1 Introduction

1.1 Motivation

The history of electronics has been inextricably linked with the growth of the communications industry. Electronic communication served as a major enabling technology for the industrial revolution. When scientists and engineers learned to control electricity and magnetism, it did not take long for people to realize that the electromagnetic force would enable long-range communication. Even though the basic science of Maxwell's equations was well understood, it took much longer for practical applications to fully exploit all the fantastic possibilities such as radio, television, and personal wireless communication.

At first only crude wires carrying telegraph signals were rolled out sending Morse code,[1] digital signals at speeds limited by human operators. In this regard it is ironic that digital communication predates analog communication. Telegraph wires were laid alongside train tracks, making long-range communication and transportation a practical reality. Sending signals faster and further ignited the imagination of engineers of the time and forced them to study carefully and understand the electromagnetic force of nature. Today we are again re-learning and inventing new digital and analog communication systems that are once again compelling us to return to the very fundamental science of electricity and magnetism.

The topic of this book is the high-frequency electromagnetic properties of passive and active devices. For the most part, passive devices are resistors, capacitors, transformers, and inductors, while active devices are transistors. Most applications we draw from are high-frequency circuits. For example, radio frequency (RF) circuits and high-speed digital circuits both depend on a firm understanding of passive devices and the environment in which they operate.

Circuit theory developed as an abstraction to electromagnetics. Circuit theory is in effect the limit of electromagnetics for a circuit with negligible dimension. This allows spatial variations and time delay to be ignored in the analysis of the circuit. As such, it allowed practicing engineers to forego solving Maxwell's equations and replaced them with simple concepts such as KCL and KVL. Even differential equations were eliminated and replaced with algebraic equations by employing Laplace transforms. The power and popularity of circuit theory was due to its simplicity and abstraction. It allowed generations of engineers to

[1] Or as Paul Nahin suggests in [41] we should more correctly call this "Vail" code.

solve difficult problems with simple and yet powerful tools. In effect, it allowed generations of engineers to forego reading a book such as this one.

So why read another book on electromagnetics? Why bother learning all this seemingly complicated theory when your ultimate goal is to build circuits and systems for communication and information management?

We live today at the intersection of several interesting technologies and applications. Integrated circuit technology has enabled active devices to operate at increasingly higher frequencies, turning low-cost Si technology into a seemingly universal panacea for a wide array of applications. CMOS digital circuits are switching at increasingly higher rates, pushing multi GHz operation. Si CMOS, bipolar, and SiGe technology have also enabled a new class of low-cost RF and microwave devices, with ubiquitous deployment of cellular phones in the 800 MHz–2 GHz spectrum, and high-speed wireless LAN in the 2–5 GHz bands. There seems to be very little in the way of enabling Si technology to exploit the bandwidths up to the limits of the device technology. In a present-day digital 130 nm CMOS process, for instance, circuits are viable up to 60 GHz [55] [13].

At the same time, wired communication is pushing the limits. Gigabit Ethernet and high-speed USB cables are now an everyday reality, and people are already pursuing a 10 Gb/s solution. Optical communication is of course at the forefront, with data rates in the 40 Gb/s range now commercially viable and at relatively low cost.

The simultaneous improvement in active device technology, miniaturization, and a host of new applications are the driving force of today's engineering. As integrated circuits encompass more functionality, many traditionally off-chip components are pushed into the IC or package, blurring the line between active devices and circuits and passive devices and electromagnetics. This is the topic of this book.

Technology enhancements

The limitations in frequency and thus speed of operation is usually set by the active device technology. One common figure of merit for a technology is, fr, the unity-gain frequency f_T, the frequency at which the short-circuit current gain of the device crosses unity. Another important figure of merit is f_{max} the maximum frequency of oscillation, or equivalently the frequency where the maximum power gain of a transistor drops to unity. Since f_{max} is a strong function of layout and parasitics in a process, it is less often employed. In contrast, the f_T depends mostly on the dimensions of the transistor and the transconductance

$$f_T = \frac{1}{2\pi}\frac{g_m}{C_\pi + C_\mu} \tag{1.1}$$

It can be shown [50] that the device f_T is inversely related to the transistor dimensions. For a long-channel MOSFET the key scaling parameter is the channel length L

$$f_T \approx \frac{1}{2\pi}\frac{3\mu(V_{gs} - V_t)}{2L^2} \tag{1.2}$$

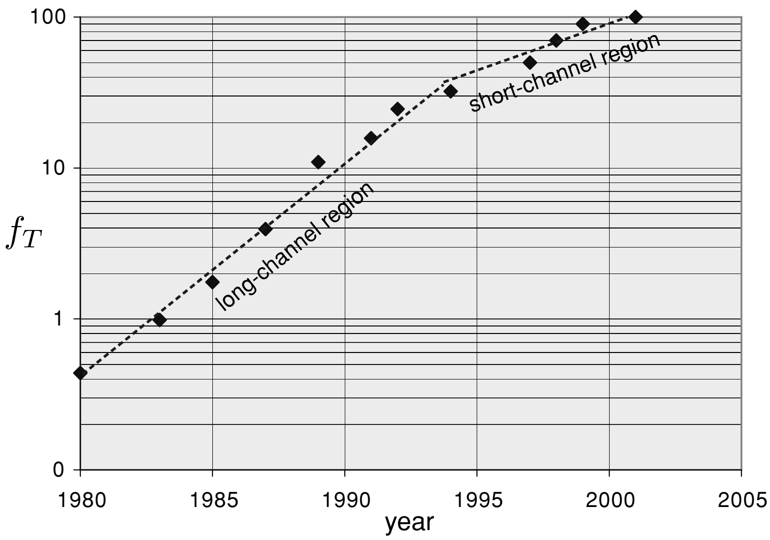

Figure 1.1 The improvements in device unity–gain frequency f_T over the past two decades due to device scaling.

while in the limit for short channel transistors the scaling changes to L^{-1} since the current is limited by velocity saturation

$$I_{ds,sat} = W Q_i v_{sat} = W C_{ox}(V_{gs} - V_t) v_{sat} \tag{1.3}$$

resulting in

$$f_T \propto \frac{v_{sat}}{L} \tag{1.4}$$

For a bipolar junction transistor (BJT) the critical dimension is the base width. In the limit that base transit limits the frequency of operation

$$f_T \approx \frac{1}{2\pi}\frac{1}{\tau_B} \propto \frac{1}{W_B^2} \tag{1.5}$$

As integrated circuit manufacturing technology has improved exponentially in the past three decades, so has the f_T of the device, giving circuit designers increasingly faster devices. A plot of the device f_T over the years for a MOSFET device is shown in Fig. 1.1, and the exponential growth in technological advancements can be seen clearly.

It is important to note that this improvement in performance only applies to the intrinsic device. Early circuits were in fact limited by the intrinsic transistor and not the parasitic routing and off-chip environment. As circuit technology has advanced, though, the situation has reversed and now the limitation is set by the parasitics of the chip and board environment, as well as the performance of the passive devices. This is why the material of this book is now particularly relevant. It can be shown that a good approximation to the CMOS device f_{max} is given by [42]

$$f_{max} \approx \frac{f_T}{2\sqrt{R_g(g_m C_{gd}/C_{gg}) + (R_g + r_{ch} + R_s)g_{ds}}} \tag{1.6}$$

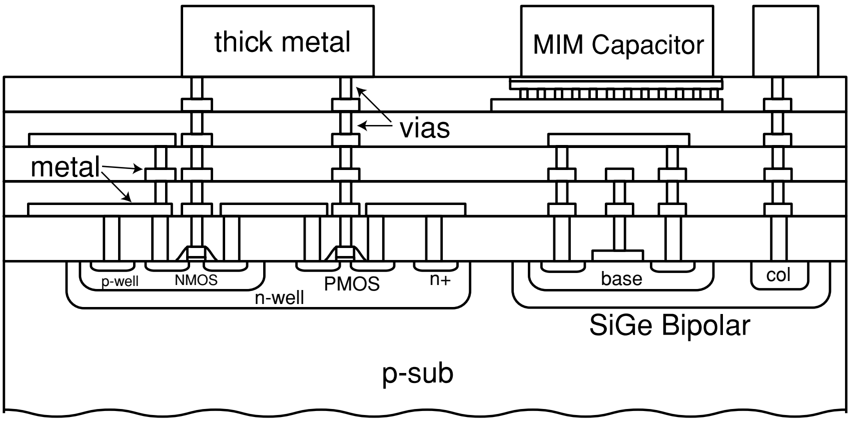

Figure 1.2 Cross section of a SiGe BiCMOS process.

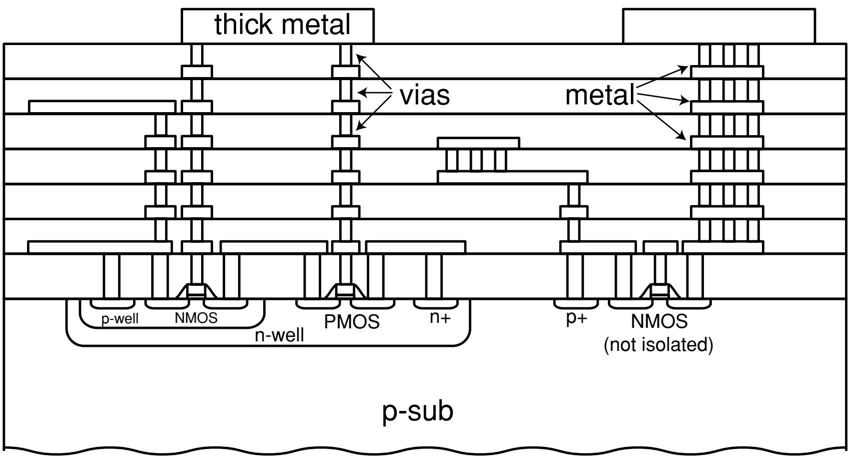

Figure 1.3 Cross section of an advanced CMOS process.

where the device performance is a strong function of the loss, such as the drain/source resistance R_s, R_d, and the gate resistance R_g. These parasitics are in large part determined by layout and the process technology.

While early integrated circuit technologies were limited to a few types of different active devices and a few layers of aluminum interconnect metal, present-day process technology has a rich array of devices and metal routing. In an advanced Si process, shown in Fig. 1.2, high-performance SiGe HBT devices are complemented by MOS and PN-junction varactors, metal-insulator-metal (MIM) high-density and high-quality capacitors, and thick-metal for low-loss interconnect and inductors/transformers. Even a digital CMOS process, as shown in Fig. 1.3, has many advanced capabilities. In addition to several flavors of MOS active devices (fast thin oxide, thick oxide, high/low V_T), there are also enhanced isolation structures and triple-well (deep n-well) devices, and many layers of interconnect that allow construction of high-quality, high-density capacitors and reasonably high-quality inductors.

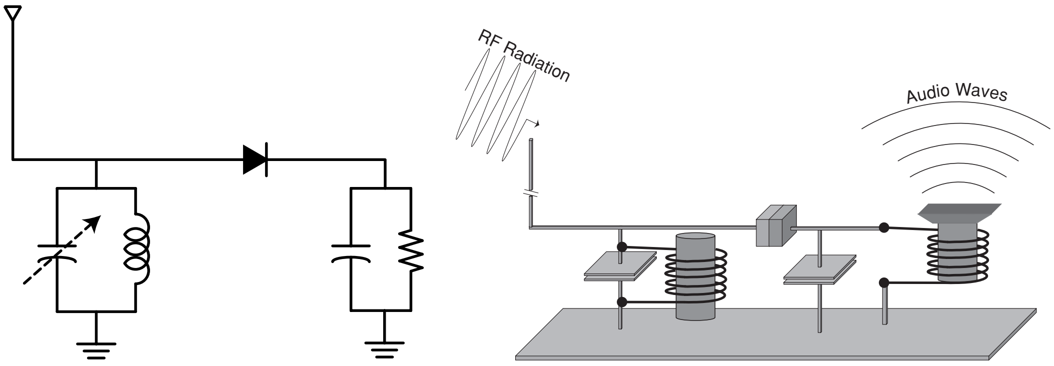

Figure 1.4 (a) A simple AM receiver circuit. The resistor represents a high-input impedance earphone. (b) A physical realization of the simple AM receiver circuit.

Radio and wireless communication

Early radio systems were essentially all passive. To see this look into the back of an old radio where a few active devices (vacuum tubes or transistors) are surrounded by tens to hundreds of passive devices. Consider the circuit diagram of a very simple AM receiver shown in Fig. 1.4a. The antenna drives a resonant tank tuned to the center frequency of the transmitting station. This signal is fed into a peak detector that follows the peak of the RF signal. The low-pass filter time constant is only fast enough to follow the low-frequency audio signal (generically the baseband signal) and yet too slow to follow the RF, thus removing the RF signal and retaining the low-frequency audio. This received signal is usually too weak to drive a speaker but can be heard through a sensitive headphone. A simple audio amplifier can be used to strengthen the signal.

It is interesting to note that this AM receiver can be physically realized by merely using contacts between a few different pieces of metal and semiconductors. This is shown in Fig. 1.4b. The resonant tank is simply a piece of wire wound into a coil which contacts with the capacitor, two metal plates in close proximity. The diode can be realized as the junction of a metal and semiconductor. Finally, to convert electric energy into sound we can use another large inductor coil and use the time-varying magnetic force to move a paper thin cone driven by a magnetic core. Magnetic materials have been known since ancient times and therefore since the metal age we have had the capability to build radio receivers! In fact, it is not surprising that radios often crop up accidentally.[2]

Most modern radios operate based on an architecture invented by Edwin Armstrong. The block diagram of such a system, called a super-heterodyne receiver, is shown in Fig. 1.7. This receiver incorporates a local oscillator (LO), a block that primarily converts DC power into RF power at the oscillation frequency. A mixer takes the product of this signal and the

[2] For instance my old answering machine also picked up the radio. Sometimes you could hear it as you were waiting for the tape recorder to rewind. At least this was a desirable parasitic radio.

signal received by the antenna. Recall the following trigonometric identity[3]

$$2\cos(\omega_{L0}t)\cos(\omega_{RF}t) = \cos((\omega_{L0} + \omega_{RF})t) + \cos((\omega_{L0} - \omega_{RF})t) \tag{1.7}$$

Note that the product of the received RF signal and the local oscillator signal produces two new signals, one centered at the difference frequency and one centered at the sum frequency. If we put a bandpass filter at one of these frequencies, call it the intermediate frequency, IF, we can electronically tune the radio by simply changing the LO frequency. This is accomplished by using a frequency synthesizer (a PLL or phase locked loop), and thus we avoid building a variable filter common to the early radios. The important point is that the IF is fixed and we can build a very selective filter to pinpoint our desired signal and to reject everything else. Why not simply set LO equal to RF to move everything to DC? This is in fact the direct-conversion or zero-IF architecture. It has some shortcomings such as problems with DC offset,[4] but its main advantage is that it lowers the complexity of the RF section of a typical radio.

At the heart of the frequency synthesizer is the voltage controlled oscillator (VCO). The VCO is an oscillator where the output frequency is a function of a control voltage or current.[5] To build a VCO we need a way to change the center frequency of a resonant tank. The resonant tank is simply an inductor in series or in parallel with a capacitor. One typical realization is to use varactor, a variable capacitor. A reversed biased diode serves this purpose nicely, as the depletion region width, and thus the small-signal capacitance, is a function of the reverse bias. It seems that a super-heterodyne receiver has simply moved the variable resonant tank from the antenna front end to a variable resonant tank in the VCO! Have we gained anything? Yes, because the frequency of the VCO can be controlled precisely in a feedback loop (using an accurate frequency reference such as a crystal), eliminating any problems associated with absolute tolerances in components in addition to drift and temperature variation.

The radio has once again emerged as a critical application of passive devices spawned by the growth and popularity of wireless telephones, in particular the cellular phone. By limiting the transmitter powers and taking advantage of spatial diversity (re-using the same frequency band for communication for points far removed – for non-adjacent cell sites), a few hundred radio channels can be used to provide wireless communication to millions of people. Modern cell phones employ complicated radio receivers and transmitters (transceivers) employing hundreds and thousands of passive devices. Early cell phones used simple architectures such as the super-heterodyne receiver but the demand for low-cost and small footprints has prompted a re-investigation of radio architectures.

The layout of a modern 2.4 GHz transceivers for 802.11b wireless LAN (WLAN) is shown in Fig. 1.5 [7]. The IC is implemented in a 0.25 μ CMOS process and employs several integrated passive devices such as spiral inductors, capacitors, and resistors. The spiral inductors comprise a large fraction of the chip area. The next chip shown in Fig. 1.6

3 I recall asking my trig teacher about the practical application of the subject. After scratching her head and pondering the question, her response was that architects use trig to estimate the height of buildings! A much better answer would have been this equation.

4 And $1/f$ noise in MOS technology.

5 This makes a nice AM to FM modulator, as well.

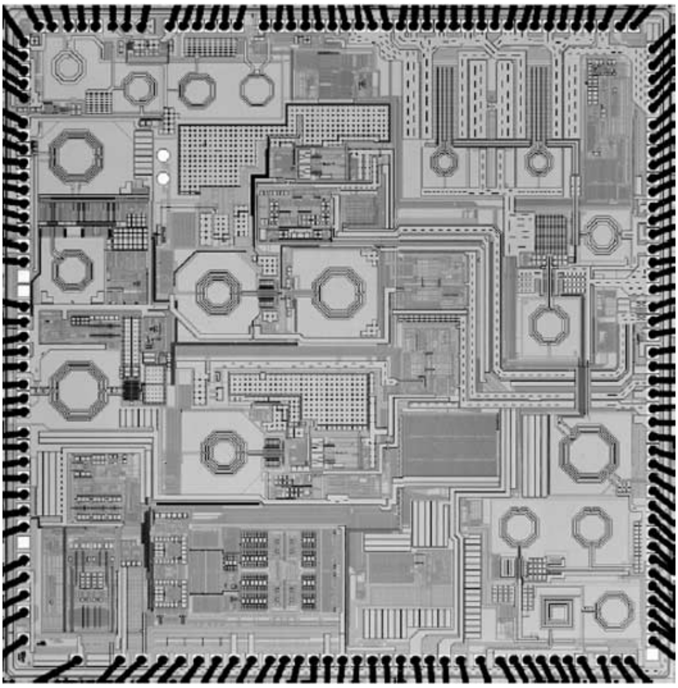

Figure 1.5 A 2.4 GHz CMOS 802.11b Wireless LAN Transceiver [7]. (Copyright 2003, IEEE)

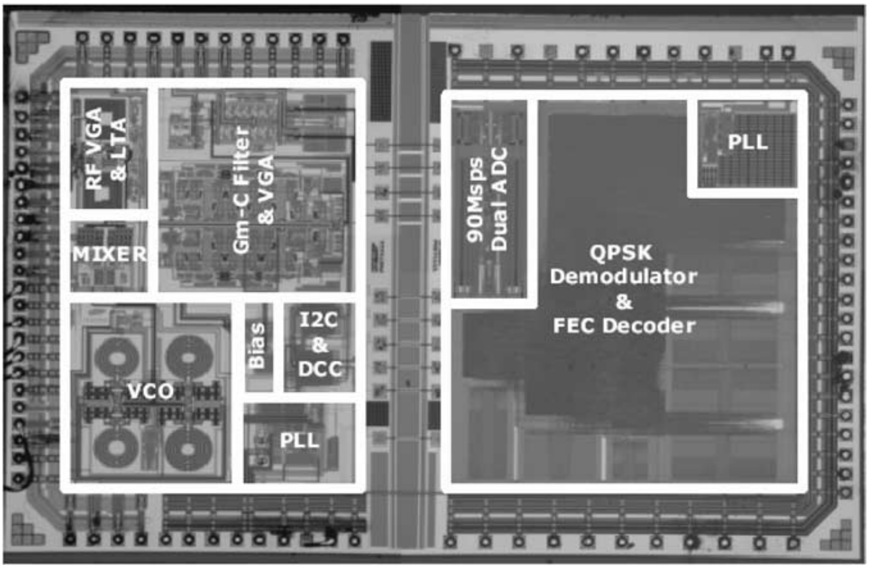

Figure 1.6 A direct-conversion satellite broadband tuner-demodulator SOC [17] operates from 1–2 GHz. (Copyright 2003, IEEE)

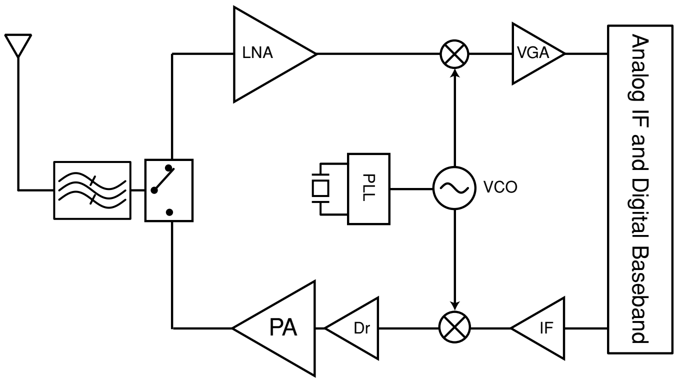

Figure 1.7 The block diagram of an Armstrong super-heterodyne transceiver.

[17] is an integrated direct-conversion satellite broadband tuner-demodulator "system-on-a-chip" (SOC). The chip is implemented in a $0.18\,\mu$ CMOS process and employs MIM capacitors and spiral inductors. It operates in the 1–2 GHz band, requiring broadband operation and high linearity. Notice that the digital baseband has been integrated on to a single chip along with the sensitive analog and RF blocks. This brings about several important challenges in the design due to the parasitic coupling between the various blocks. A triple-well process and lead-less package technology are used to maximize the isolation.

In general, integrating an entire transceiver on to a single chip has many challenges. The power amplifier (PA) or PA driver can injection lock the VCO through the package and substrate, causing a spurious modulation. Digital circuitry can couple seemingly random switching signals into the analog path, effectively increasing the noise floor of the sensitive RF and analog blocks. As the level of integration increases, a single chip or package may contain several systems in operation simultaneously, requiring further understanding and modeling of the coupling mechanisms.

Computers and data communication

Computers and data communication, particularly the Internet, have given rise to a new tidal wave in the information revolution. The speed of computers has improved drastically due to technological improvements in transistor, microprocessor, memory, and system bus architectures. Computer circuits move and process discrete time signals at a frequency determined by the system clock. For instance, in the current generation of computers the clock speed inside the microprocessor is several GHz, while the speed of the system bus and memory lag behind by a factor of 2–3. This is because inside the microprocessor everything is small and dense and signals travel short distances in the presence of small parasitics (mainly capacitance). Off-chip, though, the system bus environment is characterized by much longer distances and much larger parasitics, such as non-ideal dispersive transmission lines along the board traces. Modern computer networks, like gigahertz Ethernet LAN, also operate at high frequencies over wires, necessitating a complete understanding of distributed transmission line effects. These topics are covered in Chapters 9 and 12.

High-speed wireless data communication is the focus of much research and development. The next and future generations of cellular technology will bring the Internet from our homes and offices into virtually every location on earth. Wireless LAN systems enable short-range high-speed data communication without the expensive network infrastructure. A physical network infrastructure requires time-consuming distribution of cables to every office in a building. A wireless system can be up and running in minutes or hours as opposed to days or months.[6]

In such systems cost and size will force many external passive components on to the chip environment, where knowledge of parasitic coupling and loss is critical in a successful low-cost implementation. In this book we spend a great deal of time discussing inductors, capacitors, transformers, and other key passive elements realized in the on-chip environment.

Microwave systems

Microwave systems employ higher frequencies where the wavelength $\lambda = c/f$ is of the order of centimeters or millimeters. Thus the lumped circuit approach fails since these structures are a significant fraction of a wavelength and spatial variation begins to play as important a role as time variation. Such systems were first employed in World War II for radar systems.[7] In a radar system, the small wavelength allows us to construct a highly directional antenna to focus a beam of radiation in a given direction. By observing the reflection, we can compute the time-of-flight and hence the distance to an object. By also observing the Doppler frequency shift, we can compute the speed of the object.

Perhaps the greatest difficulty in designing microwave systems below 10 GHz is that the operating frequency is in an intermediate band where lumped element circuit techniques do not strictly apply and microwave methodology results in prohibitively large circuits. At 3 GHz, the wavelength is 10 cm in air and about 5 cm in silicon dioxide, while an integrated circuit has dimensions of the order of millimeters, thus precluding distributed elements such as quarter-wave transmission lines. But using advances pseudo-lumped passive devices such as inductors, transformers, and capacitors, microwave ICs can be realized with minimal off-chip components.

Many early microwave systems were designed for military applications where size and cost were of less concern in comparison to the quality and reliability. This led to many experimental and trial-and-error design approaches. Difficult system specifications were met by using the best available technology, and often expensive and exotic processes were employed to fabricate high-speed transistors. New microwave systems, in contrast, need to be mass produced and cost and size are the main concerns. Fortunately high-volume process

6 I seem to recall that it took a year for a network upgrade to occur in Cory Hall at Berkeley!
7 It is ironic that the EEs of the time lacked the necessary skills to build such systems and the project was handed off to the physicists at the MIT Radiation Lab.

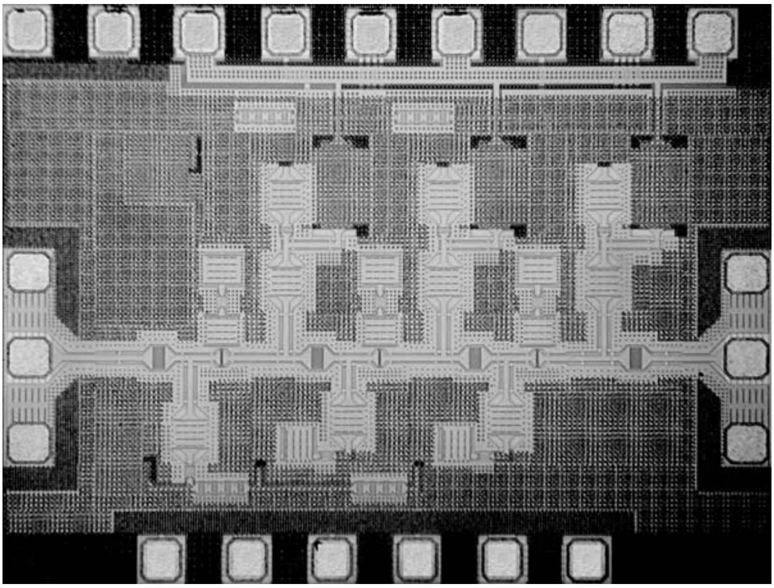

Figure 1.8 A three-stage 60 GHz CMOS LNA implemented in a digital 130 nm process.

technology using silicon is now readily available. The speed is now sufficient to displace many specialized technologies. Since high-volume microwave systems are primarily being designed by circuit engineers as opposed to microwave engineers, the lack of knowledge of electromagnetics and distributed circuits can be an impediment to successful integration and implementation.

Higher-frequency bands offer new opportunities to exploit sparsely used spectrum. The 60 GHz “oxygen absorption” band is a prime example, providing 7 GHz of unlicensed bandwidth in the US. An example of a 60 GHz multi-stage low-noise amplifier (LNA) is shown in Fig. 1.8. Here transmission lines play a key role as inductors, interconnectors, and resonators. A 60 GHz single-transistor mixer, shown in Fig. 1.9, employs a hybrid coupler (see Section 15.7) to combine the RF and LO signal. Spiral inductors are also employed in the IF stages. Both of these chips were fabricated in a digital 130 nm CMOS process. Another CMOS microwave circuit is shown in Fig. 1.10. This is a circular standing-wave 10 GHz oscillator, employing integrated transmission lines in the resonator [11]. There is a beautiful connection between this oscillator and the orbit of an electron in a hydrogen atom. Similar to the wave function of an electron, the electromagnetic mode must satisfy the periodic boundary condition, and this determines the possible resonant modes of the structure.

Optical communication

Fiber-optic communication systems allow large amounts of data to be transmitted great distances with relatively little attenuation. At optical frequencies, metals are too lossy for long-haul communication without amplification, and so the energy is confined inside a thin fiber of glass by total internal reflection. The flexibility and low cost of this material has displaced more traditional waveguides made of rigid or semi-rigid and expensive materials.