

Acceleration of Three-Dimensional Device Simulation with the 3D Convolutional Neural Network

Seung-Cheol Han
 School of EECS
 Gwangju Institute of
 Science and Technology
 Gwangju, Republic of Korea

Jonghyun Choi
 AI Graduate School
 Gwangju Institute of
 Science and Technology
 Gwangju, Republic of Korea

Sung-Min Hong
 School of EECS
 Gwangju Institute of
 Science and Technology
 Gwangju, Republic of Korea
 E-mail: smhong@gist.ac.kr

Abstract—We propose to use a 3D convolutional neural network to accelerate three-dimensional device simulation by generating an electrostatic potential profile. In the training phase, the deep neural network is trained with the simulation results for various 3D MOSFETs in a supervised manner. The generated potential profile is used as an initial guess at a non-equilibrium condition, while carrier densities are estimated by the frozen field simulation. By numerical examples for three-dimensional MOSFETs, we show that the proposed method significantly reduces the number of the Newton iterations.

I. INTRODUCTION

Artificial neural networks have attracted much research attention, mainly due to their superior performance in various application fields. In semiconductor industry, applications of the neural network are rapidly expanding. For example, the time-consuming chip floor planning can be automated by adopting a deep reinforcement learning approach [1]. It allows chip design to be performed by artificial agents. Another example can be found in the compact modeling [2]. The neural network model can reproduce the current-voltage and charge-voltage characteristics of advanced FETs with excellent accuracy.

When the semiconductor device technology is concerned, various attempts, where the neural networks are applied to the TCAD (Technology Computer-Aided Design) device simulation, have been reported [3], [4], [5], [6], [7], [8], [9]. In these attempts, the neural network is used to describe the complicated input-output relation efficiently without considering the internal physical quantities. Simulation results are collected from several (typically hundreds or thousands) runs and the neural network is trained to generate the output characteristics directly from the input conditions.

Obviously, in these attempts, the computational cost is quite low, since no actual device simulation is performed once after the training phase is finished. However, the accuracy of the predicted result cannot be guaranteed in advance, because the underlying physical equations (*e.g.*, the drift-diffusion model in the device simulation) are not solved. Moreover, when a

new physical model is introduced (*e.g.*, the mobility model parameters are changed), the training data is also changed and we may need to train the neural network again.

On the other hand, there are other attempts where the internal physical quantities are considered. In our previous works, the deep neural networks were trained to generate approximate potential profiles for various devices such as diodes [10], BJTs [11], and two-dimensional MOSFETs [12]. We use the generated approximate potential profile as a good initial solution for the self-consistent device simulation. Therefore, the computational time can be reduced significantly, compared with the conventional solution method. When the converged solution is obtained, the accuracy of the final solution is identical with that of the conventional solution method. It is expected that the predicted electrostatic potential may be reused even with a new physical model. It is noted that a similar approach has been independently applied to the NEGF (Non-Equilibrium Green's Function) calculation in [13].

Although the second approach seems to be promising, it has been applied only to the one- [10] or two-dimensional [11], [12] device structures. Since the practical device structures such as FinFETs and nanosheet MOSFETs are three-dimensional, extension to the three-dimensional structure is in need. Here, we demonstrate that our method can be applied to three-dimensional MOSFETs with appropriate modifications.

Organization of this extended abstract is as follows. In Section II, the proposed neural network is described. Numerical examples for three-dimensional planar MOSFETs are shown in Section III. The conclusion is made in Section IV.

II. NEURAL NETWORKS

Since the electron and hole continuity equations become linear or at most locally nonlinear under a fixed electrostatic potential profile, carrier densities can be easily evaluated once the electrostatic potential is given. Compared with the full set of unknown variables (the electrostatic potential and two carrier densities), it is much more efficient to specify only the electrostatic potential. Therefore, the proposed deep

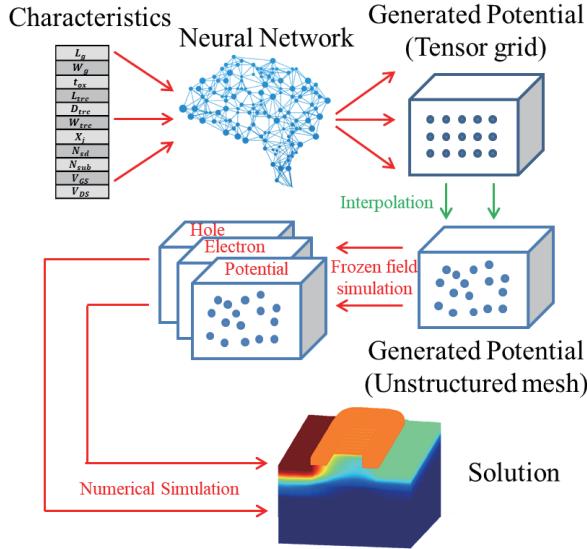


Fig. 1. Conceptual diagram for the proposed method. We consider three-dimensional structures in this work.

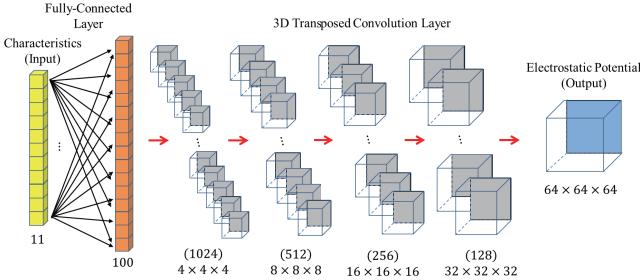


Fig. 2. Layer structure of the CNN structure adopted in the three-dimensional problem. The output layer generates a 64-by-64-by-64 tensor corresponding to the three-dimensional simulation domain.

neural network is trained to generate the electrostatic potential profiles.

Figure 1 shows a conceptual diagram for the proposed neural network, specifically the inference phase is presented. The device characteristics are provided to the trained neural network as the input parameters. Then, the electrostatic potential profile is generated as an output image.

For a two-dimensional device structure, a two-dimensional image was sufficient. However, for a three-dimensional device structure which we consider here, output needs to be three-dimensional (or a set of two-dimensional images, in other words). To address this, we employ a **3D convolution layer to a generative neural network** [14]. Figure 2 overviews the neural network structure we use. The output layer generates a 64-by-64-by-64 tensor.

Our deep neural network can generate potential profiles suitable for a tensor grid. On the other hand, a general device simulator adopts unstructured meshes. Since the meshes are different, the generated potential profile cannot be directly imported into the device simulator. In order to address such difficulty, we perform the **interpolation** between a tensor grid

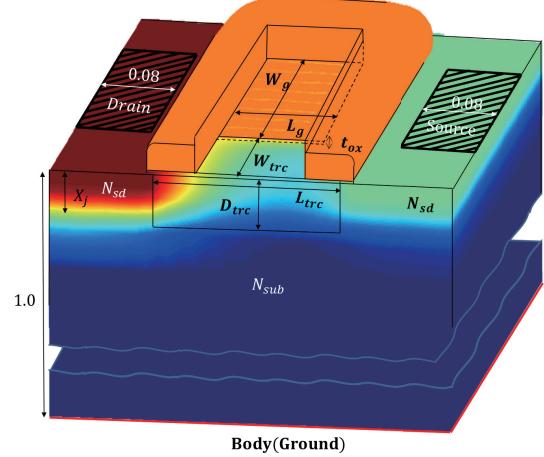


Fig. 3. Device template of structures under consideration for three-dimensional planar MOSFETs. Numbers represent lengths in microns.

of the deep neural network and an unstructured mesh of the device simulator, as discussed in [12]. The sampling and interpolation capability is implemented in our in-house device simulation framework (G-Device) [15], [16]. We implement the neural network using PyTorch [17].

III. THREE-DIMENSIONAL MOSFETs

The device template of three-dimensional planar MOSFETs is shown in Fig. 3. Several control parameters used in the device template are also shown. The gate length (L_g), gate width (W_g), oxide thickness (t_{ox}), trench length (L_{trc}), trench depth (D_{trc}), trench width (W_{trc}), and junction depth(X_j) are used as the length control parameters. The source/drain arsenic doping (N_{sd}) and substrate boron doping (N_{sub}) are used as the doping control parameters. The gate bias (V_{GS}) and drain bias (V_{DS}) are used as the bias control parameters. In total, 11 control parameters are introduced in this work. These control parameters are used as the input parameters of the deep neural networks, as shown in Fig. 1.

Table I shows ranges of control parameters in the training dataset. The training data set contains 1,000 instances of sampled potential profiles.

The distributions of input parameters are shown in Fig 4. As shown in the figure, the input parameters are randomly selected in order to consider the entire parameter range uniformly.

In training phase, the training and validation losses are measured as functions of the learning epoch in Fig. 5. Throughout the training phase, it is observed that the error is reduced for both the training dataset and the validation dataset. After 100 epochs, the losses are reduced considerably.

After the training phase, the trained convolutional network can generate an approximate potential profile in the inference phase. Figure 6a shows an example of the electrostatic potential profile generated by the trained convolutional neural network when input parameters are $L_g = 0.20 \mu\text{m}$, $W_g = 0.28 \mu\text{m}$, $t_{ox} = 1.6 \text{ nm}$, $L_{trc} = 0.32 \mu\text{m}$, $D_{trc} = 0.17 \mu\text{m}$, $W_{trc} = 0.091 \mu\text{m}$, $X_j = 0.13 \mu\text{m}$, $N_{sd} = 1.39 \times 10^{19} \text{ cm}^{-3}$, $N_{sub} =$

TABLE I
RANGES OF DEVICE CONTROL PARAMETERS

| Parameter | Minimum | Maximum |
|-----------|------------------------------------|------------------------------------|
| L_g | $0.11\mu\text{m}$ | $0.28\mu\text{m}$ |
| W_g | $0.1\mu\text{m}$ | $0.3\mu\text{m}$ |
| t_{ox} | 1.2nm | 3.0nm |
| L_{trc} | $0.2\mu\text{m}$ | $0.38\mu\text{m}$ |
| D_{trc} | $0.1\mu\text{m}$ | $0.3\mu\text{m}$ |
| W_{trc} | $0.05\mu\text{m}$ | $0.15\mu\text{m}$ |
| X_j | $0.05\mu\text{m}$ | $0.15\mu\text{m}$ |
| N_{sd} | $1.0 \times 10^{19}\text{cm}^{-3}$ | $1.0 \times 10^{19}\text{cm}^{-3}$ |
| N_{sub} | $5.0 \times 10^{16}\text{cm}^{-3}$ | $6.0 \times 10^{17}\text{cm}^{-3}$ |
| V_{GS} | 0.0V | 1.1V |
| V_{DS} | 0.0V | 1.1V |

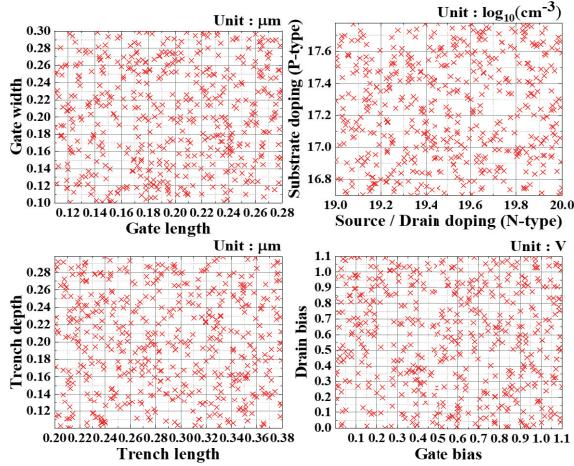


Fig. 4. Distributions of input parameters in the training dataset.

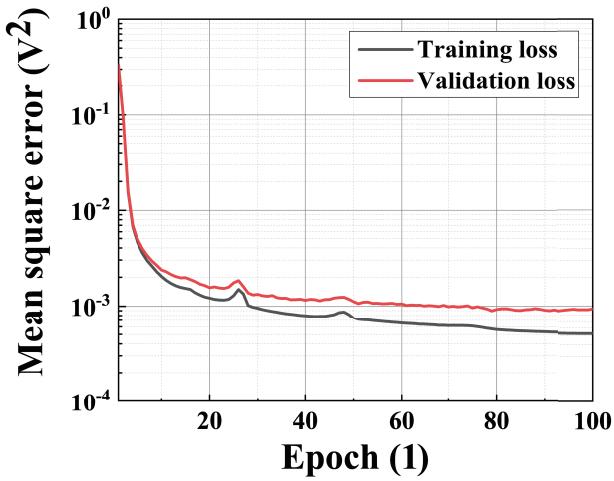


Fig. 5. Training and validation losses of a convolutional neural network, which is trained for the three-dimensional planar MOSFETs.

$8.80 \times 10^{16} \text{ cm}^{-3}$, $V_{GS} = 1.1 \text{ V}$, and $V_{DS} = 1.1 \text{ V}$. Since the structure is three-dimensional, visualizing the entire potential profile is not very convenient. Instead, quantities on a two-

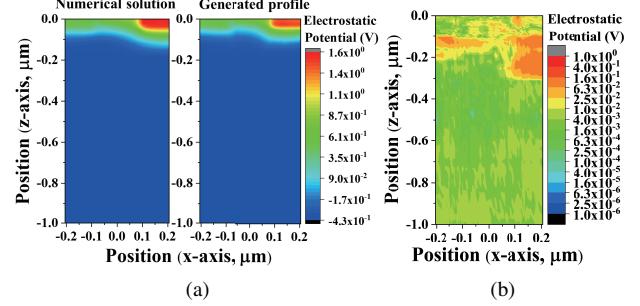


Fig. 6. (a) Numerical solution (Left) and a generated potential profile by the 3D convolutional neural network (Right) and (b) its error when $V_{GS} = 1.1 \text{ V}$ and $V_{DS} = 1.1 \text{ V}$. Device control parameters can be found in the main text. Quantities on a two-dimensional cross-section are drawn. The two-dimensional cross-section is located at the center position along the width direction.

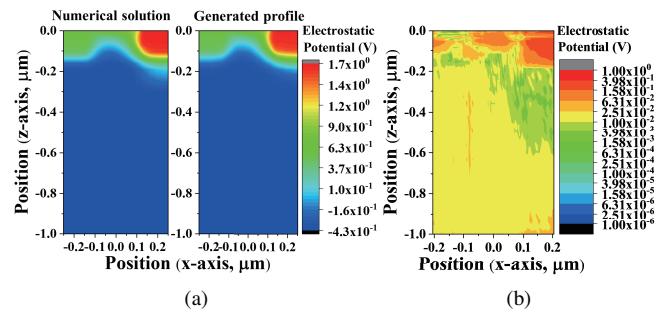


Fig. 7. (a) Numerical solution (Left) and a generated potential profile by the 3D convolutional neural network (Right) and (b) its error when $V_{GS} = 1.1 \text{ V}$ and $V_{DS} = 1.1 \text{ V}$. Device control parameters can be found in the main text. The MOSFET has a longer channel length than the one in Fig. 6.

dimensional cross-section are drawn. Its error, the difference between the generated potential profile and the numerical solution, is shown in Fig. 6b.

Fig. 7a shows another example with a longer gate. Its input parameters are $L_g = 0.11 \mu\text{m}$, $W_g = 0.1 \mu\text{m}$, $t_{ox} = 1.2 \text{ nm}$, $L_{trc} = 0.2 \mu\text{m}$, $D_{trc} = 0.1 \mu\text{m}$, $W_{trc} = 0.05 \mu\text{m}$, $X_j = 0.05 \mu\text{m}$, $N_{sd} = 1.39 \times 10^{19} \text{ cm}^{-3}$, $N_{sub} = 8.80 \times 10^{16} \text{ cm}^{-3}$, $V_{GS} = 1.1 \text{ V}$, and $V_{DS} = 1.1 \text{ V}$. Its error is shown in Fig. 7b.

As shown in Figs. 6b and 7b, the generated potential profiles are not perfectly matched with the numerical solutions. The maximum absolute error in Figs. 6b and 7b is as large as 0.36 V , which is not negligible at all. Nevertheless, they can serve as good initial solutions to accelerate the device simulation. A test dataset, including 500 device simulation results which are not included in the training dataset, is prepared. For all test cases, the numerical solutions are obtained without any convergence failure. The distribution of Newton iterations for the test dataset is shown in Fig. 8. In many cases, only 7 or 8 Newton iterations are needed to obtain the converged solution. It is noted that the typical number of Newton iterations with the conventional solution method, where the bias ramping from the equilibrium condition is adopted, is about 100. It means that the proposed method can accelerate the device simulation

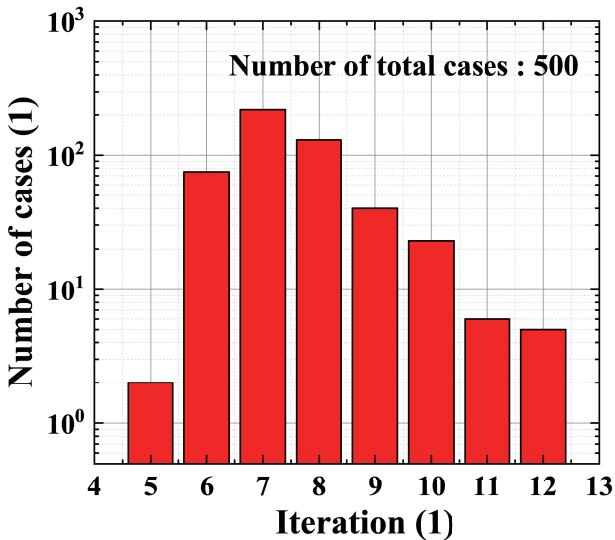


Fig. 8. Distribution of the Newton iterations when the generated potential profiles are used as initial solutions for 500 test cases.

significantly.

IV. CONCLUSION

We have shown that our previous approach to accelerate the device simulation can be extended to the three-dimensional device structures without much difficulty. The sampling and interpolation capability enables us to connect the tensor grid and the unstructured grid. The 3D convolution layer is adopted to generate the three-dimensional potential profiles. Significant acceleration has been achieved for the three-dimensional MOSFETs.

In this work, numerical results for the planar MOSFETs are shown. Of course, the multi-gate MOSFETs such as FinFETs or nanowire MOSFETs are of practical interest. Application to those multi-gate MOSFETs will be reported elsewhere.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2019R1A2C1086656 and NRF-2020M3H4A3081800). This work was also supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2019-0-01351, Development of Ultra Low-Power Mobile Deep Learning Semiconductor with Compression/Decompression of Activation/Kernel Data, 30 %)

REFERENCES

- [1] A. Mirhoseini, A. Goldie, M. Yazgan, J. W. Jiang, E. Songhori, S. Wang, Y.-J. Lee, E. Johnson, O. Pathak, A. Nazi, J. Pak, A. Tong, K. Srinivasa, W. Hang, E. Tuncer, Q. V. Le, J. Laudon, R. Ho, R. Carpenter, and J. Dean, “A graph placement methodology for fast chip design,” *Nature*, pp. 207–212, 2021.
- [2] J. Wang, Y.-H. Kim, J. Ryu, C. Jeong, W. Choi, and D. Kim, “Artificial neural network-based compact modeling methodology for advanced transistors,” *IEEE Transactions on Electron Devices*, vol. 68, no. 3, pp. 1318–1325, 2021.
- [3] H. Carrillo-Nuñez, N. Dimitrova, A. Asenov, and V. Georgiev, “Machine learning approach for predicting the effect of statistical variability in Si junctionless nanowire transistors,” *IEEE Electron Device Letters*, vol. 40, no. 9, pp. 1366–1369, 2019.
- [4] Y. S. Bankapalli and H. Y. Wong, “TCAD augmented machine learning for semiconductor device failure troubleshooting and reverse engineering,” in *2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2019, pp. 1–4.
- [5] H. Yun, J.-S. Yoon, J. Jeong, S. Lee, H.-C. Choi, and R.-H. Baek, “Neural network based design optimization of 14-nm node fully-depleted SOI FET for SoC and 3DIC applications,” *IEEE Journal of the Electron Devices Society*, vol. 8, pp. 1272–1280, 2020.
- [6] T. Wu and J. Guo, “Speed up quantum transport device simulation on ferroelectric tunnel junction with machine learning methods,” *IEEE Transactions on Electron Devices*, vol. 67, no. 11, pp. 5229–5235, 2020.
- [7] S. S. Raju, B. Wang, K. Mehta, M. Xiao, Y. Zhang, and H. Y. Wong, “Application of noise to avoid overfitting in TCAD augmented machine learning,” in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2020, pp. 351–354.
- [8] S. Myung, J. Kim, Y. Jeon, W. Jang, I. Huh, J. Kim, S. Han, K.-h. Baek, J. Ryu, Y.-S. Kim, J. Doh, J.-h. Kim, C. Jeong, and D. S. Kim, “Real-time TCAD: a new paradigm for TCAD in the artificial intelligence era,” in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2020, pp. 347–350.
- [9] H. Dhillon, K. Mehta, M. Xiao, B. Wang, Y. Zhang, and H. Y. Wong, “TCAD-augmented machine learning with and without domain expertise,” *IEEE Transactions on Electron Devices*, 2021.
- [10] S.-C. Han and S.-M. Hong, “Deep neural network for generation of the initial electrostatic potential profile,” in *2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2019, pp. 1–4.
- [11] S.-C. Han, J. Choi, and S.-M. Hong, “Electrostatic potential profile generator for two-dimensional semiconductor devices,” in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2020, pp. 297–300.
- [12] ———, “Acceleration of semiconductor device simulation with approximate solutions predicted by trained neural networks,” *IEEE Transactions on Electron Devices*, 2021.
- [13] S. Souma and M. Ogawa, “Acceleration of nonequilibrium Green’s function simulation for nanoscale FETs by applying convolutional neural network model,” *IEICE Electronics Express*, vol. 17, p. 20190739, 2020.
- [14] J. Wu, C. Zhang, T. Xue, W. T. Freeman, and J. B. Tenenbaum, “Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling,” 2017.
- [15] S.-M. Hong and J.-H. Jang, “Numerical simulation of plasma oscillation in 2-D electron gas using a periodic steady-state solver,” *IEEE Transactions on Electron Devices*, vol. 62, no. 12, pp. 4192–4198, 2015.
- [16] ———, “Transient simulation of semiconductor devices using a deterministic Boltzmann equation solver,” *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 156–163, 2018.
- [17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “PyTorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.