# Design Considerations for Future RF Circuits [1]

Behzad Razavi
Electrical Engineering Department
University of California, Los Angeles

## Abstract

**The RF design paradigm will change significantly as CMOS technology scales and integration levels rise to accommodate multi-band, multi-mode transceivers and baseband processors. This paper describes technology scaling issues such as low supply voltages, high gate leakage currents, and low transistor output impedances. Also, design techniques for low-voltage mixers are presented, and stacked and nested inductors are proposed to achieve compact layouts for multi-band systems.**

## I. Introduction

The evolution of RF circuits has crossed an inflection point defined by two important trends, one at the device level and the other at the system level. While early efforts in RF CMOS design sought acceptable speed and noise performance with slow, noisy transistors, the speed of today's transistors well exceeds that necessary for most applications, but other aspects of scaled devices have begun to pose challenges. These challenges will become increasingly more serious as the designs migrate to the 65-nm and 45-nm generations in the next five years. At the system level, the concept of "high integration" has changed its meaning from a chip containing a single transceiver in the 1990s to system-on-chips (SOCs) comprising multiple transceivers that accommodate various bands, modes, and space-time diversity techniques.

This paper describes the design issues facing future RF systems and proposes a number of techniques that ease the design of highly-integrated transceivers in scaled technologies. Section II deals with three scaling issues, namely, reduction of the supply voltage and the output impedance of transistors and the rise in the gate leakage current. Section III presents design techniques for low-voltage active mixers and multi-band circuits.

## II. Impact of Technology Scaling

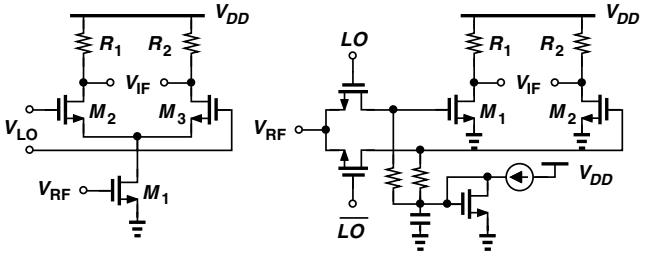
While providing $f_T$'s well above 100 GHz, scaling has been plagued by three adverse trends: lower supply voltages, higher gate leakage currents, and lower transistor output impedances. We consider the impact of these trends on RF design.

[1] This work was supported by Realtek Semiconductor and Skyworks, Inc.

### A. Supply Voltage Scaling

Not surprisingly, lower supply voltages exacerbate the trade-offs in both the RF signal path and the local oscillator (LO) path. More specifically, the linearity-gain-noise trade-offs in RF mixers and the tuning range-phase noise trade-off in voltage-controlled oscillators (VCOs) become markedly more severe.

Figure 1(a) shows a generic single-balanced active mixer. Since the overdrive voltage of $M_1$ must be chosen high enough




(a) (b)

Fig. 1. (a) Active and (b) passive mixers operating with a 1-V supply.

to provide the required $IP_3$, and since the voltage drop across the load resistors is limited by the headroom, the circuit suffers from a low conversion gain at low supply voltages. By contrast, the topology of Fig. 1(b) exhibits greater linearity and compensates for the loss of the passive mixer by a linear baseband amplifier. In fact, simulation of these two circuits for operation at 5 GHz and with a 1-V supply in 90-nm technology yields the results shown in Fig. 1, where the noise figure and power dissipation are kept constant and each circuit is optimized for gain and linearity. These results suggest that the passive configuration exhibits a 9.6-dB advantage in conversion gain and a 6.8-dB advantage in $IP_3$. These observations suggest the superiority of passive mixers at low supply voltages. Nonetheless, other techniques described in Section III can be employed to improve the performance of active mixers.

The mixers in the transmit path also face linearity and gain degradation. Figure 2 depicts a folded quadrature topology with inductive loads that can potentially operate with low supply voltages. However, as the baseband voltage swings increase and the PMOS differential pairs experience greater current steering, eventually the current of one PMOS transis-

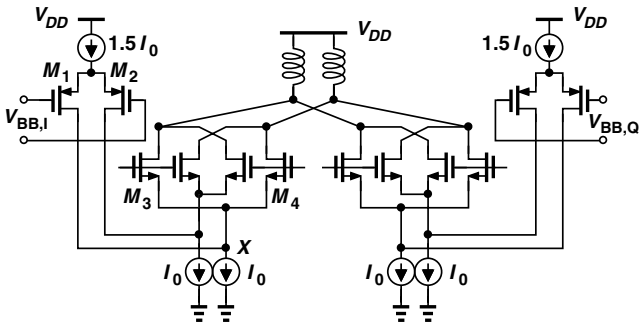



Fig. 2. Folded quadrature upconversion mixer.

tor, e.g., $M_1$ exceeds $I_0$, starving the NMOS differential pair $M_3$-$M_4$. As a result, $V_X$ rises, $M_1$ enters the triode region, and the mixer becomes very nonlinear.

To resolve this issue, the baseband voltage swings must be reduced and/or the PMOS pairs must be resistively degenerated, thereby lowering the output swing and costing greater power dissipation in subsequent stages. A more linear upconversion mixer is proposed in [1] that incorporates passive mixers along with a virtual-ground summing node.

Supply scaling also impacts the design of oscillators. For the generic negative-$G_m$ oscillator shown in Fig. 3, the thermal

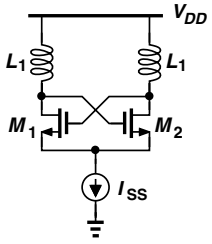



Fig. 3. Simple oscillator.

noise contribution of $M_1$ and $M_2$ to the phase noise can be obtained from Leeson's equation:

$$S(\Delta f) = \frac{2kTR_P}{V_{osc}^2 Q^2}\left(\frac{f_0}{\Delta f}\right)^2, \qquad (1)$$

where $R_P$ denotes the equivalent parallel resistance of the tank ($= QL_1\omega_0$) and $V_{osc}$ is the peak-to-peak differential output swing. Thus, the scaling of the supply and hence the oscillation amplitude directly degrades the phase noise.[2] It is also important to note that if the supply scales down by a factor of, say, two while $I_{SS}$ is constant, then $S(\Delta f)$ rises by a factor of four. Now, if the tank admittance, the transistor widths, and $I_{SS}$ are scaled up by a factor of two, then $S(\Delta f)$ drops by only a factor of 2. That is, for a given power dissipation, supply scaling still degrades the phase noise.[3]

Reduction of the voltage headroom also raises the contribution of current source $I_{SS}$ in Fig. 3 to phase noise. This effect becomes especially acute if $I_{SS}$ is increased according to the power scaling described above.

[2]It is assumed that the tank design cannot be improved and hence $Q$ and $R_P$ remain constant.

[3]Interestingly, the high $f_T$ of transistors cannot be traded for noise here.

The design of VCOs faces additional difficulties at low supply voltages. Suppose gate oxide and supply scaling occurs such that MOS varactors provide a constant capacitance range. Even in such an ideal case, the gain of the VCO, $K_{VCO}$, must scale up [Fig. 4(a)], making the circuit more sensitive to the

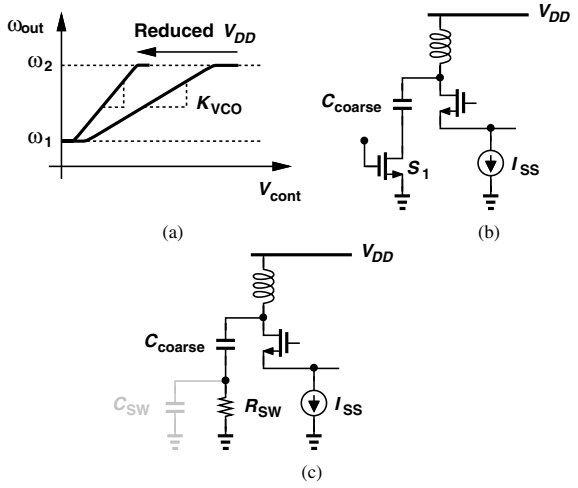



Fig. 4. (a) Effect of supply scaling on VCO sensitivity, (b) coarse tuning, (c) effect of switch resistance and capacitance.

mismatches and $1/f$ noise in the charge pump (CP). This issue can be alleviated by means of discrete (coarse) tuning [Fig. 4(b)] but at the cost of degrading the $Q$ as a result of the on-resistance of the MOS switches. To quantify this effect, we assume that each switch yields a sufficiently high capacitor $Q$ when it is on and contributes a small capacitance in series with the switched capacitors when it is off [Fig. 4(c)]:

$$Q_{sw} = 9Q_{tank} \quad \text{for 1 dB phase noise penalty} \qquad (2)$$
$$C_{sw} \ll C_{coarse}, \qquad (3)$$

where $Q_{sw} = (R_{sw}C_{coarse}\omega_0)^{-1}$ and $C_{sw}$ includes the drain junction and gate-drain overlap capacitances of the switch. Since $R_{sw}C_{sw}$ is independent of the transistor width and hence a constant of the technology, we define $\omega_{sw} = (R_{sw}C_{sw})^{-1}$ and combine Eqs. (2) and (3) to obtain the following requirement:

$$\omega_{sw} \gg 9Q_{tank}\omega_0. \qquad (4)$$

In 90-nm technology, $\omega_{sw} \approx 2\pi(422\text{ GHz})$, a seemingly high value.But a VCO designed for operation at 5 GHz with $Q_{tank} = 6$ would yield $9Q_{tank}\omega_0 = 2\pi(270\text{ GHz})$, failing to satisfy (4). In other words, either the $Q$ must degrade considerably or the discrete tuning steps and hence the overall tuning range must become narrower.

### *B. Transistor Output Impedance*

The output impedance of scaled transistors has fallen to such values that it can lower the $Q$ of oscillators. Consider the circuit depicted in Fig. 5 and suppose $I_{SS} = 1$ mA and $R_P = 500\ \Omega$ so as to provide an output swing of about 500 mV. For $r_O$ not to degrade the $Q$ of the tank, we require $r_O \gg R_P$. But a 90-nm device biased at 0.5 mA exhibits

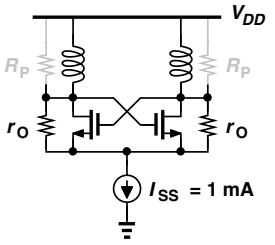



Fig. 5. Oscillator including the output impedance of transistors.

$r_O = 2.3$ kΩ. Fortunately, this resistance appears for only a fraction of the period (when $M_1$ and $M_2$ are in the vicinity of equilibrium), but twice per period. That is, the deQing effect of $r_O$ may be negligible in 90-nm technology but is likely to become problematic in 65-nm and 45-nm generations.[4]

### C. Gate Leakage Current

The gate leakage current [2] has reached 10-100 pA/$\mu$m$^2$ in 90-nm technology, and its dependence on the gate-source and gate-drain voltages makes cancellation difficult. This leakage readily manifests itself if the low-pass filter in a phase-locked loop (PLL) incorporates MOS capacitors. As illustrated in Fig. 6, the leakage current $I_G$ discharges the loop filter while

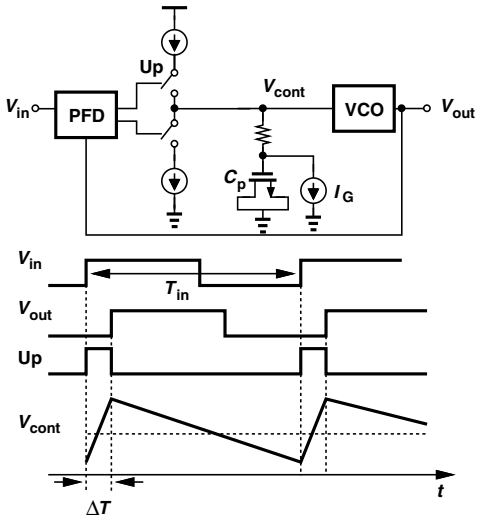



Fig. 6. Effect of gate leakage on PLL performance.

the charge pump is off. In the steady stage, the PLL develops a phase offset, $\Delta T$, during which the CP replenishes the charge drained by $I_G$. Thus, the peak-to-peak ripple on the control voltage is given by $(I_G/C_P)T_{in}$, where it is assumed $\Delta T \ll T_{in}$. The "self-droop" rate $I_G/C_P$ is independent of the MOS lateral dimensions and hence a constant of the technology, reaching 70-700 $\mu$V/ns for 90-nm devices. For example, if $T_{in} = 50$ ns, then a ripple of 3.5-35 mV$_{pp}$ appears on the control voltage, yielding large sidebands at the VCO output.

## III. Design Techniques

### A. Low-Voltage Active Mixers

A number of techniques can relax the trade-offs between noise, linearity, and gain of active mixers operating with low supply voltages. The underlying principle is to allocate a smaller bias current to the switching pair and the loads than to the RF transconductor device. Shown in Fig. 7(a) is a classic

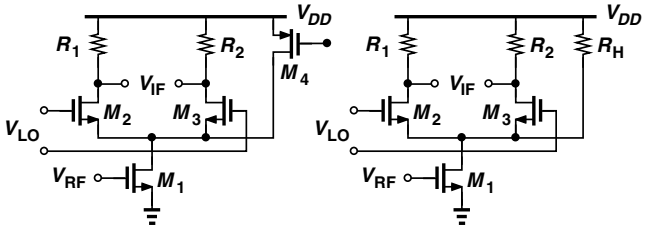



Fig. 7. Reduction of switching pair current by means of (a) PMOS current source, (b) resistor.

example [3], where $M_4$ carries 50%-75% of $I_{D1}$, thus allowing $M_2$ and $M_3$ to switch more abruptly and $R_1$ and $R_2$ to assume a higher value. As $V_{DD}$ falls, the contribution of $M_4$ to the noise becomes more pronounced. In fact, the ratio of the noise currents of $M_4$ and $M_1$ can be expressed as

$$\frac{\overline{I_{n4}^2}}{\overline{I_{n1}^2}} \approx \frac{\dfrac{2|I_{D4}|}{|V_{GS4}-V_{TH4}|}}{\dfrac{2I_{D1}}{V_{GS1}-V_{TH1}}}. \quad (5)$$

For example, if $|I_{D4}| = 0.75I_{D1}$, $V_{GS1} - V_{TH1} = 0.4$ V, and $|V_{GS4} - V_{TH4}| = 0.6$ V (with $V_{DD} = 1$ V), then $\overline{I_{n4}^2}/\overline{I_{n1}^2} = 0.5$, a significant rise in the noise figure. At very high frequencies, the capacitance contributed by $M_4$ also degrades the performance.

Figure 7(b) illustrates an alternative where $R_H$ provides the additional current [4]. Using the same assumptions as above, we obtain $\overline{I_{nRH}^2}/\overline{I_{n1}^2} = 0.25/\gamma$, observing at least a two-fold advantage. In fact, simulations reveal that this technique can raise the conversion gain of the design in Fig. 1(a) by 6 dB with no overall noise penalty.

A third topology is shown in Fig. 8(a). Here, $I_H$ sources the additional current while $L_H$ resonates with the total capacitance seen at node $X$ [5]. Since in typical designs, the equivalent parallel resistance of $L_H$ (at resonance) is much greater than $1/g_{m1}$, the thermal noise due to the loss in $L_H$ remains negligible. Thus, the auxiliary branch consisting of $I_H$, $L_H$, and $C_H$ improves the performance while contributing negligible noise. For example, if the design of Fig. 1(a) incorporates this technique, the noise figure falls to 7.1 dB and the conversion gain rises to 25 dB.

The only drawback of this approach relates to the area occupied by $L_H$. Note, however, that for quadrature downconversion, a single inductor suffices [Fig. 8(b)].

### B. Multi-Band Techniques

The demand for multi-mode, multi-band operation along with space-time diversity dictates very compact transceiver design so as to minimize the silicon area and avoid long interconnects at high frequencies.

A critical observation here is that the digital baseband processor both benefits from technology scaling and involves min-

[4]Note that these observations are independent of the $Q$ of the tank.

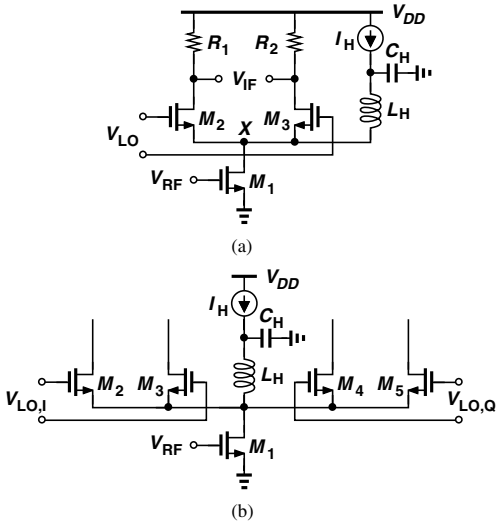



Fig. 8. (a) Use of inductor to improve the performance of an active mixer, (b) quadrature version of (a).

imal overhead for handling multi-band, multi-mode operation. As a result, in such systems, the RF and analog sections have begun to occupy a substantial portion of the total SOC area. That is, each transceiver can no longer incorporate tens of on-chip inductors and transformers.

Interestingly, recent work in the area of ultra-wideband (UWB) design can be leveraged into multi-band systems. Many concepts are similar: the receive and transmit paths must accommodate multiple bands, and so must the LO frequencies.

In this section, we introduce a number of layout compaction concepts that prove useful in multi-band systems.

Figure 9(a) shows an example of a low-noise amplifier

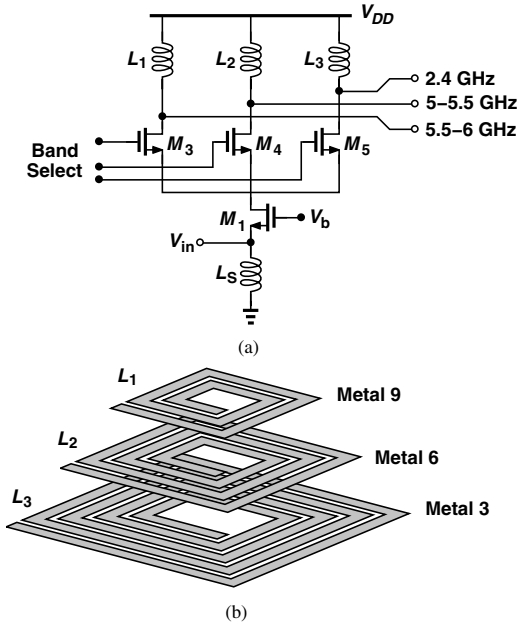



Fig. 9. (a) Multi-band LNA, (b) stacked inductors to save area.

(LNA) [4] designed for the 2.4-GHz and 5-GHz bands. The common-gate stage provides an $|S_{11}|$ of greater than 10 dB from 2 to 6 GHz and the inductive loads cover the bands of interest. The three load inductors would ordinarily occupy a large area. However, they can be constructed as depicted in Fig. 9(b), thus yielding a compact layout. The electric and magnetic coupling between the inductors must, of course, be taken into account, but the effect is relatively benign as the two inductors tied to the disabled cascode transistors carry little current.

The stacking technique of Fig. 9(b) can also be applied to multi-band VCOs. For example, such a topology can incorporate three inductors and three negative-$G_m$ cores. The outputs may or may not be multiplexed depending on the number of signal paths chosen for the receiver or the transmitter. With three inductors, a three-band VCO does not provide quadrature outputs. The speed of today's devices makes it preferable to operate VCOs at twice the required frequency and employ a $\div 2$ circuit to obtain quadrature phases. This technique consumes half as many inductors as quadrature oscillators do.

Another layout compaction method involves "nested" inductors. As shown in Fig. 10, a large spiral can enclose a

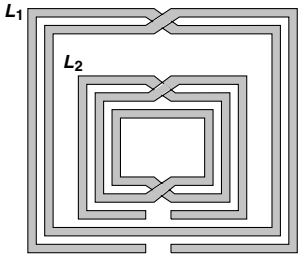



Fig. 10. Nested inductors.

smaller structure in the same metal level. Note that $L_1$ and $L_2$ need not differ in value significantly as $L_1$ has a larger diameter but a smaller number of turns. For example, $L_1$ and $L_2$ may belong to a dual-band oscillator. In cases where $L_1$ and $L_2$ carry currents simultaneously, e.g., in a two-stage amplifier, the polarity of the magnetic coupling factor between them can be chosen so as to improve the stability, frequency response, or other aspects of the circuit.

## References

[1] B. Razavi et al, "Multiband UWB Transceivers," *Proc. CICC*, pp. 141-148, September 2005.

[2] Yeo et al, "Direct Tunneling Gate Leakage Current in Transistors with Ultrathin Silicon Nitride Gate Dielectric," *IEEE Electron Device Letters,* vol. 21, pp. 540-542, Nov. 2000.

[3] W. H. Sansen and R. G. Meyer, "Distortion in Bipolar Transistor Variable-Gain Amplifiers," *IEEE J. Solid-State Circuits,* vol. 8, pp. 275-282, August 1973.

[4] B. Razavi et al, "A UWB CMOS Transceiver," *IEEE J. Solid-State Circuits,* vol. 40, pp. 2555-2562, Dec. 2005.

[5] B. Razavi, "A 60-GHz CMOS Receiver Front End," *IEEE J. Solid-State Circuits*, vol. 41, pp. 17-22, Jan. 2006.