

Engin
TK
6553
.B633
2008

RIS BOW



LYLER AND CHERYL AJLUNI

RF CIRCUIT design

SECOND EDITION

A best-seller now thoroughly revised

Two new chapters on RF Front-End
Design and RF Design Tools

Perfect for the practical, hard-working
RF professional



BOWICK

RF CIRCUIT DESIGN

SECOND EDITION



Cover image by iStockphoto

Newnes is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
Linacre House, Jordan Hill, Oxford OX2 8DP, UK

Copyright © 2008, Elsevier Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, E-mail: permissions@elsevier.com. You may also complete your request online via the Elsevier homepage (<http://elsevier.com>), by selecting "Support & Contact" then "Copyright and Permission" and then "Obtaining Permissions."



Recognizing the importance of preserving what has been written, Elsevier prints its books on acid-free paper whenever possible.

Library of Congress Cataloging-in-Publication Data

Bowick, Chris.

RF circuit design / Christopher Bowick. — 2nd ed.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-7506-8518-4

ISBN-10: 0-7506-8518-2

1. Radio circuits Design and construction. 2. Radio frequency. I. Title.

TK6553.B633 2008

621.384'12—dc22

2007036371

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 978-0-7506-8518-4

For information on all Newnes publications
visit our web site at <http://books.elsevier.com>

Typeset by Charon Tec Ltd (A Macmillan Company), Chennai, India
www.charontec.com

07 08 09 10 10 9 8 7 6 5 4 3 2 1

Printed in the United States of America

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

CONTENTS

Preface	ix
Acknowledgments	xi
 CHAPTER 1	 1
Components and Systems	
Wire – Resistors – Capacitors – Inductors – Toroids – Toroidal Inductor Design – Practical Winding Hints	
 CHAPTER 2	 23
Resonant Circuits	
Some Definitions – Resonance (Lossless Components) – Loaded Q – Insertion Loss – Impedance Transformation – Coupling of Resonant Circuits – Summary	
 CHAPTER 3	 37
Filter Design	
Background – Modern Filter Design – Normalization and the Low-Pass Prototype – Filter Types – Frequency and Impedance Scaling – High-Pass Filter Design – The Dual Network – Bandpass Filter Design – Summary of the Bandpass Filter Design Procedure – Band-Rejection Filter Design – The Effects of Finite Q	
 CHAPTER 4	 63
Impedance Matching	
Background – The L Network – Dealing With Complex Loads – Three-Element Matching – Low- Q or Wideband Matching Networks – The Smith Chart – Impedance Matching on the Smith Chart – Software Design Tools – Summary	
 CHAPTER 5	 103
The Transistor at Radio Frequencies	
RF Transistor Materials – The Transistor Equivalent Circuit – Y Parameters – S Parameters – Understanding RF Transistor Data Sheets – Summary	
 CHAPTER 6	 125
Small-Signal RF Amplifier Design	
Some Definitions – Transistor Biasing – Design Using Y Parameters – Design Using S Parameters	

CHAPTER 7	169
RF (Large Signal) Power Amplifiers	
RF Power Transistor Characteristics – Transistor Biasing – RF Semiconductor Devices – Power Amplifier Design – Matching to Coaxial Feedlines – Automatic Shutdown Circuitry – Broadband Transformers – Practical Winding Hints – Summary	
CHAPTER 8	185
RF Front-End Design	
Higher Levels of Integration – Basic Receiver Architectures – ADC'S Effect on Front-End Design – Software Defined Radios – Case Study—Modern Communication Receiver	
CHAPTER 9	203
RF Design Tools	
Design Tool Basics – Design Languages – RFIC Design Flow – RFIC Design Flow Example – Simulation Example 1 – Simulation Example 2 – Modeling – PCB Design – Packaging – Case Study – Summary	
APPENDIX A	227
APPENDIX B	229
BIBLIOGRAPHY	233
INDEX	237

RF FRONT-END Design

Throughout most of this book, we have studied the components that make up RF circuits. Now we put all these pieces together to form one of the most critical subsystems in any communication system—namely, the *RF front end*. We'll start by showing where the RF front end fits in today's modern applications, and then decompose or "tear down" this subsystem into its basic components and functions, many of which have already been covered in earlier chapters. This approach will also provide a logical way to introduce key receiver types and associated performance specifications, such as signal-to-noise ratio (SNR), receiver sensitivity, and selectivity.

The RF front end is part of an overall radio receiver-transmitter or transceiver system. It is generally defined as everything between the antenna and the digital baseband system. For a receiver, this "between" area includes all the filters, low-noise amplifiers (LNAs), and down-conversion mixer(s) needed to process the modulated signals received at the antenna into signals suitable for input into the baseband analog-to-digital converter (ADC). For this reason, the RF front end is often called the *analog-to-digital* or *RF-to-baseband* portion of a receiver.

Radios work by receiving RF waves containing previously modulated information sent by a RF transmitter. The receiver is basically a low noise amplifier that down converts the incoming signal. Hence, sensitivity and selectivity are the primary concerns in receiver design.

Conversely, a transmitter is an up converts an outgoing signal prior to passage through a high power amplifier. In this case, non-linearity of the amplifier is a primary concern. Yet, even with these differences, the design of the receiver front end and transmitter back end share many common elements—like local oscillators. In this chapter, we'll concentrate our efforts on understanding the receiver side.

Thanks to advances in the design and manufacture of integrated circuits (ICs), some of the traditional analog IF signal-processing tasks can be handled digitally. These traditional analog tasks, like filtering and up-down conversion, can now be handled by means of digital filters and digital signal processors (DSPs). Texas Instruments have coined the term *digital radio processors* for this type of circuit. This migration of analog into digital circuits means that the choice of what

front-end functions are implemented by analog and digital means generally depends on such factors as required performance, cost, size, and power consumption. Because of the mix of analog and digital technologies, RF front end chips using mixed-signal technologies may also be referred to as RF-to-digital or RF-to-baseband (RF/D) chips.

Why is the front end so important? It turns out that this is arguably the most critical part of the whole receiver. Trade-offs in overall system performance, power consumption, and size are determined between the receiver front end and the ADCs in the baseband (middle end). In more detail, the analog front end sets the stage for what digital bit-error-rate (BER) performance is *possible* at final bit detection. It is here that the receiver can, within limits, be designed for the best potential SNR.

HIGHER LEVELS OF INTEGRATION

Look inside any modern mobile phone, multimedia device, or home-entertainment control system that relies on the reception and/or transmission of wireless signals and you'll find an RF front end. In the RIM Blackberry PDA, for example, the communication system consists of both a transceiver chip and RF front-end module (see Fig. 8-1). The front-end module incorporates several integrated circuits (ICs) that may be based on widely different semiconductor processes, such as conventional silicon CMOS and advanced silicon germanium (SiGe) technologies. Functionally, such multichip modules provide most if not all of the analog signal processing—filtering, detection, amplification and demodulation via a mixer. (The term "system-in-package" or SIP is a synonym for multichip module or MCM.)

Multichip front-end modules demonstrate an important trend in RF receiver design, namely, ever-increasing levels of system integration required to squeeze more functionality into a single chip. The reasons for this trend—especially in consumer electronics—come from the need for lower costs, lower power consumption (especially in mobile and portable products), and smaller product size. Still, regardless of the level of integration, the basic RF architecture remains unchanged: signal filtering, detection, amplification and demodulation. More specifically, a modulated RF carrier signal couples with an antenna designed for a specific band of frequencies. The antenna passes the modulated signals along to the RF

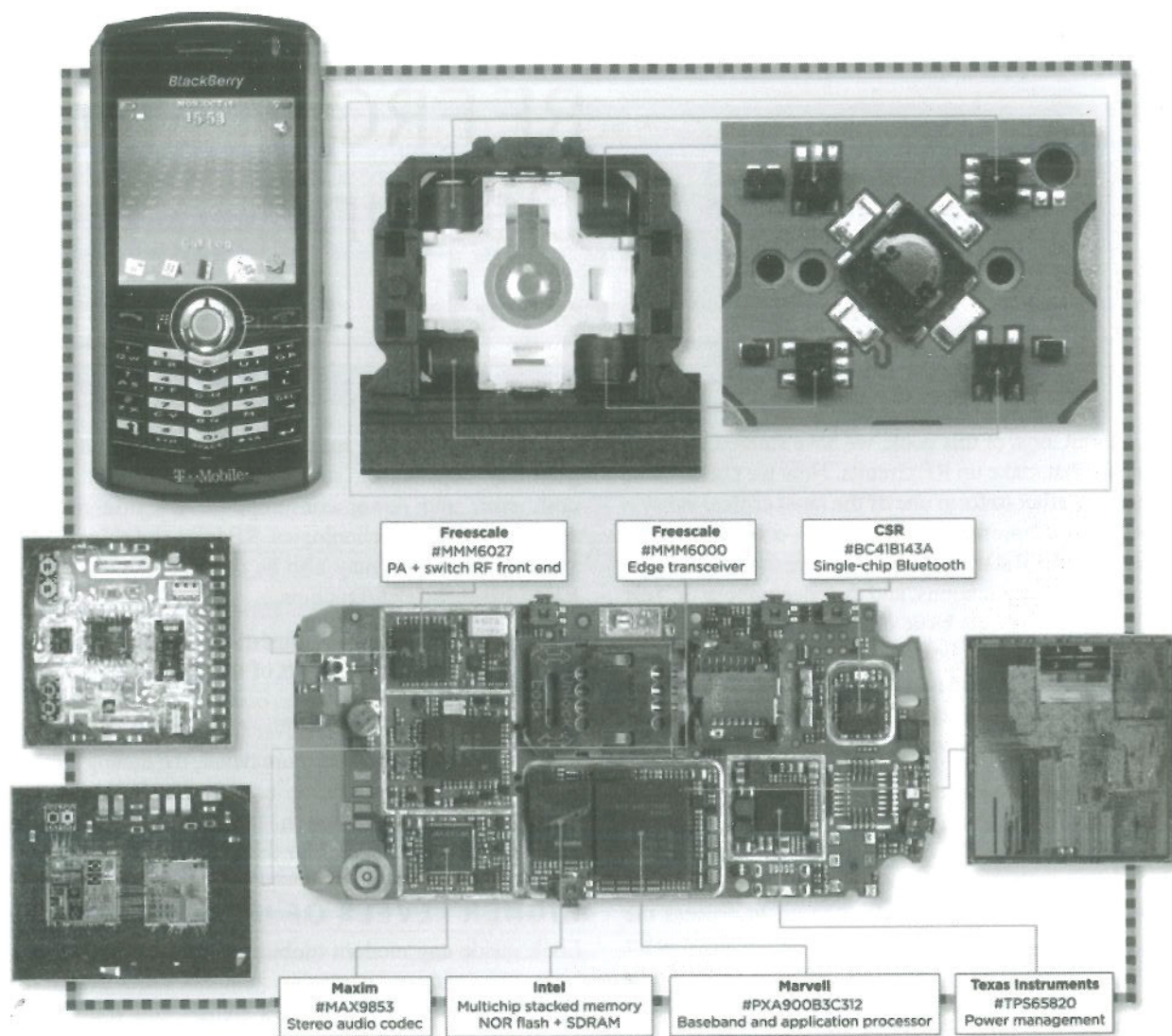


FIG. 8-1. Tear down of modern mobile device reveals several RF front-end chips. (Courtesy of iSuppli)

receiver's front end. After much conditioning in the front-end circuitry, the modulation or information portion of the signal—now in the form of an analog baseband signal—is ready for analog-to-digital conversion into the digital world. Once in the digital realm, the information can be extracted from the digitized carrier waveforms and made available as audio, video, or data.

Before the advent of such tightly integrated modules, each functional block of the RF front end was a separate component, designed separately. This means that there were separate components for the RF filter, detector, mixer-demodulator, and amplifier. More importantly, this meant that all of these physically independent blocks had to be connected together. To prevent signal attenuation and distortion and to minimize signal reflections due to impedance differences between function blocks, components were standardized for a characteristic impedance of 50 ohms, which was also the impedance of

high-frequency test equipment. The 50-ohm coaxial cable [1] interface was a trade-off that minimized signal attenuation while maximizing power transfer—signal energy—between the independently designed RF filter, LNA, and mixer. Before higher levels of functional integration and thus lower costs could be achieved, it was necessary to design and manufacture these RF functional blocks using standard semiconductor processes, such as silicon CMOS IC processes.

Unfortunately, one of the drawbacks of CMOS technology can be the difficulty in achieving a 50-ohm input impedance. Still, it is only necessary to have the 50-ohm matched input and output impedances when the connection lines between the sub-circuits is long compared to the wavelength of the carrier wave. For ICs and MCMs at GHz frequencies, connections lines are short, so 50-ohm between sub-circuits isn't a problem. It is necessary to somehow get to 50 ohms to connect to the (longer) printed circuit board traces.

This is but one example of the changes that have taken place with modern integrated front ends. We will not cover all the changes here. Instead, we'll focus on the important design parameters that can affect the design of an RF front end, including the signal-to-noise ratio (SNR), receiver sensitivity, receiver and channel filter selectivity, and even the bit resolution of the ADC (covered later). This high-level description of the RF front end reveals not only the basic functioning but also the potential system trade-offs that must be considered.

As mentioned earlier, the basic stages of an RF front end include an antenna, filter, detector-demodulator, and amplifier. Each of these signal-conditioning stages contains unique circuit components, many of which have already been covered in earlier portions of this book. None of these components work in isolation and the performance of one component may well affect the performance of another. This is why we'll look at each of these key component function blocks in the context of several different radio architectures: detector, direct-conversion, and superheterodyne receiver configurations.

BASIC RECEIVER ARCHITECTURES

The fundamental operation of an RF front end is fairly straightforward: it detects and processes radio waves that have been transmitted with a specific known frequency or range of frequencies and known modulation format. The modulation carries the information of interest, be it voice, audio, data, or video. The receiver must be tuned to resonate with the transmitted frequency or frequencies in order to detect them. Those received signals are then filtered from all surrounding signals and noise and amplified prior to a process known as demodulation, which removes the desired information from the radio waves that carried it.

These three steps—filtering, amplification and demodulation—detail the overall process. But actual implementation of this process (i.e., designing the physical RF receiver printed-circuit board (PCB)) depends upon the type, complexity, and quantity of the data being transmitted. For example, designing an RF front end to handle a simple amplitude-modulated (AM) signal requires far less effort and hardware (and even software) than building an RF front end for the latest third-generation (3G) mobile telecommunications handset.

Because of the enhanced performance of analog components due to IC process improvements and decreasing costs of more powerful digital-signal-processing (DSP) hardware and software functions, the ways that different RF front-end architectures are realized has changed over the years. Still, the basic requirements for an RF front end, such as the frequency range and type of carrier to be received, the RF link budget, and the power, performance, and size restrictions of the front-end design, remain relatively the same in spite of the differences in radio architectures. Let's start by looking at the simplest of radio architectures or implementations.

AM Detector Receivers

One of the basic RF receiver architectures for detecting a modulated signal is the amplitude modulation (AM) detector receiver (see Fig. 8-2). The name comes from the fact that information like speech and music could be converted into amplitude (voltage) modulated signals riding on a carrier wave. Such an RF signal could be de-modulated at the receiving end by means of a simple diode detector. All that is needed for a basic AM receiver—like a simple crystal radio—is an antenna, RF filter, detector, and (optional) amplifier to boost the recovered information to a level suitable for a listening device, such as a speaker or headphones. The antenna, which is capacitive at the low frequencies used for AM broadcasting, is series matched with an inductor to maximize current through both, thus maximizing the voltage across the secondary coil (see Chapter 2). A variable capacitance filter may be used to select the designed frequency band (or channel) and to block any unwanted signals, such as noise. The filtered signal is then converted to demodulate the AM signal and recover the information. Fig. 8-3 represents a schematic version of the block diagram shown in Fig. 8-2.

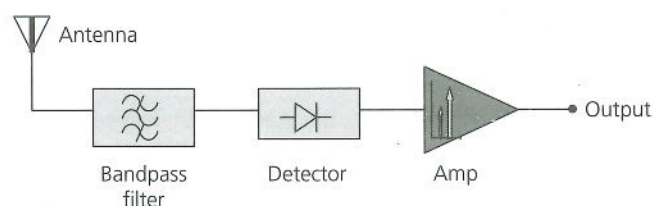


FIG. 8-2. Simple amplitude modulation (AM) radio architecture.

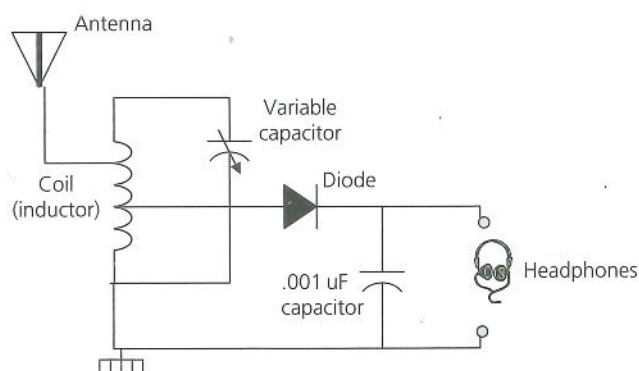


FIG. 8-3. Circuit schematic of a simple AM radio.

The heart of the AM architecture is the detector demodulator. In early crystal radios, the detector was simply a fine metal wire that contacted a crystal of galena (lead sulfide), thus creating a point contact rectifier or "crystal detector." In these early designs, the fine metal contact was often referred to as a "catwhisker." Although point-contact diodes are still in use today in communication receivers and radar, most have been

replaced by pn-junction diodes, which are more reliable and easier to manufacture.

For a simple AM receiver, the detector diode acts as a half-wave rectifier to convert or rectify a received AC signal to a DC signal by blocking the negative or positive portion of the waveform (see Fig. 8-4). A half-wave rectifier clips the input signal by allowing either the positive or negative half of the AC wave to pass easily through the rectifier, depending upon the polarity of the rectifier. A shunt inductor is typically placed in front of the detector to serve as an RF choke (see Chapter 2). The inductor maintains the input to the detector diode at DC ground while preserving a high impedance in parallel with the diode, thus maintaining the RF performance.

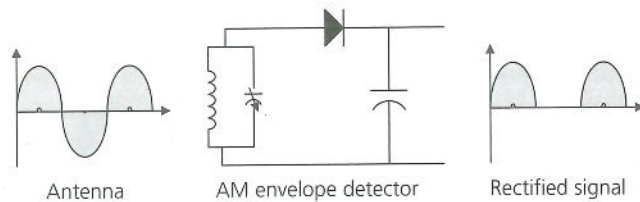


FIG. 8-4. Circuit schematic of a half-wave rectifier.

In a simple detector receiver, the AM carrier wave excites a resonance in the inductor/tuned capacitor (LC) tank subcircuit. The tank acts like a local oscillator (LO), to the current through the diode is proportional to the amplitude of the resonance and this gives the baseband signal (typically analog audio). The baseband signal may be in either analog or digital format, depending upon the original format of the information used to modulate the AM carrier. As we shall see, this process of translating a signal down or up to the baseband level becomes a critical technique in most modern radios. The exception is time domain or pulse position modulation. Interestingly, this scheme dates back to the earliest (spark gap) radio transmitters. It's strange how history repeats itself. Another example is that the earliest radios were digital (Morse code), then analog was considered superior (analog voice transmission), now digital is back!

The final stage of a typical AM detector system is the amplifier, which is needed to provide adequate drive levels for an audio listening device, such as a headset or speaker. One of the disadvantages of the signal diode detector is its poor power transfer efficiency. But to understand this deficiency, you must first understand the limitation of the AM design that uses a half-wave rectifier at the receiver. At transmission from the source, the AM signal modulation process generates two copies of the information (voice or music) plus the carrier. For example, consider an AM radio station that broadcasts at a carrier frequency of 900 kHz. The transmission might be modulated by a 1000-Hz (1-kHz) signal or tone. The RF front end in an AM radio receiver will pick up the 900-kHz carrier signal along with the 1-kHz plus and minus modulation around the carrier, at frequencies of 901 and 899 kHz, respectively (see Fig. 8-5). The modulation frequencies are also known as the upper and lower

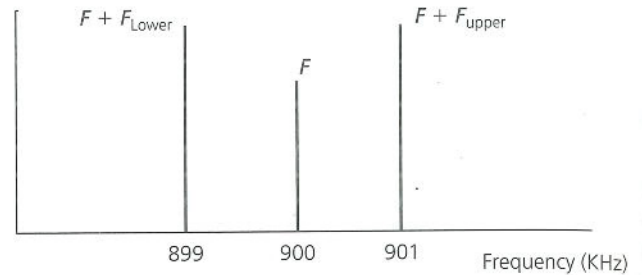


FIG. 8-5. Half-wave rectified AM radio produces upper and lower sidebands.

sideband frequencies, respectively. But only one of the sidebands is needed to completely demodulate the received signal. The other sideband contains duplicate information. Thus, the disadvantages of AM transmissions are twofold: (1) for a given information bandwidth, twice that bandwidth is needed to convey the information, and (2) the power used to transmit the unused sideband is wasted (typically, up to 50% of the total transmitted power).

Naturally, there are other ways to demodulate detector-based receiver architectures. We have just covered an approach used in popular AM receivers. Replacing the diode detector with another detector type would allow us to detect frequency-modulated (FM) or phase-modulated (PM) signals, this latter modulation commonly used in transmitting digital data. For example, many modern telecommunication receivers rely heavily on *phase-shift keying* (PSK), a form of phase (angle) modulation. The phrase "shift keying" is an older expression (from the Morse code era) for "digital."

All detector circuits are limited in their capability to differentiate between adjacent signal bands or channels. This capability is a measure of the *selectivity* of the receiver and is a function of the input RF filter to screen out unwanted signals and to pass (select) only the desired signals. As you might have guess from earlier discussions (see Chapter 2), selectivity is related to the *quality factor* or *Q* of the RF filter. A high *Q* means that the circuit provides sharp filtering and good differentiation between channels—a must for modern communication systems. Unfortunately, tuning the center carrier frequency of the filter across a large bandwidth while maintaining a high differentiation between adjacent channels is very difficult at the higher frequencies found in today's mobile devices. Selectivity across a large bandwidth is complicated by a receiver's *sensitivity* requirement, or the need to detect very small signals in the presence of system noise—noise that comes from the earth (thermal noise), not just the receiver system itself. The sensitivity of receiving systems is defined as the smallest signal that leads to an acceptable *signal-to-noise ratio* (SNR).

Receiver selectivity and sensitivity are key technical performance measures (TPMs) and will be covered in more detail in this chapter. At this point, it is sufficient to note that the AM diode detector architecture is limited in selectivity and sensitivity.

TRF Receiver

Moving up the scale in complexity, we come to the next evolutionary RF architecture: the tuned-radio-frequency (TRF) receiver (see Fig. 8-6). This early design was one of the first to use amplification techniques to enhance the quality of the signal reception. A TRF receiver consisted of several RF stages, all simultaneously tuned to the received frequency before detection and subsequent amplification of the audio signal. Each tuned stage consisted of a bandpass filter—which need not be an LC tank filter but could also be a Surface Acoustic Wave (SAW) filter or a dielectric cavity filter—with an amplifier to boost the desired signal while reducing unwanted signals such as interference. The final stage of the design is a combination of a diode rectifier and audio amplifier, collectively known as a grid-leak detector. In contrast to other radio architectures, there is no translation in frequency of the input signals, and no mixing of these input signals with those from a tunable LO. The original input signal is demodulated at the detector stage. On the positive side, this simple architecture does not generate the image signals that are common to other receiver formats using frequency mixers, such as superheterodynes (covered later in this chapter).

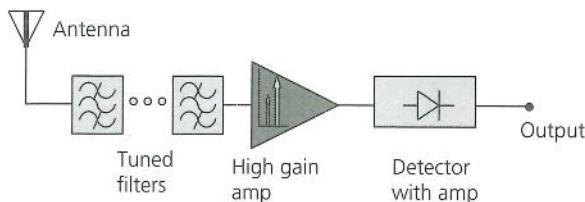


FIG. 8-6. Tuned Radio Frequency (TRF) architecture emphasizes multiple tuned filters.

The addition of each LC filter-amplifier stage in a TRF receiver increases the overall selectivity. On the downside, each such stage must be individually tuned to the desired frequency since each stage has to track the previous stage. Not only is this difficult to do physically, it also means that the received bandwidth increases with frequency. For example, if the circuit Q was 50 at the lower end of the AM band, say 550 kHz, then the receiver bandwidth would be $500/50$ or 11 kHz—a reasonable value. However at the upper end of the AM spectrum, say 1650 kHz, the received bandwidth increases to $1650/50$ or 33 kHz.

As a result, the selectivity in a TRF receiver is not constant, since the receiver is more selective at lower frequencies and less selective at higher frequencies. Such variations in selectivity can cause unwanted oscillations and modes in the tuned stages. In addition, amplification is not constant over the tuning range. Such shortcomings in the TRF receiver architecture have led to more widespread adoption of other receiver architectures, including direct-conversion and superheterodyne receivers, for many modern wireless applications.

Direct-Conversion Receiver

A way to overcome the need for several individually tuned RF filters in the TRF receiver is by directly converting the original signal to a much lower baseband frequency. In the direct-conversion receiver (DCR) architecture, frequency translation is used to change the high input frequency carrying the modulated information into a lower frequency that still carries the modulation but which is easier to detect and demodulate. This frequency translation is achieved by mixing the input RF signal with a reference signal of identical or near-identical frequency (see Fig. 8-7). The nonlinear mixing of the two signals results in a baseband signal prior to the detection or demodulating stage of the front-end receiver.

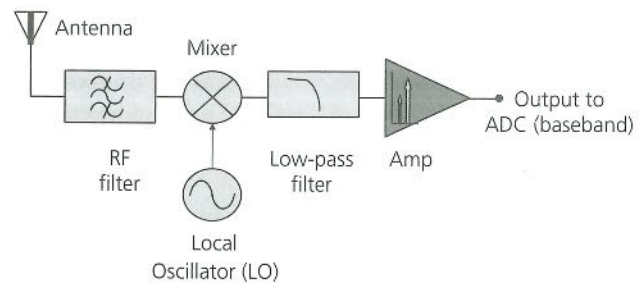


FIG. 8-7. Direct Conversion Receiver (DCR) architecture.

The reference signal is generated by a local oscillator (LO). When an input RF signal is combined in a nonlinear device, such as a diode or field-effect-transistor (FET) mixer, with an LO signal, the result is an intermediate-frequency (IF) signal that is the sum or difference of the RF and LO signals. When the LO signal is chosen to be the same as the RF input signal, the receiver is said to have a homodyne (or “same frequency”) architecture and is also known as a zero-IF receiver. Conversely, if the reference signal is different from the frequency to be detected, then it’s called a heterodyne (or “different frequency”) receiver. The terms superheterodyne and heterodyne are synonyms (“super” means “higher” or “above” not “better”).

In either homodyne or heterodyne approaches, new frequencies are generated by mixing two or more signals in a nonlinear device, such as a transistor or diode mixer. The mixing of two carefully chosen frequencies results in the creation of two new frequencies, one being the sum of the two mixed frequencies and the other being the difference between the two mixed signals. The lower frequency is called the beat frequency, in reference to the audio “beat” that can be produced by two signals close in frequency when the mixing product is an actual audio-frequency (AF) tone.

For example, if a frequency of 2000 Hz and another of 2100 Hz were beat together, then an audible beat frequency of 100 Hz would be produced. The end result is a frequency shifting from a higher frequency to lower—and in the case of RF receivers—baseband frequency.

Direct conversion or homodyne (zero-IF) receivers use an LO synchronized to the exact frequency of the carrier in order to directly translate the input signals to baseband frequencies. In theory, this simple approach eliminates the need for multiple frequency downconversion stages along with their associated filters, frequency mixers, and LOs. This means that a fixed RF filter can be used after the antenna, instead of multiple tuned RF filters as in the TRF receiver. The fixed RF filter can thus be designed to have a higher Q .

In direct-conversion design, the desired signal is obtained by tuning the local oscillator to the desired signal frequency. The remaining unwanted frequencies that appear after downconversion stay at the higher frequency bands and can be removed by a lowpass filter placed after the mixer stage.

If the incoming signal is digitally encoded, then the RF receiver uses digital filters within a DSP to perform the demodulation. Two mixers are needed to retain both the amplitude and phase of the original modulated signal: one for the in-phase (I) and another for a quadrature (Q) baseband output. Quadrature downconversion is needed since two sidebands generally form around any RF carrier frequency. As we have already seen, these sidebands are at different frequencies. Thus, using a single mixer, for a digitally encoded signal, would result in the loss of one of the sidebands. This is why an I/Q demodulator is typically used for demodulating the information contained in the I and Q signal components.

Unfortunately, many direct-conversion receivers are susceptible to spurious LO leakage, when LO energy is coupled to the I/Q demodulator by means of the system antenna or via another path. Any LO leakage can mix with the main LO signal to generate a DC offset, possibly imposing potentially large DC offset errors on the frequency-translated baseband signals. Through careful design, LO leakage in a direct-conversion receiver can be minimized by maintaining high isolation between the mixer's LO and RF ports.

Perhaps the biggest limitation of direct-conversion receivers is their susceptibility to various noise sources at DC, which creates a DC offset. The sources of unwanted signals typically are the impedance mismatches between the amplifier and mixer.

As noted earlier in this chapter, improvements in IC integration via better control of the semiconductor manufacturing process have mitigated many of the mismatch-related DC offset problems.

Still another way to solve DC offset problems is to downconvert to a center frequency near, but not at, zero. Near-zero IF receivers do just that, by downconverting to an intermediate frequency (IF) which preserves the modulation of the RF signal by keeping it above the noise floor and away from other unwanted signals. Unfortunately, this approach creates a new problem, namely that the image frequency and the baseband beat signals that arise from inherent signal distortion, can both fall within the intermediate band. The image frequencies, to be covered later, can be larger than the desired signal frequency,

thus causing resolution challenges for the analog-to-digital converter.

Superheterodyne Receivers

In contrast to the simplicity of the direct-conversion receiver, the superheterodyne receiver architecture often incorporates multiple frequency translation stages along with their associated filters, amplifiers, mixers, and local oscillators (see Fig. 8-8). But in doing so, this receiver architecture can achieve unmatched selectivity and sensitivity. Unlike the direct-conversion receiver in which the LO frequencies are synchronized to the input RF signals, a superheterodyne receiver uses an LO frequency that is offset by a fixed amount from the desired signal. This fixed amount results in an intermediate frequency (IF) generated by mixing the LO and RF signals in a nonlinear device such as a diode or FET mixer.

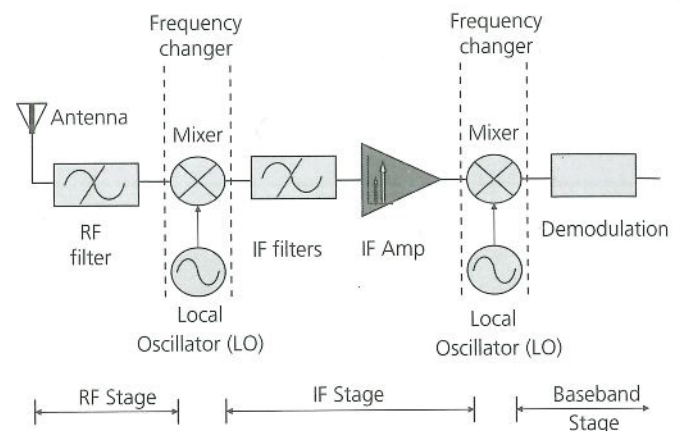


FIG. 8-8. Superheterodyne Architecture.

Generating local oscillators

The LO is often a phase-locked voltage-controlled oscillator (VCO) capable of covering the frequency range of interest for translating incoming RF signals to a desired IF range. In recent years, a number of other frequency-stabilization techniques, including analog fractional-N frequency synthesis and integer-N frequency synthesis as well as direct-digital-synthesis (DDS) approaches, have been used to generate the LO signals in wireless receiver architectures for frequency translation.

Any LO approach should provide signals over a frequency band of interest with the capability of tuning in frequency increments that support the system's channel bandwidths. For example, a system with 25-kHz channels is not well supported by a synthesized LO capable of tuning in minimum steps of only 1 MHz. In addition, the LO should provide acceptable single-sideband (SSB) phase-noise performance, specified at an offset frequency that coincides with the system's channel spacing. Referring to an LO's SSB phase noise offset 1 MHz from the carrier will not provide enough information about the phase noise that is closer to the carrier and that may affect communications systems

performance in closely spaced channels. Phase noise closer to the carrier is typically specified at offset frequencies of 1 kHz or less.

The LO source should also provide adequate drive power for the front-end mixers. In some cases, an LO buffer amplifier may be added to increase the signal source's output to the level required to achieve acceptable conversion loss in the mixer. And for portable applications, the power supply and power consumption of the LO become important considerations when planning for a power budget.

Mixers

Mixers are an integral component in any modern radio front end (see Fig. 8-9). Frequency mixers can be based on a number of different nonlinear semiconductor devices, including diodes and field-effect transistors (FETs). Because of their simplicity and capability of operation without DC bias, diode mixers have been prevalent in many wireless systems. Mixers based on diodes have been developed in several topologies, including single-ended, single-balanced, and double-balanced mixers. Additional variations on these configurations are also available, such as image-reject mixers and harmonic mixers which are typically employed at higher, often millimeter-wave, frequencies.

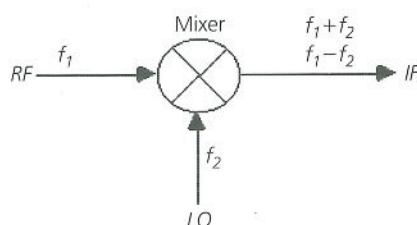


FIG. 8-9. Circuit symbol for radio mixer—or a diode multiplier in this example.

The simplest diode mixer is the single-ended mixer, which can be formed with an input balanced-unbalanced (balun) transformer, a single diode, an RF choke, and a lowpass filter. In a single-diode mixer, insertion loss results from conversion loss, diode loss, transformer loss. The mixer sideband conversion is nominally 3 dB, while the transformer losses (balun losses) are about 0.75 dB on each side, and there are diode losses because of the series resistances of the diodes. The equivalent circuit of a diode consists of a series resistor and a time-variable electronic resistor.

Moving up slightly in complexity, a single-ended mixer consists of a single diode, input matching circuitry, balanced-unbalanced (balun) transformer or some other means for injecting a mixing signal with the RF input signal, and a lowpass or bandpass filter to pass desired mixer products and reject unwanted signal components. Single-ended mixers are inexpensive and often used in low-cost detectors, such as motion detectors. The input balun must be highly selective to prevent radiation of the LO signal back into the RF port and out of the antenna.

Although the behavior of the diode changes with LO level, it can be matched for impedance at a particular frequency, such as the

LO frequency, to achieve fairly consistent conversion-loss performance and flatness. The desired frequency converted signals are available at the IF port; the filter eliminates the unwanted high-frequency signal components generated by the mixing process. The LO drive level can be arbitrary, although different types of mixers and their diodes generally dictate an optimum LO drive level for mixer operation. The dimensions of the diode will dictate the frequency of operation, allowing use through millimeter-wave frequencies if the diode is made sufficiently small.

Some single-ended mixers use an antiparallel diode pair in place of the single diode to double the LO frequency and use the second harmonics of the LO's fundamental frequency, somewhat simplifying the IF filtering requirements. The trade-off involves having to supply higher LO power in order to achieve sufficient mixing power by means of the LO's second-harmonic signals.

A single-balanced mixer uses two diodes connected back to back. In the back-to-back configuration, noise components from the LO or RF that are fed into one diode are generated in the opposite sense in the other diode and tend to cancel at the IF port.

A double-balanced mixer is typically formed with four diodes in a quad configuration (see Fig. 8-10). The quad configuration provides excellent suppression of spurious mixing products and good isolation between all ports. Because of the symmetry, the LO voltage is sufficiently isolated from the RF input port and no RF voltage appears at the LO port. With a sufficiently large LO drive level, strong conduction occurs in alternate pairs of diodes, changing them from a low to high resistance state during each half of the LO's frequency cycle.

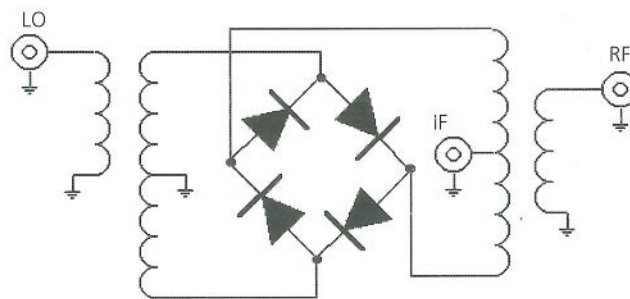


FIG. 8-10. Circuit schematic for a diode ring, double-balanced mixer.

Because the RF voltage is distributed across the four diodes, the 1-dB compression point is higher than that of a single-balanced mixer, although more LO power is needed for mixing. The conversion loss of a double-balanced mixer is similar to that of a single-balanced mixer, although the dynamic range of the double-balanced mixer is much greater due to the increase in the intercept point (recall IP discussion from earlier chapters).

By incorporating FET or bipolar transistors into monolithic IC mixer topologies, it is possible to produce active mixers with conversion gain rather than conversion loss. In general, this class of mixer can be operated with lower LO drive levels than passive FET or diode mixers, although active mixers will also distort when fed with excessive LO drive levels.

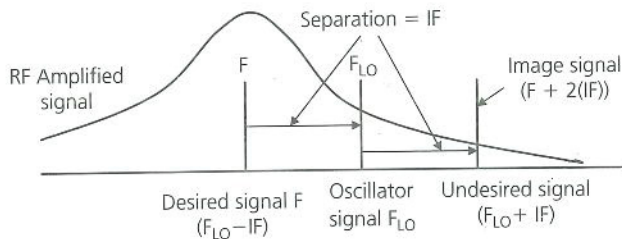


FIG. 8-11. Mixing results in unwanted frequencies, including image signals.

For RF front ends, wireless receivers, or even complete transceivers fabricated using monolithic IC semiconductor processes, the Gilbert cell mixer is a popular topology for its combination of low power consumption, high gain, and wide bandwidth. Originally designed as an analog four-quadrant multiplier for small-signal applications, the Gilbert-cell mixer can also be used in switching-mode operation for mixing purposes. Because it requires differential signals, the Gilbert-cell mixer is usually implemented with input and output transformers in the manner of double-balanced mixers.

Intermodulation and intercept points

The mixer generates IF signals that result from the sum and difference of the LO and RF signals combined in the mixer:

$$f_{IF} = f_{LO} \pm f_{RF} \quad (\text{Eq. 8-1})$$

These sum and difference signals at the IF port are of equal amplitude, but generally only the difference signal is desired for processing and demodulation so the sum frequency (also known as the image signal—see Fig. 8-11) must be removed, typically by means of IF bandpass or lowpass filtering. A secondary IF signal, which can be called f_{IF}^* , is also produced at the IF

port as a result of the sum frequency reflecting back into the mixer and combining with the second harmonic of the LO signal. Mathematically, this secondary signal appears as:

$$f_{IF}^* = \pm[2f_{LO} - (f_{LO} - f_{RF})] \quad (\text{Eq. 8-2})$$

This secondary IF signal is at the same frequency as the primary IF signal. Unfortunately, differences in phase between the two signals typically result in uneven mixer conversion-loss response. But flat IF response can be achieved by maintaining constant impedance between the IF port and following component load (IF filter and amplifier) so that the sum frequency signals are prevented from re-entering the mixer. In terms of discrete components, some manufacturers offer constant-impedance IF bandpass filters that serve to minimize the disruptive reflection of these secondary IF signals. Such filters attenuate the unwanted sum frequency signals by absorption. Essentially, the return loss of the filter determines the level of the sum frequency signal that is reflected back into the mixer.

If a mixer's IF port is terminated with a conventional IF filter, such as a bandpass or lowpass type, the sum frequency signal will re-enter the mixer and generate intermodulation distortion. One of the main intermodulation products of concern is the two-tone, third-order product, which is separated from the IF by the same frequency spacing as the RF signal. These intermodulation frequencies are a result of the mixing of spurious and harmonic responses from the LO and the input RF signals:

$$f_{LO} = \pm(2f_{RF1} - f_{RF2}) \quad (\text{Eq. 8-3})$$

$$f_{LO} = \pm(2f_{RF2} - f_{RF1}) \quad (\text{Eq. 8-4})$$

But by careful impedance matching of the IF filter to the mixer's IF port, the effects of the sum frequency products and their intermodulation distortion can be minimized.

EXAMPLE: INTERMODULATION AND INTERCEPT POINTS

To get a better understanding of intermodulation products, let's consider the simple case of two frequencies, say f_1 and f_2 . To define the products, we add the harmonic multiplying constants of the two frequencies. For example, the second order intermodulation products are $(f_1 + f_2)$; the third order are $(2f_1 - f_2)$; the fourth order are $(2f_1 + f_2)$; the fifth order are $(3f_1 - f_2)$; etc. If f_1 and f_2 are two frequencies of 100 kHz and 101 kHz—that is, 1 kHz apart—then we get the intermodulation products as shown in Table 8-1.

From the table it becomes apparent that only the odd order intermodulation products are close to the two fundamental frequencies of f_1 and f_2 . Note that one third order product $(2f_1 - f_2)$ is only 1 kHz lower in frequency than f_1 and another

Order	Intermodulation Products		Ex: $f_1 = 100$ kHz, $f_2 = 101$ kHz	
1st Order	f_1	f_2	100 kHz	101 kHz
2nd Order	$f_1 + f_2$	$f_2 - f_1$	201 kHz	1 kHz
3rd Order	$2f_1 - f_2$	$2f_2 - f_1$	99 kHz	102 kHz
	$2f_1 + f_2$	$2f_2 + f_1$	301 kHz	302 kHz
4th Order	$2f_2 + 2f_1$	$2f_2 - 2f_1$	402 kHz	2 kHz
5th Order	$3f_1 - 2f_2$	$3f_2 - 2f_1$	98 kHz	103 kHz
	$3f_1 + 2f_2$	$3f_2 + 2f_1$	502 kHz	503 kHz
Etc				

TABLE 8-1. Intermodulation Products

Continued on next page

EXAMPLE INTERMODULATION AND INTERCEPT POINTS—(Cont)

$(2f_2 - f_1)$ is only 1 kHz above f_2 . The fifth order product is also closer to the fundamentals than corresponding even order products.

These odd order intermodulation products are of interest in the first mixer state of a superheterodyne receiver. As we have seen earlier, the very function of a mixer stage—namely, forming an intermediate lower frequency from the sum/difference of the input signal and a local oscillatory—results in the production of nonlinearity. Not surprisingly, the mixer stage is a primary source of unwanted intermodulation products. Consider this example: A receiver is tuned to a signal on 1000 kHz but there are also two strong signals, f_1 on 1020 kHz and f_2 on 1040 kHz. The closest signal is only 20 kHz away. Our IF stage filter is sharp with a 2.5-kHz bandwidth, which is quite capable of rejecting the unwanted 1020-kHz signal. However, the RF stages before the mixer are not so selective and the two signals f_1 and f_2 are seen at the mixer input. As such, intermodulation components are readily produced, including a third order intermodulation component $(2f_1 - f_2)$ at $(2 \times 1020 - 1040) = 1000$ kHz. This intermodulation product lies right on our input signal frequency! Such intermodulation components or out-of-band signals can easily cause interference within the working band of the receiver.

In terms of physical measurements, the two-tone, third-order intermodulation is the easiest to measure of the intermodulation interferences in an RF system. All that is needed is to have two carriers of equal power levels that are near the same frequency. The result of this measurement is used to determine the third-order intermodulation intercept point (IIP3), a theoretical level used to calculate third-order intermodulation levels at any total power level significantly lower than the intercept point.

The IP3 is the theoretical point on the RF input vs. IF output curve where the desired input signal and third-order products become equal in amplitude as the RF input is raised. Table 8-2 explains the meaning of the three critical intercept points in RF receiver design.

Order Intercept Points	Meaning
1 dB—1dB compression point	Compression is a measure of the linearity of a device. As the RF input signal level increases the IF output should follow for a mixing device. However, when the IF output cannot follow the RF input linearly and deviates by 1dB, this is referred to as the 1dB compression point.
IP2—Second-order intercept point	The second-order intercept point is the theoretical point on the RF input vs. IF output curve where the desired input signal and second order products become equal in amplitude as the RF input is raised.
IP3—Third-order intercept point	The third-order intercept point is the theoretical point on the RF input vs. IF output curve where the desired input signal and third-order products become equal in amplitude as the RF input is raised. Note: IIP3 is the input-referred IP3. IIP3 is just the output-referred IP3 (OIP3) divided by the small-signal gain.

TABLE 8-2. Intercept Points

Preselection filters

In a typical superheterodyne receiver architecture, the LO is offset by a precise fixed amount equal to the IF. The “image” version of the generated IF signal is separated from the desired signal frequency by a difference equal to twice the IF. In order to minimize the effects of unwanted signals entering the frequency-conversion mixer, a superheterodyne receiver employs a preselection filter in the front end to remove these unwanted signal products. An input matching network provides the impedance matching required to maximize energy flow from the receive antenna to the preselector filter and to the RF amplifier in a superheterodyne receiver. One of the goals of this matching circuitry is to minimize losses from the antenna coupling

circuitry to the preselector filter, since such losses can decrease overall receiver sensitivity.

The impedance match between the receive antenna and the receiver’s front-end circuitry may be subject to change. Some receivers, for example, may be used with different antennas, in which case the match would change with each antenna. And in some mobile applications, the effects of buildings, foliage, and other environmental factors can alter the impedance of the antenna and the consequent match with the front-end components. In addition, the impedance match between the antenna and front-end components is frequency-dependent and will be a function of the particular frequency bands for the communications system. In some cases, when the antenna must

cover a broad range of frequencies, it may be necessary to employ mechanical or electrical tuning to achieve an optimum impedance match for different receiver frequencies of interest.

Filter losses can impact overall superheterodyne receiver noise figure. Because high selectivity is needed to remove unwanted spurious and image signals from the receiver signal chain, preselection filtering is often broken into several stages with intervening LNAs between stages to minimize the effects of filter losses on receiver noise figure. In essence, this balance between filter losses and LNA gain in a superheterodyne receiver front end determines the overall selectivity and sensitivity of the receiver.

In a superheterodyne receiver, any filtering pertaining to the communications channel is performed at IF. Typically, fixed rather than tunable filters are used at IF since it is much more difficult to maintain constant bandwidth in a tunable filter than in a fixed filter. When different types of signals and modulation must be handled, the IF bandwidth can be readily changed by switching between different fixed filters. In general, a superheterodyne receiver enables high selectivity and high sensitivity because of the amount of processing that is possible at IF: filtering and amplification is simply easier and cheaper to achieve at IF than at microwave frequencies.

Although a superheterodyne receiver introduces problems with spurious responses that do not exist with the other receiver architectures, it is the ease and cost-effectiveness of filtering and boosting signals at IF that makes this configuration so attractive for wireless applications. In reviewing those two key parameters—sensitivity and selectivity—that characterize the performance of a superheterodyne front end, let's examine the factors that impact each parameter and how they affect overall receiver performance.

The sensitivity of a superheterodyne receiver is relative to an acceptable level of performance, such as the sensitivity for a given SNR in an analog radio front end or for a given bit-error rate (BER) in a digital radio front end. The sensitivity refers to the weakest possible signal level that a radio front end can process to achieve that agreed-upon SNR or BER. Sensitivity is affected by the level of noise external to the receiver and generated within the receiver itself. If the external noise is low enough, then the limit to receiver sensitivity will be established by the noise generated by the receiver's own components. Noise is contributed to a receiver by means of the receiving antenna, by contributions from the RF preselector filter, and from the active devices in the receiving system, including the front-end LNA.

System sensitivity and noise

The noise from each component in the front end adds to the receiver's noise floor, which sets the limit on the minimum signal level that can be detected. Noise can be characterized by its power spectral density (PSD), which is the power contained within a

given bandwidth and is presented in units of watts per hertz. Every electronic component contributes some amount of noise to a receiving system, with the minimum amount of noise related to temperature known as the system's thermal noise, or kTB , where k is Boltzmann's constant 1.38×10^{-20} mW/K, T is the temperature in degrees Kelvin (K), and B is the noise bandwidth (in Hz).

At room temperature, the thermal noise generated in a 1-Hz bandwidth is

$$\begin{aligned} kTB &= (1.38 \times 10^{-23} \text{ J/K})(293\text{K})(1 \text{ Hz}) \\ &= 4.057 \times 10^{-21} \text{ W} = -174 \text{ dBm} \\ &\text{or } -174 \text{ dBm/Hz in terms of PSD.} \end{aligned}$$

With an increase in bandwidth comes an increase in noise power and thus the importance of filtering in a superheterodyne receiver as a means of limiting the noise power. For this reason, the final IF filter in a superheterodyne receiver is made as narrow as possible to support the channel reception and to limit the amount of noise in the channel just prior to demodulation and detection. The final IF filter determines the noise bandwidth of the receiver, since it will be the most narrowband component in the front-end analog signal chain prior to detection.

Front-end receiver components are characterized in terms of noise by several parameters, including noise figure (NF) and noise factor (F). For the receiver as a whole, the noise factor is simply a ratio of the SNR at the output of the receiver compared to the SNR at the source of the receiver. For each component, similarly, the noise factor is the ratio of the SNR at the output to the SNR at the input. The noise figure is identical to the noise factor, except that it is given in dB. The noise factor is a pure ratio:

$$\text{Noise factor} = (\text{Output SNR}_2)/(\text{Input SNR}_1) \quad (\text{Eq. 8-5})$$

while the noise figure is presented in logarithmic form as

$$\text{NF} = 10 \log(\text{SNR}_2/\text{SNR}_1) \quad (\text{Eq. 8-6})$$

where SNR_2 is the output SNR of a component, device, or receiver and SNR_1 is the input SNR of the component, device, or receiver.

If an amplifier was ideal or a component completely without noise, its noise figure would equal 0 dB. In reality, the noise figure of an amplifier or component is always positive. For a passive device, the noise figure is equal to the insertion loss of the device. For example, the noise figure of a 1-dB attenuator without losses beyond the attenuation value is 1 dB.

In a superheterodyne front end, the noise power of the components that are connected or cascaded together rises from the input to the output as the noise from succeeding stages is added to the system. In a simple calculation of how the noise contributions of front-end stages add together, there is the well-known Friis's equation:

$$\text{NF}_{\text{cascade}} = 10 \log\{[F_1 + (F_2 - 1)/A_1]\} \quad (\text{Eq. 8-7})$$

where F = the noise factor, which is equivalent to $10^{NF/10}$ and A is the numerical power gain, which is equal to $10^{G/10}$ where G is the power gain in dB. From this equation, it can be seen how the noise factor of the first stage in the system (F_1) has a dominant effect on the overall noise performance of the receiver system.

Noise factor can be used in the calculation of the overall added noise of a series of cascaded components in a receiver, using the gain and noise factor values of the different components:

$$F = F_1 + \frac{F_2 - 1}{A_1} + \frac{F_3 - 1}{A_1 A_2} + \frac{F_n - 1}{A_1 A_2 \dots A_{n-1}} \quad (\text{Eq. 8-8})$$

where the F parameters represent the noise factor values of the different front-end stages and the A parameters represent the numeric power gain levels of the different front-end stages. A quick look at this equation again shows the weight of the first noise stage on the overall noise factor. In a receiver with five noise-contributing stages ($n = 5$), for example, the noise of the final stages is greatly reduced by the combined gain of the components.

The noise floor of a receiver determines its sensitivity to low-level signals and its capability of detecting and demodulating those signals. The input referred noise level (noise at the antenna prior to the addition of noise by the other analog components in the receiver front end) is sometimes referred to as the minimum detectable signal (MDS). In some cases, a parameter known as *signal in noise and distortion* (SINAD) may also be used to characterize a receiver's noise performance, especially with a need to account for signals with noiselike distortion components. This parameter includes carrier-generated harmonics and other nonlinear distortion components in an evaluation of receiver sensitivity.

In a digital system, it is simpler to measure the bit-error rate (BER) induced by noise when a signal is weak. The BER affects the data rate so it is a more useful performance measure than the SNR for evaluating receiver sensitivity. With BER, the receiver's sensitivity can be referenced to a particular BER value. Typically a BER of 0.1%—e.g., in the GSM standard—is specified and the sensitivity of the receiver is measured by adjusting the level of the input signal until this BER is achieved at the output of the receiver.

A front end's noise floor is principally established by noise in components such as thermal noise, shot noise and flicker noise. At the same time, any decrease in gain will increase the noise floor. Thus, there must be enough margins in the system SNR to allow for a reduction in gain when making adjustments in gain for larger-level signals.

Front-End Amplifiers

The RF front-end component most commonly connected to an RF or IF filter is an RF or IF amplifier, respectively. Depending upon its function in the system, this amplifier may be designed for high output power (in the transmitter) or low-noise performance (in the receiver). At the receiver antenna, the receiver sensitivity will be a function of the ability of the preselector filter

to limit incoming wideband noise and the front-end's low-noise amplifier (LNA) to provide enough gain to boost signal levels to an acceptable signal-to-noise ratio (SNR) for subsequent signal processing in the RF front end by mixers, demodulators, and/or ADCs.

As with the filters, an RF front-end's LNAs are specified depending on their location in the signal chain, either for relatively broadband use or for channelized use at the IF stages. An LNA is specified in terms of bandwidth, noise figure, small-signal gain, power supply and power consumption, output power at 1-dB compression, and linearity requirements. The linearity is usually judged in terms of third-order and second-order intercept points to determine the expected behavior of the amplifier when subjected to relatively large-level input signals. Ideally, an LNA can provide sufficient gain to render even low-level signals usable by the RF front-end's mixers and other components, while also handling high-level signals without excessive distortion.

At one time, LNAs fabricated with gallium arsenide (GaAs) process technology provided optimum performance in terms of noise figure and gain in RF and microwave communications systems. But ever-improving performance in silicon-germanium (SiGe) heterojunction-bipolar-transistor (HBT) now provides comparable or better noise-figure and gain performance in LNAs at frequencies through about 10 GHz.

In contrast to a superheterodyne receiver's noise, the other end of the dynamic range is the largest signal that the receiver can handle without distortion or, in the case of a digital receiver, degradation of the BER. In a receiver, excessively high signal levels can bring the onset of nonlinear behavior in the receiver's components, especially the mixers and LNAs. Such nonlinear effects are evidenced as gain compression, intermodulation distortion, and cross modulation, such as AM-to-PM conversion.

At large signal levels, harmonic and intermodulation distortion cause compression and interference that limit the largest signals that a receiver can handle. A receiver's dynamic range refers to the difference between the MDS and the maximum signal level. In a single-channel system, the dynamic range is essentially the difference between the 1-dB compressed output power and the output noise floor. The spurious-free dynamic range (SFDR) is defined as the range of input power levels from which the output signal just exceeds the output noise floor, and for which any distortion components remain buried below the noise floor.

IP3

The input third-order intercept point is often used as a measure of component and receiver power-handling capability. As mentioned earlier, it is defined as the extrapolated input power level per tone that would cause the output third-order intermodulation products to equal the single-tone linear fundamental output power. The output power at that point is the output third-order intercept point. The intercept point is fictitious in that it is necessary to extrapolate the fundamental component in a linear fashion and assume that the third-order intermodulation products increase forever with a 3:1 slope. In reality, the difference

between a component's actual output power at 1-dB compression and the third-order intercept point can be as little as 6 dB and as much as 20 dB. Along with the third-order intercept point, the second-order intercept point is also used as a measure of power-handling capability of dynamic range. It refers to the fictitious intersection of the second-harmonic output power with the fundamental-frequency output power.

In analyzing a receiver's dynamic range, it is important to note how the definitions of larger signals can vary. For example, for multiple-carrier communications systems, the peak power level will be much greater than the average power level because of the random phases of the multiple carriers and how they combine in phase. In a multicarrier system, the specified average power may be within the linear region of the system but the peaks may push the system into nonlinear behavior. This nonlinear behavior includes a phenomenon known as spectral regrowth and is characterized by such parameters as adjacent-channel power ratio (ACPR) where the power of a transmitted signal can literally leak into nearby channels because of intermodulation distortion.

Automatic gain control (AGC) can be used in a superheterodyne front end to decrease the gain when strong signals can cause overload or distortion, although there may be trade-offs for the SNR performance. If attenuation is added before the LNA in a receiver front end, for example, it can reduce the risk of nonlinearities caused by large signals at the cost of an increase in noise figure, as noted earlier with the 1-dB attenuator example. An AGC tends to sacrifice small-signal performance to achieve large-signal handling capability.

Selectivity

So far, this front-end discussion has covered sensitivity and dynamic range. Another important aspect of receiver design is achieving good selectivity, so that signals of interest are processed and unwanted spurious and interference signals are properly rejected or attenuated. Selectivity can be thought of as the capability of a receiver to separate a signal or signals at a given frequency or bandwidth from all other frequencies. Selectivity is generally a function of the matching networks between components, the filters, and the amplifiers in a front-end design. Achieving good selectivity with a filter, for example, requires that the filter achieve a wide enough passband to channel the designed signals and their modulation, but provide a sharp enough response with adequate rejection to eliminate unwanted signals not falling in the desired passband. Selectivity can also be thought of as the amount of rejection needed to reduce the level of an unwanted signal to some required amount at some nominal frequency from the desired passband. Selectivity can be achieved at different stages in a superheterodyne receiver, by using selective components and devices at the RF, IF, and baseband stages.

Receiver front-end selectivity should be as high as possible close to the antenna to remove large interfering signals before they enter the active devices in later stages and cause problems

with distortion and overloading. For selective filters, impedance matching can be difficult and can adversely affect the performance of the other components in the front-end signal chain, notably the mixers.

Filters come in essentially four types: bandpass, band-reject, lowpass, and highpass filters. (For a more detailed discussion of filters, see Chapter 3.) A bandpass filter channels signals with minimal attenuation through a range of frequencies known as the passband, and rejects signals at frequencies above and below the passband. A band-reject filter (also known as a notch filter) is essentially the opposite of a bandpass filter. It rejects signals across one band (known as the stop band) and allows signals to pass with minimal attenuation at frequencies above and below the stop band. A lowpass filter channels signals with minimal attenuation below a specified cutoff frequency, while rejecting signals above that cutoff frequency. The cutoff frequency is commonly a point at which signal attenuation reaches 3 dB. A highpass filter is essentially the opposite of a lowpass filter, rejecting signals below the cutoff frequency and passing signals with minimal attenuation above the cutoff frequency.

Filters are judged in terms of a number of performance parameters, including insertion loss, return loss (or VSWR), rejection, ripple, selectivity (amplitude-versus-frequency response), group delay (how long a signal takes to propagate through a filter), phase response, and quality factor (Q). In a bandpass filter, insertion loss is the amount of signal attenuation above a 0-dB level that would be represented by an ideal transmission line in place of the filter. Insertion loss occurs due to a filter's dissipative elements (the resistors, inductors, capacitors, and transmission lines). Rejection is the amount of signal attenuation at specified points above and below the passband or center frequency, including the insertion loss.

Bandpass filters are defined in terms of their center frequency and the width of their passband. The center frequency of a bandpass filter can be defined arithmetically or geometrically. The geometric definition is usually employed in the filter design process, while the arithmetic definition is used to specify a filter. The arithmetic center frequency is simply the sum of the lower and upper bandedges divided by two. For example, for a bandpass filter with 3-dB frequencies of 900 and 1000 MHz, the arithmetic center frequency is $(900 + 1000)/2 = 950$ MHz.

The Q of a filter is the ratio of the midband frequency to the bandwidth. A narrowband filter, for example, with 3-dB band edges of 950 and 1000 MHz (center frequency of 975 MHz) has a Q of $975/50 = 19.5$. A bandpass filter with 3-dB bandwidth of 500 to 1000 MHz would have a much lower Q of $750/500 = 1.5$. Filter Q is related to the bandwidth, with narrower filter bandwidths resulting in higher filter Q values.

In fabricating filters, high- Q circuit elements (such as transmission lines, capacitors and inductors) are desirable for high-performance filter responses. Low- Q circuit elements tend to yield higher passband insertion loss and lower stopband attenuation. And lower- Q elements lead to a rounding of a filter's response, with poorly defined filter skirts.

Filter out-of-band rejection or attenuation can be increased by adding sections, although this also adds complexity and increases insertion loss due to the additional resonant elements. A wide range of filter responses, such as equiripple and linear-phase responses, are available in modern designs. The equiripple response, for example, provides minimum amplitude deviations across the passband. A linear-phase filter is suitable for preserving the content of pulsed waveforms typically used in modern communications systems, such as the Orthogonal Frequency Division Multiplex (OFDM) modulation used in WiMAX systems.

ADC'S EFFECT ON FRONT-END DESIGN

Analog-to-digital converters (ADCs) are commonly used in receivers for wireless applications for either IF or baseband signal sampling. The choice of ADC is generally determined by the rest of the receiver architecture, and can be affected by the selectivity of the filters, the dynamic range afforded by the front-end amplifiers, and the bandwidth and type of modulation to be processed. For example, the level or dynamic range of signals expected to be presented to the ADC will dictate the bit resolution needed for the converter. For example, in an example double-downconversion receiver architecture developed for broadband wireless access (BWA) applications using the IEEE 802.16 WiMAX standard, IF sampling can be performed with a 12-b ADC.

For cases where a single downconversion approach, with a subsequent higher IF, is used, a higher-resolution, 14-b converter is recommended in order to compensate for the less efficient selectivity of the single-conversion receiver and to avoid ADC saturation in the presence of high-level interference signals. Along with its input bandwidth (which should accommodate the highest IF of interest for a particular receiver design) and bit resolution, an ADC can also be specified in terms of its spurious-free dynamic range (SFDR). The ADC's sensitivity is influenced by wideband noise, including spurious noise, and often can be improved through the use of an antialiasing filter at the input of the ADC to eliminate sampling of noise and high-frequency spurious products.

To avoid aliasing when converting analog signals to the digital domain, the ADC sampling frequency must be at least twice the maximum frequency of the input analog signal. This minimum sampling condition—derived from Nyquist's theorem (covered in earlier chapters)—must be met in order to capture enough information about the input analog waveform to reconstruct it accurately.

In addition to selecting an ADC for IF or baseband sampling, the choice of buffer amplifier to feed the input of the converter can affect the performance possible with a given sampling scheme. The buffer amplifier should provide the rise/fall time and transient response to preserve the modulation information of the IF or baseband signals, while also providing the good amplitude accuracy and flatness needed to provide signal amplitudes at an optimum input level to the ADC for sampling.

Now let's consider an example using lowpass signals where the desired bandwidth goes from 0 (DC) to some maximum frequency (f_{MAX}). The Nyquist criterion states that the sampling frequency needs to be at least $2f_{MAX}$. So, if the ADC is sampling at a clock rate of 20 MHz, this would imply that the maximum frequency it can accept is 10 MHz. But then how could an FM radio broadcast signal (say, at 91.5 MHz) be converted using such a relatively low sampling rate? Here's where the design of the RF front end becomes critical. The RF receiver must support an intermediate frequency (IF) architecture, which translates a range of relatively high input frequencies to a lower-frequency range output (at the IF band). Using the example of the FM radio, with a tunable bandwidth of 88 to 108 MHz, then the receiver's front end must process signals over that tunable bandwidth to a lower IF range of no higher than 10 MHz. Such a design would ensure that the previously mentioned 20-MHz ADC could handle these IF signals without aliasing.

SOFTWARE DEFINED RADIOS

Up to this point, we have focused on the hardware implementation of RF front ends. But the capability of creating RF front-end architectures that are controlled by software has become a reality, thanks to the continuing migration of analog functionality to digital chips and the performance improvements in DSP technology. Today, a designer can use software defined radio (SDR) systems to process RF signals that were traditionally handled by analog and RF front-end circuitry. This new approach is being implemented in next generation, very high frequency (~ 6 GHz) applications that require flexible reconfiguration of the front end, like wireless base stations, mobile communication (often military) devices, and IPTV set-top boxes.

SDR technologies known as "frequency agile" systems allow the conversion of any analog wireless signal directly into digital baseband data, regardless of frequency. Semiconductor manufacturing processes have typically used gallium arsenide (GaAs) materials to demonstrate the RF-to-digital (RF/D) converter chips, though future production versions will convert the design to low-cost CMOS and BiCMOS technologies.

The great advantage of SDR systems is the on-the-fly interoperability among differing communications frequencies. In other words, a single RF receiver can communicate directly with different devices on multiple different frequency bands, using only software to receive, translate, and process the different signals.

CASE STUDY—MODERN COMMUNICATION RECEIVER

In this chapter we have introduced the design architectures common in most RF front-end receivers. We have defined a number of key parameters used to characterize the response of a receiver, including sensitivity and selectivity.

Now let's see how all of the concepts and parameters fit into the development of a typical modern communications transceiver. Such a communication front-end/back-end could be used to support a common US air interface like second generation

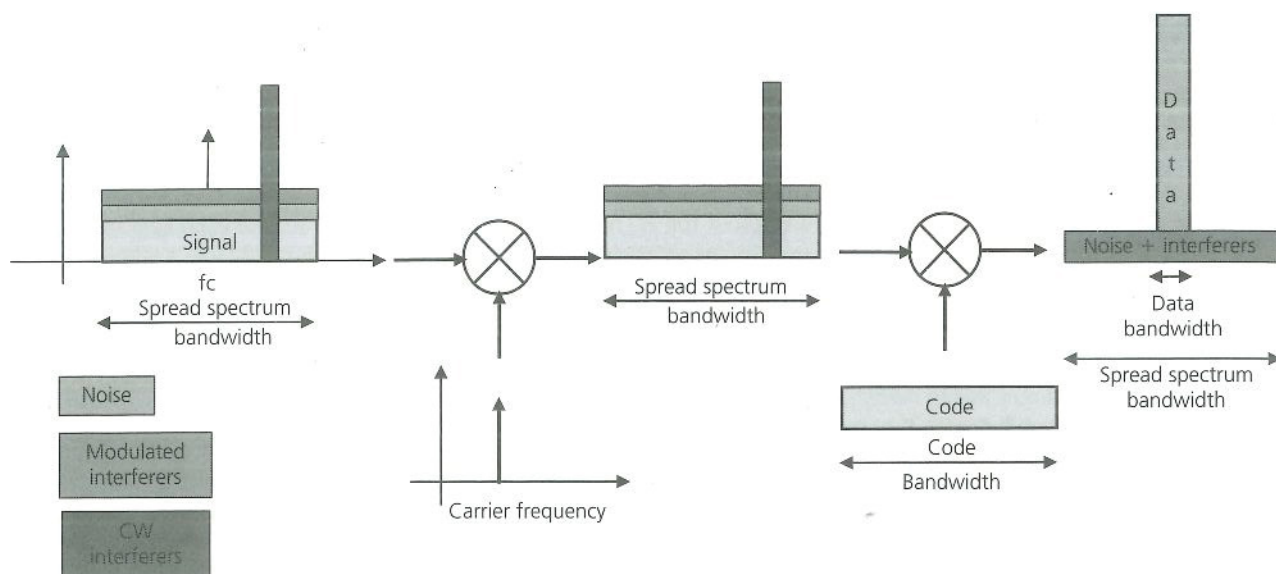


FIG. 8-12. CDMA reception process. (Courtesy of Agilent)

(2G), narrow-band Code Division Multiple Access (CDMA) or third-generation (3G), multimedia enabled wideband CDMA (W-CDMA) systems. By changing the RF tuning, this same architecture could be used for dual-band GSM (used in Europe) or TDMA systems in the same radio band, since the processing and demodulation is performed in the post-baseband, digital section.

This last point is important, since this chapter has focused on traditional analog receiver design as are used in TDMA designs. As the name implies, Time Division Multiple Access (TDMA) technology divides a radio channel into sequential time slices. Each channel user takes turns transmitting and receiving in a round-robin fashion. TDMA is a popular cellular phone technology since it provides greater channel capacity than its predecessor—Frequency Division Multiple Access (FDMA). Global System for Mobile Communications (GSM), an established cellular technology in Asia and Europe, uses a form of TDMA technology.

In this case study, though, we focus on Code Division Multiple Access (CDMA) designs for two reasons. First, the basic receiver architecture is similar to TDMA. Second, CDMA receiver designs are predominant in the US and are gaining global acceptance.

In CDMA systems, the received signal occupies a relatively narrow channel within a 60-MHz spectral allocation between 1930 MHz and 1990 MHz. W-CDMA channels operate on a wider bandwidth (3.84 MHz) than standard CDMA systems. All CDMA users can transmit at the same time while sharing the same carrier frequency. A user's signal appears to be noise for all except the correct receiver. Thus, the receiver circuit must decode one signal among many that are transmitted at the same time and at the same carrier frequency, based on correlation techniques.

The CDMA reception process is as shown in Fig. 8-12. Several mixer stages are required to separate the carrier frequency and the code bandwidth. Once complete, the desired data signal can be separated from the “noise” (other user channels) and interference.

In a modern receiver front-end communication system, the received signal is amplified, mixed down to IF, and filtered before being mixed down to baseband where it is digitized for demodulation (see Fig. 8-13). A double (multi-mixer) superheterodyne architecture is typically used in a CDMA receiver. The RF front-end consists of the typical duplexer and low-noise amplifier (LNA) to provide additional signal gain to compensate for signal losses from the subsequent image-reject filter and then the first mixer. Two downconverter stages are used between the RF and baseband subsystems. The first mixer downconverts the signal to a first IF stage of 183 MHz. The second mixer completes the downconversion from the IF stage to baseband. The *I/Q* outputs from the second mixer stage are digitally decoded and demodulated in the baseband DSP subsystem.

The receiver architecture contains an *I/Q* demodulator to separate the information contained in the *I* (in-phase) and *Q* (quadrature) signal components prior to the baseband input—recall earlier discussion on direct conversion techniques. Overall key receiver requirements (derived from the IS-95/IS-98 standards) for a CDMA system are defined by (see Fig. 8-14):

- Reference sensitivity is the minimum receiver input power, at the antenna, at which bit error rate (BER) $\leq 10^{-3}$. This results in an acceptable noise power (P_n) within the channel bandwidth of -99 dBm. The acceptable noise power (-99 dBm) within the channel bandwidth results in a receiver noise figure (NF)

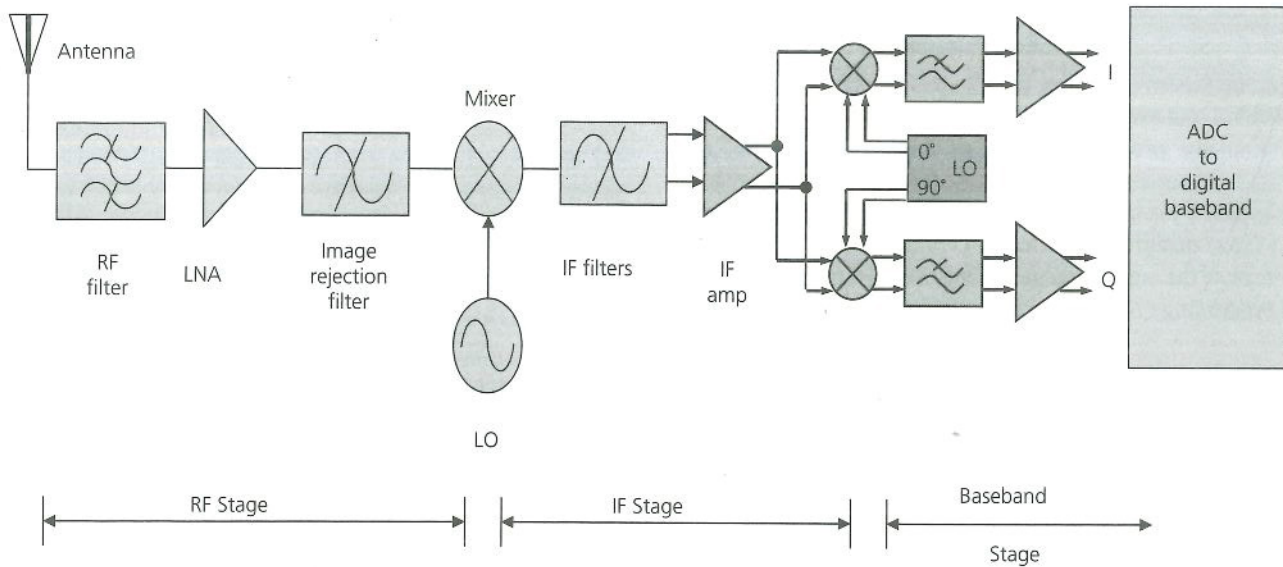


FIG. 8-13. Block diagram of a modern RF communication receiver.

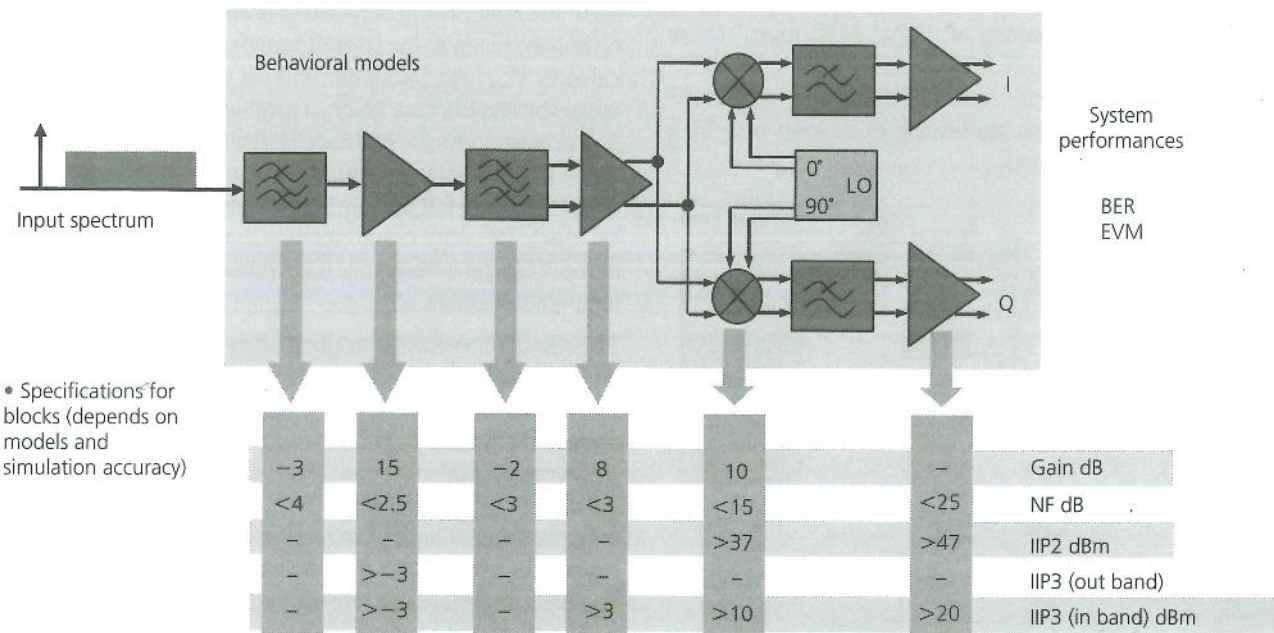


FIG. 8-14. Typical CDMA system specifications.

of 9 dB. Recall that the noise figure of a receiver is the ratio of the SNR at its input to the SNR at its output. It characterizes the degradation of the SNR by the receiver system.

- Adjacent channel selectivity (ACS) is the ratio of the receive filter attenuation on the assigned channel frequency to the receiver filter attenuation on the adjacent channel frequency.
- Intermodulation results from nonlinear modulation of two pure input signals. As mentioned in Chapter 7, when

two or more signals are input to an amplifier simultaneously, the second-, third-, and higher-order intermodulation components are caused by the sum and difference products of each of the fundamental input signals and their associated harmonics. Of particular importance to CDMA receiver design is the third-order intercept point (IP3).

Now let's consider the issue of measuring and controlling the RF signal power. On the receive side, the input signal will generally vary over some dynamic range. This may be due to weather

conditions or to the source of the received signal moving away from the receiver (e.g., a mobile handset being operated in a fast car). But as explained earlier in this chapter, we want to present a constant signal level to the analog-to-digital converter (ADC) to maintain the proper resolution of the ADC. This will also maximize the signal-to-noise ratio (SNR). As a result, receive signal systems typically use one or more variable gain amplifiers (VGAs) that are controlled by power measurement devices that complete the automatic-gain-control (AGC) loop. Recall the signal processing on the receive side occurs after the IF and ADC stages.

An inaccurate received signal strength indication (RSSI) measurement can result in a poor leveling of the signal that is presented to the ADC. This will cause either overdrive of the ADC (input signal too large) or waste valuable dynamic range (input signal too small).

IF Amplifier Design

Several amplifiers are used in the IF stage of most receivers. Consider the architecture we've been examining, noting one of these amplifiers just prior to the two-stage I/Q mixer. This amplifier can be designed as an analog or digital AGC loop. Where fast regulation of gain is required, the inherent latency of a digitally controlled automatic gain control (AGC) loop may not be acceptable. In such situations, an analog AGC loop may be a good alternative (see Fig. 8-15).

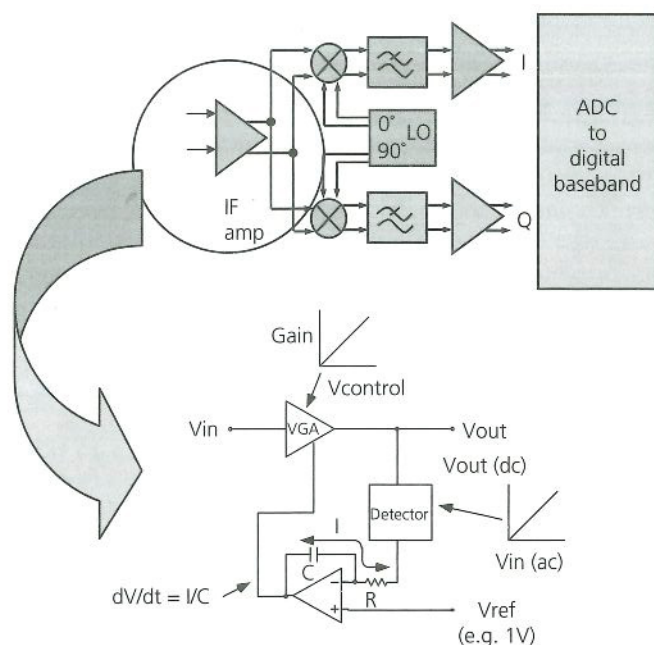


FIG. 8-15. Generic components of an AGC analog amplifier loop.

Beginning at the output of the variable gain amplifier (VGA), this signal is fed, usually via a directional coupler, to a detector. The output of the detector drives the input of an op amp, configured as an integrator. A reference voltage drives the non-inverting input of the op amp. Finally the output of the op-amp integrator drives

the gain control input of the VGA. Now, let's examine how this circuit works.

We will assume initially that the output of the VGA is at some low level and that the reference voltage on the integrator is at 1 V. The low detector output results in a voltage drop across integrator resistor R . The resulting current through this resistor can only come from the integrator capacitor C . Current flow in this direction increases the output voltage of the integrator. This voltage, which drives the VGA, increases the gain (we are assuming that the VGA's gain control input has a positive sense, that is, increasing voltage increases gain). The gain will be increased, thereby increasing the amplifier's output level until the detector output equals 1 V. At that point, the current through the resistor/capacitor will decrease to zero and the integrator output will be held steady, thereby settling the loop. If capacitor charge is lost over time, the gain will begin to decrease. However, this leakage will be quickly corrected by additional integrator current from the newly reduced detector voltage.

The key usefulness of this circuit lies in its immunity to changes in the VGA gain control function. From a static perspective at least, the relationship between gain and gain control voltage is of no consequence to the overall transfer function. Based upon the value of V_{ref} , the integrator will set the gain control voltage to whatever level is necessary to produce the desired output level. Any temperature dependency in the gain control function will be eliminated. Also, nonlinearities in the gain transfer function of the VGA do not appear in the overall transfer function (V_{out} vs. V_{ref}). The only requirement is that the gain control function of the VGA be monotonic. It is crucial however that *detector* be temperature stable.

The circuit as we have described it has been designed to produce a constant output level for varying input levels. Because this results in a constant output level, it becomes clear that the detector does not require a wide dynamic range. We only require it to be temperature stable for input levels that correspond to the setpoint voltage V_{ref} . For example, the diode detector circuits previously discussed which have poor temperature stability a low levels but reasonable stability at high levels, might be a good choice in applications where the leveled output is quite high.

If, the detector we use has a higher dynamic range, we can now use this circuit to precisely set VGA output levels over a wide dynamic range. To do this, the integrator reference voltage, V_{ref} , is varied. The voltage range on V_{ref} follows directly from the detector's transfer function. For example, if the detector delivers 0.5 V for an input level of -20 dBV, a reference voltage of 0.5 V will cause the loop to settle when the detector input is -20 dBV (the VGA output will be greater than this amount by whatever coupling factor exists between VGA and detector).

The dynamic range for the variable V_{out} case will be determined by the device in the circuit with the least dynamic range (i.e., gain control range of VGA or linear dynamic range of detector). Again it should be noted that the VGA does not need a precise gain control function. The "dynamic range" of the VGA's

gain control in this case is defined as the range over which an increasing gain control voltage results in increasing gain.

The response time of this loop can be controlled by varying the RC time constant of the integrator. Setting this at a low level will result in fast output settling but can result in ringing in the output envelope. Setting the RC time constant high will give the loop good stability but will increase settling time.

It is interesting to note that use of the term AGC (automatic gain control) to describe this circuit architecture is fundamen-

tally incorrect. The term AGC implies that the gain is being automatically set. In practice, it is the output level that is being automatically set, so the term ALC (automatic level control) would be more correct.

This case study has offered just a sample of the many issues that must be considered when design any communication receiver system. Numerous books and internet resources are available for those looking to understand more of the fascinating technology.

RF CIRCUIT DESIGN SECOND EDITION

CHRIS BOWICK WITH JOHN BLYLER AND CHERYL AJLUNI

It's Back! New chapters, examples, and insights; all infused with the timeless concepts and theories that have helped RF engineers for the past 25 years!

RF circuit design is now more important than ever as we find ourselves in an increasingly wireless world. Radio is the backbone of today's wireless industry with protocols such as Bluetooth®, Wi-Fi, WiMax, and ZigBee®. Most, if not all, mobile devices have an RF component and this book tells the reader how to design and integrate that component in a very practical fashion. This book has been updated to include today's integrated circuit (IC) and system-level design issues as well as keeping its classic "wire lead" material.

DESIGN CONCEPTS AND TOOLS INCLUDE

- **The Basics:** Wires, Resistors, Capacitors, Inductors
- **Resonant Circuits:** Resonance, Insertion Loss
- **Filter Design:** High-pass, Bandpass, Band-rejection
- **Impedance Matching:** The L Network, Smith Charts, Software Design Tools
- **Transistors:** Materials, Y Parameters, S Parameters
- **Small Signal RF Amplifiers:** Transistor Biasing, Y Parameters, S Parameters
- **RF Power Amplifiers:** Automatic Shutdown Circuitry, Broadband Transformers, Practical Winding Hints
- **RF Front-End:** Architectures, Software-Defined Radios, ADC's Effects
- **RF Design Tools:** Languages, Flow, Modeling

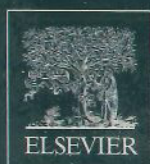
Check out this book's companion Web site at: <http://books.elsevier.com/companions/9780750685184> for full-color Smith Charts and extra content!

ALSO FROM NEWNES PRESS

Chandra, et al.: *Wireless Networking: Know It All* ISBN: 978-0-7506-8582-5

Fette, et al.: *RF & Wireless Technologies: Know It All* ISBN: 978-0-7506-8581-8

Pease: *Analog Circuits: World Class Designs* ISBN: 978-0-7506-8627-3



Newnes
An Imprint of Elsevier
www.newnespress.com



ISBN: 978-0-7506-8518-4

