

# Challenges and Trends in RF Design

Behzad Razavi

Hewlett-Packard Laboratories, Palo Alto, CA 94304

## Abstract

This paper describes the challenges and trends in today's RF industry. Beginning with a brief look at the wireless communications environment, we examine receiver architectures and their building blocks from the point of view of monolithic integration. We then present trends in circuit and architecture design, device and technology development, and wireless infrastructures.

## I. INTRODUCTION

In 1995, the worldwide sales of cellular phones exceeded \$2.5B. In the US, more than 20000 people join the cellular system *every day*. With 4.5 million customers, home satellite networks comprise a \$2.5B industry. Global positioning systems (GPS) are expected to be a \$5B market by the year 2000. In Europe, the sales of equipment and services for mobile communications will reach \$30B by 1998. The statistics are simply overwhelming . . .

The RF and wireless market has suddenly grown to unimaginable dimensions. Products such as pagers, cellular and cordless phones, GPS, home satellite networks, set-top boxes, RF IDs, and fixed satellite networks are rapidly penetrating all aspects of our lives, changing from luxury items to indispensable tools. Semiconductor companies, small and large, analog and digital, have seen the statistics and are striving to capture their own market share.

This paper describes the challenges that the RF industry faces in bringing powerful, low-cost products to the consumer market, and the strategies that are pursued in meeting this challenge. Following a brief look at the wireless communications environment, we study receiver architectures with emphasis on monolithic integration. Next, we describe the issues in designing some of the building blocks of RF systems and the relationships among circuits, architectures, and wireless standards. We conclude by considering the trends in circuit and architecture design, device and technology development, and wireless infrastructures.

## II. PROBLEM OF MULTIPLE DISCIPLINES

Unlike other areas in analog and digital electronics, RF design demands a solid understanding of many fields that are not directly related to integrated circuits (ICs). Shown in Fig. 1, the multiple disciplines contributing to RF systems have been under development for many decades. Owing to the enormity

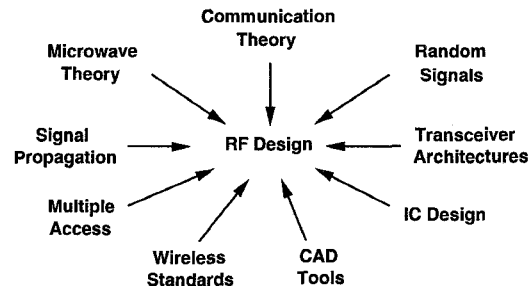


Fig. 1. Disciplines required in RF design.

of the required knowledge base, wireless systems have been traditionally designed at somewhat disjointed levels of abstraction: communication theory, RF architectures, integrated (and discrete) circuits.

With the trend toward higher integration, these boundaries need to diminish, allowing experts in various fields to speak the same language.

## III. WIRELESS COMMUNICATIONS ENVIRONMENT

The wireless communications environment, especially in urban areas, is often called "hostile" because it imposes severe constraints upon the transceiver design. Perhaps the most important constraint is the limited spectrum allocated by regulatory organizations to wireless users. From Shannon's theorem, this translates to a limited rate of information, mandating the use of sophisticated techniques such as coding, compression, and bandwidth-efficient modulation, even for voice signals.

The narrow bandwidth available to each user also impacts the design of the RF front end. As depicted in Fig. 2, the transmitter must employ narrowband amplification and filtering to avoid "leakage" to adjacent bands, and the receiver must be able to process the desired channel while sufficiently rejecting strong neighboring channels. To gain a better feeling about the latter issue, we note that if the front-end band-pass filter (BPF) in a 900-MHz receiver is to provide 60 dB of rejection at 45 kHz from the center of the channel, then the equivalent  $Q$  of the filter is on the order of  $10^7$ , a value difficult to achieve even in surface acoustic wave (SAW) filters.

The existence of large unwanted signals in the vicinity of the band of interest even after filtering creates difficulties in the design of the following circuits, in particular the front-end low-noise amplifier (LNA). As shown in Fig. 3, if the LNA exhibits nonlinearity, then the "intermodulation products" of

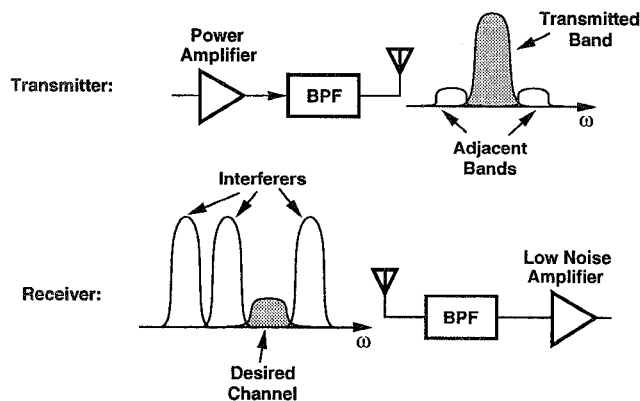


Fig. 2. Simple RF front end.

two strong unwanted signals may appear in the desired band, thereby corrupting the reception. It is interesting to note that this type of nonlinearity is important even if the signal carries no information in its amplitude.

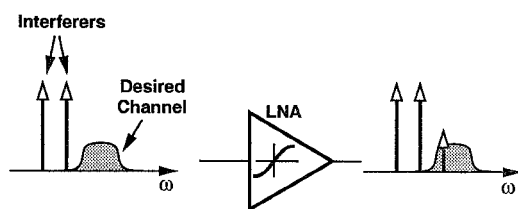


Fig. 3. Effect of third-order nonlinearity in LNA.

Another important issue in the design of wireless receivers is the dynamic range of the input signal, typically around 100 dB. Chosen so as to allow acceptable reception over a long distance in the presence of effects such as multipath fading, the dynamic range is in practice limited by a lower bound due to noise and an upper bound due to nonlinearities and saturation. The minimum detectable signal in today's wireless terminals is in the vicinity of  $-100$  dBm, thus demanding very low noise in the receive path. Saturation effects at high input levels often mandate the use of automatic gain control (AGC) in various parts of receivers.

#### IV. RECEIVER ARCHITECTURES

Complexity, cost, power dissipation, and the number of external components have been the primary criteria in selecting receiver architectures. As IC technologies evolve, architectures that once seemed impractical may return because, when they are implemented in today's advanced processes, their advantages outweigh their drawbacks. Since filter requirements prohibit channel filtering at RF, receivers first translate the input spectrum to a much lower frequency.

##### A. Heterodyne Architectures

Heterodyne receivers downconvert the input to an "intermediate frequency" (IF), perform band-pass filtering and amplification, and translate the spectrum to a lower frequency again (Fig.4). In the case of phase or frequency modulation, the last downconversion may generate both in-phase (I) and

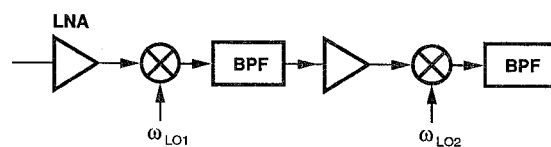
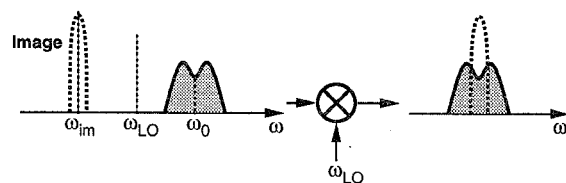


Fig. 4. Simple heterodyne architecture.

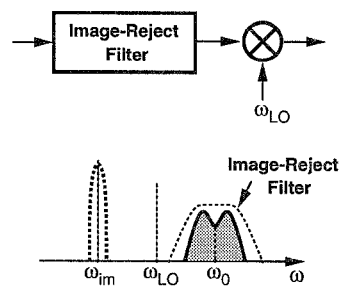
quadrature (Q) components of the signal.

Perhaps the most important feature of the heterodyne receiver is its selectivity, i.e., the capability to process and select small signals in the presence of strong interferers. While selecting a 30-kHz channel at a center frequency of 900 MHz requires prohibitively large Q's, in Fig. 4 band-pass filtering is performed at progressively lower center frequencies.

Heterodyning nonetheless entails a number of drawbacks, especially with respect to monolithic implementation. The most significant problem is the "image frequency." Since a simple mixer does not preserve the polarity of the difference between its input frequencies, it translates the bands both *above* and *below* the carrier to the same frequency [Fig. 5(a)]. Thus, the mixing operation must be preceded by an "image reject"



(a)



(b)

Fig. 5. (a) Problem of image, (b) image rejection by filtering.

filter [Fig. 5(b)]. Built as a passive off-chip device, this filter is relatively expensive and bulky and requires the LNA to drive a 50- $\Omega$  load, thereby adding another dimension to the trade-offs among noise, linearity, gain, and power dissipation of the amplifier.

Another issue in monolithic integration of heterodyne topologies is that the first IF filter must operate at typically several tens or even hundreds of megahertz while exhibiting low noise and high linearity. Consequently, this filter is also a passive, external device, further increasing the cost and complexity.

Despite these drawbacks, heterodyne architectures have been the dominant choice in RF receivers for many decades.

##### B. Homodyne Architectures

Also called the "direct-conversion" or "zero-IF" architecture, the homodyne topology translates the band of interest

directly to DC (Fig. 6). In contrast to heterodyne receivers,

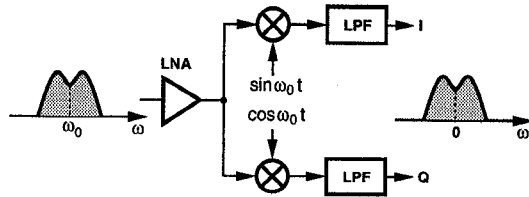


Fig. 6. Homodyne architecture.

this architecture neither suffers from the problem of image nor requires high-frequency band-pass filtering after downconversion. For these reasons, homodyne circuits are attractive for compact, efficient implementation of RF receivers.

The simplicity of direct conversion nevertheless comes with a number of design issues. First, as shown in Fig. 7, DC offsets due to mixing of the local oscillator (LO) leakage with itself

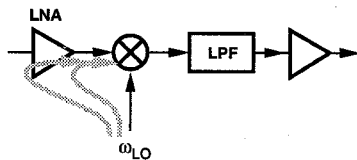


Fig. 7. DC offsets due to LO self-mixing.

corrupt the baseband signal and, more importantly, saturate the following gain stages [1]. At the output of the mixer, such offsets can be as high as 10 mV whereas the signal may be as low as a few tens of microvolts. If the modulation scheme contains significant energy near DC, as is the case in most cellular and cordless phone standards, then AC coupling with practical values of capacitors and time constants severely degrades the signal. To remove DC offsets, either periodic offset cancellation can be performed during idle times or “DC-free” modulation schemes can be adopted [2].

Second, even-order distortion in the RF signal path downconverts adjacent interferers and also demodulates amplitude-modulated components. In the presence of mismatches and hence asymmetry in the mixer, such low-frequency components appear at the output, thus degrading the signal-to-noise ratio. This effect can be reduced through the use of differential circuits or high-pass filtering the unwanted signals.

Owing to the limited gain provided by the LNA and the mixer, the downconverted signal is quite sensitive to noise. Especially problematic here is the flicker noise of baseband amplifiers and filters. The low-frequency components of noise can be suppressed by means of periodic offset cancellation (i.e., correlated double sampling) or AC coupling in conjunction with DC-free modulation.

In addition to the above issues, the leakage of the LO signal to the antenna creates interference in the band of other users and must therefore be sufficiently small (typically between  $-60$  and  $-80$  dBm). Furthermore, imbalances in the gain and phase of the I and Q paths in Fig. 6 increase the bit error rate of the detected signal.

It appears that most of these difficulties can be overcome by correction techniques in both analog and digital domains. For

this reason, homodyne architectures are currently the topic of active research.

### C. Image-Reject Architectures

The issues related to the image-reject filter have motivated RF designers to seek other techniques of rejecting the image in a heterodyne receiver. One such technique originates from a single-sideband modulator introduced by Hartley [3]. Illustrated in Fig. 8, Hartley’s circuit mixes the RF input with the

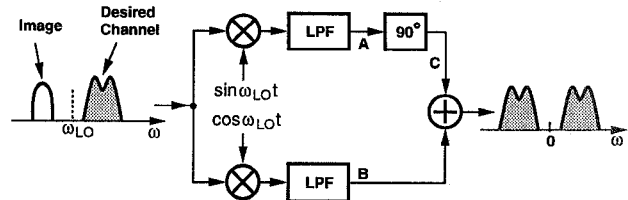


Fig. 8. Image rejection using single-sideband mixing.

quadrature outputs of the local oscillator, low-pass filters the resulting signals, and shifts one by  $90^\circ$  before adding them together. It can be shown that the spectra at points B and C contain the desired band with the same polarity and the image with opposite polarity. The summed output is therefore free from the image.

The principal drawback of the Hartley architecture is its sensitivity to mismatches. If the phase difference between the LO phases deviates from  $90^\circ$  or the gains in the upper and lower paths of Fig. 8 are not identical, then the image is not completely cancelled. Note that the effect of I/Q mismatch is much more severe here than in homodyne topologies. Also, the loss and noise of the shift-by- $90^\circ$  stage and the linearity of the adder are critical parameters.

Shown in Fig. 9 is another image-reject architecture introduced by Weaver [4]. Replacing the  $90^\circ$  shift of Hartley’s cir-

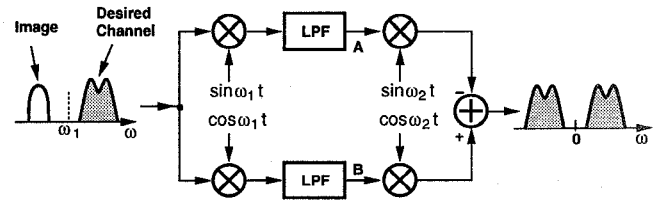


Fig. 9. Weaver architecture.

cuit with a second quadrature mixing operation, this technique provides an arbitrary translation of the signal band without image interference. It can be shown that the subtraction of the spectrum at point C from that at point D produces the signal while suppressing the image.

In addition to incomplete image cancellation resulting from I/Q mismatches, the Weaver technique exhibits high noise due to the use of four mixers. Consequently, the second set of mixers must be preceded by low-noise linear amplifiers.

## V. BUILDING BLOCKS

The performance of RF circuits is generally confined to an envelope determined by the hexagon shown in Fig. 10. In

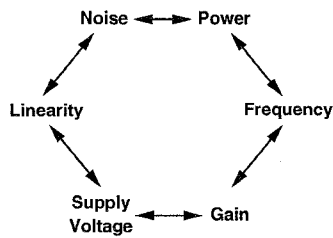


Fig. 10. RF design hexagon.

addition to low-noise amplifiers, mixers, and modulators, two of the building blocks of transceivers continue to challenge the designers: the frequency synthesizer and the transmit power amplifier.

### A. Frequency Synthesizers

The transmit and receive local oscillators are embedded in one or more frequency synthesizers. The synthesizer must generate precisely-spaced carrier frequencies according to a digital (channel select) input, e.g., 30-kHz channels from 869 MHz to 894 MHz in the receive path of the North American standard, IS54. If the frequency error is 10 parts per million, then each channel is offset by 9 kHz, an appreciable value with respect to the bandwidth of 30 kHz.

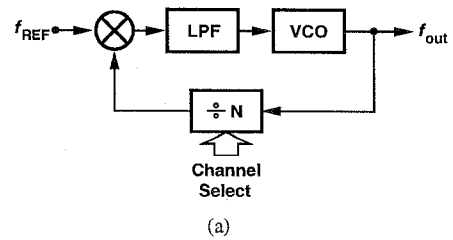
In addition to precision, the “purity” of the carrier signal is also critical. Wireless standards impose severe constraints on the phase noise and spurious contents of the synthesizer output. In fact, the principal difficulty in monolithic implementation of synthesizers is achieving low noise and spurs without the use of external resonators (e.g., inductors).

Another issue in the design of synthesizers is the lock time, i.e., the time it takes for the output frequency to change from one channel to another and settle within a specified error band around its final value. This issue is especially important in frequency-hopping spread-spectrum applications.

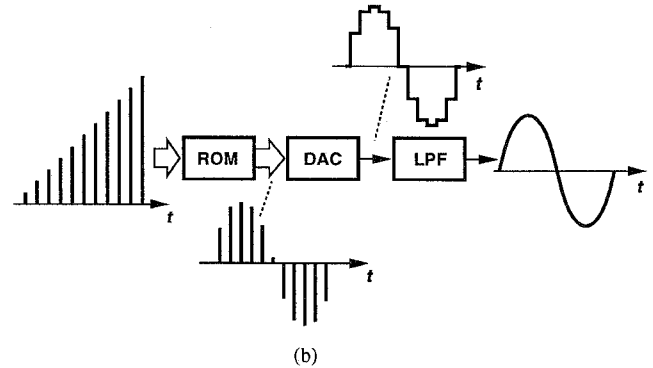
Two synthesizer architectures are shown in Fig. 11. In the phase-locked topology of Fig. 11(a), the output frequency of a voltage-controlled oscillator (VCO) is divided by a programmable factor  $N$  and locked to an accurate reference frequency. With a bandwidth roughly equal to  $f_{REF}/10$ , this configuration typically suffers from a long lock time because  $f_{REF}$  must be equal to the channel spacing. Since the phase noise of a VCO can be lowered by embedding the oscillator in a wideband phase-locked loop, it is desirable to increase the loop bandwidth of this synthesizer by about two orders of magnitude through architectural innovations.

Called “direct digital synthesis” (DDS), the approach illustrated in Fig. 11(b) first generates a *digital* sinusoid using a look-up ROM (or more efficient techniques) and subsequently converts the result to analog form by means of a digital-to-analog converter (DAC). While achieving low phase noise and fast lock, DDS requires substantial power dissipation owing to the large number of circuits operating at high speeds. Furthermore, the precision-speed-power trade-off of the DAC also limits the maximum output frequency.

To alleviate some of the above issues, DDS can be combined



(a)



(b)

Fig. 11. (a) Phase-locked synthesizer, (b) DDS synthesizer.

with phase-locking. An example is shown in Fig. 12, where a 900-MHz carrier,  $f_{PLL}$ , is generated by a wideband phase-

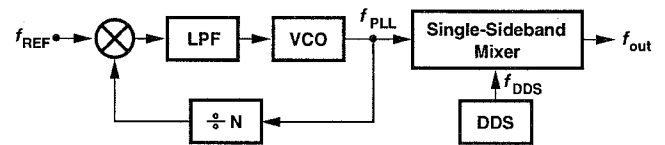


Fig. 12. Phase-locked synthesizer with DDS offset mixing.

locked loop and the required band is spanned by adding a variable DDS-based frequency,  $f_{DDS}$ , ranging from 0 to 25 MHz. For “frequency addition,” a single-sideband mixer performs the operation  $\cos \omega_1 t \cos \omega_2 t - \sin \omega_1 t \sin \omega_2 t = \cos(\omega_1 + \omega_2)t$ . Note that the PLL bandwidth can be quite large because  $f_{PLL}$  need not be variable.

The architecture of Fig. 12 faces two issues. First, if the DAC output or the mixer input circuit exhibit nonlinearity, then the SSB output is corrupted by unwanted sidebands. Second, phase and amplitude mismatches in the SSB mixer yield an undesirable component at  $f_{PLL} - f_{DDS}$ .

For all the above reasons, frequency synthesis is actively pursued in research today.

### B. Power Amplifiers

Power amplifiers are among the most power-hungry building blocks of RF transceivers, and their design is especially difficult because of supply-efficiency-linearity trade-offs.

To understand some of these issues, note that to deliver 1 W of sinusoidal power to a 50- $\Omega$  antenna, the peak-to-peak voltage swing at the antenna is approximately 20 V. If the peak-to-peak swing provided by the PA is limited to 5 V—either constrained by the supply voltage or device breakdown voltage—then a “matching network” (e.g., a transformer) is required to interface the PA with the antenna. This network

in essence transforms the  $50\text{-}\Omega$  impedance of the antenna to  $50\ \Omega/16$  seen by the PA so that a 5-V swing generates 1 W of power.

The high current levels in the PA and the matching network can introduce considerable resistive loss, thereby degrading the efficiency. At lower supply voltages, the required current levels and hence the loss are higher. As an example, consider the simple nonlinear PA shown in Fig. 13, where  $L_1$  and  $C_1$

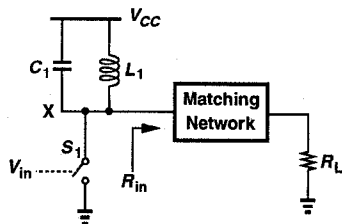


Fig. 13. Simple nonlinear power amplifier.

resonate at the carrier frequency and switch  $S_1$  pumps energy into the tank on every other transition of  $V_{in}$ . Usually implemented with silicon or GaAs field-effect transistors (FETs),  $S_1$  exhibits a finite on-resistance, dissipating power if carrying current. (Note that for the 5-V example above, the peak current drawn by the switch is approximately equal to 1.6 A.) Thus, the ideal phase relationship between  $V_{in}$  and  $V_{out}$  is such that  $S_1$  is on only when  $V_{out}$  is close to zero. Even in this case, however,  $S_1$  turns on for a greater fraction of the period because of the finite transition time of  $V_{in}$ . Now suppose the supply voltage is halved and the matching network is modified so that  $R_{in}$  decreases by a factor four. Then, to maintain the same output power, the current provided by  $S_1$  must double, and to keep the power dissipation in  $S_1$  the same, its on-resistance must be lowered by a factor of four. Since the on-resistance of FETs depends on both their gate-source overdrive voltage (i.e., the supply voltage) and their channel width, this can be accomplished by increasing the device width by a factor of 8 but at the cost of raising the input capacitance of the switch.

Certain types of phase-modulated signals experience significant band spreading (“spectral regrowth”) if they are amplified by a nonlinear PA. In these cases, the system must incorporate either linear PAs such as class A configurations or nonlinear circuits employing linearization techniques. Shown in Fig. 14 is a simplified example of the latter, where a negative feedback path consisting of a downconversion mixer, an LPF, and

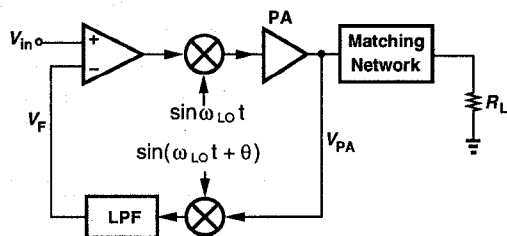


Fig. 14. PA linearization by frequency-translated feedback.

an error amplifier serves to improve the linearity of the PA [5]. As with ordinary negative feedback systems, this topology attempts to minimize the difference between  $V_{in}$  and  $V_F$ .

Thus, if the nonlinearity of the downconversion mixer and LPF is negligible,  $V_{PA}$  is a close replica of  $V_{in}$  but at the carrier frequency.

## VI. TRENDS IN RF DESIGN

The wireless fever continues to spread: even many exclusively-digital companies have come aboard the RF ship. At the same time, the wireless infrastructure is evolving to support increasingly more services and types of communication. While the standard cellular system will persist for some time, other methods continue to emerge so as to alleviate the power and cost issues of cellular phones.

Let us say that the industry’s goal is to lower the cost, power dissipation, weight, and size of the portable cellular phone so that it eventually reduces to the size of a credit card. This translates to: minimum number of (off-chip) components, low power dissipation, especially in the PA and baseband DSP(s), and fewer batteries. In this section, we consider some general approaches to achieving this goal.

### A. Microcell and Picocell Structures

Each cell in the present cellular system has a diameter of roughly 20 km. Thus, the portable phone must produce enough RF power to allow communication with the base station over a 10-km distance. Cost and efficiency issues of power amplifiers, especially as the supply voltage decreases to minimize the number of batteries, make it desirable to reduce the size of each cell. This applies particularly to the cases where most of the communication occurs within a small, but densely-populated area. Reduction of cell size can lead to “microcell” (less than 1 km in radius) and even “picocell” (less than 100 m in radius) structures, with the required RF power going down to only a few tens of milliwatts.

The principal barrier in reducing the cell size is the need for additional base stations. Complexity, cost, and *real estate* issues make it difficult to divide a crowded metropolitan area into smaller cells, each of which must be served by a new base station. Nevertheless, the concept is particularly well-suited to communication within large buildings and is also actively pursued for a wider range of applications.

### B. More Digital, Less Analog

The complexity of the typical heterodyne receiver in Fig. 4, especially the need for several off-chip, expensive, and bulky filters, has created a dream for RF system designers: discard all the analog signal processing, directly digitize the signal received by the antenna, and perform all the operations in the digital domain (Fig. 15). Since the ADC would need a dy-

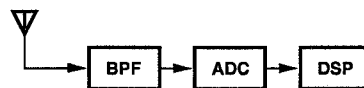


Fig. 15. Direct digitization of RF signal.

namic range of more than 100 dB and an input bandwidth of greater than 1 GHz, this dream is not practical in today’s

technology. Furthermore, even if the ADC were not the bottleneck, the required speed and power dissipation of the DSP would be excessively high. However, the idea of moving the A/D interface closer to the antenna seems promising.

A step in this direction is “IF sampling,” wherein the last mixer in the heterodyne chain is replaced with an A/D converter (Fig. 16). Since typical ADCs perform sampling before

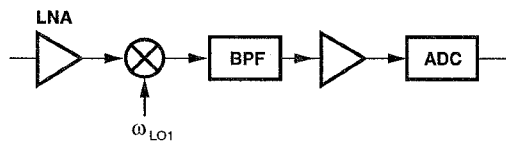


Fig. 16. IF digitization.

quantization, they can operate as downconversion mixers as well. Another possibility is to simply digitize the signal and carry out the (quadrature) mixing digitally.

As the A/D interface comes closer to the antenna, two of its parameters must simultaneously improve: the dynamic range and the input bandwidth. Since A/D converters suffer from the same trade-offs as those depicted in Fig. 10, increasing both the resolution and the input bandwidth requires substantial power dissipation penalty. As a consequence, with present technology this interface may remain at the second IF in Fig. 4, but further advances in ADC design will bring it closer to the front end.

### C. CMOS Radio

While MOSFETs were considered noisy, slow devices up to about a decade ago, scaling has dramatically improved their performance. The lower cost and faster advance of CMOS with respect to silicon bipolar and III-V technologies has motivated great efforts in designing RF CMOS circuits [6, 7]. Even though the reported performance of these circuits may not be adequate for stringent applications such as cellular phones, the potential has created a promising picture.

### D. One-Chip Radio

Another dream of RF designers is to place the entire receive and transmit paths on one chip, reducing the overall cost, power dissipation, and number of external components. The feasibility of this idea will depend on various factors such as the wireless infrastructure (e.g., picocell environment), modulation schemes, RF architectures, and IC technologies. Furthermore, the problem of crosstalk between signals whose amplitudes differ by 100 dB remains to be studied and quantified. In silicon technology, the substrate noise generated by large voltage swings can corrupt the signals at various points in the system. As depicted in Fig. 17, the digital machinery in the DSP produces noise in the analog input of the ADCs, the digital sections of the ADCs degrade the purity of the synthesizer output, and the digital counters in the synthesizer corrupt the microvolt signal received by the LNA. Careful modeling and quantification of these effects as well as methods of suppressing their contributions will be needed.

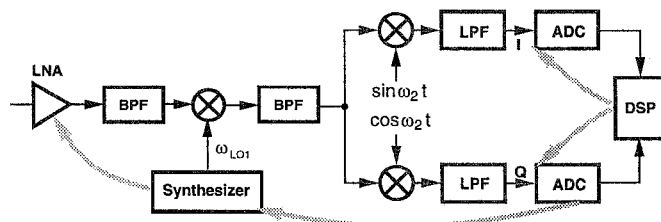


Fig. 17. Effect of substrate noise in a receiver.

### E. RF IC technologies

While cost and yield considerations make it desirable to design RF transceivers in mainstream digital VLSI technologies, tailoring an IC process to the needs of RF applications proves tremendously beneficial. Monolithic components such as inductors and transformers—considered exotic and impractical in silicon a few years ago—have appeared in recent circuits despite their inferior performance with respect to discrete counterparts. Moreover, technologies that exhibit less substrate noise, e.g., silicon-on-insulator (SOI), are under investigation for highly-integrated RF systems.

A serious difficulty in utilizing IC technologies for RF applications is poor device modeling. The high level of integration and large-signal operation prohibit the use of S parameters while typical SPICE models suffer from many shortcomings in representing the high-frequency behavior of transistors. Another problem is that RF CAD tools are still in their infancy, forcing designers to rely more on experience and intuition. Further work in RF device modeling and CAD tools is essential.

### REFERENCES

- [1] A. A. Abidi, “Direct-conversion radio transceivers for digital communications,” *IEEE Journal of Solid-State Circuits*, vol. 30, pp. 1399-1410, Dec. 1995.
- [2] C. D. Hull, R. R. Chu, and J. L. Tham, “A direct-conversion receiver for 900 MHz (ISM band) spread-spectrum digital cordless telephone,” *ISSCC Dig. Tech. Papers*, pp. 344-345, Feb. 1996.
- [3] R. Hartley, “Single-sideband modulator,” US Patent: 1,666,206, April 1928.
- [4] D. K. Weaver, “A third method of generation and detection of single-sideband signals,” *Proc. IRE*, vol. 44, pp. 1703-1705, 1956.
- [5] M. Johansson and T. Mattsson, “Transmitter linearization using Cartesian feedback for linear TDMA modulation,” *Proc. IEEE Veh. Tech. Conf.*, pp. 542-546, 1990.
- [6] A. Rofougaran *et al.*, “A 1-GHz CMOS RF front-end IC with wide dynamic range,” *Proc. ESSCIRC*, pp. 250-253, 1995.
- [7] A. Karanicolas, “A 2.7 V 900 MHz CMOS LNA and Mixer,” *ISSCC Dig. Tech. Papers*, pp. 50-51, Feb. 1996.