Silicon Technology Tradeoffs for Radio-Frequency/Mixed-Signal "Systems-on-a-Chip"

Lawrence E. Larson, *Fellow, IEEE*

Invited Paper

***Abstract*—Silicon technology has progressed over the last several years from a digitally oriented technology to one well suited for microwave and RF applications at a high level of integration. Technology scaling, both at the transistor and back-end metallization level, has driven this progress. CMOS technology is ideally suited for low-noise amplification and receiver applications, but the fundamental breakdown voltage is lower than that of equivalent Si/SiGe HBTs. High-quality passive devices are equally important, and improvements in metallization technology are resulting in higher quality inductors. This paper summarizes the silicon technology issues associated with RF "system-on-a-chip" applications.**

***Index Terms*—Capacitor, HBT, inductor, linearity, low-noise amplifier (LNA), mixer, MOSFET, noise figure, power amplifier, RF CMOS, SiGe, third-order input-referred intercept point (IIP3), transceiver, voltage-controlled oscillator (VCO).**

I. Introduction

THE DESIRE to communicate quickly and reliably with our family, friends, and colleagues is one of the most widespread of human needs, and wireless telephony has exploded in the last decade as a ubiquitous tool to fulfill that desire. Over 400 million cellular handsets were sold each in the years 2000 and 2001, and the market is expected to grow to nearly a billion phones per year within the next decade. The cellular telephone has become so common and widespread, such an integral part of modern existence, that it is easy to forget that it was considered an expensive novelty just 15 years ago. At the same time, wireless local-area networks (WLANs) are an emerging technology promising untethered communication within computer networks.

This revolution in communications has resulted from the confluence of a variety of technological factors: advances in communications theory, networking architectures, semiconductor technology, and transceiver design. The combination of stunning advances in semiconductor technology, e.g., Moore's law, combined with improved approaches to handset transceiver design, have enabled the size, cost, and battery life of the wireless handset to be shrunk to that of a typical consumer item, within reach of half of the population on the planet.

This paper will outline the key developments and trends in silicon semiconductor technology applied to RF/mixed-signal "system-on-a-chip" implementations of these wireless communications devices. The communications medium that a mobile wireless transceiver typically finds itself in is often referred to as "hostile," since the path—or channel—from the transmitter to the receiver is subject to time-varying obstructions and multipath fading as well as Doppler effects. This is in contrast to "point-to-point" communications links—either wireless, fiber-optic, or free-space optical—where the channel is essentially nontime-varying or "stationary." This hostile channel affects both the design of the transmitter and receiver in profound ways, placing extreme performance constraints on the technology required for their implementation.

One way to view these constraints is through the "football field" metaphor [1]. As an example, the popular European GSM system [2] has a minimum received signal sensitivity level (the smallest level of the desired received signal) at the antenna of -102 dBm ($10^{-13.2}$ W), but the largest (undesired) interferer also received by the antenna has a level of 0 dBm (10^{-3} W). If the desired signal power is normalized to the size of the head of a pin—approximately one mm in diameter—then the largest interferer is roughly the size of two (U.S.) football fields, 100 m by 100 m. Following this metaphor further, receiving and demodulating a GSM signal is analogous to finding the head of a pin in a football field! In addition, this task has to be done in less than 100 ms, which is typically the time it takes for the cellular handset to receive a call. Viewed through this lens, the modern cellular handset is truly a technological marvel.

The architecture of a general handheld wireless transceiver (transmitter *and* receiver) is shown in Fig. 1. In the receiver design, the radio signal is sent from the receiving antenna to a low-noise amplifier (LNA), whose purpose is to boost the signal level without reducing the signal-to-noise ratio (SNR) significantly. The signal level at the antenna can range between 1 μV rms to nearly 100 mV rms—over a 100-dB variation! At the low end of the signal range, the LNA performance is fundamentally limited by thermodynamic and electron transport issues, while at the high end of the signal range, the challenge is to minimize the effects of nonlinearities on receiver

Manuscript received July 18, 2002; revised November 11, 2002. The review of this paper was arranged by Editor H. Shichijo.

The author is with the Center for Wireless Communications, Electrical and Computer Engineering Department, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: larson@ece.ucsd.edu).

Digital Object Identifier 10.1109/TED.2003.810482

0018-9383/03$17.00 © 2003 IEEE

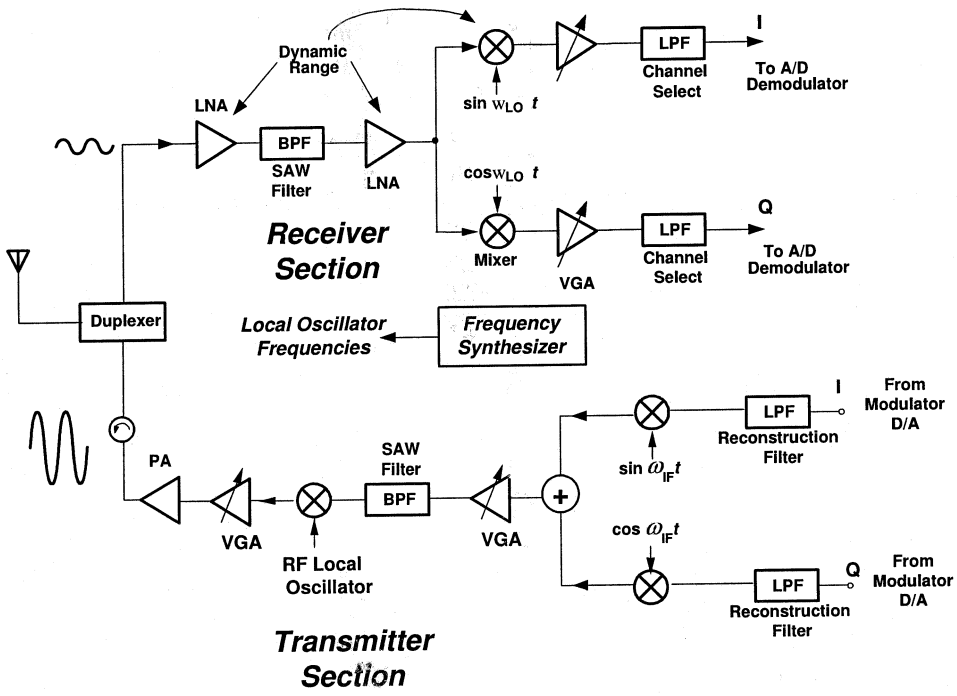


Fig. 1. Architecture of a typical wireless handset transceiver. The receiver employs a direct (single-step) downconversion approach, while the transmitter employs a "two-step" heterodyne technique.

performance. These diverse requirements are often referred to as the "LNA Bottleneck" [3]. As a result, the high-frequency LNA must exhibit excellent performance over both small-signal and large-signal conditions. Modern CMOS FETs and SiGe HBTs are particularly well suited to this application in the 1–5-GHz range, because of their outstanding f_T and low gate/base resistance.

In addition, the LNA is typically "on" all the time—listening for transmitted signals of interest—so it is constantly draining power, and therefore it must dissipate as little dc power as possible. The combination of extremely high-performance and low-power requirements result in the LNA being one of the most significant power drains in the system.

Following the LNA, the signal is typically passed through a mixer, which essentially multiplies the input signal by a local oscillator signal of constant frequency, producing an output signal whose frequency is the difference between the two inputs—the so-called intermediate frequency (IF)—and whose amplitude is proportional to the original input signal. In the case of Fig. 1, the receiver architecture employs the direct-conversion (or *homodyne*) approach, and the IF is at dc. Preceding the mixer, an analog filter eliminates the response to an undesired input signal that would "jam" the receiver and compress its gain. This filter is typically implemented with a physically large off-chip surface acoustic wave (SAW) device. In addition to their excessive size, these filters have extremely unforgiving sensitivities to variations in source impedance and ground loops, to name a few.

Substantial progress has been made recently in the area of direct-conversion approaches for wireless receivers, which are well suited to monolithic integration. The advantage of this particular architecture, compared to the more traditional two-step *heterodyne* approach, is that it is uniquely well suited to monolithic integration, due to its low-frequency filtering, lack of serious image responses, and its intrinsically simple architecture [4]. The output signals from the I and Q paths are converted into digital values with an A/D converter and then sent to the digital baseband section for synchronization and data recovery.

On the transmitter side of Fig. 1, the goal is to modulate the I and Q data from the digital baseband section onto the high-frequency carrier with near perfect fidelity and vary the output power transmitted over nearly 80 dB of dynamic range; a high output power is transmitted when the handset is far from the base station, and low power when the handset is close to the base station. This is known as the "power-control loop" in a CDMA handset and is necessary in order to solve the "near–far" problem in multi-user CDMA systems [5]. The maximum output power level at the antenna is approximately 1 W at a frequency of roughly 2 GHz.

In this case, the filtered digital data is modulated onto an IF carrier (typically several hundred megahertz), partial gain control is applied at that frequency, and then a second stage of upconversion is applied to the signal to transfer it to the final frequency, where the remaining gain control occurs. This "two-step" approach allows for the gain control to be split over several different frequencies, improving isolation between the bands. In this case, the key requirements are the maintenance of the fidelity, or linearity, of the signal at the final output stage and dc efficiency to maximize battery life.

The reference frequencies required for upconversion and downconversion of the transmitted and received signals are generated by a *frequency synthesizer*, which uses a precise reference (usually produced by crystal oscillator) to synthesize the necessary local oscillator frequencies. In this case, the phase noise of the synthesized signal must be as low as possible to

TABLE I
SUMMARY OF SOME REPRESENTATIVE WIRELESS SYSTEMS FOR VOICE AND DATA APPLICATIONS

Standard	System	Frequency (MHz)	Data Rate (Mb/sec)	Peak Transmit Power (dBm)	Minimum Sensitivity (dBm)
GSM	Cellular	900	0.014	33	-102
CDMA	Cellular	900/1900	0.014	23	-104
WCDMA	Cellular	2000	2	27	-117
802.11b	Data	2400	11	20	-80
802.11a	Data	5200	54	23	-82
Bluetooth	Data	2400	0.723	0	-70
Zigbee	Data	900/2400	0.02/0.25	—	—

accurately modulate and demodulate the signal. Furthermore, the synthesizer itself is a complex RF/analog/digital circuit, which generates copious amounts of digital switching noise and harmonics. Historically, the synthesizer circuit was contained on a separate integrated circuit but, with system-on-chip (SOC) implementations, this noise must be isolated from the sensitive receiver circuits despite the fact that they share a common substrate and package environment. This presents a fundamental challenge to the integration level of these complex circuits.

The digital portion of the communication system performs the key functions of modulation and demodulation (the so-called "modem"), carrier recovery, timing recovery, symbol recovery, equalization, channel coding, power detection, and calibration, among others [5]. Separate digital controllers also perform media access control (MAC) functions as well as a variety of other control functions. The eventual goal is to include all these digital functions on the same integrated circuit substrate as the RF and analog circuits in order to realize a true "single-chip" communications system implementation. This goal has been realized on a number of ultralow-cost digital systems—such as those conforming to the well-known Bluetooth standard [6]—but remains elusive for higher performance systems involving cellular telephones or wireless LANs.

The range of wireless communication systems employing RF techniques that are amenable to SOC implementation has grown dramatically and continues to expand with new standards and applications being developed all the time. Consumer demand for the applications available using untethered wireless devices continues to grow, and Table I summarizes some of the data rates and bandwidths being developed on a worldwide basis. The key to the widespread adoption of these systems is developing low-cost highly integrated implementations of the key radio and digital functions.

The goal of this paper is to highlight the key technological tradeoffs in silicon integrated circuit technologies for these RF SOC applications. As a result, it is meant to be a "survey" of existing results in a variety of disciplines, which, when presented together, present a more complete picture of the technological challenges that face the RF-SOC community. Based on the improvements in device performance achieved in recent years, it is clear that, CMOS or BiCMOS technology, where high-quality active and passive devices are integrated on a common substrate along with a high level of digital integration, is the preferred medium for RF-SOC implementation.

The paper begins in Section II at the level of substrate improvements and interdevice isolation developments that are required for RF-SOC applications. Section III discusses transistor level performance and explores how transistor scaling enhances the performance of key building blocks for wireless communications systems. This is related to the f_T and f_{MAX} of the transistor, as well as the achievable breakdown voltage. The characteristics of other important factors, such as passive device performance are also explored. Next, the performance of several key RF SOC *circuits* is analyzed in terms of the active and passive device performance at the lower level. LNA performance is discussed in Section V, and voltage-controlled oscillator (VCO) performance is discussed in Section VI. Finally, conclusions are presented concerning the challenges to RF/SOC implementation as silicon technology is scaled in the future.

II. SUBSTRATE AND ISOLATION TECHNOLOGIES FOR RF-SOC APPLICATIONS

The substrate plays an intimate role in determining the performance of an RF-SOC. This is because the desired signal levels are so small, and the frequencies are so high, that undesired spurious signals can leak into the sensitive receiver portions through almost any path, particularly capacitive coupling through the conductive substrate. Returning to the example of the CDMA transceiver of Fig. 1, the received signal strength can be a as low as -104 dBm, but the transmitted signal strength can be as high as $+23$ dBm—a nearly 130-dB difference! Clearly, the isolation between the receiver and transmitter on an SOC is a significant challenge. Similar isolation considerations apply for the required isolation between the frequency synthesizer and the receiver, where digital switching noise can couple into the receiver through the substrate.

In addition, the conductive silicon substrate increases the eddy losses in monolithic inductors and increases the losses associated with high-frequency transmission-line structures. Fortunately, these problems are well known, and many enhancements have been suggested to improve substrate loss and interdevice isolation for RF-SOC applications. These improvements can be grouped into the categories of substrate resistivity

enhancements, implant blocking layers, and layout-dependent improvements.

Digitally oriented bulk CMOS processes typically rely on a low-resistivity substrate in order to minimize latch-up considerations. The resulting conductive losses in monolithic inductors and other high-frequency circuits are usually considered to be excessive for RF applications, and so lightly-doped p-type substrates are more typically used in both CMOS and BiCMOS approaches for RF-SOC applications. In this case, the typical bulk resistivity is roughly 10 Ω·cm, corresponding to a doping density of approximately $5 \times 10^{15}\text{cm}^{-3}$.

Interdevice isolation can be improved through a variety of approaches at the substrate level. Simply increasing the resistivity of the silicon substrate is the most conservative approach, and can provide for a significant improvement in isolation. The resistivity of production Czochralski (CZ) wafers is currently limited to a maximum of roughly 10–20 Ω·cm, although a new technique known as Magnetic Czochralski (MCZ) has demonstrated resistivities up to 1 kΩ·cm [7]. In addition, Float-Zone (FZ) silicon wafers can achieve resistivities of up to 10 kΩ·cm, although the cost of FZ material is currently substantially higher than that of CZ wafers [8]. These highly resistive substrates have historically exhibited manufacturing problems under the high-temperature stress of subsequent wafer processing, but great progress has been made recently in this area [9].

Other, more exotic bulk approaches to improving substrate resistivity or isolation have also been proposed for RF-SOC applications. These include the use of silicon-on-insulator (SOI) [10], silicon-on-sapphire (SOS) [11], silicon-on-anything (SOA) [12], porous silicon [13], through substrate vias [14], and bulk micromachining [15]. None of these techniques have found their way into widespread use, although there have been some notable successes in niche applications. Clearly, the manufacturability and cost-effectiveness of these more exotic technologies in a high-volume consumer-oriented marketplace must be carefully considered.

A more traditional approach to the problem of improving device isolation is the use of grounded "guard rings" that surround the sensitive active devices. This approach is shown in Fig. 2(a). The effectiveness of this technique depends on the width of the guard ring, substrate resistivity, and the inductance between the guard ring and ground. In general, the isolation improves with increasing spacing, guard ring width, and substrate resistivity. An extreme example of the use of guard rings to improve isolation is the Bluetooth chip presented by ISSCC2002 [16]. In this case, a 300-μm guard ring completely surrounded the sensitive RF portions of the circuit, separating it from the digital portions. This enabled the chip to exceed the Bluetooth 2.4-GHz receive sensitivity level of -70 dBm; in fact the designed achieved a sensitivity of -82 dBm on a single die containing all of the RF, analog, and digital functions.

With a fixed substrate resistivity, a further improvement in isolation can be achieved through the use of a deep low-resistivity n-well placed underneath the active device; when biased to a low-impedance and low-noise potential, it acts as an effective shield to signals injected from nearby sources. This approach is shown in the cross section in Fig. 2(b). The addition of a deep well can improve isolation between adjacent devices

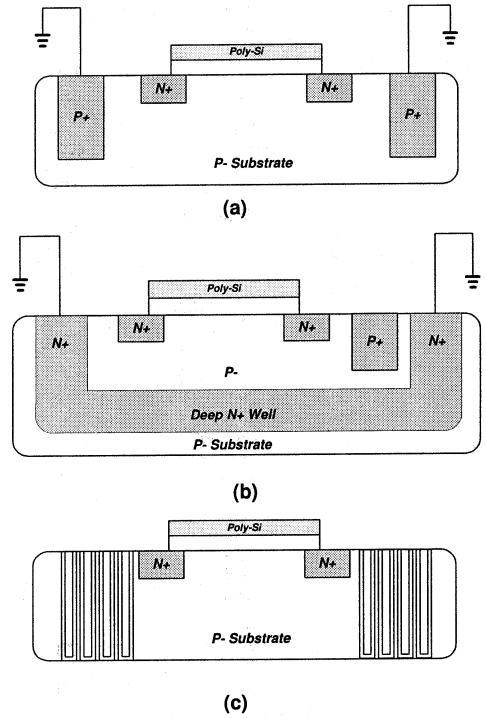


Fig. 2. Different isolation techniques used for RF-SOC applications in bulk technologies: (a) P+ substrate guard-ring isolation, (b) buried n-layer isolation, and (c) deep trench isolation.

by roughly 20 dB (from 40 to 60 dB) at 2 GHz [17]. The effectiveness of this technique at high frequencies depends on the common inductance of the signal line and the grounding structure of the n-well, and an inductance of as little as 0.5 nH can significantly degrade the improvement at frequencies above 1 GHz [18]. This "triple-well" technology is now a standard option of many sub-0.18-μm CMOS processes, using both low- and high-resistivity substrates.

Deep trench isolation techniques are a standard feature of many advanced BiCMOS technologies, and the use of deep trenches has proved effective in improving inductor quality factor by reducing eddy losses. It can also be employed to improve isolation between devices, as shown in Fig. 2(c), although the improvement in device isolation is modest compared to the other two approaches.

Clearly, a combination of lightly doped substrates, deep n-wells, and generous guard rings can provide for improved isolation in an RF-SOC environment. The challenge then becomes integrating these features into a *design environment* that can predict the isolation prior to fabrication. This is highly challenging, as it requires integrating an accurate physically based equivalent-circuit model of the substrate and the isolation structures with the simulation of the rest of the circuit. Several tools have recently been introduced to accomplish this, although more work remains to be done in this area [19].

III. Transistor Scaling for RF-SOC Applications

The key active device parameters for enhanced circuit performance of noise and linearity for most RF-SOC applications are the *short-circuit unity current gain frequency* (f_T) and the

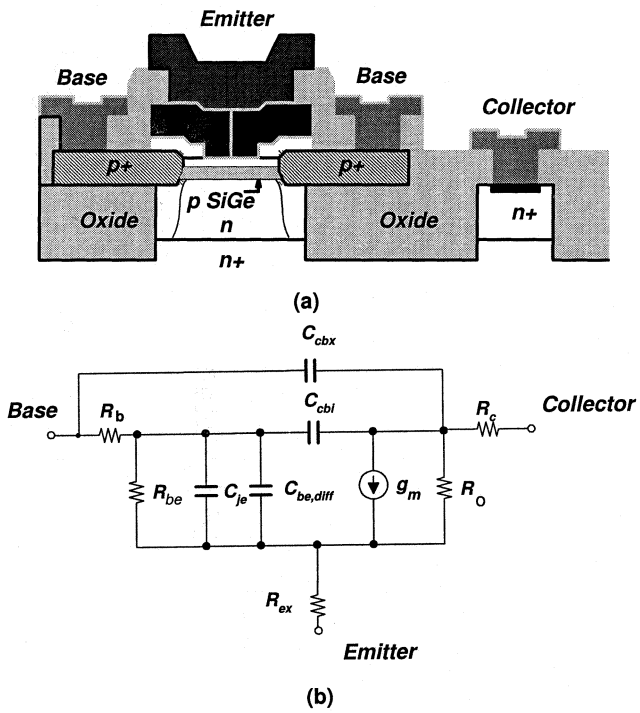


Fig. 3. Si/SiGe HBT. (a) Cross section of the device. (b) Equivalent circuit model of the transistor.

maximum unity power gain frequency (f_{MAX}). These two parameters have made astonishing progress in recent years in both HBTs and MOSFETs, with recently reported values for both devices in excess of 200 GHz [20], [21]. The next most important issue is breakdown voltage, which together with noise considerations sets the dynamic range limitation of most circuits.

If we examine the Si/SiGe HBT first, using the physical cross section and equivalent circuit model of the device shown in Fig. 3, the f_T is given by

$$\frac{1}{2\pi f_T} = \tau_B + \tau_C + \frac{kT}{q}\left(C_{\mathrm{je}} + C_{\mathrm{cb}}\right) + \left(R_{\mathrm{ex}} + R_c\right) C_{\mathrm{cb}} \quad (1)$$

where R_{ex} and R_c are the parasitic emitter and collector resistances, C_{cb} is the collector-base junction capacitance, C_{je} is the emitter-base junction capacitance, τ_B is the base transit time, and τ_C is the collector transit time.

In most high-frequency applications, the base and collector transit times dominate the f_T, and the other parasitic-related terms have a secondary effect. For this same physical structure and equivalent circuit model, the f_{MAX} of the transistor is given by [22]

$$f_{\mathrm{MAX}} \approx \sqrt{\frac{f_T}{8\pi\tau_{\mathrm{cb}}}} \quad (2)$$

where τ_{cb} is *approximately* $R_b C_{\mathrm{cb}}$, but can be more accurately described as a weighted average of the distributed base resistance and base-collector capacitance [23]. Equation (2) is slightly pessimistic in cases of large collector junction width, but represents a good starting point for discussions of device scaling issues.

The base transit time τ_B is given by [24]

$$\tau_B = \frac{T_b^2}{D_n}\left(\frac{kT}{\Delta E}\right) - \frac{T_b^2}{D_n}\left(\frac{kT}{\Delta E}\right)^2\left(1 - e^{-\Delta E/kT}\right) + \frac{T_b^2}{v_{\mathrm{exit}}}\left(\frac{kT}{\Delta E}\right)\left(1 - e^{-\Delta E/kT}\right) \quad (3)$$

where T_b is the base thickness, v_{exit} is the base exit velocity (roughly kT/m^*, where m^* is the electron effective mass), D_n is the base minority carrier diffusivity, and ΔE is the grading in the base bandgap energy.

The collector transit time is the average delay of the electron transit through the collector depletion region and is given by [25]

$$\tau_c \approx \frac{T_c}{2v_{\mathrm{eff}}} \quad (4)$$

where T_c is the collector thickness, and v_{eff} is the effective saturated electron velocity ($v_{\mathrm{eff}} \approx 0.5 \times 10^7$ cm/s).

These expressions highlight the critical role of vertical scaling to improve the f_T and f_{MAX} for bipolar device performance. At the same time, *lateral* scaling of the devices is equally critical, to further reduce extrinsic base resistance and collector-base capacitance.

The base resistance R_b, which has a large impact on f_{MAX} is a result of the sum of several components, including the spreading resistance underneath the emitter R_{spread}, the base-emitter gap resistance R_{gap} and the contact resistance R_c. Given a contact resistivity ρ_c and a base sheet resistance ρ_b, the resulting base resistance is [23]

$$R_b = R_{\mathrm{spread}} + R_{\mathrm{gap}} + R_c \quad (5a)$$

$$R_{\mathrm{spread}} = \rho_b W_e / 12 L_e \quad (5b)$$

$$R_{\mathrm{gap}} = \rho_b W_{\mathrm{gap}} / 2 L_e \quad (5c)$$

$$R_C = \sqrt{\rho_c \rho_b} / 2 L_e \quad (5d)$$

where W_e is the emitter width, L_e is the emitter length, and W_{gap} is the gap width between the emitter and base.

It is clear that increasing the f_T of the device through reducing the base thickness will improve the f_{MAX}, but the base resistance can rise from the increase in ρ_b, minimizing the overall improvement. Most scaling efforts with HBT structures aim to keep the f_{MAX} equal to or slightly larger than the f_T.

The dependence of transistor f_T and f_{MAX} on base width can be seen clearly from the plots of measured devices in Fig. 4, where the clear dependence of transit time on base width has a significant effect on f_T [26]. The effect of base width on f_{MAX} is less pronounced, due to the additional necessity to keep base resistance equally low.

Although Si/SiGe HBTs have historically been leading MOS devices in terms of peak reported f_T, super-scaled MOS devices

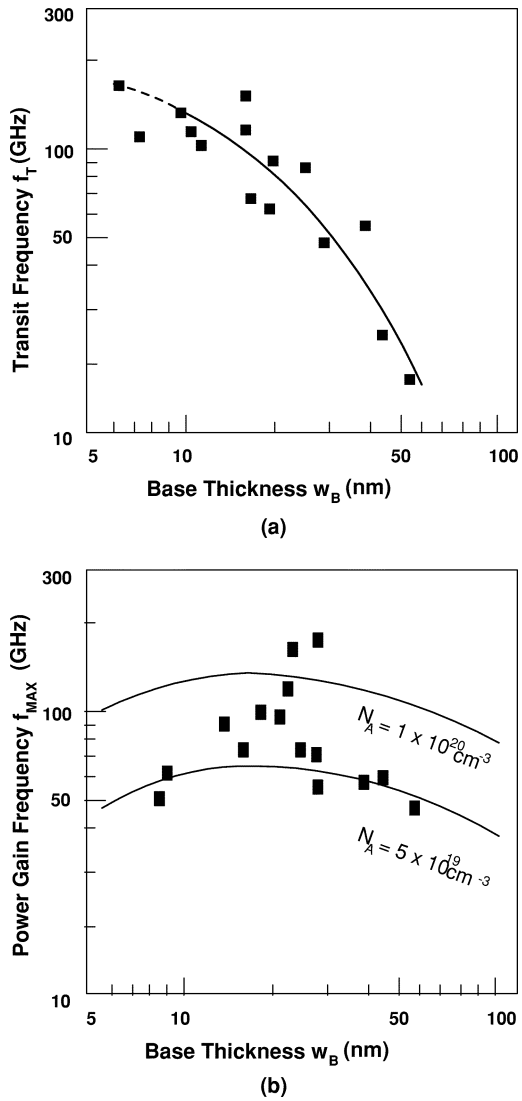


Fig. 4. Si/SiGe HBT speed as a function of base width [26]. (a) The f_T demonstrates a clear base width dependence, and (b) the device f_{MAX} is also affected by the base resistance but the improvement with decreasing base thickness is much less pronounced.

have recently demonstrated outstanding results as well, and we are now at a point where laboratory results of f_T are nearly comparable. Similar scaling expressions can be derived for the f_T and f_{MAX} of the MOSFET of Fig. 5, where

$$\frac{1}{2\pi f_T} = \frac{C_{\mathrm{gs}}}{g_m}\left(1 + g_{\mathrm{ds}}(R_d + R_s)\right) + (R_s + R_d)\, C_{\mathrm{gd}} \tag{6}$$

where g_m is the transistor transconductance, C_{gs} is the gate–source capacitance, C_{gd} is the gate–drain capacitance, R_d is the parasitic drain resistance, R_s is the parasitic source resistance, and g_{ds} is the drain–source conductance.

For this physical structure and equivalent circuit model, the f_{MAX} is given by

$$f_{\mathrm{MAX}} \approx \sqrt{\frac{f_T}{8\pi\tau_{\mathrm{gd}}}} \tag{7}$$

where τ_{gd} is *approximately* $R_g C_{\mathrm{gd}}$ but is more accurately described as a weighted average of the distributed gate resistance

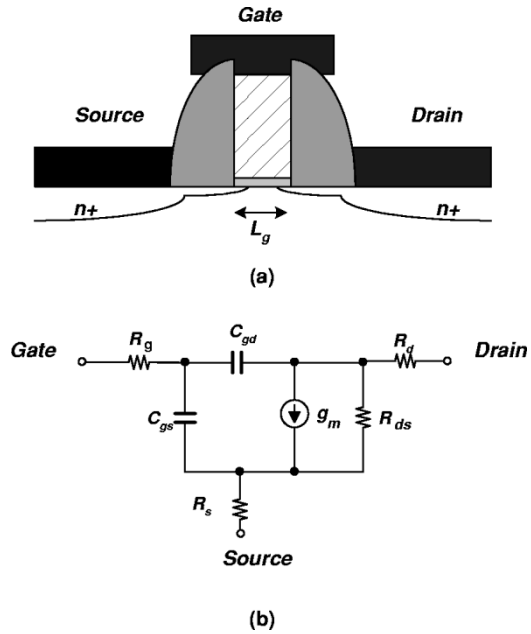


Fig. 5. Scaled NMOS FET. (a) Cross section of the device. (b) Equivalent circuit model of the transistor.

and gate–drain capacitance, in a manner analogous to that of the HBT. From a physical perspective, modern MOSFETs operate in a heavily velocity saturated regime, where transit time effects dominate the f_T characteristics.

As MOSFETs scale to smaller and smaller dimensions, the gate resistance effect on f_{MAX} and noise can become increasingly problematic. The dc gate resistance (per finger) is given by

$$R_{g,\,\mathrm{dc}} = \frac{\rho W}{L \cdot h} \tag{8}$$

where ρ is the gate resistivity, h is the gate thickness, W is the gate finger width, and L is the gate length. Due to capacitive shunting effects, the per-finger effective series resistance at high frequencies is given by [27]

$$R_{g,\,rf} \approx \frac{R_{g,\mathrm{dc}}}{3}. \tag{9}$$

Since h scales along with L as the devices are reduced to smaller and smaller dimensions (in order to maintain a roughly constant aspect ratio), the dc gate resistance can rise nearly as fast as the square of the scaling factor [28]. This effect has been addressed through a variety of proposed improvements, including the use of "T-gate" structures [29] and parallel gate strapping approaches [30].

As Fig. 6 demonstrates, the f_T of modern scaled MOS devices approaches a value of [31]

$$\frac{1}{2\pi f_T} \approx \frac{L_g}{\upsilon_{\mathrm{eff}}}. \tag{10}$$

Fig. 6 also demonstrates that the *ratio* of f_{MAX}/f_T for the MOSFET has been falling as the speed of the devices rises, in response to the increased gate resistance of the ultrashort gate

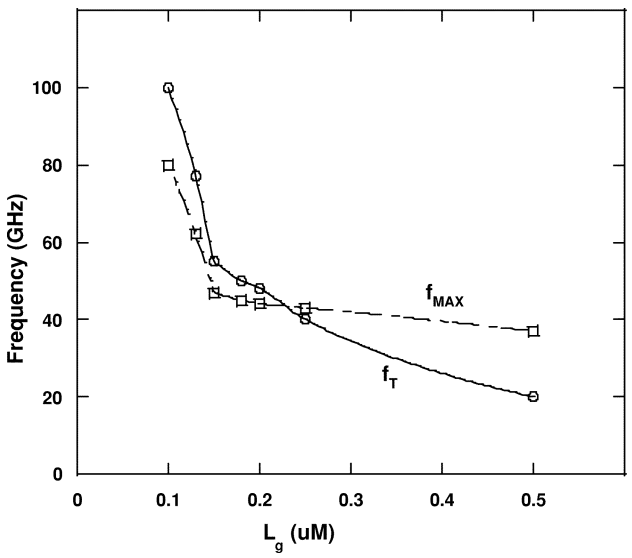


Fig. 6. MOSFET speed as a function of gate length [31]. The f_T and f_{MAX} demonstrate a clear gate length dependence. Note that the ratio of f_{MAX}/f_T decreases with decreasing gate length, demonstrating the increasing impact of parasitic gate resistance.

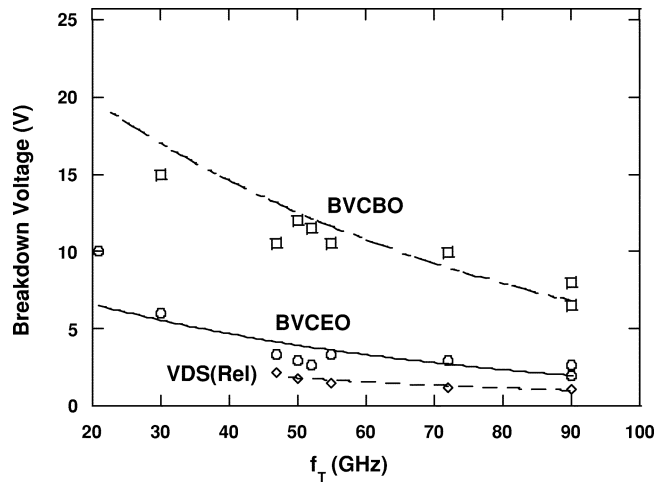


Fig. 7. Comparison of voltage limitations of MOSFETs and HBTs as a function of f_T [31], [35]. The VDS(Rel) of the MOSFET is the recommended operating voltage to minimize long-term degradation of the transistor. The Si/SiGe HBT BVCEO and BVCBO maintain a roughly 1 : 3 relationship from 20 to 90 GHz.

length in the sub-0.25-μm region. In the case of a half-micrometer 20-GHz f_T device, the ratio is nearly two, but it drops to less than one for the 0.1-μm design.

The other absolutely key issue for RF applications of scaled transistors is the breakdown voltage of the device, which influences the dynamic range of operation. The breakdown voltage of a transistor is mostly an issue for the implementation of power amplifiers in the transmitter section, although other circuit areas can benefit from a high breakdown voltage as well. The breakdown voltage issue is complicated by the physics of the device at high electric fields, the varied physical mechanisms that lead to device failure, and the interaction of the breakdown mechanisms with the external circuit.

The bipolar device is fundamentally limited by avalanche multiplication in the collector-base region [32]. This breakdown effect is traded off against the increasing f_T of the transistor, and the $BV * f_T$ product is the key consideration for most high-frequency applications and is a material-related constant known as the *Johnson limit* [33]. In the bipolar device, the collector-base junction typically experiences avalanche breakdown first, and the device can be characterized by the collector-emitter breakdown voltage when the base is shorted to the emitter (BVCBO) or when the base is open-circuited (BVCEO). The former is usually larger than the latter, due to current gain in the emitter-base region, and can be approximated by [34]

$$\mathrm{BVCEO} \approx \frac{\mathrm{BVCBO}}{\beta^{1/m}} \tag{11}$$

where β is the dc current gain of the transistor and m is a constant that varies from 2 to 5, depending on a variety of physical factors. When the devices have very shallow doping (as in the high f_T case), the transistors exhibit nonlocal avalanche, and the $BV * f_T$ of the device can exceed its value seen for lower frequency devices [35]. Fig. 7 plots the BVCEO and BVCBO for modern bipolar devices, and the effects of nonlocal avalanching on breakdown voltage can clearly be seen at the higher f_T values, where the breakdown voltage does not change significantly as the f_T increases. In the operation of a power amplifier *circuit*, the device can typically operate at peak voltages in excess of BVCEO, but less than BVCBO, due to the time-dependent nature of the carrier multiplication process [36] and the impedances presented at each terminal.

This last issue of terminal impedances is crucial in the operation bipolar devices for power amplifiers, since the current gain at the emitter-base junction influences the breakdown characteristics. A simplified view of typical power amplifier operation is shown in Fig. 8, and the collector-base avalanche current can be modeled by

$$i_{\mathrm{av}} = C_{\mathrm{av}} {v_{CB}}^m i_c' \tag{12}$$

where C_{av} is a technology-dependent avalanche breakdown constant.

The transistor exhibits breakdown when $\partial i_c'/\partial v_{CE} \rightarrow \infty$ and therefore

$$\frac{\partial i_C'}{\partial v_{\mathrm{CE}}} \approx \frac{C_{\mathrm{av}} m {v_{\mathrm{CB}}}^{m-1} g_m' r_{\mathrm{in}} i_C'}{(1 - C_{\mathrm{av}} {v_{\mathrm{CB}}}^m g_m' r_{\mathrm{in}})} \tag{13}$$

where g_m' is the effective transconductance of the device, including the feedback effects of any extrinsic emitter impedance, and r_{in} is the input impedance consisting of the *parallel* combination of the extrinsic source impedance (including the base resistance R_b) and the input impedance due to the finite β.

Breakdown occurs when

$$C_{\mathrm{av}} {v_{\mathrm{CB}}}^m g_m' r_{\mathrm{in}} = 1. \tag{14}$$

In the limiting case of a low source impedance, r_{in} is simply the transistor base resistance R_b. Then (14) reduces to

$$C_{\mathrm{av}} {v_{\mathrm{CB}}}^m g_m' R_b = 1 \tag{15}$$

and

$$\mathrm{BV_{CBO}} \approx \left(\frac{1}{C_{\mathrm{av}} g_m' R_b}\right)^{1/m} \tag{16}$$

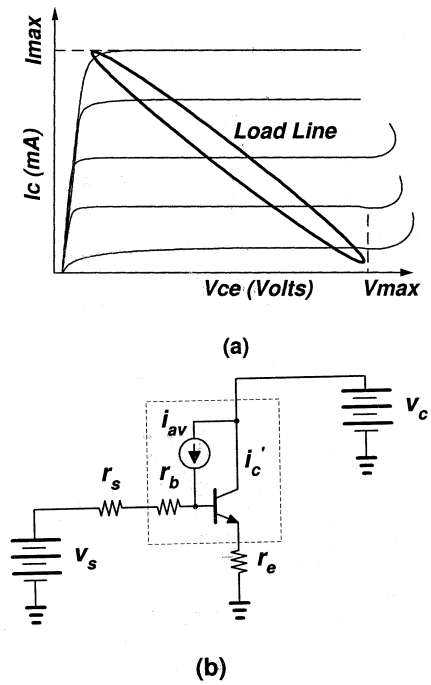


Fig. 8. Illustration of breakdown mechanisms in bipolar RF power amplifiers. (a) Current–voltage excursions of amplifier during large-signal operation. (b) Current behavior of bipolar transistor in the breakdown-limited region.

which illustrates the dependence of breakdown voltage on base resistance; as the base resistance increases, the internal feedback shunts more and more of the avalanche current to the emitter, increasing the positive feedback that leads to breakdown.

In the limit of a high source impedance ($\mathrm{BV_{CEO}}$), r_{in} increases to approximately β/g_m and

$$\mathrm{BV_{CEO}} \approx \mathrm{BV_{CBO}} \left(\frac{g_m' r_b}{\beta}\right)^{1/m} \tag{17}$$

which illustrates the well-known relationship between BVCBO and BVCEO in the bipolar transistor. The dependence of bipolar breakdown voltage on source impedance can be exploited in power amplifier design to significantly increase the safe operating voltage range.

The breakdown voltage mechanisms limiting MOSFET performance are complicated by the diverse breakdown mechanisms, primarily time-dependent dielectric breakdown (TDDB) due to impact ionization in the drain region, gate-oxide rupture, drain avalanche breakdown, parasitic bipolar transistor operation, and punchthrough [37].

From a reliability perspective, TDDB presents the most significant limitation on dynamic range in scaled MOSFETs. This effect is a result of damage to the silicon–oxide interface due to injection of hot electrons at the drain. This shifts the threshold voltage of the device over an extended period of time [38]. The recommended voltage limitations are typically based on dc or transient reliability tests, but in many RF applications the instantaneous dc voltage can significantly exceed the dc voltage, with potentially deleterious consequences. This phenomena has recently been observed to degrade the output power of a 0.18-μm CMOS power amplifier over a matter of days of operation [39].

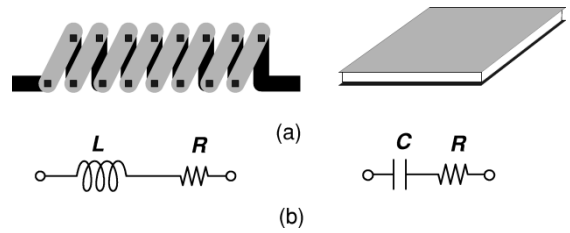


Fig. 9. Passive component scaling issues for monolithic implementation. (a) Cross section of toroidal inductor and MIM capacitor. (b) Simplified equivalent circuit model of inductor and capacitor.

A comparison of the HBT BVCEO and BVCBO and the recommended operating voltage for a MOSFET as a function of f_T is shown in Fig. 7. There seems to be a small but significant advantage for the bipolar device in this high-voltage regime, which is attributed to the fact that there is a cumulative degradation mechanism when the MOSFET is operated in the weak avalanche range of operation (due to the long-term shift in the threshold voltage). By comparison, bipolar devices appear to recover without any degradation in performance from weak avalanche breakdown in the collector-base junction. This will have a significant impact on the design of power amplifiers in these technologies, although it should be noted that LDMOS devices exhibit excellent performance in high-power base station amplifier applications [40]. In this case, the device is engineered to exhibit a very high breakdown voltage as well as acceptable gain at microwave frequencies, which is very different from design considerations that go into typical digital CMOS device scaling.

IV. Passive Device Performance Tradeoffs for RF SOCs

The required circuits for the implementation of these "systems-on-a-chip" require a wide variety of elements, over and above the n-channel and p-channel MOSFET and NPN of a typical BiCMOS digital ASIC. This includes the need to include high-quality inductors, capacitors, varactor diodes, transmission lines, and resistors. This section will discuss the tradeoffs and limitations of inductors and capacitors, which are two of the most challenging components to implement in monolithic form.

The implementation of high-quality monolithic inductors on silicon was considered to be an intractable problem until recently. The *fundamental* problem for integrated inductors is that they need to store much more energy than they dissipate per cycle, and it is very difficult to store a large amount of energy in the small volume of an integrated circuit die. The maximum stored energy per cycle is ($Li_{\max}^2/2$) and the average dissipated energy per cycle is ($i_{\max}^2 R/2\omega$), where L is the inductance, R the series resistance, $i_{\max}$ is the maximum current flowing through the inductor, and ω is the radian frequency.

This limitation on the energy can be seen by examining the fundamental scaling properties of the classical toroidal inductor shown in Fig. 9(a) and its equivalent circuit of Fig. 9(b) [41]. This scaling behavior of this structure is closely related to that of the planar inductors on an integrated circuit die.

In this case, the inductance is given by

$$L = n^2 F\,(d/l)\,d \tag{18}$$

where n is the number of turns, l is the length, d is the inductor diameter, and $F(d/l)$ is the "form factor" (which depends on the ratio of d to l).

The finite resistance of the wiring creates an equivalent series resistance of

$$R = \rho \frac{l_w}{A} \tag{19}$$

where ρ is the metal resistivity and l_w is the total length of the wire ($\approx n\pi d$).

The resulting quality factor of the inductor (the ratio of the stored to dissipated energy per cycle) is

$$Q_L = \frac{\omega L}{R}. \tag{20}$$

If we now adjust every dimension of the inductor by a factor of x, as would be the case for scaling a large inductor down to the size compatible with an integrated circuit, then the inductance becomes

$$L_x = n^2 F\left(xd/xl\right) xd = xL \tag{21}$$

and the series resistance becomes

$$R_x = \rho \frac{x l_w}{x^2 A} \tag{22}$$

and the resulting quality factor of the scaled inductor becomes

$$Q_{L_x} = \frac{\omega L_x}{R_x} = \frac{\omega x L}{R/x} = x^2 Q_L. \tag{23}$$

So, decreasing the physical size of the toroidal inductor by a factor of ten will reduce the resulting quality factor by a factor of one hundred. This argument accounts for the historically low Q of monolithic inductors compared with their discrete board level counterparts. For example, the quality factor of discrete surface mount RF inductors is at least a factor of ten higher than their integrated circuit counterparts [42]. This scaling argument also applies for inductors fabricated on an integrated circuit die, with some small modifications. For example, on an integrated circuit die, the metal thickness does not scale with the size of the inductor (the metal thickness is determined by the fabrication technology), so the decrease in quality factor due to scaling by an order of magnitude will be closer to a factor of ten than one hundred.

By comparison, the scaling properties of monolithic capacitors are much more advantageous. In this case, as the physical cross section and equivalent circuit of Fig. 9(a) and (b) shows, we have

$$C = \frac{\varepsilon A}{t_{ox}} \tag{24}$$

where A is the capacitor plate area, ϵ is the dielectric constant of the interlayer dielectric, and t_{ox} is the dielectric thickness. The equivalent series resistance of the capacitor, which is dominated by the metal resistance in a monolithic circuit, is

$$R = \frac{\rho K(t_m)}{A} \tag{25}$$

where ρ is the metal resistivity, and $K(t_m)$ is a factor that accounts for the contact resistance to the metal. The resulting Q for the series equivalent circuit of the capacitor—both before and after scaling—is therefore

$$Q = \omega C R \tag{26}$$

which is independent of scaling, since the resistance rises as the capacitance falls. As a result, the size of a monolithic capacitor can be dramatically reduced without affecting the resulting Q, and this is borne out in the measured data.

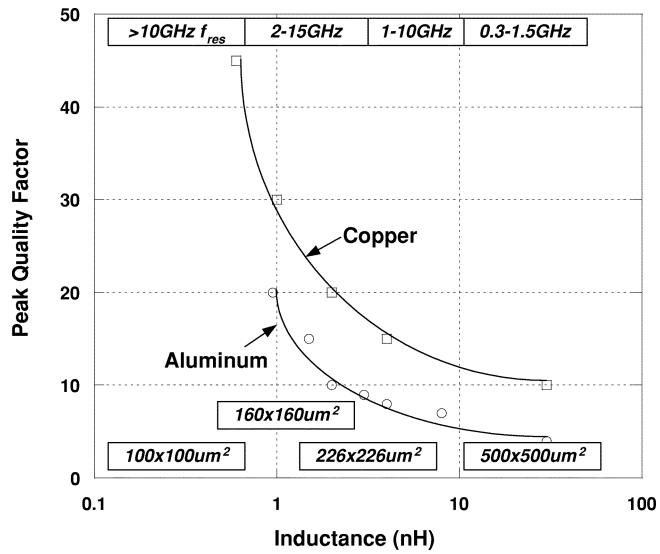


Fig. 10. Comparison of inductor peak Q as a function of inductance for Al-based and Cu-based metallization. Note that the area increases and the self resonant frequency decreases as the inductance grows [44].

A. Monolithic Inductors

Inductors are a crucial part of any RF-SOC implementation, and they are especially important for high-performance frequency synthesizers and LNAs. The major improvements in inductor performance have occurred through the application of more lightly doped silicon substrates, thicker dielectric layers, thicker metallization, as well as a move to copper metallization [43]. The improvement in inductor Q as a result of the migration from aluminum-based metallization to copper-based metallization is illustrated in Fig. 10 [44]. In the case of Cu metallization, the ohmic losses due to metal winding resistance are greatly reduced compared to Al-based structures. However, in most cases, the inductor performance then becomes dominated by losses in the silicon substrate. Further improvements in inductor Q will then require even larger dielectric stacks (to further separate the metallization from the lossy substrate) or some sort of transferred substrate approach to completely separate the inductor from the silicon [45].

There have been some incremental improvements in the design of monolithic spiral inductors using more exotic techniques as well, including the use of patterned ground shields to reduce eddy current losses [46] "hairpin" designs [47], micromachining techniques [48], and "solenoid"-based designs [49].

B. Monolithic Capacitors

Monolithic capacitors represent a relatively straightforward implementation of modern MOS technology to the problem of

a high-performance passive component and do not suffer from the quality factor limitations of monolithic inductor structures. As an example, reported Q values of ~80/f(GHz)/C(pF) for MIM caps with 0.7 fF/μm^2 and ~20/f(GHz)/C(pF) for MOS caps with 1.4 fF/μm^2 were reported in 1997 [50]. More recently, MIM capacitance densities of 2.7 fF/μm^2 with Q's of 150 were reported using a sputtered plasma-enhanced chemical vapor deposition (PECVD) nitride dielectric [51]. In the case of MIM capacitors, the challenge is to reduce the area of the capacitor, in order to reduce the area of the overall die, through the use of thinner dielectrics and higher dielectric constant materials.

Recently, in an attempt to provide for a high-capacitance-per-unit area in a standard digital CMOS process, several groups have reported "fractal" capacitors, where fringing fields are used to provide the capacitance [52], [53]. Although the capacitance values per unit area are rather modest compared to the above MIM structures (0.2–0.5 fF/μm^2), they provide an alternative approach for the realization of high-quality capacitors without the need for an extra process step, as in more traditional MIM capacitance structures.

V. Comparative LNA Performance of MOS and Bipolar Transistor Circuits for RF-SOC Applications

The front-end LNA of Fig. 1 is one of the key determiners of SOC performance, since the overall SNR of the final received signal is set by the noise performance of this particular amplifier. Typical wireless application frequencies today are in the 1–5-GHz range. Both bipolar and MOS transistors have been utilized recently for front-end applications in wireless systems. Fortunately, the microwave noise performance of both bipolar and MOS transistors has improved dramatically in recent years, thanks to aggressive technology scaling that was largely designed to improve digital circuit performance.

The input referred noise performance of a radio receiver determines the *minimum* signal level that can be reliably demodulated. As a result, it is a key factor in determining the range and power dissipation of the entire communications system. The *noise factor* (F)—defined as the degradation of the SNR of an input signal as it passes through the amplifier—is the standard metric for determining the noise performance of an RF receiver and is given by (27), shown at the bottom of the page, and the noise figure (NF) is defined as F in decibels [i.e., $NF = 10\log_{10}(F)$].

An equivalent circuit diagram of modern MOS and bipolar transistors, showing the major contributors to microwave noise performance, is shown in Fig. 11. In the case of the deep submicrometer MOSFET, the main contributors to the noise factor are the drain current noise and the thermal noise contributed by the extrinsic gate resistance. These two noise contributors are given by [54]

$$\overline{\nu_g^2} = 4kTR_g\Delta f$$
$$\overline{i_d^2} = 4kT\gamma_{\rm sat}g_{ds0}\Delta f \tag{28}$$

where k is Boltzman's constant, T is temperature, Δf is the measurement bandwidth, g_{ds0} is a conductance term that is equal to the drain–source conductance at $V_{DS} = 0$, and $\gamma_{\rm sat}$ is the "channel noise factor" [55].

The channel noise factor ($\gamma_{\rm sat}$) is a complicated factor of device design and bias conditions and can be approximated by [56]

$$\gamma_{\rm sat} \approx \frac{2}{3}\left[1 + \frac{I_D}{g_m n(kT/q)}\frac{\nu_{\rm eff}\tau_e}{L_{\rm eff}}\right] \tag{29}$$

where $\nu_{\rm eff}$ is the saturated electron drift velocity, τ_e is the carrier relaxation time, $L_{\rm eff}$ is the effective gate length, and n is the ratio between bulk transconductance and gate transconductance (approximately unity). The quantity $\gamma_{\rm sat}$ is approximately 2/3 for long-channel devices, but rises to nearly two for short-channel devices, is a strong function of applied gate-to-source voltage, and is also a weak function of drain-to-source voltage [57].

Given this noise model, the minimum noise factor (when the device is presented with an optimized source reactance) as a function of the source resistance is [58]

$$F_{\rm min} \approx 1 + \frac{R_g}{R_s} + \gamma_{\rm sat}g_{ds0}\frac{f^2}{f_T^2}\frac{(R_g+R_s)^2}{R_s} \tag{30}$$

and the noise figure is minimized at a source resistance of

$$R_{s,\,\rm opt} \approx \sqrt{\frac{R_g}{\gamma_{\rm sat}g_{ds0}}}\frac{f_T}{f} \tag{31}$$

and the minimum noise factor at that source impedance is approximately

$$F_{\rm min} \approx 1 + 2\frac{f}{f_T}\sqrt{\gamma_{\rm sat}g_{ds0}R_g}. \tag{32}$$

Therefore, the noise figure of the MOSFET is primarily determined by the gate resistance and the f_T of the transistor. As technology scales to shorter and shorter gate lengths, the gate resistance may become an increasingly dominant factor, even as the noise figure itself improves. In a practical circuit, the noise figure is limited by technological factors, as well as by the fact that the optimum source impedance is rising as the gate length shrinks, and this impedance may not be achievable using practical circuit components. This simplified model leaves out the induced gate noise due to the drain current, which has been analyzed by several authors but the overall trend in the results remains the same [59].

$$F = \frac{\text{Total Noise Power Delivered to the Load Impedance}}{\text{Total Noise Power Delivered to the Load Impedance due } only \text{ to the Source}} \tag{27}$$

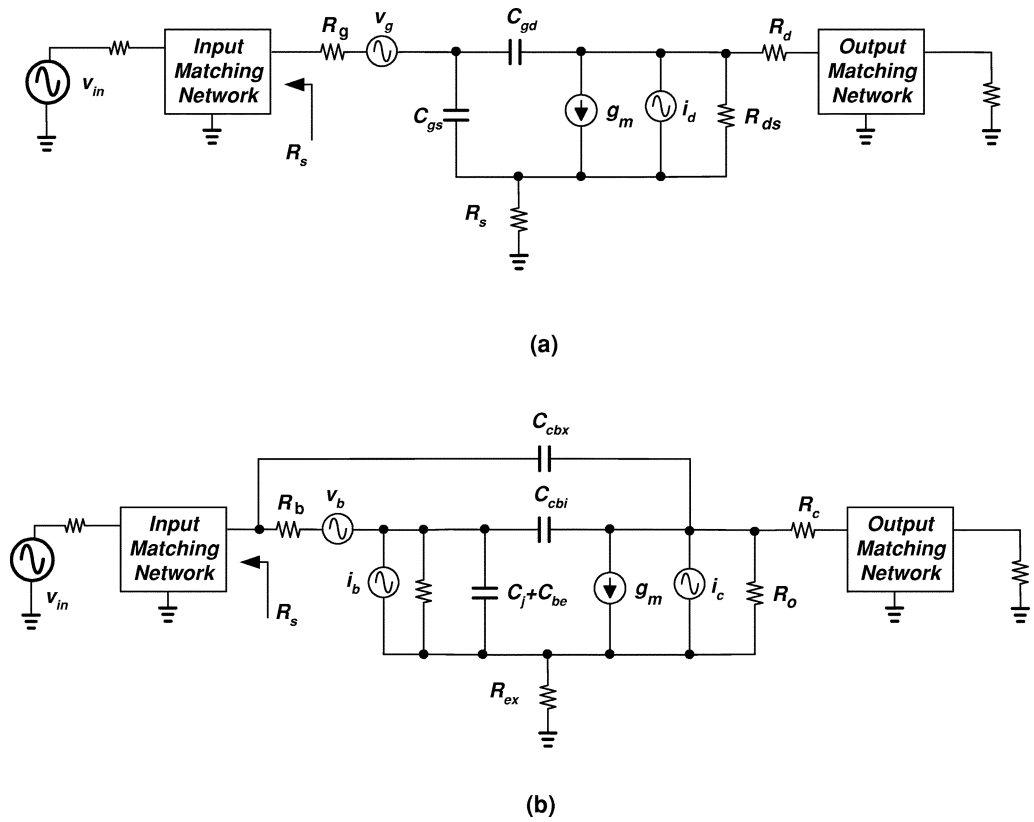


Fig. 11. Equivalent circuit noise and low-noise circuit models of (a) NMOS FET and (b) Si/SiGe HBT LNA.

The equivalent circuit model for a bipolar transistor results in a similar set of design tradeoffs for device design and noise optimization. In this case, there are three dominant broad-band noise sources given by [60]

$$\overline{v_b^2} = 4kTR_b\Delta f \tag{33a}$$

$$\overline{i_b^2} = 2qI_{\mathrm{BQ}}\Delta f \tag{33b}$$

$$\overline{i_c^2} = 2qI_{\mathrm{CQ}}\Delta f \tag{33c}$$

where R_b is the base resistance and I_{CQ} and I_{BQ} are the dc collector and base currents, respectively. The quantities $\overline{i_b^2}$ and $\overline{i_c^2}$ are normally considered to be uncorrelated, but at high frequencies, the correlation between the two is given by [60]

$$\langle i_b^* i_c \rangle = 2qI_c(e^{-j\omega\tau} - 1). \tag{34}$$

Given this noise model of the bipolar transistor, the transistor will exhibit the following minimum noise factor as a function of source impedance when presented with the optimum source reactance $X_s \approx 1/\omega C_\pi$ [58]:

$$F_{\min} \approx 1 + \frac{R_b}{R_s} + \frac{g_m}{2}\frac{f^2}{f_T^2}\frac{(R_b+R_s)^2}{R_s} + \frac{g_m}{2\beta_0}\frac{f^2}{f_T^2}\frac{\left[(R_b+R_s)^2 + X_s^2\right]}{R_s}. \tag{35}$$

Note that—in the limit of high β_0—this result is quite similar to the minimum noise factor in the MOS case, with R_b replacing R_g and $g_m/2$ replacing $\gamma_{\mathrm{sat}} g_{ds0}$. So, the keys to lowering the noise figure of the bipolar device are the reduction in R_b and/or an increase in the f_T. The main difference between the MOS and bipolar case is the final term in (35), which is dependent on the current gain of the device, and results in the well-understood role of base current in limiting the noise performance of bipolar amplifiers. In most cases, this final term is small relative to the first two, especially for modern HBT devices with high dc current gains at high frequencies (larger than the frequency where the current gain begins to decline from β_o) [61].

A comparison of the reported noise figure characteristics of SiGe HBT and MOS devices confirms the analysis given above, as shown in Fig. 12. The key determinant of noise figure performance is the f_T of the device, with MOS devices demonstrating roughly a 0.5-dB improvement for a given f_T compared to an equivalent HBT device. The difference is mostly attributed to the relatively higher base resistance of the HBT compared to the MOSFET. However, as Voinigescu *et al.* pointed out, this advantage in *intrinsic* noise performance of the MOSFET is difficult to realize in *practice*, because the optimum source impedance for the MOS device is much higher than that of the HBT, making the noise figure of a MOSFET LNA very sensitive to source impedance mismatch [62]. One solution to this dilemma is to increase the size of the MOSFET, at the expense of higher power dissipation. In this case, the bipolar LNA would exhibit a slightly lower power dissipation than the MOSFET implementation for a given noise figure.

Direct comparisons between the noise figure performance of MOS and bipolar LNAs are difficult to perform, due to inevitable circuit interaction effects but a recent result [63] where a 0.25-μm MOSFET amplifier (whose peak f_T was

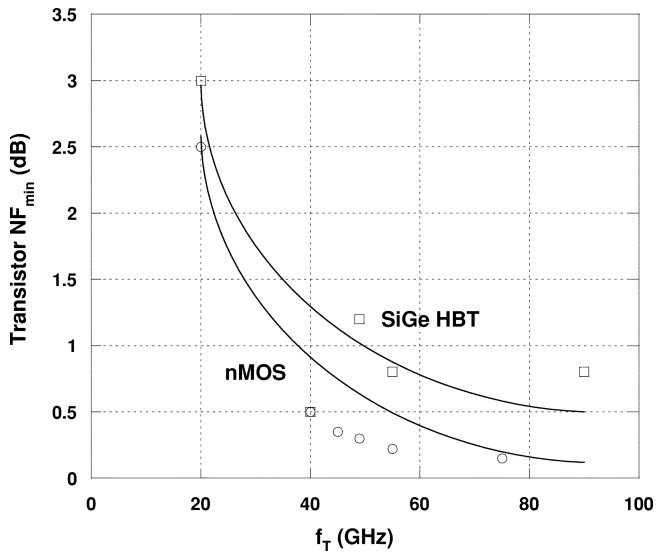


Fig. 12. Comparison of reported SiGe HBT and MOSFET minimum device noise figures as a function of peak f_T. For an equivalent intrinsic device speed, the MOSFET typically has an approximately 0.5-dB advantage, but this is difficult to realize in practice in a monolithic circuit due to the higher source impedance required.

roughly 50 GHz) was compared with a 50-GHz HBT amplifier showed essentially equivalent noise figures at 2.4 GHz (2.9 dB), with power dissipation roughly 20% higher in the MOS case. Both MOS and bipolar transistors exhibit relatively broad $F_{\min}$ versus current behavior, and so device level power comparisons between the two technologies are complicated by the simultaneous requirement for low noise and high linearity, as the next section will demonstrate.

A. Comparative Linearity Performance of MOS and Bipolar Transistor Circuits for RF-SOC Applications

Circuit linearity affects the performance of both the transmitter and receiver sections of the RF SOC and the requirements of the two sections differ significantly. An RF receiver is typically operated well below its 1-dB compression point, and therefore *small-signal* linearity is the key performance metric. In the GSM receiver case, the circuit must be able to amplify a signal of roughly 10^{-13} W while simultaneously receiving an undesired signal many orders of magnitude larger. The key figures-of-merit (FOMs) here are the input intercept point and cross-modulation sensitivity. Transmitters are typically operated at high levels of output power, and so their *large-signal* linearity is the key consideration.

From the perspective of receiver design, which encompasses the low-noise amplification stages as well as the downconversion mixer, circuit nonlinearity arises from weak nonlinearities both in the dependent sources (principally the transconductance) and charge storage elements (capacitors) within the transistor; at low frequencies, the former consideration dominates. In the low-frequency case, the small-signal output voltage of a weakly nonlinear circuit can be described by a power series of the form

$$i_o = a_1 v_i + a_2 {v_i}^2 + a_2 {v_i}^3 + \cdots \tag{36}$$

where i_o is the output current, v_i is the input voltage, and $a_1 \cdots a_3$ are the power-series coefficients of the amplifier response [64]. Intuitively, the linearity of the circuit will be improved if the higher order power series coefficients a_2, a_3 are reduced compared to a_1. Unlike in the case of low-noise performance, the linearity behavior of scaled bipolar and MOS transistors in the low-frequency regime are very different.

The low-frequency collector current of bipolar devices continues to be determined by the well-known exponential relationship to base-emitter voltage ($I_c = I_s \exp(V_{be}/V_T)$), even as the device is scaled into the regime where f_T exceeds 200 GHz [20]. In this case, with the typical common-emitter amplifier circuit with load impedance R_L, operated with an ideal voltage source input, the power-series coefficients are given by

$$a_n = I_{C0}/(n! V_T) \tag{37}$$

which shows that the relative ratio of the various power series coefficients are *independent* of the dc operating current. This point will become more significant when we introduce linearity FOMs.

The low-frequency linearity behavior of MOS devices is not as simply described as that of the bipolar transistor and, unlike the bipolar device, exhibits significant changes as a result of technology scaling. A simple expression for MOSFET drain current in strong inversion and saturation, which will help to illustrate this effect, is given by [65]

$$i_D = \frac{\mu_0 C_{\rm ox} W}{aL} \frac{(v_{GS} - V_{\rm TH})^2}{1 + (\theta + \mu_0/v_{\rm eff} L)(v_{\rm GS} - V_{\rm TH})} \tag{38}$$

where μ_0 is the low-field electron mobility, a is a dimensionless body-effect parameter that is close to unity, $C_{\rm ox}$ is gate oxide capacitance per unit area, W and L are the gate width and length, $V_{\rm TH}$ is the threshold voltage, and θ is a mobility reduction factor due to the normal gate field [66].

In this case, the power series coefficients for the nonlinear MOSFET amplifier response are given by [67]

$$a_1 = \frac{\mu_0 C_{\rm ox} W}{2aL}(v_{\rm GS} - V_{\rm TH}) \cdot \frac{2 + (\theta + \mu_0/v_{\rm SAT} L)(v_{\rm GS} - V_{\rm TH})}{\{1 + (\theta + \mu_0/v_{\rm SAT} L)(v_{\rm GS} - V_{\rm TH})\}^2} \tag{39a}$$

$$a_2 = \frac{\mu_0 C_{\rm ox} W}{2aL} \frac{1}{\{1 + (\theta + \mu_0/v_{\rm SAT} L)(v_{\rm GS} - V_{\rm TH})\}^3} \tag{39b}$$

$$a_3 = -\frac{\mu_0 C_{\rm ox} W}{2aL} \frac{\theta + \mu_0/v_{\rm SAT} L}{\{1 + (\theta + \mu_0/v_{\rm SAT} L)(v_{\rm GS} - V_{\rm TH})\}^4}. \tag{39c}$$

In order to put these results into context, we need to explore how these nonlinearity coefficients (both bipolar and MOS) affect the linearity of the receiver circuit. The standard small-signal linearity FOM for a receiver amplifier is the third-order input-referred intercept point (IIP3). This is defined as the input power level of two input signals (at frequencies f_1 and f_2) where the extrapolated undesired third-order output

nonlinear response intersects the desired first-order linear response [68]. The third-order responses are particularly insidious in a narrow-band communication system, especially because one of them appears at $2f_1 - f_2$ and another at $2f_2 - f_1$; both frequencies are close to the original frequencies f_1 and f_2. Although this figure of merit has many limitations in practical situations, its ease of measurement and calculation make it a perennial favorite among microwave engineers.

The second-order input-referred intercept point (IIP2)—the input power level where the extrapolated second-order response intersects the desired first-order response—is also sometimes specified, although it is usually less important than the IIP3. This is due to the fact that the frequency of the second-order distortion product is well away from the desired signal (at $f_1 + f_2$ or $f_1 - f_2$), whereas the third-order response frequency is nearly the same as that of the two original input tones. The *input* intercept points can be referred to the *output* by simply multiplying by the gain of the circuit.

With a power series model of the amplifier, the IIP3 *voltage* is given by

$$\mathrm{IIP}_3 = \sqrt{\frac{4}{3}\frac{|a_1|}{|a_3|}}. \tag{40}$$

The IIP3 of the bipolar transistor *at low frequencies and without feedback* is then simply given by $2\sqrt{2}V_T$—roughly 75 mV at room temperature.

By contrast, the IIP3 of the MOS device is "theoretically" infinite in the long gate length regime (where L and t_{ox} are both relatively large). Even in the short gate length regime the linearity of the MOSFET is excellent, and the low-frequency IIP3 voltage of the MOSFET amplifier reduces to

$$IIP3 \approx \frac{\{1 + (\theta + \mu_0/v_{\mathrm{eff}}L)(v_{\mathrm{GS}} - V_{\mathrm{TH}})\}^2}{\theta + \mu_0/v_{\mathrm{eff}}L} \tag{41}$$

which shows that the intrinsic linearity performance of the short-channel MOSFET exhibits a moderate increase with bias voltage, unlike the bipolar transistor. At relatively large values of v_{GS} and short gate lengths, the IIP3 can be further simplified to

$$IIP3 \approx (\theta + \mu_0/v_{\mathrm{eff}}L)(v_{\mathrm{GS}} - V_{\mathrm{TH}})^2. \tag{42}$$

Some typical values for these parameters are $\mu_0 = 0.048$ m^2/(V $\cdot$ s), $\theta = 0.42$ V^{-1}, and $v_{\mathrm{eff}} = 1.95 \times 10^5$ m/s, and $L = 0.25$ μm which yields an IIP3 of approximately 1.5 V with a $v_{\mathrm{GS}} - V_{\mathrm{TH}} \approx 1$ V. This is substantially higher than a bipolar transistor operated under conditions of equal current, which accounts for the improved low-frequency linearity of MOS devices compared to their bipolar counterparts. This improvement was confirmed by the experimental results in [63].

This analysis provides a starting point for a comparison of the two device technologies, illustrating the intrinsic differences between the MOSFET and bipolar transistor, but the linearity performance will change at higher frequencies, due to the nonlinear behavior of the stored charge, and circuit impedances will provide *feedback* that further alters the linearity.

If we examine the case of the higher frequency performance, the situation is complicated by the nonlinear stored charge effects and the impedances at each terminal of the transistor. These nonlinearities introduce a *frequency dependence* to the nonlinearity, which considerably complicates the analysis. The situation can be simplified if we consider resistive terminations *only* at each terminal of the transistor. In this case, the work of Vaidyanathan *et al.* [69] employing a Volterra series analysis clarifies the relationship between the high-frequency linearity of the bipolar transistor and its physical design, particularly the relationship between the high-frequency linearity and the behavior of its "loaded" unity current-gain frequency $\widehat{f_T}$, where the loaded unity current-gain frequency is defined as the frequency where the current-gain drops to unity *with the appropriate terminating impedances*.

As an example, at sufficiently high frequencies, and without avalanche breakdown occurring, the OIP2 of a bipolar transistor is given by the relatively simple relationship [69]

$$OIP2(2f) \approx \left|\frac{4\widehat{f_T}}{\widehat{f_T}'}\right| \tag{43}$$

where $\widehat{f_T}'$ is the derivative of $\widehat{f_T}$ with respect to collector current (in the case of the bipolar transistor). To minimize the second-order intermodulation distortion, the transistor should be designed to have as constant an $\widehat{f_T}$ as possible, and the device will have the highest OIP2 near the peak of the $\widehat{f_T}$ versus i_C curve.

The important OIP3 behavior is more complicated than in the OIP2 case, but some important generalizations can be derived from the analysis of device operation. At sufficiently high frequencies, and when the device is operated at the peak of the $\widehat{f_T}$ curve, the OIP3 of the bipolar transistor is given by [69]

$$\mathrm{OIP3_{HF}}\,(2f_1 - f_2)\Big|_{\widehat{f_T}'=0} \approx \left|\frac{8\widehat{f_T}}{\widehat{f_T}''}\right|^{1/2} \tag{44}$$

where $\widehat{f_T}''$ is the second derivative of the $\widehat{f_T}$ with respect to collector current. This result implies that, when the device is operated at the peak of its $\widehat{f_T}$ versus collector current curve, the best distortion performance is obtained when the device has a high $\widehat{f_T}$ and when the $\widehat{f_T}$ curve is as "flat" as possible. Both the OIP2 and OIP3 results cited above demonstrate that the "ideal" bipolar transistor—defined as one with very low junction capacitances and hence nearly constant $\widehat{f_T}$—will have outstanding high-frequency linearity and that this intrinsic linearity can improve with future device scaling. As the devices scale to higher f_T, avalanche breakdown in the collector region becomes a significant factor and can also have a deleterious effect on linearity, and this has been examined in several recent papers [70], [71].

The distortion results presented so far are *device-oriented* in the sense that they do not include frequency-dependent circuit impedance termination effects. These can improve (or degrade) the performance of the actual amplifier, depending on a variety of factors [72]. Although the circuit interaction effects for linearity of microwave amplifiers are very complicated—far more so than for noise factor determination—there are some general results that can be used [73]. In particular, the IIP3

of a low-noise bipolar amplifier was improved by over 10 dB through the use of optimized termination impedances at the *sum* and *difference* frequencies of the two-tone input signals [74]. This nonlinear cancellation effect is especially useful in bipolar amplifiers because, unlike MOSFETs, the a_2 and a_3 coefficients in (37) have a well defined relationship, and so the cancellation of third-order nonlinearities can be nearly perfect when proper terminations are chosen [74].

VI. Comparative Performance of MOS and Bipolar VCOs for RF SOC Applications

The VCO provides the frequency reference for the upconversion of the transmitted signal or downconversion of the received signal, as shown in Fig. 1. The VCO frequency is usually not accurate enough by itself to provide the correct downconversion or upconversion frequency and so is usually *phase-locked* to a more precise reference frequency. The key performance issues with this circuit are *phase noise*, *power dissipation*, and *frequency tuning range*. Unlike many other circuits, the performance of the *passive* devices can have a significant impact on the performance of this circuit.

The phase noise of the oscillator is the ratio of the power in the desired output (the carrier) to the output power in a 1-Hz bandwidth at a given frequency offset from the carrier, when the amplitude variation on the carrier has been removed through a limiting process. So, the phase noise is expressed in units of dBC/Hz at a specified offset frequency. Ideally, the spectrum of the VCO output is a delta-function in the frequency domain, so the ideal VCO phase noise would be infinite dBc/Hz at all offset frequencies. Phase noise contributes to a variety of deleterious effects in radio systems, including a rise in the receiver noise floor and reciprocal mixing [68].

A simplified schematic of a bipolar transistor monolithic differential LC-tuned VCO, along with its most significant noise sources is seen in Fig. 13. The cross-coupled differential transistor pair presents a negative impedance to the resonator, cancelling the resistive losses in the resonator and enabling sustained oscillation. Frequency variation is achieved with a reverse-biased pn-junction diode or accumulation-mode MOS varactor, which changes the resonant frequency of the circuit.

The close-in phase noise behavior at an offset f_m from the carrier frequency f_0 in the differential LC-tuned VCO is determined from the well-known Leeson's model to be [75]

$$\mathcal{L}(f_m) = 10\log\left\{\frac{2kTR_{eq}F}{A_0^2}\left(\frac{f_0}{2Qf_m}\right)^2\left(1+\frac{\Delta f_{1/f^3}}{f_m}\right)\right\} \tag{45}$$

where k is Boltzman's constant, T is the absolute temperature, A_0 is the amplitude of oscillation, Q is the resonator loaded quality factor, $\Delta f_{1/f^3}$ is the corner frequency where the $1/f$ noise is no longer significant, and F is the excess noise factor.

Leeson's model shows that phase noise is reduced as the amplitude of oscillation is increased. However, once the amplitude of oscillation drives the transistors in the cross-coupled differential pair into saturation the loaded quality factor of the resonator is lowered and phase noise degrades significantly. It also

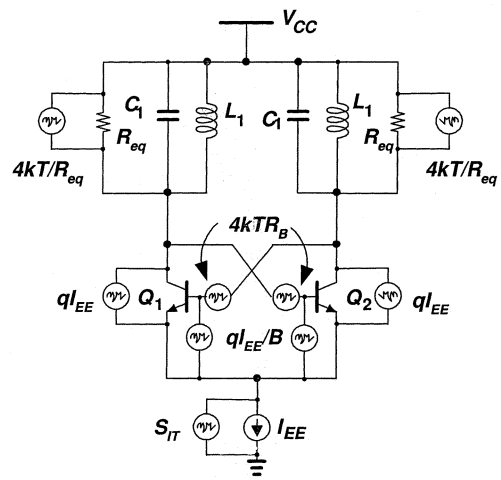


Fig. 13. Simplified schematic of a monolithic bipolar transistor LC-tuned VCO with noise sources.

illustrates the tradeoff between the *power dissipation* and *phase noise*, since a large amplitude will lead to both lowered phase noise and higher power dissipation.

A benefit of a bipolar transistor VCO design is that the corner frequency $\Delta f_{1/f^3}$ will be very low due to the excellent linearity of transistors and the low level of $1/f$ noise in the devices, rendering the frequency upconversion process from $1/f$ device noise relatively insignificant [76]. The contribution of $1/f$ noise from MOS-based VCOs is expected to be much larger, due to their intrinsically higher level of $1/f$ noise [77]. However, the upconversion of $1/f$ noise in CMOS VCOs can be dramatically reduced through symmetric circuit operation, as illustrated in [78].

Leeson's Equation clearly shows the importance of maximizing the Q factor of the resonating circuit, through the techniques described in Section IV. The excess noise factor F is determined by the wideband noise from the cross-coupled differential transistor pair and the dc current noise source, taking the nonlinear operation of the oscillator into account. In the case of a bipolar VCO, the excess noise factor F is given by [79]

$$F \approx 1 + \frac{R_b}{2R_{\rm eq}}\left(\frac{f_T}{f_0}\right) + \frac{qI_TR_{\rm eq}}{4kT}\left(\frac{\Delta V}{\pi A_0}\right)^2 \cdot\left(1+\operatorname{sinc}^2\left(\frac{\Delta V}{2A_0}\right)\right) + \frac{S_{I_T}R_{\rm eq}}{8kT} \tag{46}$$

where ΔV is the signal level required to make the cross-coupled differential transistor pair switch completely to one side, $R_{\rm eq}$ is the parallel equivalent impedance of the resonator, I_T is the dc current, and S_{I_T} is the mean square current noise power spectral density.

The odd harmonics around f_0 of the base resistance R_{bb} thermal noise are modulated into the resonator passband with approximately equal weights. The contribution to the excess noise factor from this effect becomes $\approx(R_{bb}/2R_{\rm eq})f_T/f_0$. This can be estimated by assuming that the wide-band noise spectrum has been sampled with a periodic impulse train at twice the oscillation frequency [80]. This illustrates the importance of minimizing base resistance for low-phase-noise

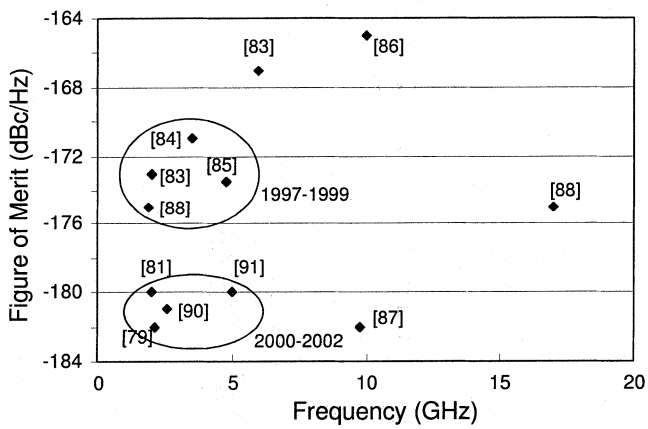


Fig. 14. Comparison of published VCO phase-noise FOM from a variety of sources. [79], [81]–[91], Note that there has been a significant improvement in recent years due to improved back-end metallization and circuit design approaches.

operation as well as the slight penalty incurred through the use of a high f_T device.

The shot noise contribution of the individual transistors (Q_1 and Q_2) takes place in a short time during the zero crossings of the output waveform. The shot noise contributes a factor of $(qI_T R_{\rm eq}/4kT)(\Delta V/\pi A_0)^2(1+\mathrm{sinc}^2(\Delta V/2A_0))$ to the excess noise factor. Low-frequency noise from the dc current source results in amplitude modulation of the carrier and therefore little phase noise contribution from this source. However, dc current source noise at frequencies near the even harmonics of the oscillator creates both amplitude and phase noise. If the dc current source noise has a spectral density S_{I_T}, it contributes $S_{I_T} R_{\rm eq}/8kT$ to the excess noise factor [80]. The contributions to F from this last noise source have been reduced through the use of *noise filtering* techniques, which essentially reduce the noise transfer function to the output of the second harmonic contributions [81].

The performance of monolithic VCOs is affected by so many diverse factors that it is difficult to draw meaningful comparisons between various technology and circuit approaches. A VCO FOM can be defined that provides for some insight into this issue [80], where

$$\mathrm{FOM_{VCO}} = \mathcal{L}(f_m) - 20\log\frac{f_o}{f_m} + 10\log\frac{P_{\rm diss}}{1\ \mathrm{mW}} \qquad (47)$$

where $P_{\rm diss}$ is the dc power dissipated by the VCO.

Fig. 14 is a plot of the measured FOM for a variety of reported monolithic VCOs (in both bipolar and CMOS technology) as a function of frequency. There is no clear trend in the comparative performance of bipolar versus CMOS technologies—their comparative performance is comparable. However, the plot shows that the performance of VCOs in both technologies has improved significantly in recent years—by roughly 8 dB in the last five years. This is attributed to improvements in inductor quality factor, as was discussed in Section IV, as well as to improved circuit design techniques that filter away much of the noise created by the dc biasing circuits [81].

VII. Conclusion

The performance of RF-oriented "systems-on-a-chip" has been historically limited by the performance of the active and passive devices available from a typical CMOS or BiCMOS integrated circuit technology. In recent years, advances in process technology—mostly intended to improve the performance of digital integrated circuits—have improved the performance of these higher frequency RF circuits as well.

The *fundamental* requirements of these circuits are those of low noise and (simultaneously) high linearity. This paper has outlined the effect that semiconductor scaling will have on these two performance issues in the coming years. The improvement in device speed (through reduction in lithographic dimensions) will continue to enhance RF circuit performance for many years, although limitations of gate and/or base resistance are becoming increasingly dominant in the sub-0.1-μm regime. At the same time, the dynamic range of the circuits will become increasingly challenged—more so with MOSFET than with HBT technology—because voltage limits are being reduced along with gate dimensions. HBT's appear to have some advantages in this regard compared to MOSFETs, since they can accommodate weak avalanche effects without long-term degradation.

The performance of other important RF circuits—such as the VCO—is primarily limited by the performance of the *passive* device technology, particularly the monolithic inductors, as well as improved circuit design techniques.

Acknowledgment

The author would like to acknowledge many valuable discussions with Prof. P. Asbeck and Prof. I. Galton from the University of California, San Diego (UCSD), Prof. M. Rodwell and S. Long from the University of California, Santa Barbara, Prof. J. Long, Prof. L. DeVreede, and Prof. J. Burkhartz from the Technical University of Delft, and Dr. B. Meyerson and Dr. D. Harame from IBM. The author would further like to acknowledge many valuable insights gained from discussions with S. Rosenbaum of Magis Networks, Dr. M. Vaidyanthan of UCSD, T. Johansen of the Technical University of Denmark, D. Rowe of Sierra Monolithics, and Dr. P. Gudem and V. Aparin of Qualcomm, Inc.

References

[1] P. C. Davis, "Merits and requirements of a few RF architectures," in *Proc. 1999 Bipolar/BiCMOS Circuits and Technology Meeting*, Minneapolis, MN, Sept. 1999, pp. 62–66.

[2] S. Redl, M. Weber, and M. Oliphant, *An Introduction to GSM*. Norwell, MA: Artech House, 1995.

[3] B. Gilbert, "The design of Si/SiGe LNA's from the ground up," in *IEEE BCTM Short Course Notes*, 1996.

[4] B. Razavi, "Design considerations for direct-conversion receivers," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 428–435, June 1997.

[5] J. Groe and L. Larson, *CDMA Mobile Radio Design*. Norwell, MA: Artech House, June 2000.

[6] J. Cheah, E. Kwek, E. Low, C. Quek, C. Yong, R. Enright, J. Hirbawi, A. Lee, H. Xie, L. Wei, L. Luong, J. Pan, S. Yang, W. Lau, and W. Ngai, "Design of a low-cost integrated 0.25 μm CMOS Bluetooth SOC in 16.5 mm^2 silicon area," in *Proc. 2002 IEEE Int. Solid-State Circuits Conf.*, vol. 45, Feb. 2002, pp. 64–66.

[7] T. Ohguro, T. Ishikawa, T. Kimura, S. Samata, A. Kawasaki, T. Nagano, T. Yoshitomi, and T. Toyoshima, "High performance digital-analog mixed device on an Si substrate with resistivity beyond 1 kΩ-cm," in *IEDM Tech Dig.*, 2000, pp. 757–760.

[8] R. H. Rasshofer, E. M. Biebl, K. M. Strohm, and J. F. Luy, "Long-term stability of passive millimeterwave circuits on high-resistivity silicon substrates," in *Proc. 1999 IEEE MTT-S Int. Microwave Symp.*, vol. 2, pp. 585–588.

[9] J. Burkhartz, M. Bartek, B. Rejaei, P. Sarro, A. Polyakov, N. Pham, E. Boullard, and K. Ng, "Substrate options and add-on process modules for monolithic RF silicon technology," in *Proc. 2002 IEEE Bipolar Circuits and Technology Meeting*, pp. 17–20.

[10] J. Raskin, R. Gillon, J. Chen, D. Vanhoenacker-Janvier, and J. Colinge, "Accurate SOI MOSFET characterization at microwave frequencies for device performance optimization and analog modeling," *IEEE Trans. Electron Devices*, vol. 45, pp. 1017–1025, May 1998.

[11] M. Stuber, M. Megahed, J. Lee, T. Kobayashi, and H. Domyo, "SOI CMOS with high-performance passive components for analog, RF, and mixed signal design," in *Proc. 1998 IEEE Int. SOI Conf.*, pp. 99–100.

[12] L. Nanver, H. van Zeijl, H. Schellevis, R. Mallee, J. Slabbekoorn, R. Dekker, and J. Slotboom, "Ultra-low-temperature low-ohmic contacts for SOA applications," in *Proc. 1999 Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 137–140.

[13] K. Han-Su, C. Kyuchul, X. Ya-Hong, M. Devincentis, T. Itoh, A. Becker, and K. Jenkins, "A porous Si based novel isolation technology for mixed-signal integrated circuits," in *Proc. 2002 Symp. VLSI Technology*, pp. 160–161.

[14] J. Wu, J. Scholvin, J. del Alamo, and K. Jenkins, "A Faraday cage isolation structure for substrate crosstalk suppression," *IEEE Microwave Wireless Components Lett.*, vol. 11, pp. 410–412, Oct. 2001.

[15] N. Pham, P. Sarro, K. Ng, and J. Burghartz, "IC-compatible two-level bulk micromachining for RF silicon technology," in *Proc. 30th Eur. Solid-State Device Research Conf.*, pp. 204–207.

[16] P. van Zeijl, J. Eikenbroek, P. Vervoort, S. Setty, J. Tangenberg, G. Shipton, E. Kooistra, I. Keekstra, and D. Belot, "A Bluetooth radio in 0.18 μm CMOS," in *Proc. 2002 IEEE Int. Solid-State Circuits Conf.*, vol. 45, Feb. 2002, pp. 86–87.

[17] H. Hsu, J. Chang, J. Su, C. Tsai, S. Wong, C. W. Chen, K. R. Peng, S. P. Ma, C. H. Chen, T. H. Yeh, C. H. Lin, Y. C. Sun, and C. Y. Chang, "A 0.18 μm foundry RF CMOS technology with 70 GHz f_T for single-chip system solutions," in *Proc. 2001 IEEE MTT-S Int. Microwave Symp.*, June 2001, pp. 1869–1872.

[18] T. Blalack, Y. Leclercq, and C. Yue, "On-chip RF isolation techniques," in *Proc. 2002 Bipolar/BiCMOS Circuits and Technology Meeting*, Oct. 2002, pp. 205–211.

[19] A. Cathelin, D. Saias, D. Belot, Y. Leclerq, and F. Clement, "Substrate parasitic extraction for RF integrated circuits," in *Proc. Design Automation and Test in Europe*, Mar. 2002, poster 4C-2.

[20] B. Jagannathan, M. Khater, F. Pagette, J. Rieh, D. Angell, H. Chen, H. J. Florkey, F. Golan, D. Greenberg, R. Groves, S. Jeng, J. Johnson, E. Mengistu, K. Schonenberg, C. Schnabel, P. Smith, A. Stricker, D. Ahlgren, G. Freeman, K. Stein, and S. Subbanna, "Self-aligned site NPN transistors with 285 GHz f_{MAX} and 207 GHz f_T in a manufacturable technology," *IEEE Electron Device Lett.*, vol. 23, pp. 258–260, May 2002.

[21] A. Chatterjee, D. Mosher, S. Sridhar, Y. Kim, and M. Nandakumar, "Analog integration in a 0.35μm Cu metal pitch, 0.1μm gate length, low-power digital CMOS technology," in *IEDM Tech. Dig.*, Dec. 2001, pp. 10.1.1–10.1.4.

[22] M. Vaidyanythan and D. Pulfrey, "Extrapolated f_{MAX} of heterojunction bipolar transistors," *IEEE Trans. Electron Devices*, vol. 46, pp. 301–309, Feb. 1999.

[23] M. Rodwell, M. Urteaga, T. Mathew, D. Scott, D. Mensa, Q. Lee, J. Guthrie, Y. Betser, S. Martin, R. Smith, S. Jaganathan, S. Krishnan, S. Long, R. Pullela, B. Agarwal, U. Bhattacharya, L. Samoska, and M. Dahlstrom, "Submicron scaling of HBTs," *IEEE Trans. Electron Devices*, vol. 48, pp. 2606–2624, Nov. 2001.

[24] H. Kroemer, "Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region," *Solid-State Electron.*, vol. 28, no. 11, pp. 1101–1103, Nov. 1985.

[25] T. Ishibashi, "Influence of electron velocity overshoot on collector transit times of HBTs," *IEEE Trans. Electron Devices*, vol. 37, pp. 2103–2105, Sept. 1990.

[26] U. Konig, A. Gruhle, and A. Schuppen, "SiGe devices and circuits: Where are the advantages over III-V?," in *Proc. 1995 IEEE GaAs IC Symp.*, pp. 14–18.

[27] B. Razavi, Y. Ran-Hong, and K. Lee, "Impact of distributed gate resistance on the performance of MOS devices," *IEEE Trans. Circuits Syst. I*, vol. 41, pp. 750–754, Nov. 1994.

[28] L. Kim and R. Dutton, "Modeling the distributed gate RC effect in MOSFETs," *IEEE Trans. Computer-Aided Design*, vol. 8, pp. 1365–1367, Dec. 1989.

[29] A. Schmitz, R. Walden, L. Larson, S. Rosenbaum, R. Metzger, J. Behnke, and P. Macdonald, "A deep-submicrometer microwave/digital CMOS/SOS technology," *IEEE Electron Device Lett.*, vol. 12, pp. 16–17, Jan. 1991.

[30] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. New York: WCB McGraw-Hill, 1999.

[31] H. Iwai, S. Ohmi, H. Sasaki Momose, T. Ohguro, and Y. Katsumata, "Current status and future trend of microwave Si IC technologies," *Trans. IEICE C*, vol. J83-C, no. 10, pp. 911–919, Oct. 2000.

[32] A. Huang and B. Zhang, "The future of bipolar power transistors," *IEEE Trans. Electron Devices*, vol. 48, pp. 2535–2543, Nov. 2001.

[33] E. O. Johnson, "Physical limitation on frequency and power parameters of transistors," in *IEEE Int. Conv. Record*, 1965, p. 27.

[34] D. Harame, "Si/SiGe epitaxial-base transistors—Part I: Materials, physics, and circuits," *IEEE Trans. Electron Devices*, vol. 42, pp. 455–468, Mar. 1995.

[35] R. Jos, "Future developments and technology options in cellular phone power amplifiers: From power amplifier to integrated RF front-end module," in *Proc. 2000 Bipolar/BiCMOS Circuits and Technology Meeting*, Minneapolis, MN, pp. 118–125.

[36] A. Inoue, S. Nakatsuka, R. Hattori, and Y. Matsuda, "The maximum operating region in SiGe HBT's for RF power amplifiers," in *Proc. 2002 IEEE MTT-S Int. Microwave Symp.*, Seattle, WA, June 2002, pp. 1023–1026.

[37] H. Wong, "Drain breakdown in submicron MOSFETs: A review," *Microelectron. Reliabil.*, vol. 40, no. 1, pp. 3–15, Jan. 2000.

[38] A. Acovic, G. La Rosa, and Y. Sun, "A review of hot-carrier degradation mechanisms in MOSFETs," *Microelectron. Reliabil.*, vol. 36, no. 7–8, pp. 845–869, July–Aug. 1996.

[39] T. Sowlati and D. Leenaerts, "A 2.4 GHz 0.18μm CMOS self-biased cascode power amplifier with 23 dBm output power," in *Proc. IEEE ISSCC 2002*, pp. 108–109.

[40] M. van der Heijden, H. de Graaff, L. de Vreede, J. Gajadharsing, and J. Burghartz, "Theory and design of an ultra-linear square-law approximated LDMOS power amplifier in class-AB operation," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 2176–2184, Sept. 2002.

[41] G. Temes and J. LaPatra, *Introduction to Circuit Synthesis and Design*. New York: McGraw-Hill, 1977.

[42] L. Larson, "Integrated circuit technology options for RFIC's—Present status and future directions," *IEEE J. Solid-State Circuits*, vol. 33, pp. 387–399, Mar. 1998.

[43] J. Burghartz, D. Edelstein, M. Soyuer, H. Ainspan, and K. Jenkins, "RF circuit design aspects of spiral inductors on silicon," in *IEEE J. Solid-State Circuits*: IEEE, Dec. 1998, vol. 33, pp. 2028–2034.

[44] J. Burghartz, "Spiral inductors on silicon-status and trends," *Int. J. RF Microwave Computer-Aided Eng.*, vol. 8, no. 6, pp. 422–432, Nov. 1998.

[45] ——, "Status and trends of silicon RF technology," *Microelectron. Reliabil.*, vol. 41, no. 1, pp. 13–19, Jan. 2001.

[46] F. Rotella and J. Zachan, "Modeling and optimization of inductors with patterned ground shields for a high performance fully integrated switched tuning VCO," in *Proc. IEEE 2002 Custom Integrated Circuits Conf.*, May 2002, pp. 221–224.

[47] D. Y. C. Lie, X. Yuan, L. E. Larson, T. Robinson, A. Senior, X. Wang, J. Mecke, and M. Case, "Phase noise analysis of fully-integrated digitally-tuned wideband Si/SiGe BiCMOS VCOs," in *Proc. 2002 IEEE Bipolar Circuits and Technology Meeting*, pp. 128–131.

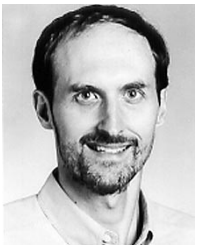
[48] J. Chang, A. Abidi, and M. Gaitan, "Large suspended inductors on silicon and their use in a 2 μm CMOS RF amplifier," *IEEE Electron Device Lett.*, vol. 14, pp. 246–248, May 1993.

[49] D. Laney, L. Larson, P. Chan, J. Malinowski, D. Harame, S. Subbanna, R. Volant, and M. Case, "Microwave transformers, inductors and transmission lines implemented in an Si/SiGe HBT process," *IEEE Trans. Microwave Theory Tech.*, vol. 49, pp. 1507–1510, Sept. 2000.

[50] J. Burghartz, M. Soyuer, K. Jenkins, M. Kies, M. Dolan, K. Stein, J. Malinowski, and D. Harame, "Integrated RF components in a SiGe bipolar technology," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1440–1445, Sept. 1997.

[51] D. Coolbaugh, E. Eshun, R. Groves, D. Hararnel, J. Johnson, A. Harnmad, Z. He, V. Ramachandran, K. Stein, S. St. Ongel, S. Subbanna, D. Wang, R. Volant, X. Wang, and K. Watson, "Advanced passive devices for enhanced integrated RF circuit performance," in *Tech. Dig. 2002 IEEE MTT-S Int. Microwave Symp.*, Seattle, WA, June 2002, pp. 187–191.

[52] H. Samavati, A. Hajimiri, A. Shahani, G. Nasserbakht, and T. Lee, "Fractal capacitors," *IEEE J. Solid-State Circuits*, vol. 33, pp. 2035–2041, Dec. 1998.
[53] R. Aparicio and A. Hajimiri, "Capacity limits and matching properties of lateral flux integrated capacitors," in *Proc. 2001 Custom Integrated Circuits Conf.*, San Diego, CA, May 2001, pp. 365–368.
[54] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[55] A. van der Ziel, *Noise in Solid State Devices and Circuits*. New York: Wiley, 1986.
[56] D. Pehlke, M. Schroter, A. Burstein, M. Matloubian, and M. Chang, "High-frequency application of MOS compact models and their development for scalable RF model libraries," in *Proc. IEEE 1998 Custom Integrated Circuits Conf.*, May 1998, pp. 219–222.
[57] J. Deen, C. Chen, and Y. Cheng, "MOSFET modeling for low-noise RF circuit design," in *Proc. IEEE 2002 Custom Integrated Circuits Conf.*, May 2002, pp. 201–205.
[58] D. K. Schaeffer and T. Lee, "A 1.5 V 1.5 GHz CMOS low-noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745–759, May 1997.
[59] R. Kraus and G. Knoblinger, "Modeling the gate-related high-frequency and noise characteristics of deep submicron MOSFETs," in *Proc. IEEE 2002 Custom Integrated Circuits Conf.*, May 2002, pp. 209–212.
[60] G. Niu, J. Cressler, Z. Shiming, W. Ansley, C. Webster, and D. Harame, "A unified approach to RF and microwave noise parameter modeling in bipolar transistors," *IEEE Trans. Electron Devices*, vol. 48, pp. 2568–2574, Nov. 2001.
[61] J. Cressler, "SiGe HBT technology: A new contender for Si-based RF and microwave circuit applications," *IEEE Trans. Microwave Theory Tech.*, pt. 2, vol. 46, pp. 572–589, May 1998.
[62] S. Voinigescu, S. Tarasewicz, T. MacElwee, and J. Ilowski, "An assessment of the state-of-the-art 0.5 μm bulk CMOS technology for RF applications," in *IEDM Tech. Dig.*, 1995, pp. 721–724.
[63] X. Li, T. Brogan, M. Esposito, B. Meyers, and K. O , "A comparison of CMOS and SiGe LNA's and mixers for wireless LAN application," in *Proc. IEEE 2001 Custom Integrated Circuits Conf.*, May 2001, pp. 531–534.
[64] S. Maas, *Nonlinear Microwave Circuits*. New York: IEEE Press, 1997.
[65] F. Klaassen and W. De Groot, "Modeling of scaled-down MOS transistors," *Solid-State Electron.*, vol. 23, no. 3, pp. 237–242, Mar. 1980.
[66] F. Klaassen, "A MOS model for computer-aided design,", Phillips Res. Rep., vol. 31, 1976.
[67] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*. Boston, MA: Kluwer Academic, 1998.
[68] R. Carson, *Radio Communications Concepts: Analog*. New York: Wiley Interscience, 1994.
[69] M. Vaidyanthan, L. Larson, M. Iwamoto, P. Asbeck, and P. Gudem, "A basic theory of high frequency distortion in bipolar transistors," IEEE Trans. Microwave Theory Tech., vol. 51, pp. 448–461, Feb. 2003, to be published.
[70] L. C. N. de Vreede, H. C. de Graaff, J. A. Willenmen, W. van Noort, H. F. F. Jos, L. E. Larson, J. W. Slotboom, and J. L. Tauritz, "Bipolar transistor epilayer design using the MAIDS mixed level simulator," *IEEE J. Solid-State Circuits*, vol. 34, pp. 1331–1338, Sept. 1999.
[71] G. Niu, Q. Liang, J. Cressler, C. Webster, and D. Harame, "RF linearity characteristics of SiGe HBT's," *IEEE Trans. Microwave Theory Tech.*, vol. 49, pp. 1558–1565, Sept. 2001.
[72] K. L. Fong, "High-frequency analysis of linearity improvement technique of common-emitter transconductance stages using a low-frequency-trap network," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1249–1271, Aug. 2000.
[73] J. Vuolevi and T. Rahkonen, "The effects of source impedance on the linearity of BJT common-emitter amplifiers," in *Proc. ISCAS 2000*, vol. VI, May 2000, pp. 197–200.
[74] V. Aparin and C. Persico, "Effect of out-of-band terminations on intermodulation distortion in common-emitter circuits," in *IEEE MTT-S Dig.*, vol. 3, June 1999, pp. 977–980.
[75] D. Leeson, "Simple model of a feedback oscillator noise spectrum," *Proc. IEEE*, vol. 54, Feb. 1956.
[76] G. Niu *et al.*, "Transistor Noise in SiGe HBT RF Technology," in *IEEE 2000 BCTM*, 2000, pp. 207–210.
[77] S. Vora and L. Larson, "Noise power optimization of monolithic CMOS VCOS," in *1999 Radio Frequency Integrated Circuit Symp. Dig.*, vol. 2, pp. 163–166.
[78] A. Hajimiri and T. H. Lee, "Design issues in CMOS differential LC oscillators," *IEEE J. Solid-State Circuits*, vol. 34, pp. 717–724, May 1999.
[79] T. Johansen and L. Larson, "Optimization of SiGe VCO's for wireless applications," in *Proc. 2002 Asia-Pacific Microwave Conf.*, Kyoto, Japan.
[80] C. Samori, A. L. Lacaita, F. Villa, and F. Zappa, "Spectrum folding and phase noise in LC tuned oscillators," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 781–790, July 1998.
[81] E. Hegazi, H. Sjoland, and A. Abidi, "A filtering technique to lower LC oscillator phase noise," *IEEE J. Solid-State Circuits*, vol. 36, pp. 1921–1930, Dec. 2001.
[82] B. Buck, M. Green, A. Dearn, A. Long, S. Melvin, and J. Clifton, "GaAs MMIC's for 5.2 GHz HIPERLAN," in *Proc. 5th Annu. Wireless Symp.*, Feb. 1997, pp. 62–66.
[83] B. Jansen, K. Negus, and D. Lee, "Silicon bipolar VCO family for 1.1 to 2.2 GHz with fully integrated tank and tuning circuits," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 1997, pp. 392–393.
[84] P. Baltus, A. Wagermans, R. Dekker, A. Hoogstrade, H. Maas, A. Tombeur, and J. Sinderen, "A 3.5 mW 2.5 GHz diversity receiver and a 1.2 mW 3.6 GHz VCO in silicon-on-anything," *IEEE J. Solid-State Circuits*, vol. 33, pp. 2074–2080, Dec. 1998.
[85] P. Kinget, "A fully integrated 2.7 V 0.35 μm CMOS VCO for 5 GHz wireless applications," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 1998, pp. 226–227.
[86] J. Plouchart, "Design of HBT MMIC's for PLL applications," Ph.D. dissertation, Univ. of Paris, France, May 1994.
[87] H. Wong, "A 9.8 GHz back-gate tuned VCO in 0.35 μm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 1999, pp. 406–407.
[88] J. Plouchart, B. Klesper, H. Ainspan, and M. Soyeur, "Fully monolithic 3-V SiGe differential voltage-controlled oscillators for 5 GHz and 17 GHz wireless applications," in *Proc. Eur. Solid-State Circuits Conf.*, Sept. 1998, pp. 332–335.
[89] F. Herzel, M. Pierschel, P. Weger, and M. Tiebout, "Phase noise in a differential CMOS voltage-controlled oscillator in RF applications," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 11–15, Jan. 2000.
[90] C. Samori, A. Zanchi, S. Levantino, and A. Lacaita, "A fully integrated low-power low-noise 2.6 GHz bipolar VCO for wireless applications," *IEEE Microwave Wireless Components Lett.*, vol. 11, pp. 199–201, May 2001.
[91] J. Plouchart, H. Ainspan, M. Soyeur, and A. Ruehli, "A fully monolithic SiGe differential voltage-controlled oscillators for 5 GHz wireless applications," in *Proc. 2000 IEEE Radio Frequency Integrated Circuits Symp.*, June 2000, pp. 57–59.



Lawrence E. Larson (M'82–SM'90–F'00) received the B.S. degree in electrical engineering and the M.Eng. degree from Cornell University, Ithaca, NY, in 1979 and 1980, respectively, and the Ph.D. degree in electrical engineering from the University of California, Los Angeles, in 1986.

He joined Hughes Research Laboratories, Malibu, CA, in 1980, where he directed work on high-frequency InP, GaAs, and silicon-integrated circuit development for a variety of radar and communications applications. While at Hughes, he led the team that developed some of the first MEMS-based circuits for RF and microwave applications, the first InP HEMT MMIC foundry, and developed some of the first MMICs in Si/SiGe HBT technology with IBM. He was also an Assistant Program Manager of the Hughes/DARPA MIMIC Program from 1992 to 1994. From 1994 to 1996, he was at Hughes Network Systems, Germantown, MD, where he directed the development of RF integrated circuits for wireless communications applications. He joined the faculty at the University of California of San Diego (UCSD), La Jolla, in 1996, where he is the inaugural holder of the Communications Industry Chair and Director of the UCSD Center for Wireless Communications at the Jacobs School of Engineering. While on academic leave in 2000–2001, he was Director of the IBM West Coast Design Center of Excellence where he led the development of RF integrated circuits for 3G applications. He has published over 150 papers and has received 24 U.S. patents.

Dr. Larson was co-recipient of the 1996 Lawrence A. Hyland Patent Award of Hughes Electronics for his work on low-noise millimeter-wave HEMTs, the HRL Sector Patent Award for his work on RF MEMS technology, and the IBM Microelectronics Excellence Award for his work on Si/SiGe HBT technology. He is a member of Eta Kappa Nu and Sigma Xi.