A Novel Source/Drain-to-Gate Non-overlap MOSFET to Reduce Gate Leakage Current in Nano Regime

Ashwani K. Rana, Narottam Chand, Vinod Kapoor

Abstract—In this paper, gate leakage current has been mitigated by the use of novel nanoscale MOSFET with Source/Drain-to-Gate Non-overlapped and high-k spacer structure for the first time. A compact analytical model has been developed to study the gate leakage behaviour of proposed MOSFET structure. The result obtained has found good agreement with the Sentaurus Simulation. Fringing gate electric field through the dielectric spacer induces inversion layer in the non-overlap region to act as extended S/D region. It is found that optimal Source/Drain-to-Gate Non-overlapped and high-k spacer structure has reduced the gate leakage current to great extent as compared to those of an overlapped structure. Further, the proposed structure has improved off current, subthreshold slope and DIBL characteristic. It is concluded that this structure solves the problem of high leakage current without introducing the extra series resistance.

Keywords—Gate tunneling current, analytical model, spacer dielectrics, DIBL, subthreshold slope.

I. INTRODUCTION

LEAKAGE power has been recently recognized as a major challenge in the electronic industry. The desire to improve device performance has resulted in aggressive scaling of gate oxide thickness to below 2nm. At such oxide thicknesses, there occurs an increased probability of tunneling of charge carriers through the gate oxide[1]. This has resulted in an alarming increase in gate leakage current. Gate leakage is predicted to increase at a rate of more than 500X per technology generation [2]. Thus, reduction of gate leakage power dissipation is a primary concern of current research in the field of integrated circuits especially for low power battery-operated portable applications [3].

Numerous effective techniques for controlling gate leakage have been proposed in the past. The work in [4] presents an approach to reduce $I_{sub}$, but not $I_{gate}$. The impact of $I_{gate}$ on delay is discussed in [1], but its impact on leakage power is not addressed. In [5], the authors presented circuit-level techniques for gate leakage minimization. In each of these reports, extensive SPICE simulations were performed to obtain estimates of gate leakage. In [6], authors addressed various leakage mechanisms including gate leakage and presented circuit level technique to reduce the leakage. However, this can be extremely time-consuming, especially for large circuits. In [7], the authors examine the interaction between $I_{gate}$ and $I_{sub}$ and their state dependencies. This work applies pin reordering to minimize $I_{gate}$. In [8], Lee et al., developed a method for analyzing gate oxide leakage current in logic gates and suggested pin reordering to reduce it. Sultania et al., in [9], developed an algorithm to optimize the total leakage power by assigning dual $T_{on}$ values to transistors. In [10], Sutin et al. and Roy use multiple channel lengths and multiple gate oxide thickness for reduction of leakage. Mohanty et. al. [11] have presented analytical models and a data path scheduling algorithm for reduction of gate leakage current. In [12], conventional offset gated MOSFET structure has been widely used to reduce subthreshold leakage but gate leakage reduction has not been addressed in the literature so far. Thus, the general problems of gate leakage reduction techniques are the need for additional devices (e.g. sleep transistors) and the reduction of only one component of leakage. Moreover, transistor level approaches are not applicable for standard cell designs and require long calculation time. Further, gate level DVT-DTOCMOS methods do not offer the best possible solution as the number of gate types limits the improvement.

To solve this problem, we propose a novel source/drain-to-gate non-overlap nano CMOS device structure for the first time to reduce the gate leakage current effectively because gate leakage current through the source/drain overlap region has been identified as the principal source of power dissipation in VLSI chips especially in sub-1V range [13]. By adopting high-k dielectric spacers, we can induce low resistance inversion layer as a S/D extension region in the non-overlap region and report various results regarding proposed structure. An effective and compact model has been developed for analyzing the gate tunneling current of Gate to Source/Drain Non-overlap N-MOSFET by considering the NSE (nano scale effect) effect that are difficult to ignore at nano scale regime. The NSE effect include (i) the non-uniform dopant profile in poly-gate in vertical direction resulted due to low energy ion implantation, (ii) additional depletion layer at the gate edges due to gate length scaling down and (iii) gate oxide barrier lowering due to image.
charges across the Si/SiO₂ interface. We, also, adopted advanced physical models in the Simulation (sentaurus simulator) to see other device characteristics such as DIBL (drain induced barrier lowering), SS (subthreshold slope), on current and off current.

The rest of the paper is organized as follows. Section 2 establishes the gate current model. Section 3. concentrates on the device design of novel source/drain-to-gate non-overlap nano transistor. The simulation set up is presented in Section 4. The results and discussion are presented in Section 5. Finally, conclusions and directions for future work are summarized in Section 6.

II. THEORETICAL GATE CURRENT MODEL

Modeling of the direct tunneling current analytically has largely been based on the WKB approximation [14]. The discrepancies that were present in the original WKB approximation [14] have been rectified in [15] by introducing few adjusting parameters. In our work, we adopt this model to evaluate the direct tunneling current from channel and overlap region in nano scale regime where poly gate dopant non-uniformity, gate length effects and gate oxide barrier lowering due to image charges across the Si/SiO₂ interface cannot be ignored. This model is more consistent and covers the entire gate bias range. The robustness of the model has been verified with Sentaurus simulation data. Because of non overlap region between gate-source and gate-drain, the edge direct tunneling current is absent and hence total gate leakage current is given by

\[ I_G = I_{ge} = J_{ch} \times L_g \]  

(1)

where \(L_g\) is the total gate length. The channel current density \(J_{ch}\) are modeled as follows.

\[ J_{ch} = A C_{F(ch)} T_{WKB(ch)} \]  

(2)

where \(A = \frac{q^2}{8\pi\phi_{eff}^2}\) & \(C_{F(ch)}\) is the correction term incorporated in [15] and \(T_{WKB(ch)}\) is the modified WKB transmission probability given as follows

\[ C_{ch} = \exp \left[ \frac{20}{\phi_{eff}} \left( \frac{\phi_{off}}{\phi_{eff}} \right) \right] \]  

\[ T_{WKB(ch)} = \exp \left[ \frac{-8\pi\sqrt{2m_{eff}q\phi_{off}^2}}{3h\pi |E_{ox}|} \left( 1 - \frac{|V_{ox}|}{\phi_{off}} \right) \right]^{3/2} \]  

(3)

(4)

\[ N_{DTC(ch)} = \frac{e_{ox}}{L_{ox}} n_{acc} V_{t} \ln \left[ 1 + \exp \left( \frac{V_g - V_{FB}}{n_{acc} V_{t}} \right) \right] \text{ for } V_g < 0 \]  

\[ N_{DTC(ch)} = \frac{e_{ox}}{L_{ox}} n_{inv} V_{t} \ln \left[ 1 + \exp \left( \frac{V_g - V_{FB}}{n_{inv} V_{t}} \right) \right] \text{ for } V_g > 0 \]  

(5)

where \(\alpha_{ch}\) is the fitting parameter for channel tunneling and in this scheme, the value of \(\alpha_{ch}\) for channel region has been used as 0.73 to match the overall best fit with Sentaurus simulation. The \(T_{ox}\) refers to the physical oxide thickness and effective mass of the carrier in the oxide has been used as 0.40m₅, throughout this work. The \(n_{inv}\) and \(n_{acc}\) are the swing parameters, \(V_{FB}\) represents the flat band voltage, \(N_{DTC(ch)}\) denotes the density of carrier in channel region depending upon MOSFET biasing condition and \(V_{ge}\) is the effective gate voltage excluding poly gate non-uniformity and gate length effect and is equal to \(V_g - V_{poly}\) where \(V_{poly}\) is the voltage drop due to poly depletion layer in the poly-Si gate. The default values of \(n_{inv}\) and \(n_{acc}\) are \(\frac{S}{V_{t}}\) (S is the sub threshold swing and \(V_{t}\) is thermal voltage, equal to \(K_T/q\)) and 1 respectively. The voltage across the gate oxide for different region of operation is as follows

\[ V_{ox} = \begin{cases} \left( V_{ge} - \phi_s - V_{FB} \right) & \text{for } V_g < 0 \\ \left( V_{ge} - \phi_s - V_{FB} \right) & \text{for } V_g > 0 \end{cases} \]  

(6)

Where \(\phi_s\) is the surface band bending of the substrate and is calculated for channel region depending upon the biasing condition of the MOSFET including the poly non-uniformity, gate length effects and image force barrier lowering. The accurate surface potentials expressions in case of channel in weak inversion/depletion, strong inversion and in accumulation can be taken from [16]. The gate effective voltage including the effect of nonuniform dopant distribution in the gate is derived as follows

\[ V_{ge} = (V_{FB} + \phi_{so} - \Delta V_{p1} - \Delta V_{p2}) + \frac{q e_{pol} N_{poly} T_{g2}}{e_{gi}^2} \left[ \frac{2 e_{gi}^2}{q e_{gi} N_{poly} T_{g2}^2} \left( V_{ge} - V_{FB} - \phi_{so} \right) \right]^{1/2} \]  

\[ \text{for } V_{ge} \]  

(7)

By taking the quantization effect into account, \(\phi_{so}\) is as follows

\[ \phi_{so} = 2\phi_b + \Delta \phi_{OM} - V_{BS} \]  

(8)

where \(\Delta \phi_{OM}\) can be taken from [16]. The equation (7) includes the non uniformity in the gate dopant profile through a term \(\Delta V_{p1}\) and fringing field effect \(i.e\) gate length effect through a term \(\Delta V_{p2}\). The potential drop \(\Delta V_{p1}\) due to non uniform dopant profile in poly Si gate, caused by low energy implantation, is given by

\[ \Delta V_{p1} = \frac{kT}{q} \ln \frac{N_{poly \_top}}{N_{poly \_bottom}} \]  

(9)

The \(N_{poly \_top}\) and \(N_{poly \_bottom}\) are the doping concentration at top and bottom of the polysilicon gate. The potential drop \(\Delta V_{p2}\) due to gate length effect, caused by very short gate lengths is given as below

\[ \Delta V_{p2} \]
\[ \Delta V_{p2} = \frac{\Delta Q}{C_d} = \frac{2q4N_d}{L_gC_d} \left( \frac{\sqrt{V}}{cm} \right) \]

\[ C_d = \delta \frac{\varepsilon_{ox}}{\pi} \ln \left( \frac{3 - \cos \left( \frac{\pi}{2} \left( \frac{T_F - T_{ox}}{T_F} \right) \right)}{1 + \cos \left( \frac{\pi}{2} \left( \frac{T_F - T_{ox}}{T_F} \right) \right)} \right) \]

where \( A \) denotes the triangular area of the additional charge, \( L_g \) is the gate length, \( C_d \) is the depletion capacitance in the sidewalls [18], \( \varepsilon_{ox} \) is the permittivity of the gate oxide, \( T_F \) is the thickness of the field oxide, \( T_{ox} \) is the thickness of the gate oxide and \( \delta \) is fitting parameter equal to 0.95 normally.

III. DEVICE DESIGN

The N-MOSFET device cross-section with gate to source/drain non-overlap, designed for the analysis of the gate tunneling current characteristics, is shown in Fig. 1.

![N-MOSFET device cross-section](image)

The MOSFET has \( n^+ \) poly-Si gate of physical gate length \( L_g \) of 35 nm, gate oxide of 1.0 nm, buffer oxide of 1.0 nm under high-k spacer. The buffer oxide is used to minimize the stress between spacer and substrate. The HfO\(_2\), high-k dielectric, is used as spacer after optimization. Here \( L_{no} \) represents the non-overlap length between gate to source/drain. The source/drain extension region are created with the help of fringing gate electric field by inducing an inversion layer in the non overlap region. The halo doping around the S/D also reduces short-channel effects, such as the punch-through current, DIBL, and threshold voltage roll-off, for different non-overlap lengths.

III. SIMULATION SET UP

Fig. 2 shows the Sentaurus simulator schematic N-MOSFET which has non-overlapped gate to S/D (source/drain) region. The doping of the silicon S/D region is assumed to be very high, \( 1 \times 10^{20} \text{ cm}^{-3} \), which is close to the solid solubility limit and introduces negligible silicon resistance. The dimension of the silicon S/D region is taken as 40 nm long and 20 nm high. This gives a large contact area resulting in a small contact resistance.

![Sentaurus simulator schematic N-MOSFET](image)

The doping concentration in silicon channel region is assumed to be graded due to diffusion of dopant ions from heavily doped S/D region with a peak value of \( 1 \times 10^{19} \text{ cm}^{-3} \) and \( 1 \times 10^{17} \text{ cm}^{-3} \) near the channel. The poly-silicon doping has been taken to be \( 1 \times 10^{20} \text{ cm}^{-3} \) at the top and \( 1 \times 10^{18} \text{ cm}^{-3} \) at bottom of the polysilicon gate i.e. interface of oxide and silicon. The MOSFET was designed to have \( V_T \) of 0.23 V. We determined \( V_T \) by using a linear extrapolation of the linear portion of the \( I_{DS}-V_{GS} \) curve at low drain voltages. The operating voltage for the devices is 1V. The simulation study has been conducted in two dimensions, hence all the results are in the units of per unit channel width.

The simulation of the device is performed by using Sentaurus design suite [19], with drift-diffusion, density gradient quantum correction and advanced physical model being turned on.
magnitude of vertical electric field increases with the increase in dielectric constant of the spacer. The vertical electric field is responsible to induce an inversion layer in the non-overlap region. Result shows that approximately three times higher vertical electric field is obtained in the non-overlap region under HfO$_2$ high-k spacer compared to the oxide spacer. This implies that the on-state current of the high-k spacer non-overlapped gate to S/D MOSFET can be significantly larger than that of the oxide spacer MOSFET. This guides the use of compatible high-k spacer to induce the sufficient inversion layer in non-overlap region. It also shows that vertical electric field magnitude decreases significantly with the distance of non overlap region from the gate edge thereby limiting the non-overlap length ($L_{\text{no}}$).

Fig 4. plots the electron concentration along channel for such as HfO$_2$ (k=22), Si$_3$N$_4$ (k=7.5) and SiO$_2$ (k=3.9) in the non overlap region. Electron concentration below the spacer increases as the dielectric constant of the spacer material increases. This is due to the fact that electron concentration under the spacer strongly depends on the intensity of vertical fringing field, obvious from Fig 3. At $V_{gs}=1.0$ V, electron concentration more than $3 \times 10^{19}$/cm$^3$ is induced under the spacer which can act as extended S/D region.

Fig 5. represents the variation of electron concentration along the channel direction for different $V_{ds}$ from 0.0 V to 1.2 V with a 0.2 V step and at $V_{ds}$ of 0.05 V. It is observed from the Fig 5. that electron concentration below the spacer increases as the gate bias increases. This figure shows that gate bias controls inversion layer effectively by the gate fringing field.

Thus, a reasonable amount of electron concentration was induced for HfO$_2$ spacer with $L_{\text{no}}$ less than 8 nm. $L_{\text{no}}$ was also optimized with DIBL and subthreshold slope (SS) (as shown in Fig. 9) and found to be 5 nm.

Fig. 4. Electron concentration along channel for different spacer in the non overlap region

V. RESULTS AND DISCUSSION

Computation has been carried out for a n-channel nanoscale non overlapped gate to S/D MOSFET to estimate the gate tunneling current. This model is computationally efficient and easy to realize.

The comparison between the simulated data and the model value for gate tunneling current is presented in Fig. 6. The figure shows the gate tunneling current versus gate bias for non overlapped gate to S/D MOSFET with HfO$_2$ spacer above the non-overlap region at an equivalent oxide thickness (EOT) of 1 nm and non-overlap length of 5 nm. It is shown in Fig. 6 that result calculated by our model has better agreement with the simulated results certifying the high accuracy of the proposed analytical modelling.

Fig. 5. Electron concentration along channel for different $V_{gs}$ from 0.0 V to 1.0 V with a 0.2 V step and at $V_{ds}$ of 0.05 V

Fig. 6. Comparison of analytical model data with Sentaurus simulated data for HfO$_2$ based high-k spacer MOSFET with equivalent oxide thickness (EOT) of 1.0 nm, physical gate length of $L_g=35$ nm and S/D to gate non overlap length of $L_{\text{no}}=5$ nm.

Fig. 7 shows the variation of the gate tunneling current with gate bias for different non overlap length with HfO$_2$ spacer.

Fig. 7. Gate tunneling current for different $L_{\text{no}}$
above non overlap region at an EOT of 1.0 nm. It is observed that gate leakage current decreases significantly with non overlap structure as compared to overlapped structure especially at low gate bias range. At low gate bias, channel is about to form so that gate leakage current is mainly due to carrier tunneling through gate to S/D overlap region. The gate to S/D overlap region is absent in our designed MOSFET, so gate tunneling (leakage) current is reduced to greater extent. However, at higher gate bias range the gate tunneling (leakage) current is mainly due to the carrier tunneling through the channel region to the gate. Due to this region, gate tunneling current is almost same for overlapped gate to S/D MOSFET structure and non overlapped gate to S/D MOSFET structure.

Fig. 8 plots the gate tunneling current vs gate bias with and without buffer oxide layer. It is observed that gate leakage current increases without buffer oxide because without buffer oxide vertical E-field of the region is slightly enhanced.

Fig. 9 shows the variation of DIBL and SS with gate to S/D non overlap length of a NMOSFET. It is observed in Fig 9 that DIBL is maximum for overlapped gate to S/D MOSFET structure. It is due to the fact that the effect of fringing field on channel is maximum. Due to this decrease in gate control, the drain electrode is tightly coupled to the channel and the lateral electric field from the drain reaches a larger distance into the channel. Consequently, this electrically closer proximity of drain to source gives rise to higher drain-induced barrier lowering (DIBL) in overlapped gate to S/D MOSFET structure. In non-overlapped gate to S/D MOSFET structure, DIBL improves as the gate to S/D non overlap length ($L_{no}$) increases up to 5.0 nm because lateral electric field from the drain reaches a smaller distance into the channel. This is due to increase in metallurgical gate length as compared to conventional MOSFET structure in the same physical gate length. For $L_{no}$ larger than 5 nm, DIBL degrades due to poor turn-on characteristics. The poor turn-on characteristics at low $V_{ds}$ (0.05 V) are degraded by the barrier peaks under HfO2 spacer (shown in Fig.4).

Fig. 7. Gate tunneling current vs gate bias for different gate to S/D non overlap length with EOT of 1.0 nm.

Fig. 8. Gate tunneling current vs gate bias with and without buffer oxide

Fig. 9. DIBL, SS vs gate to S/D non overlap length of a NMOSFET with EOT of 1.0 nm

It is also shown in Fig. 9 that subthreshold characteristic improves for non overlapped gate to S/D MOSFET structure from $L_{no}$ of about 0 nm to 5 nm as compared to overlapped gate to S/D MOSFET structure. This represents one of the positive aspect of non overlapped gate to S/D MOSFET structure. For $L_{no}$ larger than 5 nm, SS degrades because the channel area to be depleted by the vertical and fringing electric field from the gate becomes wide thereby increasing the depletion capacitance in the subthreshold equation. This increased depletion capacitance, in turn, degrades the SS. The result indicates that non overlapped gate to S/D non NMOSFET with the $L_{no}$ of 5 nm is reasonable to achieve very small DIBL of 27 mV/V and subthreshold slope of 80.75 mV/dec. and is capable to suppress the SCE.
The effect on non overlap length on off and on current is plotted in Fig.10. It is observed in Fig. 10 that on current slightly degrades with increase in non overlap length due to increase in threshold voltage ($V_{th}$). The off current decreases due to increase in threshold voltage and improved SS thereby resulting in reduced subthreshold leakage. The decrement in off current is more as compared to decrement in on current which in turn results in enhanced $I_{on}/I_{off}$ ratio greater than $3 \times 10^5$ at $L_{no}$ of 5nm.

VI. CONCLUSION

In this work, we have developed a simplified and compact analytical gate current model including nano scale effect (NSE) for a new 35 nm non-overlapped gate to S/D non NMOSFET structure. It is found that with a fixed metallurgical gate length of 35 nm, non-overlapped gate to S/D MOSFET structure have shown smaller gate tunneling current compared to overlapped gate to S/D MOSFET structure. The gate tunneling current also degrades slightly with non-overlap length. It was found that the gate to channel control ability of non-overlapped MOSFET structure can be significantly enhanced by using high-k HfO$_2$ spacer. It is also shown that non-overlapped MOSFET structure can be significantly reduced gate tunneling current also deceases slightly with non-overlap compared to overlapped gate to S/D MOSFET structure. The MOSFET structure have shown smaller gate tunneling current to greater extent in addition to suppress the SCE.

REFERENCES


Ashwani K. Rana was born in Kangra, H.P., India in 1974. He received his B.Tech degree in Electronics and Communication Engineering INIT, Hamirpur, India in 1998 M.Tech degree in VLSI Technology from Indian Institute of Technology, Roorkee, India in 2006. He is currently pursuing the Ph.D degree in nano devices from National Institute of Technology, Hamirpur, India.

Presently he is with Department of Electronics and Communication Engineering, National Institute of Technology, Hamirpur, India, as an Assistant Professor.

His research interest include modeling of semiconductor devices, low power high performance VLSI circuit design and emerging integrated circuit technologies. He has more than 10 publications in International/National Journal & conferences and guided more than 10 M.Tech students in these areas. He is member of ISTE.
Dr. Narottam Chand received his PhD degree from IIT Roorkee in Computer Science and Engineering. Previously he received MTech and BTech degrees in Computer Science and Engineering from IIT Delhi and NIT Hamirpur respectively. Presently he is working as Head, Department of Computer Science and Engineering, NIT Hamirpur. He also served as Head, Institute Computer Centre, NIT Hamirpur from February 2008 to July 2009. He has coordinated different key assignments at NIT Hamirpur like Campus Wide Networking, Institute Web Site, Institute Office Automation. His current research areas of interest include mobile computing, mobile ad hoc networks and wireless sensor networks. He has published more than 50 research papers in International/National journals & conferences and guiding six PhDs in these areas. He is member of ISTE, CSI, International Association of Engineers and Internet Society.

Dr. Vinod Kapoor was born at Mandi town in Himachal Pradesh, India. He received his BE Degree in Electronics & Communication Engineering from National Institute of Technology (formerly Regional Engineering College), Durgapur, Bengal, in the year 1987 and Masters Degree Electronics & Control from Birla Institute of Technology & Science, Pilani (Rajasthan) in the year 1996. He did his Ph.D. from Kurukshetra University, Kurukshetra in the field of Optical Fiber Communication in Sept.2006. He also obtained his MBA degree with specialization in Human Resources Management in the year 2002. He is presently Professor in the Department of Electronics & Communication Engineering NIT Hamirpur, India. He has coordinated different key assignments at NIT Hamirpur like Training & Placement Officer, Chief Warden (Hostels), organizing short term courses. His research interest includes optical fiber communication and optoelectronics/nano devices. He has published more than 15 research papers in International/National journal & conferences and guiding four PhDs in these areas. He is member of ISTE, IETE and IE.