Gate Oxide Scaling and High-κ Dielectric Integration in Bulk MOSFETs: A Sentaurus TCAD Study

L S Sankalana, H V Ranasinghe, M K Jayananda, and L S Liyanage

Department of Physics, Faculty of Science, University of Colombo, Sri Lanka

luckshitha@phys.cmb.ac.lk

1. ABSTRACT

Transistor scaling is crucial for optimizing device performance and improving power efficiency. However aggressive scaling can compromise electrostatic integrity and exacerbate short-channel effects. This paper presents a simulation study of n-channel bulk MOSFET scaling using Synopsys Sentaurus TCAD in which the gate length (L_g) was scaled across four gate-stack dielectric configurations; baseline device with 1.2 nm physical SiO_2, scaled variants with 0.9 and 1.5 nm SiO_2, and a high-κ stack with an Equivalent Oxide Thickness of 0.9 nm. Oxide scaling gives rise to a monotonic increase in I_{on} and I_{off} while V_{th} exhibits a decrease indicating a progressive loss of gate control. High-κ devices result in better V_{th} roll-off trend and lower I_{off} compared to the SiO_2 devices. A fundamental structural limitation was identified where scaling L_g below 20 nm results in negative threshold voltages. In conclusion high-κ dielectrics extend the viability of the bulk MOSFET while smaller gate lengths demand alternative device architecture to maintain effective switching capability.

2. INTRODUCTION

The continuous scaling of the Metal Oxide Semiconductor Field Effect Transistor (MOSFET) has been the primary driver of performance improvements in digital electronics and integrated circuits for decades [1]. The scaling of MOSFET dimensions, the gate length and the gate oxide thickness (T_{ox}) has been used to enhance electrostatic control and drive current. For many years, Silicon Dioxide (SiO_2) served as the ideal gate dielectric due to its large bandgap and high-quality interface with silicon. However, as transistor scaling pushed SiO_2 thickness below 2.0 nm, fundamental physical limits emerged. At a SiO_2 thickness of 1.2 nm, direct quantum mechanical tunneling from gate to channel led to exponential increase in gate leakage current, resulting in unacceptable standby power dissipation [2].

To overcome this "tunneling wall" and maintain the capacitance scaling required for performance, the industry has shifted toward high permittivity (high-κ) dielectrics, such as Hafnium Oxide (HfO_2) and Zirconium Oxide (ZrO_2). These materials allow a physically thicker gate insulator layer thereby suppressing leakage current while achieving a smaller Equivalent Oxide Thickness (EOT) for higher capacitance [3]. While high-κ integration effectively reduces leakage and improves Short Channel Effects (SCEs), it introduces new

trade-offs, including carrier mobility degradation due to remote phonon scattering and interface trap formation[4], [5]. Consequently, optimizing the balance between physical oxide thickness, material choice, and EOT is critical for minimizing off current (I_{off}) and Subthreshold Swing (SS) while maximizing on current (I_{on}) for optimal performance.

In this work, we present a comparative simulation study of bulk MOSFET scaling using Synopsys Sentaurus TCAD. Even though the state-of-the-art CMOS logic has recently adapted FinFETs and 3D architecture, bulk MOSFET technology is still widely used in analog, RF, IoT and power applications [6], [7], [8]. In addition, it is the most heavily studied canonical structure that is utilized as a reference or benchmark to study short channel effects effectively. Thus, in this study we focused on understanding the electrostatics of the scaled bulk MOSFET that would be a foundation for future advanced architectures.

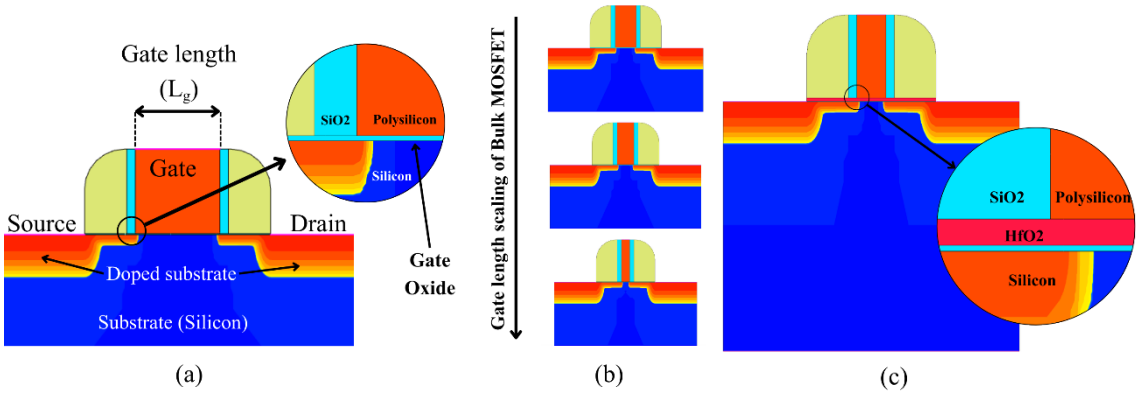

Figure 1: (a) Simulated bulk MOSFET structure with the inset highlighting the section of gate oxide positioned between polysilicon gate and the Silicon substrate (b) Illustration of the gate length scaling procedure (c) Introduction of high-κ/SiO_2 stack replacing the conventional SiO_2 gate oxide layer.

We investigate the electrostatic integrity and transport characteristics of a baseline device with a 1.2 nm SiO_2 gate dielectric against scaled variants (0.9 nm and 1.5 nm) and a high-κ stack with an EOT of 0.9 nm. By analyzing key performance metrics V_{th}, I_{on}, I_{off} and Subthreshold Swing across varying gate lengths (L_g), we presented the performance benefits of high-κ dielectric integration in scaling bulk silicon transistors. Furthermore, the I_{on} and I_{off} plots were analyzed for the 0.9 nm EOT device families to understand the benefits of the high-κ stack transition.

3. THEORY (PHYSICAL MODELS)

The device simulation was carried out using Sentaurus Device Simulations software. The software mainly utilizes the drift-diffusion transport model to solve the Poisson equation coupled with the continuity equations for electrons and holes [9].

Poisson Equation

$$\nabla \cdot \left(\epsilon \nabla \phi - \vec{P}\right) = -q(p - n + N_D - N_A) - \rho_{trap} \quad (1)$$

In this equation ϵ represents the electrical permittivity, ϕ is the electrostatic potential, and $\vec{P}$ is the polarization vector. The term ρ_{trap} accounts for the charge density of traps and fixed charges. Carrier densities are denoted by n (for electron concentration) and p (hole concentration), while N_D and N_A represent the ionized donor and acceptor concentrations respectively.

Electron and Hole Continuity Equations

$$\nabla \cdot \overrightarrow{J_n} = q\left(R_{\text{net},n} - G_{\text{net},n} + \frac{\partial n}{\partial t}\right) \tag{2}$$

$$-\nabla \cdot \overrightarrow{J_p} = q\left(R_{\text{net},p} - G_{\text{net},p} + \frac{\partial p}{\partial t}\right) \tag{3}$$

In the equations above, $\overrightarrow{J_n}$ and $\overrightarrow{J_p}$ are the electron and hole current densities, q is the elementary charge, R_{net} is the net recombination rate and G_{net} is the net generation rates for electrons and holes [9].To simulate submicron level physics, the simulation incorporates Bandgap Narrowing (OldSlotboom method) to adjust intrinsic carrier concentration in highly doped regions. Mobility Degradation is modeled using the Enhanced Lombardi (Enormal) model (accounting for surface roughness and acoustic phonon scattering) and the Canali model (for high-field velocity saturation). Additionally, quantum confinement effects in the thin gate oxide are treated using the Density Gradient model, which introduces a quantum potential correction to the transport equations [9].

To study the electrostatic characteristics of high-permittivity (high-κ) dielectrics relative to conventional SiO_2, the Equivalent Oxide Thickness (EOT) has been utilized. The bilayer EOT definition is as follows.

$$\text{EOT}_{\text{stack}} = t_{SiO_2} + t_{\text{high-}\kappa} \cdot \frac{\kappa_{SiO_2}}{\kappa_{\text{high-}\kappa}} \tag{4}$$

EOT is defined as the thickness of a theoretical SiO_2 layer that would yield the same gate capacitance per unit area as the actual dielectric stack [10]. The equation explains that by inserting a higher permittivity (κ-value) oxide we could obtain a physically thicker layer (thereby suppressing leakage) while maintaining the same EOT that results in the same gate to channel capacitance.

4. DEVICE STRUCTURE AND SIMULATION METHODOLOGY

4.1 Device Architecture and Design

The device studied is a symmetrical n-channel bulk MOSFET designed using Synopsys Sentaurus Structure Editor. It was designed by first creating a "half-FET" geometry which was then reflected along the vertical axis to obtain the full structure. This ensures perfect symmetry between structure components and keeps symmetric meshing in both sides.

The total substrate thickness of the device was set to 300 nm to ensure sufficient bulk depth. The rest of the dimensions for the structure are as follows: Polysilicon gate has a thickness of

100 nm and Nitride spacers (Si_3N_4) have a lateral width of 60 nm. To simulate realistic process conditions, the spacers were modeled with a rounded corner profile with a fillet radius of 50 nm. The source and drain kept at 100 nm lateral width while the gate length (L_g) served as the primary variable for characterization and it was changed from 50 nm to 15 nm with 5 nm steps. The baseline MOSFET device has a SiO_2 gate oxide thickness of 1.2 nm. It was later varied systematically from 0.9 nm to 1.5 nm in order to understand the impact of the oxide scaling on transistor properties. Finally, SiO_2/high-κ gate stack with an EOT of 0.9 nm and a total physical oxide thickness of 2.7 nm was included in TCAD simulations. This allowed us to directly compare the performance of high-κ gate stack against the ultra-thin (0.9 nm) SiO_2 bulk device for a range of channel lengths.

The device has a uniformly doped p-type silicon substrate with a Boron concentration of $1.5 \times 10^{18}\text{cm}^{-3}$. Highly n-doped regions were formed using Arsenic for source and drain regions with a peak concentration of $6 \times 10^{20}\text{cm}^{-3}$. These regions utilize a Gaussian profile with a junction depth (X_j) of 50 nm and a lateral rolloff factor of 0.4. To suppress hot-carrier effects and manage electric fields, shallow source/drain extensions were included. These follow a Gaussian Arsenic profile with a peak concentration of $2 \times 10^{20}\text{cm}^{-3}$ and a significantly shallower junction depth of 12.5 nm ($0.25 \times X_j$). The polysilicon gate is highly n-doped (Arsenic) with a concentration of $1 \times 10^{20}\text{cm}^{-3}$.

To ensure numerical convergence and accurate capture of quantum confinement effects, a mesh refinement strategy was applied. The grid resolution in the channel region was restricted to 1 nm. Furthermore, interface refinement was applied at the Si/SiO_2 boundary with a resolution of 0.1 nm to accurately solve the physics in interfaces.

4.2 Simulation setup and Parameter Extraction

2D numerical simulations were performed using Synopsys Sentaurus Device. The transfer characteristics ($I_d - V_{gs}$) were obtained by ramping the gate voltage in the linear (drain bias of $V_{ds} = 0.1$ V) and saturation ($V_{ds} = 1.1$) regions. Key performance metrics were extracted using Synopsys Sentaurus Visual. Threshold voltage (V_{th}) was determined using the transconductance method, I_{on} was extracted at gate supply voltage of $V_{gs} = 1.1\,V$ using saturation curve data and I_{off} was measured at $V_{gs} = 0\,V$ using linear curve data. The subthreshold swing (SS) was calculated as the inverse slope of the $I_d - V_{gs}$ curve in the subthreshold regime.

5. RESULTS AND DISCUSSION

In this study we systematically studied critical transistor device properties such as I_{on}, I_{off}, subthreshold slope and V_{th} with respect to gate length and gate oxide. Each marker in **Figure 2** and **Figure 4** represents the results of a bulk MOSFET structure with a unique gate length and oxide thickness.

The effects on I_{on} and I_{off} when gate width scales down are presented in **Figure 2** (a) and (b). To understand the gains and compromises of all device families, the I_{off} vs I_{on} characteristics plotted in **Figure 2** (c). When the gate length decreases both I_{off} and I_{on} will be increased, and when EOT of SiO_2 devices increases I_{on} will decrease. The better performing devices are the

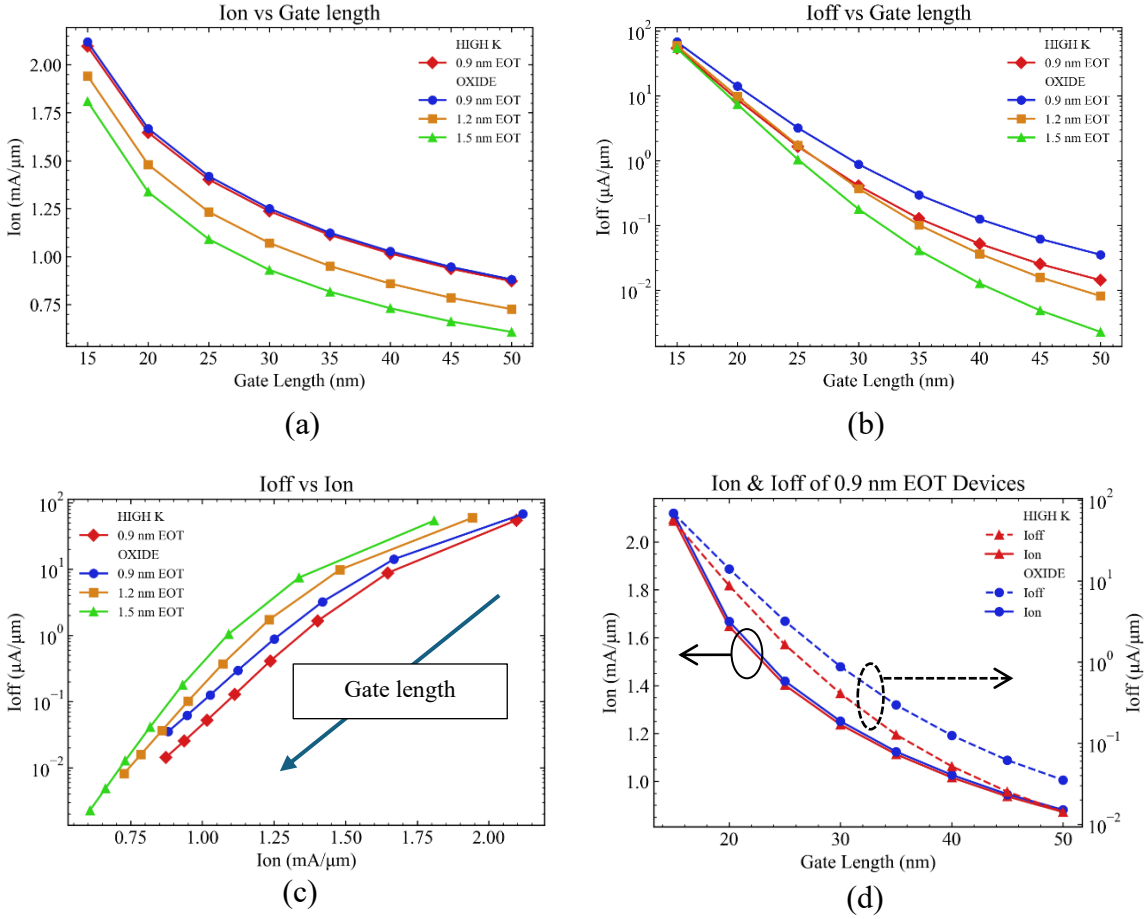

Figure 2: Bulk MOSFET characteristics of the four device families (a) I_{on} vs gate length (b) I_{off} vs gate length (c) I_{off} vs I_{on} of all device families (d) Comparison of 0.9 nm EOT high-κ device family and 0.9 nm EOT SiO_2 device family.

ones with lower I_{off} and a higher I_{on}, comparing overall characteristics high-κ device family has given better results.

A significant reduction in I_{off} is observed for the high-κ device compared to the 0.9 nm SiO_2 variant as depicts in the **Figure 2** (d). Although both devices have the same EOT, the high-κ

stack utilizes a physically thicker layer, which suppresses the direct tunneling component of the gate leakage while maintaining high I_{on}.

To illustrate gate control failure when scaling, cross sectional electron current density contours of the off state were extracted at the of $L_g = 15$ nm and $L_g = 50\,nm$ as shown in the **Figure**

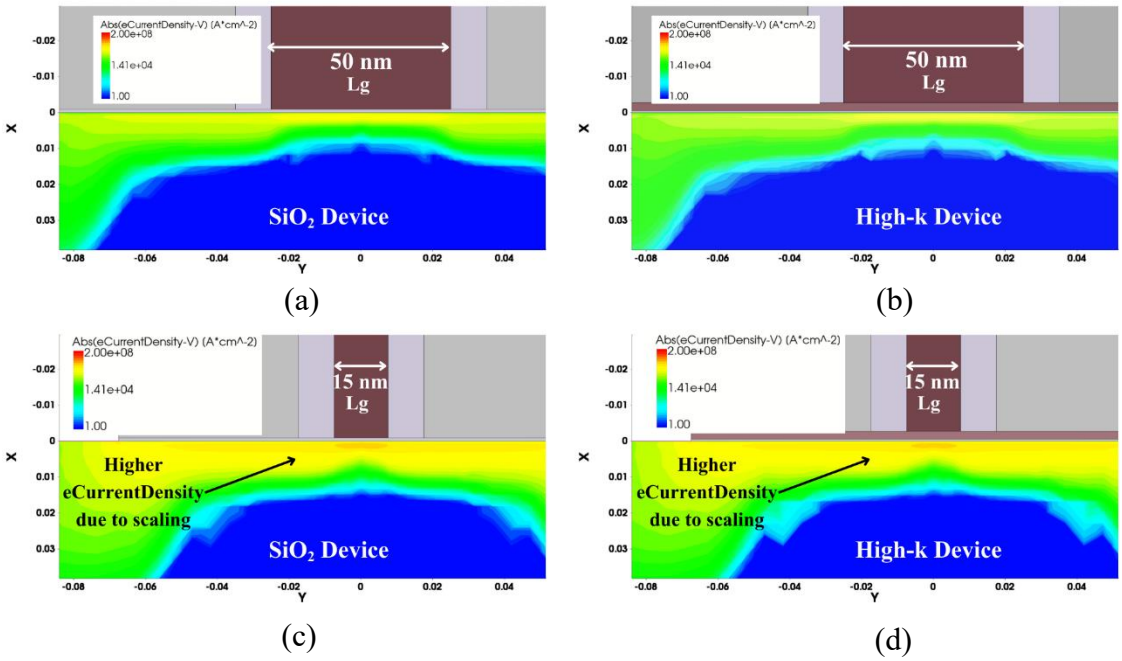

Figure 3: Off state absolute electron current density (eCurrentDensity) of 0.9 nm EOT devices. Smaller channel length devices have larger electron current density at off state. (a) 50 nm oxide device with lower leakage (b) 50 nm high-κ device with lower leakage (c) Scaled 15 nm oxide device with higher leakage (d) Scaled 15 nm high-κ device with higher leakage

3. When comparing (a) with (c) and (b) with (d) broadening of the channel (yellow region) was observed for scaled 15 nm devices with higher electron current density at off state. Additionally, at 15 nm the difference of I_{off} in oxide (c) and high-κ (d) devices is not significant. This can be observed in **Figure 2** (d) as L_g scales the difference between the I_{off} of two device families decreases (red and blue dashed lines).

The impact of gate oxide scaling on electrostatic integrity is presented in **Figure 4** (a), which plots the V_{th} as a function of L_g. For all device family variations, decrease in V_{th} is observed as L_g scales down, indicating the progressive onset of short channel effects (SCEs) that are detrimental to gate control and transistor functionality. The 1.5 nm SiO_2 device family exhibits the steepest degradation trend of V_{th}, while the high-κ family exhibits relatively low steep which implies applying high-κ materials provide better gate control and a positive impact on

transistor scaling. At and below 20 nm the V_{th} for this bulk structure drops below 0 V. This implies that the devices have a complete breakdown in electrostatic control.

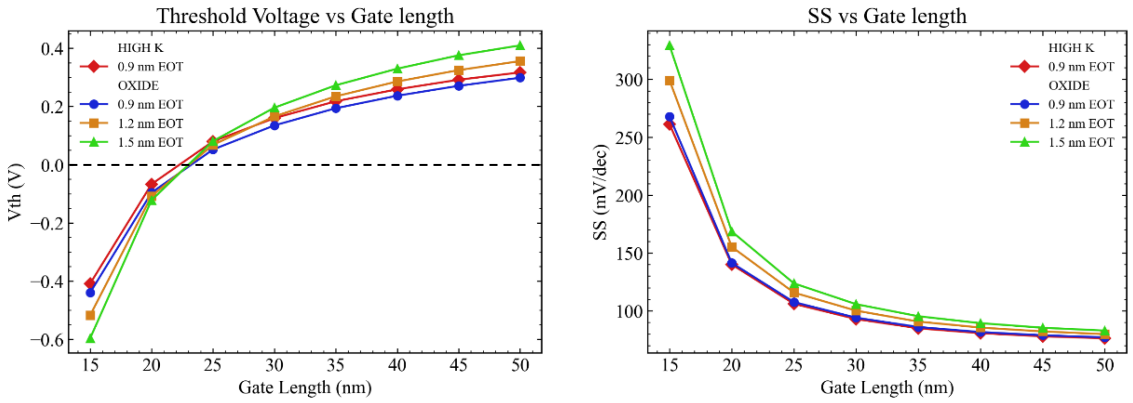

Figure 4: (a) Threshold voltage (using transconductance method) vs Gate length (b) Sub threshold swing vs Gate length, of four Bulk MOSFET device families.

Figure 4 (b) compares the Subthreshold Swing (SS) across the simulated device families. Ideally, SS should remain close to the theoretical limit of 60 mV/dec for faster switching speeds. The simulation results indicate that the high-κ dielectric stack (EOT = 0.9 nm) achieves the smallest SS, comparable to the physical 0.9 nm SiO_2 device, maintaining similar characteristics throughout. This confirms that the high-κ stack successfully replicates the capacitive control of the ultra-thin SiO_2 without the associated physical thinning limits.

6. CONCLUSION

In this study four device families with different gate stacks and channel lengths were compared to one another, and the effects of scaling have been shown using the standard MOSFET characteristics. While scaling down a particular device family, progressive loss of gate control was observed using the V_{th} roll off curves. Introduction of high-κ dielectric stack to gate oxide layer gives rise to noticeable improvement in I_{off} parameter by reducing leakage currents, while keeping I_{on} equivalent to the same EOT oxide devices. Considering all cases we conclude scaling transistor L_g beyond 20 nm for this specific device structure and specification is not feasible due to high leakage current at off state and loss of gate control. Therefore, bulk MOSFET while suitable for cost-effective applications, needs novel channel materials or advanced architectures for high performance logic that need scaling beyond 20 nm to sustain effective switching performance.

7. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Farazy Fahmy, Dr. Achintha Kondarage at Synopsys Sri Lanka and Dr. Sankalp Singh in Synopsys India, for providing access to the Sentaurus TCAD Educational License. Their assistance and guidance were essential for carrying out the device simulations in this work.

8. REFERENCES

[1] G. E. Moore, "Cramming More Components Onto Integrated Circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, Jan. 1998, doi: 10.1109/JPROC.1998.658762.

[2] J. Robertson and R. M. Wallace, "High-K materials and metal gates for CMOS applications," *Materials Science and Engineering: R: Reports*, vol. 88, pp. 1–41, Feb. 2015, doi: 10.1016/j.mser.2014.11.001.

[3] M. Salmani-Jelodar, H. Ilatikhameneh, S. Kim, K. Ng, P. Sarangapani, and G. Klimeck, "Optimum High-k Oxide for the Best Performance of Ultra-Scaled Double-Gate MOSFETs," *IEEE Trans. Nanotechnol.*, vol. 15, no. 6, pp. 904–910, Nov. 2016, doi: 10.1109/TNANO.2016.2583411.

[4] S.-H. Kuk *et al.*, "Channel Mobility With Higher-k Doped-HfO_2 for CMOS Logic," *IEEE Trans. Electron Devices*, vol. 71, no. 11, pp. 6534–6538, Nov. 2024, doi: 10.1109/TED.2024.3466843.

[5] R. Yadav, A. Kaushik, and K. Goyal, "Effect of Deposition of Different Dielectric Materials with Different Oxide Thickness on the Performance of Square Gate All Around MOSFET," in *2022 International Conference on Futuristic Technologies (INCOFT)*, IEEE, Nov. 2022, pp. 1–4. doi: 10.1109/INCOFT55651.2022.10094532.

[6] V. Subramanian *et al.*, "Planar Bulk MOSFETs Versus FinFETs: An Analog/RF Perspective," *IEEE Trans. Electron Devices*, vol. 53, no. 12, pp. 3071–3079, Dec. 2006, doi: 10.1109/TED.2006.885649.

[7] Y. S. Chauhan *et al.*, "BSIM6: Analog and RF Compact Model for Bulk MOSFET," *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 234–244, Feb. 2014, doi: 10.1109/TED.2013.2283084.

[8] N. Mendiratta and S. L. Tripathi, "A review on performance comparison of advanced MOSFET structures below 45 nm technology node," *Journal of Semiconductors*, vol. 41, no. 6, p. 061401, Jun. 2020, doi: 10.1088/1674-4926/41/6/061401.

[9] "Sentaurus™ Device User Guide, Version W-2024.09, Mountain View, California: Synopsys, Inc., 2024."

[10] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 4th ed. Hoboken, NJ: John Wiley & Sons, 2022.