5-GHz CMOS Wireless LANs

Thomas H. Lee, *Member, IEEE*, Hirad Samavati, and Hamid R. Rategh

Invited Paper

***Abstract*—This paper first provides an overview of some recently ratified wireless local-area network (WLAN) standards before describing an illustrative 5-GHz WLAN receiver implementation. The receiver, built in a standard 0.25-μm CMOS logic technology, exploits several recent developments, including lateral-flux capacitors, accumulation-mode varactors, injection-locked frequency dividers, and an image-reject low-noise amplifier. The receiver readily complies with the performance requirements of both IEEE 802.11a and ETSI HiperLAN. It exhibits a 7.2-dB noise figure, as well as an input-referred third-order intercept and 1-dB compression point of −7 and −18 dBm, respectively. Image rejection for this double conversion receiver exceeds 50 dB throughout the frequency band without using external filters. Leakage out of the RF port from the local oscillators is under −87 dBm, and all synthesizer spurs are below the −70-dBm noise floor of the instrumentation used to measure them. The receiver consumes 59 mW from a 1.8-V supply and occupies only 4 mm^2 of die area, in no small measure due to the use of fractal capacitors for ac coupling.**

***Index Terms*—802.11, 802.11a, 802.11b, CMOS, CMOS RF, HiperLAN, HiperLAN/1, HiperLAN/2, LAN, receiver, RF, RF CMOS, transceiver, transmitter, WiFi, wireless, WLAN.**

I. Introduction to Wireless LANs

THE growing demand for wireless connectivity has motivated the industry to evolve beyond today's voice-based cellular services. Data-centric third-generation (3G) services now under development seek to provide substantially higher data rates to supplement, and occasionally supplant, wired networks. At the same time, there is a constant desire to keep power consumption and size of the communication devices to a minimum. Fortunately, continuing advances in integrated circuit (IC) technology have made possible the low-cost, compact implementation of transceivers capable of operating at multiple-GHz carrier frequencies with data rates competitive with established wired alternatives. Although the main focus of this paper is the implementation of an integrated 5-GHz wireless local-area network (WLAN) receiver in CMOS technology, the rationale underlying many of the design objectives is best appreciated after understanding the WLAN standards to which the receiver must conform. As a consequence, we undertake here a brief history of, and introduction to, some wireless local area network (LAN) standards.

After working for nearly a decade, the IEEE ratified in 1999 two wireless networking communications standards, dubbed 802.11a (for operation at 5 GHz) and 802.11b (at 2.4 GHz). Underscoring the demand for such products, a generous array of 11-Mb/s 802.11b-compliant devices became available from a multitude of vendors within a year of ratification.

It is not surprising that 802.11b should have been implemented before 802.11a (although the reverse order seemingly implied by the nomenclature can be confusing); building a product for use at 2.4 GHz is substantially easier than building one for 5 GHz. The 802.11b standard specifies operation in the 2.4-GHz industrial–scientific–medical (ISM) band, using direct-sequence spread-spectrum (DS SS) modulation, whereas the 802.11a standard specifies operation in the 5-GHz unlicensed national information infrastructure (UNI) band recently allocated in the U.S. Unlike 802.11b, however, 802.11a does not use DS SS. Instead, it employs orthogonal frequency division multiplexing (OFDM) in order to contend more effectively with the vagaries of indoor propagation. Regrettably the differing physical layers makes it difficult to implement a low-cost transceiver capable of complying with both 802.11a and 802.11b [1].

The 802.11a standard specifies operation over a generous 300-MHz allocation of spectrum for unlicensed operation in the 5-GHz block [2]. Of that 300-MHz allowance, there is a contiguous 200-MHz portion extending from 5.15 to 5.35 GHz, and a separate 100-MHz segment from 5.725 to 5.825 GHz. These allocations are further split into three equal domains distinguished by allowable transmit powers. The bottom 100-MHz domain is restricted to a maximum power output of 50 mW, the next 100 MHz to 250 mW, and the top 100 MHz to a maximum of 1 W (this last domain is largely intended to support outdoor communications). In all three cases, antennas with up to 6 dBi gain are allowed, increasing the effective isotropic radiated powers (EIRP) by a factor of four. Furthermore, antennas with even higher gain may be used as long as the actual transmitter power is reduced 1 dB for every 1 dBi of additional antenna gain. In any case, this rigid segmentation by power level is in contrast to 802.11b where, for example, all transmitters can radiate a continuum of powers up to 1 W (in the United States).

The 300-MHz aggregate spectrum available for 802.11a devices is nearly quadruple the 83 MHz available for 802.11b. The difference in utility implied by that raw bandwidth ratio is compounded by the radically different occupancy rates of the

Manuscript received May 20, 2001. This work was supported by the David and Lucile Packard Foundation, by IBM, and by the Stanford Graduate Fellowship Program.

The authors are with the Center for Integrated Systems, Stanford University, Stanford, CA 94305-4070 USA.

Publisher Item Identifier S 0018-9480(02)00837-2.

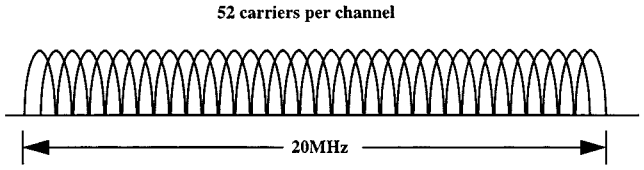

Fig. 1. Detail of subchannels for 802.11a.

two allocations. The 802.11b spectrum has become increasingly crowded by various wireless technologies, such as cordless telephones, remote sensing devices, and recently, Bluetooth. These technologies not only vie with each other for spectrum, but also suffer interference from ubiquitous microwave ovens. The spectrum at 5 GHz, on the other hand, is relatively free of interference, at least for the present. Furthermore, note that Friis' famous propagation formula shows that free-space attenuation is a constant per *wavelength*. On that basis alone, one would expect 5-GHz signals to propagate more poorly than those at 2.4 GHz, mitigating interference to a certain extent. The difference is compounded by the considerably higher indoor attenuation, which additionally increases with frequency in this general frequency range. Of course, the less favorable propagation physics also implies a need for higher radiated powers to achieve a given communication radius.

Increased power alone is not enough to maintain 802.11b-like distances in an 802.11a environment, however. To help compensate, and support higher data rates at the same time, orthogonal frequency division multiplexing (OFDM) is used. OFDM subdivides a carrier into several individually modulated orthogonal subcarriers, all of which are subsequently transmitted in parallel. In 802.11a, each carrier is 20 MHz wide and is subdivided into 52 subchannels, each about 300 kHz wide (not factoring in guardbands; see Fig. 1). Forty-eight of these subchannels are used for data, and the remaining four for error correction.

This subdivision provides a convenient means for accommodating a variety of data rates, permitting different levels of service, and adapting behavior in the face of changing propagation conditions. At the lowest data rate, binary phase-shift keying (BPSK) encodes 125 kb/s of data per channel, resulting in a 6-Mb/s data rate. Using quadrature phase-shift keying (QPSK), the data rate doubles to 250 kb/s per channel, yielding a 12-Mb/s data rate. With 16-level quadrature amplitude modulation (16-QAM) the rate increases further to 24 Mb/s. All 802.11a-compliant devices must support at least these three data rates in order to guarantee a level of interoperability. The standard also allows evolution to rates beyond 24 Mb/s, and several manufacturers have already proposed methods for doing so. The most straightforward in principle is simply to use higher-order QAM, propagation conditions permitting. For example, using 64-QAM theoretically permits an increase to 54 Mb/s. Furthermore multiple channels may also be combined to provide still higher aggregate data rates, of the same order as Fast Ethernet.

At present, 802.11b devices enjoy global deployment because the 2.4-GHz band is available in nearly every country. In parts of the world in which spectrum conflicts exist, easily-implemented software can prevent a device from operating on disallowed frequencies. Regrettably, however, the situation is different in the 5-GHz band. For example, Japan shares only the lowest 100-MHz domain with the U.S., which implies that 802.11a devices in Japan will face a higher likelihood of contention for the fewer available channels. In Europe, the situation is a little better, as the lower 200-MHz band coincides with the Federal Communications Commission's (FCC) 5-GHz allotment. Regrettably, however, the top 100-MHz domain is not as available around the world. For example, various military and government organizations use portions of the 5-GHz space for ground tracking stations and satellite communications. To ensure that unlicensed applications don't interfere with these important existing applications, the European Telecommunication Standards Institute (ETSI) requires the implementation of two additional protocols before allowing operation in Europe. These protocols, dubbed dynamic frequency selection (DFS) and transmit power control (TPC), allow WLAN nodes to respond dynamically to radio interference by some combination of changing channels and reducing power. These protocols are designed to ensure that any incumbent signal gets highest priority when a new signal is introduced in a given area. Both DFS and TPC implementations for 802.11a are being discussed as 802.11h, and an addendum to the 802.11a standard will likely include these features as options.

To complicate an already complex situation, a competing 5-GHz standard, HiperLAN2, is nearing ratification, with many adherents in Europe. HiperLAN2 is an evolutionary step beyond HiperLAN1 (see [3]–[5]). The latter divides the spectrum allocation into 24-MHz-wide channels, each of which nominally provides a 24-Mb/s maximum data rate using Gaussian minimum shift keying (GMSK, with $BT = 0.3$), in addition to a low-bit-rate (LBR) 1.5-Mb/s mode using simple FSK. The latter mode confers on HiperLAN1 the ability to tolerate with ease the large delay spreads that characterize many propagation environments. Measurements show that mean delay spreads of 150 ns or less may be expected in many indoor environments. Since this value is short compared with an LBR mode symbol period, receivers can demodulate such data with simple channel equalization. HiperLAN2 specifies OFDM, bringing it much closer in structure to 802.11a.

Just as in 802.11a, HiperLAN transceivers are divided into classes as a function of permissible transmit powers and required receive sensitivities (defined here as the minimum input power required for a block error rate of 10^{-2}). For HiperLAN1, the lowest class, i.e., Class A, is restricted to a maximum output power (EIRP) of 10 mW and requires a sensitivity of at least -50 dBm, Class B to 100 mW/-60 dBm, and Class C to a maximum of 1 W/-70 dBm. Each transmitter class must provide all of the power levels of the class(es) below it, and each receiver must be able to measure signal strength down to -75 dBm in order to determine better if the channel is clear, but is not required to decode such weak signals. Furthermore, for any given transceiver, it is not permissible for the transmitter class to exceed the receiver class. At the same time, receivers must be able to decode signals as large as -25 dBm to accommodate situations involving WLAN nodes in close proximity.

For HiperLAN2, 200mW EIRP is permitted in the 5.15–5.35-GHz band, with an increase to 1 W in the 5.47–5.725-GHz band. The required sensitivities are

−85 dBm@6 Mb/s data rate and −68 dBm@54 Mb/s, corresponding to 3 dB more stringent requirements than for 802.11a.

Although HiperLAN2 and 802.11a do share some superficial similarities, there are important differences, particularly in the medium access control (MAC) protocol used to give multiple users access to a shared medium. HiperLAN2 derives its MAC largely from asynchronous transfer mode (ATM) and possesses a mechanism for guaranteeing quality of service (QoS) by tagging packets with priority data.

By contrast, 802.11's MAC uses a contention resolution mechanism that traces its heritage to that of 802.3 Ethernet. In the latter, each transmitter first listens to the channel to establish whether the medium is free (that is, it senses whether a carrier is already present). If the medium is free, the transmitter then sends data while simultaneously monitoring its own transmissions. Hearing anything other than the intended transmission is assumed to arise from a collision with data from another transmitter. Upon detecting a collision, the transmitter ceases operation and simply tries sending the data again after waiting a random interval.

The protocol, known as carrier sense multiple access with collision detection (CSMA/CD) is simple and works remarkably well for wired LANs (e.g., 802.3 Ethernet, which pioneered this MAC protocol). However, note that success depends on the ability both to sense the presence of a carrier before transmitting and to detect corrupted data during transmission. In the wired case, the relatively low attenuation of the medium assures that transmitted and received signals are similar in amplitude, facilitating this detection. But wireless propagation involves considerably more, and more variable, attenuation than through a cable. As a result, both carrier sensing and collision detection may fail in numerous ways. As one simple example, consider three linearly arrayed WLAN nodes, labeled A, B, and C. Suppose that B can communicate with both A and C but, because of fading, A and C are unable to communicate directly with each other. In this case, it is possible for both A and C to attempt communication with B simultaneously, each unaware of the presence of the other. Both carrier sense and collision detection fail in such an instance.

Note also that, aside from being largely ineffective, listening while transmitting would impose a severe implementation penalty for wireless nodes because receive and transmit circuit blocks could no longer be shared. Because of such problems, the 802.11 WLAN MAC protocol differs in several important respects from 802.3. It includes a slot reservation mechanism and does not require that a transmitter also listen to its own transmissions. The resulting scheme is called carrier sense multiple access with collision *avoidance* (CSMA/CA). Here, a node first listens before transmitting, just as in CSMA/CD. If it detects no carrier signal, it can safely conclude only that the medium *might* be free. However, there are two additional possibilities: either an out-of-range station may be in the process of requesting a slot, or such a station may already be using a slot reserved for it.

To reserve a slot, a WLAN node sends to the intended receiver a request to send (RTS) message specifying the duration of the requested slot. At the same time, stations within range of the sender also note the request. The receiver replies with a clear to send (CTS) message confirming the duration of the slot. Other stations within the range of the receiver also note the transfer duration specified in the CTS response. All of the stations within range of both sender and receiver use the information contained in the RTS and CTS packets to refrain from transmitting during the requested transfer slot. At the end of the transfer, the receiver acknowledges receipt of the data by transmitting an acknowledgment (ACK) packet.

The RTS and CTS frames are designed to be short to keep small the probability of collision during their transmission. If one occurs nonetheless, or if an RTS does not result in a CTS for some reason, a random backoff interval prior to retransmit is used, just as in the 802.3 Ethernet wired LAN standard.

Although CSMA/CA functions well, it burdens the transceivers with considerable overhead, causing 802.11 WLANs to have slower performance than that of an otherwise equivalent Ethernet LAN. Under favorable conditions the 802.11 MAC is about 70% efficient, so true data throughput at 54 Mb/s is under 40 Mb/s in practice. Additional inefficiencies in drivers, combined with propagation vagaries, may reduce the actual typical throughput to about 25–30 Mb/s, based on experience with 802.11b systems, where 11-Mb/s links generally supply about 6 Mb/s in practice.

Other enhancements to 802.11a currently under discussion include 802.11e, which adds support for multimedia mechanisms to guarantee QoS, and 802.11i, whose focus is enhancement of network security.

Not to be outdone, improvements to 802.11b are also under consideration, as 802.11g, which proposes to double the peak data rate to 22 Mb/s, while maintaining backward compatibility with existing 11-Mb/s 802.11b devices. The higher data rate is enjoyed when propagation and interference conditions permit it. Again, actual peak throughput is likely to be below 15 Mb/s, and concern about the crowded nature of the 2.4-GHz ISM band raises questions about the fraction of time users may realistically enjoy the boosted rates.

A. Other WLANs

Although the highest data rates are promised by the WLAN systems discussed in the foregoing section, there are other emerging WLAN standards (see, e.g., [6] and [7]). Already mentioned is Bluetooth, designed as a short-range, low-data-rate system. Originally intended largely as a cable replacement technology, target applications for it have evolved to overlap with those of 802.11. Named for Harald Bluetooth, a 10th-century Viking who ruled over Denmark and Norway, it operates in the 2.4-GHz ISM band. Bluetooth employs frequency hopping for interference mitigation and nominally provides a 1-Mb/s peak data rate over short distances (e.g., $<$10 m). Its promise of low cost has generated much interest, despite the relatively low data rate, and wide deployment of Bluetooth-compliant devices has been forecast.

Somewhat in competition with both Bluetooth and 802.11, the HomeRF Networking Group has produced a set of specifications known as the shared wireless access protocol (SWAP) that also uses frequency hopping in the 2.4-GHz ISM band to

provide peak data rates of 10 Mb/s, over distances intermediate between those targeted by 802.11 and Bluetooth.

In addition to HiperLAN, ETSI also supports the digital European cordless telecommunications (DECT) standard. Both the standard and abbreviation have been co-opted in the United States as "digitally enhanced cordless telecommunications." The modulation technique used for DECT is Gaussian frequency shift keying (GFSK), allowing DECT to offer data rates up to 1.152 Mb/s in a channelized frequency band that spans from 1.88 to 1.99 GHz. Countless DECT cordless telephones are currently in use in Europe and elsewhere.

II. Performance Requirements for a 5-GHz WLAN Receiver

Although the MAC layers for HiperLAN2 and 802.11a differ significantly, performance requirements for the RF signal processing blocks are quite similar. This commonality should not be surprising in view of the similar frequency bands, data rates, and intended deployment scenarios. Consequently, it is possible for a single receiver design to comply with both sets of specifications.

To determine the precise target values, we first compute the specifications for both HiperLAN and 802.11a separately, and select the more stringent of the two in every case. Here we reduce the specification set to frequency range, noise figure, maximum input signal level (or input-referred 1-dB compression point), and limits on spurious emissions.

For frequency range, it is often acceptable to cover only the lower 200-MHz band. The upper 100-MHz domain is not contiguous with that allocation, so its coverage would complicate somewhat the design of the synthesizer. Furthermore, that upper 100-MHz spectrum is not universally available. Hence the choice here is to span 5.15–5.35 GHz.

The worst-case noise figure requirement for HiperLAN1 is not directly specified, but may be readily estimated from the fact that a Class C receiver must exhibit a -70-dBm sensitivity over a channel bandwidth of 24 MHz. Assuming conservatively that the predetection SNR must exceed 12 dB, the overall receiver noise figure must be better than about 18 dB.

Strictly speaking, the required noise figure for HiperLAN2 and 802.11a receivers is a function of data rate. Since it would be cumbersome to specify (let alone design for) individual noise figures for each possible data rate, the specification for 802.11a instead simply recommends a noise figure of 10 dB, with a 5-dB implementation margin, to accommodate the worst-case situation. As this target is more demanding than that of HiperLAN2, a 10-dB maximum noise figure is the design goal for the present work.

As stated previously, HiperLAN1 specifies -25 dBm as the maximum input signal that a receiver must accommodate (for a 1% block error rate), whereas 802.11a specifies a value of -30 dBm (for a 10% packet error rate). Consequently, -25 dBm is the target maximum input level. Converting these specifications into a precise IIP3 target or 1-dB compression requirement is nontrivial. However, as a conservative rule of thumb, the 1-dB compression point of the receiver should be about 4 dB above the maximum input signal power level that must be tolerated successfully. Based on this approximation, we target a worst-case input-referred 1-dB compression point of -21 dBm.

Finally, the spurious emissions generated by the receiver must not exceed -57 dBm for frequencies below 1 GHz, and -47 dBm for higher frequencies, in order to comply with FCC regulations.

III. Silicon Technologies for 5-GHz Wireless Circuits

There are presently three silicon IC technologies suitable for realizing circuits in the 5-GHz frequency range. Silicon, and silicon–germanium (SiGe), bipolar devices currently provide the highest performance and enjoy the customary advantage of a high g_m/I ratio, in addition to process refinements specifically intended to enhance analog and RF performance. These latter improvements often include special resistor and capacitor options that possess some combination of tighter tolerance, reduced parasitics, and higher Q.

A significantly less expensive technology is conventional digital CMOS. Although it suffers from decidedly inferior device physics, continued investment on a vast scale has increased its suitability for use at high frequencies. Transistors in the 0.13-μm generation of technology now making a transition into production typically possess peak f_T values of 80–90 GHz.[1] This performance is not a limit by any means, as transistors with extrapolated f_T values of triple this value for the 45-nm process generation have already been demonstrated in the laboratory, with further improvements anticipated [20]. Although its inferior g_m/I ratio makes CMOS circuit performance more sensitive to wiring parasitics at a given level of power consumption than for bipolar technologies, the superior linearity of short-channel MOS transistors typically confers a somewhat higher *dynamic range* per power than that of bipolars, and this quality is often extremely important for wireless systems.

Another noteworthy factor is the large number of interconnect layers now commonly available in CMOS logic processes. There the obsession with circuit density has driven the development of chemical–mechanical polishing (CMP) to make practical the fabrication of an almost arbitrary number of interconnect levels. On average, three interconnect levels are added every four process generations, and leading-edge processes currently provide seven or eight layers of metal [8]. For RF applications, these additional layers are indispensable for fabricating inductors and linear capacitors of high quality.

IV. Receiver Implementation

A. Architectural Considerations

Within the general family of superheterodyne receivers lie numerous variants, each of which is capable of satisfactorily meeting the electrical specifications. There is always a desire to minimize cost, so architectures that reduce the need for costly

[1]As an *extremely* crude approximation, the effective channel length may be taken as about 50% of the drawn channel length for all CMOS process generations from 0.13 μm on. Furthermore, the product of peak f_T and effective channel length may be treated as a constant (again, as an extremely crude approximation) in the deep submicrometer regime, with a value of about 5–6 GHz-μm.

external filters, such as the direct-conversion (or homodyne) and low IF, are particularly attractive.

The homodyne is the degenerate case of the superheterodyne in which the IF is chosen as zero. Since an incoming RF signal and its image are separated by twice the intermediate frequency, the homodyne theoretically sidesteps the image-rejection problem by making the signal its own image. Subsequent signal processing also takes place at the lowest possible frequencies, relaxing speed demands on critical blocks such as A/D converters and baseband filters.

Despite these attributes, implementation of a direct-conversion receiver is not without serious difficulties, and these have inhibited widespread adoption of this architecture [10]. These problems are all a consequence of a homodyne's inherently high sensitivity to dc and low-frequency signals. For example, typical dc offsets are easily much larger than the downconverted RF signals. Additionally, radiation from the local oscillator (LO) may couple back into the RF input port with a random phase, producing additional dc offset after mixing. Even-order nonlinearities can also create signal-dependent offsets as well. Regardless of origin, these offsets may also change dramatically when the LO frequency changes value during frequency hopping or channel selection, making offset removal additionally challenging. Finally, $1/f$ noise is unfortunately of a nature to produce the maximum negative effect in homodyne receivers. Regrettably, CMOS exhibits inferior matching and $1/f$ noise. Even if offsets and noise are not large enough to overload subsequent stages, they can readily reduce sensitivity to poor values. Implementing successful and cost-effective solutions is sufficiently difficult that there is yet no abundance of commercially significant existence proofs despite numerous (and ongoing) earnest efforts.

A low-IF architecture possesses many of the attributes of a homodyne receiver (namely, relaxed speed demands on IF circuit blocks), but has low sensitivity to dc offsets and $1/f$ noise. The tradeoff, however, is that the image rejection problem reappears. If the goal is to avoid the use of expensive filters, the burden of image rejection must be borne architecturally.

The Weaver architecture (Fig. 2) is a well-known textbook solution to the image rejection problem [9]. Since a signal and its image may be distinguished by their differing phase, cancellation of the image signal while simultaneously passing the RF signal is possible. As with any system reliant on miraculous cancellations, a high degree of image rejection depends on exquisite matching of gains and phase throughout the receiver chain. If the radian phase-matching error ε and fractional gain mismatch θ are both small, the image-rejection ratio (IRR) (defined as the power ratio of signal to image) may be expressed approximately as [15]

$$\text{IRR} \approx \frac{4}{\varepsilon^2 + \theta^2}. \tag{1}$$

To underscore the tight requirements on matching, consider that errors of 0.1% in gain and 1° in phase bound the IRR to below 41 dB. At a carrier frequency of 5 GHz, note that a 1° phase error corresponds to a time mismatch of under 0.6 ps, or the time it takes light to travel about 200 μm in free space. Consequently, image rejection ratios for 5-GHz receivers typically fall short of the 41-dB value calculated and generally lie within the range of 25–35 dB. Unfortunately, higher values may be needed for practical systems. While automatic calibration techniques can improve the practically achievable image rejection ratios, implementations can be somewhat cumbersome or power-consumptive.

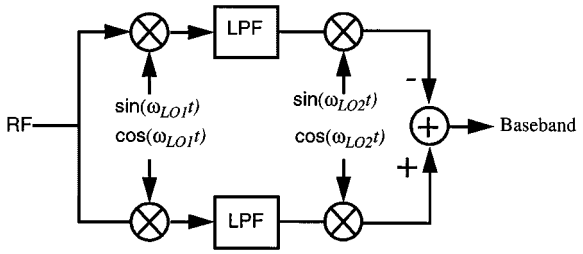

Fig. 2. Weaver architecture.

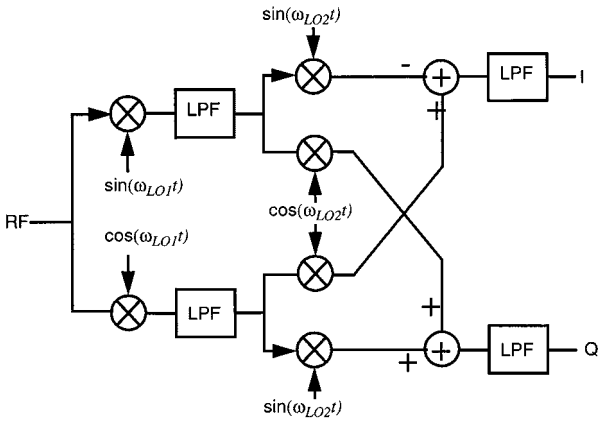

Fig. 3. Quadrature Weaver architecture.

With a simple modification, the Weaver architecture readily provides quadrature outputs, as is needed for many modulation types (see Fig. 3), and it is this architecture that is used in this receiver.

In this work, double conversion to a zero-frequency baseband uses first and second LO frequencies that are 16/17 and 1/17 that of the RF input, respectively. These choices have several attributes. One is that the second LO is readily derived from the first LO through a simple binary divider. Another is that the image signal happens to lie within the downlink spectrum of an existing satellite system and is consequently relatively weak.

The overall receiver architecture is shown in Fig. 4. As seen, it consists of an LNA that is integrated with a tracking notch filter controlled by a phase-locked loop (PLL), a quadrature Weaver image-reject core, and ac-coupled baseband buffers. In addition, the receiver contains a frequency synthesizer that provides coverage for the 200-MHz span of the lower two domains. The synthesizer provides quadrature outputs at both 16/17 and 1/17 the RF input frequency and reduces power consumption by replacing a standard flip-flop based divider with an injection-locked frequency divider. Implementation details for these and other blocks are discussed in the following sections.

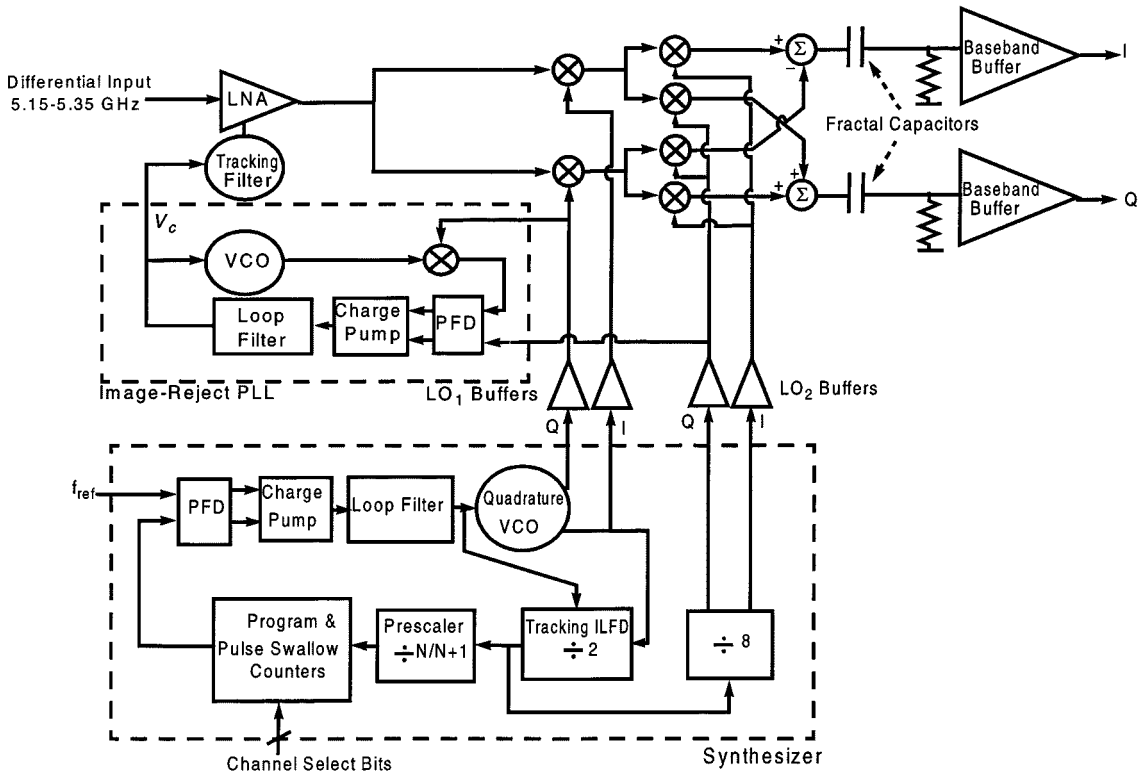

Fig. 4. Architecture for the 5-GHz CMOS WLAN receiver.

B. LNA With Tracking Notch Filter

Since mismatches in a Weaver circuit can degrade IRR to unacceptably low values, many practical receivers need to supplement the image rejection beyond what a Weaver architecture can provide. One way to augment IRR is through the use of an external bandpass filter, of course, but such a solution is contrary to the cost-conscious philosophy that motivates consideration of the architecture in the first place. An alternative is to use a notch filter, which is more easily integrated than a conventional bandpass filter. This ease of integration stems from the fact that a deep notch may be provided by simple low-order networks. The major drawback of this approach is the need for tuning, owing to the narrowness of the notch. Hence, automatic tuning is mandatory if a notch filter is to be used for image cancellation [17]–[19].

To save area and power, the notch filter is merged here with a popular source-degenerated low-noise amplifier. Because the sensitivity requirements are modest, the low-noise amplifier need not exhibit extraordinarily low noise figures. Rather, the focus is on low power consumption and providing sufficient linearity. To understand how the notch may be implemented with minimal overhead, first consider modifying the transfer function of this LNA by an LC tank circuit, as shown in Fig. 5.

The series resonant frequency of the LC circuit is chosen equal to that of the image. At the image frequency, the LC circuit steals current away from M_3, thus reducing the gain at that frequency. Regrettably, the impedance of the LC circuit at the signal frequency is still finite, so noise figure and gain suffer.

As shown in Fig. 6, parasitic capacitance at node X further degrades the noise performance of the cascode structure.

This parasitic capacitance C_x lowers the impedance at node X and reduces the gain of the cascode structure. The presence of this capacitance increases the noise contribution of M_3 while simultaneously reducing the signal contribution of M_1. To reduce the resulting noise figure penalty, this capacitance must be nullified. Ignoring for the moment the issue of biasing, an inductor placed in parallel with this parasitic capacitance is a remedy to the problem. In Fig. 6, noise figure is plotted versus frequency, showing the improvement obtained with the help of the inductor.

Combining the ideas shown in Figs. 5 and 6 results in the circuit shown in Fig. 7.

The filter comprises an inductor, a capacitor, and a varactor. The filter has a low impedance at the frequency of the image and a high impedance at the frequency of the signal. Formally, the input impedance of the filter, Z_f, can be written as

$$Z_f(s) = \frac{s^2 L_5(C_3 + C_1) + 1}{s^3 L_5 C_1 C_3 + sC_3}. \tag{2}$$

The filter has imaginary zeros at

$$\omega_z = \pm \frac{1}{\sqrt{L_5(C_3 + C_1)}} \tag{3}$$

and imaginary poles at

$$\omega_p = \pm \frac{1}{\sqrt{L_5 C_1}}. \tag{4}$$

The location of the pole-zero pair on the imaginary axis is controlled by an accumulation mode varactor (Fig. 8) whose small-

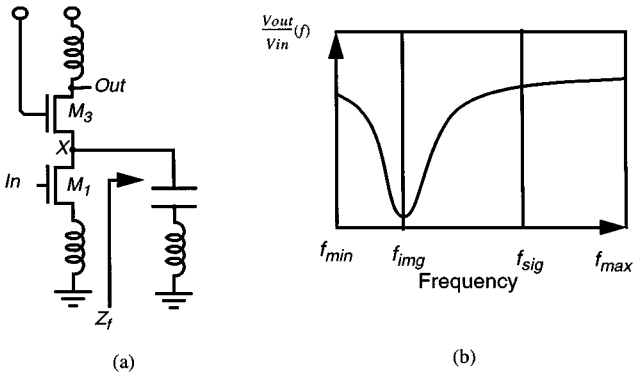

Fig. 5. (a) Image-reject LNA and (b) input–output transfer function.

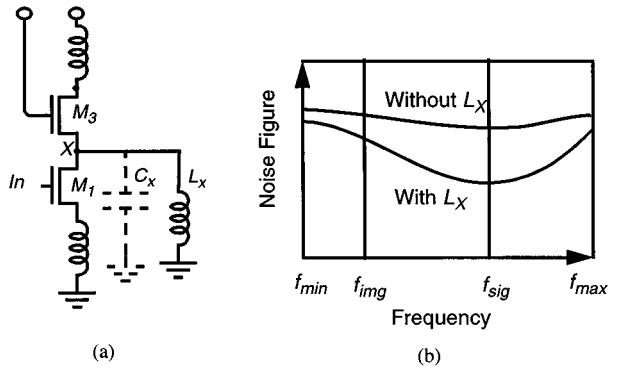

Fig. 6. (a) Improving the noise figure of a standard LNA. (b) Noise figure versus frequency.

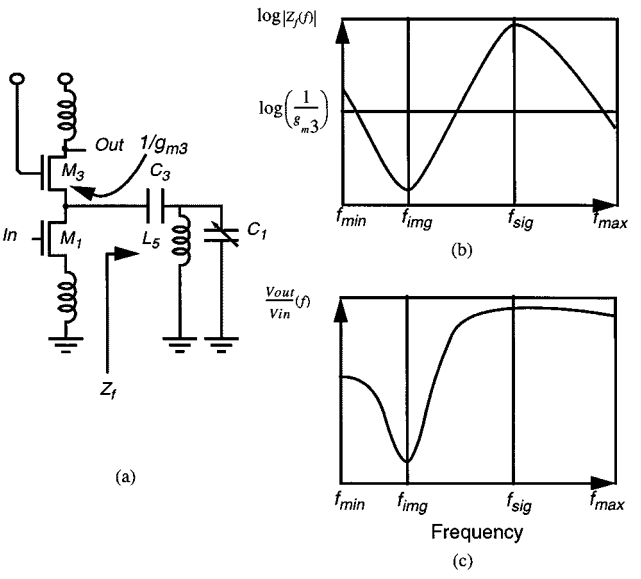

Fig. 7. (a) Circuit diagram of the LNA with the filter. (b) Input impedance of the filter versus frequency. (c) The transfer function of the LNA/filter combination.

signal tuning characteristics are shown in Fig. 9 [13]. This structure is inherently available in all CMOS processes and exhibits Q-frequency products in excess of 200 GHz at the 0.25-μm process generation.

Also shown in Fig. 7 is the input impedance of the filter, $|Z_f|$, as a function of frequency. The resistance looking into the source of the cascode device, $1/g_{m3}$, has also been marked on

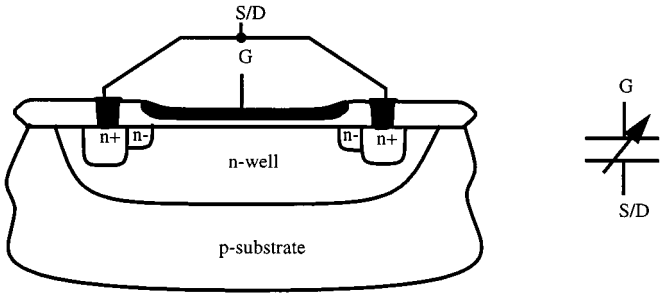

Fig. 8. Accumulation mode varactor.

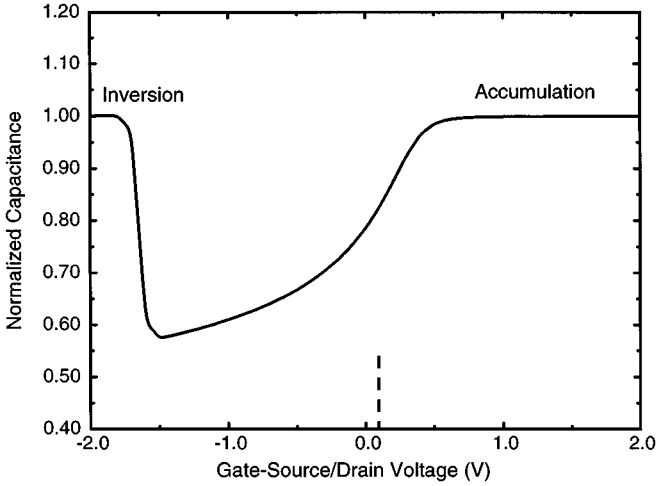

Fig. 9. Tuning characteristics of varactor.

the same graph for comparison. For frequencies close to the location of the zero, the filter has an impedance lower than $1/g_{m3}$ and steals the ac current away from M_3, thus reducing the LNA gain. Near the pole frequency, $|Z_f|$ is larger than $1/g_{m3}$ and the LNA gain is consequently high. As seen from the figure, the resulting overall transfer function has a narrow valley, so for correct image cancellation the zero must occur at the correct frequency. On the other hand, the peak is wideband and so the exact location of the pole is less important.

The third-order filter thus not only boosts image rejection but also diminishes the effect of the parasitic capacitance at node X. Thus the filter simultaneously provides good image rejection and good noise performance. Although, rigorously speaking, (2)–(4) need to be modified slightly to include the effect of this parasitic capacitance, the foregoing argument is still valid in its essential features.

Fig. 10 shows in greater detail the combined LNA/filter as actually implemented. A differential architecture is chosen for its better rejection of on-chip interference and for its insensitivity to parasitic inductance between the common-source connection and ground. To achieve the desired linearity, the LNA consists of only one stage, formed by transistors M_1–M_4. Inductive degeneration is employed in the sources of M_1 and M_2 to produce a real term in the LNAs input impedance [12].

Capacitors C_1–C_4 and inductors L_5 and L_6 form a differential version of the third-order filter. To accommodate the requirement for precise tuning of the notch, accumulation-mode MOS varactors C_1 and C_2 are varied by control voltage V_c.

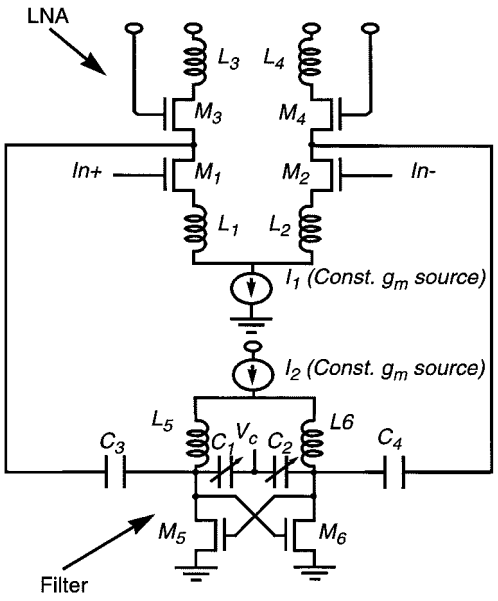

Fig. 10. Simplified circuit diagram of the LNA/filter combination.

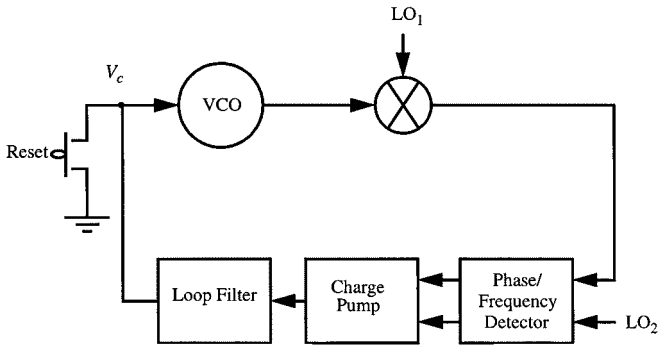

Fig. 11. Image-reject PLL.

The negative resistance generated by the cross-connected differential pair, M_5–M_6, deepens the notch by canceling filter losses arising mainly from the finite Q of the inductors. To decrease the sensitivity of this negative impedance (and consequently of the LNA gain) to temperature and process variation, a constant-g_m biasing source is employed. Here, bias current I_2 is chosen to boost the Q by a factor 5. This value is high enough to provide a significant notch depth, but not nearly high enough to endanger loop stability in any condition. Simulations show, and measurements confirm, that the circuit tolerates more than a tripling of the nominal bias current without instability.

The control voltage for the notch filter is generated by a low power image-reject phase-locked loop (IR PLL) (Fig. 11). This PLL is a simple offset synthesizer which achieves lock when the internal IR PLL's voltage-controlled oscillator (VCO) frequency equals the difference between the two LO frequencies. To prevent parasitic locking at the sum frequency instead, the lock range is restricted, and acquisition always starts from the low-frequency side (using the reset switch shown), assuring that the loop first encounters the desired difference frequency condition.

The IR PLL's VCO (Fig. 12) and the LNA's notch filter are topologically identical, differing only in bias current. Consequently, tuning the VCO to the image frequency also tunes the notch frequency, assuring process independence of the notch location.

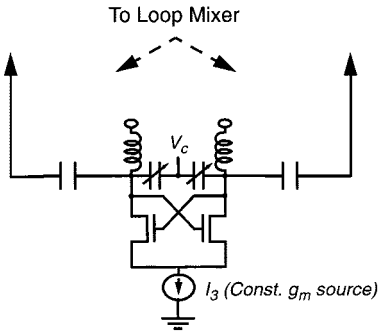

Fig. 12. Schematic of IR PLL VCO.

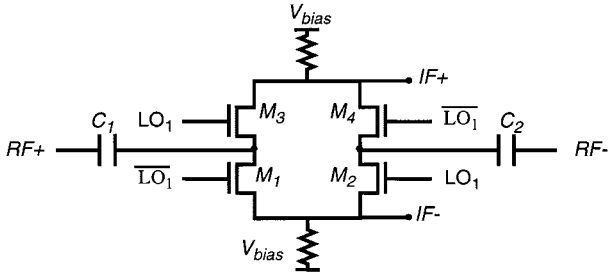

Fig. 13. Passive ring mixer.

The LNA proper consumes 6.7 mW and exhibits a noise figure of 4.3 dB. The IR PLL adds 3.1 mW, for a total consumption here of just under 10 mW. The image rejection enhancement provided by the notch filter is 16 dB.

C. Mixers

The six mixers of the quadrature Weaver architecture are implemented two ways. Since CMOS transistors are good voltage-mode switches, the first pair of mixers are simple passive ring mixers for good linearity and low power (Fig. 13) [15]. The outputs of this first pair drive a quad of Gilbert type mixers (Fig. 14). Although these mixers exhibit worse linearity than passive rings, the attenuation provided by the passive mixers relaxes the requirements. Furthermore, their differential current mode nature makes it easy to implement addition and subtraction of the output signals. Finally, the gain provided by these active mixers is desirable by itself and for reducing the noise figure contribution of subsequent stages.

The common-source connection of the input transistors is grounded to reduce supply voltage requirements and also to mitigate any second-order distortion which might contribute to the generation of beat components [10].

D. AC Coupling for Offset Mitigation

The outputs of the second set of mixers is ac coupled to the baseband circuitry. Although ac coupling confers relative freedom from offsets compared to an otherwise equivalent single-conversion homodyne receiver, important issues remain nonetheless. For example, the coupling capacitors must be linear. Additionally, the pole frequency of the coupling net-

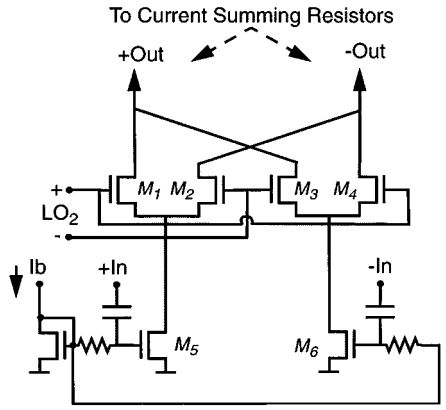

Fig. 14. Gilbert-type double-balanced mixer.

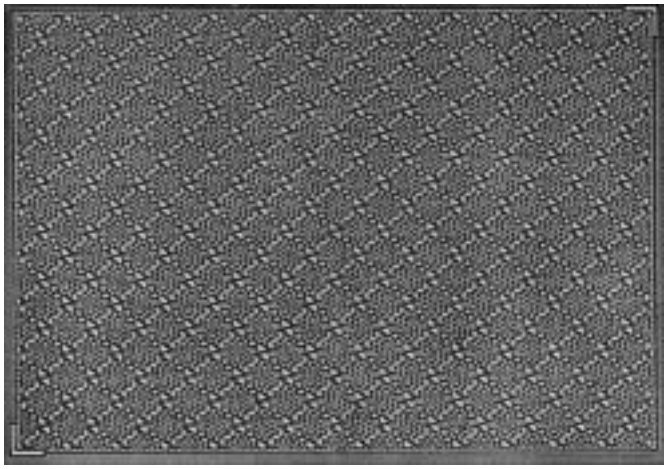

Fig. 15. Minkowski-sausage-based fractal capacitor.

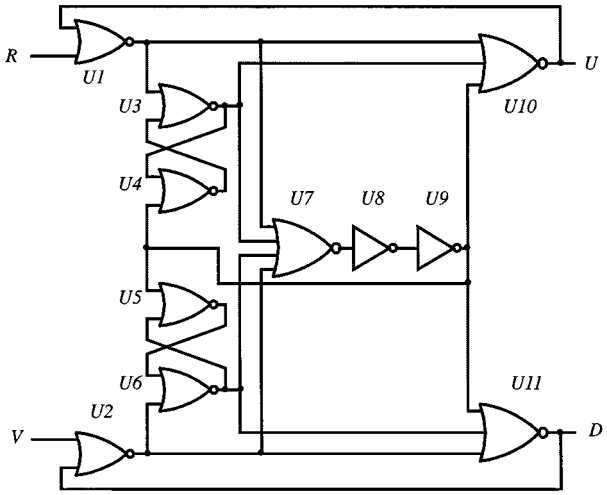

Fig. 16. Phase-frequency detector.

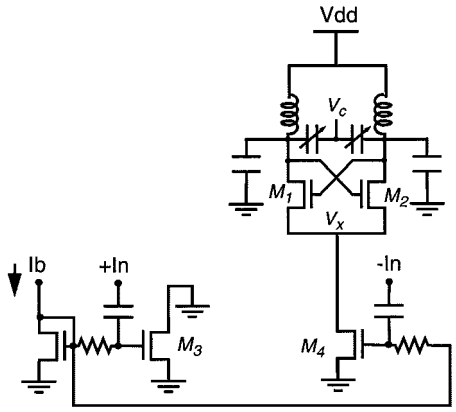

Fig. 17. Injection-locked frequency divider.

work must be high enough to assure sufficiently fast recovery from overload, but low enough to avoid causing excessive intersymbol interference. This latter consideration usually demands the use of coupling capacitors that are relatively high in value. The linearity requirement is best satisfied by metal–metal structures, rather than gate capacitance, but high value metal–metal capacitors consume significant die area. In this work, 15-pF coupling capacitors are used to produce a 5-kHz corner frequency. By using lateral flux capacitors with quasi-fractal boundaries, the capacitance density is boosted by a factor of 3.5 (relative to a standard parallel-plate sandwich with the same number of metal layers), to 700 af/μm^2 [14]. This boost factor increases as lithography scales and is also a function of the particular fractal geometries chosen. Here, the layout is based on a Minkowski sausage (shown in Fig. 15), chosen for its reasonable boost factor, as well as its ability to fill a rectangular space.

A subsidiary benefit of exploiting lateral flux is a reduction in bottom-plate capacitance. This reduction arises from two contributions. One is a direct effect resulting from the simple reduction in plate area needed to produce a given overall capacitance. The second is that some flux that would have terminated in the substrate instead terminates on adjacent metal. For the capacitor used here, the bottom-plate capacitance per terminal is only 8% of the total value, a value lower than that found in many processes with special structures dedicated to enhance analog and RF performance.

For many OFDM-modulated systems, settling time requirements are more stringent, so more sophisticated offset cancellation techniques than used here would probably be more appropriate.

E. Frequency Synthesizer

The LO signals are generated by an integer-N frequency synthesizer. The loop employs a conventional phase-frequency detector with the standard delay (via $U8$ and $U9$) in the reset path to mitigate dead-zone effects arising from runt pulses (Fig. 16).

Not shown is additional circuitry for generating low-skew complementary representations of the U and D outputs. To reduce power, the feedback divider is implemented as a cascade of an injection-locked frequency divider and a more conventional prescaler. The injection-locked divider is actually an oscillator whose free-running frequency is approximately 2.5 GHz, nominally one-half the synthesized output frequency (Fig. 17). Such a circuit can consume considerably less power than an analogous flip-flop based divider because resonant circuits can be used. The tradeoff is a reduction in operational frequency range but, since most commercial systems are narrowband in nature, this limitation does not preclude the use of such circuits. Here,

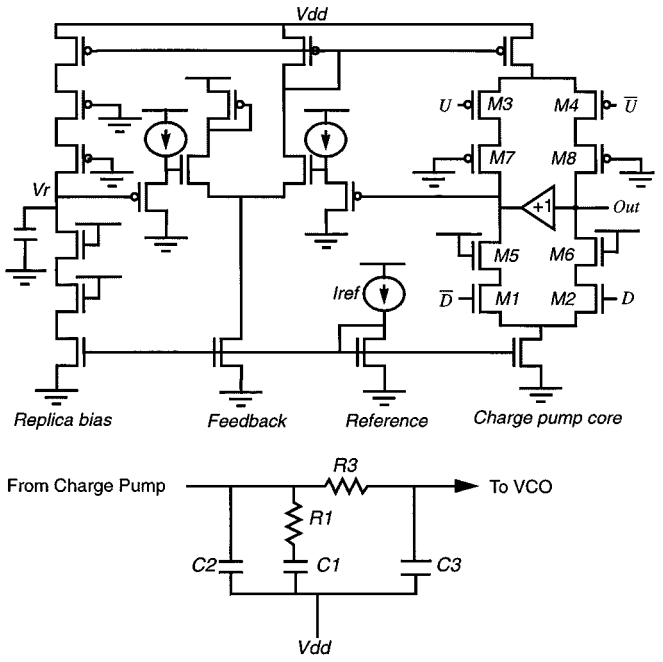

Fig. 18. Schematic of charge pump and loop filter.

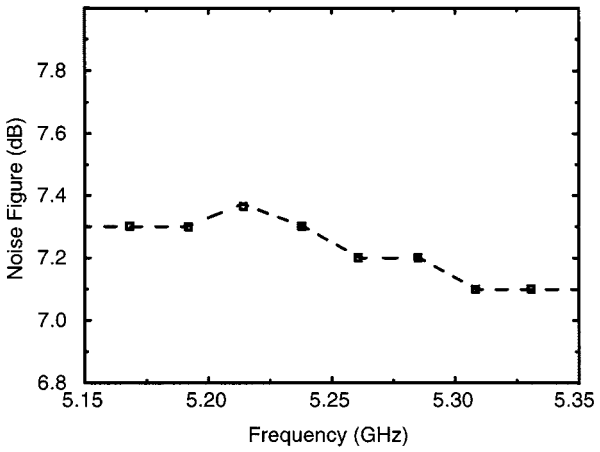

Fig. 19. Measured receiver noise figure.

cross-coupled differential pair M_1–M_2 synthesizes a negative resistance to overcome the loss in the LC drain network to sustain oscillation; the output of the divider is taken from the drains of these transistors. Even with an externally applied perturbation, the Barkhausen conditions on loop gain and phase must be satisfied. Obtaining the desired divide-by-two action is facilitated by enhancing the second-order nonlinearity in the loop. Such a nonlinearity produces an intermodulation component at a frequency equal to the difference between the frequency of the oscillator and the injection signal. If these frequencies are in a precise 2 : 1 ratio, a self-consistent solution to the loop equations can exist, and synchronization results [11].

In this differential circuit, the common-source node V_x contains a strong spectral component at twice the oscillation frequency. As this double-frequency component may be regarded as due to the action of a second-order nonlinearity, injecting a 5-GHz signal into this node should result in synchronization at

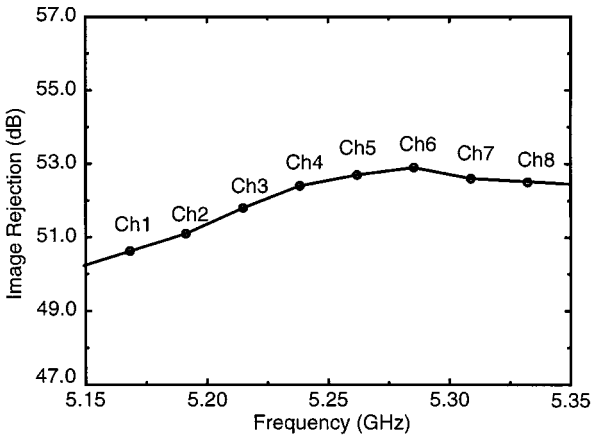

Fig. 20. Measured image rejection.

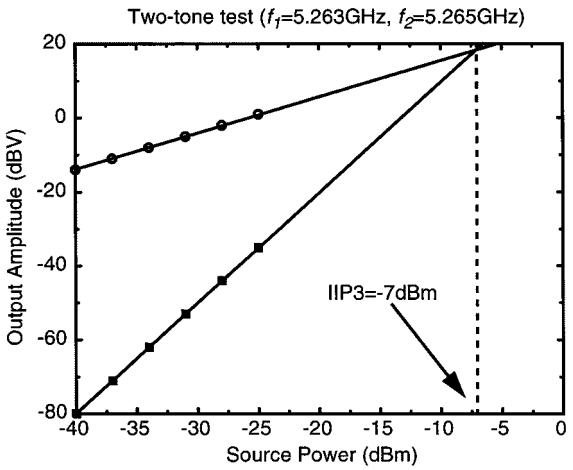

Fig. 21. Two-tone intermodulation test results.

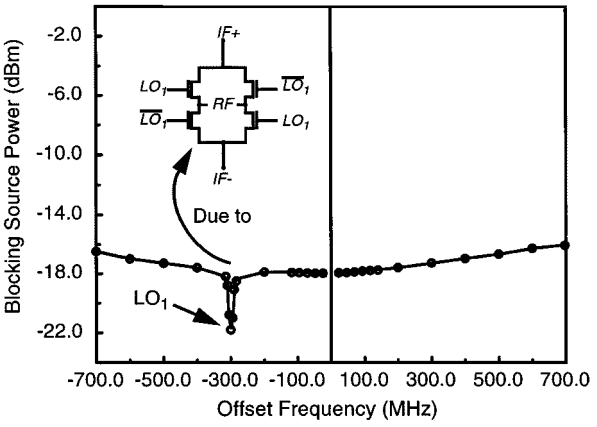

Fig. 22. Measured blocking performance.

half the injection frequency. This conjecture is confirmed in a quantitative analysis of the circuit [11].

The same analysis reveals that maximizing the locking range requires maximizing the tank inductance L. However, power consumption is inversely related to the tank impedance at resonance and, hence, to the QL product. Unfortunately, there is no guarantee that maximizing L automatically maximizes QL at the same time. Hence, the divider implemented here maximizes

TABLE I
SUMMARY OF 5-GHz CMOS WLAN RECEIVER CHARACTERISTICS

Signal path performance	**Achieved**	**Required**
Noise figure	7.2dB	18.3dB
Voltage gain	26dB	
S_{11}	< -14dB	
Image rejection (filter only)	16dB	
Image rejection (total)	53dB	
Input-referred IP3	-7dBm	
1-dB compression point	-18dBm	-21dBm (est)
LO_1 Leakage to RF	-87dBm	-47dBm
LO_2 Leakage to RF	-88dBm	-57dBm
Power dissipation		
Synthesizer	25.3mW	
Divide-by-8 (for LO_2)	6.0mW	
Signal path	18.5mW	
Image-reject PLL	3.1mW	
LO buffers	5.0mW	
Biasing	0.9mW	
Total power	58.8mW	
Supply voltage	1.8V	

the inductance, subject to a somewhat arbitrary power consumption limit of 1 mW. This power is one-fifth that consumed by a conventional flip-flop divider built in this technology.

Because the optimization objective does not directly accommodate a specification on the tuning range, there is always a danger of insufficiency. To solve this problem, the center frequency of the divider is made to track automatically the frequency of the VCO. Rather than having to accommodate the entire tuning range of the receiver, the divider now only has to have a tuning range sufficient to accommodate component mismatches, a considerably simpler requirement. The tuning capacitance is implemented as an accumulation mode varactor, as is the tuning element in the LNA notch filter.

To minimize spurs, the phase detector drives a differential charge pump that is designed for low leakage and low feedthrough of up and down command pulses, as well as the removal of all sources of systematic offset (Fig. 18). Although the charge pump output is taken as a single-ended signal, a bootstrapping buffer forces the unused output in the charge pump core to the same voltage as the main output. Furthermore, a replica bias circuit assures minimum sensitivity of the pump current mismatch to the common-mode output voltage. If the charge pump output voltage differs from Vr, an op-amp adjusts the pull-up current until equality is restored. The loop order is also increased to four to enhance filtering of the control voltage, as seen in the figure. As a result of these combined strategies, all synthesizer spurs are below the -70-dBc noise floor of the instrumentation and well below any values needed to meet performance objectives.

Thanks in part to the use of the resonant frequency divider, the complete synthesizer consumes 25 mW, including the VCO and all conventional dividers (whose power consumption now dominates).

The measured phase noise of the synthesizer is -134 dBc/Hz at the center of the adjacent channel (22-MHz offset). The VCO noise integrated over two adjacent channels is -58 dBc, implying that an adjacent channel interferer can be 48 dB stronger than the desired signal while maintaining a 10-dB signal-to-interference ratio.

Finally, the settling time after changing channels is under 35 μs, comfortably better than the 1-ms requirement [16].

F. Performance Measurements

The overall receiver noise figure is shown as a function of frequency in Fig. 19. As can be seen, the stages after the LNA increase the noise figure by about 3 dB to a value of about 7.2 dB. Despite the relatively large second stage contribution, this noise figure remains well below the 18-dB target value for HiperLAN, and still comfortably below the 10-dB figure for 802.11a.

The image rejection is seen from Fig. 20 to lie between 50–53 dBtt0 over the entire band. About 16 dB of this rejection is due to the notch filter in the LNA, and another 35 dB comes from the Weaver architecture itself. These values are robustly achieved without implementing calibration of any kind. If the specifications on image rejection were significantly tighter, it may be desirable to implement some autocalibration method.

Linearity is evaluated with a two-tone test in Fig. 21. The input-referred IP3 is -7 dBm, with a -1-dB compression point of -18 dBm. The latter value is comfortably better than the -21-dBm target. This performance is obtained at relatively low bias currents, thanks to the high linearity of short-channel MOSFETs.

A revealing test is a 1-dB blocking desensitization evaluation. As seen in Fig. 22, the receiver generally tolerates blockers larger than -18 dBm over the entire frequency range. Since HiperLAN specifies that receivers must tolerate in-band blockers as large as -25 dBm, there is evidently substantial margin. Note that, at a frequency equal to that of the first LO, there is a noticeable dip in the blocking performance (to a still-satisfactory -22 dBm). The passive ring mixer used to implement the RF mixers is the reason for this diminished performance. When a strong blocker at a frequency LO_1

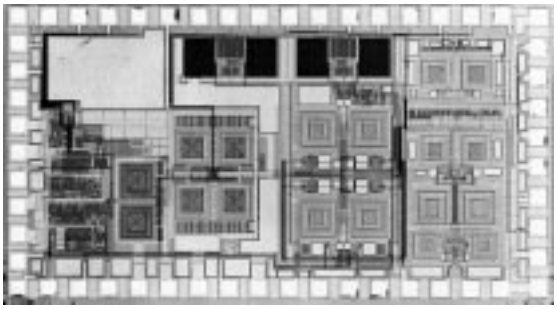

Fig. 23. Die micrograph.

appears at the RF port, it generates a dc voltage at the output of the RF mixers. The resulting bias shift reduces the gain [16].

The overall characteristics of the receiver are summarized in Table I, and the die micrograph is shown in Fig. 23.

V. Summary

From the results, it should be clear that CMOS is a credible medium for implementing high-performance, low-power RF circuits in the low-GHz frequency range. Although CMOS suffers from decidedly inferior device physics, these deficiencies can be largely overcome by a combination of raw process scaling, appropriate architectural choices, exploitation of the large number of available interconnect layers, and the judicious application of appropriate circuit techniques.

Acknowledgment

The authors are grateful to Agilent Technologies, Palo Alto, CA, for equipment, and to Cadence, San Jose, CA, and Tektronix, Beaverton, OR, for supplying the computer-aided design (CAD) tools. They are particularly grateful for the generous fabrication services provided by National Semiconductor, Santa Clara, CA.

References

[1] J. Conover. (2001, Jan.) 802.11a: Making space for speed. TechWeb. [Online]. Available: http://www.planetit.com/techcenters/docs/networks-network_systems/technology_feature/PIT20 010 111S0019.

[2] K. Pahlavan, A. Zahedi, and P. Krishnamurthy, "Wideband local access: Wireless LAN and wireless ATM," *IEEE Commun. Mag.*, pp. 34–40, Nov. 1997.

[3] *Radio equipment and systems (RES); high performance radio local area network (HIPERLAN); system definition*, ETSI TC-RES, 1994.

[4] *Radio equipment and systems (RES); high performance radio local area network (HIPERLAN); functional specifications*, ETSI TC-RES, 1995.

[5] *Radio Equipment and systems (RES); high performance radio local area network (HIPERLAN); Type 1: functional specifications, technical correction*, ETSI TC-RES, 1996.

[6] A. Dutta-Roy, "Networks for homes," *IEEE Spectrum*, vol. 36, pp. 26–33, Dec. 1999.

[7] H. Ahmadi, A. Krishna, and R. LaMaire, "Design issues in wireless LANs," *J. High-Speed Networks*, vol. 5, no. 1, pp. 87–104, 1996.

[8] B. Kleveland, T. Lee, and S. Wong, "50 GHz interconnect design in standard silicon technology," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 3, 1998, pp. 1913–1916.

[9] D. K. Weaver, "A third method of generation and detection of single sideband signals," *Proc. IRE*, vol. 44, pp. 1703–1705, Dec. 1956.

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

[11] H. Rategh and T. Lee, "Superharmonic injection locked frequency dividers," *IEEE J. Solid-State Circuits*, vol. 34, pp. 813–821, June 1998.

[12] D. Shaeffer and T. Lee, "A 1.5 V, 1.5 GHz CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745–759, May 1997.

[13] T. Soorapanth, C. P. Yue, D. Shaeffer, T. Lee, and S. Wong, "Analysis and optimization of accumulation-mode varactors for RF ICs," in *Symp. VLSI Circuits Dig.*, 1998, pp. 32–33.

[14] H. Samavati, A. Hajimiri, A. Shahani, G. Nasserbakht, and T. Lee, "Fractal capacitors," *IEEE J. Solid-State Circuits*, pp. 743–752, Dec. 1998.

[15] T. Lee, *The Design of CMOS Radio Frequency Integrated Circuits*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[16] H. Samavati, H. Rategh, and T. Lee, "A fully integrated 5GHz CMOS wireless LAN receiver," in *Int. Solid-State Circuits Conf. Tech. Dig.*, Feb. 2001, pp. 208–209.

[17] V. Aparin and P. Katzin, "Active GaAs MMIC bandpass filters with automatic frequency tuning and insertion loss control," *IEEE J. Solid-State Circuits*, vol. 30, pp. 1068–1073, Oct. 1995.

[18] J. Macedo and M. Copeland, "A 1.9-GHz silicon receiver with monolithic image filtering," *IEEE J. Solid-State Circuits*, vol. 33, pp. 378–386, Mar. 1998.

[19] M. Copeland *et al.*, "5-GHz SiGe HBT monolithic radio transceiver with tunable filtering," *IEEE Trans. Microwave Theory Tech.*, vol. 48, pp. 170–181, Feb. 2000.

[20] R. Chau, "30nm and 20nm physical gate length transistors," in *Silicon Nanoelectron. Workshop*, Kyoto, Japan, June 2001. [Online]. Available: http://www.intel.com/research/silicon/20nmtransistor.pdf.

Thomas H. Lee (S'87–M'87) received the S.B., S.M., and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1983, 1985, and 1990, respectively.

In 1990, he joined Analog Devices, where he was primarily engaged in the design of high-speed clock recovery devices. In 1992, he joined Rambus Inc., Mountain View, CA, where he developed high-speed analog circuitry for 500-MB/s CMOS DRAMs. He has also contributed to the development of PLLs in the StrongARM, Alpha, and K6/K7 microprocessors. Since 1994, he has been a Professor of electrical engineering at Stanford University, Stanford, CA, where his research focus has been on gigahertz-speed wireline and wireless ICs built in conventional silicon technologies, particularly CMOS. He authored the textbook *The Design of CMOS Radio-Frequency Integrated Circuits* (Cambridge, U.K.: Cambridge Univ. Press, 1998) and co-authored three books on RF circuit design. He holds 14 U.S. patents. He also cofounded Matrix Semiconductor.

Prof. Lee was a two-time recipient of the Best Paper Award presented at the International Solid-State Circuits Conference (ISSCC), corecipient of a Best Student Paper Award presented at the ISSCC, and the Best Paper Prize presneted at the Custom Integrated Circuits Conference (CICC). He is also the holder of a Packard Foundation Fellowship. He is a Distinguished Lecturer of the IEEE Solid-State Circuits Society and the IEEE Microwave Theory and Techniques Society (IEEE MTT-S).

Hirad Samavati received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1994 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1996 and 2001, respectively.

He is currently with Atheros Communications Inc., Sunnyvale, CA, where he is involved with a 5-GHz wireless-LAN transceiver chipset. During the summer of 1996, he was with Maxim Integrated Products, where he designed building blocks for a low-power infrared transceiver IC. His current research interests include RF circuits and analog and mixed-signal very large scale integration (VLSI), particularly integrated transceivers for wireless communications.

Dr. Samavati was the recipient of a 1995 departmental fellowship presented by Stanford University and a 1998 fellowship presented by the IBM Corporation. He was the recipient of the 1998 International Solid-State Circuits Conference Jack Kilby Outstanding Student Paper Award.

Hamid R. Rategh was born in Shiraz, Iran, in 1972. He received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1994, the M.S. degree in biomedical engineering from Case Western Reserve University, Cleveland, OH, in 1996, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2001.

During the summer of 1997, he was with Rockwell Semiconductor Systems (Conexant), Newport Beach, CA, where he was involved in the design of a CMOS dual-band GSM/DCS1800 direct conversion receiver. He co-foundeed Tavanza Inc. His current research interests are in RF ICs for high data-rate wireless LAN systems.

Dr. Rategh was a member of the Iranian team in the 21st International Physics Olympiad, Groningen, The Netherlands, in 1990. He was the recipient of the 1997 Stanford Graduate Fellowship. He was also the recipient of the 2000 Analog Devices Outstanding Student Designer Award.