Extended Abstracts of the 2003 International Conference on Solid State Devices and Materials, Tokyo, 2003, pp. 434-435

P2-12

Design Issues for Sub-100nm Analog CMOS

Mayank Garg, Sushant S. Suryagandh and Jason C.S. Woo
Department. of Electrical Engineering
University of California, Los Angeles, CA 90095, USA.
woo@icsl.ucla.edu, Tel:(310)206-3279

1. Introduction

Continuous scaling of CMOS has resulted in cut-off frequencies (f_T) well above 50GHz [1, 2]. However, the improvement in f_T comes at the cost of degraded output resistance (R_{out}) and reduced intrinsic gain, $G=g_mR_{out}$, of the transistor. Fig. 1(a) shows the intrinsic gain vs. f_T trade-offs for various channel length (L_g) devices. Each point corresponds to a particular device's intrinsic gain and f_T at a bias condition of I_{ds}=100μA/μm (rather than at fixed gate over-drive) and V_{ds}=0.8V. For a given channel length, different points are obtained by varying the substrate doping (N_A) and the source/drain extension (SDE) depth, X_j. The overall gain vs. f_T trade-offs improve with scaling. However, if a constant threshold voltage (V_{th}) is maintained, the gain degrades considerably. Degradation of linearity with scaling is a major concern for many analog circuits. Linearity is another concern. $VIP_3 = \sqrt{0.75 g_m / g_{m3}}$ is a standard way of measuring the third order non-linearity in g_m, where $g_{m3} = \partial^3 I_{ds}/\partial V_{gs}^3$ [3]. Fig. 1(b) shows the effect of scaling L_g on overall VIP_3 vs. f_T trade-offs. Similar to the gain vs. f_T trade-offs, the improvement in VIP_3 vs. f_T trade-offs is limited by the maximum V_{th}. Thinner oxide thickness (T_{ox}) or smaller X_j can be used to reduce the degradation in gain. However, reducing T_{ox} improves the trade-offs only slightly while X_j has nearly no impact on the trade-offs (Fig. 1(c, d)).

2. Analytical modeling

In order to gain insight into the analog behavior of scaled MOSFETs, analytic models with accurate incorporation of transport and 2-D electric fields have been formulated. After incorporating the effect of degeneration due to source resistance (R_S) and mobility degradation due to vertical electric field ($E_\perp$) using parameter θ (αT_{ox}^{-1}), the transconductance (g_m) of short channel MOSFETs is given by

$$g_m = \frac{W v_{sat} C_{ox} V_{gt}(V_{gt}(1+\theta E_{sat}L_g)+2E_{sat}L_g)(1+\lambda V_{ds})}{(E_{sat}L_g(1+\theta V_{gt})+V_{gt})^2(1+W v_{sat}C_{ox}R_S/(1+\theta E_{sat}L_g))} \quad (1)$$

where v_{sat} is the saturation velocity, V_{gt} ($=V_{gs}-V_{th}$) is gate overdrive, W is the device width and E_{sat} ($=2v_{sat}/\mu_o$). μ_o is the channel mobility when $E_\perp$ = 0. To calculate R_{out}, a concept similar to early voltage (V_A) in BJT can be applied to MOSFETs [4],

$$R_{out} = \frac{V_A}{I_{dsat}} = \frac{1}{I_{dsat}}\left(V_{Asat} + \left(1+\alpha\frac{L_n}{l}\right)(V_{ACLM} \parallel V_{ADIBL})\right) \quad (2)$$

$$V_{ACLM} = (E_{sat}L_g(1+\theta V_{gt})+V_{gt})\frac{(V_{ds}-V_{dsat})}{E_{sat}l(1+\theta V_{gt})} \quad (3)$$

$$V_{ADIBL} = \frac{(E_{sat}L_g(1+\theta V_{gt})+V_{gt})}{\theta'(L_g)(1+(2E_{sat}L_g(1+\theta V_{gt}))/V_{gt})} \quad (4)$$

$$\text{with } \theta'(L_g) = \frac{(\sqrt{(\Delta V_{bs})/(\Delta V_{bs}+V_{ds})})\sqrt{(\exp(\Gamma)-1)}+(1-\exp(-\Gamma))}{4\sinh^2(\Gamma/2)} \quad (5)$$

$\Delta V_{bs}=V_{bi}-\Phi_s$ and $\Gamma=L_g/l_t$. V_{bi} is the built in potential of substrate to source/drain pn junction and Φ_s is the surface potential. l_t ($=\sqrt{3T_{ox}(X_{dep}/\eta)}$) is the characteristic length. l in (2) is the inversion charge thickness at the drain end and is used as a fitting parameter. The early voltage components, V_{ACLM} and V_{ADIBL}, are caused by channel length modulation (CLM) and drain induced barrier lowering (DIBL) respectively. The substrate current induced body effect is negligible for short channel MOSFETs and is not included in Eq. 2.

3. Results and Discussion

It is well known that the intrinsic gain reduces as the channel length is scaled, mainly due to worse SCE. As seen in Fig. 2, R_{out} is dominated by V_{ACLM} for longer L_g and by V_{ADIBL} for shorter L_g devices. While V_{ACLM} (Eq. 3) decreases less than proportionally with L_g, V_{ADIBL} (Eq. 4,5) decreases exponentially with scaling. Therefore, to alleviate the degradation in R_{out} when L_g is reduced, l_t need to be decreased proportionally. Hence, T_{ox} and X_{dep}/η (governed by N_A and X_j), should be scaled with L_g to control SCE. However, increasing N_A decreases mobility while decreasing X_j degrades the MOSFET performance caused by the increase in R_S. Therefore, g_m and hence, f_T degrade. In the case of oxide scaling, C_{gate} increases nearly linearly with decreasing T_{ox} while g_m increases only sub-linearly due to a reduction in mobility caused by the higher $E_\perp$. Therefore f_T degrades slightly for thinner T_{ox}.

The improvement in gain and only marginal degradation in f_T with T_{ox} (Fig. 3) leads to an overall improvement in gain vs. f_T trade offs (Fig. 1(a)). Note that f_T does not degrade severely with thinner T_{ox} because while θ increases, V_{gt} is smaller for the same I_{ds}, resulting in a smaller increase in $E_\perp$. Hence, the reduction in mobility is somewhat eased. However, T_{ox} is ultimately limited by the gate leakage and input swing constraints. Also, linearity degrades with scaling T_{ox} (inset of Fig. 3) due to increased effect of $E_\perp$ on g_m for thinner T_{ox} device.

Fig. 4 shows the impact of increasing N_A on gain and f_T. Gain and f_T values corresponding to $N_A=8x10^{17}/cm^3$ are used as the reference point to calculate percentage changes. As N_A is increased to $6x10^{18}/cm^3$, L_g=50nm devices show more that 200% improvement in gain (due to better SCE control) with only 30% degradation in f_T. f_T degrades with an increase in N_A because μ_o decreases (due to increase in coulomb scattering). Short L_g devices degraded less for higher N_A since g_m in short-channel device is largely limited by velocity saturation. The increase in E_{sat} (due to decrease in μ_o) leads to better linearity as velocity saturation effects are reduced (Fig. 5). However, higher N_A leads to higher V_{th} with conventional n+ poly-Si gate. Therefore, methods (other than reducing N_A) which can reduce V_{th}, such as gate material engineering and active bias, need to be employed to enhance the performance of scaled devices.

Scaling X_j improves linearity through the increase in R_S, which is a significant component of R_{out} (Fig. 5). R_S is only slightly non-linear with V_g, but it still contributes to the overall non-linearity of the device. Fig. 6 shows the variation in gain and f_T as X_j is reduced. Smaller L_g devices show more improvement in gain as they are affected by SCE more severely. The degradation in g_m however limits the improvement in gain and also degrades the f_T of the devices. The impact of R_S is more serious for smaller L_g devices as can be seen from Eq. 1 (the coefficient of R_S increases with scaling). To reduce the effect of R_S for scaled devices, an increase in θ or E_{sat} (decrease in μ_o) is required. Both of which will lead to a lower g_m as the ON resis-

tance (R_{ON}) of the channel increases. The only way to get around the problem of higher R_S/R_{ON} is to reduce R_S itself which imposes constraints on scaling X_j.

4. Summary

In this paper, analog behavior of scaled MOSFET was studied. Detailed physics based analytical models were developed for the understanding of the fundamental causes behind the problems. In order to maintain acceptable gain values and simultaneously improving f_T with scaling, thinner T_{ox}, higher bulk doping, and ultra shallow X_j with low sheet resistance are required. The latter will require advance shallow doping technology. Linearity has similar dependence on X_j and N_A. However, thicker T_{ox} is needed for improved linearity. Higher substrate doping can be effectively used to alleviate the degradation in gain and linearity but it leads to higher V_{th}. Therefore, methods (other than reducing N_A) which can reduce V_{th}, such as gate material engineering and active bias, need to be employed.

References

[1] G. Weinberger, ISSCC Tech. Dig., pp. 20-25, Feb. 2000.
[2] R-H. Yan, et. al., IEEE EDL, vol. 13(5), pp. 256-8, 1992.
[3] P. H. Woerlee, et. al., IEEE TED, vol. 48(8), p. 1776, Aug. 2001.
[4] J. H. Huang, et. al., IEDM, pp. 569-72, Dec. 1992
[5] Z. H. Liu, et. al., IEEE TED, vol. 40(1), pp. 86-95, Jan. 1993.

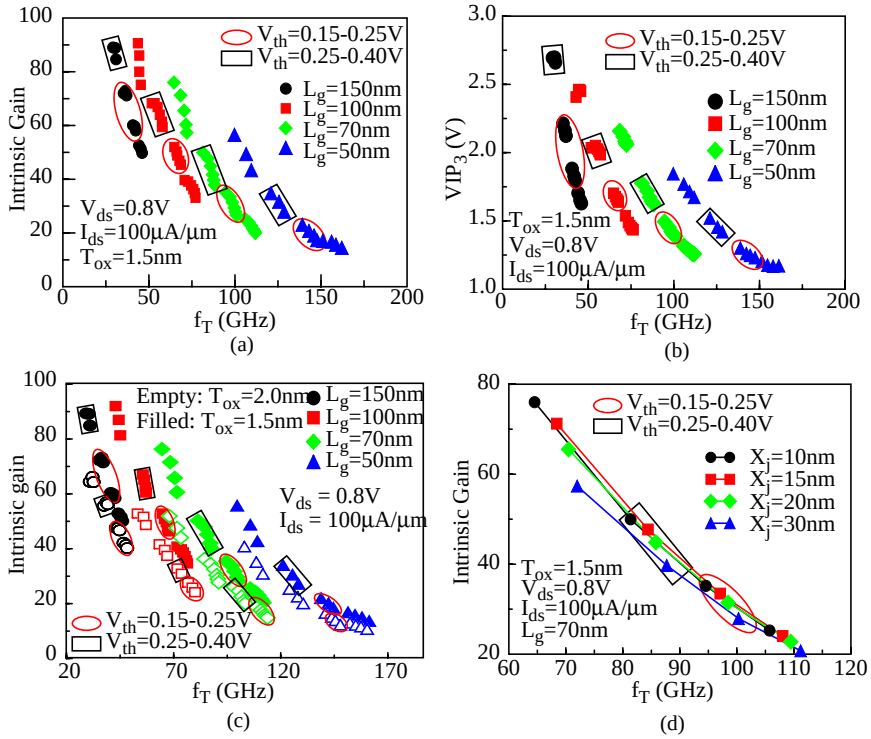

Fig. 1: (a) gain vs. f_T and (b) VIP_3 vs. f_T trade-offs for various channel lengths. Impact of (c) T_{ox} and (d) X_j on gain vs. f_T trade offs. In (a), (b) and (c) N_A as well as X_j are varied for a given L_g and T_{ox} to obtain various points. In (d) only N_A is varied for a given X_j, to obtain the trade-off curves.

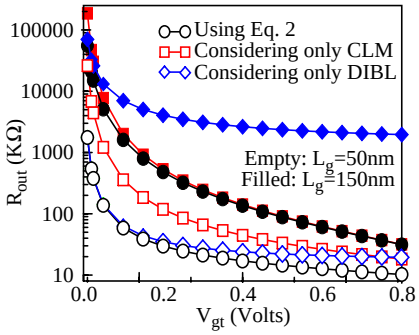

Fig. 2: CLM and DIBL components of R_{out}.

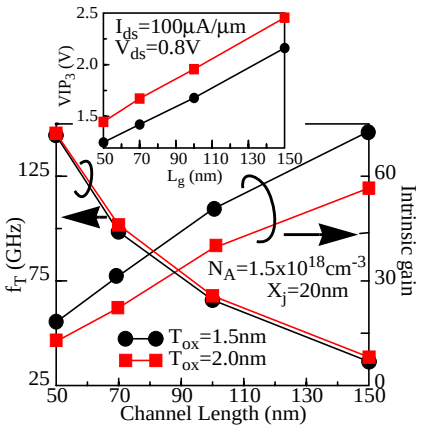

Fig. 3: Gain, f_T and VIP_3 (inset) vs. L_g for T_{ox}=1.5 and 2nm.

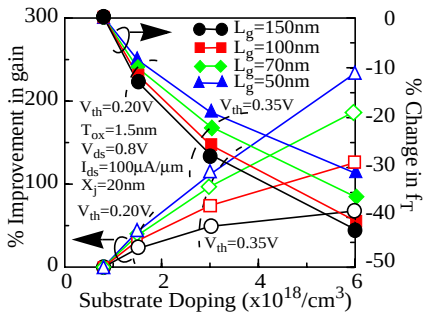

Fig. 4: Percentage change in gain and f_T with increase in substrate doping. Gain and f_T values corresponding to N_A=8x10^{17} are used as base values for calculating the percentage changes. Dotted lines show the constant V_{th} devices.

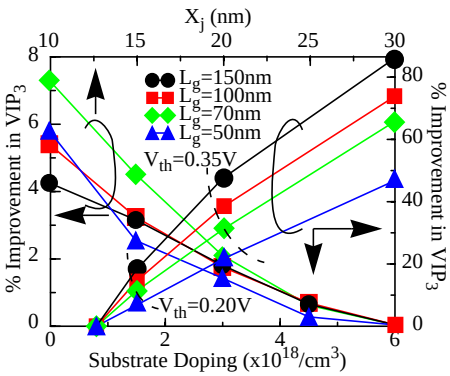

Fig. 5: Percentage change in VIP_3 with increase in N_A and decrease in X_j. VIP3 values at N_A=8x10^{17} and X_j=30nm are used to calculate corresponding percentage changes. Dotted lines show the constant V_{th} devices.

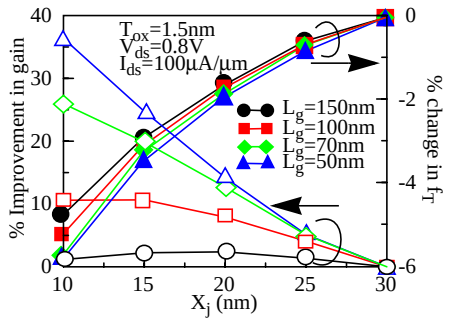

Fig. 6: Percentage change in gain and f_T with decrease in X_j. Gain and f_T values corresponding to X_j=30nm are used as base values for calculating the percentage changes.