The Impact of Scaling Down to Deep-Submicron on CMOS RF Circuits

Qiuting Huang, Francesco Piazza, Paolo Orsatti and Tatsuya Ohguro*

Integrated Systems Laboratory, Swiss Federal Institute of Technology
Zurich, Switzerland
*Toshiba Corporation, Tokyo, Japan

ABSTRACT

Recent papers reporting CMOS RF building blocks with very low current consumption have aroused much expectation in RF receivers using deep-submicron technologies. This paper examines the trend in CMOS scaling, in order to establish the required current levels for different feature sizes if robust designs are to be implemented. The need to match to 50Ω and to limit voltage gain in the input passive matching network is emphasized because it is often overlooked. At 1GHz, 0.25μm CMOS appears to be the threshold for robust, low-NF RF Front-ends with current consumption competive to today's BJT implementations.

INTRODUCTION

Scaling of CMOS technologies have defied all predictions of technology limitations and continued unabatedly towards the deep-submicron minimum feature size. This not only promises gigabit integration, gigahertz clock rate and systems on a chip, but also arouses great expectations for RF CMOS circuits to reach the same performances as bipolar transceivers. Various CMOS RF chips have been implemented in the last couple of years, with the ultimate goal of integrating a radio transceiver on a single chip [1,2]. The successful inroads of CMOS into selective RF applications such as campus-wide wireless LAN and cordless phone prompt the question if deep submicron CMOS could be used for other more stringent applications. Examples of such more demanding applications include GSM handset, very low power pagers etc. It is the purpose of this contribution to illustrate the device parameter improvements that can be expected as CMOS scales towards deep submicron, and describe the significance of such improvements to the key RF design parameters and building blocks.

For RF designs, the most critical device parameters are the transconductance coefficient $\mu C_{ox}'$ and parasitic capacitances. The former determines the amount of current required for a given RF frequency and noise figure, which tends to be high for CMOS implementations. The parasitic capacitances often form undesirable coupling paths and reduce various isolation requirements in addition to limiting the speed achievable for a given current. The use of salicided gate and contacts are widespread even in today's CMOS processes, so that gate and contact resistances can be made negligible at low gigahertz range by proper layout.

The improvements in device parameters are dictated by certain major scaling constraints in CMOS (driven mainly by digital circuit considerations), which are summarized here [3]. The primary objective is to decrease both the minimum channel length and gate-oxide thickness, in order to improve the achievable transconductance and increase integration density. Reduction of oxide thickness, however, lowers the threshold voltage, which must be compensated by increasing channel and, consequently, drain and source diffusion doping concentrations. From 1μm to 0.25μm CMOS, the doping concentration in the last decade has increased by an order of magnitude. Fig.1 shows parasitic capacitances of some typical CMOS processes of different minimum feature sizes, along with some of the surface doping concentrations. It is clear that junction capacitances increase rapidly as process scales down, while the overlap capacitance stays roughly constant, due to reduction of the gate-drain overlap. Also worth noting is that although the minimum feature size L_{min} has improved by a factor of 3, the design rules for minimum drain and source width have decreased by only 1.5 times. This means the

capacitance per µm of gate width due to CJ remains roughly constant as the technology scales. Since the gate oxide thickness scales by the same ratio as L_{min}, gate capacitance per µm of gate-width also remains constant. Thus depending on the capacitance that is involved, the capacitive load per µm gate-width either remains constant or increases. The benefits of improved performance therefore only comes from increased transconductance. Fig.1 shows the mobility at zero bias UO. Being a weak function of doping concentration, it more or less remains constant as the latter increases. The effect of the decrease in oxide thickness, which increases the transconductance coefficient K' by increasing the C_{ox}', is somewhat offset by mobility degradation due to increased vertical field. Instead of tripling as the oxide scales from 18nm to 6nm, K' is less than doubled from 100 to 180 $\mu A/V^2$, for ~200mV gate over-drive. It can be concluded that for the same current and gate width, the transconductance improvement ratio is less than that of minimum gate-length scaling. Since the capacitance does not decrease, the same conclusion holds for frequency improvement for RF applications.

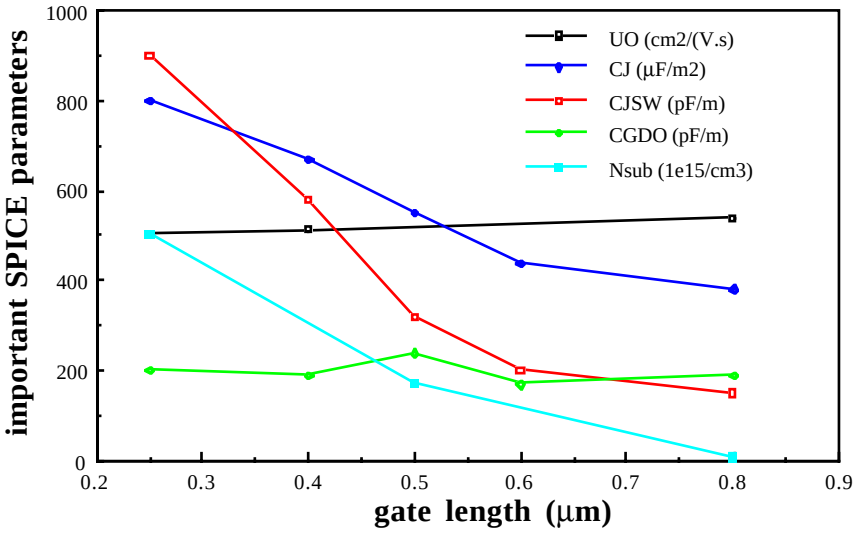

Fig. 1, Circuit parameters of typical (known) CMOS technologies

IMPLICATION OF SCALING ON LNA

At the low gigahertz range, the most important design considerations for the low noise amplifier are (50Ω) impedance matching, noise figure and gain. Although impedance matching is an integral part of LNA design, it has not always been discussed in a manner consistent with circuit performance in papers reporting CMOS LNAs. Mismatched input impedance results in a high reflection coefficient. Since there is some physical distance x_R (1-5cm depending on PCB layout) between the antenna/RF filter and the LNA input, the standing wave between the two can cause appreciable gain variation at 1-2GHz. It can be shown that the relationship between the voltage gain variation, input reflection coefficient $|\Gamma|e^{j\theta}$ and the distance is given by

$$\left|\frac{V_R}{V_s}\right| = \sqrt{\frac{1+2|\Gamma|\cos(\theta)+|\Gamma|^2}{1+2|\Gamma|\cos(\theta-4\pi x_R \div \lambda)+|\Gamma|^2}} \approx 1-\frac{1}{2}\frac{\sin(\theta)\sin(4\pi x_R \div \lambda)}{1+2|\Gamma|\cos(\theta)+|\Gamma|^2} \qquad (1)$$

It then follows that even with a LC matching network, the tolerance on its center frequency places a limit on the network's quality factor Q. If we accept 1dB variation in gain, take 10% tolerance on the LC components and assume the distance between the antenna and LNA to be less than 2cm, the Q allowed for the matching network is around 2 for 1GHz signals. This constraint, though ultimately important, is rarely discussed in recent papers reporting CMOS LNAs and mixers [4-6]. In [4], for example, 10% LC value (thus ω_o) variations can cause 6dB variation in gain, which suggests that the Q of the matching network, which provides part of the overall LNA voltage gain and reduces the apparent noise figure, is too high.

Bearing the matching network in mind, let us examine the first stage of the common-source LNA configuration in Fig.2, which is superior to a common-gate configuration because of its lower achievable noise figure. The matching components are shown to include L_1, which is essential for producing the real part of the input impedance for matching. Implementations with M1's source directly connected to ground tend to have a purely capacitive input which is

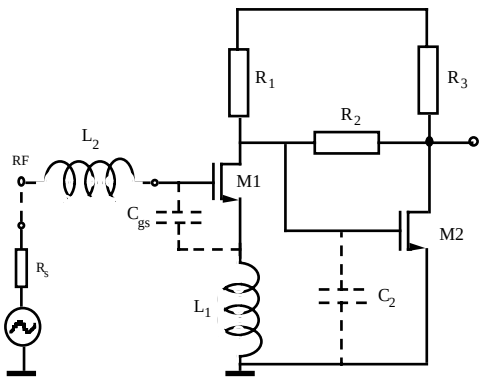

Fig.2 LNA circuit consisting of gm-stage followed by a transimpedance stage

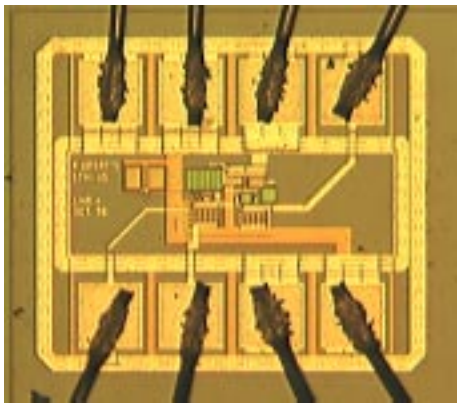

Fig.3 Die-photo of a 0.25μm CMOS LNA

Table 1 Summary of Prelim. Measurements

Vdd	2.25V	$\omega_o/2\pi$	900Mhz
Im1	6.6mA	Im2	10.8mA
S21	15.1dB	S12	-34.5dB
S11	-5dB	S22	-20.2dB
CP	-20dBm	IP3	-2dBm
NF	2.3dB	ext. com	1 (L_2)

very hard to match. Neglecting M1's Miller capacitance and the pad capacitance, the matching conditions are $\omega_o^2(L_1+L_2)C_{gs}=1$ and $g_m L_1 = R_s C_{gs}$, where R_s is the (50Ω) source resistance. The quality factor of the matching network, $Q=1/(\omega_o g_m L_1)$, provides a voltage gain at ω_o between the RF input and the gate-source branch of M1. This means that the overall transconductance G_m from the RF input to M1's drain current is given by $G_m=Qg_m=1/(\omega_o L_1)$. Once the acceptable Q is decided, M1's g_m is determined by L_1, whose best implementation is by a bondwire having a value of 1~2nH. As technology scales, the same W≈600μm is required to give the correct C_{gs} to match to 50Ω, taking into account the neglected Miller and pad capacitances. This enables us to find the minimum current required for each technology for a given Q. For 0.8μm CMOS (K'≈ 100μA/V^2), 40mA is required for g_m=80mS (L_1= 1nH). It scales down to 20 mA for a 0.5μm CMOS (K'=130μA/V^2) and 7.5mA for a 0.25μm CMOS (K' =180μA/V^2). Risking higher Q does enable lower current to be used. For example, 4mA was used in [4] with a 0.5μm CMOS, at the expense of the aforementioned 6dB gain variation for 10% mismatch. Only at 0.25μm CMOS or better, a robust design can be realized by a current consumption competitive to bipolar realizations. At g_m=80mS it can be shown that the achievable NF for the first stage is of the order of 1dB. Fig.3 shows the die-photo of a 0.25 μm LNA designed to be robust against normal parameter tolerances. Preliminary measured performance confirm the above analysis. Detailed discussion of 2nd-stage and justification of current consumption is omitted here as it is rather involved and not all related to scaling. Worthing noting here is that the gain variation for 10% tolerance in ω_o is less than 1dB.

Impact of Scaling on Mixers

The key requirements in a RF mixer design are noise figure and linearity. To have less than 15dB noise figures (required in cellular applications), Gilbert cell structure (Fig.4) is usually used. It is essentially a g_m stage (MA) followed by commutating switches. The g_m stage determines the linearity and scales the noise figure, which makes it the critical part. If off-chip filter is used to improve blocking performance and minimize noise figure, much of the comment above on matching still applies here, except that common-gate input is much more acceptable in the context of the mixer's higher noise figure. We shall concentrate here on linearity. Describing the I-V relationship [7] of MA as in (2), where $\alpha = \dfrac{\theta}{1+\theta(V_{gso}-V_T)}$,

$$I_D(t) = \mu C'_{ox}\frac{W}{L}\frac{[V_g(t)+V_{gso}-V_T]^2}{\{1+\theta[V_{gs}(t)-V_T]\}} = I_{DO}\frac{1+2\beta V_g(t)+\beta^2 V_g^2(t)}{1+\alpha V_g(t)} \quad (2),\ \beta = \frac{1}{[V_{gso}-V_T]}.$$

it can be shown that $IM3 = \dfrac{3}{2\sqrt{2}}HD3 = \dfrac{3}{8\sqrt{2}}\dfrac{\alpha(\beta-\alpha)^2}{2\beta-\alpha}A^2 \approx \dfrac{3\alpha\beta}{16\sqrt{2}}A^2 = \dfrac{3\alpha\sqrt{K'W}}{32\sqrt{L\cdot I}}A^2.$

where A is the signal amplitude. Since θ is inversely proportional to oxide thickness [7], linearity worsens as CMOS scales towards lower feature size. Once the acceptable IM3 limit is reached, the W/L ratio can't be increased and the current level needs to be maintained, in order not to degrade linearity, even when the minimum feature size becomes much smaller. Fortunately, CMOS is inherently more linear than BJT transistors, so that there is still some room in reducing current and increasing W/L ratio before similar limit is reached. One factor that does decrease with decreasing feature size is the noise amplification around clock transitions when the switches in the mixer are not completely turned on or off. Shorter-channel switches respond faster, thereby reducing the period in which noise is amplified.

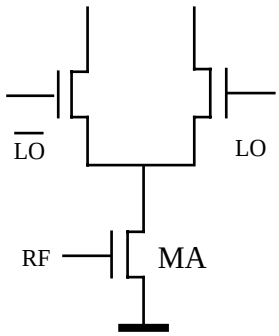

Fig.4 Single-balanced mixer

Impact of Scaling on VCOs

The key requirements of a VCO, typically implemented as a tuned LC oscillator, are phase noise and signal amplitude. It can be shown that once the transconductance is 5~10 times (it usually is) higher than the critical transconductance required for start-up, the amplitude is determined only by the passive resonator (which does not improve with technology scaling) and the bias current [8]. Since LC oscillators at 1~2GHz can already be implemented easily with 1μm CMOS, the benefits of scaling are minor in terms of current consumption.

Impact of Scaling on Frequency Prescalers

Source-coupled logic is well understood. Because it is one of the most power hungry RF frequency blocks for technologies above 0.5μm, the impact of scaling on power is briefly reviewed here for completeness. The bias current I of a source-coupled inverter is on the one hand determined by the required logic swing divided by the load resistance R, on the other by required g_m to have a voltage gain $g_mR \approx 1.5\sim2$. The latter requires a minimum width for the transistors, which results in a load capacitance that limits R for a given corner frequency. As feature size decreases, the improvement of average K' is insignificant because of the large gate overdrive in a digital circuit. For the same aspect ratio W is allowed to scale with L, so that parasitic load capacitance scales down as the feature size. This allows R to increase by the same ratio for a given corner frequency. It is now possible to scale current down by the same factor without losing swing nor gain. In fact, since g_m only scales as the square-root of current, W can be scaled down somewhat faster than L. The saving in current is therefore faster than the feature-size reduction.

Conclusions

As CMOS scales to deep-submicron, the RF building blocks that benefit most in terms of current consumption are the LNA and the Prescaler. For both the required current scales down somewhat faster than the feature size. Quarter micron CMOS is the threshold where robust designs can be realized with current consumption competitive to BJT implementations.

REFERENCES

[1] A. Abidi et al, "The Future of CMOS Wireless Transceivers", IEEE ISSCC, Digest of Technical papers, pp118-119, Feb. 1997, San Francisco, USA
[2] J. Rudell et al, "A 1.9GHz Wide-band IF Double Conversion CMOS Integrated Receiver for Cordless Telephone Applications", IEEE ISSCC, Digest of Technical papers, pp304-305, Feb. 1997, San Francisco, USA
[3] R. K. Watts, edited, *Submicron Integrated Circuits*, Wiley-Interscience 1989
[4] A. Karanicolas, "A 2.7V 900MHz LNA and Mixer", IEEE JSSC, Vol. 31, pp.1939-1944, Dec. 1996
[5] P. Sullivan et al, "A Low Voltage Evaluation of a 1.9GHz Silicon MOSFET Gilbert Cell Downconversion Mixer" Proc. ESSCIRC'96, pp.212-215, Sept. 1996, Neuchatel, Switzerland
[6] A. Shahani et al, "A 12mW Wide Dynamic Range CMOS Front-End for a Portable GPS Receiver, IEEE ISSCC, Digest of Technical papers, pp368-369, Feb. 1997, San Francisco, USA
[7] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill 1987
[8] Q. Huang, "Power Consumption vs LO Amplitude for CMOS Colpitts Oscillators", Proc. CICC, 1997, USA