Impact of Technology Scaling on RF CMOS

Hassan Hassan Mohab Anis Mohamed Elmasry
VLSI Research Group, University of Waterloo, Waterloo, ON N2L 3G1, Canada
h3hassan, manis, elmasry@vlsi.uwaterloo.ca

Abstract

Inspired by the huge improvement in the RF properties of CMOS, RF designers are invading the wireless market with all-CMOS RF transceivers and system-on-chip implementations. In this work, the impact of technology scaling on the RF properties of the CMOS; frequency properties, noise performance, linearity, stability, and non-quasi static effects, is investigated to provide RF designers with an insight to the capabilities of future CMOS technologies. Using the BSIM4 model it is found that future CMOS technologies have high prospects in the RF industry.

1. Introduction

The scaling of CMOS technologies has pushed the performance of CMOS devices into limits that were thought to be unreachable in the past. With technology scaling, silicon CMOS is expected to became a strong candidate for personal communication systems offering robust, high-density, high-performance, and low-cost system-on-chip implementations. Thus, the performance of future CMOS technologies is expected to challenge bipolar implementations for RF applications. In the industry, CMOS performance improvements have enabled CMOS to be used almost exclusively in some applications such as Bluetooth [1], DECT [2], GPS [3], and WLAN [4].

In order to utilize the capabilities of CMOS, RF circuits designers need to fully understand the impact of technology scaling on the RF performance of CMOS. In this work, the RF performance improvement of CMOS devices is investigated for several future technologies [5] (70nm, 65nm, 45nm) and compared to some existing state-of-the art technologies (500nm, 350nm, 180nm, 130nm, 100nm). The performance parameters investigated include; frequency performance, high-frequency noise performance, linearity, stability, and non-quasi static effects. The improvement in performance is analyzed to give the designers an insight about the impact of technology scaling on RF CMOS performance. To model the increasing nonlinearities of future CMOS devices, the BSIM4 model [6] is used to model all of the parameters under investigation.

In Section 2, the scaling trends of the CMOS parameters that directly affect the device RF performance are studied. The impact of technology scaling on the RF frequency response of CMOS is investigated in Section 3. The improvement in the noise performance of RF CMOS is depicted in Section 4. The linearity and stability of RF CMOS as technology scales down are analyzed in Section 5. Finally, a study of the onset of non-quasi static effects in RF CMOS and how it scales with the technology is presented in Section 6.

2 Scaling Trends of CMOS Parameters Relevant to RF Design

The key CMOS parameters that affect the device RF performance are: oxide thickness t_{ox}, total gate capacitance C_g, and gate resistance R_g. t_{ox} controls the oxide capacitance C_{ox}, threshold voltage V_{th}, and transconductance g_m. On average, t_{ox} scales down with every technology generation by 30% [5] (Figure 1(a)), thus giving new technologies larger driving capabilities and pushing them to higher speeds.

On the other hand, the CMOS parasitics, C_g and R_g, increase by an average of 15% and 25%, respectively with each technology generation (Figures 1(b) and 1(c)). The increase in parasitics limits the improvement in the performance of CMOS devices with scaling.

3 Impact of Technology Scaling on CMOS Frequency Response

The RF frequency parameters of the CMOS device are: the cutoff frequency f_t, maximum oscillation frequency f_{max}, and maximum gain G_{max}.

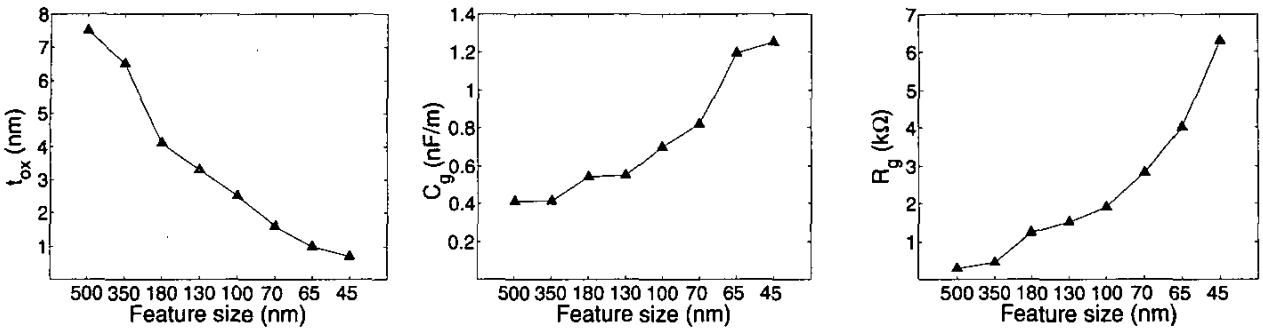
3.1 The Cutoff Frequency (f_t)

f_t is defined as the frequency at which the current gain of the device is unity. For a CMOS, f_t is calculated by

$$f_t = \frac{g_m}{2\pi \times (C_{gs} + C_{gd})} = \frac{g_m}{2\pi \times C_g}, \qquad (1)$$

where C_{gs} and C_{gd} are the gate-source and gate-drain overlap capacitances, respectively. The importance of f_t as a figure of merit for the frequency response of CMOS devices comes from the fact that it represents the maximum fundamental frequency component the device can generate.

0-7803-8445-8/04/$20.00 ©2004 IEEE



(a) t_{ox} vs. Technology (b) C_g vs. Technology (c) R_g vs. Technology

Figure 1. Impact of technology scaling on CMOS parameters.

By examining the impact of technology scaling on the variables in (1), one can have a rough estimate of the f_t trend for future CMOS technologies. Firstly, the gate capacitance C_g scales up by 15% for every technology generation. Secondly, due to the decrease in t_{ox} by 30%, g_m increases by the same percentage. Hence, f_t is expected to increase by an average of 13% every technology generation.

In Figure 2, f_t of a nominal gate length NMOS is plotted against the DC drain current per unit width for the different technologies under investigation. f_t is expected to increase approximately by 20% per technology generation. For a 45nm CMOS technology, a maximum f_t of 200GHz is achieved, which conforms with the figures predicted by the ITRS [7].

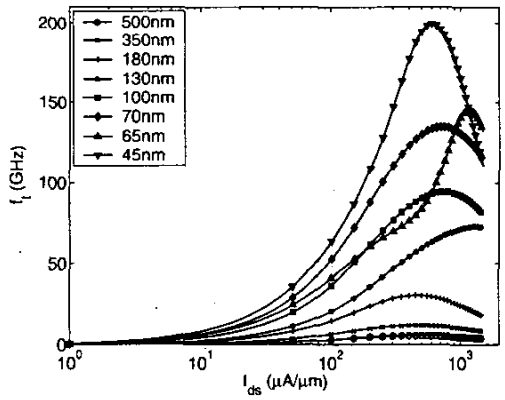


Figure 2. Impact of technology scaling on f_t.

From Figure 2, it is also noticed that f_t increases with the technology for both low- and high-drain currents (i.e., in moderate and strong inversion regions of operation, respectively), while at very low-drain currents (sub-threshold region), the improvement in f_t is not so noticeable. This is because the increase in f_t is due to the rise in g_m, and at very low-currents, g_m is compensated by the increase in C_g. However, at high currents, the g_m term starts to dominate f_t, until saturation is reached, where g_m starts to be independent of the current, and f_t begins to drop again by the effect of C_g.

It should be noted that the 65nm technology needs larger values of current to start exhibiting the same increase in f_t as the other technologies (Figure 2). This can be justified by the fact the C_g of the 65nm [5] technology is scaled up more than the other technologies (as shown in Figure 1(b)), hence larger values of current are needed to overcome the increase in C_g in the denominator of f_t in (1).

3.2 Maximum Oscillation Frequency (f_{max})

f_{max} is the frequency at which the maximum power gain is unity. For a CMOS device, f_{max} is given by

$$f_{max} = \frac{f_t}{2\sqrt{g_{ds}(R_g + R_s) + 2\pi f_t R_g C_{gd}}}, \qquad (2)$$

where R_g and R_s are the gate and source resistance, respectively, and g_{ds} is the output conductance. From (2), it is noticed that f_{max} is heavily dependent on the parasitic resistance and capacitance of the CMOS device. Hence, the layout of the device can significantly affect the value of f_{max}, and there is no fixed trend for scaling f_{max}.

Figure 3 shows the scaling of f_{max} with the technology, which indicates that a 220GHz f_{max} can be achieved for a 45nm CMOS technology. It can also be noticed that the increase in f_{max} with scaling is not as big as the f_t case, which can be accounted to the dependence of f_{max} on the CMOS parasitics (C_g, R_g) which increase with technology scaling as indicated in Figures 1(b) and 1(c). It is also noticed that f_{max} has a similar response due to the change in the current, which can be accounted to the dependence of f_{max} on f_t.

3.3 Maximum Power Gain (G_{max})

The maximum power gain is defined as the power delivered by the device when both the input and output ports are matched to the source and load impedances, respectively. It sets a limit to how much power gain a CMOS device can provide. For a CMOS device, G_{max} is calcu-

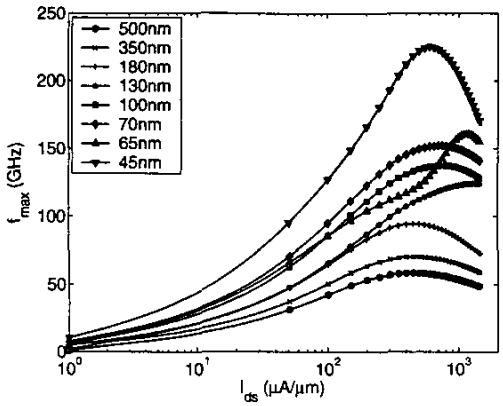


Figure 3. Impact of technology scaling on f_{max}.

lated by

$$G_{max} = \frac{f_t}{8\pi R_g C_{gd} f^2}, \quad (3)$$

where f is the frequency of operation of the device. From (3), it is expected that G_{max} will scale similar to f_t. This fact is depicted by the simulation results of G_{max} with technology scaling in Figure 4 for a frequency of 2GHz. Figure 4 shows that 45nm CMOS can achieve a G_{max} of more than 25dB at 2GHz.

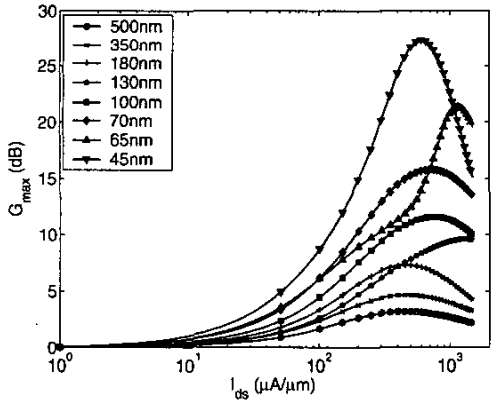


Figure 4. Impact of technology scaling on G_{max} at a frequency of 2GHz.

4 Microwave Noise Properties

There are two main sources of noise in CMOS devices; flicker noise and thermal noise. Flicker noise mainly affects the low-frequency performance of the device and is ignorable at high-frequencies. Thermal noise is the main noise source for RF CMOS and is dominated by the channel thermal noise, and is thus the focus of this work. A figure of merit for the device thermal noise properties is the minimum noise figure NF_{min}. NF_{min} is given by

$$NF_{min} = 1 + K\frac{f}{f_t}\sqrt{g_m(R_g + R_s)}. \quad (4)$$

Thus, NF_{min} is a function of the bias current and frequency of operation.

In Figure 5, NF_{min} is plotted against the drain-source current per unit width at a frequency of 2 GHz for the CMOS technologies under investigation. It can be noticed that NF_{min} decreases strongly with scaling even at low drain-source currents. For low values of bias current, NF_{min} decreases with the current until a certain point where it will start increasing, this is the point where g_m starts being independent of I_{ds} (i.e., the device enters saturation). This trend can be justified by the fact that NF_{min} is proportional to $\sqrt{L_{eff}/W_{eff}}$, which becomes smaller with technology scaling. Hence, NF_{min} is expected to decrease with scaling giving future CMOS devices a better noise performance, NF_{min} below 0.5dB can be achieved by 45nm processes.

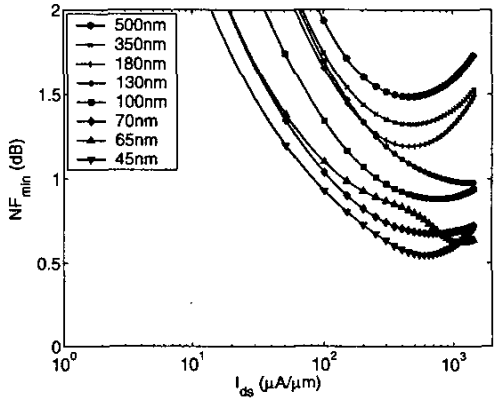


Figure 5. Impact of technology scaling on NF_{min} at a frequency of 2GHz.

5 Linearity and Stability of RF CMOS Devices

In this Section the linearity and stability issues of CMOS technologies under investigation are characterized.

5.1 Impact of Scaling on the Linearity of RF CMOS

In future RF circuits, requirements for linearity will get severer due to the use of wideband communication systems and very high transmission frequencies. Hence, the investigation of linearity of CMOS devices with technology scaling is important to ensure that RF circuits built with future CMOS processes will be able to satisfy the expected demands for linearity. A parameter that is used to characterize the linearity of CMOS devices is the V_{IP3}. V_{IP3} is defined as the input gate bias amplitude at which the first-and third-order output amplitudes of drain current are equal. To achieve low-levels of distortion, V_{IP3} should be as high as possible [8].

For a CMOS device, V_{IP3} is characterized as

$$V_{IP3} = \sqrt{24 \times \frac{g_m}{g_{m3}}}, \quad (5)$$

where $g_{m3} = \partial^3 I_{ds}/\partial V_{gs}^3$. In order to find an expression for the third-order transconductance g_{m3}, special distortion modeling techniques should be used. In this work, the distortion modeling presented in [9] is used to model g_{m3}. In a CMOS device, g_{m3} is given by

$$g_{m3} = \mu_0 C_{ox} \frac{W}{L} \theta^2 \frac{1}{(1+\theta f_\mu)^3} \frac{1}{\sqrt{1+(\frac{V_{ds}}{LE_c})^2}}, \quad (6)$$

where

$$f_\mu = (V_{gs} - V_{fb} - \phi) - \frac{1}{2}V_{ds} + \frac{2}{3}\gamma \frac{(\phi + V_{sb} + V_{ds})^{3/2} - (\phi + V_{sb})^{3/2}}{V_{db} - V_{sb}}, \quad (7)$$

$$\theta = \frac{\mu_0}{2 t_{ox} v_{norm}}, \quad (8)$$

where V_{fb} is the flat band voltage, f_μ is a mobility fitting function, θ is the mobility-reduction coefficient [9].

Using the modeling equations (6)-(8), V_{IP3} is plotted for the 100nm, 70nm, 65nm, and 45nm CMOS technologies in Figure 6. Only these technologies are plotted for the clarity of the figure. The first thing to be noticed from Figure 6 is that V_{IP3} values drop as technology scales down, thus leading to an increase in distortion in future RF CMOS devices. This can be justified by the fact that g_{m3} scales up in a faster rate than g_m with the technology [10]. The second thing about Figure 6 is the shift in the peak towards higher currents as technology scales down. This implicates that future CMOS devices will need larger currents to preserve linearity [11].

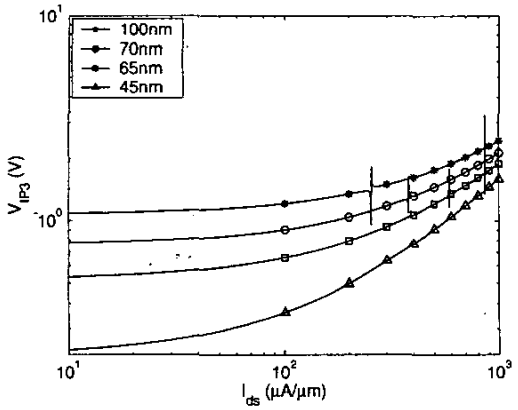


Figure 6. Impact of technology scaling on V_{IP3}.

5.2 RF CMOS Stability

As the scaling of CMOS technologies proceeds with decreasing the threshold and supply voltages, the issue of CMOS stability needs to be given more attention. Stability considerations for RF circuits are crucial, especially in the design of RF amplifiers. For CMOS devices to be unconditionally stable, the operation frequency should be larger than the unity stability-factor frequency, f_k [12]. For a CMOS device, f_k is given by

$$f_k = \frac{f_t d}{\sqrt{8 g_{ds} R_g ds (g_m s^2 R_g + ds(3 g_m R_g + 1))}}, \quad (9)$$

where $d = C_{gd}/C_g$ and $s = C_{gs}/C_g$. By inspecting (9), f_k is expected to follow a similar trend as f_t and increase with scaling, but the presence of the $\sqrt{g_{ds} \times g_m}$ term in the denominator cancels out that effect ($f_t, g_{ds}, g_m \propto W/L$). Hence, the scaling trend of f_k is mainly inversely controlled by the parasitics d, s, and R_g. Since these parasitics increase with technology scaling, then f_k is expected to decrease with scaling. Figure 7 plots f_k for the technologies under investigation, and from which it confirms that f_k decreases with scaling. As a result of that CMOS devices with smaller feature size will have a larger unconditionally stable bandwidth.

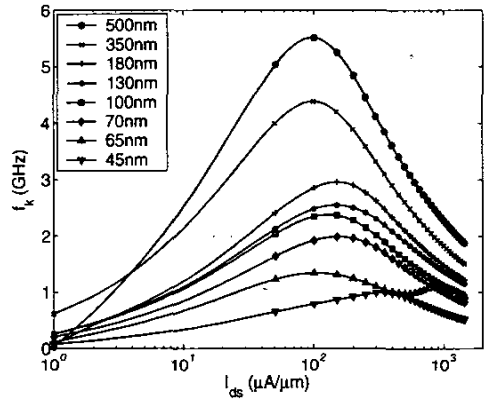


Figure 7. Impact of technology scaling on the unity stability-factor frequency f_k.

6 Non-Quasi Static Effects

The quasi-static (QS) approximation used in MOS devices modeling assumes that the channel charge is a unique function of the instantaneous biases, thus the charge responds to a change in voltage with infinite speed. For low-frequency applications, the QS assumption can be valid because the charging delay is small compared to the rate of change in the input signal. On the other hand, in high-frequency RF applications, the device response time is comparable to the rate of input signal change. This effect is called the Non-Quasi-Static (NQS) effect. NQS effects set upper frequency bounds to analog RF circuits due to reduced amplifications and phase shift effects.

NQS effects are encountered in CMOS devices when the ratio of the instantaneous inversion charge to the inversion charge at low frequency drops below unity [13]. The onset frequency of NQS effects is given by [14]

$$f_{NQS} = n \frac{\mu_{eff}(V_{gs} - V_{th})}{2\pi L_{eff}^2}, \quad (10)$$

where n is a fitting parameter that depends on the accuracy required ($0 < n < 1$). To achieve accurate results, n is chosen as small as possible. In this work, n is selected to be 0.1 [13]. To investigate the impact of scaling on the onset of NQS effects in RF CMOS devices, f_{NQS} is evaluated for the technologies under investigation and the results are plotted in Figure 8.

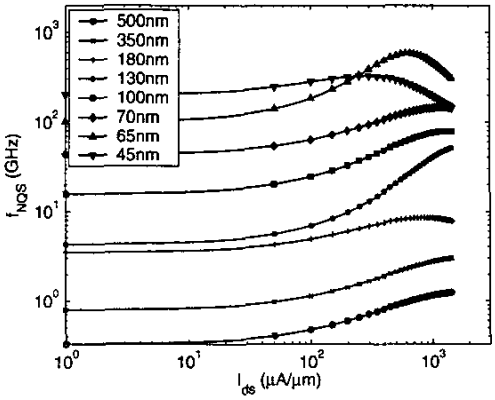


Figure 8. Impact of technology scaling on f_{NQS}.

From Figure 8 it is noticed that f_{NQS} increases with the technology. This is because of the dependency of f_{NQS} on $1/L_{eff}^2$. From these results one can conclude that future CMOS technologies will suffer less from NQS effects than those with larger feature size.

As a further study for the impact of scaling on NQS effects, f_{NQS} is normalized by dividing it over f_t

$$R_{NQS} = \frac{f_{NQS}}{f_t} . \qquad (11)$$

When R_{NQS} is less than 1, this means that the device will be more likely to suffer from NQS effects during its operation. Devices with $R_{NQS} > 1$ will probably not suffer from NQS effects, and QS modeling equations can be used to model their behavior. In Figure 9, R_{NQS} is plotted against the drain current per unit width. Figure 9 confirms that future CMOS devices will less likely to suffer from NQS effects.

7 Conclusion

In this work, it is found that the RF performance of CMOS will improve notably with technology scaling, especially in terms of frequency and noise performance. 45nm CMOS technology technology is expected to achieve f_t of 200GHz. On the other hand, future CMOS will only suffer from increased non-linearities, which can be compensated for using circuit techniques [12]. Hence, with the down scaling of CMOS minimum feature size, CMOS is expected to dominate future wireless applications implementations.

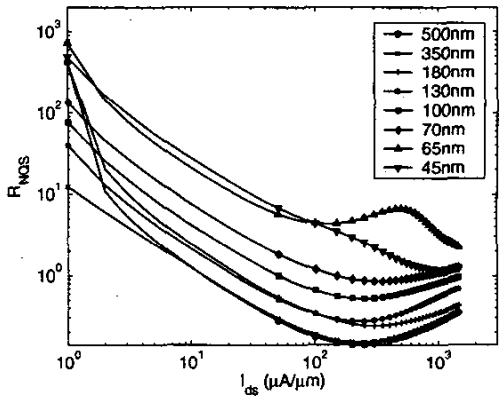


Figure 9. Impact of technology scaling on R_{NQS}.

References

[1] F. O. Eynde et al., "A Fully Integrated Single-Chip SOC for Bluetooth", *IEEE ISSCC Tech. Dig.*, pp. 196–197, 2001.

[2] A. Leeuwenburgh et al., "A 1.9GHz Fully Integrated CMOS DECT Transceiver", *IEEE ISSCC Tech. Dig.*, page 450, 2003.

[3] F. Behbahani et al., "A 27mW GPS Radio in 0.35μm CMOS", *IEEE ISSCC Tech. Dig.*, page 398, 2002.

[4] H. Darabi et al., "A Dual Mode 802.11b/Bluetooth Radio in 0.35μm CMOS", *IEEE ISSCC Tech. Dig.*, page 86, 2003.

[5] University of California Berkeley, Predective Technology Model, "http://www-device.eecs.berkeley.edu/~ptm/".

[6] University of California Berkeley, BSIM4 SPICE MOS Device Model for Circuit Design, "http://www-device.eecs.berkeley.edu/~bsim3/bsim4.html".

[7] International Technology Roadmap for Semiconductors 2003, "http://public.itrs.net".

[8] W. Y. Choi et al., "Stable Extraction of Linearity (V_{IP3}) for Nanoscale RF CMOS Devices", *IEEE Microwave Wireless Compon. Lett.*, vol. 14, n. 2, pp. 83–885, February 2004.

[9] P. Wambacq et al., *Distortion Analysis of Analog Integrated Circuits*, Kluwer Academic Publishers, 1998.

[10] R. Van Langevelde et al., "RF-Distortion in Deep-Submicron CMOS Technologies", *Proc. IEDM*, pp. 807–810, 2000.

[11] P. H. Woerlee et al., "RF-CMOS Performance Trends", *IEEE Trans. Electron Devices*, vol. 48, n. 8, pp. 1776–1782, August 2001.

[12] J.-G. Su et al., "New Insights on RF CMOS Stability Related to Bias, Scaling, and Temperature", *Proc. IEDM*, pp. 40–43, 2000.

[13] A. F.-L. Ng et al., "Determining the Onset Frequency of Non-quasistatic Effects of the MOSFET in AC Simulation", *IEEE Elec. Dev. Letters*, vol. 23, n. 1, pp. 37–39, January 2002.

[14] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, 1987.