Physics-Informed Neural Network for Predicting Out-of-Training-Range TCAD Solution with Minimized Domain Expertise

Albert Lu[1], Yu Foon Chau[2], Hiu Yung Wong[1*]
[1]San Jose State University, USA, [2]University of California, Irvine, USA, [*]hiuyung.wong@sjsu.edu

Abstract

In this paper, a Si nanowire transistor is used to demonstrate the possibility of using a physics-informed neural network to predict out-of-training-range TCAD solutions without accessing internal solvers and with minimal domain expertise. The machine can predict a 10 times larger range than the training data and also predict the inversion region behavior with only subthreshold region training data. The physics-informed module is trained without human-coded differential equations making this extendable to more sophisticated systems.

Keywords: Physics Informed Neural Networks (PINN), ML, TCAD, Nanowire, Out-of-training-range prediction

Introduction

Machine learning (ML) is promising in assisting technology computer-aided design (TCAD) simulations to alleviate difficulty in convergence and prolonged simulation time. Roughly, ML is used in TCAD in three approaches. In the first approach, ML is used to improve solver performance by providing an initial guess [1]-[2]. This requires access to the internal solver of a TCAD tool and significant domain expertise by coding physical equations (including differential operators) in the loss functions. The second approach is to use TCAD to generate terminal data such as currents and voltages and then use ML to perform reverse engineering [3]-[6] or simulation emulation [7]-[9]. This does not require access to the internal solver and has been proven to be able to predict out-of-training range data [9]. If an appropriate machine learning method is used, it can also minimize the requirement for domain expertise [7]-[9]. The third approach is to solve the TCAD problem by training a machine with the spatial distribution of physical quantities such as electron density and potential [10][11] without using device physics (as the solver is not accessible) and mostly can only predict the physical quantities within the training range.

Physics-informed neural network (PINN), in which the physics of the problem is incorporated into the NN, is expected to improve predictability [12]. In this paper, we study a new approach in which a machine is trained by the spatial distribution of physical quantities with PINN and minimal domain expertise *without accessing the solver of the tool* using a Si nanowire (NW) transistor as an example.

Data Generation

An n-type Si NW transistor is studied using TCAD Sentaurus. Half of the domain was simulated due to its symmetry. Fig. 1 shows the structure with a radius of 3 nm, a gate length of 18 nm, and an oxide thickness of 1 nm. The source/drain and body doping are 10^{20} cm^{-3} (n-type) and 10^{10} cm^{-3} (p-type), respectively. Tensor mesh is used (129 × 17 = 2193 mesh points). The electrostatic potential (ϕ), electron density (n), and net space charge after device simulation at each point are extracted using a TDX script in TCAD Sentaurus. The extracted quantities are verified with an in-house Python-coded Poisson equation solver (Fig. 2).

The structure is then simulated in SDevice with a setup that includes Fermi-Dirac statistics. The Poisson equation is solved with the drain and source voltages at 0 V. The gate voltage, V_G, is ramped from 0 to 0.75 V, and snapshots of the electrostatic potential and electron density are taken for every 7.5 mV for a total of 101 snapshots. Note that each 'snapshot' (profile) records ϕ and n distributions in the 2193-point mesh.

Physics-Informing Module

Firstly, a physics-informing module machine (not PINN yet) is created. This machine predicts the ϕ profile for a given n profile. This is equivalent to solving the Poisson equation for a given space charge (ρ) profile in the NW,

$$\nabla \cdot (\varepsilon \nabla \phi) = -\rho \qquad (1)$$

where ε is the permittivity. Hole and acceptor concentrations are ignored due to their small values in this problem.

The module is trained using linear regression (LR) and the first 40 snapshots (*only* from V_G = 0V to 0.3V). The input is the n profile and the output is the ϕ profile. The electron density is normalized by adding 10^{10} cm^{-3} and then dividing by 10^{19} cm^{-3}. Adding 10^{10} cm^{-3}, which is the noise level in this problem, is to avoid division by zero in the oxide region. Fig. 3 plots ϕ predicted by the machine against the ϕ calculated by TCAD for every mesh point from V_G = 0V to 0.75V for the corresponding n profiles from TCAD and they agree with each other. *Note again the machine has not seen the profiles for V_G > 0.3V*. This demonstrates the ability of the machine to learn sophisticated physics where ϕ in the channel vary significantly and non-linearly while those in the source and drain stay almost constant. Moreover, the training data are in the subthreshold regime which has a very different V_G dependency than the inversion regime. LR is sufficient because when the n profile is given, Eq. (1) is just a linear matrix multiplication after discretization. Therefore, the machine models the inverse of the matrix at the left-hand side in Eq. (1). Systems of differential equations are generally discretized and linearized. Therefore, this methodology is expected to be applicable to more sophisticated problems.

PINN Architecture

A PINN is then built. It consists of a convolutional neural network (CNN) and the aforementioned physics-informing module. Fig. 4 shows the architecture. The input is the desired V_G and the outputs are the n and ϕ profiles. For any given V_G, the n profile is predicted by a CNN. The output layer of the CNN is chosen to be the exponential linear unit (ELU). It will then be passed to the physics-informing LR module which will calculate the corresponding ϕ profile, from which the gate voltage, V_G', can be extracted. If the solution is correct, $V_G' = V_G$. This comparison forms the first loss function and enforces the gate boundary condition.

The predicted ϕ profile is then further used to predict the n profile using the Fermi-Dirac (FD) module in Fig. 3, which is also a physics-informing module but much more straightforward. It calculates the n profile based on the ϕ profile through the approximation of the Fermi Integral of order ½ for Fermi-Dirac statistics [13] and is compared to the n profile predicted by the CNN. This forms the second constraint and is used to ensure that the predicted electron density obtained through the approximate Fermi Integral matches the predicted electron density from the CNN. This essentially enforced the Fermi-Dirac statistics. Fig. 5 shows that the approximate Fermi Integral calculation is correct by plotting the approximation against the actual value obtained from TCAD. In the loss calculation, the logarithmic value of n is taken so that low-density values also have enough weight.

For each V_G of interest, the two losses will be minimized using the Adam optimizer. To speed up the process, the learning rate is initially 10^{-3} but as it plateaus it is changed adaptively until a minimum of 10^{-5} is reached. This process can be treated as a replacement for TCAD solvers. This can also be regarded as *self-supervised learning*. Note again that, in the whole process, only the LR module has been trained with the first 40 V_G snapshots. Moreover, *unlike regular TCAD simulation, one can calculate V_G = 0.75V directly without ramping V_G.*

Results

Fig. 6 and Fig. 7 show the predicted ϕ profile and n profile compared to TCAD simulation results for V_G = 0.75V, respectively. The maximum error is less than 0.3% for the ϕ profile and less than 0.6% for the n profile in the logarithmic scale. Fig. 8 shows the prediction of ϕ by PINN at a point (see Fig. 1) in the middle of the structure under the gate region across all V_G. This point was chosen because it experiences the transition from the depletion to the inversion regime. The result matches the TCAD simulation well. This demonstrates the ability of the PINN to learn the physics of this device and predict out-of-training-range results and is not based on simple extrapolation. Fig. 9 further shows the predicted vs. actual ϕ and n of every mesh point across all V_G. The n comparison is in log scale and shows how the PINN is able to learn across several orders of magnitude successfully. The number of epochs in the previous figures is 200,000. Fig. 10 shows that <30,000 epochs are enough for most applications.

The out-of-range predictability is further studied using the PINN (trained up to V_G = 0.3V). Fig. 11 shows that it can predict up to V_G = 3V well with 20,000 epochs.

Conclusion

A novel framework based on a PINN is demonstrated to have effectively learned the underlying physics by using specific physics-based loss functions. It can predict out-of-training-range simulation results as it has learned the underlying physics. A machine trained by the TCAD data from V_G = 0V to 0.3V (below the subthreshold region) is shown to be able to predict the ϕ and n profiles for any given V_G (up to 3V).

Acknowledgments

Part of the work was supported by the National Science Foundation under Grant No. 2046220.

References

[1] S. -C. Han et al., "Deep Neural Network for Generation of the Initial Electrostatic Potential Profile," 2019 SISPAD, pp. 1-4.

[2] S. -C. Han et al., "Acceleration of Semiconductor Device Simulation With Approximate Solutions Predicted by Trained Neural Networks," in IEEE TED, vol. 68, no. 11, pp. 5483-5489, Nov. 2021.

[3] C. -W. Teo et al., "TCAD-Enabled Machine Learning Defect Prediction to Accelerate Advanced Semiconductor Device Failure Analysis," 2019 SISPAD, pp. 1-4.

[4] Y. S. Bankapalli and H. Y. Wong, "TCAD Augmented Machine Learning for Semiconductor Device Failure Troubleshooting and Reverse Engineering," 2019 SISPAD, pp. 1-4.

[5] H. Y. Wong et al., "TCAD-Machine Learning Framework for Device Variation and Operating Temperature Analysis With Experimental Demonstration," in IEEE JEDS, vol. 8, pp. 992-1000, 2020.

[6] K. Mehta and H. -Y. Wong, "Prediction of FinFET Current-Voltage and Capacitance-Voltage Curves Using Machine Learning With Autoencoder," in IEEE EDL, vol. 42, no. 2, pp. 136-139, Feb. 2021..

[7] H. Dhillon et al., "TCAD-Augmented Machine Learning With and Without Domain Expertise," in IEEE Transactions on Electron Devices, vol. 68, no. 11, pp. 5498-5503, Nov. 2021.

[8] K. Mehta et al., "Improvement of TCAD Augmented Machine Learning Using Autoencoder for Semiconductor Variation Identification and Inverse Design," in IEEE Access, vol. 8, pp. 143519-143529, 2020.

[9] V. Eranki et al., "Out-of-Training-Range Synthetic FinFET and Inverter Data Generation Using a Modified Generative Adversarial Network," in IEEE EDL, vol. 43, no. 11, pp. 1810-1813, Nov. 2022.

[10] J. Lee et al., "Device simulations with a U-Net model predicting physical quantities in two-dimensional landscapes," Sci. Rep., vol. 13, no. 1, p. 731, Jan. 2023.

[11] W. Jang et al., "TCAD Device Simulation With Graph Neural Network," in IEEE Electron Device Letters, vol. 44, no. 8, pp. 1368-1371, Aug. 2023, doi: 10.1109/LED.2023.3290930.

[12] B. Kim and M. Shin, "A Novel Neural-Network Device Modeling Based on Physics-Informed Machine Learning," in IEEE Transactions on Electron Devices, vol. 70, pp. 6021-6025, Nov. 2023.

[13] D. Bednarczyk, et al., "The approximation of the Fermi-Dirac integral F½(η)," Physics Letters A, 64(4):409-410 (1978)

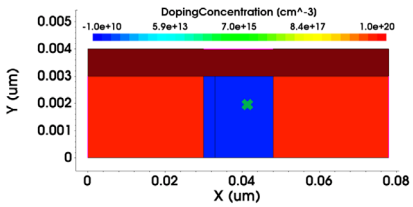

Fig. 1: Si NW structure used in this study. The cross shows the location used in Fig. 8.

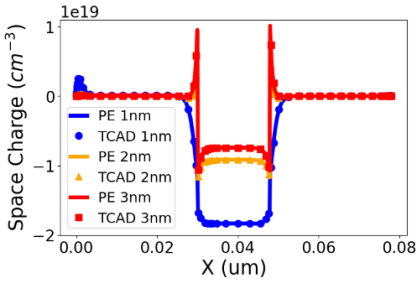

Fig. 2. Comparison between TCAD TDX extracted and in-house Poisson equation (PE) solved space charge at horizontal cutlines at different distances from SiO_2/Si interface (Fig. 1).

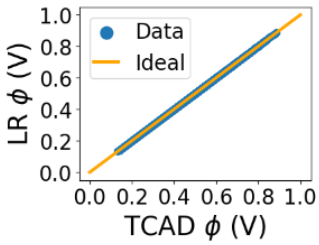

Fig. 3. Comparison between the electrostatic potential predicted using the physics-informing module and the electrostatic potential from TCAD. The blue points consist of all mesh points at all V_G from 0V to 0.75V.

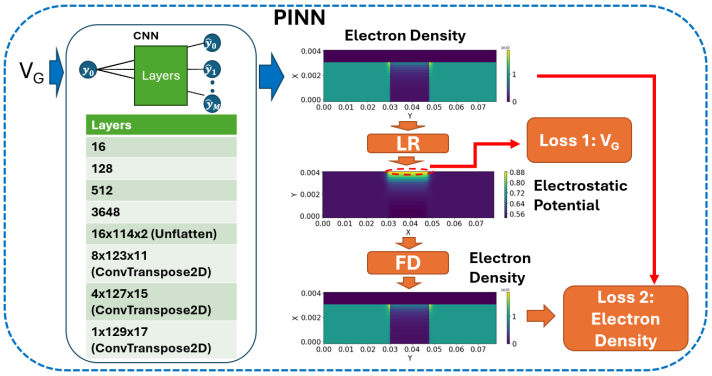

Fig. 4. The PINN used in this study consists of a CNN and 2 loss functions. V_G is input to the PINN and ϕ and n are the outputs. The physics-informing module is trained only using snapshots from V_G = 0V to 0.3V. The predicted n is passed through the LR machine to obtain ϕ. The gate contact ϕ matches the input V_G through loss function 1. The predicted ϕ is also passed through Fermi Integral to obtain n. Loss function 2 is used to obtain accurate n.

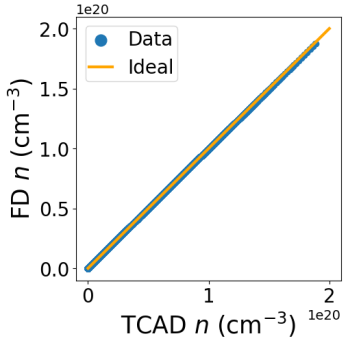

Fig. 5. Comparison between the electron density predicted using FD module and TCAD. The blue points consist of all mesh points at all V_G from 0V to 0.75V.

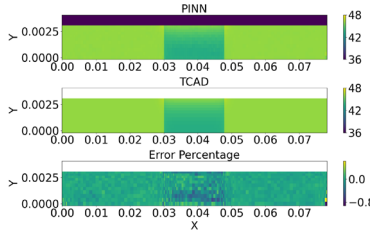

Fig. 6. Predicted electrostatic potential profile using the PINN (top) and the electrostatic potential from TCAD (middle) when V_G=0.75V. The max error percentage between them is less than 0.3% (bottom).

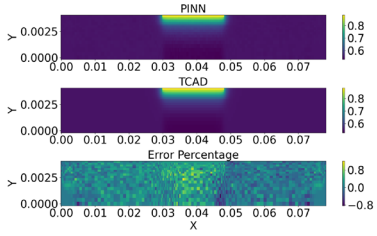

Fig. 7. Predicted electron density profile using the PINN (top) and the electron density from TCAD in log scale (bottom) when V_G=0.75V. The max error percentage between them is less than 0.6% (bottom).

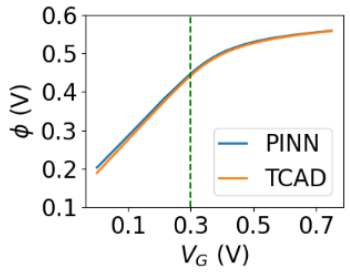

Fig. 8. Comparison of ϕ at x=40.5nm and y = 2nm (Fig. 1) predicted by PINN and TCAD at various V_G. The PINN was only trained by data before the green dotted line (V_G < 0.3V).

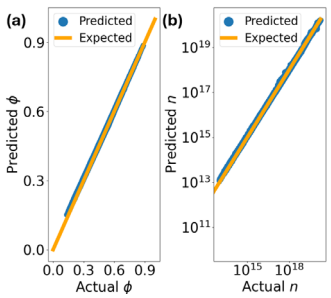

Fig. 9. PINN-predicted vs. TCAD (a) ϕ and (b) n. The blue points consist of all mesh points at all V_G from 0V to 0.75V.

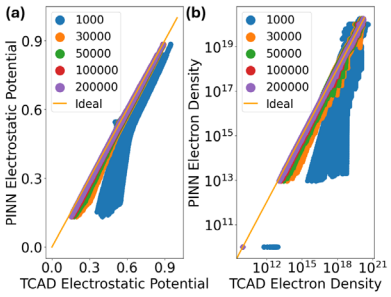

Fig. 10. PINN-predicted vs. TCAD (a) ϕ and (b) n with respect to the number of epochs.

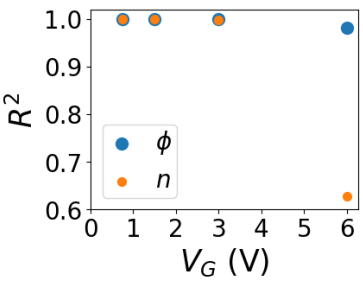

Fig. 11. R^2 values of ϕ and n against V_G for a machine trained with data only up to 0.3V.