Analog CMOS Performance Degradation due to Edge Direct Tunneling (EDT) Current in sub-100nm Technology

Navakanta Bhat and Chandrabhan Singh Thakur

Abstract— **We report the results of extensive mixed mode simulations and theoretical analysis to quantify the contribution of the edge direct tunneling (EDT) current on the total gate leakage current of 80nm NMOSFET with SiO2 gate dielectric. It is shown that EDT has a profound impact on basic analog circuit building blocks such as sample-hold (S/H) circuit and the current mirror circuit. A transistor design methodology with zero gate-source/drain overlap is proposed to mitigate the EDT effect. This results in lower voltage droop in S/H application and better current matching in current mirror application. It is demonstrated that decreasing the overlap length also improves the basic analog circuit performance metrics of the transistor. The transistor with zero gate-source/drain overlap, results in better transconductance, input resistance, output resistance, intrinsic gain and unity gain transition frequency.**

Index Terms— **reliability, CMOS direct tunneling, short channel, transition frequency**

I. INTRODUCTION

The aggressive scaling of CMOS devices has been accomplished by decreasing the oxide thickness to obtain high current drive and good control on short channel effect. Several limiting factors associated with the ultra thin gate oxides have been identified. Among them, the direct tunneling current is the most sensitive one to the oxide thickness [1]. For conventional CMOS devices, dominant leakage mechanism is mainly due to short channel effects owing to Drain-Induced Barrier Lowering (DIBL). However, in the sub-100nm technology regime with ultra-thin gate oxide, the gate leakage current can contribute significantly to the off state leakage [2]. Although alternate gate dielectrics are being explored to replace SiO_2, a viable alternative is yet to emerge for manufacturability [3-5]. In the meanwhile the SiO_2 thickness has been aggressively scaled down, in spite of leakage current, for high performance digital applications. In this context, it is important to characterize the effect of such aggressive scaling of oxide thickness on analog circuits for applications requiring the implementation of mixed signal circuits.

Recent studies have shown that the direct tunneling current appearing between the source-drain extension (SDE) and gate overlap, so-called the Edge Direct Tunneling (EDT), dominates the off-state current [6-8], especially in short channel devices. In this work we have looked at the scaling of gate to source/drain overlap length from gate leakage and analog circuit performance perspective. A commercial TCAD software from ISE [9] has been used for extensive mixed mode simulation work. In particular we have shown that the large gate-source/drain overlap is detrimental to analog circuits such as sample-hold and sub-threshold current mirrors. The MOS small signal equivalent circuit model is modified accordingly. It is demonstrated that the EDT has a significant impact on analog circuit performance. In order to mitigate the EDT problem, the transistor

Manuscript received August 3, 2003; revised September 2, 2003.
Department of Electrical Communication Engineering Indian Institute of Science, Banagalore-560012
E-mail : navakant@ece.iisc.ernet.in

design option with zero gate-source/drain overlap length is proposed. The decreasing overlap length not only improves the leakage related reliability problems in analog circuit, but also improves the basic analog circuit performance metrics of the MOS transistor.

In the next section we describe the effect of overlap length on total leakage current and the resulting small signal equivalent MOS model. Section III demonstrates the impact of EDT on sample-hold circuit and current mirror circuit. Section IV presents the effect of different transistor design on basic analog circuit performance metrics. Section V concludes by summarizing the important results.

II. Effect Of Gate Overlap Length On Total Gate Leakage Current Of Cmos Device

There are fundamentally two components of direct tunneling current. One is the Edge Component (EC) or Edge Direct Tunneling (EDT, tunneling of carriers between gate and source/drain overlap regions) and another one is Channel Component (CC) (tunneling of carriers between gate and channel region). EDT has been identified as the principal component of gate tunneling current.

Figure 1 shows the two components of tunneling in ultra-thin gate oxide NMOS structure. To explore the effect of gate overlap length on total gate leakage current, two device structures were simulated. One having 20nm gate to source/drain overlap length and the other having zero overlap region (0nm gate to source/drain overlap). The two devices were matched for subthreshold leakage current, by turning off the gate leakage model in the simulator. The subthreshold current matching was achieved by adjusting the pocket halo doping concentration. Figure 2 shows the I_g-V_g characteristics of two devices for oxide thickness of 15Å and 10Å. From figure 2 it becomes clear that reducing the gate to source/drain overlap length will certainly help in reducing total gate leakage current especially in the sub threshold regime. We have demonstrated earlier that the magnitude of the total gate leakage current is directly proportional to the length of the overlap region for sub

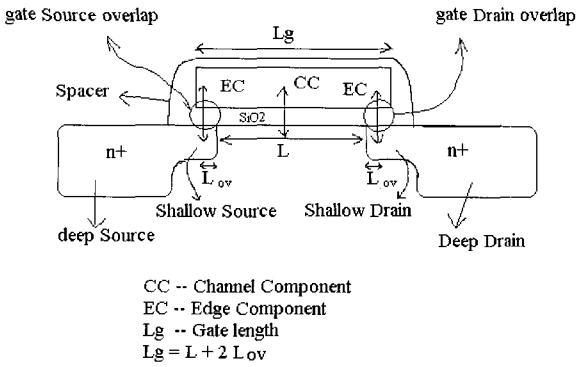


Fig. 1. Gate direct tunneling current components of a short-channel NMOSFET.

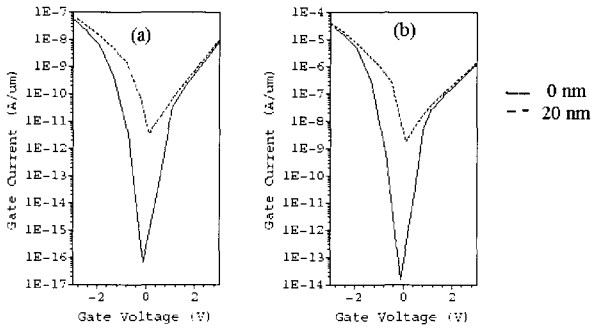


Fig. 2. I_g-V_g characteristics of NMOS with oxide thickness of (a) 15 Å and (b) 10 Å.

100nm gate length [8]. This effect is due to the fact that the flat-band voltage for the overlap region is almost zero whereas the flat-band voltage for channel region is negative. This results in higher vertical electric field for the overlap region when the transistor is off, thus resulting in more leakage current.

The equivalent small signal model for the transistor is shown in Fig. 3 in the presence of direct tunneling gate leakage. The gate leakage results in three additional conductance parameters namely g_{gs}, g_{gd} and g_{gb} which are dependent on their respective terminal voltages. When the transistor is in above threshold region, the g_{gs}, g_{gd} components dominate over g_{gb}. In the conventional transistor design with large gate-source/drain overlap, g_{gs}, g_{gd} components dominate over g_{gb} even in the sub threshold region. The transistor with zero overlap length, essentially retains only the g_{gb} in the equivalent circuit, which is very small. All these conductance values depend on DC bias voltages similar to the other small signal parameters.

The effect of these conductances should be appre-

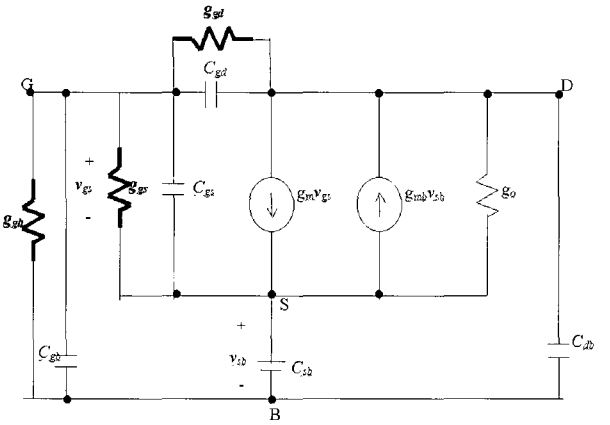


Fig. 3. Small signal equivalent circuit model in presence of gate leakage.

ciated, in terms of their effect on output (I_{ds}) and the input (R_{in}) terminal behaviour, depending on the different regions of operation. The drain current is given by

$$I_{ds} = I_{dso} + g_m v_{gs} + g_{gd} v_{dg} + g_o v_{ds} \qquad (1)$$

where I_{ds0} is DC bias current and the other voltages are small signal voltages. Notice that the effect of $g_{gd}v_{dg}$ term on the drain current becomes very prominent, especially in the subthreshold region. This will have a significant impact on sample-hold application as well as low power sub-threshold current mirror application. The effect of g_{gs} is to degrade the low frequency input impedance in applications such as OPAMPs. It should be further noticed that g_{gs} can result in an additional non linearity in the gain, when the amplifier is driven by a source with nonzero resistance. This is because g_{gs} itself will be dependent on input voltage.

III. Effect Of EDT On Analog Circuit Performance

3.1 Sample and Hold circuit

The circuit diagram of sample and hold circuit consisting of an NMOS pass-gate and a holding capacitor is shown in Fig. 4(a). DESSIS software is used to simulate the circuit. The gate leakage models including direct tunneling have been activated during simulation. When the control gate pulse is high, the input

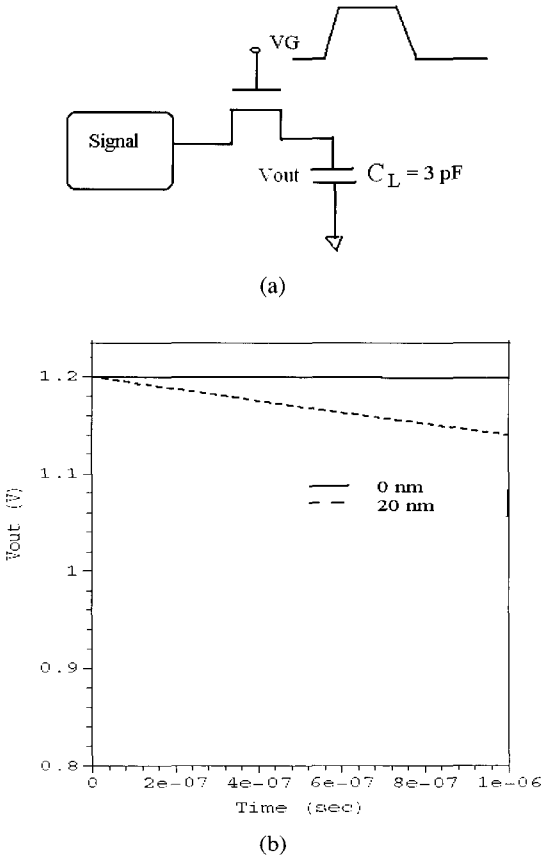


Fig. 4. (a) Sample and Hold circuit used in mixed mode simulation (b) Voltage droop in the hold mode.

signal value is passed to the holding capacitor at the output. When the gate pulse goes low, the capacitor should retain its charge and hence the sampled value of the input voltage. But the output voltage droops in the presence of gate leakage current since the charge stored in the capacitor will leak off. Further it should be noticed that during the hold phase, the biasing condition is such that the EDT current due to g_{gd} component is significantly enhanced. Figure 4(b) shows V_{out} versus time plot of the circuit, only during the holding phase. In the 20 nm overlap device, a voltage droop of 60.2 mV has been observed over 1 μs, whereas the 0 nm overlap device has a voltage droop of 2.1 mV. For an 8-bit A/D converter, the maximum allowed voltage droop is given by $\triangle/2$ where $\triangle = 1/2^8$, i.e. for V_{dd} of 1.2 V the allowed voltage droop will be 2.34 mV. Now if this sample hold circuit is used for 8-bit A/D converter application, the voltage droop will impose limitation on minimum frequency that can be used. A minimum frequency, f_{min} , of 27.58 MHz is possible with 20 nm overlap device

whereas f_{min} = 1 MHz for 0 nm overlap device.

3.2 Current mirror

The simple current mirror shown in Fig. 5 is used to illustrate the current matching issue. The sub-threshold operation of MOS transistors is very attractive for low power applications. The exponential non linearity between I_{ds} and V_{gs} is exploited in analog neural network applications [10]. In the absence of gate leakage and for the matched channel lengths, the output current of the current mirror is given by

$$I_o = I_R \frac{W_2}{W_1}[1 + \lambda(V_o - V_R)] \quad (2)$$

The deviation from the ideal mirroring is essentially due to channel length modulation parameter λ. For the sub-threshold operation, the impact of λ is not significant since the current saturates when the drain voltage is more than a few times the thermal voltage. In fact this characteristic results in excellent current mirroring for sub-threshold operation. However, in the presence of gate leakage current, the output current equation has to be modified

$$I_o = (I_R - I_{gs1} - I_{gs2})\frac{W_2}{W_1}[1 + \lambda(V_o - V_R)] + g_{gd2}(V_o - V_R) \quad (3)$$

This is because the part of the reference current I_R is lost through gate to source EDT of M1 and M2 and hence the I_{ds1} is essentially I_R-I_{gs1}-I_{gs2}. This current is reflected on to the drain of M2. However, there is an additional component due gate to drain EDT of M2. Hence the matching performance of the current mirror is drastically impacted.

IV. Impact Of Gate OverlAp scaling On Analog Performance Metrics

The discussion in the previous section motivates us to go for a transistor design with zero gate-source/drain overlap for reliable performance of analog circuits. In this section we investigate, the impact of such a transistor design on basic analog circuit performance

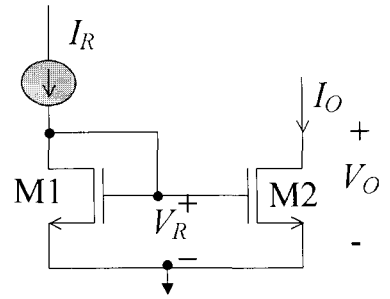


Fig. 5. Simple current mirror circuit operating in low power sub-threshold analog circuit.

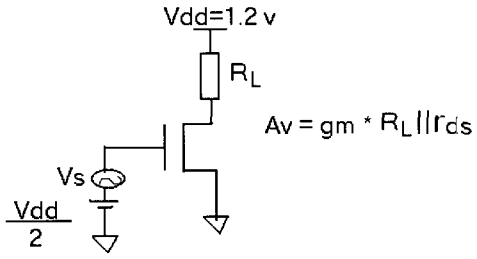


Fig. 6. Schematic of circuit used to calculate gm and A_{vi}.

Table 1. Comparison of analog performance metrics of two different transistor designs.

Overlap length	*Voltage gain (A_v)*	*Small signal output resistance (r_{ds})*	*Transconductance (g_m)*	*Intrinsic Voltage gain(A_{vi})*
0 nm	*3.935*	*35.33 K*	*$5.048x10^{-4}$ A/V*	*17.837*
20 nm	*2.755*	*18.71 K*	*$4.227x10^{-4}$ A/V*	*7.91*

metrics such as transconductance (g_m), output resistance (r_{ds}), low frequency voltage gain (A_{vi}), and transition frequency (f_t). We demonstrate that the zero overlap, in fact, helps in improving these metrics.

The circuit simulated is quite simple and is shown in Figure 6. The gate is biased at $V_{dd}/2$ to ensure that the device always remains in saturation. A sinusoidal signal V_s of frequency 1 KHz and peak-to-peak voltage of 0.2 V is applied. The load resistance R_L of 10 K has been used. The small signal voltage gain of the circuit has been calculated from the ratio of output signal V_0 to the input signal V_s. The small signal output resistance, r_{ds}, has been calculated from I_{ds}-V_{ds} characteristics of the devices under the same DC biasing conditions. Then g_m and A_{vi} were calculated using the formula $A_v = g_m(\frac{r_{ds}R_L}{r_{ds}+R_L})$, and $A_{vi} = g_m r_{ds}$. Table 1 summarizes the various circuit parameters extracted from simulation

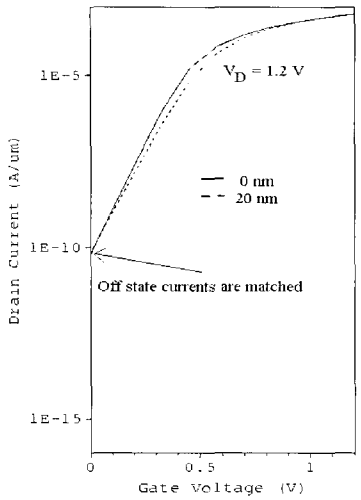


Fig. 7. Effect of mobility degradation and high DIBL on I_{ds}-V_{gs} characteristics.

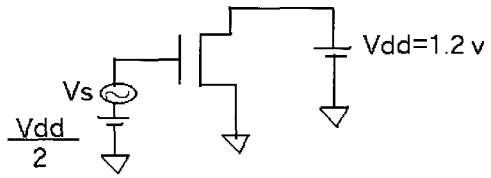


Fig. 8. Schematic of circuit used to calculate the f_t.

results.

It is interesting to note that the 0 nm overlap device has higher transconductance at this biasing condition. It is generally believed that decreasing the overlap length increases the parasitic source/drain resistance and hence decreases the drive current [11]. In fact the drain current of 0 nm overlap device is slightly higher for V_{gs}=V_{ds}=V_{dd} bias condition. However, the resistance effect has been more than compensated by the mobility degradation and DIBL effect at the V_{gs}=V_{dd}/2 biasing. In order to maintain the same sub-threshold leakage in two transistors for fair comparison, the halo doping of 20 nm overlap device is more than the 0 nm overlap device. Higher halo doping in 20nm device will increase the ion scattering and hence reduce the mobility of electrons and thus will result in less drive current. In addition, the higher DIBL results in higher linear V_t which also contributes to lower drain current. This is illustrated in Fig. 7. The small signal output resistance of 20 nm overlap device is less because its metallurgical gate length is actually 40 nm, that means short channel effects will be more in this device.

Transition frequency is defined as the frequency at which the short circuit current gain becomes unity, i.e. $\frac{I_d}{I_g}=1$ at f_t. The transition frequency is expressed as

$$f_t = \frac{g_m}{2\pi(C_{gs}+C_{gd})} \quad (4)$$

f_t depends on transconductance and internal capacitances of the device. The circuit simulated to find the f_t is shown in Figure 8. The transition frequency (f_t) of 0 nm overlap device is found to be 80 GHz in contrast to 65 GHz for 20 nm overlap device. This is because the 0 nm overlap device has both higher g_m and reduced overlap capacitance compared to the 20 nm overlap device.

V. Conclusions

The effect of EDT on the total gate leakage has been quantified for ultra thin gate oxides. The effect of gate to source/drain overlap on the EDT and the analog circuit performance is investigated. It is demonstrated that the S/H circuit with a 20nm overlap transistor results in large voltage droop. The current matching in current mirror application is also affected in the 20nm overlap device. Further, it is demonstrated that the zero nm overlap device outperforms the 20nm overlap device in terms of transconducatnce, out put resistance, intrinsic low frequency voltage gain and transition frequency. These are the basic performance metrics, which an analog circuit designer looks for. We suggest that, in order to build the reliable analog circuits using ultra thin tunneling SiO_2 gate oxides, the transistor design should be optimized to provide the lowest gate-source/drain overlap that is allowed by the process technology.

References

[1] D.A. Buchanan and S.-H. Lo, "Reliability and integration of ultra-thin gate dielectrics for advanced CMOS", *Microelectronic Engineering*, vol. 36, pp13-20, 1997

[2] N.Yang, W.K.Henson, and J.Wortman, "A comparative study of gate direct tunneling and drain leakage currents in N-MOSFETs with sub-100 nm gate oxides," *IEEE Trans. Electron Devices*, vol. 47, p.1636 (2000).

[3] G.D. Wilk, R.M. Wallace, J.M. Anthony, "High-k gate

dielectrics: current status and material properties considerations," *Journal of Applied Physics,* vol. 89, 5243 (2001).

[4] M.P. Singh, C.S. Thakur, K. Shalini, N. Bhat, and S.A. Shivashankar, "Structural and Electrical Characterization of Erbium Oxide Films Grown on Si(100) by Low-pressure Metalorganic Chemical Vapour Deposition", *accepted in Applied Physics Letters*

[5] K. Onishi, C.S.Kang, R. Choi, H.-J. Lo, S. Gopalan, R.E.Nieh, S.A.Krishnan, J.C.Lee, "Improvement of Surface Carrier Mobility of HfO2 MOSFETs by High Temperature Forming Gas Annealing", *IEEE Trans. Electron Devices,* vol. 50, pp.384-390, 2003.

[6] K.N. Yang, H.T. Huang, M.J. Chen, Y.M. Lin, M.C. Yu, S.M. Jang, Douglas, C.H. Yu, and M.S. Liang, "Characterization and modeling of Edge Direct Tunneling (EDT) Leakage in Ultrathin Gate oxide MOSFETs", *IEEE Trans. On Electron Devices*, vol.48, pp. 1159-1164, 2001.

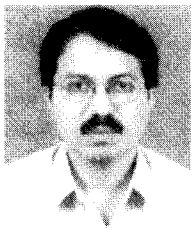
[7] K.N. Yang, H.T. Huang, M.J. Chen, Y.M. Lin, M.C. Yu, S.M. Jang, C. H. Yu, and M.S. Liang, "Edge hole direct tunneling in off-state ultrathin gate oxide p-channel MOSFETs," IEDM Tech Dig., 2000.

[8] K. Maitra and N. Bhat, "Analytical approach to integrate the different components of direct tunneling current through ultrathin gate oxides in n-channel metal oxide semiconductor field-effect transistors", *Journal of Applied Physics*, Vol 93, No. 2, 15 January 2003

[9] ISE-TCAD User Manuals, version 6.1, Integrated systems Engineering, Zurich.

[10] Amit Gupta and Navakanta Bhat, "Hardware realisation of a digitally controllable neuron activation function and its derivative for extremely low power applications", *6th IEEE VLSI Design & Test workshop-2002*

[11] K. Maitra and N. Bhat, "Impact of Gate to Source/Drain Overlap Length on 80 nm CMOS Circuit Performance", submitted to *IEEE Transactions on Electron Devices*



Navakanta Bhat received his M.Tech from IIT Bombay and Ph.D. from Stanford University. He is currently A/Professor in the ECE Department of Indian Institute of Science, Bangalore, India and responsible for the Device Analysis in the IISc. His areas of research include VLSI Design and processing and RF MEMS. He is a Senior Member of IEEE.