SiGe Radio Frequency ICs for Low-Power Portable Communication

JOHN R. LONG

Invited Paper

The range and impact of SiGe bipolar and BiCMOS technologies on wireless transceivers for portable telephony and data communications are surveyed. SiGe technology enables transceiver designs that compare favorably with competing technologies such as RF CMOS or III–Vs, with advantages in design cycle time and performance versus cost. As wireless devices continue to increase in complexity using conventional battery technology as the power source, the desire to reduce current consumption in future transceivers continues to favor SiGe technology. Examples are drawn from contemporary wireless communications ICs. The performance of on-chip passive components in silicon technologies are also reviewed in this paper. Greater understanding of the limitations of passive devices coupled with improved models for their performance are leading to circuits offering wider RF dynamic range at ever higher operating frequencies. The innovations in on-chip passive design and construction currently being pioneered in mixed-signal SiGe technologies are enabling circuits operating deep into millimeter-wave frequency bands (i.e., well above 30 GHz). In addition, sophisticated on-chip magnetic components combined with deep submicrometer SiGe active devices in a transceiver front end are envisioned that enable single-volt SiGe circuits, with even lower current consumption than is achievable today. Relevant examples from the recent literature are presented.

Keywords*—Coplanar waveguide, injection locking, integrated circuits (ICs), low-noise amplifier (LNA), microstrip, millimeter wave, mixer, monolithic inductor, radio frequency, slow wave, transformer, transmission line, voltage-controlled oscillator (VCO), wireless communication transceivers.*

I. Introduction

Marconi's demonstration that electromagnetic waves could transmit information across vast distances without the aid of wires precipitated a revolution in communication technology. During the century since radio technology was first used, mobile communication has evolved from spark-gap telegraphy to today's third-generation (3-G) mobile handsets equipped with cameras, MP3 audio players and Internet access. These devices must meet aggressive RF performance specifications in a small-sized, low-cost solution that consumes little power, while operating at data rates up to 2 Mb/s (and beyond) as new applications are added to wireless portables. The growing legacy of existing standards which continue to be supported by these handsets increases hardware complexity and power consumption, while giving little additional utility to the user (e.g., multiband and multistandard telephony). Advanced IC technologies such as SiGe-BiCMOS are addressing the challenge of designing the third and future generation transceivers, which are destined for the mass consumer market.

In the early 1990s, silicon transistors—both bipolar and field-effect—crossed the 10-GHz unity-gain frontier. This led advocates of silicon IC technology to propose it as a platform for radio and high-speed applications [1]. Despite initial skepticism from many RF designers, silicon technology rapidly displaced others in the RF sections of a mobile telephone by the end of the decade. Innovations in circuit and system design also played a major role in this. Direct conversion or homodyning reduced the number of passive interstage filters at the cost of more elaborate circuit implementations and greater complexity in the baseband circuits [2], [3]. Also, system standards geared to silicon IC technology, such as Bluetooth, played to silicon's strengths and attraction as a low-cost technology for high-volume applications. Aggressive scaling of transistor dimensions from 1.0 μm to 90 nm has moved the unity-gain frequency, or f_T, up to 120 GHz [4]. These faster transistors give RF circuit designers even more freedom to innovate. While far from being the perfect base for microwave frequency circuit development, previous experience has shown that silicon often plays a leading role because of its inherent economies of scale. It will continue to dominate RF IC technology for the foreseeable future.

Today's handsets are smaller and lighter, cost much less to manufacture, and run longer on a fresh battery charge.

Manuscript received August 21, 2004; revised March 20, 2005.

J. R. Long is with the Electronics Research Laboratory/DIMES, Faculty of EEMCS, Delft University of Technology, Delft 2628CD, The Netherlands (e-mail: j.r.long@ewi.tudelft.nl).

Digital Object Identifier 10.1109/JPROC.2005.852227

Many of the technologies seen in the earlier handsets are still there, but they are scaled down in size or assembled using less expensive and more efficient packaging techniques. Silicon ICs, which were extensively used for baseband and IF signal processing in early sets, have now migrated toward the antenna. Depending upon the system requirements, almost all of the RF transceiver functions can now be implemented on a single chip (e.g., GSM). While cost or performance constraints do not always favor a fully monolithic approach, the general trend toward greater integration driven by cost, size, and battery life has been steady and relatively rapid over the past decade.

The application of silicon–germanium (SiGe) bipolar and SiGe-BiCMOS technologies to the front end of wireless transmitters and receivers for portable telephony and data communications are surveyed in this paper. It will be seen that SiGe technology enables transceiver designs that compare favorably with competing technologies such as RF CMOS or III–Vs, but with advantages in design cycle time and performance versus cost. As wireless devices continue to increase in complexity using conventional battery technology as the power source, the requirement to further reduce current consumption in future transceivers continue to favor advanced technologies such as SiGe. Examples are given which are drawn from contemporary wireless communication ICs.

A. Growing Demand for Wireless Systems

Wireless communications systems, of which cordless telephones, pagers, and mobile cellular telephones are familiar and successful examples, have grown enormously in commercial significance over the past decade. Wireless office information networks are now being selected by business and manufacturing concerns that want to avoid the high cost of reconfiguring voice and/or data networks. Data rates for these applications vary from the low kilobits per second range to over 50 Mb/s for LANs using the 802.11a/g standard. Wireless networks can also provide access to information in places where wires cannot go or are prohibitively expensive to install. In a commercial office building, for example, wireless networks are easily augmented without moving existing wires or installing new ones. Furthermore, users can access shared information anywhere in the local area and remain free to roam, since they are no longer tied to a physical network access point. In addition, wireless systems can adapt to a large and frequently varying number of users with modern multiple-access techniques. They can be organized in an *ad hoc* (i.e., peer-to-peer) topology for a few users, or in a cellular topology to accommodate thousands of nodes over a larger area without replacement of existing hardware at the access points. The untethering of computing devices may even overcome the problem of the narrow-band "last mile" in wireline telephone systems, replacing a bottleneck that has delayed widespread access to broad-band networks with broad-band data capability.

As a result of these factors and others, wireless is projected to surpass wireline as the principal method of communication worldwide by 2008 [5]. Low-cost radio transceiver parts that can be manufactured in large volumes and that consume little power are required for these new wireless applications. This is a natural market for highly integrated monolithic circuits at RF and microwave frequencies.

B. RF IC Design

Analog RF circuits demand much more than fast switching speeds between binary states and the capability to pack an enormous number of devices onto a single chip. For example, the interface between the RF channel and the baseband digital signal processor in a cellular telephone must detect microvolt signal levels, while keeping harmonic and intermodulation distortion produced by much stronger interferers within acceptable limits. Also, tetherless communications devices such as cellular telephones and wireless LAN interfaces are light and portable, which makes factors such as size and battery lifetime important to the RF circuit designer.

There are many aspects of circuit design at radio frequencies that differ from circuit design for lower frequency applications. Some are related to frequency, such as operating a transistor or circuit close to its bandwidth limit, or the difficulties of testing and measuring circuits and components in the gigahertz frequency range. In addition, many of the specifications and terminology used to describe RF and high-speed circuits are unique to the field. Many simulators are designed specifically to solve RF design problems, so designers must be familiar with the various simulation techniques and design methodologies—especially when verifying operation across system interfaces (e.g., antenna to baseband in a transceiver).

Many circuits might not appear to involve "high frequencies" or "high operating speeds," but these terms are related to the bandwidth limitations of the active devices used in the design. The frequency range where instability occurs in most circuits lies between the -3-dB bandwidth $f_{\rm bw}$ and the unity-gain frequency $f_{\rm unity\text{-}gain}$. For an SiGe heterojunction bipolar transistor (HBT) with a current gain β of 100 at kHz frequencies and a unity-gain–bandwidth product (f_T) of 100 GHz, the *high-frequency* range lies above f_T/β, or above about 1 GHz. *High speed* implies a device or circuit with an operating range extending into the high-frequency regime, where the maximum frequency of operation typically lies in the range from 0.1 to 0.5 f_T.

As transistor f_T in production technologies moves beyond 100 GHz, it is clear that the interconnect bandwidth and passive component parasitics are not scaling as rapidly as transistor performance. As a result, improvements in RF circuit performance that could capitalize on advances from Moore's Law (e.g., a shorter MOSFET gate length) continue to be constrained by the environment surrounding nanometer-sized active devices. This includes on-chip interconnects, imperfections in passive components that process the voltages and currents in an analog circuit, packaging parasitics (e.g., bondwires), as well as the printed circuit or other component subassemblies. Despite these problems, there are components available at RF that are not realizable at lower frequencies as illustrated by the chip micrograph

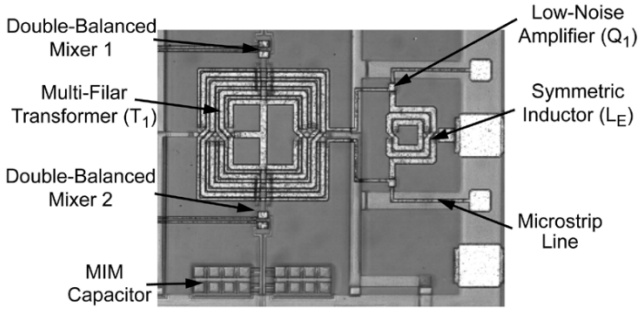


Fig. 1. 5.1–5.8-GHz single-sideband receiver front end built around a monolithic trifilar transformer [6].

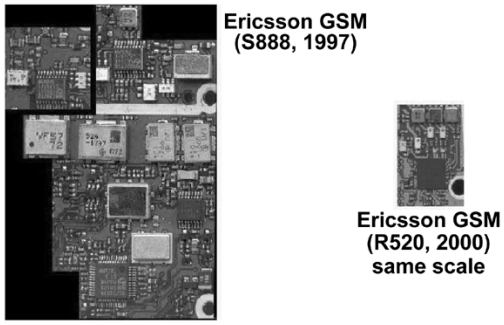


Fig. 2. Evolution of RF integration in GSM mobile phone handsets (courtesy Ericsson).

of Fig. 1 [6]. In this circuit, a symmetric inductor (synthesized from on-chip transmission lines) improves input impedance matching (i.e., maximizes power transfer from an antenna) and also minimizes the noise figure of a differential low-noise amplifier (LNA). The microstrip line interconnection between the bondpad and LNA form part of the matching network. Interstage signal coupling and bias isolation between the LNA and two balanced mixers are implemented using an on-chip three filament (trifilar) transformer. These magnetic components (i.e., inductor and transformer) enable sub-1-V operation with reduced current consumption and wider dynamic range (lower noise and distortion) at the cost of additional circuit area. Superior device matching and the potential for large-scale integration on silicon chips enables system architectures using single-sideband converters (e.g., as in Fig. 1) with baseband signal processing and self-calibration schemes. This ultimately reduces the overall component cost and size while also improving system reliability and manufacturability.

II. Integrated Wireless Transceivers

As a result of increased functional integration, the number of components required to implement the RF front end in a GSM-standard handset has decreased steadily from many hundreds to well below 50 components today. The Ericsson S888 and R520 handset RF electronics shown in Fig. 2 are an example of this evolution. The RF sections of the dual-band S888 from 1997 require 2.5 mm^2 of circuit board area and consist of 310 parts, of which five are application-specific ICs (ASICs) operating from a 3.6-V supply. Many of the surface mount resistors, capacitors, ICs, and RF subassemblies from the S888 are replaced by smaller outline filters, a handful of surface mount passives, and a BiCMOS ASIC in the triband R520 introduced in 2000 [Fig. 2(b)]. The RF board area shrinks to 1 mm^2, and only 90 RF parts including two ASICs operating at 2.7 V are used to implement the GSM phone. This trend toward denser integration in transceivers using silicon IC technology is continuing, and some of today's handsets operate not only in multiple frequency bands within one standard (e.g., GSM at 900, 1800, and 1900 MHz), but also together with 3-G (e.g., WCDMA or CDMA2000) standards in multistandard/multiband handsets [7]–[9].

Continuing this trend toward fewer RF parts while at the same time increasing the performance and functionality of a portable handset by using IC technology alone—as in the previous example—may prove difficult. Although integration of complete single-chip transceivers [i.e., an RF system on a chip (RF-SoC)] for wireless standards such as Bluetooth has been demonstrated, RF-SoC designs are difficult to shrink between technology nodes because analog circuits do not scale as readily in chip area and supply voltage as digital ICs. Analog passive components such as capacitors and inductors do not scale with feature size (note that a typical on-chip inductor consumes 10–100 times more area than a transistor). Also, lowering the supply voltage as demanded by scaling [10] reduces the dynamic range of most RF circuits. Reliability of thin-oxide devices and poor on-chip device matching are also considerations when analog designers use state-of-the-art nanometer-scale devices for RF-SoC applications. An RF system in a package (RF-SiP) is becoming an attractive alternative to the RF-SoC, because it avoids some of these difficulties.

Multiple technologies are integrated in an SiP, which can be just a scaled version of the RF system or subsystem currently being integrated on a printed circuit board. However, a multitechnology design integrated into the same package as an SiP does not benefit from CMOS economies of scale, unless batch processing is used extensively to manufacture and package individual components. However, an RF-SiP is more flexible, in that the most effective technology for a given task can be selected (e.g., FET RF switch, passive filter, IC downconverter, etc.), thereby lowering the risk of making an inappropriate technology selection. It also makes it possible to scale and port only part of the system between technology nodes (e.g., aggressively scale the baseband DSP and memory) while keeping the analog designs unchanged.

The choice of SiGe-BiCMOS or RF-CMOS [either bulk or silicon-on-insulator (SOI)] depends upon the application and performance requirements. SiGe-BiCMOS is suited to applications which demand the highest level of RF performance (e.g., wide dynamic range with low current consumption) but do not require the lowest cost or the highest level of digital integration. Bulk RF-CMOS has a 20%–25% cost advantage over SiGe-BiCMOS and has the advantage of integration density for digital blocks because it is at the leading edge of silicon technology scaling. This makes it the logical choice for commodity RFIC applications such as Bluetooth and 802.11 WLAN. However, SiGe-BiCMOS

Table 1
Wireless System Evolution

Generation	Standard	Frequency Band, in MHz	Channel Spacing, in MHz	Modulation	Data Rate
2	GSM	900	0.2	GMSK	9.6kbps
	DCS	1800			
	PCS	1900			
	CDMA	900	1.25	OQPSK	14.4kbps
2.5	GPRS	shared with GSM/DCS	0.2	GMSK (multislot)	115kbps
	EDGE			8-PSK	384kbps
	CDMA IS-95B	900/1900			114kbps
WLAN	802.11b	2400		DSSS	11Mbps
	802.11a	5300	20	OFDM	54Mbps
PAN	Bluetooth	2400		GFSK	1Mbps
3	W-CDMA	1900-2200	5	HPSK	384k/2Mbps
3.5	W-CDMA+	-			2Mbps
	Packet Data	-			10Mbps
4	MC-CDMA	-			100Mbps

technology could be an appropriate choice in either RF-SoC or RF-SiP scenarios where digital integration density on the same chip is not a top priority, and cost can be traded off for higher performance (i.e., lower power consumption or improve RF specifications).

A. Next-Generation Handsets

The path toward a fourth-generation (4-G) multimedia portable handset passes through the current 3-G standards. Third-generation handsets incorporate wide-band CDMA, or WCDMA (which can be either UMTS or CDMA2000 standards) as listed in Table 1, in various segments of the frequency spectrum from 1.9 to 2.2 GHz. The channel bandwidth of 5 MHz allows data rates up to 2 Mb/s, ramping up from the 114 kb/s or 384 kb/s possible with 2.5-G packet data systems. Growing demand for high-speed data services will necessitate further increases in data rates (perhaps to 10 Mb/s) for a 3.5-G handset as a stepping stone to the full multimedia capabilities and 100 Mb/s data rates supported by the 4-G phone of the future.

The general trend toward higher data rates implies a drop in the receiver sensitivity (consequently lowering the maximum distance between handset and base station) due to the greater receiver bandwidth required at higher data rates. The receiver linearity must also improve to compensate for increased interference as the number of users within a cell site rises. Improving the receiver's dynamic range requires a high-performance radio architecture, circuit technology, and passives components. This favors more flexible implementations incorporating multiple technologies rather than highly integrated SoC implementations that usually compromise system performance. It also favors high-performance IC technologies like SiGe-BiCMOS.

A block diagram for a 3-G portable is shown in Fig. 3. It must support multiple standards, including operation in multiple frequency bands and handover from one cell to another. Frequency bands range from approximately 800 to 2200 MHz for today's GSM and WCDMA standards. The WCDMA radio front end transmits and receives continuously in different bands (e.g., in Europe: 2110–2170 MHz Rx and 1920–1980 MHz Tx) with data transmission rates of up to 2 Mb/s. Since a frequency rather than time-duplexing scheme is used to separate the WCDMA transmit and receive signals, the transceiver is operating continuously. GSM is a legacy standard from 2-G digital mobile telephony. The transmitter and receiver operate at different times (i.e., time duplexed using a time-division multiple access protocol). GSM is approved for use in frequency bands ranging from approximately 824 to 1990 MHz, and worldwide coverage requires a transceiver that can operate in four frequency bands. A 3-G handset supports applications other than telephony, so high-speed data transfer (for Internet access, video streaming or up/downloading of images) is facilitated by a WLAN or PAN standard [either 802.11 or Bluetooth (BT)]. A global positioning system (GPS) receiver at 1575 MHz locates and tracks the portable for navigation, billing, etc. It is clear that much of the digital signal processing can be shared using a programmable DSP engine. However, direct implementation of the analog RF interfaces requires at least four (and possibly more) radio transceivers; one each for GSM, WCDMA, 802.11, and GPS standards. This is illustrated for GSM and WCDMA in the block diagram

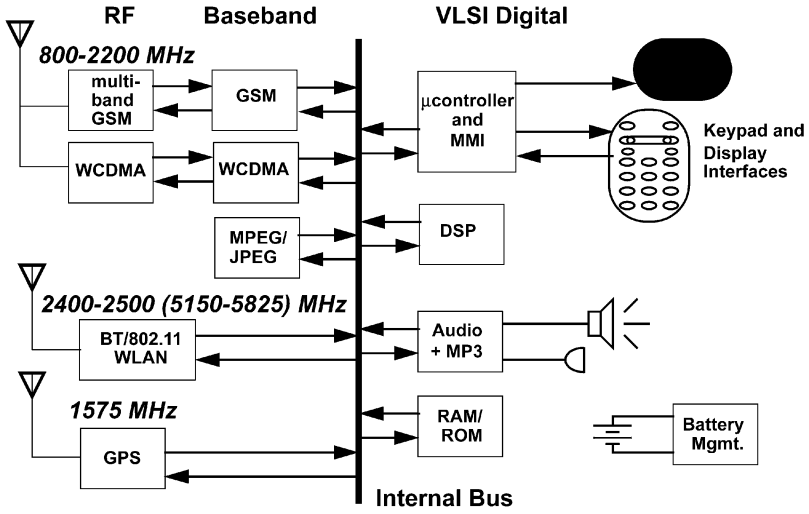


Fig. 3. Block diagram of a 3-G multistandard/multiband portable handset.

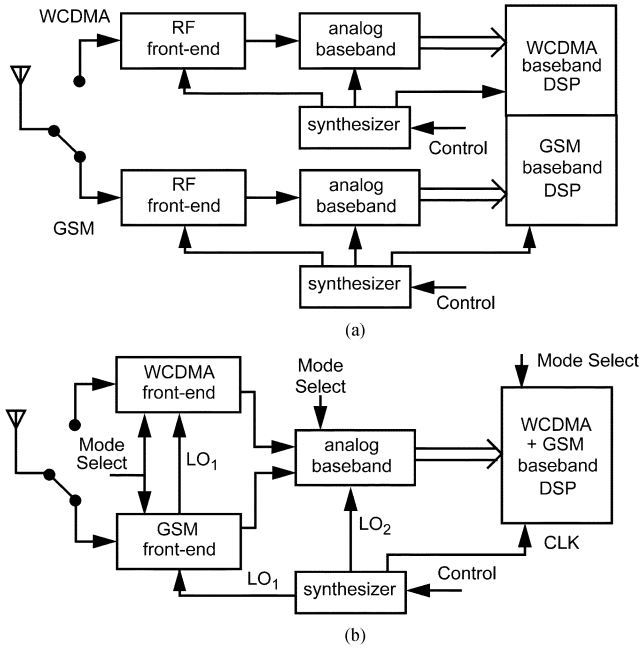


Fig. 4. Multistandard/multiband receiver integration. (a) Parallel receiver paths. (b) Integrated adaptive receiver.

of Fig. 4(a). Duplicating the radio paths (i.e., one path for each standard) leverages existing designs, but the resulting handset is bulky, relatively complex, and expensive to manufacture compared to 2-G or 2.5-G designs.

Consumers expect modern mobile telephone handsets to be small, light, and inexpensive, and to operate for many hours on a single battery charge with reasonable quality of service. Integrating more RF standards using the same number of components as older handsets requires a transceiver architecture that can exploit the economies of scale in manufacturing, and miniaturization offered by monolithic integration. Passive filters and other passive components (e.g., switches, matching elements, bandselect filters, diplexers, etc.) must also be integrated and packaged cheaply, preferably using a batch process similar to that used to manufacture ICs. Therefore, selection of the appropriate technology for each function in the transceiver is more important now than it was for previous generation handsets. An inappropriate technology choice could compromise performance, increase cost, delay production, and affect future scaling and cost reductions. Consequently, the risks involved in technology selection are driving many manufacturers toward conservative technology choices—favoring SiP designs—while at the same time cost and complexity constraints are forcing them to innovate in the transceiver architecture to share resources.

Power consumption in a portable wireless transceiver is dominated by the transmit power amplifier. Therefore, low power consumption requires a very efficient RF power amp, likely implemented in a specialized technology and built as a separate component or module in order to dissipate heat efficiently.

Managing power consumption and complexity are two of the challenges facing 3-G handset designers. Unfortunately, most of the radio front ends in a 3-G portable operate concurrently because one (or more) of the applications may be active at any time. Note that this is not always under the control of the user. For example, a telephone service subscriber will want to initiate or receive a call over either GSM and WCDMA systems, while possibly browsing the Internet using an 802.11 LAN connection. For WCDMA, the handset and base station are in continuous contact with each other in order to control key parameters (e.g., the RF output power level), and as a result power consumption during a GSM phone call cannot be conserved by simply turning the WCDMA transceiver "off." Therefore, unless the radio architecture is carefully crafted to manage power consumption (e.g., by sharing resources as in Fig. 4(b), the power consumed by the handset can grow almost linearly with the number of standards implemented.

The need to support many standards operating in numerous frequency bands implies that radios should share resources wherever possible (i.e., at RF as well as baseband). This would also help to scale down power consumption, reduce the number of parts needed to build a handset, and lower costs. An illustration of such an integrated receiver is shown in Fig. 5. The RF front end uses separate LNAs,

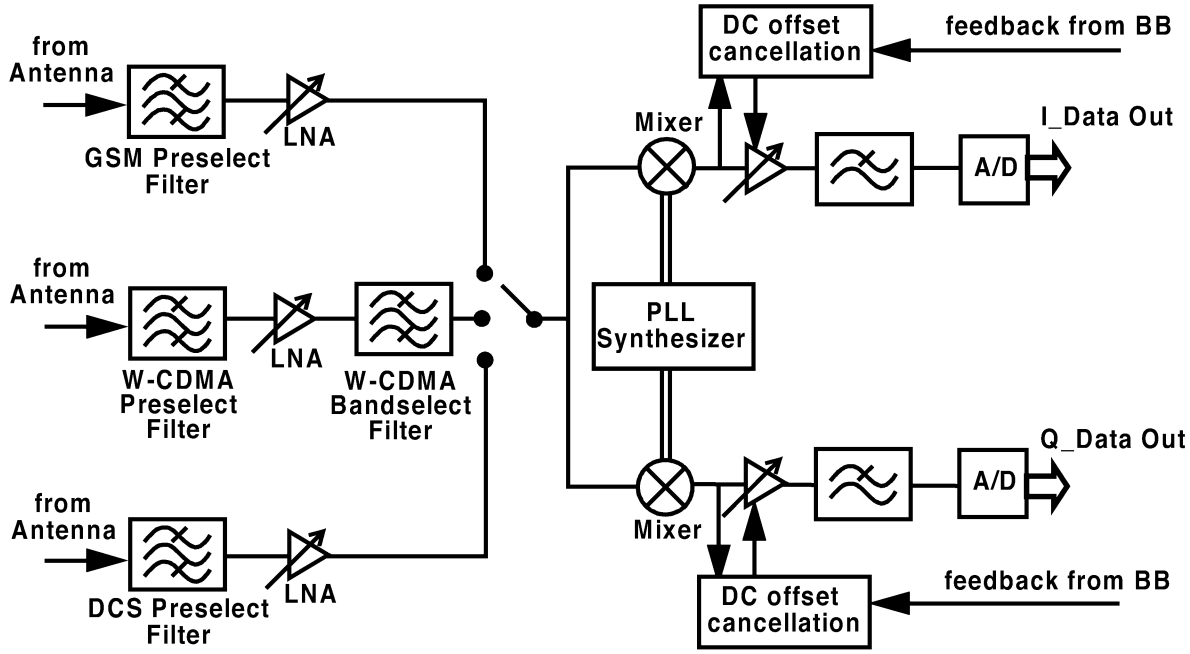


Fig. 5. Reconfigurable 3G direct-conversion receiver example.

band-select filters, etc., as required for each frequency band and standard. However, the downconversion and analog baseband functions are shared and programmed on the fly for each standard supported by the transceiver. Control and adjustment/calibration of components in the receive chain is periodically required. To optimize performance, the receiver electronics are adaptive, in that the dynamic range and gain are optimized for each standard by adjusting circuit operating parameters to minimize power consumption. Similar arguments could also be made in favor of an adaptive transmitter architecture to reduce power consumption, cost, and the overall parts count.

A direct conversion receiver is shown as the example in Fig. 5 because it requires less baseband hardware and is one of the most amenable to monolithic integration. However, implementation issues such as the correction of static and dynamic dc offsets at baseband caused by leakage and self-mixing of the local oscillator [11] must be considered.

SiGe-BiCMOS technology is well-suited to the implementation of the 3-G transceiver RF front end, because it offers the opportunity to optimize power consumption, analog performance, and technology selection (e.g., using bipolar and/or MOS devices, or other passive components). The level of circuit integration is not as high as might be achieved by using bulk CMOS, but it is likely that the risks inherent in designing and maintaining or evolving a 3-G handset design favor a solution that partitions the transceiver into multiple ICs (e.g., digital baseband and analog RF) that are implemented in different technologies.

III. Monolithic Building Blocks for RFIC Design

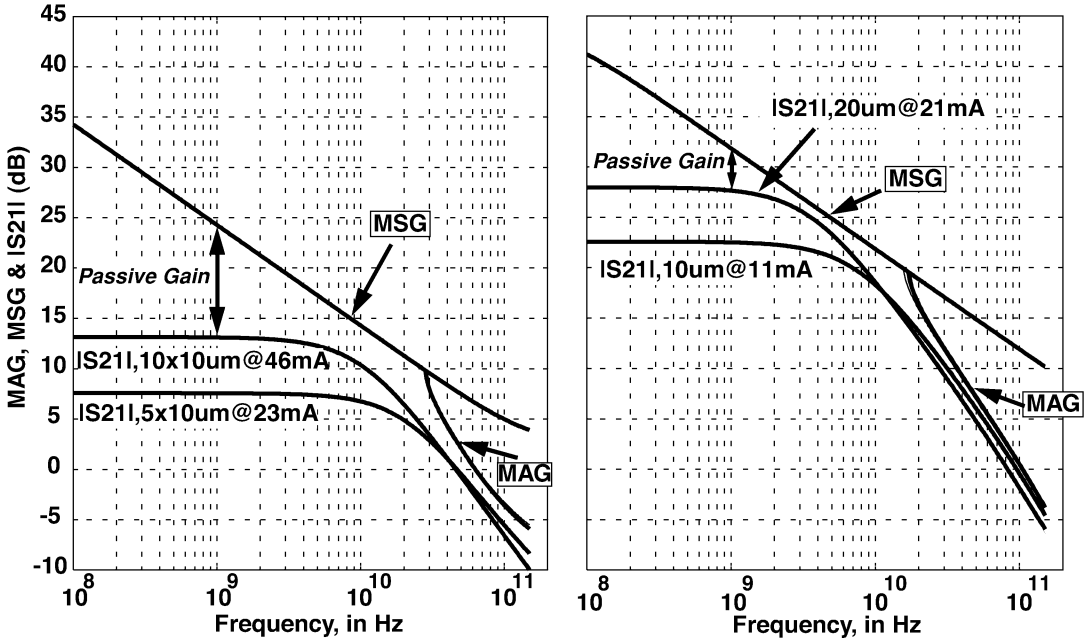
Building blocks for wireless applications (e.g., LNA, mixer, voltage-controlled oscillator (VCO), etc.) as shown in Figs. 4 and 5 have stringent requirements on sensitivity, distortion, bandwidth, spectral purity, and power output. Circuit performance can be optimized for a given task when a wide variety of on-chip components are available for designers to choose from. Many SiGe-BiCMOS technologies offer processing options (e.g., thicker interconnect metal, linear capacitor, polysilicon resistor, etc.) that improve analog/RF performance beyond what is available from a high-volume technology intended solely for digital applications.

Table 2
Silicon MOSFET and BJT Comparison (0.18 μm)

Parameter	MOS	BJT
I/gm	$V_{eff}/2$	V_T
Cin/Cμ	3	10
1/f corner	1 MHz	1 kHz
ro @ 5mA bias	1 kΩ	10 kΩ
Breakdown Voltage	1.8 V	2-5 V

A. Bipolar and CMOS Transistors

The relevant RF performance attributes of 0.18-μm silicon MOS and bipolar (BJT) devices are compared in Table 2. The ratio of bias current to transconductance (I/gm) is an indication of the gain that can be realized at a given supply current (the smaller the ratio the better). For a MOSFET, this depends on the effective gate-source voltage (V_{GS}-V_{th}), which is on the order of 200–300 mV, compared to the thermal voltage ($V_T = kT/q$, approximately 25 mV at 27 °C) for the bipolar. The factor of 8–12 in gives a substantial advantage in gain for a given bias current to the bipolar transistor when used in a typical amplifier circuit. When used in a negative feedback circuit, the larger loop gain available from a bjt gain stage can be traded off for other performance benefits (e.g., linearity, stability, controlled terminal impedances, etc.).



(a) (b)

Fig. 6. Gain versus frequency for 10- and 20-μm emitter length BJTs compared to MOS devices with 50- and 100-μm total gate widths. (a) 0.18-μm nMOS biased at peak f_T. (b) 0.2-μm BJT biased at peak f_T.

The impedance levels in RF and high-speed circuits are usually kept low because of bandwidth restrictions, so transconductance defines the "active" gain at RF (note that voltage or current gain can also arise from impedance matching; this is a "passive" gain). The product of the transconductance and output resistance $(\mathrm{gm}^{*}\mathrm{ro})$ sets the maximum active gain at low frequency (i.e., up to a few hundred kilohertz). Gain at RF also depends upon transistor parasitics. The ratio of input to Miller capacitance $(C_{\mathrm{in}}/C_{\mu})$ in common-source (or common-emitter) configuration indicates the relative importance of the parasitic capacitances, which affect bandwidth. As seen from the data in Table 2, bipolar devices have a clear advantage in both gain and bandwidth. In addition, the input impedance of a BJT at RF is relatively low compared to a MOSFET. This is a subtle advantage of the bipolar device when used in an RF circuit like an LNA. As shown in Fig. 6, MOS devices can realize much more passive (i.e., voltage) gain at the input if a matching network is used to match the transistor input impedance to an RF source (e.g., a 50-Ω antenna). However, low loss matching elements are needed. This gives a narrow-band frequency response that is more susceptible to production tolerances and usually requires trimming in manufacture. Note that losses affect the quality of a passive component, and reduce its figure of merit (FOM) or Q-factor.

A bipolar transistor, on the other hand, has a relatively low input impedance making the transistor easier to match to the typical RF source off-chip using a simple, low-Q passive network. This is illustrated in Fig. 6, where both BJT and MOSFETs are biased at maximum f_T. The forward transmission coefficient $|S_{21}|$ is proportional to the voltage gain of the common-emitter BJT or common-source MOSFET amplifier when driven by a 50-Ω source and driving a 50-Ω load. The maximum stable gain (MSG) is realized when both input and output are perfectly matched to a 50-Ω source and load, so the difference between $|S_{21}|$ and MSG is the passive gain of the matching network. From the plots shown in Fig. 6, it is clear that higher gain is produced by the BJT with less bias current (e.g., $|S_{21}|$ of 22 dB at 1 GHz for a 10-μm-long BJT at 11 mA versus 7 dB for a 50-μm-wide FET biased at 23 mA). Also, the difference between MSG and $|S_{21}|$ (i.e., the need for a matching network) is much smaller for a BJT amplifier compared to the MOSFET stages (e.g., a 5-dB difference for the 20-μm-long BJT compared to 11 dB for a 50-μm-wide FET at 1 GHz).

Taking an example from the recent literature, a two-stage 60-GHz BJT amplifier fabricated and tested in a 0.12-μm SiGe technology realized 17-dB gain while drawing 6 mA from a 1.8-V supply [16]. A design study of a three-stage 60-GHz amplifier in a 0.13-μm CMOS technology predicts an 11-dB gain and draws 36 mA from a 1.5-V supply (simulated, [17])—a sixfold increase in current consumption for 6 dB less gain using the MOSFET.

The impact of these device metrics on RF circuits are now examined in more detail.

B. Power Consumption

Modern silicon devices in production have f_T exceeding 100 GHz (refer to Fig. 7), which is well above the operating frequencies for most of today's commercial wireless products (e.g., mobile telephones or wireless computer networking equipment). Therefore, some of this bandwidth can be traded off for reduced power consumption in circuits such as the LNA. Gain is not severely affected, because enough transconductance (gm) can be realized from a BJT with a relatively low I/gm ratio even at low bias currents. The LNA

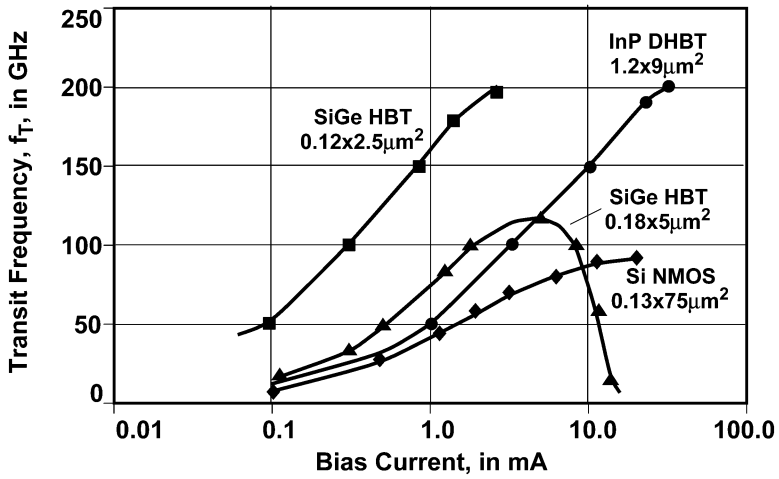


Fig. 7. Gain–bandwidth product (f_T) versus bias current for transistors fabricated using various semiconductor technologies (year 2003).

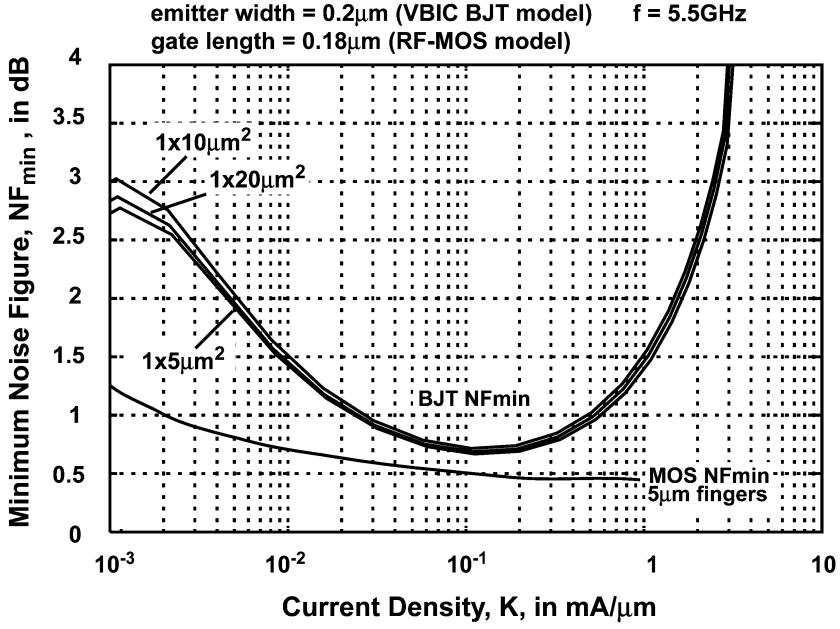


Fig. 8. NF_{min} versus current density for various emitter lengths at $V_{CE} = 1.5$ V; MOS devices with 50- and 100-μm gate widths at $V_{DS} = 1.8$ V, $f = 5.5$ GHz.

suppresses noise introduced by the mixer and subsequent stages in the receive chain. The other design objectives for the preamplifier in a radio receiver are adequate gain, low intermodulation distortion production, and an input impedance which matches the antenna impedance (50 Ω is usually assumed). In addition, the BJT collector-base feedback capacitance (which causes the Miller effect) in a typical common-emitter preamplifier is low enough that a single transistor can be used to realize a gain of 15–20 dB with a bias current of just a few milliamperes. This is a consequence of the relatively high Cin/Cμ ratio indicated in Table 2. When on-chip resonant circuits are used as loads, the transistor can be biased from a supply as low as one base-emitter voltage drop (i.e., V_BE—approximately 0.9 V), thereby realizing a true low-voltage, low-power LNA with excellent RF performance. By comparison, a MOSFET LNA typically uses the cascode topology (to suppress the Miller effect) and requires more bias current (and possibly a higher supply voltage) in order to realize the same transconductance and gain as a BJT equivalent.

C. Noise

Thermal and shot noise generated by active devices degrades the SNR and noise figure (a measure of the SNR degradation, expressed in decibels) of LNAs and mixers. The minimum noise figure (NFmin) occurs when a unique source impedance minimizes the noise added to the signal by a transistor (or amplifier/mixer) compared with the noise inherent in the source impedance itself. NFmin is both frequency and bias current dependent. Example plots of the minimum noise figure for MOSFET and BJT devices in a 0.18-μm SiGe-BiCMOS technology are plotted (from simulation) as a function of current density (i.e., per millimeter of emitter length or per millimeter of gate width) in Fig. 8.

The minimum noise figure for a MOSFET continually decreases with increasing current density (and hence with bias current) and reaches a minimum at the border of the triode/saturation region (i.e., the highest bias current that can flow through the transistor while still in pinchoff). Consequently, low noise for a MOSFET amplifier implies operation at the highest possible current density—which is also desired in order to realize enough RF gain. This has further consequences for power consumption (with relatively large FET areas) or sensitivity of the input to impedance mismatch (for small area transistors) in a low-noise amplifier or mixer constructed using MOS devices.

The flat portion of the noise figure curve for the BJT (i.e., below 0.3 mA/μm) corresponds to the region where the noise figure is dominated by the extrinsic base resistance of the transistor. As the emitter area of the BJT increases, its base resistance decreases. However, the collector current also increases for a given current density, and so the minimum noise figure remains almost constant, as predicted by theory [12]. At large collector currents, the f_T of the BJT begins to decrease, causing a rise in the minimum noise figure. Note that the minimum noise figure is almost independent of the emitter area selected, and that there is a broad range of bias points over which the noise figure of a typical common emitter amplifier can be minimized. Also, peak gain and minimum noise occur at different current densities, so gain and noise performance can be traded off in a way that is not possible with MOS devices.

The noise figure of a transistor is also frequency dependent. As the operating frequency begins to approach the device f_T, the gain of the transistor begins to decrease and the SNR degrades rapidly. For a constant bias current, the transistor transit frequency is lowest for larger area devices, and therefore the noise figure of a large area transistor will increase more rapidly with frequency than for smaller transistors.

Since the relative contribution of the transistor noise sources to the overall SNR depends upon the source impedance seen at the transistor input terminal, there is an optimum source impedance which results in the lowest noise figure. In general, this optimum noise match is not equal to the conjugate of the transistor input impedance required for maximum power transfer. An inductor placed in series with the emitter lead of the BJT (or source of a MOSFET) modifies the optimum noise match. Under certain conditions, the minimum noise figure and the maximum power transfer at the input can be achieved simultaneously, making this approach very attractive [13]. However, the amount of feedback which can be applied to the amplifier is limited by the constraints of gain and power consumption, and in most cases only a small amount of feedback can be applied [14]. Nevertheless, this technique is very useful when optimizing the performance of any LNA or mixer input stage.

D. Linearity

The minimum input signal level or sensitivity of a receiver is dominated by the LNA noise figure. The upper limit of the dynamic range is usually set by the distortion (e.g., intermodulation distortion) produced by the mixer. The linearity of the LNA must be greater than the mixer linearity when referred to the receiver input (note that high linearity implies low distortion).

Third-order intermodulation distortion (e.g., frequencies $2f_1 - f_2$ and produced by closely spaced input tones f_1 and f_2) falls within the intermediate frequency (IF) bandwidth and could interfere with signal reception in a heterodyne (or low-IF) radio [15]. Second-order intermodulation distortion products (e.g., frequencies $f_1 - f_2$ and produced by closely spaced input tones f_1 and f_2) have a similar effect in homodyne and low-IF receivers. Intermodulation distortion is characterized by the intercept specification (e.g., third-order intercept IP_3 or second-order intercept IP_2). The intercept point is measured by applying two equal-amplitude sinusoids at the mixer RF input (the two-tone test) and observing the intermodulation distortion products at the output. The intercept point referred to the input (e.g., IIP_3) is related to the output intercept (OIP_3) by the component (e.g., amplifier) gain. Whereas odd-order intermodulation distortion is generated by device nonlinearities or dynamic range limitations in RF circuits, such as the finite supply voltage, even-order distortion mainly arises from device mismatch, as differential circuits (which reject even-order harmonics) are typically used in practical implementations. Mismatch between bipolar transistors are typically lower than for MOSFET devices (3 σ offset voltages of ± 1 mV for bipolar V_{BE} and $\pm 4\%$ in collector current, versus ± 10 mV for MOS V_T and $\pm 3\%$–4% for drain current).

When distortion is considered, the linearity of the common-emitter bipolar amplifier alone (i.e., without feedback) is poor, with the input-referred third-order intercept (IIP_3) typically in the -12- to -20-dBm range. Intermodulation and harmonic distortions can be avoided by increasing the transistor bias current and the power supply voltage. However, a large bias current and a high supply voltage (which would be chosen to achieve high gain and good linearity) cannot be used in a low-power/low-voltage amplifier design. It is generally the amplitude of the base-emitter voltage that drives the device-related distortion produced by the transistor.

In theory, MOSFETs produce less distortion than bipolar transistors for the same bias current and drive signal. However, BJT-based LNAs reported in the literature do not appear to underperform compared to their MOS counterparts. This may be due in part to the fact that feedback is normally applied to the LNA in order to improve the overall linearity. Negative feedback is often applied to an amplifier to reduce the base-emitter voltage for a given output voltage swing and improves the linearity at the expense of lower gain. Feedback is a useful alternative to increasing the bias current for a low-power design, since it is well known that negative feedback can improve the linear input signal range of an amplifier, even though the active device itself might be operating in a relatively nonlinear manner.

The impedances presented at the input and output ports of an RF amplifier at other than the desired operating frequency

(i.e., other than the fundamental) also have a profound effect on the intermodulation (IM) distortion observed in RF amplifiers [18]–[21]. The amplitude and phase of unwanted sum and difference frequencies caused by device nonlinearities contribute to third-order intermodulation distortion by mixing with the RF signal at the input. One source of distortion is the difference frequency between two in-band RF input tones (i.e., second-order intermodulation distortion). This relatively low frequency signal can be attenuated by keeping the impedance of the bias path at the amplifier input as small as possible at low frequencies. However, it has been shown that the optimum value of the impedance seen at low frequencies is not zero, but finite and complex [19]. Also, the second harmonic of the RF signal is fed from the output back to input via the collector-base capacitance, where it mixes with the fundamental tones thereby generating third-order IM distortion. Attenuating the second harmonic at the RF input will reduce this source of distortion. Neutralizing the feedback from output back to input also diminishes this source of distortion. Circuitry (often passive) which eliminates the low frequency and second harmonic signal at the input is called an out-of-band harmonic termination.

Fifth-order intermodulation also has a greater effect on signal fidelity, as third-order intermodulation distortion is reduced. While impedances placed at the input to terminate out-of-band harmonics demonstrably reduce the IM distortion, experimental measurements also show that the improvement decreases with input amplitude and operating frequency. As the fundamental frequency approaches the bandwidth limitations of the transistor, the IM distortion increases. Moreover, the IM distortion improvement is realized only in the small-signal regime and distortion increases rapidly with RF input amplitude (power levels greater than about -30 dBm for a 50-Ω input impedance).

E. Other Considerations

There are other considerations that are also important to the overall performance of an RF IC chip that depend heavily upon the technology and the maturity/sophistication of the technology selected. Some of these are low frequency or flicker $(1/f)$ noise, breakdown voltages, circuit isolation and substrate noise effects, device matching properties, integration density, quality of the computer-aided design (CAD) tools and models, etc.

Flicker noise is important at frequencies up to the flicker corner frequency ($1/f$ corner), where the thermal and flicker noise levels in a device are equal. Aside from baseband circuity—especially in homodyne and low-IF radios—$1/f$ noise affects oscillator phase noise and mixer noise figure in homodyne receivers with intermediate frequency stages that operate below the flicker corner frequency. Therefore, it is desirable that the $1/f$ corner frequency should be as kept as low as possible. The flicker $(1/f)$ noise corner frequency for a BJT is typically three orders of magnitude lower than for a MOS transistor (see Table 2).

Breakdown voltage is important in circuits where either high voltages or high power outputs are anticipated. For MOS devices, the breakdown limit depends mainly upon the gate oxide, which will fail when the applied voltage exceeds the breakdown strength. This is 1.8 V for 0.18-μm gate length devices, but downscaling of the gate length in future generations of MOS devices will lower the breakdown voltage even further (e.g., 1.2 V for 0.13-μm technologies). In bipolar technology, breakdown voltage is limited by the collector-emitter breakdown voltage (V_{CEO}) when there is a high-impedance path for current flow from the base terminal to ground. The collector-emitter breakdown voltage with a finite resistance in the base lead (V_{CER} which can be two to three times higher than V_{CEO}) is more relevant to RF circuits, since excess electrons created by impact ionization in the base region often have a relative low-impedance path to flow out of the transistor's base terminal. The Johnson limit, which is the product of breakdown voltage (V_{CEO}) and f_T for the BJT, is approximately constant and is approaching 300 GHz-V for SiGe devices [47]. As a result, SiGe BJTs from a standard BiCMOS process flow are used to implement high-efficiency commercial power amplifiers with peak power outputs on the order of 1 W or more, while specialized LDMOS devices are required in order to improve the breakdown voltage and optimize the power output and reliability of an equivalent MOSFET power amplifier. It should be noted that integration of a final power amplifier stage together with other transceiver circuits is normally considered undesirable because of thermal loading and packaging constraints.

The common-mode rejection inherent in differential and symmetric circuit topologies helps improve the isolation between RF blocks in integrated radio circuits, which is one of the main reasons why the higher power consumption in these circuits is tolerated. For example, the differential pair has (ideally) the same noise figure and gain as the single-ended equivalent, but doubles the power consumption compared to a single transistor amplifier. As the RF input power is now split across two transistor base-emitter junctions, the distortion produced by the amplifier for a given input power level is lower than for a single transistor amplifier. Also, an explicit ground connection is not required, as there is a virtual ground at the common terminal in a differential amplifier driven by a balanced (i.e., differential) signal, so ground path parasitics (e.g., bondwire inductance and package inductance) do not affect amplifier gain. In addition, most mixed-signal technologies offer highly doped buried layers, heavily doped guard rings, trench isolation, and metal shielding schemes as ways of improving on-chip circuit-to-circuit isolation.

Other considerations for high-frequency are CAD model accuracy and the accuracy of the CAD design kits that are available to the designer. Deep submicrometer MOS models (e.g., the industry standard, BSIM3) are not scalable or accurate enough for RF analog design without extensive modification. Newer generations of MOS models (e.g., BSIM4 or Philips' MOS Model 13 and above) offer substantial improvements in small- and large-signal simulation accuracy. Bipolar models have improved considerably over the past decade. The latest bipolar models (VBIC, MEXTRAM, and HiCuM) are fully scalable and have demonstrated excellent accuracy in RF and high-speed applications.

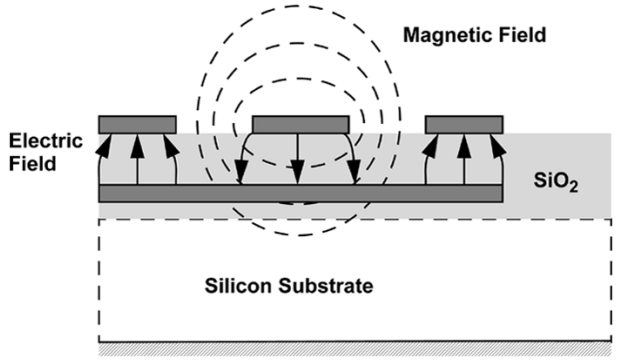


Fig. 9. Advances in silicon transistor unity-gain frequency (f_T) with technology.

IV. Magnetically Coupled Monolithic Components

On-chip passive magnetic components could be further leveraged to develop analog circuits which operate from supply voltages below 1 V with maximum efficiency. Also, innovations in on-chip passive design and construction beyond simple inductors are currently being pioneered in mixed-signal silicon technologies such as SiGe. The state of the art in on-chip magnetic components is reviewed in this section.

Digital applications continue to push CMOS development forward, with integration levels increasing by a factor of two every 18 months, as predicted by Moore's Law. Aggressive scaling of transistor dimensions for digital applications not only increases the number of devices that can be packed into a given area, but also their operating speed. In addition, innovations in the structural design of electronic devices has further reduced parasitics and increased speed/bandwidth. The progressive improvement over time in the unity-gain or f_T for silicon active devices is illustrated in Fig. 9; transit frequencies over 300 GHz are ultimately possible.

Are these transistor developments alone sufficient to address the challenges posed by wireless applications beyond 20–30 GHz, such as a 60-GHz integrated broad-band transceiver? All silicon-based technologies have common elements such as barriers to transistor downsizing (or scaling), the technology used for interconnecting on-chip components, and the conductive silicon substrate (even in SOI technology, where devices and the insulating layer are bonded to or fabricated on a semiconducting substrate). In digital applications, the electrical performance and packing density of on-chip interconnections is the greatest bottleneck to improving performance aside from transistor speed. Interconnections also restrict radio and high-speed circuit performance in silicon-based technologies, and the difficulties currently faced by designers attempting to extend the reach of silicon circuits toward millimeter-wave frequencies will now be considered.

A. On-Chip Interconnect—When Is a Wire No Longer a Wire?

Analog RF applications have circuit requirements that demand much more than fast switching speeds between binary states and the capability to pack an enormous number of devices onto a single chip. The passive devices (including interconnections) are just as important to the ultimate circuit performance as the transistors, especially as the operating frequency increases. This challenge has resulted in a push toward lower loss interconnect metals (e.g., switching materials from aluminum to copper), reducing parasitic capacitance (using dielectrics with lower permittivity than SiO_2), and increasing the interconnect density by implementing more wiring planes (from six to eight metal levels and perhaps ten or more in the near future).

One of the benefits of the small physical dimensions inherent in IC technology is that propagation delays along interconnections are usually dominated by component parasitics, which makes them relatively easy to analyze. For example, a 1-GHz signal sees a phase shift of less than 3° across a 1-mm connection corresponding to a travelling time or "time of flight" of 8 ps. This is tiny compared to the delays caused by resistances and capacitances typically present in CMOS circuits. These conditions have prevailed since the early years of IC development, and popular circuit analysis programs such as SPICE and Cadence-Spectre implicitly assume that all circuit components behave as lumped elements.

The "time of flight" over interconnect wiring is normally neglected but becomes important at multigigahertz frequencies. This is partly due to fact that average wiring length in circuits is not shrinking as quickly as the transistor dimensions. The physical space required by resistors, capacitors and inductors, power supply wiring, and interconnect bondpads, as well as the need to circuits physically separate as a way of reducing electrical interference between blocks, keeps wiring lengths between 10 and 100 μm within circuit blocks and approaching 0.5–1 mm between blocks of circuitry.

The millimeter-wave frequency range (wavelengths of 10 mm or less) corresponds to a frequency of 12.5 GHz or more for signals travelling on a conventional silicon IC. Above these frequencies, the wavelength encountered on-chip approaches the length of a typical interconnect wire and time-of-flight delays become important. Wires no longer behave as a simple lumped resistor–capacitor filter, but begin to look like transmission lines. This indicates that a change in the computer-aided design tools used to analyze circuits is needed, because the SPICE-type lumped-element models are no longer valid. A shift toward the methods used by microwave circuit designers to synthesize higher frequency circuits is likely necessary in cases where distributed-parameter design is required.

One solution to deteriorating signal quality caused by the semiconductor is building the IC on an insulating or higher resistivity substrate (e.g., SOI). The most common production process for SOI (SIMOX) incorporates a thin insulating film (typically 4–6 μm thick) by implanting oxygen into a conventional low- or medium-resistivity silicon substrate. SIMOX forms uniform insulating islands for the active devices. The UNIBOND technique bonds together two wafers with a thin oxide film trapped between them. A semiconductor film for transistor fabrication is then produced by

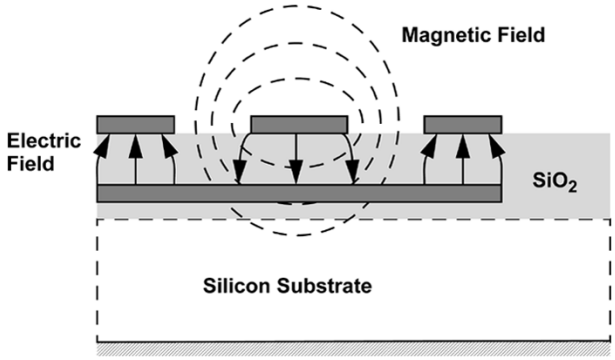


Fig. 10. Shielded coplanar waveguide (S-CPW) on silicon.

mechanically thinning one of the two bonded wafers. In either case, electromagnetic energy easily penetrates the thin insulating layer near the surface of the wafer. As a result, interconnect performance still suffers at gigahertz frequencies where energy is coupled into the underlying semiconducting layer if low- or medium-resistivity substrates are used. SOI wafers up to 5 kΩ-cm resistivity with diameters exceeding 200 mm are available for production, but they are more expensive to produce. Charge accumulation at the oxide semiconductor interface must be suppressed by additional processing steps for resistivities lower than about 10 kΩ-cm in order to maximize the benefit of the semi-insulating wafer on interconnect performance [50]. This increases manufacturing costs further. Therefore, other methods of improving interconnect performance with conventional process flows (e.g., shielding the interconnect from the lossy substrate) are important for high-volume, low-cost applications like wireless transceivers.

B. Shielded Interconnect for Lower Losses

Signal integrity can also be improved by shielding the interconnect from the substrate. This comes at the cost of lowering the characteristic impedance of the transmission line when a solid metal shield layer is used (e.g., top metal signal conductor over a grounded first metal shield, or metal–insulator–metal (MIM) line). It increases the overall power consumption, since lower impedance levels imply higher current consumption generally. It should be noted that low impedances are generally needed to achieve wide bandwidth because of capacitive parasitics, so more power must be consumed despite the development of faster and smaller devices as we migrate toward higher operating frequencies.

The slow-wave coplanar waveguide (S-CPW [53]) configuration of Fig. 10 overcomes many of the limitations of existing on-chip transmission line designs [49]. It consists of three coplanar top conductors where the gap between signal and ground is wide enough to achieve an inductance per unit length (L) comparable to a single-metal microstrip-on-silicon line. Floating metal strips are placed beneath the top conductors and oriented to minimize current induced by the flow of signal current in the topmetal. This results in a capacitance per unit length C, similar to the microstrip line with a solid metal ground shield. As a result, the wave speed $(v = 1/\sqrt{LC})$ and wavelength are lower than for other configurations. Distributed microwave devices such as quadrature and hybrid couplers [54] will use less chip area when designed with S-CPW. In addition, only a fraction of the electric field enters the silicon substrate via the floating strips which further reduces signal attenuation at high frequency.

The measured attenuation per millimeter of length for four transmission line configurations are compared in Fig. 11 [53], [55]: microstrip and coplanar waveguide on silicon, MIM microstrip, coplanar waveguide on an insulating substrate (CPW on alumina), and S-CPW lines. A 50-Ω CPW (55- and 200-μm-wide signal and grounds, respectively) made from 2.5-μm-thick gold conductors separated by a 20-μm gap on an insulating alumina substrate is used as a reference for comparison. Attenuation for the reference CPW (gold on alumina) increases continually with frequency, and 0.2–0.3 dB/mm at 50 GHz is typical for this technology.

The other transmission line measurements shown in Fig. 11(b) are for a topmetal microstrip-on-silicon line, consisting of 4.5-μm-wide, 2.1-μm-thick aluminum topmetal without a first-metal ground sheet, a 2.5-μm-wide line and 10-μm gap coplanar waveguide on silicon (CPW on silicon), and a 50-Ω MIM transmission line with topmetal width of 14.5 μm. The microstrip-on-silicon and CPW lines are fabricated from 2.1-μm-thick aluminum topmetal over 5.75-μm oxide/200-μm-thick silicon substrate. The MIM line is made from 6.7-μm oxide sandwiched between 2.1-μm-thick topmetal and 0.5-μm first metal (both aluminum) layers on a 350-μm-thick silicon substrate. For the S-CPW, topmetal (2.3 μm thick) and 1-μm-thick shielding strips in an all-copper analog wiring scheme are used. A 6.73-μm-thick oxide isolates the shield strips from the semiconducting substrate. The S-CPW line uses 420-μm-wide floating strips with minimum length and spacing (1.6 μm). Coplanar topmetal grounds are placed 20 μm away from a 16-μm-wide center conductor.

Silicon substrates with the same resistivity (10 Ω-cm) were used for fabrication of an all-silicon-based interconnect (microstrip and CPW on silicon, MIM, and S-CPW) shown in Fig. 11(b). To summarize the results, attenuation increases rapidly due to energy loss to the substrate for the microstrip and CPW fabricated over bare silicon, giving severe attenuation above 10 GHz. The metal-insulator-metal sandwich (MIM) structure has lower attenuation, but characteristic impedance which can be realized on-chip is restricted to below 100 Ω for topmetal trace widths greater than approximately 8 μm. The S-CPW configuration allows the magnetic field to fill a larger volume, and so a wider topmetal signal trace can be used compared to either MIM or conventional CPW designs [53]. This is advantageous for high-current applications (e.g., power amplifiers), where the metal traces must be wide enough to support high ac and dc current flows without reliability problems caused by electromigration of metal over time. As a result, the S-CPW has very low attenuation per millimeter of length (as seen in Fig. 11). The S-CPW line also shows reduced wavelength because of its lower wave velocity. Thus, an on-chip quarter-wavelength

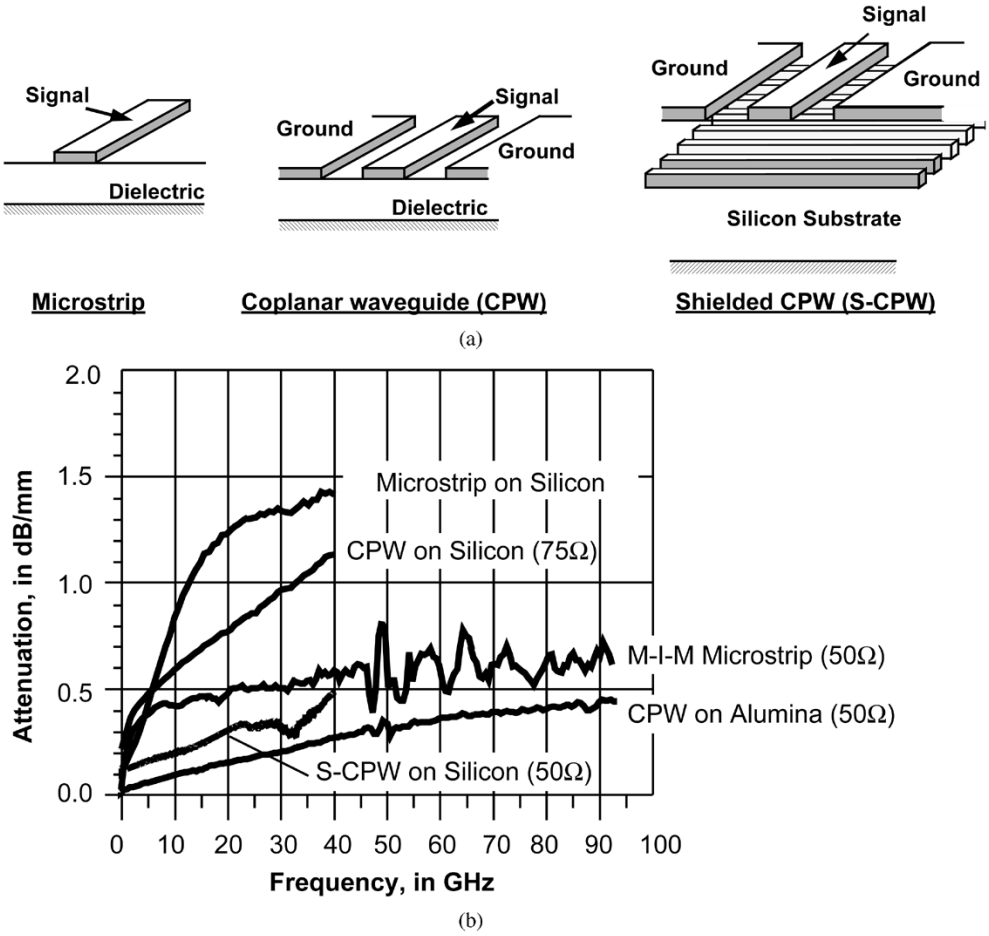


Fig. 11. Performance of on-chip interconnect at millimeter-wave frequencies compared to an off-chip reference standard. (a) Monolithic transmission line designs. (b) Measure attenuation (per millimeter of length).

(90°) phase shifter at 30 GHz shrinks by more than a factor of two when using S-CPW compared to conventional CPW on silicon. The quality FOM (Q-factor) for a transmission line resonator [54] is defined by the ratio of phase shift and attenuation per unit length, and it quantifies the energy stored to energy dissipated per cycle of a sine wave by the line. The Q is improved by 2 $\times$ over most of the 40-GHz frequency range for the S-CPW compared to the MIM test lines from Fig. 11, and 3 $\times$ in the millimeter-wave range between 30 and 35 GHz [53]. This improvement in Q arises from the combination of lower attenuation and wavelength reduction (i.e., more phase shift per unit length) with S-CPW.

Transmission lines are used for pulse shaping [56], [57], delay lines for VCOs, 90° phase shifters, as synthetic inductors and capacitors on-chip at millimeter-wave frequencies, and as signal processing elements [54]. When transistors are embedded in a transmission line [58], the effect of capacitive parasitics are minimized, resulting in circuit topologies (such as distributed amplifiers and oscillators) that can operate as close to the unity-gain bandwidth of the devices as possible. They are also the foundation for other passive components such as on-chip inductors and transformers.

C. On-Chip Inductors

Resonant-tuned (*LC*) circuits offer many benefits to the designer of high-frequency circuits. Operation at a low supply voltage, simplified impedance matching between stages, and low energy dissipation for reduced circuit noise are just a few of the properties of *LC* circuits that can be exploited to achieve a higher level of performance. At radio and microwave frequencies, a purely passive inductor is often preferable to synthesis of an inductive reactance with an active circuit. Passive components introduce less noise, consume less power, and have a wider bandwidth and linear operating range than their electronic equivalents, such as the gyrator. Prior to the mid-1990s, silicon IC technology was rarely used for analog applications in the radio and microwave range of frequencies, in part because transmission line structures performed poorly on a semiconducting substrate. However, suitable performance can be realized when the limitations imposed by the technology are understood by designers and the components are accurately modeled and characterized.

Planar inductors for monolithic circuits are useful for interstage matching and coupling, as resonant loads, and for biasing and bias circuit isolation in RF IC applications [60]. They can be realized in a number of configurations, all implemented (at minimum) using a single-layer metallization scheme and as on-chip transmission lines (i.e., a microstrip line). The total line length must be kept at a small fraction of a wavelength; otherwise, the conductor cannot be treated as a lumped element. These components only approximate

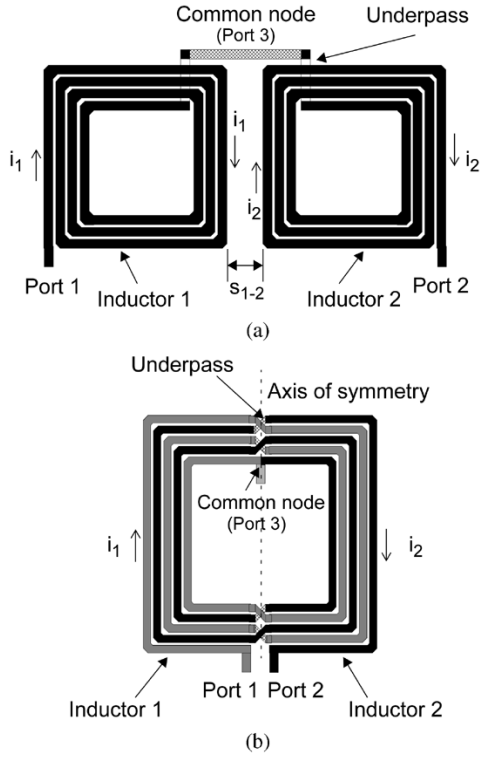


Fig. 12. Microstrip inductor physical layouts for differential drive, after [33]. (a) Two asymmetric square spiral inductors. (b) Symmetric square spiral inductor.

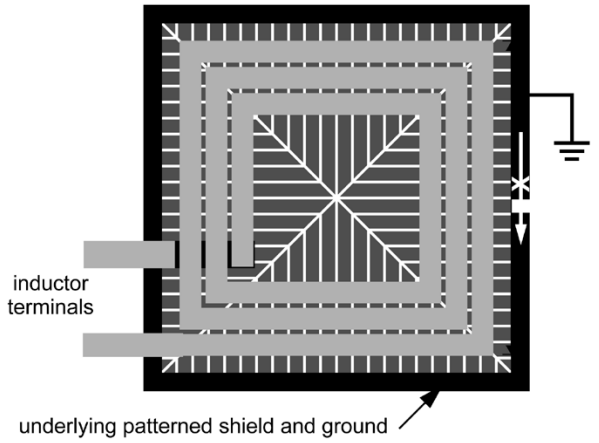


Fig. 13. Patterned ground substrate shielding for a spiral inductor.

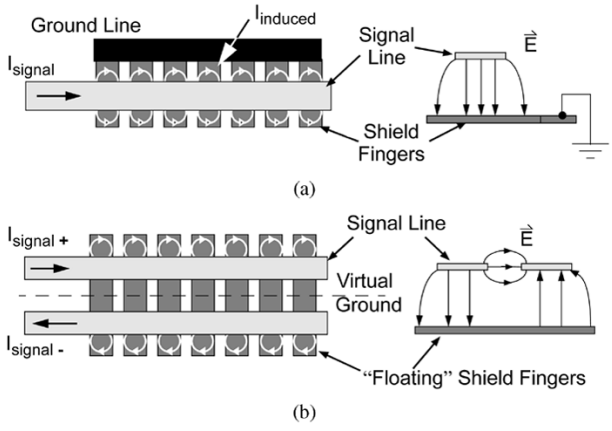


Fig. 14. Simplified drawing of substrate shielding methods. (a) Grounded shield. (b) Differential shield.

lumped inductors, and in this sense they do not have a low-frequency equivalent for monolithic design, unlike on-chip capacitors, which are also used in monolithic circuits for signal processing from audio to radio frequencies. They can also be used to implement components that are unique to RF circuits, such as delay lines and transformers [35], [59], [61], [62]. The silicon monolithic inductor performs best in a balanced circuit, because substrate parasitics have less effect on the self-resonant frequency and the Q-factor FOM [33].

The Q-factor is often used to compare passive components such as inductors and capacitors. Q-factor for an inductor is easily determined from a resonant *LC*-tank built using a an ideal capacitor at the resonant frequency ω_o. The Q is then given by the ratio of -3-dB bandwidth to ω_o, or $Q = \omega_{-3\ \mathrm{dB}}/\omega_o$. Energy dissipated in the conductor metals or the substrate or radiated to the surrounding environment reduces the Q.

The fully symmetric spiral inductor [33] of Fig. 12(b) is designed for differential excitation (i.e., voltages and currents at the terminals are 180° out of phase). The differentially driven symmetric inductor can be viewed as a transformer (i.e., magnetically coupled windings) where two of the transformer's terminals are connected together at the midpoint of the winding. When driven differentially, the voltages on adjacent conducting strips are anti-phase, however, current flows in the same direction along each adjacent conductor shown in Fig. 12(b) (i.e., signal currents i1 and i2 flow in the same direction on any side). This reinforces the magnetic field produced by the parallel groups of conductors and increases the overall inductance per unit area. A pair of asymmetric inductors (as in Fig. 12(a) must be spaced far enough apart to limit unwanted coupling (both magnetic and electric) between the inductor pair. The symmetric inductor does not face this restriction, so it consumes less chip area and simplifies connections to differential circuits.

Spiral inductors fabricated in production silicon technologies suffer from losses caused by energy dissipated in both the metal spiral and the semiconducting substrate. When implemented on medium resistivity silicon (i.e., 1–100-Ω-cm material), it is electric coupling of energy to the substrate that limits the inductor Q-factor. Grounded substrate shielding layers have been devised [77] that can block the electric field from entering the silicon, thereby improving the inductor Q. An effective substrate shield blocks current flow in the shield layer itself by magnetic induction, so that the inductance of the coil is not affected by the shield. However, the shield fingers capture almost all of the electric field and a capacitive current flows to the ground. An example shield pattern, which is often implemented using silicided (e.g., MOS gate) polysilicon, is shown in Fig. 13. Gaps between the shield fingers block current induced by the magnetic field produced by the signal current (limited to small eddy currents, as shown in the simplified diagram of Fig. 14). Each finger is connected to a common ground line, which must be properly grounded and carefully designed so that induced

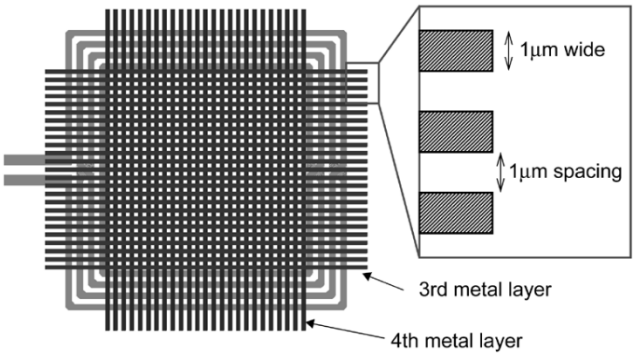


Fig. 15. Differentially shielded symmetric inductor.

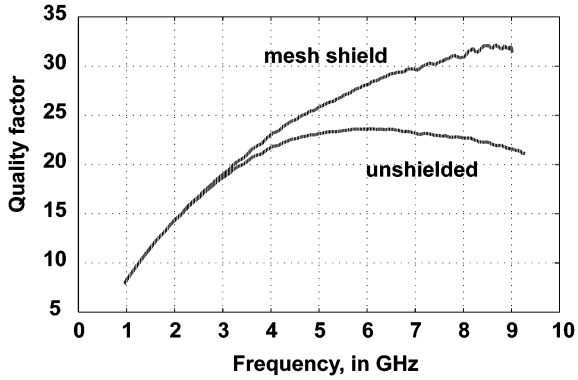


Fig. 16. Measured inductor Q-factor on a 10-Ω-cm substrate.

current is minimized. However, directly grounding the shield layer adds parasitic capacitance that reduces the inductor's self-resonant frequency and limits the useful frequency range of the inductor in a circuit. Also, a reference "ground" is ambiguous and ill-defined on-chip at RF, because of the effect of packaging parasitics (such as bondwires and lead inductance) which isolate the on-chip ground from circuitry external to the IC. Nevertheless, patterned ground shields may improve the Q-factor of an inductor by 25%–30% [77].

Differential circuits are often used in RF ICs to improve isolation between on-chip circuits and also to limit impairments caused by packaging parasitics. For differential circuit applications, a "differential shield" can be used to improve inductor Q. One advantage of the differential shield is that it does not require an explicit ground connection for the shield layer, leading to a simpler implementation. Fig. 14(b) shows a simplified drawing of a differential shield implementation. The balanced signal lines are capacitively coupled to orthogonal metal strips placed beneath it. Due to the balanced excitation, a virtual ground (the net voltage induced by electric coupling to the underlying metal strips) shields the signal conductors from the underlying silicon substrate.

An example of a differentially shielded inductor is illustrated in Fig. 15 [34]. The "mesh" shield pattern is composed of horizontal and vertical strips that span the length and width of the inductor. A voltage induced on a shield strip along one side of the coil is compensated for by an equal but opposite voltage induced at the other end of the strip when the inductor is driven differentially. Since the metal strips have little loss, the net voltage induced on the strips is very close to zero. Induced current is inhibited by placing the strips orthogonal to the inductor winding.

The Q-factor for differential shielding is compared to an unshielded design in Fig. 16. The measured low-frequency inductance is about 7.4 nH for both inductors. The parasitics of a differential shield lower the self-resonant frequency of the inductor by less than 3%. There is less than 2% difference in the inductance between shielded and unshielded inductors, which indicates that current induced in the shielding strips from the coil's magnetic field is very small. In other words, the addition of a differential shield does not diminish the useful frequency range or the inductance value. Fig. 16 shows that differential shielding improves the Q-factor by 35%.

A summary of the performance from various inductors fabricated in silicon technology is listed in Table 3. It should be noted that smaller inductance values have higher peak Q-factors, which makes direct comparisons between designs difficult to interpret. In any case, the data shows that thicker topmetal, or stacking the upper layers of metal (i.e., using many vias to connect metal layers in parallel), effectively increases the winding thickness and improves the Q-factor. However, the oxide layer separating the inductor winding from the silicon substrate must be thick enough to isolate the inductor from the silicon substrate. As a result, stacking all metal layers usually results in poorer performance than using the top one or two layers. High-resistivity substrates, or complete substrate removal by etching, increases the Q-factor substantially at the expense of higher production costs cost and potential difficulties with manufacturability and packaging (e.g., protection of the silicon from contamination after substrate removal via etching). The effects of differential excitation and shielding over a patterned ground shield for the inductor are also clear from the data listed in Table 3. Differential drive improves the Q by almost 50% for the same inductor design, while differential shielding increases the Q-factor by reducing energy loss to the conductive substrate with only a negligible effect on the self-resonant frequency.

D. On-Chip Transformers

The on-chip transformer relies upon the mutual magnetic coupling between two or more windings. It is designed to couple alternating current from one winding to another with minimal loss, and impedance levels between windings are *transformed* in the process (i.e., the ratio of terminal voltage to current flow can be changed between windings). In addition, direct current is blocked by the transformer, allowing each current loop connected to a winding to be biased independently. Consequently, the transformer can be used to ac couple RF stages, with the added advantage that each stage can use resonant tuning to mitigate the deleterious effects of parasitic capacitance and extend operating bandwidth. Each transformer winding (ideally) has negligible dc voltage drop, which should prove useful in the coming era of low-voltage analog circuits where voltage headroom must be preserved in order to operate at sub-1-V supplies. The ability to couple

Table 3
Silicon Technology Inductor Comparison

Inductor Type	Reference	Resistivity, in Ohm-cm	Metal Thickness, in μm	Inductance, in nH	Peak Q-factor
1-level metal	[61]	10	1-3	1.88	6-10@4GHz
2-level metal	[76]	2,000	2	13	12@3GHz
5-level metal	[77]	12	4.3	2.2	16@2GHz
Patterned Ground Shield	[78]	10-20	2	8	7.2@1.5GHz
Membrane	[79]	2,000	1 (Au)	0.9	20@4.3GHz
Differential	[33]	15	2	8	9.3@2.5GHz
Single-ended					6.6@1.6GHz
Differential Mesh Shield	[34]	10	2.3 (Cu)	7.8	28@6.4GHz

multiple circuit loops magnetically using a multifilar transformer enables circuit functions on-chip that are compact and difficult to implement with similar quality using conventional IC components.

Interwinding microstrip spiral inductors to magnetically couple independent conductors results in a monolithic transformer. Note that like silicon monolithic inductors, transformers on silicon chips typically use wide metal lines spaced closely together in order to minimize ohmic losses and maximize the coupling of magnetic flux between windings. Frlan and Rabjohn [35], [63] both demonstrated compact physical layouts for spiral transformers and developed circuit simulation tools based upon the extraction of lumped element models from physical and geometric parameters. This modeling technique was later extended to the analysis of planar structures on silicon substrates [59]–[62], although commercial electromagnetic simulation tools are now widely available which can be used to analyze these components. In the recent literature, there are many examples of monolithic transformers fabricated in silicon IC technology for use in RF circuits, such as preamplifiers [64], oscillators [65], mixers [22], and power amplifiers [66]–[68].

Transformer Baluns: Multifilament transformers can also be constructed on-chip. These devices are used to implement power dividers/combiners and baluns. A balun is a device which couples a balanced (i.e., differential) circuit to an unbalanced (i.e., single-ended) one. There are many structures used to implement baluns at RF and microwave frequencies, although a differential amplifier is the most commonly used circuit for unbalanced to balanced signal conversion on-chip. Microwave balun structures such as the Lange, rat-race, and branch line coupler require physical dimensions on the order of the signal wavelength and therefore consume too much chip area when operating below approximately 15 GHz. A disadvantage of asymmetric physical layouts is that the electrical and physical center of an asymmetric winding are not coincident.

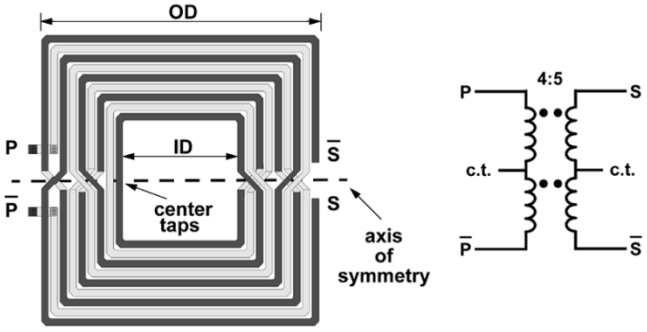


Fig. 17. Square symmetric (Rabjohn) balun.

The square symmetric Rabjohn winding of Fig. 17 solves this problem. The balun is constructed from two groups of interwound microstrip lines that are divided along a line of symmetry running horizontally, as shown in the figure. The groups of lines are interconnected so that all four terminals to the outside edge of the transformer layout, which is an advantage when connecting the transformer to other circuitry. Also, the midpoint between the terminals on each winding, or the center-tap, can be located precisely in the layout as indicated in Fig. 17. The turns ratio for the example shown is 4 : 5 between primary and secondary.

In practice, there is a slight difference in the magnitude response at the inverting and noninverting secondary ports due to the effect of interwinding capacitance. This effect cannot be reduced by adding extra capacitance in shunt with the transformer ports when tuning a resonant circuit. However, tuning does reduce the loss, where close to an ideal 3-dB power split at the balun outputs has been achieved [62]. This phase/amplitude error is typically on the order of 1°/0.5 dB in the desired passband for a well-designed balun.

The decreasing breakdown voltage of SiGe bipolar transistors with each new technology generation presents a challenge to the designers of RF power amplifiers. To increase power output and amplifier efficiency, an on-chip power combiner is often used to combine the power from

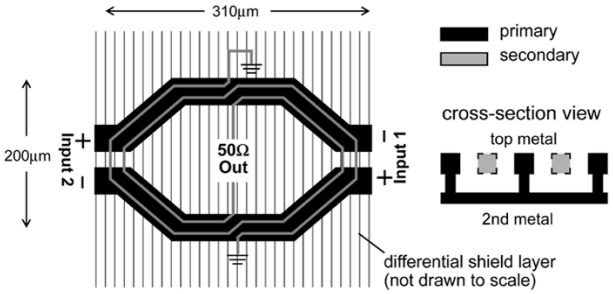


Fig. 18. Power combining balun for a broad-band millimeter-wave power amplifier [68].

multiple or n transistor units to a single output (i.e., n-way combining). In order to maximize the output from each amplifier, the combiner must present a controlled impedance to each transistor and at the single-ended output over the required bandwidth. Any load imbalance can cause some transistors in the amplifier to operate on the verge of breakdown, while the others are operating well away from their maximum output rating. This results in an overall reduction in the maximum output power. Losses and bandwidth of the combiner also have a direct impact on the efficiency of a power amplifier.

A four-way power combining balun (see Fig. 18) with 0.9-dB loss for 21–27 GHz medium power (i.e, >20-dBm output) applications has been developed in a 0.18-μm SiGe-BiCMOS technology [67]. Based on a four-way transformer combiner, the symmetrical design transforms a 50-Ω output load to four equal 13-Ω loads with only 3% load mismatch. The self-shielded transformer windings achieve a high magnetic coupling coefficient ($k > 0.85$) between a single-turn primary and two secondary coils of the transformer. A self-shielding winding (also shown in Fig. 18) together with a differential shield, is used to lower energy leakage to the 10-Ω-cm silicon substrate. A multifilament secondary also overcomes asymmetry seen in earlier transformer balun designs developed for power combining applications [69]. Moreover, the design minimizes the skin effect and resultant power lost due to heat in the balun winding.

V. Magnetically Coupled Circuits for Broad-Band and Low-Power Applications

The next generations of wireless and hybrid wireless/wireline communication systems are demanding further innovations to meet the performance constraints imposed by compatibility and regulatory requirements. Rising consumer demand for wider bandwidth services such as video streaming or video on demand will quickly stress the capabilities of today's wireless systems, so additional frequency allocations will be necessary. Many gigahertz of spectrum for back-haul links, point-to-multipoint distribution, and other broad-band services may be opened-up by new RF circuit developments aimed at the millimeter-wave band (i.e., wavelength <10 mm in air, or frequencies >30 GHz). The 60-GHz band is of interest for dense local-area communication, because there is 3 GHz of spectrum available worldwide near 60 GHz (i.e., from 59 to 62 GHz). This far exceeds the combined spectrum from 2.4- and 5–6-GHz bands currently available for wireless networking. While the high attenuation at 60 GHz (approximately 10 dB/km) makes it unsuitable for spans greater than about 1 km, attenuation has little impact on communication between devices locally (e.g., within the same room). Approximately 5 GHz of spectrum has been allocated in Europe (59–62 GHz for WLAN and 62–63/65–66 GHz for mobile broad-band systems) and North America (59–64 GHz, unlicensed), while 7 GHz has been allocated in Japan (59–66 GHz) for high-speed data communication [48]. Fast, low-cost circuit technologies such as 200–300-GHz f_T SiGe-BiCMOS provide a relatively simple way of unlocking this bandwidth for local networking incorporating broad-band applications such as multimedia.

In the near future, wireless networks will connect people as well as devices, making information available anywhere and at any time. Wireless hardware is a key component for the successful deployment of ubiquitous connectivity in applications such as multimedia services, automated traffic control, ambient intelligence, and home-based medical care. However, the dramatic reduction in form factor at chip level derived by migrating from conventional hybrid RF transceivers to integrated solutions has not been fully realized. Part of the reason for this is that packaging, which accounts for roughly 40% of the cost of the final radio and significantly limits the size of the component, has not been developed with more economical batch processing methods. There are, however, attractive wireless system concepts that rely on miniature-size radio components using multihop signal processing rather than point-to-point communication. These network can either be constantly "on" or woken up periodically (ad hoc networks). Both low-data-rate sensor and broad-band communication ad hoc networks are conceivable, but RF ICs operating at supply voltages compatible with digital baseband circuits (i.e., below 1 V for the latest CMOS technologies) will be required for these applications [51]–[52].

Greater understanding of the limitations of passive devices coupled with improved models for their performance are leading to circuits offering wider RF dynamic range and operation at higher frequencies and lower voltages with reduced current consumption and greater efficiency. Sophisticated magnetic components that also perform signal processing functions, combined with deep submicrometer SiGe bipolar or MOSFET active devices are envisioned that enable single-volt transceiver circuits which consume less supply current. Some of the results from the first efforts at building circuits using these types of devices will be reviewed in Sections V-A–F.

A. Transformer Feedback Amplifier

As the supply voltage of digital circuitry shrinks with technology scaling, RF circuit topologies require sub-1-V operation. This is because integration of analog/RF and digital circuitry on the same die is desirable from both cost and

packaging considerations. In addition, as operating frequencies increase, amplifier designers can no longer neglect the effects of the collector-base capacitance C_μ (or drain–gate capacitance in a MOSFET) on performance. The effects of parasitic feedback via C_μ are reduced using a cascode configuration; however, a two-transistor stack is not optimal for operation at the lowest possible supply voltage. The transformer feedback amplifier employs reactive negative feedback through an on-chip transformer that can neutralize C_μ, while also allowing a transistor bias voltage equal to the supply voltage (i.e., $V_{\rm CE} = V_{\rm CC}$). As a result, gain and dynamic range are not compromised when only a single active device is used [22], [64].

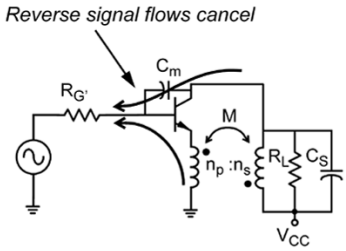
Circuit techniques that mitigate the effect of C_μ are usually grouped into two categories: unilateralization and neutralization. Unilateralization decreases reverse signal flow and coupling between output and input ports of an amplifier. Neutralization cancels the parasitic feedback by adding signal paths around the amplifier that cancel signal flow via C_μ. This technique increases the forward gain and reverse isolation for a given power consumption, but does not necessarily reduce the effect of C_μ on the input capacitance.

Feeding back a portion of the output signal via a transformer as shown in Fig. 19(a), can effectively cancel the feedback from output to input through the Miller capacitance (C_μ) and neutralize a single-stage amplifier. This increases the amplifier gain for a given bias current and improves the isolation between output and input. Neutralization is achieved when

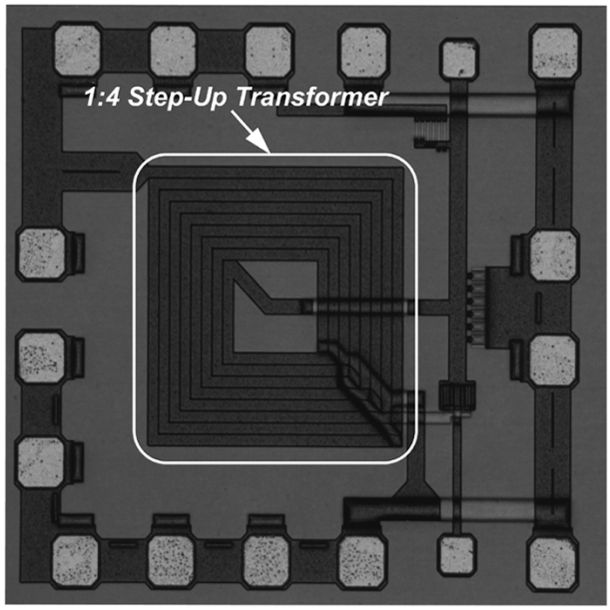
$$\frac{n}{k_m} = \frac{C_{\rm be}}{C_\mu} \tag{2}$$

using a transformer with imperfect magnetic coupling. Note that there is no frequency dependence involved in the neutralization equation [i.e., (2)], which implies that transformer feedback can be used as a wide-band neutralization technique that is limited only by the bandwidth of the transformer itself. With $0.7 < k_m < 0.9$ being typical for a transformer fabricated on a silicon IC, bandwidths greater than 2–3 octaves are possible. For a given LNA design, the transformer turns ratio (n) is often constrained by physical constraints (e.g., what is physically realizable), linearity, gain, and noise specifications. In these cases, the coupling coefficient (k_m) is the extra degree of freedom that can be adjusted to achieve amplifier neutralization. This is accomplished by adjusting the spacing between the transformer primary and secondary windings.

An example of a transformer feedback amplifier fabricated in a 0.5-μm SiGe bipolar technology is shown in Fig. 19 [23]. The 2.4-GHz LNA draws 2.5 mA from a 0.9-V supply. The step-up ratio between primary and secondary of the transformer is realized by sectioning one winding (e.g., the primary) into a number of single turns rather than one continuous winding. These single-turn windings are then connected in parallel to form the step-up ratio between primary and secondary of the transformer. The 1 : 4 step-up design shown in Fig. 19(b) consists of eight turns of



(a)



(b)

Fig. 19. Transformer feedback amplifier. (a) Neutralization using a feedback transformer. (b) 1 : 4 feedback LNA in 0.5-mm SiGe technology.

10-μm-wide topmetal with a 3-μm conductor spacing and measures 350 μm on each side. The step-up transformer is an almost ideal feedback element for an RF amplifier and can be used as a narrow-band alternative to a broad-band resistive network. The impedance match for the LNA input is off-chip so that either 50-Ω or minimum noise figure matching between the source and the amplifier input could be selected. When matched for optimal noise performance, the measured preamplifier noise figure is 0.95 dB at a gain of 10.5 dB (biased at 2.5 mA from 0.9 V). When a 50-Ω impedance match is used, the gain rises to 11 dB at a noise figure of 1.75 dB. At the same bias point, the third-order intercept point (IIP3) of the preamplifier is -4.5 dBm for both matching situations. This excellent combined performance (i.e., noise figure, gain, power consumption, and IIP3) was achieved with a preproduction 0.5-μm SiGe bipolar process that was not optimized for RF performance from the active or passive components.

B. Balun-Coupled Mixer

One application of the on-chip balun is in an RF mixer, where the balun couples the RF input to a quad of transistors as used in the traditional Gilbert multiplier [22]. This circuit configuration (see Fig. 17) retains many of the benefits of the Gilbert multiplier, such as the balanced topology and

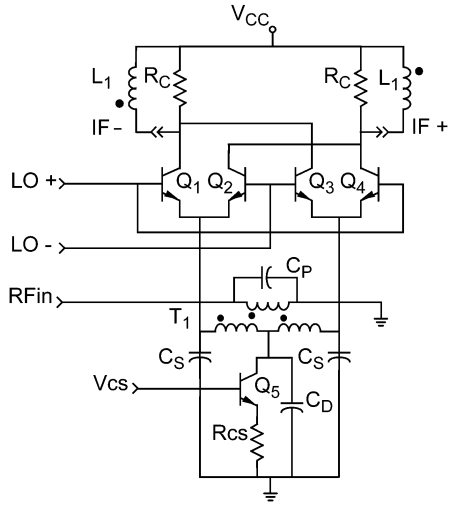


Fig. 20. Transformer-coupled Gilbert mixer using an input RF balun [22].

compatibility with silicon IC technology, while allowing wider dynamic range and low-voltage/low-power operation.

A balanced circuit is required to obtain good port-to-port isolation and rejection of spurious signals, and the switching quad from the Gilbert multiplier is well-suited to monolithic integration. Transformer coupling of the RF input signal retains the advantages of a doubly balanced topology while improving upon the performance of the active circuit typically used to generate differential RF signal currents in Gilbert's multiplier. The RF input signal to the mixer (refer to Fig. 20) is split into in-phase and anti-phase components by balun T_1. These signals are then fed to the cross-coupled switching quad of transistors, Q_1 to Q_4. Bias current is fed from current source Q_5 to the switching quad through the center-tap in the balun secondary. The signal current is chopped by the transistor quad at the LO rate in order to downconvert the input signal from RF to the desired intermediate frequency (IF). Package and bondwire parasitics have a relatively small influence on the IF port matching, and a good impedance match (if necessary) can be achieved through the proper selection of the collector load resistance, R_C.

Close to ideal transformer behavior can be realized over a narrow bandwidth when the balun primary and secondary windings are resonant tuned. The LO inputs of the switching quad are driven with a large amplitude signal, and therefore two of the transistors in the quad are biased in the active region, and the other transistors in the quad are cut off for a large portion of each cycle (e.g., when LO+ is much larger than LO−, Q_1 and Q_4 are "on" and Q_2 and Q_3 are cut off in Fig. 17). The transistors biased in the active mode operate in the common-base configuration and amplify each phase of the received signal to the intermediate frequency output. The transformed source resistance (r_{SRC}) degenerates the common-base amplifier and extends its linear range of operation. The transformer matches the source impedance to the mixer and improve the mixer linearity without causing a significant increase in the overall noise figure. This occurs because no additional dissipation has been added to the circuit other than the losses in the transformer windings, which are relatively small. Linearization of conventional IC mixers, such as the Gilbert-type balanced demodulator, requires degeneration resistance which degrades the mixer noise figure when high linearity is desired.

The noise introduced by the mixing process is difficult to determine analytically, but can be simulated. Operating the mixer at a low bias current reduces the shot noise contributed by each active device. However, fast switching speed of the transistors in the quad is also important when attempting to realize a lower noise figure. Careful selection of the emitter area for the transistors in the switching quad is therefore required in order to achieve a good mixer noise figure. A compromise is needed between a small transistor which can switch quickly between states and a larger transistor with less thermally generated noise from the transistor extrinsic base resistance, $r_{\mathrm{bb'}}$.

The transformer balun-coupled mixer demonstrates low-voltage operation and low-current consumption (5 mA from a 1.2-V supply). The supply voltage could be pushed below 1 V if the bias current source (Q_5 from Fig. 17) were replaced by a current limiting resistor. The 4 : 5 balun step-up ratio is chosen for a 50-Ω match to the RF input.

C. Transformer-Coupled RF Stages

Conventional RF IC front ends often use a cascode LNA and a mixer consisting of a differential pair input cascoded with a four-transistor switching quad. The power supply voltage of these configurations is limited to (minimum) 1.8 and 2.7 V for the LNA and mixer, respectively (using BJTs). Cascading the LNA and mixer using a transformer to couple the two circuits together [e.g., T_1 in Fig. 21(a)] allows independent dc bias paths so that the full supply voltage (i.e., V_{CC}) appears across the active devices in each stage. Each circuit can now operate down to a minimum supply voltage of $V_{\mathrm{BE}} \approx 0.9$ V for a BJT, or even lower for a deep submicrometer MOSFET. In addition, the inductance at each port of the transformer can be resonated with the parasitics of the active circuitry to maximize the RF performance of the circuit.

If a low supply voltage is not required (note that the power amplifier and many other analog circuits require at least 3.3 V), then power can be conserved by sharing bias current between the circuit blocks operating from the same supply voltage using the transformer. This is illustrated for an LNA/mixer combination in Fig. 21(b). The supply voltage is now split between the LNA and mixer, which share the same bias current (i.e., $V_{\mathrm{CC}} = V_{\mathrm{LNA}} + V_{\mathrm{Mix}}$) because the primary and secondary windings of the transformer are connected in series. Capacitor C_D forces one terminal of the transformer to ac ground, thereby placing the LNA and mixer in cascade for the RF signal path.

D. Multifilament Transformer Application Examples

The transformer-coupled approach has the potential to realize a variety of other proven mixer topologies, including the Hartley image-reject topology shown in Fig. 22. This eliminates a passive (off-chip) filter connected between

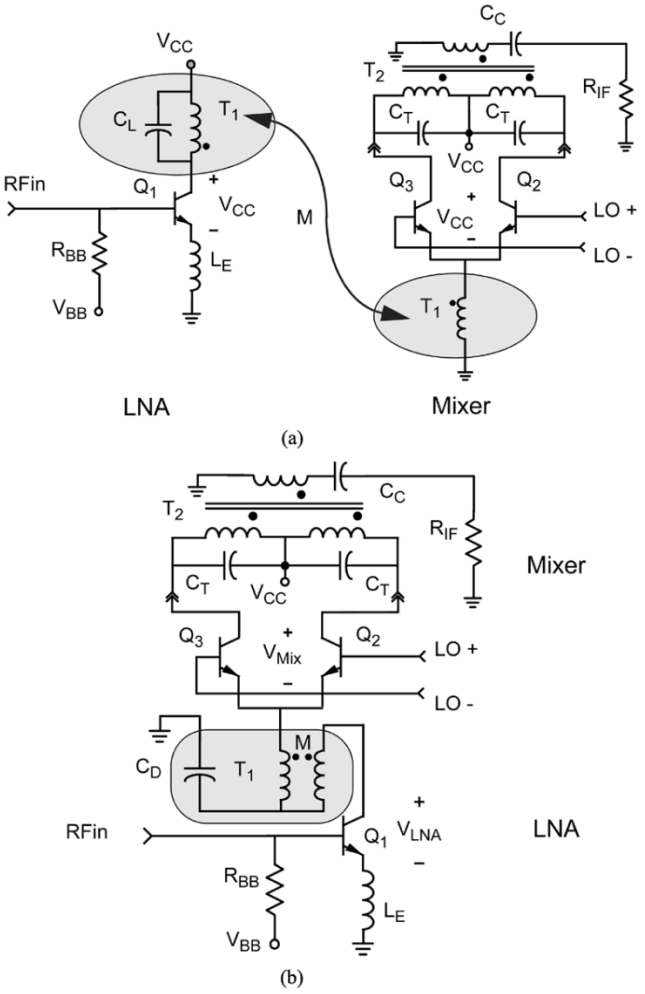


Fig. 21. Simplified schematics of independent bias and current-sharing (cascode) receiver topologies. (a) LNA/mixer cascade. (b) LNA/mixer bias cascode.

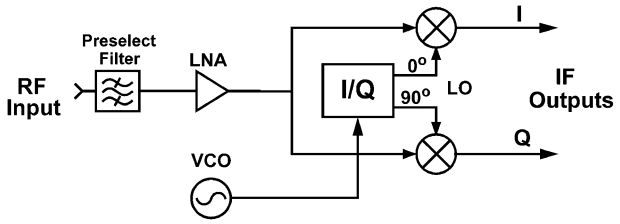


Fig. 22. Single-sideband downconverter block diagram.

the preamplifier (LNA) and the mixer needed to reject spurious signals at the "image" of the desired signal band in a heterodyne radio. When implemented monolithically, these components offer a significant advance in transceiver performance and integration level.

An image-reject receiver RF front end can be used in either a heterodyne or homodyne (i.e., direct conversion) receiver. The image-reject approach adds additional complexity to the receiver but can benefit from the potential for matching between components fabricated on the same chip to achieve good performance (e.g., 30–40 dB of image rejection). Parameter values also track each other closely with temperature and bias variations in an IC implementation.

A complete circuit schematic of a transformer-coupled image-reject receiver front end using independent bias paths for the LNA and mixers is shown in Fig. 23. The interstage coupling transformer (T_1) is a three winding (trifilar) design with a noninteger turns ratio. A single transformer with a fully differential signal path rather than two single-ended transformers is desirable in order to conserve chip area. However, the electrical center tap for each winding must be identified so that the ac grounds introduced by the supply connections to the circuit do not imbalance the RF signal path. In order to realize 15 dB of gain with a few milliamperes of bias current in the LNA, the impedance reflected from the secondary winding back to the primary winding of the transformer must be on the order of 300–400 Ω. This defines the turns ratio of the transformer, which is 4 : 1 : 1 as drawn in the physical layout (e.g., see T_1 in Fig. 1). However, the primary winding has a narrower line width (5 μm) than the secondaries (10 μm), which results in an additional step-up factor. The metal-to-metal spacing is 3 μm, and the transformer measures 300 μm on a side. The transformer has three ports: one on the primary (LNA) side and two for the mixing quads. DC bias is fed to the LNA and mixer circuits via the center taps. A program written to simulate arbitrary configurations of microstrip lines on silicon [59], [60] was used to generate a SPICE model for circuit simulations of the RF path.

Bias current for the doubly balanced mixers (Q_3–Q_{10} in Fig. 23) is selected to provide the desired load impedance for the preamplifier when reflected from secondary to the primary side of the transformer. LNA transistors Q_1 and Q_2 are driven differentially from the RF inputs. Differential drive improves the Q-factor of the trifilar transformer and emitter degeneration inductor by approximately 50% [6]. The impedance match at the RF input realizes both minimum noise figure and maximum power transfer by employing both series feedback from the emitter via inductor $L_{\rm EE}$ and shunt feedback via the collector-base (i.e., Miller) capacitance [14].

A prototype receiver designed in a 25-GHz silicon bipolar technology realized over 45 dB of image-rejection with a measured SSB (50 Ω) noise figure of 5.1 dB, IIP3 of -4.5 dBm and 17 dB conversion gain at 5.3 GHz [6]. The fully differential RF front end consumes less than 22 mW from a 2.2-V supply. A large improvement in performance over conventional RF receiver circuits was achieved in this case through the use of on-chip magnetic components. A SiGe HBT equivalent would realize even wider dynamic range with lower power consumption using the same architecture.

Perfect linkage of the magnetic flux produced by a transformer built in a production silicon VLSI technology is not possible because there is substantial flux leakage. A coupling coefficient ranging from 0.75–0.85 is possible depending upon the conductor width (w), spacing between conductors (s) and 3–4 turns on each winding (Nt) of a 1 : 1 transformer (i.e., $n = 1$). However, for step-up and step-down transformers $0.6 < k_m < 0.7$ are typical. Higher

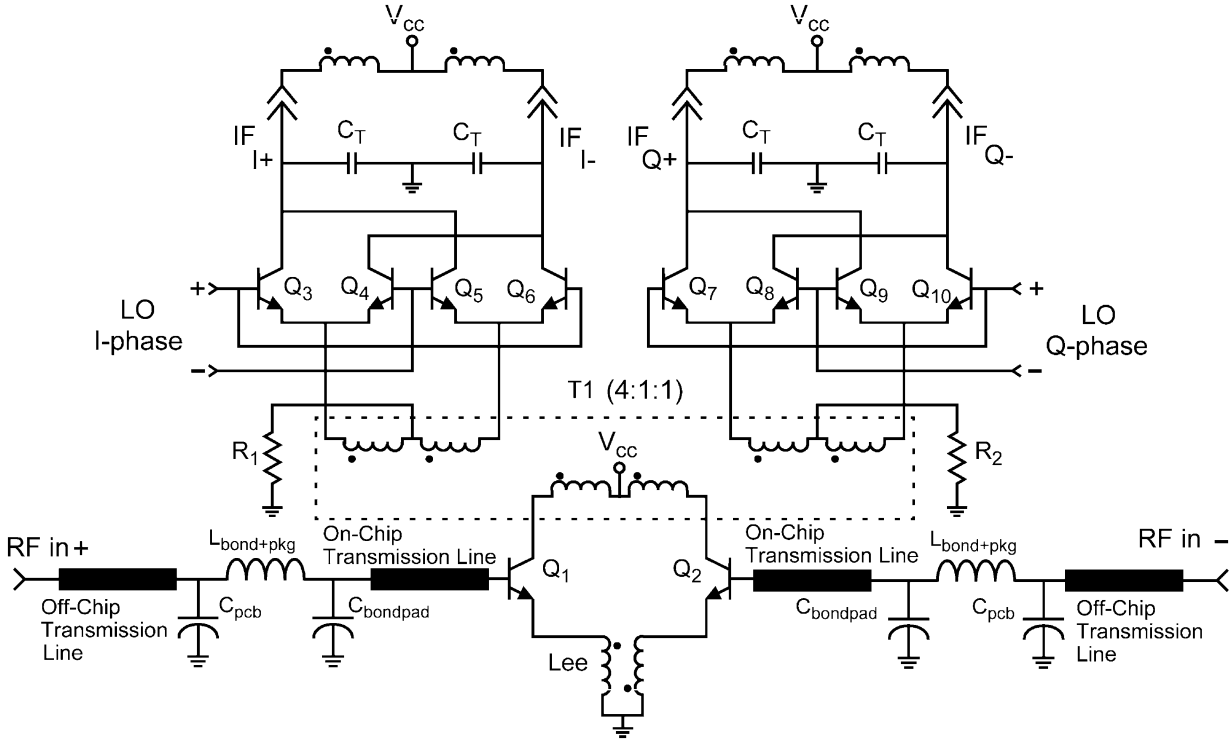


Fig. 23. Trifilar coupled RF front end schematic diagram.

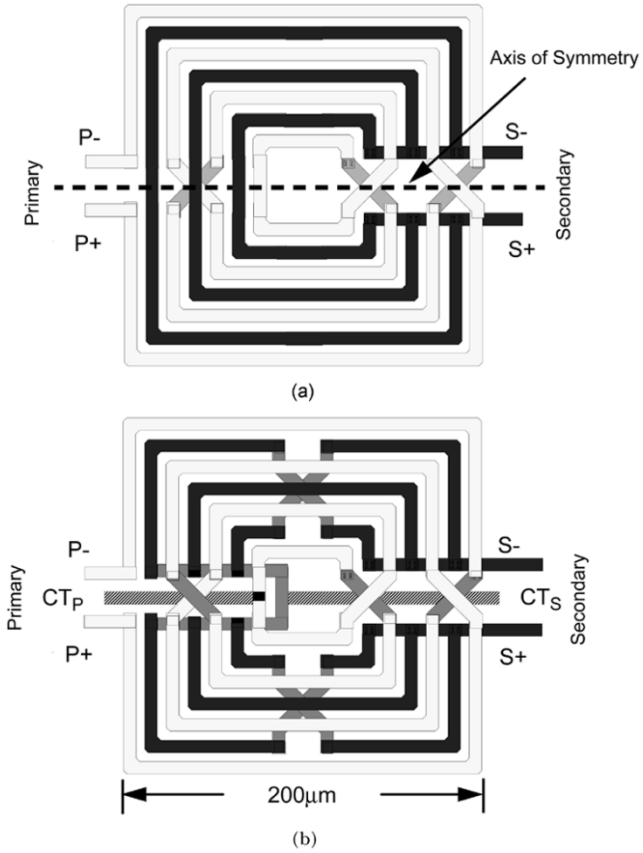


Fig. 24. Symmetrical 3 : 1 step-down transformers for the 17–17.5-GHz band. (a) Conventional winding style. (b) Balanced winding style.

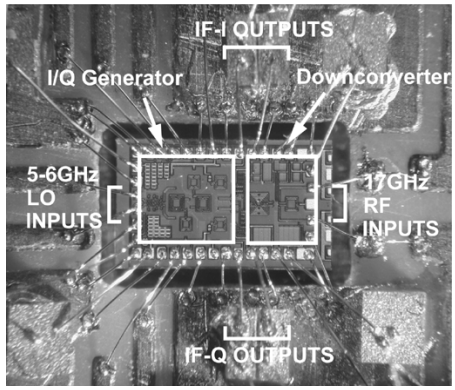


Fig. 25. HiperLINK (17–17.3 GHz) integrated RF receiver front end [26].

coupling is desirable in order to improve the efficiency of the transformers in RF circuits.

Fig. 24(a) shows the top-view of a conventional planar step-down transformer layout. The terminals of the primary winding (lightly shaded top metal) are grouped closely together on the left side and the secondary terminal (black winding) are grouped on the right. Current flowing between $P+$ and $P-$ terminals will induce a positive voltage between $S+$ and $S-$ terminals on the secondary. In order to realize a step-down between primary and secondary, the three individual turns on the secondary winding are connected in parallel, giving a $4:1$ turns ratio transformer as drawn (i.e., four turns on the primary for one effective turn on the secondary). However, the impedance of each turn on the secondary is related to the inductance and resistance per unit length of the winding, which varies from turn to turn because each turn has a different total path length. Current flow in the secondary is therefore concentrated along the innermost turns because they are shorter in total length than the outermost turn. This results in a poor magnetic coupling coefficient ($k_m = 0.6$) and inefficient ac coupling between stages in the RF receiver application.

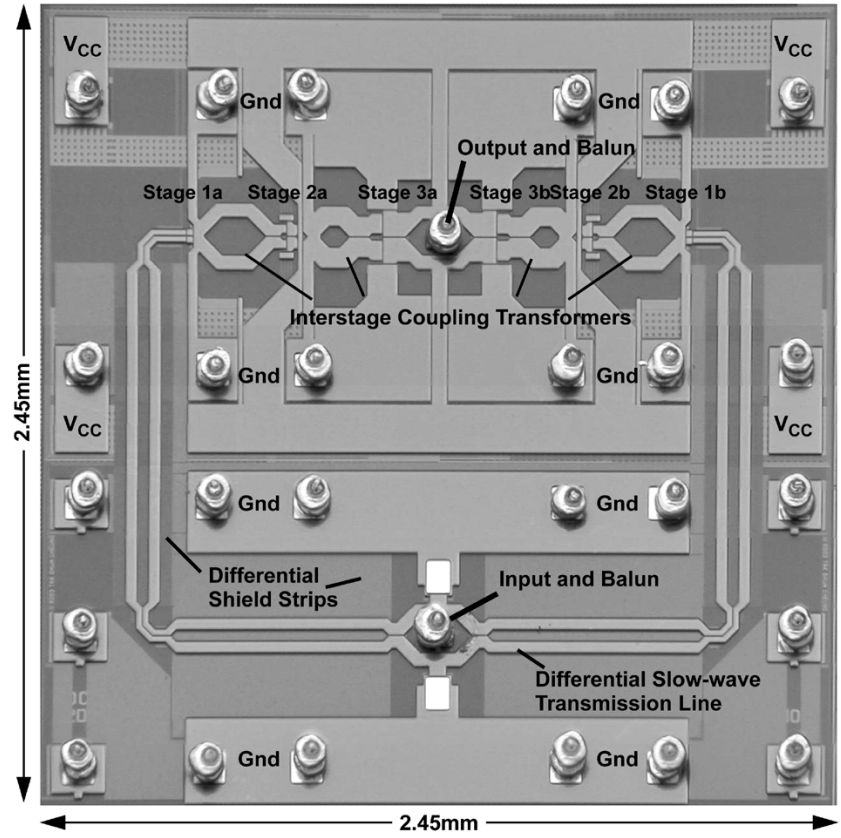


Fig. 26. SiGe 21–26-GHz power amplifier chip micrograph.

A high-k_m step-down transformer developed for a 17-GHz integrated downconverter is shown in Fig. 24(b) [26]. The primary winding is identical to the design of Fig. 24(a), however, the secondary winding style is different. Each half-turn (i.e., portions of the winding either above or below the horizontal axis of symmetry) is now equal in length. Segments of the inner and outer half-turns are connected so that the physical length of each half-turn is the same for all three turns on the secondary winding. This results in balanced inductance and resistance among all three turns in the secondary winding and an accompanying increase in primary to secondary coupling ($k_m = 0.78$, a 20% increase). The electrical turns-ratio determined from simulation is $1:0.27$ or $3.71:1$.

A 17-GHz image-reject receiver IC (i.e., LNA, two mixers and quadrature LO generator) incorporating this new transformer design is shown in Fig. 25 [26]. The image rejection of the downconverter can be optimized by adjusting the phase of the LOs generated by a subharmonically injection-locked oscillator incorporating on-chip passive delay lines. Over 75 dB of image-rejection was demonstrated by the 1.9×1.0 mm^2 IC in a production 100-GHz f_T SiGe-BiCMOS technology. It consumes a total of 62.5 mW from a 2.2-V supply.

E. A SiGe Millimeter-Wave Power Amplifier

Future broad-band wireless networks will process RF signals in the 24-GHz ISM band, where wavelengths are just a few millimeters. At present, most monolithic microwave ICs (MMICs) are fabricated in technologies which are three to five times more expensive than silicon (e.g., GaAs or InP). Silicon integration enables new applications incorporating wireless and computing technologies in products affordable to the average consumer. Operation in the 24-GHz ISM band reduces congestion in lower frequency bands and supports data services up to hundreds of megabits per second, enabling 4-G wireless access and connectivity.

A linear integrated power amplifier that operates from 21 to 26 GHz with approximately 125-mW $(+21\text{-dBm})$ output power using 1.8-V breakdown (V_{CEO}) 100-GHz f_T SiGe bipolar transistors is described in [68]. This three-stage, single-ended design uses interstage step-down transformers and input/output baluns to optimize the gain in each stage and preserve signal swing, with minimal energy loss to the medium resistivity (10–15 Ω-cm) substrate. A photomicrograph of the testchip is shown in Fig. 26.

Three stages of amplification (approximately 6 dB small-signal gain/stage) provide 15 dB gain at 1 dB gain compression per stage (i.e., large signal), and the amplifier produces full power with a 6-dBm RF input. An interstage step-down transformer ($2:1$ as drawn turns ratio) is used to maximize power transfer. The low-voltage secondary coil (emitter side) of the coupling transformer forms a self-shielding structure around the higher-voltage primary coil (collector side) to minimize substrate loss and skin effect (see Fig. 18). The input and output baluns are identical. At the input, a shielded differential transmission line connects the input balun to the first gain stage. As the output balun

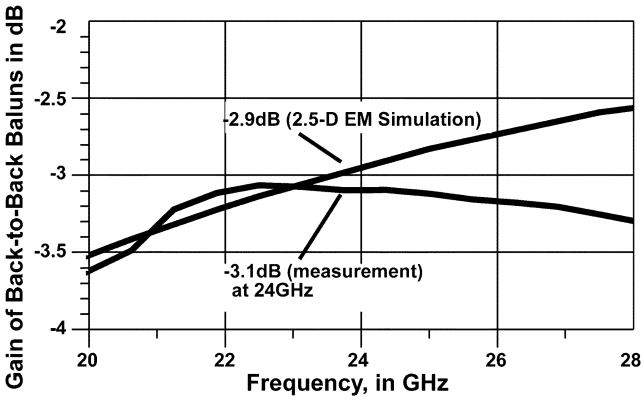


Fig. 27. Measurement versus simulation for the shielded PA balun (back-to-back test) [67].

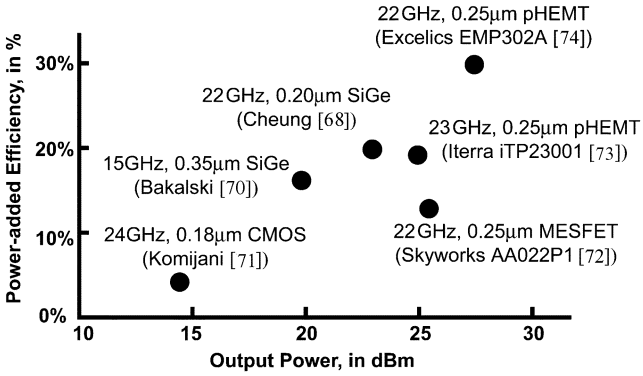


Fig. 28. Millimeter-wave power amplifier comparison.

must handle up to 400 mA of current, multiple metal layers are used with an effective width of 90- and 5-μm thickness. A plot of the measured and simulated response of the power combining (i.e., output) balun is shown in Fig. 27, where excellent agreement is seen [67]. Self-shielding is used in this balun to minimize substrate losses despite the relatively high operating frequency. A back-to-back test simplifies characterization of the balun, as the measurement equipment is single-ended. The loss of a single balun is the one-half the measured value from a back-to-back test from Fig. 27 (i.e., 3.1 dB/2, or 1.55 dB at 24 GHz).

Power added-efficiency for an amplifier is defined by the ratio of RF power added to the signal (i.e., P_{out}-P_{in}), to dc power supplied to the circuit (expressed as a percentage). It is plotted in Fig. 28 for a number of millimeter-wave amplifiers—both commercial products and examples recently reported in the literature. The 0.2-μm SiGe power amplifier with on-chip coupling baluns offers efficiency, operating bandwidth, and output power that is competitive with other technologies (e.g., GaAs pHEMT). The SiGe amplifier's maximum power-added efficiency (PAE) of 19.7% is obtained at 22 GHz, while PAE at 24 GHz is 13%. Peak output power of 23 dBm is achieved at 22 GHz, and over 20.8 dBm output power is available between 20 and 25 GHz.

F. Prospective Applications

Highly integrated radio transceiver chips that operate at low power levels over a short- to medium-range span ($<$ 1 km) are an enabling technology for remote monitoring and sensing networks. Potential applications for these chips, such as health monitoring of patients from their homes, traffic control, and security monitoring depend on the mobility inherent in wireless devices. Batch integration of high-quality RF filters, antennas, low-voltage analog, and state-of-the-art DSP electronics into the same package (i.e., an SiP) is a flexible, high-performance solution. SiGe BiCMOS technology can play an important role in this scenario because the wide range of proven devices and components available to the designer helps reduce design time and improves design quality. True single-chip (SoC) solutions, however, are more likely to be implemented in CMOS (rather than BiCMOS) technology to minimize production costs.

Many gigahertz of frequency spectrum for collision avoidance radar, point-to-point, and point-to-multipoint distribution of broad-band services could be unlocked by new RF circuit developments in the millimeter-wave band. These devices are of strategic and economic importance to companies that wish to maintain a competitive position in the expanding telecom and computer data networking markets.

Low-cost implementation of such high-speed or high-frequency functions requires integration at levels which has not been possible in the past, but offers savings in power, size, and increased reliability. Silicon is a cost-effective technology platform that could combine high-speed or RF interfaces with other signal processing capability (e.g., at baseband), resulting in compact devices with a wide operating bandwidth. While far from being the perfect technology base for millimeter-wave frequency development, previous experience (e.g., mobile telephony) has shown that silicon devices play a leading role because of their inherent low production costs. Emerging technologies such as 0.13-μm SiGe-BiCMOS processes with transistor transit frequencies (f_T) exceeding 200 GHz are a logical choice for design, as they offer volume production capability on 200- or 300-mm-diameter wafers at relatively low cost compared to other millimeter-wave semiconductor technologies (e.g., InP or other III–Vs) manufactured on more expensive 100–150-mm wafers.

Magnetic components (e.g., transmission lines, inductors, and transformers) enable operation below 1-V supply with reduced current consumption (to save battery power) and offer wider dynamic range (i.e., lower noise and distortion) at the cost of additional circuit area. These benefits are more compelling at millimeter-wave frequencies because the size of each passive element can be reduced in proportion to the shrinking wavelength. High-quality passive components with millimeter-wave bandwidth, and high-performance circuits incorporating on-chip magnetics could enable silicon ICs operating above 24 GHz that exploit emerging technologies with transistor bandwidths in excess of 100 GHz. These circuits are affected by the distributed nature of on-chip interconnects at millimeter-wave frequencies, and therefore require adaptation of impedance matching and transmission line design methodologies currently used on insulating substrates such as GaAs and InP. They can benefit

from the incorporation of interconnect such as S-CPW and differentially shielded on-chip magnetic components in their design.

Aside from broad-band data networks, automated vehicular control is an important new application area for millimeter-wave circuits. Radar ranging and wireless vehicle tagging systems have been proposed as ways to reduce fatalities and road congestion in many countries around the world. In the proposed 77/79-GHz bands, the transceiver antenna size is reduced to just a few square centimeters, allowing units to be built unobtrusively into the front bumper of a vehicle. Production volume for this application is in the millions of units per year, as one or multiple units would be built into every new automobile. The cost/volume benefits of silicon IC technology can make this happen within the next decade.

VI. Conclusion

Highly integrated low-current/low-voltage radios are essential to the integration of multiband and multistandard transceivers in a small form factor at the lowest possible cost. Where high performance is critical, SiGe technologies are used extensively.

In addition, broad-band networks that offer interactive access to information and multimedia over the Internet have revolutionized the way we do business and spend our leisure time. Wireless broad-band devices are the next step in the evolution of these networks, but economic viability again depends upon low-cost silicon microelectronic realizations.

Recent developments of 100–200-GHz SiGe technologies bring high-performance applications (both broad-band and narrow-band) within our reach, but the bandwidth capabilities of silicon devices will again be pushed to the point where analog effects dominate. It is expected that on-chip passive components are essential as operating frequencies exceed 10 GHz, because of limitations present in scaling on-chip interconnect technology. New passive element designs, models, and circuit design techniques are needed to cross this threshold even as the f_T from active devices like SiGe bipolar transistors eclipse the 200-GHz barrier. Successful demonstration of SiGe bipolar and BiCMOS circuit implementations well into the millimeter-wave frequencies will open up many new avenues for creative engineering and product development.

Acknowledgment

Fabrication services were provided by Nortel and IBM Microelectronics. The author would like to thank R. Hadaway of Nortel Networks, Ottawa, ON, Canada, and D. Harame of IBM Microelectronics, Burlington, VT, for fabrication support. The author would also like to acknowledge the research contributions of former graduate students D. Cheung, A. Chung, M. Danesh, D. Ma, and S. Yue, and their help in preparing illustrations, and W. Straver, R. Klerks, and L. van Schie at the Delft University of Technology, Delft, The Netherlands, for providing technical support for packaging and testing.

References

[1] A. A. Abidi, P. R. Gray, and R. G. Meyer, *Integrated Circuits for Wireless Communications*. Piscataway, NJ: IEEE Press, 1999.

[2] A. A. Abidi, G. J. Pottie, and W. J. Kaiser, "Power conscious design of wireless circuits and systems," *Proc. IEEE*, vol. 88, pp. 1528–1545, Oct. 2000.

[3] T. H. Lee and S. S. Wong, "CMOS rf integrated circuits at 5 GHz and beyond," *Proc. IEEE*, vol. 88, pp. 1560–1571, Oct. 2000.

[4] B. A. Orner *et al.*, "A 0.13 μm BiCMOS technology featuring a 200/280 GHz ($f_T/f_{\max}$) SiGe HBT," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, 2003, pp. 203–206.

[5] K. Biesecker, "The promise of broad-band wireless," *IEEE IT Prof.*, vol. 2, no. 6, pp. 31–39, Nov./Dec. 2000.

[6] J. P. Maligeorgos and J. R. Long, "A 5.1–5.8 GHz image-reject receiver with wide dynamic range," *IEEE J. Solid-State Circuits*, vol. 35, no. 12, pp. 1917–1926, Dec. 2000.

[7] K. Gard, K. Barnett, J. Dunworth, T. Segoria, B. Walker, J. Zhou, D. Maldonado, A. See, and C. Persico, "Direct conversion dual-band SiGe BiCMOS transmitter and receive PLL IC for CDMA/WCDMA/AMPS/GPS applications," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2003, pp. 272–273.

[8] E. Duvivier, S. Cipriani, L. Carpineto, P. Cusinato, B. Bisanti, F. Galant, F. Chalet, F. Coppola, S. Cercelaru, G. Puccio, N. Mouralis, and J. C. Jiguet, "A fully integrated zero-IF transceiver for GSN-GPRS quad band application," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2003, pp. 274–275.

[9] V. Leung, L. Larson, and P. Gudem, "Digital-IF WCDMA handset transmitter IC in 0.25 μm SiGe BiCMOS," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2004, pp. 182–183.

[10] (2004) 2003 International Technology Roadmap for Semiconductors. Semiconductor Industry Association (SIA). [Online]. Available: http://www.semichips.org

[11] J. Sevenhans, F. Op't Eynde, and P. Reusens, "The silicon radio decade," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 1, pp. 235–244, Jan. 2002.

[12] H. Fukui, "The noise performance of microwave transistors," *IEEE Trans. Electron Devices*, vol. ED-13, no. 3, pp. 329–341, Mar. 1966.

[13] S. P. Voinigescu, M. C. Maliepaard, J. L. Showel, G. E. Babcock, D. Marchesan, M. Schroeter, P. Schvan, and D. L. Harame, "A scalable high-frequency noise model for bipolar transistors with application to optimal transistor sizing for low-noise amplifier design," *IEEE J. Solid-State Circuits*, vol. 32, no. 9, pp. 1430–1439, Sep. 1997.

[14] J. R. Long, "A low-voltage 5.5–5.8 GHz image-reject downconverter RF IC," *IEEE J. Solid-State Circuits*, vol. 35, no. 9, pp. 1320–1328, Sep. 2000.

[15] A. A. Abidi, "Low-power radio-frequency ICs for portable communications," *Proc. IEEE*, vol. 83, no. 4, pp. 544–569, Apr. 1995.

[16] S. Reynolds, B. Floyd, U. Pfeiffer, and T. Zwick, "60 GHz transceiver circuits in SiGe bipolar technology," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2004, pp. 442–443.

[17] C. H. Doan, S. Emami, A. M. Niknejad, and R. W. Broderson, "Design of CMOS for 60 GHz applications," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2004, pp. 440–441.

[18] S. A. Maas, *Nonlinear Microwave and RF Circuits*. Boston, MA: Artech House, 1997.

[19] V. Aparin and C. Persico, "Effect of out-of-band terminations on intermodulation distortion in common-emitter circuits," in *Proc. IEEE Int. Microwave Symp.*, 1999, pp. 977–980.

[20] K. L. Fong, "High-frequency analysis and linearity improvement technique of common-emitter transconductance stage using a low-frequency-trap network," *IEEE J. Solid-State Circuits*, vol. 35, no. 8, pp. 1249–1252, Aug. 2000.

[21] V. Aparin and L. E. Larson, "Modified derivative superposition method for linearizing FET low-noise amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 2, pp. 571–581, Feb. 2005.

[22] J. R. Long and M. A. Copeland, "A 1.9 GHz low-voltage silicon bipolar RF front-end for wireless personal communications systems," *IEEE J. Solid-State Circuits*, vol. 30, no. 12, pp. 1438–1448, Dec. 1995.

[23] J. R. Long, M. Copeland, S. Kovacic, D. Malhi, and D. L. Harame, "RF analog and digital circuits in SiGe technology," *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pp. 82–83, 1996.

[24] V. Aparin, P. Gazzerro, J. Shou, B. Sun, S. Szabo, E. Zeisel, T. Segoria, S. Ciccarelli, C. Persico, C. Narathong, and R. Sridhara, "A highly-integrated tri-band/quad-mode SiGe BiCMOS RF-to-baseband receiver for wireless CDMA/WCDMA/AMPS applications with GPS capability," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2002, pp. 234–235.

[25] M. Tanabe, T. Fukuda, and K. Nishii, "Step gain amplifier with impedance unchanged in gain control," in *Proc. IEEE Radio Frequency Integrated Circuits Symp.*, 2002, pp. 209–212.
[26] S. Y. Yue, D. Ma, and J. R. Long, "A 17.1–17.3-GHz image-reject downconverter with phase-tunable LO using 3X subharmonic injection locking," *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2321–2332, Dec. 2004.
[27] S. Kumar, M. Vice, H. Morkner, and W. Lam, "Enhancement mode GaAs PHEMT LNA with linearity control (IP3) and phased matched mitigated bypass switch and differential active mixer," in *Proc. IEEE Int. Microwave Symp.*, 2003, pp. 1577–1580.
[28] P. Leroux, M. Steyaert, V. Vassilev, and H. Maes, "A 1.3 dB NF CMOS LNA for GPS and 3 kV HBM ESD-protection," in *Proc. ESSCIRC*, 2002, pp. 335–338.
[29] J. Durec, "An integrated silicon bipolar receiver subsystem for 900-MHz ISM band applications," *IEEE J. Solid-State Circuits*, vol. 33, no. 9, pp. 1352–1372, Sep. 1998.
[30] X. Li, T. Brogan, M. Esposito, B. Meyers, and K. K. O, "A comparison of CMOS and SiGe LNAs and mixers for wireless LAN application," in *Proc. IEEE CICC*, 2001, pp. 531–534.
[31] D. B. Leeson, "A simple model of feedback oscillator noise spectrum," *Proc. IEEE*, vol. 54, no. 2, pp. 329–330, Feb. 1966.
[32] J. N. Burghartz, D. C. Edelstein, M. Souyer, H. A. Ainspan, and K. A. Jenkins, "RF circuit aspects of spiral inductors on silicon," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 2028–2034, Dec. 1998.
[33] M. Danesh and J. R. Long, "Differentially-driven symmetric spiral inductors," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 1, pp. 332–341, Jan. 2002.
[34] T. S. D. Cheung and J. R. Long, "Differentially shielded monolithic inductors," in *Proc. IEEE CICC*, 2003, pp. 95–98.
[35] G. G. Rabjohn, "Monolithic microwave transformers," M.Eng. thesis, Carleton Univ., Ottawa, ON, Canada, 1991.
[36] D. Coolbaugh, E. Eshun, R. Groves, D. Harame, J. Johnson, M. Hammad, Z. He, V. Ramachandran, K. Stein, S. St. Onge, S. Subbanna, D. Wang, R. Volant, X. Wang, and K. Watson, "Advanced passive devices for enhanced integrated RF circuit performance," in *IEEE RFIC Symp. Tech. Dig.*, 2002, pp. 341–344.
[37] P. Andreani and S. Mattisson, "On the use of MOS varactors in RF VCOs," *IEEE J. Solid-State Circuits*, vol. 35, no. 6, pp. 905–910, Jun. 2000.
[38] J. W. M. Rogers, J. A. Macedo, and C. Plett, "The effect of varactor nonlinearity on the phase noise of completely integrated VCOs," *IEEE J. Solid-State Circuits*, vol. 35, no. 9, pp. 1360–1367, Sep. 2000.
[39] N. H. W. Fong, J.-O. Plouchart, N. Zamdmer, L. Duixian, L. F. Wagner, C. Plett, and N. G. Tarr, "A 1-V 3.8 –5.7-GHz wide-band VCO with differentially tuned accumulation MOS varactors for common-mode noise rejection in CMOS SOI technology," *IEEE Trans. Microw. Theory Tech.*, vol. 51, no. 8, pp. 1952–1959, Aug. 2003.
[40] J.-M. Mourant, J. Imbornone, and T. Tewksbury, "A low phase noise monolithic VCO in SiGe BiCMOS," in *Proc. IEEE Radio Frequency Integrated Circuits Symp.*, 2000, pp. 65–68.
[41] M. A. Margarit, J. L. Tham, R. G. Meyer, and M. J. Deen, "A low-noise, low-power vco with automatic amplitude control for wireless applications," *IEEE J. Solid-State Circuits*, vol. 34, no. 6, pp. 761–771, Jun. 1999.
[42] M. Zannoth, J. Fenk, A. Springer, and R. Weigel, "A single-chip Si-bipolar 1.6 GHz VCO with integrated bias network," in *Proc. IEEE Radio Frequency Integrated Circuits Symp.*, 1999, pp. 117–120.
[43] M. Soyuer, J. N. Burghartz, H. A. Ainspan, K. A. Jenkins, P. Xiao, A. R. Shahani, M. S. Dolan, and D. L. Harame, "An 11-GHz 3-V SiGe voltage controlled oscillator with integrated resonator," *IEEE J. Solid-State Circuits*, vol. 32, no. 9, pp. 1451–1454, Sep. 1997.
[44] H. Li, H.-M. Rein, R. Kreienkamp, and W. Klein, "47 GHz VCO with low phase noise fabricated in a SiGe bipolar production technology," *IEEE Microw. Compon. Lett.*, vol. 12, no. 3, pp. 79–81, Mar. 2002.
[45] F. Svelto, S. Deantoni, and R. Castello, "A 1 mA, −120.5 dbc/hz at 600 kHz from 1.9 GHz fully tunable *LC* CMOS VCO," in *Proc. IEEE CICC*, 2000, pp. 577–580.
[46] B. DeMuer, N. Itoh, M. Borremans, and M. Steyaert, "A 1.8 GHz highly-tunable low-phase-noise CMOS VCO," in *Proc. IEEE CICC*, 2000, pp. 585–588.
[47] A. Gruhle, "Prospects for 200 Ghz on silicon with SiGe heterojunction bipolar transistors," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, 2001, pp. 19–25.
[48] P. Smulders, "Exploiting the 60 GHz band for local wireless multimedia access: Prospects and future directions," *IEEE Commun. Mag.*, vol. 40, no. 1, pp. 140–147, Jan. 2002.
[49] H. Hasegawa, M. Furukawa, and H. Yanai, "Properties of microstriplines on Si-SiO_2 system," *IEEE Trans. Microw. Theory Tech.*, vol. 19, no. 11, pp. 869–881, Nov. 1971.
[50] E. Valletta, J. v. Beek, A. den Dekker, N. Pulsford, H. Jos, L. de Vreede, L. Nanver, and J. Burghartz, "Design and characterization of integrated passive elements on high ohmic silicon," in *Proc. Int. Microwave Symp.*, 2003, pp. 1235–1238.
[51] J. M. Rabaey *et al.*, "Pico radios for wireless sensor networks—the next challenge in ultra-low power design," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2002, pp. 200–201.
[52] I. O'Donnell *et al.*, "An integrated, low-power, ultra-wideband transceiver architecture for low-rate, indoor wireless systems," presented at the IEEE CAS Workshop Wireless Communications and Networking, Pasadena, CA, 2002.
[53] T. S. D. Cheung, J. R. Long, K. Vaed, R. Volant, A. Chinthakindi, C. M. Schnabel, J. Florkey, and K. Stein, "On-chip interconnect for mm-wave applications using an all-copper technology and wavelength reduction," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2003, pp. 396–397.
[54] D. Pozar, *Microwave Engineering*, 2nd ed. New York: Wiley, 1998, ch. 7.
[55] J. Zhang, M. K. Jackson, and J. R. Long, "Millimeter-wave characteristics of SiGe heterojunction bipolar transistors and monolithic interconnects in silicon technologies," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2001, pp. 274–275.
[56] M. Case, "Nonlinear transmission lines for picosecond pulse impulse and millimeter-wave harmonic generation," Ph.D. dissertation, Univ. California, Santa Barbara, 1993.
[57] E. Afshari and A. Hajimiri, "Non-linear transmission lines for pulse shaping in silicon," in *Proc. IEEE CICC*, 2003, pp. 91–94.
[58] T. T. Y. Wong, *Fundamentals of Distributed Amplification*. Norwood, MA: Artech House, 1993.
[59] J. R. Long and M. A. Copeland, "Modeling of monolithic inductors and transformers for silicon RFIC design," in *Proc. IEEE MTT-S Int. Symp. Technologies for Wireless Applications*, 1995, pp. 129–134.
[60] ——, "The modeling, characterization and design of monolithic inductors for silicon RF ICs," *IEEE J. Solid-State Circuits*, vol. 32, no. 3, pp. 357–369, Mar. 1997.
[61] A. M. Niknejad and R. G. Meyer, "Analysis, design, and optimization of spiral inductors and transformers for Si RF ICs," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1470–1481, Oct. 1998.
[62] J. R. Long, "Monolithic transformers for silicon RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, no. 9, pp. 1368–1382, Sep. 2000.
[63] E. Frlan, S. Meszaros, M. Cuhaci, and J. Wight, "Computer-Aided design of square spiral transformers and inductors," *Proc. IEEE MTT-S*, pp. 661–664, 1989.
[64] D. Cassan and J. R. Long, "A 1 V transformer-feedback low noise amplifier for 5–6 GHz WLAN in 0.18 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 427–435, Mar. 2003.
[65] K.-C. Kwok and H. Luong, "A 0.35 V 1.46 mW low-phase-noise oscillator with transformer feedback in standard 0.18 μm CMOS process," in *Proc. IEEE CICC*, 2003, pp. 551–554.
[66] W. Simbuerger, H.-D. Wohlmuth, P. Weger, and A. Heinz, "A monolithic transformer coupled 5 W silicon power amplifier with 59% PAE at 0.9 GHz," *IEEE J. Solid-State Circuits*, vol. 34, no. 12, pp. 1881–1892, Dec. 1999.
[67] T. S. D. Cheung, J. R. Long, Y. V. Tretiakov, and D. L. Harame, "A 21–27 GHz self-shielded 4-way power-combining PA balun," in *Proc. IEEE CICC*, 2004, pp. 617–620.
[68] T. S. D. Cheung, J. R. Long, and D. L. Harame, "A 21 to 26 GHz SiGe bipolar PA MMIC," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2005, pp. 530–531.
[69] I. Aoki, S. D. Kee, D. B. Rutledge, and A. Hajimiri, "Fully integrated CMOS power amplifier design using the distributed active-transformer architecture," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 371–383, Mar. 2002.

[70] W. Bakalski, A. Vasyleyev, W. Simbuerger, R. Thueringer, H.-D. Wohlmuth, A. L. Scholtz, and P. Weger, "A fully integrated 7–18 GHz power amplifier with on-chip output balun in 75 GHz-fT SiGe-bipolar," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, 2003, pp. 61–64.
[71] A. Komijani and A. Hajimiri, "A 24 GHz, +14.5 dBm fully-integrated power amplifier in 0.18 μm CMOS," in *Proc. IEEE CICC*, 2004, pp. 561–564.
[72] *18–23 GHz GaAs MESFET MMIC Power Amplifier AA022P1 Data Specification*, 2002.
[73] *21–24 GHz Power Amplifier MMIC iTP23001 Data Specification*, 2004.
[74] *21–24 GHz Power Amplifier MMIC EMP302A Data Specification*, 2003.
[75] M. Park, S. Lee, H. K. Yu, J. G. Koo, and K. S. Nam, "High Q CMOS-compatible microwave inductors using double-metal interconnection silicon technology," *IEEE Microw. Guided Wave Lett.*, vol. 7, no. 2, pp. 45–47, Feb. 1997.
[76] J. N. Burghartz, D. C. Edelstein, M. Souyer, H. A. Ainspan, and K. A. Jenkins, "RF circuit aspects of spiral inductors on silicon," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 2028–2034, Dec. 1998.
[77] C. P. Yue and S. S. Wong, "On-Chip spiral inductors with patterned ground shields for Si-based RF ICs," *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 743–752, May 1998.
[78] C.-Y. Chi and G. M. Rebeiz, "Planar microwave and millimeter-wave lumped elements and coupled-line filters using micro-machining techniques," *IEEE Trans. Microw. Theory Tech.*, vol. 43, no. 4, pp. 730–738, Apr. 1995.

John R. Long received the B.Sc. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 1984 and the M.Eng. and Ph.D. degrees in electronics from Carleton University, Ottawa, ON, Canada, in 1992 and 1996, respectively.

He was employed for ten years by Bell-Northern Research, Ottawa (now Nortel Networks R&D), involved in the design of ASICs for Gb/s fiber-optic transmission systems, and from 1996 to 2001 as an Assistant and then Associate Professor at the University of Toronto, Toronto, ON, Canada. Since January 2002, he has been chair of the Electronics Research Laboratory, Delft University of Technology, Delft, The Netherlands. His current research interests include low-power transceiver circuitry for highly integrated wireless applications, and electronics design for high-speed data communications systems.

Prof. Long received the NSERC Doctoral Prize and Douglas R. Colton and Governor General's Medals for research excellence, and Best Paper Awards from the International Solid-State Circuits Conference (ISSCC) 2000 and the IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM) 2003. He currently serves as technical program chair for the 2005 IEEE BCTM. He also chairs the Analog/RF circuits subcommittee for the 2006 ISSCC, and is a member of the technical program committees for the European Solid-State Circuits (ESSCIRC) and GAAS2005 (EuMW) conferences. He is a former Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS.