A Finite Element Framework for Solving the Density-Gradient Model

Pengcong Mu[1,2], Tao Cui[2,3], Lijun Xu[1,2], Kun Luo[1,2], Zhiqiang Li[1,2*], Zhenhua Wu[1,4*]
[1] Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China.
[2] University of Chinese Academy of Sciences, Beijing 100049, China.
[3] LSEC, NCMIS, Institute of Computational Mathematics and Scientific Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.
[4] School of Pyhsics, Zhejiang University, Hangzhou, 310027, China.
*Email: lizhiqiang@ime.ac.cn, wuzhenhua@zju.edu.cn

2024 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD) | 979-8-3315-1635-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/SISPAD62626.2024.10732921

***Abstract*—This paper presents a positive-preserving, stable finite element scheme for the density-gradient (DG) model. An efficient Newton-Krylov solver is designed to address the nonlinear coupled discrete system. Simulation results for gate-all-around (GAA) devices demonstrate the robustness and effectiveness of the proposed scheme. Notably, the simulation time using our method is only half that of the Sentaurus Device, highlighting its superior computational efficiency.**

***Index Terms*—Density-gradient model, Finite element method, Newton-Krylov method, GAA NS FETs**

I. Introduction

As modern semiconductor devices shrink to nanometer sizes, quantum mechanical effects become increasingly significant and must be considered in numerical simulations. The density-gradient (DG) model [1], [2], a macroscopic quantum equation, offers a computationally efficient alternative to directly solving the microscopic quantum Schrödinger equation. It is widely preferred in the industry due to its lower computational costs.

The finite element (FE) method is well-suited for solving problems with complex geometries, constructing high-order formulations, and enabling adaptive computation. Investigating numerical discretization methods for DG model within the finite element framework is of great academic and practical importance [3]. Additionally, three-dimensional device simulations require considerable computational time and memory resources. Techniques such as message-passing interface (MPI) parallel programming based on distributed memory and preconditioned Krylov subspace iterative methods [4] are crucial for managing these demands.

In this work, we propose a finite element framework for solving the DG model, develop a highly scalable MPI-based parallel solver, and implement simulations for GAA devices. This framework aims to provide a robust and efficient tool for addressing the computational challenges associated with modern semiconductor device simulations.

II. Methods

A. Physical Model and Exponential Transform

We split the fourth-order equations in the DG model into two second-order equations. Consequently, the static DG model can be regarded as a strongly coupled system comprising five nonlinear elliptic equations as outlined below:

$$\begin{aligned}
&\nabla\cdot(\epsilon\nabla\phi) = -q(p-n+D),\\
&\nabla\cdot(-q\mu_n n\nabla(\phi+\mathcal{G}_n)+qD_n\nabla n) = qR,\\
&\nabla\cdot(-q\mu_p p\nabla(\phi+\mathcal{G}_p)-qD_p\nabla p) = -qR,\\
&\mathcal{G}_n = \frac{1}{\sqrt{n}}\operatorname{div}\left(2b_n\nabla\sqrt{n}\right),\\
&\mathcal{G}_p = -\frac{1}{\sqrt{p}}\operatorname{div}\left(2b_p\nabla\sqrt{p}\right),
\end{aligned}$$

where ϕ is the electrostatic potential, n and p are the electron and hole densities, $\mathcal{G}_n$ and $\mathcal{G}_p$ are the quantum potentials for electrons and holes, q is the electronic charge, ϵ is the dielectric constant of the material, D and R are the doping profile and the net recombination rate, and $b_n = \frac{\gamma_n\hbar^2}{12qm_n^*}$, $b_p = \frac{\gamma_p\hbar^2}{12qm_p^*}$ are the density-gradient coefficients.

To preserve the positivity of numerical carrier densities, an exponential transformation of veriables $n/n_i = e^{2\psi_n}$, $p/n_i = e^{2\psi_p}$ are employed [5], [6]. The DG model can be rewritten in the following form:

$$-\nabla\cdot(\epsilon\nabla\phi) = q(n_i e^{2\psi_p} - n_i e^{2\psi_n} + D), \quad (1)$$
$$\nabla\cdot(\mu_n n_i e^{2\psi_n}\nabla\phi_n) = -R, \quad (2)$$
$$\nabla\cdot(\mu_p n_i e^{2\psi_p}\nabla\phi_p) = R, \quad (3)$$
$$\nabla\cdot(2b_n\nabla e^{\psi_n}) = e^{\psi_n}(\phi_n - \phi + 2V_{th}\psi_n), \quad (4)$$
$$-\nabla\cdot(2b_p\nabla e^{\psi_p}) = e^{\psi_p}(\phi_p - \phi - 2V_{th}\psi_p), \quad (5)$$

where ϕ_n and ϕ_p are the electron and hole quasi-Fermi potentials.

B. Finite Element Discretization

The solutions of (1)-(5) are approximate in the linear Lagrange element space [7]. The poisson equation is discretized using the traditional Galerkin finite element method. To descretize the DG equations (4)-(5), we propose an interpolated-exponential finite element (IEFE) scheme. Considering only electrons, the IEFE scheme for (4) is given by

$$-2b_n(\nabla\pi_h(e^{\psi_n}),\nabla v) = (e^{\psi_n}(\phi_n - \phi + 2V_{th}\psi_n), v)_I$$

where $\pi_h(\cdot)$ is the linear finite element interpolation, v is the test function, $(u,v) = \int_\Omega uv\mathrm{d}x$ represents the inner product on $L^2(\Omega)$, and $(\cdot,\cdot)_I$ is the discrete inner product defined as

$$(u,v)_I = \frac{1}{d+1}\sum_K \sum_{i=1}^{d+1} |K| u(x_i)v(x_i),\ d = 1,2,3.$$

For the current continuity equations (2)-(3), we apply the edge-averaged finite elemnt (EAFE) discretization [8]. For electrons, the EAFE scheme for (2) is formulated as

$$\sum_{K\in\mathcal{T}_h}\sum_{E\subset\partial K}\Big(\int_K \nabla b_{\boldsymbol{A}_1^E}\cdot\nabla b_{\boldsymbol{A}_2^E}\Big)\mathcal{A}_E(\mu_n n_i)\delta_E\phi_n\delta_E v = (R,v)$$

where $\mathcal{T}_h$ is a shape-regular tetrahedral mesh, $b_{\boldsymbol{A}}$ is the basis function associated with vertex $\boldsymbol{A}$, $\boldsymbol{A}_1^E$ and $\boldsymbol{A}_2^E$ are the endpoints of edge E, $\delta_E v := v(\boldsymbol{A}_2^E) - v(\boldsymbol{A}_1^E)$ and

$$\mathcal{A}_E(a) := \Big(\frac{1}{|E|}\int_E a^{-1}\Big)^{-1}$$

with $|E|$ representing the length of edge E. This method aims to approximate an advective-diffusive flux using a piecewise constant flux vector on each element, and it has been demonstrated to be equivalent to the finite volume Scharfetter-Gummel (FVSG) scheme.

C. Newton-Krylov Solver

To address the inherent nonlinearity, we propose a Newton method for solving the nonlinear discrete coupled system. This solver focuses on computing the corrections $\delta\boldsymbol{u}$ to the solution $\boldsymbol{u}$ of the aforementioned nonlinear finite element discrete system. Subsequently, the solution $\boldsymbol{u}$ is updated with a damping factor $0 < \alpha \leq 1$,

$$\boldsymbol{u} \longleftarrow \boldsymbol{u} + \alpha\delta\boldsymbol{u},$$

ensuring stability and convergence.

At each Newton iteration step, it is necessary to solve a large-scale, asymmetric, indefinite linear equation system. For this task, we employ the algebraic multigrid (AMG) [9] preconditioned flexible generalized minimal residual (FGMRES) [10] solver, which dramatically enhances the efficiency of simulations. Unlike geometric multigrid, AMG constructs a hierarchy of grids algebraically from the matrix representing the linear system. These grids are derived from the matrix's structure rather than corresponding to any physical space, thereby increasing the robustness of multigrid methods in practical applications. Nonetheless, the parameters involved in AMG need to be delicately tuned to achieve optimal performance.

III. Results

A. Simulation of a Single Channel GAAFET

To evaluate the performance of our proposed finite element discretization scheme for the DG model, we first consider an n-type 10 nm gate length single channel GAAFET as illustrated in Fig. 1(a) [11], with source and drain defined by two 5×5nm contacts. We select a finite element discretization mesh consisting of 2,734,080 tetrahedra, as shown in Fig. 1(b). To maintain the monotonicity of the scheme, the interior angle between any two faces of each tetrahedral element is constrained to be less than or equal to $\pi/2$ [8].

Figure 2(a) compares the I_D-V_G characteristics using the drift-diffusion (DD) model and the DG model, at both low and high drain voltages of 0.05 V and 0.7 V, respectively. The electron density in the OFF and ON states across the channel is depicted in Fig. 3(a)-3(b). Figure 2(b) presents the I_D–V_D characteristics for different source-to-gate bias voltages. It is important to note that the specific values are less critical for our purposes, as our primary objective is to demonstrate the numerical stability of our scheme rather than to produce physically meaningful results.

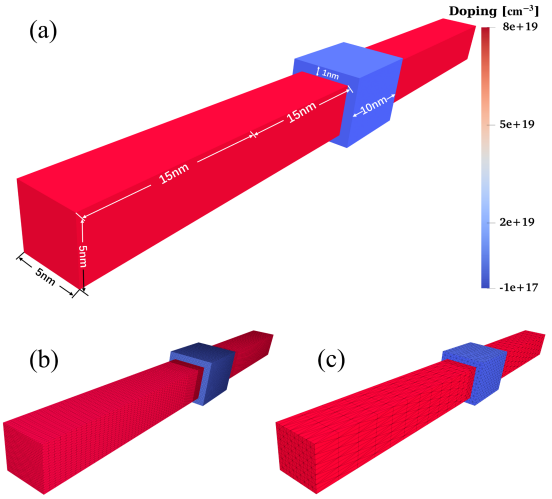

Fig. 1. (a) Structure of the n-type single channel GAAFET. (b) Tetrahedral mesh partitioning with 2,734,080 elements. (c) Tetrahedral mesh partitioning with 42,720 elements.

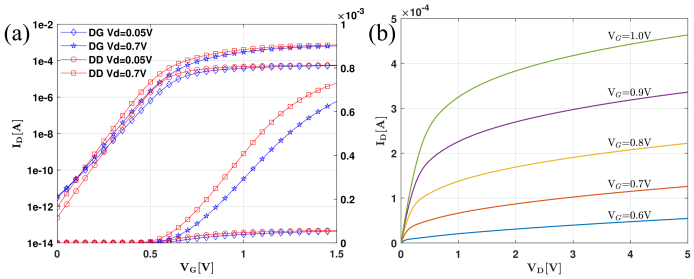

Fig. 2. (a) I_D-V_G characteristics obtained with the DG and DD model at $V_D = 0.05$ V, 0.7 V. (b) I_D-V_D characteristics obtained with the DG model at different V_G.

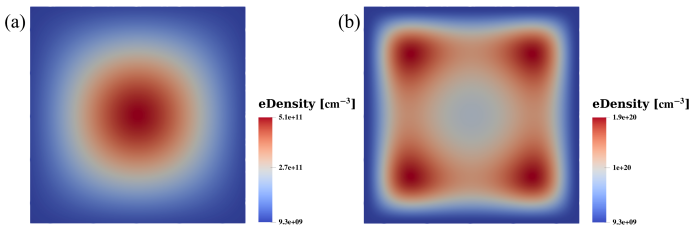

Fig. 3. Electron density along a cross-section perpendicular to the channel: (a) $V_D = 0.05$ V, $V_G = 0$ V. (b) $V_D = 0.05$ V, $V_G = 1.5$ V.

To demonstrate the stability of the discretization scheme and the efficiency of the linear iterative solver, we present the number of Newton iterations (with a relative tolerance, $rtol = 1e-5$) and the average number of FGMRES iterations (with a absolute tolerance, $atol = 1e-15$) at each bias step (fixed 0.05 V) in Fig. 4 and Fig. 5. As illustrated in these figures, the number of iterations does not significantly increase as the drain bias V_D ramped. Figure 6 shows the electron density along a line across the channel, utilizing meshes successively refined from a coarse mesh with 42,720 tetrahedra (refer to Fig. 1(c)). This figure demonstrates that the numerical solution has converged, indicating that the results are accurate, reliable, and independent of the discretization parameters. Figure 7 depicts the strong scalability of our MPI-based parallel simulator in obtaining the I_D-V_G curve, showcasing the effectiveness of our approach in large-scale simulations.

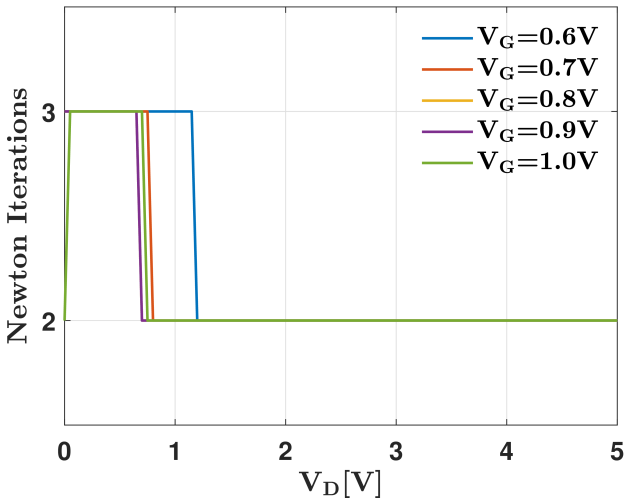

Fig. 4. Numbers of Newton iterations for increasing V_D and V_G. Setting relative error $rtol = 1e-5$, the convergence criterion is reached after $2 \sim 3$ iterations.

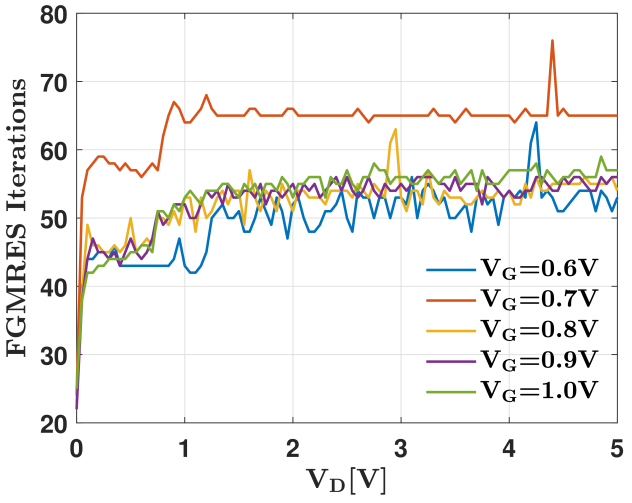

Fig. 5. Average number of FGMRES iterations in a Newton step for increasing V_D and V_G (setting absolute error $atol = 1e-15$).

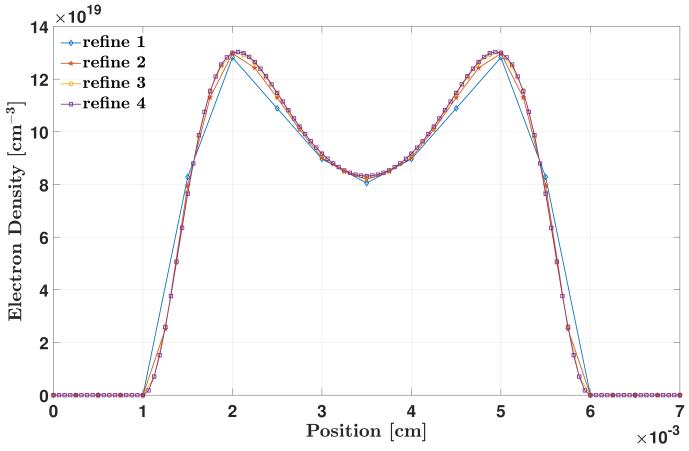

Fig. 6. Electron density along a cross-line perpendicular to the channel under different mesh refinements.

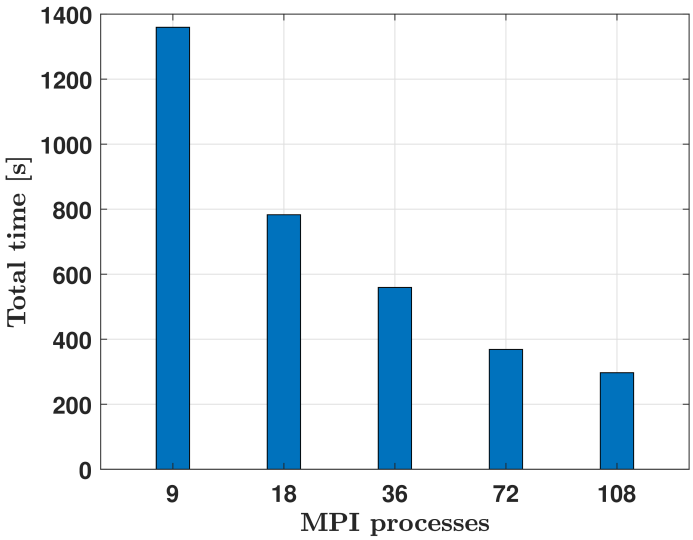

Fig. 7. Total simulation time required to obtain the I_D-V_G characteristic under different MPI processes (2,734,080 tetrahedra).

B. Simulation of a Silicon Nanoslab (NS) FET

To validate the correctness and efficiency of our program, we selected a 3nm technology node three-dimensional NMOS silicon nanoslab FET provided by Sentaurus [12] as a benchmark example, as shown in Fig. 8(a). We use the same mesh partitioning, comprising a total of 2,592,483 tetrahedral elements, as illustrated in Fig. 8(b). The relative errors of the I_D-V_G curves at drain biases of 0.05 V and 0.7 V were 12% and 13%, respectively, as depicted in Fig. 9(a). Figure 9(b) presents the time taken to compute the I_D-V_G curve using our program and Sentaurus for different processes at $V_D = 0.05$ V. The comparison of electron density with Sentaurus in the OFF and ON states across the channel is depicted in Figs. 10(a) and 10(b), demonstrating the accuracy and reliability of our program in capturing the key characteristics of the device.

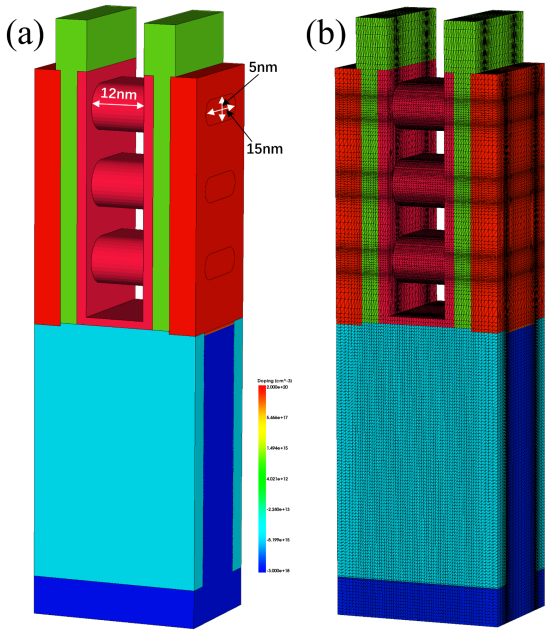

Fig. 8. (a) Structure of the n-type NSFET provided by Sentaurus [12]. (b) Tetrahedral mesh partitioning with 2,592,483 elements.

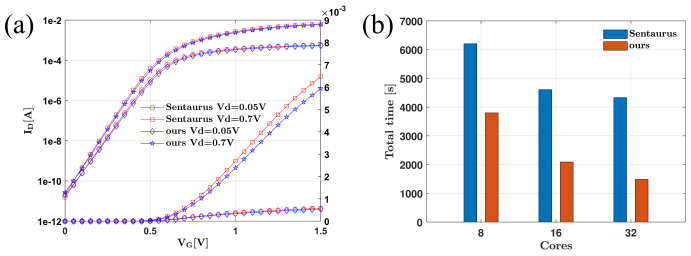

Fig. 9. (a) I_D-V_G characteristics obtained from our program and Sentaurus Device. (b) Total simulation time of the I_D-V_G characteristic under different processes.

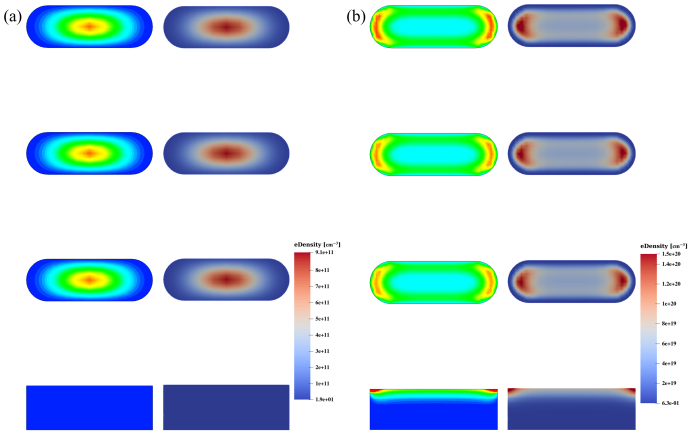

Fig. 10. Electron density along a cross-section perpendicular to the channel: (a) our method (right) vs. Sentaurus (left) at $V_D = 0.7\,\text{V}$, $V_G = 0\,\text{V}$. (b) our mehtod (right) vs. Sentaurus (left) at $V_D = 0.7\,\text{V}$, $V_G = 1.6\,\text{V}$.

IV. Conclusion

The FVSG scheme has long been recognized as the optimal computational format for simulating semiconductor devices using DD and DG models, with no finite element computational format known to rival it. In this work, we propose a stable and efficient finite element framework for solving the DG model, which is parameter-free and is also applicable to the DD model. Our designed Newton-Krylov solver significantly reduces the simulation time for each Newton iteration, even with two million elements, to the order of 10 seconds. This advancement demonstrates the framework's robustness and efficiency, offering a viable alternative to the traditional FVSG scheme.

Acknowledgment

This work was supported by the Strategic Priority Research Program of Chinese Academyof Sciences under Grant XDA0330401. Tao Cui was supported partially by National Key R & D Program of China 2019YFA0709600 and 2019YFA0709602.

References

[1] M. G. Ancona and G. J. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor," *Phys. Rev. B*, vol. 39, pp. 9536–9540, May 1989. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.39.9536

[2] C. S. Rafferty, B. Biegel, Z. Yu, M. G. Ancona, J. Bude, and R. W. Dutton, "Multi-dimensional quantum effect simulation using a density-gradient model and script-level programming techniques," in *Simulation of Semiconductor Processes and Devices 1998*, K. De Meyer and S. Biesemans, Eds. Vienna: Springer Vienna, 1998, pp. 137–140.

[3] D. J. Cummings, M. E. Law, S. Cea, and T. Linton, "Comparison of discretization methods for device simulation," in *2009 International Conference on Simulation of Semiconductor Processes and Devices*, 2009, pp. 1–4.

[4] N. Xu, Z. Jiang, P.-S. Lu, Y. H. Chang, J.-T. Li, K.-C. Ong, Z.-R. Xiao, and J. Wu, "Large-scale TCAD simulations enabled by hardware-accelerated high performance computing," in *2021 IEEE International Electron Devices Meeting (IEDM)*, 2021, pp. 18.1.1–18.1.4.

[5] S. Odanaka, "Multidimensional discretization of the stationary quantum drift-diffusion model for ultrasmall MOSFET structures," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 23, no. 6, pp. 837–842, 2004.

[6] A. J. Garcia-Loureiro, N. Seoane, M. Aldegunde, R. Valin, A. Asenov, A. Martinez, and K. Kalna, "Implementation of the density gradient quantum corrections for 3-d simulations of multigate nanoscaled transistors," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 6, pp. 841–851, 2011.

[7] S. C. Brenner, *The Mathematical Theory of Finite Element Methods*, 2008.

[8] J. Xu and L. Zikatanov, "A monotone finite element scheme for convection-diffusion equations," *Mathematics of Computation*, vol. 68, no. 228, pp. 1429–1446, 1999.

[9] J. Xu and L. T. Zikatanov, "Algebraic multigrid methods," *Acta Numerica*, vol. 26, pp. 591–721, 2017.

[10] Y. Saad, *Iterative methods for sparse linear systems*. SIAM, 2003.

[11] Y.-C. Wu and Y.-R. Jhan, *3D TCAD simulation for CMOS nanoeletronic devices*. Springer, 2018.

[12] "Coupled Sentaurus Device-Sentaurus Device QTX simulations of 3 nm silicon three-stack nanoslab FETs," *available from TCAD Sentaurus Version U-2022.12 installation, Applications_Library/AdvancedTransport/NSFET_SdeviceSBTE_3nm*.