Transworld Research Network
37/661 (2), Fort P.O., Trivandrum-695 023, Kerala, India

Si-based Semiconductor Components for Radio-Frequency Integrated Circuits (RF IC), 2006:
ISBN: 81-7895-196-7 Editor: Will Z. Cai

2 RF and millimeter-wave IC design in the nano-(Bi)CMOS era

S. P. Voinigescu, T. O. Dickson, M. Gordon, C. Lee, T. Yao
A. Mangan, K. Tang and K. H. K. Yau
Edward S. Rogers Dept. of Electrical and Computer Engineering, University of Toronto, 10 King's College Rd., Toronto, ON, M5S 3G4, Canada

Abstract

This article describes new developments in circuit topologies and design methodologies that have become necessary in the RF and millimeter wave (mm-wave) arena due to the emergence of the silicon MOSFET as a serious contender. The impact of nano-scale CMOS problems that affect digital design, such as gate and subthreshold leakage, is shown to be negligible in traditional RF/mm-wave and high-speed CML/ECL building blocks. Furthermore, it is demonstrated that, by changing their design style from

Correspondence/Reprint request: Dr. S. P. Voinigescu, Edward S. Rogers Dept. of Electrical and Computer Engineering, University of Toronto, 10 King's College Rd., Toronto, ON, M5S 3G4, Canada
E-mail: sorinv@eecg.toronto.edu

a V_{GS}- and V_T-centric to a current-density centric approach, RF/mm-wave designers can take advantage of the constant-field scaling rules that have driven CMOS process development over the past 10 years, to design low-voltage, low-power, high performance circuits that are robust to nano-CMOS process variation and are portable between foundries and technology nodes. Candidate topologies and algorithmic design methodologies for mm-wave IC building blocks such as LNAs, mixers, VCOs, and power amplifiers are discussed and recent experimental results obtained in SiGe BiCMOS and 90-nm RF CMOS technologies using inductors and transformers above 50 GHz are presented.

1. Introduction

(Bi)CMOS technology scaling to nanometer dimensions has continued unabated. Cutoff f_T and oscillation f_{MAX} frequencies in production 90-nm strained CMOS [1] and SOI technologies [2] now exceed 200 GHz at 1V supply voltages and are on par with those of production III-V technologies and of the most advanced SiGe HBTs. Power dissipation, high frequency noise figure, and phase noise performance improve with scaling. By 2010, when the 45-nm node is expected to enter early production [3], planar strained-channel MOSFET speed will likely exceed 400 GHz. At the same time, maximum output swing, linearity, and device leakage are degraded by scaling, albeit at a much slower pace than previously anticipated because gate oxide scaling has virtually come to a standstill at the 90-nm node. This situation provides an excellent incentive for the introduction of digital signal processing techniques relying on high oversampling ratios and mm-wave clock frequencies. Such techniques typically require only simple circuit topologies that can be made robust to process variation, transistor leakage, and transistor non-linearity.

Despite the tremendous progress in intrinsic device speed and high-density integration capability, RFICs and fiberoptic transceivers, the main high-speed technology drivers in the 1990's, have practically stalled for the past ten years at 2-5 GHz and 10 Gb/s, respectively. Perhaps ironically, the clock signal of microprocessors has surpassed by more than a factor of two the highest "radio frequencies" in cellular phones and is fast approaching that of OC192 systems. It should come then as no surprise that digitizing the 2-GHz cellular phone is seriously pursued by those in the industry with in-house 90-nm and 65-nm technologies [4]. Traditional RF building blocks such as low-noise amplifiers (*LNAs*), voltage-controlled oscillators (*VCOs*) and power amplifiers (*PAs*) will likely survive the digital onslaught. However, RF designers will have to settle for 1-V (or lower) supplies, circuits with low transistor count, and simple topologies with at most 2 vertically stacked transistors.

This disconcerting outlook for high-speed and RFIC research can be overcome by taking on new applications beyond 50 GHz. The latter have yet to benefit from the large scale integration capability and digital signal processing power of CMOS and SiGe BiCMOS technologies.

Rather than trying to "mimic" GaAs and InP MMIC techniques which rely on large area transmission lines, hybrid couplers, and distributed topologies, this paper proposes to extend trusted RFIC design styles and topologies to millimeter wave (mm-wave) frequencies. The goal is to apply them in emerging, potentially large volume, applications in the 60-GHz to 100-GHz range. These include, but are not limited to, wireless gigabit Ethernet, automotive radar, and mm-wave imagers.

The paper will first introduce the remarkable invariance of MOSFET characteristic current densities between technology nodes and foundries. Next, it will illustrate how this property can be applied to the algorithmic design of RF and mm-wave Si(Ge) (Bi)CMOS IC building blocks, leading to optimal performance and minimal redesign effort as circuits are ported between technology nodes and foundries.

2. Optimal sizing and biasing of active devices

With the scaling of MOSFETs in the sub-100nm range several key phenomena arise or become critical. These include (i) gate leakage, (ii) subthreshold leakage, (iii) mobility degradation due to the vertical electric field, (iv) geometry-dependent stress due to narrow active regions and the close proximity of STI regions, (v) strain engineering to enhance carrier mobility, (vi) greatly diminished scaling beyond the 90-nm node of the effective gate oxide thickness (*EOT*) due to charge quantization in the channel at the silicon-gate oxide interface, (vii) increasing contribution of the source and drain series resistance to the overall channel resistance and to the degradation of intrinsic g_m, f_T, f_{MAX}, and NF, and (viii) increased V_T variation due to larger fluctuations in the number of dopant atoms in very short channels.

Gate and subthreshold leakage does not pose significant problems for high-speed, RF, and analog design. For many years RF and high-speed designers have used GaAs p-HEMTs, InP HEMTs, InP HBTs and SiGe HBTs, all of which exhibit much larger gate or base currents than 90-nm or 65-nm MOSFETs. In fact, in many CML and RF circuits, low-V_T devices (with increased subthreshold leakage) are preferred to achieve high speed with a low-voltage power supply.

The bias voltage range over which the quadratic law that describes the DC characteristics of the transistor in the active region and which is the cornerstone of CMOS-based analog and RF circuit design today, becomes

ever smaller. For example, as illustrated in Fig. 1, in 90-nm n-channel MOSFETs, the effective gate voltage range is smaller than 300 mV. At the same time, due to the DIBL effect [5], V_T can vary by at least 50 mV when the drain bias voltage is changed between 0 and 1 V. A similar variation in threshold voltage occurs as a result of the reverse short-channel effect, when the gate length is changed from the minimum feature size to 2-3 times the minimum feature size, a common practice among analog designers. Obviously, designing circuits with a certain effective gate overdrive voltage, V_{eff}, and V_T in mind is no longer feasible. Adding to the predicament is the impact of ever-increasing process variation on V_T at nano-scale gate lengths and, with the advent of strain engineering [6], the gate length and gate finger width dependence of carrier mobility. It has become apparent that CMOS analog and RF design techniques must be overhauled to keep pace with nano-scale CMOS phenomena. Fortunately, the change is not that difficult to make. Constant-current-density design techniques, which have long been in practice in the industry for high-speed and RF circuits using bipolar transistors, can also be reliably applied to nano-CMOS circuits.

One of the most useful features of nano-scale MOSFETs is that, for most of the active bias range, the I-V characteristics are linear, resulting in almost constant transconductance and gate capacitance as a function of gate voltage. Fig. 2 reproduces the small signal transconductance g'_m, gate source capacitance C'_{gs}, and gate-drain capacitance C'_{gd}, per gate width W, measured on a 90-nm n-MOSFET after de-embedding the series parasitics R_s, R_d, and R_g. In 250nm or older generation technologies, the linear region

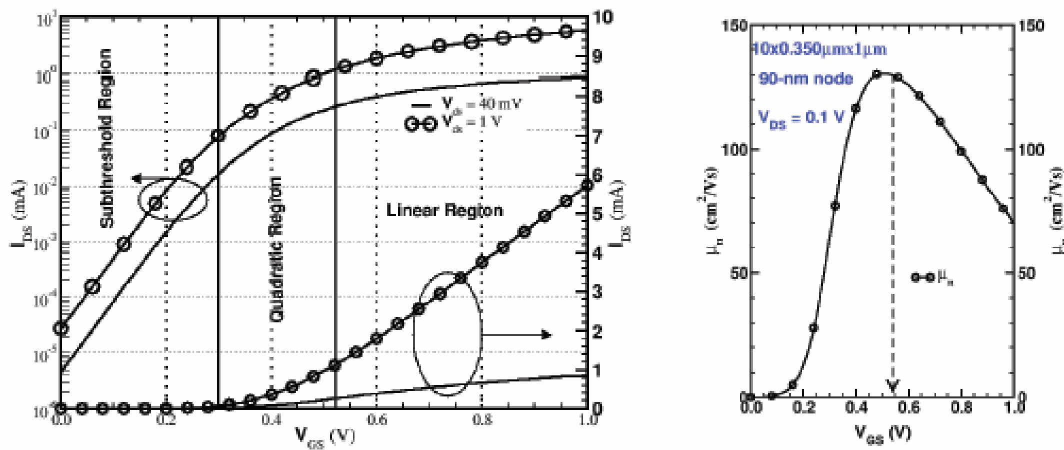

Figure 1. a) Measured transfer and subthreshold characteristics of a 10x1µ mx90-nm n-MOSFET indicating the narrow voltage range over which the square law is still valid. b) Measured electron mobility in a long channel device as a function of V_{GS} at V_{DS}=0.1 V.

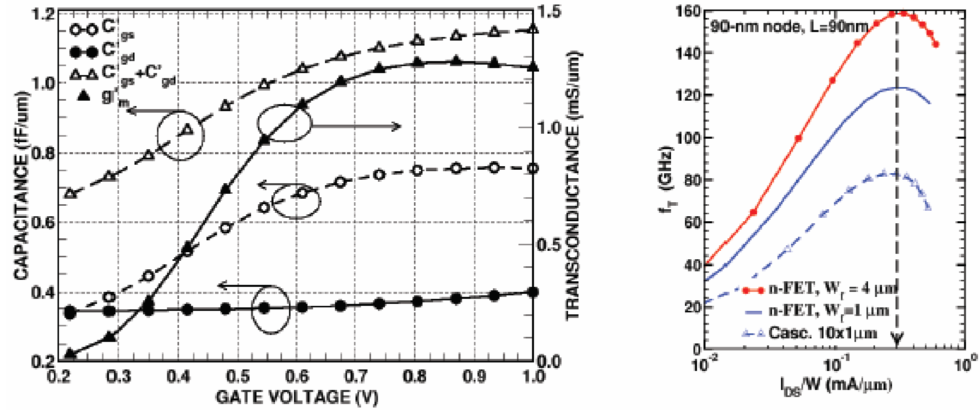

Figure 2. a) Measured small signal transconductance, gate-source, and gate-drain capacitance per unit gate width in a 90-nm n-MOSFET biased at V_{DS}=0.7 V. b) Measured f_T as a function of drain current density and gate finger width W_f for 90-nm n-MOSFETs with 10 gate fingers, and for a 90-nm n-MOSFET cascode stage.

is present only at very large gate voltages, impractical for use in most applications [8]. The main physical phenomenon responsible for the linear *I-V* characteristics of the MOSFET at moderate and high V_{GS} is the carrier mobility degradation due to the vertical electric field, on the order of 4-5 MV/cm [7]. This is a much stronger effect than velocity saturation due to the lateral drain-source field [1,5,8] which reaches a value of at most 0.2 MV/cm in a 50-nm n-MOS channel biased at V_{DS} = 1 V. As shown in Fig.1b, above V_{GS} = 0.55 V, the mobility-V_{GS} dependence can be empirically described as

$$\mu_{eff}(V_{GS}) = \frac{\mu_n}{1+\theta(V_{GS}-V_T)} \tag{1}$$

where θ captures mobility degradation due to the vertical electric field [5], and μ_n is the low field is the low field mobility. As a result, in saturation, the drain current of 90-nm MOSFETs can be approximated by

$$I_{Dsat} = \frac{\mu_n}{2}\frac{\epsilon_{ox}}{t_{ox}}\frac{W}{L}\frac{(V_{GS}-V_T)^2}{1+\theta(V_{GS}-V_T)}(1+\lambda V_{DS}) \tag{2}$$

We have recently reported that, in n-MOSFETs, as a result of constant-field scaling rules being applied rigorously by most foundries down to the 130-nm node, the peak f_T (J_{pfT}), peak f_{MAX} (J_{pfMAX}), and minimum noise figure (J_{opt}) current densities are approximately 0.3 mA/µm, 0.2 mA/µm, and 0.15 mA/µm, irrespective of foundry and technology node [8]. Furthermore, because J_{opt} is frequency-independent and close to J_{pfMAX}, the optimal noise bias and the maximum power gain bias almost coincide [9].

Measured data in Fig. 2.b. show that, in a cascode stage, the peak f_T also occurs at 0.3 mA/μm. However, depending on the gate finger width, W_f, its value is at least 40% lower than the f_T of the transistor, itself.

Fig.3 illustrates that, even in devices with different threshold voltages, the characteristic current densities remain constant while the V_{GS} at peak f_T and peak f_{MAX} varies widely. When MOSFETs are biased at a constant current density of 0.15 to 0.4 mA/μm, the impact of V_T and bias current variation on power gain and noise figure is greatly diminished. By comparison, biasing at constant V_{GS}, at low effective gate voltages, or, even

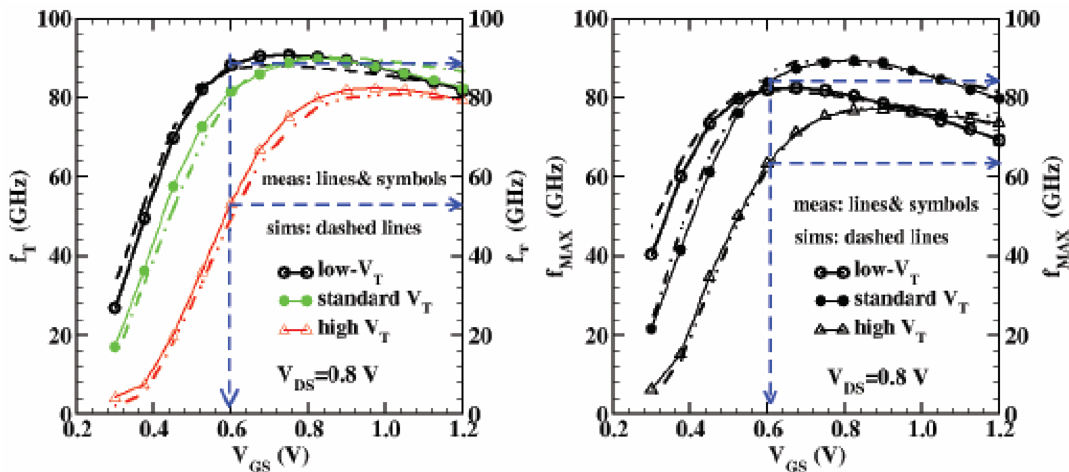

Figure 3a. f_T and f_{MAX} dependence on V_{GS} in 130-nm MOSFETs with different threshold voltages.

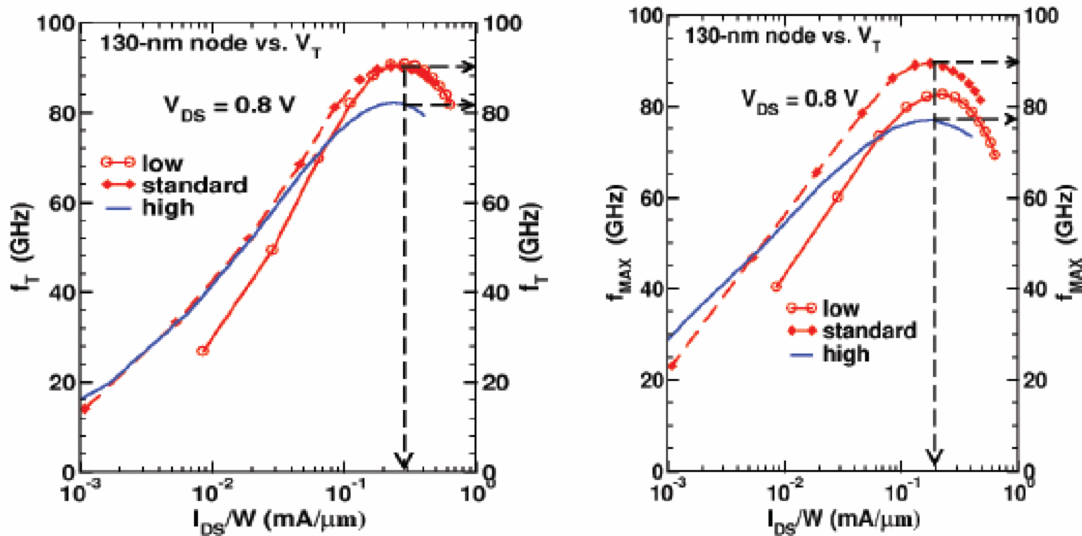

Figure 3b. f_T and f_{MAX} dependence on I_{DS}/W in 130-nm MOSFETs with different threshold voltages.

worse, in the subthreshold region, leads to significant variability in f_T and f_{MAX}. Furthermore, in a given technology node, the characteristic current densities remain invariant with gate length [8]. Therefore, the common practice of keeping V_{eff} or V_{GS} constant while increasing gate length to improve the intrinsic device voltage gain is not recommended in nano-scale CMOS. It unnecessarily slows down the transistor *over and above* the speed degradation due to the longer gate. Higher DC gain, with less degradation in speed, could be accomplished with the minimum feature size transistor biased at lower effective gate voltage.

Based on the preceding discussion, it becomes apparent that constant-current-density biasing at one of the characteristic current densities ensures circuit robustness to process variation (V_T and L), maximizes performance, and allows CMOS designs to be ported between foundries and technology nodes. Table 1 summarizes typical small signal parameters for general-purpose n-channel MOSFETs across technology nodes. Only g'_m and g'_o scale (increase) over all nodes, and through them, f_T, f_{MAX}, F_{MIN} and the noise resistance R_n, also improve with scaling. The rest of the process parameters are largely invariant between nodes down to the 130-nm node. Note that the EOT in 90-nm processes is 1.2 nm [6,7]. For a device-level comparison between state-of-the-art SiGe HBTs and 90-nm MOSFETs the reader is referred to [9].

Table 1. Measured small signal parameters for n-channel MOSFETs in different technology nodes. (simulation-only data for the 65-nm and 45-nm nodes).

Param/node	***250nm***	***180nm***	***130nm***	***90nm***	***65nm***	***45nm***
t_{ox} (nm)	5	3.5	2.5	2	2	2
V_{DD} (V)	2.5	1.8	1.2	1	1	1
g'_m (mS/μ m)	0.35	0.55	1	1.3	1.5	1.7
g'_o (mS/μ m)	0.03	0.05	0.1	0.13	0.15	0.17
C'_{gs} (fF/μ m)	1.4	1.4	1	0.75	0.5	0.35
C'_{gd} (fF/μ m)	0.45	0.45	0.45	0.35	0.3	0.25
C'_{db} (fF/μ m)	1.5	1.5	1.5	1.1	0.8	0.6
R'_s/R'_d (Ω*μ m)	100	120	150	200	200	300

In so far as the noise performance is concerned, the intrinsic (ignoring gate/base and source/emitter resistance) MOSFET and HBT can be described by two correlated noise current sources I_{n1} (gate/base noise current) and I_{n2} (drain/collector noise current) and by the Y-parameters of the noise-free small signal equivalent circuit. In the CS/CE configuration, the equivalent input noise voltage (V_n) and noise current (I_n), and the two-port noise parameters in the noise admittance formalism (R_n, G_u, Y_{cor}) can be derived as

$$V_n = \frac{-I_{n2}}{y_{21}},\ \ I_n = \frac{-y_{11}}{y_{21}} I_{n2} + I_{n1} \text{ and } I_u = \frac{Y_{cor} - y_{11}}{y_{21}} I_{n2} + I_{n1} \tag{3}$$

$$R_n = \frac{\langle I_{n2}^2 \rangle}{4kT\Delta f |(y_{21})|^2} \quad Y_{cor} = \frac{\langle I_n \bar{V}_n^* \rangle}{\langle V_n^2 \rangle} = \frac{y_{11}\langle I_{n2}^2\rangle - y_{21}\langle I_{n1} I_{n2}^* \rangle}{4kT\Delta f R_n |(y_{21})|^2} = y_{11} - \frac{\langle I_{n1} I_{n2}^* \rangle}{4kT\Delta f R_n y_{21}^*} \tag{4}$$

$$G_u = \frac{\langle I_{n1}^2 \rangle}{4kT\Delta f} + \frac{\langle I_{n2}^2 \rangle}{4kT\Delta f}\left|\frac{y_{11} - Y_{cor}}{y_{21}}\right|^2 - 2\Re\left[\frac{y_{11}^* - Y_{cor}^*}{y_{21}^*}\frac{\langle I_{n1} I_{n2}^* \rangle}{4kT\Delta f}\right] = \frac{\langle I_{n1}^2 \rangle}{4kT\Delta f} - \left|\frac{\langle I_{n1} I_{n2}^* \rangle}{4kT\Delta f R_n |Y_{21}|^2}\right|^2$$

In the CS/CE configuration, the Y-parameters of the intrinsic transistor are:

$$\left[\frac{y}{W}\right] = \begin{bmatrix} j\omega(C'_{gs} + C'_{gd}) & -j\omega C'_{gd} \\ g'_m - j\omega C'_{gd} & g'_o + j\omega(C'_{db} + C'_{gd}) \end{bmatrix};\ [y] = \begin{bmatrix} g_\pi + j\omega(C_{be} + C_{bc}) & -j\omega C_{bc} \\ g_m - j\omega C_{bc} & g_o + j\omega(C_{cs} + C_{bc}) \end{bmatrix}$$

From R_n, G_u and Y_{cor} one can obtain the familiar two-port noise parameters in the noise admittance formalism:

$$Y_{sopt} = \sqrt{G_{cor}^2 + \frac{G_u}{R_n}} - jB_{cor} \quad F_{MIN} = 1 + 2R_n(G_{cor} + G_{sopt}) \text{ and } F = F_{MIN} + \frac{R_n}{G_s}|Y_s - Y_{sopt}|^2 \tag{5}$$

Note that, for both MOSFETs and HBTs, B_{cor} is approximately equal to the imaginary part of Y_{11}, a property that is essential to achieving simultaneous noise and input impedance matching in CS/CE and cascode LNA stages.

When considering the noise from the gate and source resistance of a MOSFET, the noise parameters in CS become [10],

$$R_n = R_s + R_g + \frac{\langle i_{nd}^2 \rangle}{4kT\Delta f |(y_{21})|^2} \quad Y_{cor} = y_{11} - \frac{\langle I_{ng} I_{nd}^* \rangle}{4kT\Delta f R_n y_{21}^*} \tag{6}$$

We can repeat the noise parameter derivation for CB/CG stages:

$$V_n = \frac{-I_{n2}}{y_{21}},\ \ I_n = -\left(1 + \frac{y_{11}}{y_{21}}\right) I_{n2} - I_{n1} \text{ and } I_u = -\frac{y_{21} + y_{11} - Y_{cor}}{y_{21}} I_{n2} - I_{n1} \tag{7}$$

$$R_n = \frac{\langle I_{n2}^2 \rangle}{4kT\Delta f |(y_{21})|^2} \quad Y_{cor} = \frac{\langle I_n V_n^* \rangle}{\langle V_n^2 \rangle} = \frac{(y_{21} + y_{11})\langle I_{n2}^2 \rangle + y_{21}\langle I_{n1} I_{n2}^* \rangle}{4kT\Delta f R_n |(y_{21})|^2} = y_{21} + y_{11} + \frac{\langle I_{n1} I_{n2}^* \rangle}{4kT\Delta f R_n y_{21}^*}$$

$$G_u = \frac{\langle I_{n1}^2 \rangle}{4kT\Delta f} + \frac{\langle I_{n2}^2 \rangle}{4kT\Delta f}\left|\frac{y_{21} + y_{11} - Y_{cor}}{y_{21}}\right|^2 + 2\Re\left[\frac{y_{21}^* + y_{11}^* - Y_{cor}^*}{y_{21}^*}\frac{\langle I_{n1} I_{n2}^* \rangle}{4kT\Delta f}\right] \tag{8}$$

$$G_u = \frac{\langle I_{n1}^2 \rangle}{4kT\Delta f} - \left|\frac{\langle I_{n1} I_{n2}^* \rangle}{4kT\Delta f R_n |Y_{21}|^2}\right|^2 \tag{9}$$

with the *Y*-parameters of the transistor in CG/CB configuration given by:

$$\left[\frac{y}{W}\right]=\begin{bmatrix} g'_m+g'_o+j\omega(C'_{gs}+C'_{gd}) & -g'_o \\ g'_m-g'_o & g'_o+j\omega C'_{db}\end{bmatrix};\ [y]=\begin{bmatrix} g_m+g_o+g_\pi+j\omega(C_{be}+C_{bc}) & -g_o \\ g_m-g_o & g_o+j\omega C_{cs}\end{bmatrix}$$

R_n and G_u are similar to those of the CE/CS stage but Y_{cor} and, hence, Y_{sopt} are different. Therefore, although the minimum noise figure is almost identical in the CG/CB and CE/CS stages, the optimum noise impedance is different. We underscore that, to ease comparison between HBTs and MOSFETs, expressions for I_{n1} and I_{n2} were not substituted in eqns. (3)-(4) and (7)-(9). This also allows us to account for correlation between the base and collector noise current sources in an HBT, and between the gate and drain noise current sources in a MOSFET.

The input impedance of a MOSFET CS stage has a high Q. If the gate resistance R_g is ignored (as above), the Q is infinite. On the contrary, the input impedance of a CE stage has a low Q (typically 1) due to the presence of g_π. This property explains why circuits with bipolar transistors are easy to match over broad bandwidths and are relatively insensitive to impedance mismatch. It is the high Q of the input impedance of a MOSFET, rather than model inaccuracies, that makes MOSFET circuit design so haphazard.

3. Millimeter-wave passives

Despite the significant performance improvement and size reduction that transistors have undergone as a result of CMOS technology scaling during the last 10 years, there has been scant evidence of size scaling applying to RF inductors and transformers. Even though the quality factor of passive components has improved by taking advantage of the multi-layer copper back-end-of-line (BEOL) now available in 130-nm and 90-nm CMOS technologies, their footprint has remained largely unchanged, in locked step with the frequency of most commercial RF applications. The latter has been frozen in the 2-5 GHz range. This lack of scaling makes RF passives, by comparison with MOSFETs, more expensive with every new technology generation. It is yet another reason why digitization of RF signals in the 2-5 GHz range is beginning to make economic sense for large volume applications such as cellular phones. However, by increasing their intended frequency of application, inductors, transformers [11,12], and MIM capacitors can simply be scaled down in size to minimize their footprint above the lossy silicon substrate and, by doing so, improve their quality factor and self-resonant frequency. Q values between 10 and 20 can be easily achieved in the 50-GHz to 100-GHz range with conventional CMOS copper metallization, even as metal stripe width is reduced below 2 μm. Furthermore, by vertically-stacking inductor and transformer coils, the inductance per unit area is increased while reducing the footprint to less

than 20μm per side [11]. Above 50 GHz, it is now possible to realize good inductors and transformers whose area is comparable to smaller than that of the transistors used in the same mm-wave circuit. Thus, mm-wave circuits can benefit from the zero DC power dissipation, low-noise, and robustness to process variation, that are characteristic and desirable features of RF passive components.

4. IC building blocks

Based on the observations discussed in the preceding two sections, very simple rules can be derived for the optimal design of silicon mm-wave circuits: (i) minimize the number of transistors in order to improve circuit bandwidth, reduce noise, and maximize linearity, (ii) inductors and transformers, rather than t-lines, should be employed for impedance matching to minimize die area, (iii) in LNAs, receive mixers and VCOs, transistors must be biased at J_{opt}, and (iv) in power amplifiers and upconverters, transistors must be biased at the peak f_{MAX}/f_T current density. In MOSFET circuits, (iii) and (iv) above result in identical size and bias current, irrespective of technology node. The latter simplifies porting of designs between technology nodes, making it a rather effortless exercise.

4.1. Power amplifiers

The International Technology Roadmap for Semiconductors (ITRS) [3] figure of merit for power amplifiers, FoM_{PA}, lumps together the output power, P_{OUT}, the power gain, G, and the power added efficiency, PAE.

$$FoM_{PA} = P_{OUT} \times G \times PAE \times f^2 \tag{10}$$

Through the f^2 term, it accounts for the fact that the amplifier gain and output power degrade with frequency. In order to maximize PAE, a power amplifier, including those operating at mm-wave frequencies, typically consists of a common-emitter (CE) or common-source (CS) stage with a large transistor or a large number of smaller transistors connected in parallel. Unless the output power is lower than 20 dBm, the output matching network is implemented off chip to minimize losses. The output power is limited by the maximum allowable voltage for the safe operation of the transistor, V_{MAX}.

The linearity of a high frequency amplifier depends on the linearity of its power gain as a function of the applied input voltage or current. The most popular way of evaluating transistor linearity has been the ratio of the small signal transconductance and its second order derivative with respect to the gate voltage [13]. This approach, based on a Taylor series expansion of the I-V characteristics, is adequate only for low-frequency amplifiers

because it does not capture nonlinear capacitance effects. It has been proposed recently [9] that the most elegant way of capturing high-frequency non-linearity, without having to resort to lengthy Volterra series derivations [14], is to use the f_T or f_{MAX} characteristics of the device. Since f_{MAX} is obtained directly from the high-frequency dependence of the maximum available power gain of the transistor, it is most suitable for PA linearity assessment. Transistor linearity depends on the flatness of the $f_{MAX}(V_{GS})$ and $f_{MAX}(I_{DS})$ characteristics in the region around the peak. To facilitate the comparison, a generic linearity figure of merit (to which *IIP3* is proportional) is calculated in dB using the second order derivative of g_m, f_T, or f_{MAX} with respect to V_{GS} or I_{DS}.

$$IIP3 \propto 20\log_{10}\left[\frac{g_m}{\frac{\partial^2 g_m}{\partial V_{GS}^2}}\right], \text{ or } IIP3 \propto 20\log_{10}\left[\frac{f_{MAX}}{\frac{\partial^2 f_{MAX}}{\partial V_{GS}^2}}\right] \text{ or } IIP3 \propto 20\log_{10}\left[\frac{f_T}{\frac{\partial^2 f_T}{\partial V_{GS}^2}}\right] \quad (11)$$

$$IIP3 \propto 20\log_{10}\left[\frac{f_{MAX}}{\frac{\partial^2 f_{MAX}}{\partial I_{DS}^2}}\right], \; IIP3 \propto 20\log_{10}\left[\frac{f_T}{\frac{\partial^2 f_T}{\partial I_{DS}^2}}\right], \text{ or } IIP3 \propto 20\log_{10}\left[\frac{g_m}{\frac{\partial^2 g_m}{\partial I_{DS}^2}}\right] \quad (12)$$

Fig. 4 compares various linearity figures of merit as functions of V_{GS} and I_{DS}. It is important to note that, since the derivatives are taken with respect to different variables, the Y-axes in Fig. 4.a. and 4.b. have different values. The familiar linearity "sweet spot" can be identified in Fig. 4.a. for a narrow range of gate voltages, just above the transistor threshold voltage. However, it occurs at different V_{GS} values depending on whether it is derived from f_{MAX}, f_T or g_m. More interestingly, if we use the derivative with respect to bias current, the "sweet spot" vanishes altogether, indicating that it is most likely a mathematical artifact, and thus of little practical value. If one is to ensure that the transistor is biased for the best possible linearity, V_{GS} should be large, corresponding to a current density of 0.3 mA/µm or higher. As a result, the design of a class A PA becomes an exercise in optimally biasing the transistor for linearity at a current density corresponding to the peak f_T (0.3mA/µm in MOSFETs). In a switching PA (class D, E or F), where the MOSFET is operated as an ideal digital switch, the gate voltage waveform must be optimized such that the peak current through the transistor reaches up to 0.3mA/µm, as in a MOS-CML gate [15].

The linear voltage swing at the input/output of the transistor decreases with every new node while the current swing remains constant over nodes [8]. It follows immediately that the bias current and transistor size must be increased to generate the same power as in older nodes. For example, at the

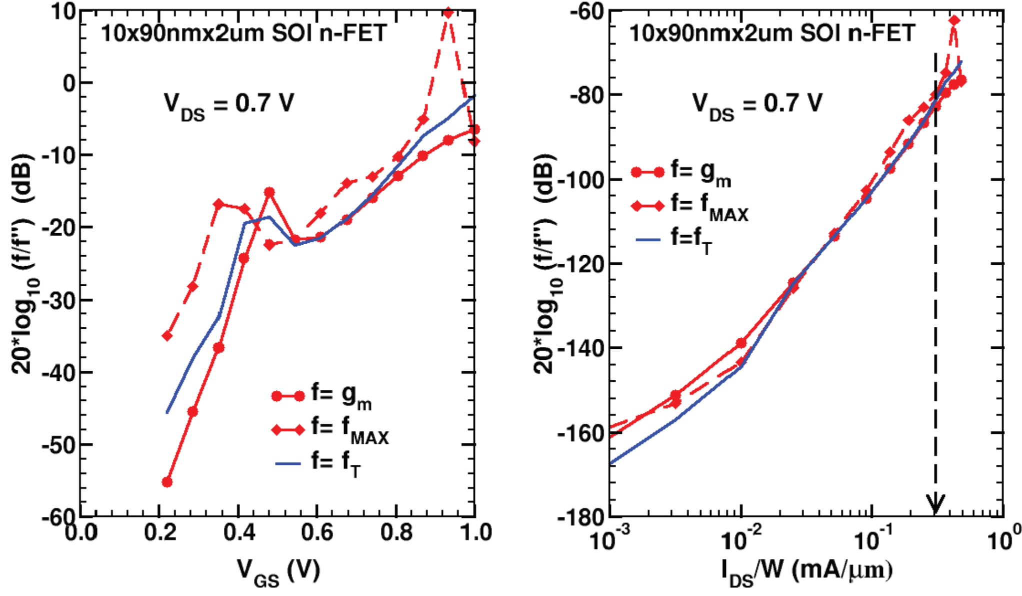

Figure 4. Linearity as a function of a) V_{GS} and of b) I_{DS}.

output compression point, the available power into a matched load for a class A power amplifier stage is

$$OP_{1dB} \approx \frac{\Delta I_{DS} V_{MAX}}{8} = 50 \frac{\mu W}{\mu m}$$ in 90-nm MOSFETs and

$$OP_{1dB} \approx \frac{\Delta I_C V_{MAX}}{8} = 376 \frac{\mu W}{\mu m}$$ in SiGe HBTs

where V_{MAX} is typically 1 V for a 90-nm MOSFET and 3 V for a 160-GHz SiGe HBT. At comparable f_T/f_{MAX}, devices that can tolerate larger voltage swing have a clear advantage over nanoscale CMOS.

4.2. Low-noise amplifiers

The linearity of a low noise amplifier can be described by the output referenced third order intercept point ($OIP3 = G \times IIP3$ where G is the gain and $IIP3$ is the input referenced third order intercept point). The minimum signal that is correctly amplified by a LNA is provided by the noise factor of the amplifier, F. However, (F-1) is a better measure of the contribution of the amplifier to the total noise, since it allows the ratio between the noise of the amplifier N_a and the noise already present at the input N_i to be directly evaluated. ITRS combines these two performance figures with the total power consumption P. The resulting figure of merit captures the dynamic range of an amplifier versus the necessary DC power. For roadmapping purposes ITRS has chosen a performance measure that is

independent of frequency and thus independent of the specific application [3]. This was achieved by assuming that the LNA is formed by a single amplification stage, so that the FoM scales linearly with operating frequency *f*. With these approximations and assumptions, a figure of merit (FoM_{LNA}) for LNAs is defined:

$$FoM_{LNA} = \frac{G \times IIP3 \times f}{(F-1) \times P} = \frac{OIP3 \times f}{(F-1) \times P} \tag{13}$$

Note that, in a well-designed LNA stage, *OIP3* only depends on the supply voltage, and that *F-1* increases linearly with frequency.

Common-base (CB) [16], MOSFET cascodes [17], bipolar cascodes [18], and common-gate (CG) [19] topologies, illustrated in Fig. 5, have been considered for LNAs operating above 50 GHz. In the 90-nm and 65-nm nodes, because of the excellent performance of the p-channel MOSFET, a CMOS inverter stage can also be employed. In all cases, the transistor size and bias current that result in optimal noise impedance matching to 50 Ω is significantly smaller, in the 1-to-5 mA range, than at 2-to-10 GHz. Although this may look attractive from the power dissipation point of view, in conjunction with the lower breakdown voltages typical of transistors suitable for the mm-wave region, it results in poor linearity [15,18] and ultimately reduces system dynamic range. Linearity is particularly bad in CB and CG stages. To satisfy input impedance matching conditions, the bias current is set to 3 mA or less, limiting the linear input range to less than 150 mV_{pp}. The difficulty of simultaneously matching the input and noise impedance of CB/CG stages adds to their list of problems. It must be pointed out, however, that in the 70-GHz to 80-GHz range concomittant input and noise impedance matching is achievable for CB/CG stages.

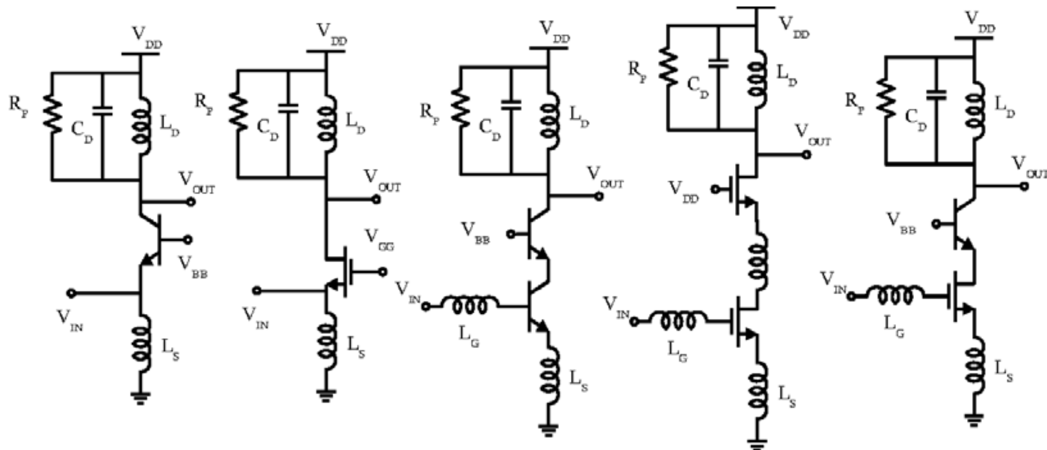

Figure 5. mm-wave LNA topologies: a) common base, b) common gate, c) bipolar cascode, d) MOS cascode, e) HBT-MOS (BiCMOS) cascode.

For a more rigorous analysis, *Y*-parameters can be used to assess the merits of each topology in terms of small-signal gain, input impedance, output impedance, and isolation. In conjunction with feedback theory and *Z*-parameters, they can also be employed to investigate the best noise matching techniques.

In general, for a two-port, the expressions of the input admittance Y_{IN} and voltage gain G_V are given by [20]:

$$Y_{IN}=Y_{11}-\frac{Y_{12}Y_{21}}{Y_L+Y_{22}} \text{ and } G_v=\frac{-Y_{21}}{Y_L+Y_{22}} \tag{14}$$

The input impedance of a CB transistor, approximately $1/g_m$, is independent of the load. However, due to its large g_o, the CG stage has relatively poor isolation. Consequently, matching common gate transistors to 50 Ω at microwave/mm-wave frequencies is difficult. When the output is a resonant circuit with large equivalent resistance, the input impedance becomes a function of the load impedance R_p.

$$Z_{IN}\approx\frac{1}{g_m}+\frac{R_P}{1+\frac{g_m}{g_o}} \tag{15}$$

As an example, if R_p=800 Ω and g_m/g_o =15, Z_{IN} will be larger than 50 Ω even if a large MOSFET with large bias current is employed.

For a cascode stage, the input admittance and voltage gain are given by (16) and (17), respectively.

$$Y_{IN}=Y_{11,CS}-\frac{Y_{12,CS}Y_{21,CS}}{Y_{22,CS}+Y_{11,CG}-\frac{Y_{12,CG}Y_{21,CG}}{Y_L+Y_{22,CG}}} \tag{16}$$

$$G_v(casc)=\frac{-Y_{21,CS}}{Y_{12,CG}-\left[\frac{Y_{11,CG}+Y_{22,CS}}{Y_{21,CG}}\right][Y_L+Y_{22,CG}]} \tag{17}$$

For tuned low-noise amplifiers it is useful to derive the *Y*-parameters of a transistor two-port with inductive degeneration and with source, drain, and gate resistance [21]:

$$Y_{11}=\frac{y_{11}+(R_s+R_d+j\omega L_S)\Delta y}{N} \quad Y_{12}=\frac{y_{12}-(R_s+j\omega L_S)\Delta y}{N} \tag{18}$$

$$Y_{21}=\frac{y_{21}-(R_s+j\omega L_S)\Delta y}{N} \quad Y_{22}=\frac{y_{22}+(R_s+R_d+j\omega L_S)\Delta y}{N} \tag{19}$$

where:

$$N=1+(R_s+j\omega L_S)(y_{11}+y_{12}+y_{21}+y_{22})+R_d y_{22}+R_g y_{11}+\Delta y+\Delta R$$

$$\Delta y=y_{11}y_{22}-y_{12}y_{21} \quad \Delta R=(R_s+j\omega L_S)R_d+(R_s+j\omega L_S)R_s+R_d R_g$$

The last three topologies in Fig. 5 employ series feedback through a degeneration inductor L_S. In order to understand how to systematically apply feedback theory to design low-noise amplifiers, the noise parameters of parallel-parallel and series-series connected two ports will be derived next. In the case of an amplifier and a feedback network connected in parallel, as in Fig. 6, it is convenient to employ *Y*-parameters and the noise admittance formalism.

$$\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \begin{bmatrix} Y_{11f} & Y_{12f} \\ Y_{21f} & Y_{22f} \end{bmatrix} + \begin{bmatrix} Y_{11a} & Y_{12a} \\ Y_{21a} & Y_{22a} \end{bmatrix} \tag{20}$$

$$i_n = i_{nf} + i_{na} + \frac{Y_{11a}Y_{21f} - Y_{21a}Y_{11f}}{Y_{21}} v_{nf} + \frac{Y_{11f}Y_{21a} - Y_{21f}Y_{11a}}{Y_{21}} v_{na} \tag{21}$$

$$v_n = \frac{Y_{21f}v_{nf} + Y_{21a}v_{na}}{Y_{21}} \tag{22}$$

A straightforward application of these formulae is to a connection of N identical two-ports in parallel. In this case one obtains: $R_n = R_{ni}/N$; $Y_{sopt} = NY_{sopti}$; $F_{MIN} = F_{MINi}$. By connecting transistors or gate fingers in parallel, the

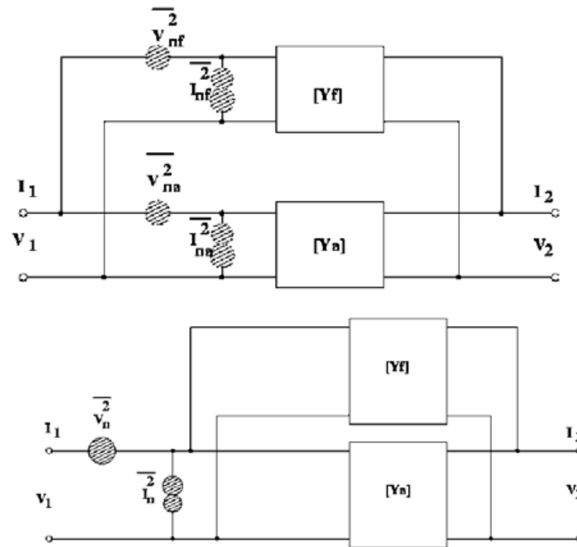

Figure 6. a) Two noisy two-ports connected in parallel. b) Noise equivalent circuit representation of the two two-ports connected in parallel.

minimum noise factor is preserved while the optimum source impedance and sensitivity to source mismatch are reduced. This provides the IC designer with a powerful tool for realizing noise matching simply by controlling the size of transistors and without (to first order) compromising the noise factor.

If the unilateral amplifier approximation holds, as in the case of an amplifier with transimpedance feedback

$$Y_{21} \approx Y_{21a};\ \ Y_{12} \approx Y_{12f} \tag{23}$$

one obtains:

$$v_n \approx v_{na} \qquad i_n = i_{nf} + i_{na} + Y_{11f}(v_{na} - v_{nf}) \qquad i_n = i_u + Y_{cor} v_n \tag{24}$$

$$R_n \approx R_{na} \qquad G_u = G_{uf} + G_{ua} + |Y_{corf} - Y_{11f}|^2 R_{nf} \qquad Y_{cor} = Y_{cora} + Y_{11f} \tag{25}$$

We note that the noise voltage of the amplifier with parallel feedback is equal to that of main amplifier. The noise currents of the amplifier and feedback networks add while Z_{in} and Z_{sopt} decrease. F_{MIN} increases if the feedback network has resistive elements. If it is purely reactive, it will not degrade the noise figure. One can conclude that parallel feedback can be used for noise matching in situations where the noise impedance of the original two-port is higher than that of the source impedance.

Another case of interest is that of the CMOS inverter which, from the small signal and noise point of view, can be represented as a parallel connection of the n-MOSFET and p-MOSFET equivalent circuits. Let us consider a CMOS inverter in which the p-MOSFET is sized twice as large as the n-MOSFET to ensure symmetrical I-V characteristics and transconductance i.e. $W_p=2W_n$, $g'_{mp}=g'_{mn}/2$, $C'_{gsp} = C'_{gsn}$; $C'_{gdp} = C'_{gdn}$, $C'_{dbp}=C'_{dbn}$; $g_{mn}=g_{mp}$; $I_{Dn}=I_{Dp}$; $R_{gn}=R_{gp}$, $R_{dsp}=2R_{dsn}$. It is relatively straightforward to arrive at the following expressions for the equivalent f_T, F_{MIN} and noise parameters of the CMOS inverter.

$$\frac{1}{2\pi f_{T,CMOS}} = \frac{3}{2}\frac{C_{gsn}+C_{gdn}}{g_{mn}} + \frac{3}{2}C_{gdn}(R_{sn}+R_{dn});\ f_{T,CMOS} = \frac{2}{3}f_{T,nMOS} \tag{26}$$

$$F_{MIN,CMOS} - 1 = \frac{3}{2}\left[F_{MIN,n-MOS} - 1\right] \tag{27}$$

$$\frac{g_m}{I_{DS}} = \frac{2g_{mn}}{I_{DS}};\ \frac{R_n}{I_{DS}} = \frac{R_{nn}}{2I_{DS}};\ \frac{R_{SOP}}{I_{DS}} = \frac{R_{SOPn}}{4I_{DS}};\ X_{SOP} = \frac{X_{SOPn}}{3};\ X_{IN} = \frac{X_{INn}}{3} \tag{28}$$

Compared to an n-MOSFET-only implementation, when using a CMOS inverter the inductor size and the bias current can be made 3 times smaller in order to realize a noise-impedance-matched LNA stage. The

slight degradation in noise figure due to the p-MOSFET will be offset by the lower loss of the matching inductors. Tables 2 and 3 compare the simulated small signal and noise parameters of 90-nm n-MOSFETs and CMOS inverters biased at the optimum noise current density.

Table 2. Noise parameters for 90-nm 20x1µm n-MOSFETs at 0.15mA/µm, f_T = 120 GHz.

f(GHz)	R_n (Ω)	R_{sopt} (Ω)	X_{sopt} (Ω)	X_{in} (Ω)	F_{MIN}	NF_{MIN} (dB)
5	55.4	1060	1353	-1109	1.04	0.17
30	53.81	179.6	224.8	-184	1.26	1
60	53.48	92.9	112.4	-92.9	1.58	1.95

Table 3. Noise parameters for 90-nm 20/(40)x1µm CMOS inverter at 0.15(0.075) mA/µm, f_T = 80 GHz.

f(GHz)	R_n (Ω)	R_{sopt} (Ω)	X_{sopt} (Ω)	X_{in} (Ω)	F_{MIN}	NF_{MIN} (dB)
5	32.99	232.1	449.45	-349.6	1.06	0.26
30	30.76	40.9	74.75	-59.15	1.4	1.47
60	28.96	22.25	37.38	-31	1.88	2.75

Similarly, to analyze circuits consisting of two-ports connected in series, one can use *Z*-parameters and the noise impedance formalism

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = \begin{bmatrix} Z_{11f} & Z_{12f} \\ Z_{21f} & Z_{22f} \end{bmatrix} + \begin{bmatrix} Z_{11a} & Z_{12a} \\ Z_{21a} & Z_{22a} \end{bmatrix} \quad (29)$$

$$v_n = v_{nf} + v_{na} + \frac{Z_{11a}Z_{21f} - Z_{21a}Z_{11f}}{Z_{21}} i_{nf} + \frac{Z_{11f}Z_{21a} - Z_{21f}Z_{11a}}{Z_{21}} i_{na} \qquad i_n = \frac{Z_{21f}i_{nf} + Z_{21a}i_{na}}{Z_{21}} \quad (30)$$

If the unilateral amplifier approximation is invoked in (30), as in the case of a tuned amplifier with inductive degeneration, then

$$Z_{21} \approx Z_{21a};\; Z_{12} \approx Z_{12f} \qquad i_n \approx \frac{Z_{21a}}{Z_{21}} i_{na} \approx i_{na} \qquad v_n = v_{nf} + v_{na} + Z_{11f}(i_{na} - i_{nf}) \qquad v_n = v_u + Z_{cor} i_n \quad (31)$$

$$G_n \approx \frac{|Z_{21a}|^2}{|Z_{21}|^2} G_{na} \approx G_{na} \qquad Z_{cor} = Z_{cora} + Z_{11f} \qquad R_u = R_{uf} + R_{ua} + |Z_{corf} - Z_{11f}|^2 G_{nf} \quad (32)$$

$$z_{sopt} = \sqrt{R_{sopta}^2 + \frac{R_{uf}}{G_{na}} + 2R_{cora}\Re(z_{11f}) + \Re^2(z_{11f}) + \frac{|Z_{corf} - z_{11f}|^2 G_{nf}}{G_{na}}} + j[X_{sopta} - \Im(z_{11f})] \quad (33)$$

$$F_{MIN}=1+2G_{na}[R_{cora}+R_{sopt}+\Re(Z_{11f})] \tag{34}$$

We note that the noise current of the amplifier with series feedback is equal to that of the main amplifier. The noise voltages add while Z_{in} and Z_{sopt} increase. F_{MIN} increases if the feedback has resistive elements. If the feedback network is purely reactive, it will not degrade the noise figure of the amplifier. One can conclude that series feedback should be employed for noise matching in situations where the noise impedance of the original two-port is lower than that of the source impedance.

For a connection of N identical two-ports in series, one obtains $G_n = G_{ni}/N$; $Z_{sopt} = NZ_{sopti}$; $F_{MIN} = F_{MINi}$. By connecting two-ports in series, the minimum noise figure is preserved while the optimum source impedance is increased and sensitivity to source mismatch is reduced. One such circuit is the differential pair which can be represented as a series connection of two identical half circuits, indexed with 1 and 2. Assuming that the noise sources in the two half circuits are not statistically correlated, one can derive the expressions for the noise parameters of the differential pair

$$F=1+\frac{\overline{v_{u1}^2}+\overline{v_{u2}^2}}{\overline{v_{ns1}^2}+\overline{v_{ns2}^2}}+\frac{|Z_{cor1}+Z_{s1}|^2\overline{i_{n1}^2}+|Z_{cor2}+Z_{s2}|^2\overline{i_{n2}^2}}{\overline{v_{ns1}^2}+\overline{v_{ns2}^2}} \tag{35}$$

$$F=1+\frac{R_{u1}+R_{u2}}{R_{s1}+R_{s2}}+\frac{|Z_{cor1}+Z_{s1}|^2G_{n1}+|Z_{cor2}+Z_{s2}|^2G_{n2}}{R_{s1}+R_{s2}} \tag{36}$$

$$Z_{s1}=Z_{s2}=\frac{Z_s}{2};\ R_{u1}=R_{u2};\ G_{n1}=G_{n2};\ Z_{cor1}=Z_{cor2} \tag{37}$$

$$F=1+\frac{2R_{u1}}{R_s}+\frac{|2Z_{cor1}+Z_s|^2\frac{G_{n1}}{2}}{R_s} \qquad R_u=2R_{u1};\ G_n=\frac{G_{n1}}{2};\ Z_{cor}=2Z_{cor} \tag{38}$$

$$X_{soptdif}=2X_{sopt};\ R_{soptdif}=\sqrt{4R_{cor1}^2+4\frac{R_{u1}}{G_{n1}}}=2R_{sop} \qquad F_{MINdif}=1+|2R_{sopt}+2R_{cor1}|\frac{G_{n1}}{2}=F_{MIN} \tag{39}$$

Based on the theoretical formulation developed above, we are now in a position to define simple LNA design scaling equations. For all topologies presented in Fig. 5, the design starts by biasing the LNA stage at the optimum noise current density. For HBTs, the optimum noise current density is frequency dependent [9], changes with the technology-node [9], and is different for a cascode stage than for a CS or CB stage [18]. For MOSFETs, the optimum noise current density, 0.15mA/μm [8], is independent of frequency, technology node, and of the LNA topology.

Next, in a CE/CS or a cascode LNA stage, while maintaining the transistors biased at the optimum noise current density, their size is set such that the noise impedance of the stage becomes equal to the signal source impedance, Z_0. At this point, the size and the bias current of the stage are fixed. In the case of a CB or CG stage, the size and bias current are chosen such that the input impedance is equal to Z_0. In general, this will conflict with noise matching conditions and with sizing and biasing for maximum linearity.

The goal of the third step in the LNA design is input impedance matching. For a cascode or a CS/CE stage, one can derive the expression of the input admittance as a function of the cutoff frequency of that particular stage, and thus elegantly account for the Miller capacitance:

$$h_{21}=\frac{y_{21}}{y_{11}}=\frac{f_T}{jf} \qquad y_{11}=\frac{f}{f_T}y_{21}\approx j\frac{f}{f_T}g_m \tag{40}$$

With this observation, fairly accurate expressions for the two matching inductors L_S and L_G are obtained

$$L_S=\frac{Z_0}{2\pi f_T} \qquad Z_{in}=Z_O+j\omega(L_G+L_S)+\frac{1}{j\omega C_{in}} \qquad L_G=\frac{1}{\omega^2 C_{in}}-L_S \tag{41}$$

as well as the upper bound on the power gain.

$$G\approx\left(\frac{f_T}{f}\right)^2\frac{R_P}{Z_0} \tag{42}$$

The last expression indicates that (i) transistors with high f_T are needed for high gain, (ii) the power gain decreases with the square of the frequency, (iii) a high Q load results in larger R_p and higher power gain, and (iv) increasing the input impedance (Z_0) will compromise gain. The latter aspect is very important because it has been suggested that one way to reduce the power dissipation of a MOSFET LNA, while retaining noise impedance matching, is to match to a higher input impedance level. Obviously, such an approach will result in lower power gain, an outcome which might not be acceptable in most practical cases.

For a MOSFET cascode or CS LNA operating at frequency f, the design equations can be recast as functions of largely technology-node independent parameters:

$$R_{sopt}=\frac{k}{2\pi f W(C'_{gs}+C'_{gd})}=Z_0-R_g,\ W=\frac{k}{2\pi f Z_0(C'_{gs}+C'_{gd})} \qquad I_{DS}=W\times 0.15\frac{mA}{\mu m} \tag{43}$$

$$L_S=\frac{Z_0}{2\pi f_T}\ \text{and}\ L_G=\frac{Z_0}{2\pi f}-L_S \tag{44}$$

where $k = 0.25 .. 1$, depending on the choice of gate finger width, W_f. The finger width dependence of the optimal total gate width W makes MOSFET LNA design rather sensitive to layout, number of gate contacts, and to the series resistance of the matching inductors. Even so, if the layout is not modified, the bias current and size of transistors do not change over nodes while f_T improves and L_S decreases, leading to increased gain and lower noise figure. One should note that, even at mm-wave frequencies, the bias current of MOS LNAs is larger than in an HBT LNA. It can be reduced only at the expense of larger noise figure. At frequencies up to 10 GHz, CMOS inverter LNAs can be used to reduce the bias current and improve the g_m/I_{DS} and I_{DS}/R_{sopt} ratios.

Just as at 2-5 GHz, the CS/CE and cascode stages can be simultaneously noise- and input-impedance matched at mm-wave frequencies [18]. Furthermore, their linearity is much better than that of CB/CG stages and can be improved simply by increasing the transistor size and current, without affecting the input impedance match and with negligible impact on noise matching. As the measured data in Fig. 2.b indicate, the MOSFET cascode has relatively low f_T. In compensation, an inductor may be placed between the two transistors to tune out the middle pole (Fig. 5.d). Nevertheless, as in PAs, for applications above 50 GHz, a CS stage should be preferred to a MOSFET cascode stage.

The bipolar cascode continues to be the LNA topology of choice because it combines the excellent isolation of the CB stage with the high power gain, good linearity and ease of simultaneous noise and input impedance matching, typical of the CE/CS stage. Furthermore, the high g_m/I_C ratio and small noise resistance make this stage, unlike a MOSFET one, insensitive to impedance mismatch and model inaccuracy.

Fig. 7 shows the schematics of a 2-stage SiGe-HBT LNA whose design was scaled from 6 GHz to 52 GHz. Unlike other LNAs reported in this frequency range [16,17,19] it uses inductors, regular 60μmx60μm pads, and it includes all bias circuitry in a 300μmx400μm die [18]. It was fabricated in a production 0.18-μm SiGe BiCMOS process with f_T/f_{MAX} of 155 GHz. More recently, the same design was scaled to 65 GHz and 77 GHz. Its layout is shown in Fig.8, side by side with that of a 90-nm RF-CMOS, single-stage cascode LNA. The CMOS LNA uses the same inductors as the SiGe LNA and, predictably, its gain peaks at the same frequency, as shown in Fig. 9. This result points to the importance of accurate passive component models. The 2.5-dB gain of the CMOS LNA, much smaller than the 20-dB gain of the HBT LNA, can be explained by the fact that only a 5-mA cascode stage is used to drive the 50-Ω load directly. Simulations indicate that, with a two-stage 90-nm CMOS cascode design, over 15 dB of gain is achievable at 65 GHz.

As illustrated in Fig. 10, the input compression point of the SiGe-HBT LNA, measured at 50 GHz, is -14 dBm, 8 dB higher than those of the CB [16] and CG [19] LNAs, and slightly better than that of the CMOS cascode LNA in [17]. The latter consumes 3 times larger current and dissipates twice the power.

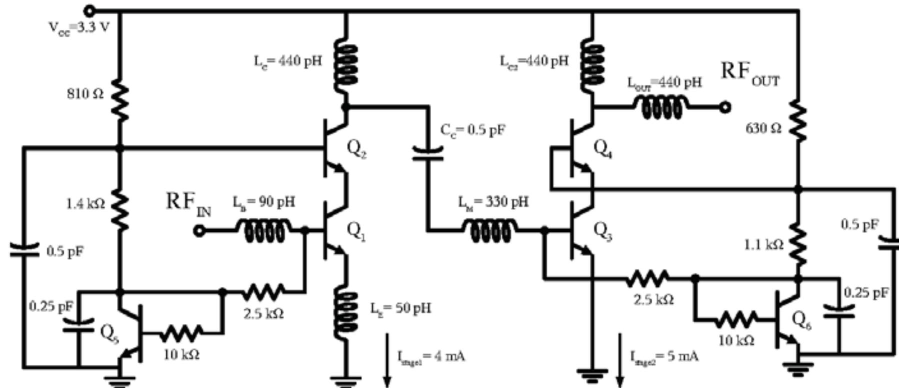

Figure 7. 52-GHz SiGe HBT LNA schematics.

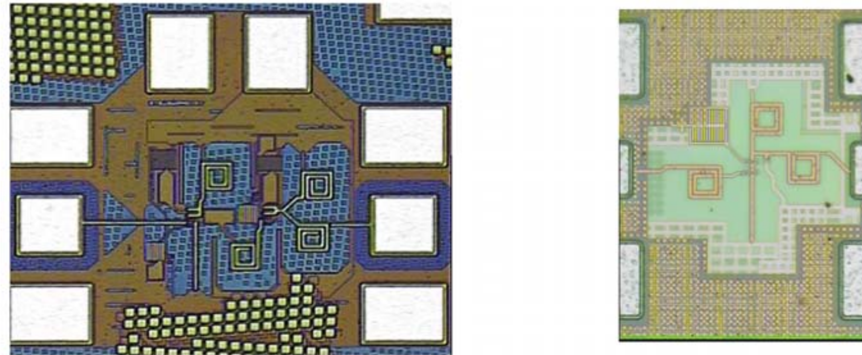

Figure 8. a) SiGe HBT 2-stage and b) 90-nm CMOS 1-stage cascode LNA layouts, drawn at similar scale.

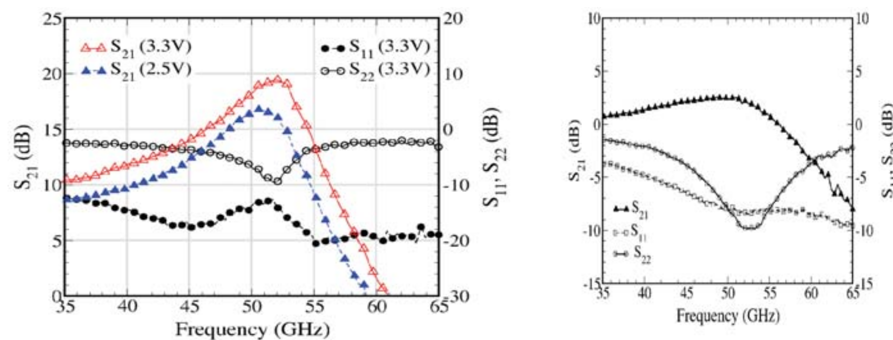

Figure 9. Measured S parameters of a) two-stage SiGe-HBT and b) single stage 90-nm CMOS LNA.

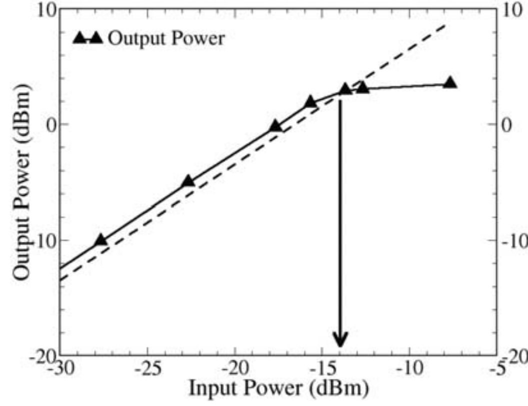

Figure 10. Input and output compression points of SiGe-HBT LNA measured at 50 GHz.

4.3. Mixers

Even though the race is on for the upconvert and downconvert mixer to be replaced by direct digital synthesis and ΔΣ ADCs at RF frequencies [4], its utility at mm-wave frequencies is firmly established. The Gilbert cell (Fig. 11) remains the topology of choice for mixers even at mm-waves. The conversion gain of a Gilbert cell mixer is, to first order, equal to that of a cascode LNA multiplied by 2/π to account for the switching action of the mixing quad.

$$G_v = \frac{2}{\pi}\frac{-Y_{21}}{Y_L + Y_{22}} \approx \frac{-2}{\pi} g_m R_p \tag{45}$$

R_p is the equivalent parallel resistance of the resonant tank consisting of $Y_L + Y_{22}$ at the frequency of interest at the output of the mixer. This corresponds to the IF frequency in a downconverter, and to the RF frequency in an upconverter. Depending on its deployment in the transmitter or in the receiver, the mixer is designed along the same guidelines as a linear, moderate power amplifier, or as a linear low-noise amplifier.

In a downconverter, the bottom (transconductor) pair in Fig. 11 is biased at the optimum noise figure current density (0.15 mA/μm in n-MOSFET irrespective of frequency, variable as a function of frequency in HBTs) or at the optimal linearity bias which corresponds to the peak f_T current density in both MOSFETs and HBTs.

$$W = \frac{I_{TAIL}}{2 J_{pfT}}; \text{ and } A_E = \frac{I_{TAIL}}{2 J_{pfT}} \tag{46}$$

In an upconverter, the bottom (transconductor) pair should be biased at the optimal linearity bias which corresponds to the peak f_T current density in

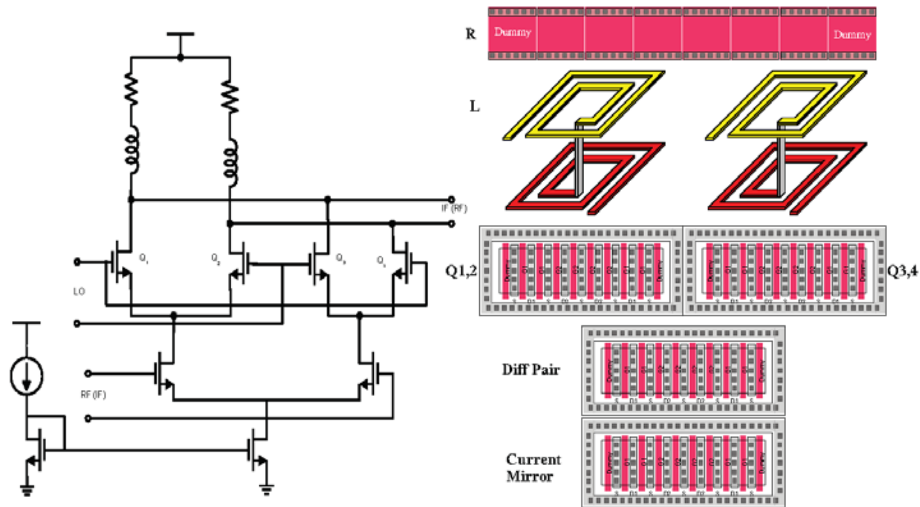

Figure 11. Schematics and layout sketch of a Gilbert cell mixer in CMOS.

both MOSFETs and HBTs. In both upconverters and downconverters, the mixing quad transistors are sized as in a CML/ECL gate [11,15], because they operate as a digital switch. Since in MOSFETs $J_{pfT} = 1.5J_{pFMAX}$ while in HBTs $J_{pfT} = J_{pFMAX}$ the transistor sizing equations become remarkably similar:

$$W=\frac{I_{TAIL}}{3J_{pfMAX}} \text{ and } A_E=\frac{I_{TAIL}}{3J_{pfMAX}} \quad (47)$$

Layout symmetry is critical in realizing high isolation between the RF, LO, and IF ports. In image reject architectures, layout symmetry, transistor and resistor matching limit the amount of image rejection. Fig. 11 illustrates a possible Gilbert-cell mixer layout that is suitable for implementation in strained channel MOSFET technologies. Note that, to avoid open-circuit terminations due to resonances at mm-wave frequencies, the load resistors are placed between the vertically stacked inductors and the positive power supply. Some of the layout details shown in Figs. 11 and 12 include (i) identical finger width for all transistors to avoid mismatch due to STI-induced strain, (ii) identical orientation and size for all resistors, (iii) identical orientation for all transistors, (iv) dummy gates and dummy resistors for optical lithography correction as well as to minimize edge strain, (v) differential pair transistors with interspersed gate fingers and sharing the same p-well for improved matching, and (vi) substrate contacts entirely surrounding transistors to minimize crosstalk, transistor substrate resistance, and to ensure model accuracy.

Fig. 13, reproduces the schematic and layout photograph of a DC-to-50 GHz broadband upconverter/downconverter implemented in 130nm SiGe

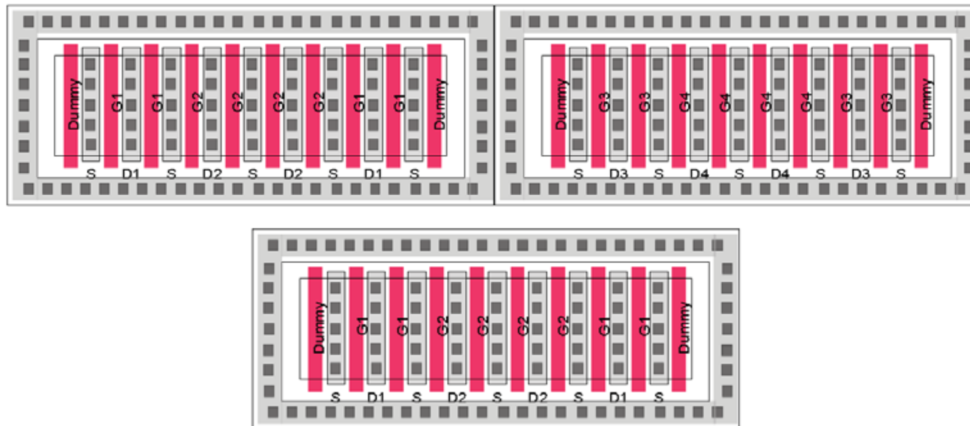

Figure 12. Layout detail of 90-nm CMOS Gilbert cell that minimizes the impact of strain, gate length, t_{ox} and doping variation.

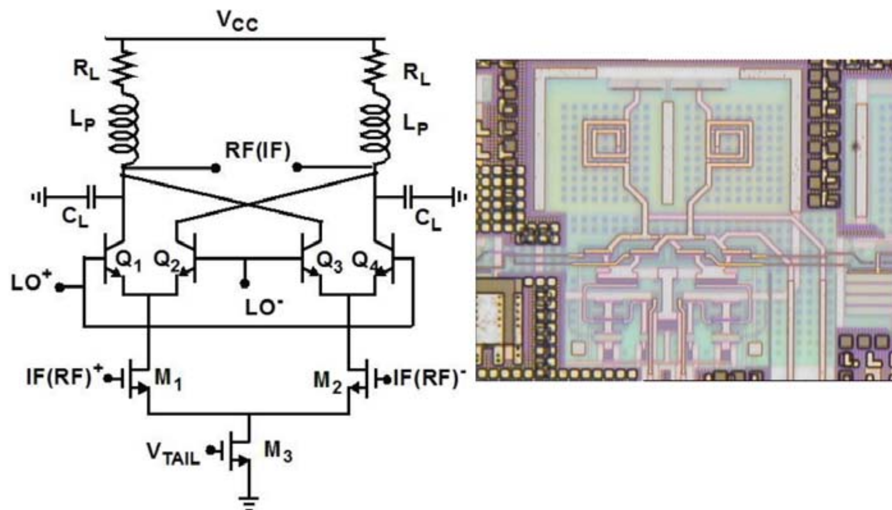

Figure 13. Schematics and layout of a DC-to-50 GHz upconvert/downconvert Gilbert cell mixer in 130-nm SiGe BiCMOS technology [11].

BiCMOS technology [21]. It employs a MOS-HBT cascode stage that maximizes linearity without compromising noise figure or power gain. At the same time, because the mixing quad is implemented using SiGe HBTs, the LO drive requirement is minimized.

4.4. Voltage-controlled oscillators

Another key component of RF and mm-wave signal processing systems is the VCO. According to the ITRS [3], the main design objectives for VCOs are to minimize the timing jitter of the generated waveform (or, equivalently, the phase noise) and to minimize the power consumption. At mm-waves, one must also consider the output power generated by the VCO

as an important design goal. Since transistor power decreases with the square of frequency, it becomes apparent that providing adequate output voltage swing to switch mixers and flip-flops is not trivial. From these parameters a figure of merit (FoM_{VCO}) is defined:

$$FoM_{VCO}=\left(\frac{f_{osc}}{\Delta f}\right)^2\frac{P_{OUT}}{L[\Delta f]P} \tag{48}$$

Here, f_{osc} is the oscillation frequency, $L[\Delta f]$ is the phase noise power spectral density measured at a frequency offset Δf from f_{0sc} and taken relative to the carrier power, P_{OUT} is the generated power, and P is the total power consumption.

The definition neglects the tuning range of the VCO since the necessary tuning range strongly depends on the application. However, just to cover process and temperature variation, a minimum of 10% to 15% of the oscillation frequency is necessary. In this tuning range, FoM_{VCO} should be evaluated at the frequency where phase noise is maximal. As equation (49) which applies to a Colpitts oscillator (Fig. 14) indicates, phase noise is mainly determined by the amplitude of the oscillation V_{osc}, the quality factor of the LC tank (which determines the oscillation frequency), thermal noise of the active and passive components in the VCO, and—close to the carrier frequency —by the *1/f* noise of the active components of the VCO. In (49) circuit noise is lumped into an equivalent noise current source, I_n.

$$L[\Delta f]=\frac{|I_n|^2}{V_{osc}^2}\times\frac{1}{\Delta f^2}\times\frac{1}{C_1^{\,2}\left(\frac{C_1}{C_2}+1\right)^2}\qquad L(f_m)=\frac{|I_n|^2\omega_{osc}^2}{I_{BIAS}^2 4Q^2}\frac{C_2^{\,2}}{C_1^{\,2}}\times\frac{1}{\left(\frac{C_1}{C_2}+1\right)^2} \tag{49}$$

As in PAs, V_{osc} is limited by the transistor V_{MAX}. The larger breakdown voltage of SiGe HBTs gives them a significant advantage. Since voltage swings of 3 V_{pp} are safe in CB stages while only 1 V_{pp} can be reliably accommodated in 90-nm MOSFETs, the phase noise and output power levels reported for SiGe-HBT VCOs [12] are typically 10 dB better than those of 60-GHz VCOs fabricated in advanced 90-nm SOI CMOS [22].

In an effort to compare the ultimate performance of different topologies and transistor technologies, expressions for the maximum oscillation frequency of Colpitts and cross-coupled VCOs can be derived as functions of transistor technology parameters. For a given technology, Colpitts VCOs exhibit higher oscillation frequencies than cross-coupled ones.

$$\omega_{osc,\,Colpitts}\le\frac{g'_m Q_{eff}}{C'_{gs}+C'_{sb}}\left[1+\frac{C'_{gd}}{C'_{gs}}+\frac{C'_{gd}}{C'_{sb}}\right] \tag{50}$$

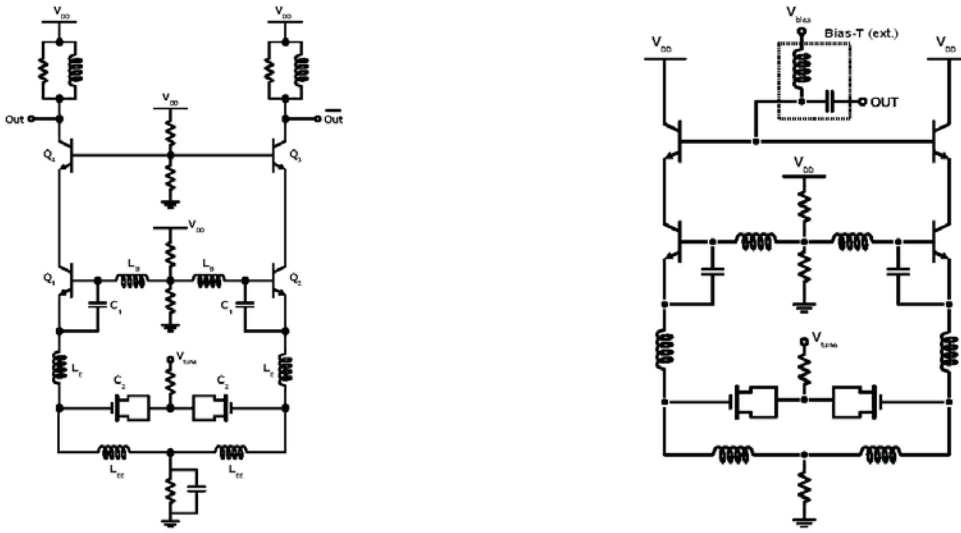

Figure 14. Schematics of fundamental 60-GHz and second harmonic 77-GHz SiGe HBT Colpitts VCOs [12].

$$\omega_{osc,n-MOS} \le \frac{g'_m Q_{eff}}{C'_{gs} + 4C'_{gd} + C'_{db} + \frac{C_L}{W}} \tag{51}$$

The size and bias current of MOSFET Colpitts and cross-coupled VCOs decreases with increasing frequency

$$W_{n-MOS,cross} \le \frac{1}{\omega_{osc}^2 L[C'_{gs} + 4C'_{gd} + C'_{db}]} \tag{52}$$

$$W_{n-MOS,Colpitts} \le \frac{1}{\omega_{osc}^2 L\left[\frac{C'_{gs} C'_{sb}}{C'_{gs} + C'_{sb}} + kC'_{gd}\right]} \tag{53}$$

Since in bipolar transistors the equivalent of C'_{sb} does not exist (i.e. $C_{ES} = 0$), even at the extreme highest frequency, achieving moderate tuning range is possible with varactor diodes:

$$A_{E,HBT,Colpitts} \le \frac{1}{\omega_{osc}^2 L\left[\frac{C_{jbA}\frac{C_{var}}{A_E}}{C'_{jbA} + \frac{C_{var}}{A_E}} + kC_{jcA}\right]} \tag{54}$$

where k accounts for the Miller effect and is typically between 1 and 2. C_{var} is the varactor diode capacitance.

At the onset of oscillation, Colpitts and cross-coupled VCOs can be described as small signal amplifiers with positive feedback. The tank impedance acts as the signal source impedance for the amplifier. The noise generated within the VCO can be minimized if the transistors are biased at the optimum noise current density (J_{opt}) and if their noise impedance is matched to that of the tank at resonance. Clearly, once the tank Q and tank voltage V_{osc} are maximized, the design for low phase noise should be conducted in much the same way as for LNAs [12]. It should be noted that Colpitts VCOs have one extra degree of freedom in the C_1/C_2 ratio which can be optimized to ensure noise matching to the tank impedance. In cross-coupled topologies noise matching is usually compromised because the transistor size and bias current are set by the voltage swing requirement. As a result, in a given transistor technology, optimally noise-matched Colpitts VCOs will exhibit lower phase noise than cross-coupled ones.

Fig. 14 shows the schematics of 60-GHz fundamental frequency and 77-GHz push-push SiGe-HBT VCOs implemented in a 155 GHz SiGe BiCMOS process [12]. A record phase noise of -104 dBc/Hz at 1 MHz offset was measured in the 60 GHz VCO. The measured tuning characteristics and chip photograph of the push-push VCO are shown in Fig. 15. The 21% tuning range is accomplished using inductors and accumulation-mode varactor diodes. Wide tuning range is necessary to cover the 5-GHz bandwidth of a 60-GHz radio, as well as process and temperature variations. Finally, Figs. 16 and 17 reproduce the schematics and spectrum of a directly modulated BPSK transmitter operating at 65 GHz. It integrates a mixing quad as the common-base part of a cascode Colpitts VCO and is implemented in the same process. The data signal is applied differentially at the input of the mixing quad while the bottom HBT differential pair acts as the fixed frequency oscillator.

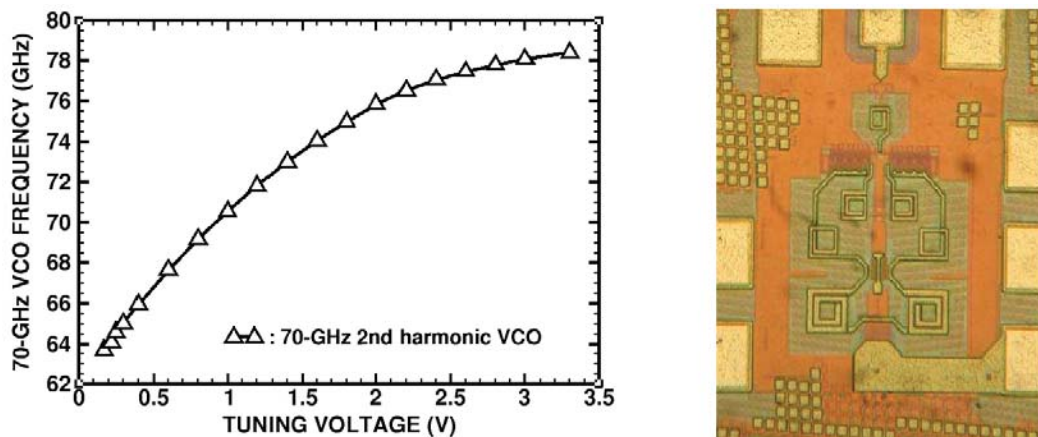

Figure 15. Tuning characteristics and layout of push-push SiGe-HBT VCO.

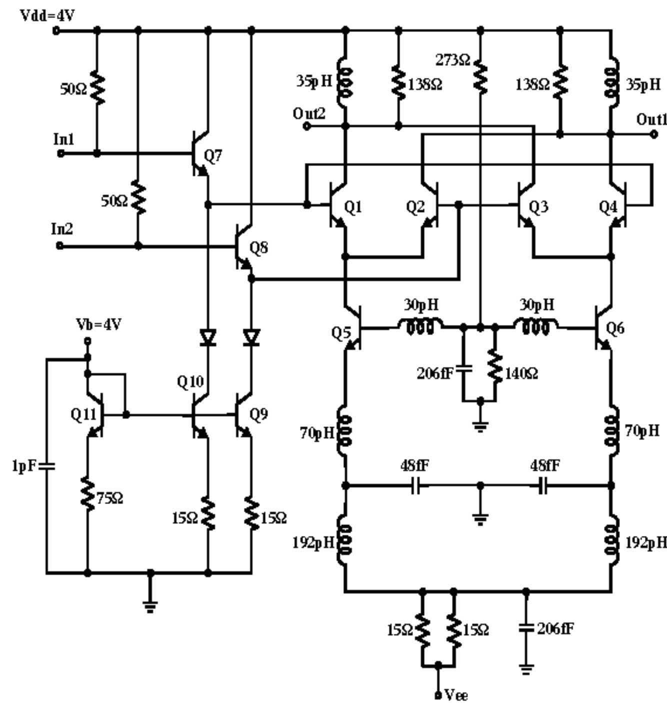

Figure 16. 65-GHz BPSK transmitter schematics.

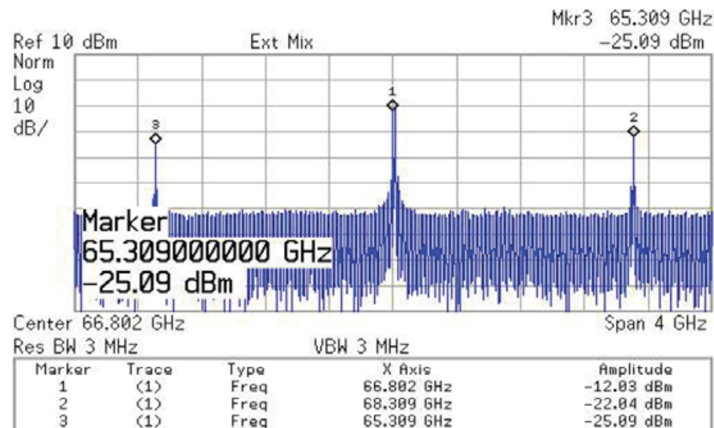

Figure 17. Schematics and measured spectrum at the output of 65-GHz BPSK transmitter for a 1.5 GHz data input.

Conclusions

It was demonstrated that, by optimally sizing and biasing transistors, simplifying circuit topologies, and taking advantage of the reduced dimensions of inductors and transformers at mm-wave frequencies, it is possible to repeatably design and fabricate high performance mm-wave SiGe BiCMOS and 90-nm CMOS ICs with smaller die area and lower cost than corresponding wireless ICs in the 2-10 GHz range. The analysis and

experimental results presented in this paper indicate that CMOS technology scaling into the nano-scale domain is mostly beneficial for RF and mm-wave applications. In situations where the maximum allowable voltage swing limits circuit performance, as in power amplifiers and low-noise oscillators, SiGe HBTs will continue to exhibit superior performance, emphasizing the need for future generation of SiGe BiCMOS technologies beyond the state-of-the-art 130-nm node.

Acknowledgments

We thank STMicroelectronics, Jazz Semiconductor, and TSMC for fabricating the circuits and, Rudy Beerkens, Bernard Sautreuil, Paul Kempf, Marco Racanelli, and M.T. Yang for their support. This work was funded by Micronet, and NSERC. We acknowledge CAD tools from CMC and equipment grants from OIT and CFI.

References

1. K. Kuhn, R. Basco, D. Becher, M. Hattendorf, P. Packan, I. Post, P. Vandervoorn, and I.Young, *Symposium on VLSI Technology* Digest of Technical Papers, pp. 224-225, June, 2004.
2. N Zamdmer, J. Kim, R. Trzcinski, J.-O. Plouchard, S. Narasimha, M. Khare, L.Wagner, and S. Chaloux, *Symposium on VLSI Technology* Digest of Technical Papers, pp. 98-99, June, 2004.
3. ITRS 2004, www.itrs.net
4. R. B. Staszewski, K. Muhammad, and D. Leipold, *5th IEEE IWSOC* Digest, Banff, Alberta, Canada, pp. 217-222, July, 2005
5. Y. Taur and T. H. Ning, "*Modern VLSI Devices*," Chapter 4, Cambridge University Press, 1998.
6. S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C-H. Jan, C. Kenyon, J. Klaus, K. Kuhn, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetti, P. Nguyen, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr, and Y. El-Mansy, *IEEE Trans. Electron Dev.*, Vol. ED-51, pp.1790-1797, 2004.
7. S. E. Thompson, R. S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M. Bohr, *IEEE Trans. Semicond. Man.*, Vol. ED-51, pp.1790-1797, 2004.
8. S.P. Voinigescu, A. M. Mangan, M. Tazlauanu, and M.T. Yang, submitted to *IEEE JSSC* in April 2005.
9. S.P. Voinigescu, T.O. Dickson, and R. Beerkens, I. Khalid, and P. Westergaard, *Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, Digest of Technical Papers, pp.111-114, Sept., 2004.
10. G. Dambrine, H. Happy, F. Danneville, and A. Cappy, *IEEE Trans. Electron Dev.*, pp.375-381, March. 1993.
11. T. O. Dickson, M.-A. LaCroix, S. Boret, D. Gloria, R. Beerkens, and S. P. Voinigescu, *IEEE Trans. MTT*, Vol.53, No.1, pp.123-133, 2005.

12. C. Lee, T. Yao, A. Mangan, K. Yau, M. A. Copeland, and S. P. Voinigescu, *IEEE CSICS Technical Digest*, pp.179-182, Oct. 2004.
13. P. H. Woerlee, M. J. Knitel, R. van Langevelde, B. M. Klaassen, L. F. Tiemeijer, A. J. Scholten, and A. T. A. Zegers-van Duijnhoven, *IEEE Trans. Electron Dev*. Vol. 48, pp.1776-1782, 2001.
14. L. Larson, *IEEE Trans. Electron Dev*. Vol. 50, pp.683-699, 2003.
15. T. O. Dickson, R. Beerkens, and S. P. Voinigescu, *IEEE Journal of Solid-State Circuits*, Vol.40, No.4, pp.994-1003, 2005.
16. S. K. Reynolds, U.R. Pfeiffer, T. Zwick, T. Beukema, and B. Gaucher, *ISSCC Digest*, pp.442-443, 2004.
17. C.H. Doan, S. Emami, A. M. Niknejad, and R.W. Brodersen, *IEEE Journal of Solid-State Circuits*, pp.144-155, Jan. 2005.
18. M. Gordon and S.P. Voinigescu, *IEEE ESSCIRC*, Leuven, Belgium, Sept. 2004.
19. B Razavi, *ISSCC-Digest*, pp.400-4001, 2005.
20. G.A.M. Hurkx, P. Agarwal, R. Dekker, E. van der Heijden, and H. Veenstra, *IEEE Trans. Electron Dev*., pp.2121-2128, Dec. 2004.
21. H. M. J. Boots, G. Doornbos, and A. Heringa, *IEEE Trans. Electron Dev*., pp.2102-2108, Dec. 2004.
22. F. Ellinger, et al., *IEEE MTT-S Digest*, pp.1329-1332, Jun. 2004.