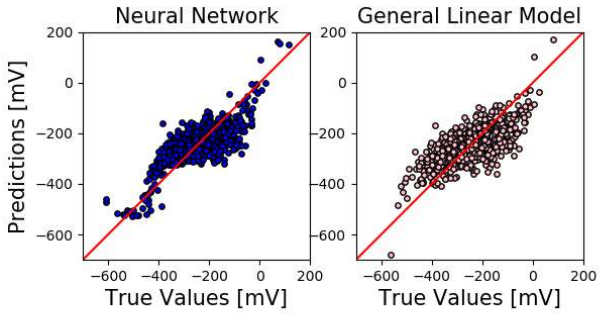

Fig. 5 True vs predicted values for the voltage threshold (V_{TH}) evaluated of the validation data set containing 680 devices which are used only for testing the NN (left hand side) and GLM (right hand side).

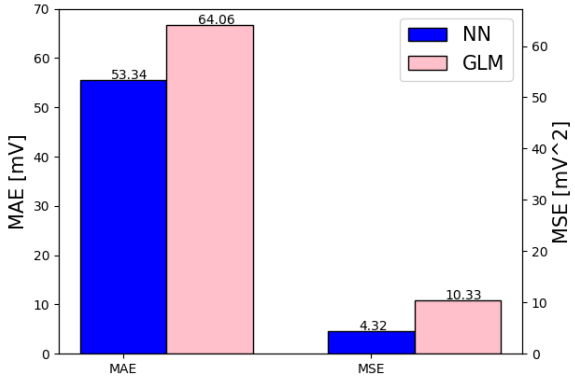

Fig. 6 MAE and MSE error bars for the NN and the GLM.

III. Conclusions

In this work, we have reported a statistical analysis of the RDD variability is the Si n-type JL-NWT. A statistical sample of 1380 microscopically different transistors has been simulated. The simulated current-voltage characteristics allowed us to extract key Figures of Merit (I_{OFF}, I_{ON}, SS) which we used to train the NN in order to predict the V_{TH}.

Results obtained from the NN can lead to the following conclusion. Using the NN instead of the TCAD simulations in principle could significantly decrease the computational time and shorten the research and development process. For example, each I_D-V_G curve takes on average 4 hours to simulate 25 V_G points. Hence, for 1380 devices this makes 5520 hours (331,20min) total simulations time. However, the computational time for the NN is less than 2 min (1.26 min). This includes running the NN and producing the figures reported here.

Since the NN shows stable training process with the Gaussian distribution of the error histogram, a better performance in comparison to the GLM and ability to predict new and unknown set of data, which is used only for testing and not training, we are confident that the presented NN is well trained. Also, similar NNs can be used to describe material properties, such resistance in metal nanowires [13], that cannot be described by nonparametric methods, such as a general linear model. However, it needs to be noted that predictivity of the NN can be improved even further by providing more data, using different pre-processing schemes and attempting alternative network architectures. Indeed, all of these options are currently under investigation.

REFERENCES

[1] J. Appenzeller, J. Knoch, M. Bjork, H. Riel, H. Schmid and W. Riess, "Toward Nanowire Electronics," *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 2827–2845, Nov. 2008. DOI: 10.1109/TED.2008.2008011

[2] J.-P. Colinge, "The SOI MOSFET: from Single Gate to Multigate," in *FinFETs and Other Multi-Gate Transistors*, J.-P. Colinge, Ed.: Springer US, 2008, pp. 1-48. DOI:10.1007/978-0-387-71752-4_1

[3] B. Tian, X. Zheng, T. J. Kempa, Y. Fang, H. Yu, G. Yu, J. Hunag and C. M. Lieber, "Coaxial Siliconnano Wires as Solar Cells and Nanoelectronic Power Sources," *Nature*, vol. 449, no. 7164, pp. 885–889, Oct. 2007. DOI: http://dx.doi.org/10.1038/nature06181.

[4] A. I. Boukai, Y. Bunimovich, J. Tahir-Kheli, J. K. Yu, W. A. Goddard and J. R. Heath, "Silicon Nanowires as Efficient Thermoelectric Materials," *Nature*, vol. 451, no. 7175, pp. 168–171, 2008. DOI:10.1038/nature06458

[5] S. Nadj-Perge, S. M. Frolov, E. P. A. M. Bakkers and L. P. Kouwen- Hoven, "Spin-orbit Qubit in a Semiconductor Nanowire," *Nature*, vol. 468, no. 7327, pp. 1084–1087, Dec. 2010. DOI: http://dx. doi.org/10.1038/nature09682.

[6] V. P. Georgiev, M. M. Mirza, F. Adamu-Lema, S. M. Towie, C. Riddet, D. MacLaren, A. Asenov and D. J. Paul , "Experimental and Simulation Study of Silicon Nanowire Transistors Using Heavily Doped Channels," *IEEE Transactions on Nanotechnology*, vol. 16, no. 5, pp. 727-735, 2017. DOI: 10.11.09/TNANO.2017.2665691

[7] H. N. Carrillo-Nuñez, M. M. Mirza, D. J. Paul, D. A. MacLaren, A. Asenov and V. P. Georgiev, "Impact of Randomly Distributed Dopants on Ω-Gate Junctionless Silicon Nanowire Transistors," *IEEE Transactions on Electron Devices*, vol. 65, no. 5, pp. 1692-1698, 2018. DOI: 10.1109/TED.2018.2817919

[8] A. Asenov, A. Brown, G. Roy, B. Cheng, C. Alexander, C. Riddet, U. Kovac, A. Martinez, N. Seoane and S. Roy, "Simulation of Statistical Variability in Nano-CMOS Transistors Using Drift-Diffusion, Monte Carlo and Non-Equilibrium Green's Function Techniques," *J. Comput. Electron.*, vol. 8, issue. 3–4, pp. 349–373, 2009. DOI: http://dx.doi.org/10.1007/s10825-009-0292-0.

[9] A. Asenov, "Random Dopant Induced Threshold Voltage Lowering and Fluctuations in sub-0.1 µm MOSFET's: A 3-D 'Atomistic' Simulation study," *IEEE Trans. Electron Devices*, vol. 45, no. 12, pp. 2505–2513, Dec. 1998. DOI: http://dx.doi.org/10.1109/16.735728.

[10] *Neural Network Design*, M. Hagan, H. Demuth, and M. Beale: PWS Publishing Company, Boston, MA, 1996. ISBN: 0-534-94332-2

[11] https://www.tensorflow.org/

[12] *Neural Networks—A Comprehensive Foundation*, S. Haykin: Prentice-Hall, Englewood Cliffs, NJ, 1999.

[13] D. Valencia, E. Wilson, Z. Jiang, G. A. Valencia-Zapata, G. Klimeck and M. Povolotskyi, "Grain-Boundary Resistance in Copper Interconnects: From an Atomistic Model to a Neural Network," *Physical Review Applied,* vol. 9, no. 4, p. 044005, 04/04/ 2018. DOI: 10.1103/PhysRevApplied.9.044005

[14] J. A. Nelder and R. W. M. Wedderburn, "*Generalized linear models*," Journal of the Royal Statistical Society. Series A (General), vol. 135, pp. 370–384, 1972. Stable URL: http://www.jstor.org/stable/2344614

[15] O. Gencel, F. Kocabas, M. S. Gok and F. Koksal, "Comparison of artificial neural networks and general linear model approaches for the analysis of abrasive wear of concrete", *Construction and Building Materials*, vol. 25, no. 8, pp. 3486 – 3494, 2011. DOI: https://doi.org/10.1016/j.conbuildmat.2011.03.040

[16] A. R. Brown, J. R. Watling, G. Roy, C. Riddet, C. Alexanger, U. Kovac, A. Martinez and A. Asenov, "Use of density gradient quantum corrections in the simulation of statistical variability in MOSFETs," *J. Comput. Electron.*, vol. 9, no. 3, pp. 187-196, 2010. DOI: 10.1007/s10825-010-0314-y