Contents

Chapter 12 — Online Content

Articles
- HF Receiver Testing by Adam Farson, VA7OJ/AB7OJ
- Noise Power Ratio (NPR) Testing on HF Receivers by Adam Farson, VA7OJ/AB7OJ
- Performance Capability of Active Mixers by Dr. Ulrich Rohde, N1UL
- SDR Simplified — Filter Design Program by Ray Mack, W5IFS
- SDR Simplified — Introduction to CIC Filters by Ray Mack, W5IFS
- SDR Simplified — More Filter Activities by Ray Mack, W5IFS
- SDR Simplified — Nyquist Meets Real World by Ray Mack, W5IFS
- VHF and UHF Receivers and UHF and Microwave Techniques

Also see the Online Content for the **Transmitters** and **Transceivers** chapters.

Projects
- 10 GHz preamp PCB template by Zack Lau, W1VT
- A Dual Band Low Noise Amplifier for 2 Meters and 70 Centimeters by Jim Kocsis, WA9PYH
- A High Performance 45 MHz IF Amplifier for an Up-Conversion HF/LF Receiver by Colin Horrabin, G3SBI
- A Long-wave Upconverter by Fred Brown, W6HPH
- A Software-Based Remote Receiver Solution by Martin Ewing, AA6E
- A Software Controlled Radio Preselector by J. Onate, MØWWA and X. de Fortuny
- All Mode 1 kHz to 1.7 GHz SDR Receiver by James Forkin, WA3TFS
- Binaural I-Q Receiver project by Rick Campbell, KK7B
- General Coverage Preselector by George Hirshfield, W5OZF
- Receiver projects from previous editions of the *ARRL Handbook*
- Rock Bending Receiver PCB template by Randy Henderson, WI5W
- Simple SDR Receiver by Michael Hightower, KF6SJ
- Tunable RF Preamplifier Using Varicap Diode by George Steber, WB9LVI
- Universal MMIC Preamp by Paul Wade, W1GHZ

Chapter 12
Receiving

Receivers have traditionally been at the forefront of Amateur Radio technology and certainly are today, even as the long-reigning analog superheterodyne architecture is being overtaken by the digital software defined radio. While commercial designs and devices abound, receiving in the crowded amateur bands is still a demanding application with strong signals immediately adjacent to signals at the noise floor, a variety of natural and man-made interferences to reject, and modes ranging from manually sent Morse to the latest experimental modulations. Like the equipment in use today, expect this chapter to change from edition to edition as technology marches on.

In order to address these challenges, this chapter focuses on the functions of receiving and how they are implemented, whether by analog or digital technologies. Metrics of receiver performance are discussed, recognizing that digital receivers behave differently than their established analog counterparts, requiring new measurement definitions and techniques. Receiver home-brew projects from past editions have been collected into a set of projects in the online content accompanying this book,

This chapter's material has been adapted and updated from or provided by a number of authors. A great deal of the receiver material was originally written by Joel Hallas, W1ZR, and the sections on mixers by Dave Newkirk, W9VES, and Rick Karlquist, N6RK. Some sections on SDR functions were adapted from "SDR: Simplified" columns in *QEX* magazine by Ray Mack, W5IFS. Material on SDR receivers, architecture, and noise reduction was contributed by Steve Hicks, N5AC, and Doug Grant, K1DG. Jim Brown, K9YC provided material on active noise canceling and diversity reception. Bob Allison, WB1GCM, and Adam Farson, VA7OJ/AB4OJ, contributed material on receiver testing. Additional material was taken from Chapters 10 and 11 in *Experimental Methods in RF Design* by Hayward, Campbell, and Larkin. The reader interested in the professional perspective and depth of detail is referred to the comprehensive *Communications Receivers — Principles and Design, 4th Edition* by Rohde, Whitaker, and Zahnd.

Coverage of VHF/UHF/microwave receivers from previous editions is included in the online content. Coverage of this topic will be revised and updated in future editions of the *ARRL Handbook*.

The major subsystems of a radio receiving system are the antenna, the receiver and the information processor. The antenna's task is to provide a transition from an electromagnetic wave in space to an electrical signal that can be conducted on wires. The receiver has the job of retrieving the information content from a particular signal coming from the antenna and presenting it in a useful format to the processor for use.

The processor typically is an operator, but can also be an automated system. When you consider that most "processors" require signals in the range of volts (to drive an operator's speaker or headphones, or even the input of an A/D converter), and the particular signal of interest arrives from the antenna at a level of mere microvolts, the basic function of the receiver is to amplify the desired signal by a factor of a million. It must do this, while in the presence of signals many orders of magnitude greater and of completely different characteristics, without distortion of the desired signal or loss of the information it carries.

12.1 Characterizing Receivers

As we discuss receivers we will need to characterize their performance, and often their performance limitations, using certain key parameters. The most commonly encountered are as described in the following sections. These are often the key performance parameters, but in many cases there are others that are important to specify, as well. Examples are audio output power, power consumption, size, weight, control capabilities and so forth.

The ARRL Lab has developed an extensive set of standardized tests that it performs on transceivers in support of *QST* Product Reviews. These tests are described in the test procedure document referenced in the **Test Equipment and Measurements** chapter and in the book *Amateur Radio Transceiver Performance Testing* by ARRL Lab Staff Engineer, Bob Allison WB1GCM.

12.1.1 Receiver Sensitivity and Noise

Sensitivity is a measure of how weak a signal the receiver can extract information from. This generally is expressed at a particular signal-to-noise ratio (SNR) since noise is generally the limiting factor. A typical specification might be: "Sensitivity: 1 μV for 10 dB SNR with 3 kHz bandwidth." The bandwidth is stated because the amount (or power) of the noise, the denominator of the SNR fraction, increases directly with bandwidth. Generally the noise parameter refers to the noise generated within the receiver, often less than the noise that arrives with the signal from the antenna. (See the **RF Techniques** chapter for a more complete discussion of noise and noise sources.)

The sensitivity of a receiver is a measure of the lowest power input signal that can be received with a specified signal-to-noise ratio. In the early days of radio, this was a very important parameter and designers tried to achieve the maximum practical sensitivity. In recent years, device and design technology have improved to the point that other parameters may be of higher importance, particularly in the HF region and below. Sensitivity remains an issue for receiving systems at VHF and above, particular with respect to noise.

Noise level is as important as signal level in determining sensitivity. This section builds on the discussion of noise in the **RF Techniques** chapter. The most important noise parameters affecting receiver sensitivity are *noise bandwidth*, *noise figure*, *noise factor*, and *noise temperature*.

Since received noise power is directly proportional to receiver bandwidth, any specification of sensitivity must be made for a particular noise bandwidth. For DSP receivers with extremely steep filter skirts, receiver bandwidth is approximately the same as the filter or operating bandwidth. For other filter types, noise bandwidth is somewhat larger than the filter's 6 dB response bandwidth.

The relationship of noise bandwidth to noise power is one of the reasons that narrow bandwidth modes, such as CW, have a significant signal-to-noise advantage over modes with wider bandwidth, such as voice, assuming the receiver bandwidth is the minimum necessary to receive the signal. For example, compared to a 2400-Hz SSB filter bandwidth, a CW signal received in a 200 Hz bandwidth will have a 2400/200 = 12 = 10.8 dB advantage in received noise power. That is the same difference as an increase in transmitter from 100 to 1200 W.

Table 12.1

Typical Noise Levels (Into the Receiver) and Their Source, by Frequency

Frequency Range	*Dominant Noise Sources*	*Typical Level (μV/m)**
LF 30 to 300 kHz	atmospheric	150
MF 300 to 3000 kHz	atmospheric/man-made	70
Low HF 3 to 10 MHz	man-made/atmospheric	20
High HF 10 to 30 MHz	man-made/thermal	10
VHF 30 to 300 MHz	thermal/galactic	0.3
UHF 300 to 3000 MHz	galactic/ thermal	0.2

*The level assumes a 10 kHz bandwidth. Data from *Reference Data for Engineers*, 4th Ed, p. 273, Figure 1.

SOURCES OF NOISE

Any electrical component will generate a certain amount of noise due to random electron motion. Any gain stages after the internal noise source will amplify the noise along with the signal. Thus a receiver with no signal input source will have a certain amount of noise generated and amplified within the receiver itself.

Upon connecting an antenna to a receiver, there will be introduction of any noise external to the receiver that is on the received frequency. The usual sources and their properties are described below. **Table 12.1** presents typical levels of external noise in a 10 kHz bandwidth present in the environment from different sources.

Atmospheric noise. This is noise generated within our atmosphere due to natural phenomena. The principal cause is lightning which sends wideband signals great distances. All points on the Earth receive this noise, but it is much stronger in some regions than others depending on the amount of local lightning activity. This source is usually the strongest noise source in the LF range and may dominate well into the HF region, depending on the other noises in the region. The level of atmospheric noise tends to drop off by around 50 dB every time the frequency is increased by a factor of 10. This source usually drops in importance by the top of the HF range (30 MHz).

Man-made noise. This source acts in a similar manner to atmospheric noise, although it is more dependent on local activity rather than geography and weather. The sources tend to be sparks from rotating and other kinds of electrical machinery as well as gasoline engine ignition systems and some types of lighting. In recent years, noise from computing and network equipment, switchmode power supplies, and appliances has increased significantly in urban and suburban environments. All things being equal, this source, on average, drops off by about 20 dB every time the frequency is increased by a factor of ten. The slower decrease at higher frequencies is due to the sparks having faster rise times than lightning. The effect tends to be comparable to atmospheric noise in the broadcast band, less at lower frequencies and a bit more at HF.

Galactic Noise. This is noise generated by the radiation from heavenly bodies outside our atmosphere. Of course, while this is noise to communicators, it is the desired signal for radio-astronomers. This noise source is a major factor at VHF and UHF and is quite dependent on exactly where you point an antenna (antennas for those ranges tend to be small and are often pointable). It also happens that the Earth turns and sometimes moves an antenna into a position where it inadvertently is aimed at a noisy area of the galaxy. If the Sun, not surprisingly the strongest signal in our solar system, appears behind a communications satellite, communications is generally disrupted until the Sun is out of the antenna's receiving pattern. Galactic noise occurs on HF, as well: Noise from the planet Jupiter can be heard on the 15 meter band under quiet conditions, for example.

Thermal Noise. Unlike the previous noise sources, this one comes from our equipment. All atomic structures have electrons that move within their structures. This motion results in very small currents that generate small amounts of wideband signals. While each particle's radiation is small, the cumulative effect of all particles becomes significant as the previous sources roll off with increasing frequency. The reason that this effect is called *thermal* noise is because the electron motion increases with the particle's temperature. In fact the noise strength is directly proportional to the temperature, if measured in terms of absolute zero (0 K). For example, if we increase the temperature from 270 to 280 K, that represents an increase in noise power of 10/270 = 0.037, or about 0.16 dB. Some extremely sensitive microwave receivers use cryogenically-cooled front-end amplifiers to provide large reductions in thermal noise.

Oscillator Noise. As noted in the **Oscillators and Synthesizers** chapter, real oscillators will have noise sidebands that extend out on either side of the nominal carrier frequency at low amplitudes. Any such noise will be transferred to the received signal and through *reciprocal mixing* create noise products from signals on adjacent channels through the mixer. A good receiver will be designed with oscillators that generate noise well below the expected level of received noise.

NOISE POWER AND SENSITIVITY

There are a number of related measures that can be used to specify the amount of noise that is generated within a receiver. If that noise approaches, or is within perhaps 10 dB of the amount of external noise received, then it must be carefully considered and becomes a major design parameter. If the internally generated noise is less than perhaps 10 dB below that expected from the environment, efforts to minimize internal noise are generally not beneficial and can, in some cases, be counterproductive.

While the total noise in a receiving system

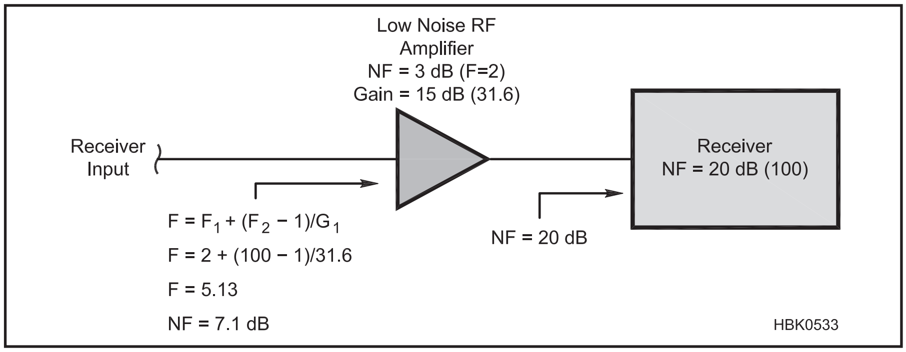


Figure 12.1 — The effect of adding a low-noise preamplifier ahead of a noisy receiver system.

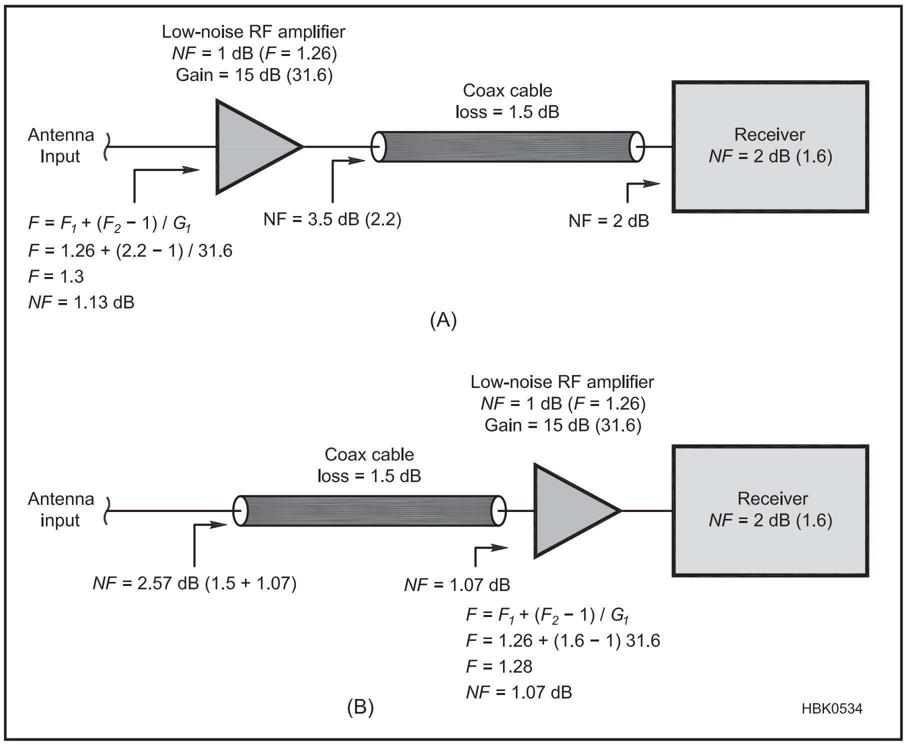


Figure 12.2 — Examples comparing the effect on input NF of placing the preamplifier at different places in the antenna system. At (A), the preamplifier is ahead of the coax cable loss, resulting in an input NF of 1.13 dB. At (B), the coax loss before the preamp results in an input NF of 2.57 dB, a reduction in sensitivity.

is, as discussed, proportional to bandwidth, the noise generating elements are generally not. Thus it is useful to be able to specify the internal noise of a system in a way that is independent of bandwidth. It is important to note that even though such a specification is useful, the actual noise is still directly proportional to bandwidth and any bandwidth beyond that needed to receive signal information will result in reduced SNR.

To evaluate the effect of noise power on sensitivity, refer to the discussion of noise in the **RF Techniques** chapter. The discussion hinges on the value of N_i, the equivalent noise power in watts at the input of a perfect receiver that would result in the same noise output. Ni is generally expressed in dBm_i:

$$dBm_i = -198.6 + (10 \times \log_{10} B) + (10 \times \log_{10} T_E)$$

where B is the system bandwidth in Hz and T_E is the equivalent noise temperature expressed in K.

If input noise (N_i) is greater than the noise generated internally by the receiver, the receiver's sensitivity is limited by the external noise. This is usually the case for HF receivers where atmospheric and man-made noise are much stronger than the receiver's internal noise floor. If N_i is within, perhaps, 10 dB greater than the receiver's internal noise, then the effect of the receiver's internal circuits on overall system sensitivity must be taken into account. At VHF and above, noise generated by the system components begins to exceed input noise.

EFFECT OF INPUT NOISE

A receiver designer needs to know how strong the signals are to establish the range of signals the receiver will be required to handle. One may compare the equivalent noise power (Ni) with the expected external noise to determine whether the overall receiver SNR will be determined by external or internal noise. A reasonable design objective is the have the internal noise be less than perhaps 10 to 20 dB below the expected noise. As noted above, this is related closely to the frequency of signals we want to receive. Any additional sensitivity will not provide a noticeable benefit to SNR, and may result in reduced dynamic range, as will be discussed in the next section.

For frequencies at which external noise sources are strongest, the noise power (and signal power) will also be a function of the antenna design. In such cases, the signal-to-noise ratio can be improved by using an antenna that picks up more signal and less noise, such as with a directional antenna that can reject noise from directions other than that of the signal. Some antennas improve SNR simply by rejecting noise, such as the Beverage antenna. Signals from a Beverage antenna are usually much weaker than from a conventional antenna but their ability to reject noise from undesired directions creates a net improvement in the SNR at the receiver output.

MINIMUM DETECTABLE (DISCERNABLE) SIGNAL (MDS)

Also referred to as the *noise floor* of a receiver, the MDS is the strength of the smallest input signal that produces a specified increase in the output noise power of a receiver. MDS depends on the required SNR (SNR_{MIN} in dB), the system bandwidth (B in Hz), the temperature of the receiver (TE), and the receiver noise figure (NF). (Measurement of MDS is discussed in the **Test Equipment and Measurements** chapter.)

An ideal receiver at room temperature with a bandwidth of 1 Hz has a theoretical MDS of –174 dBm. This is often referred to as the "1 Hz noise floor."

As bandwidth, noise figure, and temperature increase, so does MDS as follows:

$$\text{MDS (in dBm)} = -174 + 10 \log (B) + NF + SNR_{MIN} + 10 \log (T_E/290)$$

For example, if a receiver at room temperature has a bandwidth of 1 kHz and a noise figure of 2.5 dB, for an SNR_{MIN} of 3 dB:

$$MDS = -174 + 30 + 2.5 + 3 + 0 = -138.5 \text{ dBm}$$

If the receiver's equivalent temperature then

Noise Measurement Terminology

There are many ways to measure and specify noise, each with its unique name and abbreviation. Most amateurs are familiar with SNR (signal-to-noise ratio) at least in a general sense, but fewer are aware of the need to specify bandwidth and possibly temperature. Communications professionals and receiver designers use a variety of names and methods to specify noise performance. Many of these are explained in Walt Kester's excellent tutorial "Understand SINAD, ENOB, SNR, THD, THD+N, and SFDR so You Don't Get Lost in the Noise Floor" published by Analog Devices as MT-003. See the References section for the complete URL for this document.

Table 12.2
Typical Communications Bandwidths for Various Operating Modes

Mode	*Bandwidth (kHz)*
FM Voice	15
AM Broadcast	10
AM Voice	4-6.6
SSB Voice	1.8-3
Digital Voice	1.0-1.2
RTTY (170-850 Hz shift)	0.3-1.5
CW	0.1-0.5

increases to 300 K, the MDS also increases to:

MDS = –174 + 30 + 2.5 + 3 + 0.15 = –138.35 dBm

Many commercial radios achieve an MDS of –135 to –140 dBm in a 500 Hz bandwidth — which is quite good — but at HF the external noise is much higher, making the lower MDS specification somewhat irrelevant. As frequency increases through the higher HF bands and into VHF, MDS becomes more important.

12.1.2 Preamplifiers and Noise

Often we are faced with the requirement of determining the noise figure of a system of multiple stages. In general adding an amplification stage between the antenna and the rest of the system will reduce the equivalent noise figure of the system by the amount of gain of the stage but adds in the noise of the added stage directly. In general, to improve system sensitivity, the preamp's internal noise must be much lower than the noise generated by the receiver. (See the discussion of Cascaded Amplifiers in the **RF Techniques** chapter.**)** A number of preamplifier projects are included in the online content for this chapter.

In many cases elements of a receiving system exhibit loss rather than gain. Since they reduce the desired signal, while not changing the noise of following stages, they increase the noise figure at their input by an amount equal to the loss. This is the reason that a VHF *low noise amplifier* is mounted at the antenna, ahead of a noisy receiver as shown in **Figure 12.1.** As shown in the figure for fairly typical values, while the addition of a low noise amplifier does reduce the system's noise figure, as can be observed, the amplifier gain and noise figure of the rest of the receiver can make a big difference.

As an example of the difference, if the coax between the antenna and radio has a loss of 1 dB, and the preamp has a noise figure of 1 dB, the resulting noise figure with the preamp at the radio will be 2 dB, while if at the antenna, the noise added by the coax will be reduced by the gain of the preamp, resulting in a significant improvement in received SNR at VHF and above. **Figure 12.2** shows a typical example.

Along with noise figure, the preamplifier's effect on distortion and dynamic range must be considered. As discussed in later sections of this chapter, distortion products can be generated by intermodulation from in-band and out-of-band signals. These distortion products can be considered noise and can reduce receiving system sensitivity.

System *spurious-free dynamic range (SFDR)* is the difference between MDS (see previous section) and the highest level of input signal that does not create spurious signals. Although the preamp may increase the amplitude of the desired signal, it also increases the amplitude of the unwanted distortion products even more. For this reason, spurious-free dynamic range of the receiving system can be degraded by adding a preamplifier with a limited dynamic range.

Both dynamic range and noise figure must be considered when choosing a preamp. A thorough discussion of how to evaluate and balance gain and noise figure when choosing a preamp is available in Paul Shuch, N6TX's online article "Quiet! Preamplifier at Work" (**www.setileague.org/articles/ham/preamp.pdf**) listed in the References section.

12.1.3 Receiver Selectivity

Selectivity — Selectivity is just the bandwidth discussed above. This is important because to a first estimate it identifies the receiver's ability to separate stations. With a perfect "brick wall" filter in an ideal receiver, stations within the bandwidth will be heard, while those outside it won't be detected. The selectivity thus describes how closely spaced adjacent channels can be. With a perfect 3 kHz bandwidth selectivity, and signals restricted to a 3 kHz bandwidth at the transmitter, a different station can be assigned every 3 kHz across the spectrum. In a less than ideal situation, it is usually necessary to include a *guard band* between channels.

Note that the word *channel* is used here in its generic form, meaning the amount of spectrum occupied by a signal, and not defining a fixed frequency such as an AM broadcast channel. A CW channel is about 300 Hz wide, a SSB channel about 2.5-3 kHz wide, and so forth. "Adjacent channel" refers to spectrum immediately higher or lower in frequency. **Table 12.2** shows the bandwidth required for such a "channel" in various services and modes.

Selectivity under linear operating conditions is determined by the receiver's filters. In analog heterodyne designs, the filters are typically crystal filters with a fixed bandwidth. Clever frequency shifting schemes (Passband Tuning or IF Shift, for example) create a continuously adjustable filter by shifting two fixed passbands relative to each other. DSP is also used in both hybrid and direct-sampling SDR designs to provide performance unavailable to filters using discrete components. Both types of filters are discussed in the **Analog and Digital Filtering** chapter.

12.1.4 Receiver Dynamic Range

In the case of all real receivers, there is a range of signals that a receiver can respond to linearly without distortion. This is referred to as *dynamic range* and, as will be discussed in more detail, can be established based on a number of different criteria. In the most general sense, dynamic range can be defined as the ratio of the strongest to the weakest signal that a system, in this case our receiver, can respond to linearly. This range typically extends from the receiver's noise floor to a level at which some intermediate stage or stages overload in some way.

Table 12.1 gives us an idea of how small a signal we might want to receive. The designer must create a receiver that will handle signals from below the noise floor to as strong as the closest nearby transmitter can generate. Most receivers have a specified (or sometimes not) highest input power that can be tolerated, representing the other end of the spectrum. Usually the maximum power specified is the power at which the receiver will not be damaged, while a somewhat lower power level is generally the highest that the receiver can operate at without overload and the accompanying degradation of quality of reception of the desired signal.

The type and severity of the overload is often part of the specification. A straightforward example might be a 130 dB dynamic range.

The nature of the distortion will determine the observed phenomenon. If the weakest and strongest signals are both on the same channel, for example, we would not expect to be able to process the weaker of the two. However, the more interesting case would be with the strong signal in an adjacent channel. In an ideal receiver, we would never notice that the adjacent signal was there. In a real receiver with a finite dynamic range or nonideal selectivity, there will be some level of adjacent channel signal or signals that will interfere with reception of the weaker on-channel signal.

Dynamic range is also affected by preamplifiers in the receiver front-end circuitry. (Passive resistive attenuators have less effect on dynamic range as they typically remain linear over a very wide range of signals.) Turning on a preamplifier may worsen various types of dynamic range performance if the preamplifier's dynamic range is less than that of the rest of the receiver. Conversely, in some SDR receivers having the preamp on may improve performance. Dynamic range specifications and measurements must include whether a preamp is on or off.

It is important to note that the behavior of a heterodyne receiver and an SDR receiver are quite different with respect to dynamic range. The notion of an "intercept point" discussed below is based on the assumption of intermodulation products increasing in a certain way that is characteristic of analog circuitry. This assumption is not valid for direct-sampling SDR equipment. An SDR's input ADC is basically linear up to the point where the instantaneous sum of all signals present at the receiver input exceeds the full-scale range of the input ADC. (See the **DSP and SDR Fundamentals** chapter.) Beyond that point, the receiver will generate spurious products based on its software and how that specific ADC responds. Hybrid analog/digital receivers that apply DSP at IF or audio frequencies behave more like traditional analog receivers.

ON-CHANNEL DYNAMIC RANGE

The signal you wish to listen to can range from the strongest to the weakest, sometimes changing rapidly with conditions, or in a situation with multiple stations such as a net. While a slow change in signal level can be handled with manual gain controls, rapid changes require automatic systems to avoid overload and operator discomfort.

This is a problem that has been long solved with automatic gain control (AGC) systems. These systems are described in a later section of the chapter, but it is worth pointing out that the measurement and gain control points need to be applied carefully to the most appropriate portions of the receiver to maintain optimum performance. If all control is applied to early stages, the SNR for strong stations may suffer, while if applied in later stages, overload of early stages may occur in the presence of strong stations. Thus, gain control has to be designed into the receiver distributed from the input to the detector.

The next two sections illustrate a frequent limitation of receiver performance — dynamic range between the reception of a weak signal in the presence of one or more strong signals outside of the channel.

BLOCKING GAIN COMPRESSION

Because a direct-sampling SDR receiver has the same gain until the input ADC clips or overloads (see section below), these receivers do not exhibit blocking. (An exception arises if an active stage ahead of the ADC blocks prior to the onset of ADC clipping, due to incorrect gain distribution. This can be detected by testing for gain compression below the level of clipping.) Otherwise, the following section applies only to analog superheterodyne or direct-conversion receivers and receivers with a hybrid analog-DSP IF architecture.

A very strong signal outside the channel bandwidth can cause a number of problems that limit receiver performance. *Blocking gain compression* (or "blocking") occurs when strong signals overload the receiver's high gain amplifiers and reduce its ability to amplify weak signals. (Note that the term "blocking" is often used outside Amateur Radio when referring to reciprocal mixing of oscillator noise with strong local signals.)

While listening to a weak signal, all stages operate at maximum gain. If the weak signal were at a level of S0, a strong signal could be at S9 + 60 dB. Using the standard of S9 representing a 50 µV input signal, and each S unit reflecting a change of 6 dB, the receiver's front-end stages would be receiving a 0.1 µV signal and a 50,000 µV into the front end at the same time. A perfectly linear receiver would amplify each signal equally until the undesired signal is eliminated at the operating bandwidth setting stage. However, in practical receivers, after a few stages of full gain amplification, the stronger signal causes amplifier clipping, which reduces the gain available to the strong signal. This is seen as a gradual reduction in gain as the input signal amplitude increases. Gain reduction also reduces the amplitude of the weaker signal which is perceived to fade as the strong signal increases in amplitude. Eventually, the weaker signal is no longer receivable and is said to have been "blocked", thus the name for the effect.

The ratio in dB between the strongest signal that a receiver can amplify linearly, with no more than 1 dB of gain reduction, and the receiver's noise floor in a specified bandwidth is called the receiver's *blocking dynamic range* or *BDR* or the *compression-free dynamic range* or *CFDR*. In an analog superhet, BDR is established by the linear regions of the IF amplifiers and mixers. If the receiver employs DSP, the range of the analog-to-digital converter usually establishes the receiver's BDR. The spacing of the signals must also be specified so that any internal filtering is accounted for. Typical signal spacing is 2, 5, 20, and 100 kHz.

A related term is "near-far interference" which is used primarily in the commercial environment to refer to a strong signal causing a receiver to reduce its gain and along with it the strength of weak received signals.

INTERMODULATION (IMD)

Blocking dynamic range is the straightforward response of a receiver to a single strong interfering signal outside the operating passband. In amateur operation, we often have more than one interferer. While such signals contribute to the blocking gain compression in the same manner as a single signal described above, multiple signals also result in a potentially more serious problem resulting from *intermodulation (IM) products*. As noted previously, analog heterodyne and SDR receivers behave quite differently with respect to intermodulation.

If we look again at the equation for two sinusoids in the sidebar on Nonlinear Signal Combinations, we note that there are an infinite number of higher order terms. In general, the coefficients of these terms are progressively lower in amplitude, but they are still greater than zero. Of primary interest is third-order term, $K_3 \times V_{IN}^3$, when considering V_{IN} as the sum of two interfering signals (f_1 and f_2) near our desired signal (f_0) and within the first IF passband, but outside the operating bandwidth.

$$V_{OUT} = K_3 \times [A\sin(f_1)t+B\sin(f_2)t]^3$$
$$= K_3 \times \{A^3\sin^3(f_1)t + 3A^2B\,[\sin^2(f_1)t \times \sin(f_2)t] + 3AB^2\,[\sin(f_1)t \times \sin^2(f_2)t] + B^3\sin^3(f_2)t\}$$

The cubic terms (the first and last terms) result in products at three times the frequency and can be ignored in this discussion. Using trigonometric identities to reduce the remaining $\sin^2$ terms and the subsequent cos( )sin( ) products reveal individual intermodulation (IM) products, recognizing that the signals have cross-modulated each other due to the nonlinear action of the circuit. (Math handbooks such as the *CRC Standard Mathematical Tables and Formulae* have all the necessary trigonometry information.)

IM products have frequencies that are linear combinations of the input signal frequencies, written as $n(f_1) \pm m(f_2)$, where n and m are integer values. The entire group of products that result from intermodulation are broadly referred to as *intermodulation distor-*

Nonlinear Signal Combinations

Although a mixer is often thought of as nonlinear, it is neither necessary nor desirable for a mixer to be nonlinear. An ideal mixer is one that linearly multiplies the LO voltage by the signal voltage, creating two products at the sum and difference frequencies and only those two products. From the signal's perspective, it is a perfectly linear but time-varying device. Ideally a mixer should be as linear as possible.

If a signal is applied to a nonlinear device, however, the output will not be just a copy of the input, but can be described as the following infinite series of output signal products:

$$V_{OUT} = K_0 + K_1 \times V_{IN} + K_2 \times V_{IN}^2 + K_3 \times V_{IN}^3 + \ldots + K_N \times V_{IN}^N$$

What happens if the input V_{IN} consists of two sinusoids at F_1 and F_2, or A × [sin ($2\pi F_1$) × t] and B × [sin ($2\pi F_2$) × t]? Begin by simplifying the notation to use angular frequency in radians/second ($2\pi F = \omega$). Thus V_{IN} becomes $A\sin\omega_1 t$ and $B\sin\omega_2 t$ and equation A becomes:

$$V_{OUT} = K_0 + K_1 \times (A\sin\omega_1 t + B\sin\omega_2 t) + K_2 \times (A\sin\omega_1 t + B\sin\omega_2 t)^2 + K_3 \times (A\sin\omega_1 t + B\sin\omega_2 t)^3 + \ldots + K_N \times (A\sin\omega_1 t + B\sin\omega_2 t)^N$$

The zero-order term, K_0, represents a dc component and the first-order term, $K_1 \times (A\sin\omega_1 t + B\sin\omega_2 t)$, is just a constant times the input signals. The second-order term is the most interesting for our purposes. Performing the squaring operation, we end up with:

$$\text{Second order term} = [K_2A^2\sin^2\omega_1 t + 2K_2AB(\sin\omega_1 t \times \sin\omega_2 t) + K_2B^2\sin^2\omega_2 t]$$

Using the trigonometric identity:

sin á sin â = 1/2 {cos (á – â) – cos (á + â)}

the product term becomes:

$$K_2AB \times [\cos(\omega_1 - \omega_2)t - \cos(\omega_1 + \omega_2)t]$$

These products are at the sum and difference frequency of the input signals! The signals, originally sinusoids are now cosinusoids, signifying a phase shift. These signals, however, are just two of the many products created by the nonlinear action of the circuit, represented by the higher-order terms in the original series.

In the output of a mixer or amplifier, those unwanted signals create noise and interference and must be minimized or filtered out. This nonlinear process is responsible for the distortion and intermodulation products generated by amplifiers operated nonlinearly in receivers and transmitters.

tion or IMD. The ratio in dB between the amplitude of the interfering signals, f_1 and f_2, and the resulting IM products is called the *intermodulation ratio*.

If *all* of the higher-order terms in the original equation are considered, n and m can take on any integer value. If the sum of n and m is odd, (2 and 1, or 3 and 2, or 3 and 4, etc.) the result is products that have frequencies near our desired signal, for example, $2(f_1) - 1(f_2)$. Those are called *odd-order* products. Odd-order products have frequencies close enough to those of the original signals that they can cause interference to the desired signal. If the sum of n and m is three, those are *third-order IM products* or *third-order IMD*. For fifth-order IMD, the sum of n and m is five, and so forth. The higher the order of the IM products, the smaller their amplitude, so our main concern is with third-order IMD.

If the two interfering signals have frequencies of $f_0 + \Delta$, and $f_0 + 2\Delta$, where Δ is some offset frequency, we have for the third-order term:

$$V_{OUT} = K_3 \times [A \sin(f_0 + \Delta)t + B \sin(f_0 + 2\Delta)t]^3$$

A good example would be interfering signals with offsets of 2 kHz and 2 × 2 kHz or 4 kHz from the desired frequency, a common situation on the amateur bands.

Discarding the cubic terms and applying the necessary trigonometric identities shows that a product can be produced from this combination of interfering frequencies that has a frequency of exactly f_0 — the same frequency as the desired signal! (The higher-order terms of the sidebar's equation for sinusoid signals can also produce products at f_0, but their amplitude is usually well below those of the third-order products.)

Thus we have two interfering signals that are not within our operating bandwidth so we don't hear either by themselves. Yet they combine in a nonlinear circuit and produce a signal exactly on top of our desired signal. If the interfering signals are within the passband of our first IF and are strong enough the IM product will be heard.

As the strength of the interfering signals increases, so does that of the resulting intermodulation products. For every dB of increase in the interfering signals, the third-order IM products increase by approximately 3 dB. Fifth-order IM increases by 5 dB for every dB increase in the interfering signals, and so forth. Our primary concern, however, is with the third-order products because they are the strongest and cause the most interference.

IMD DYNAMIC RANGE AND INTERCEPT POINT

As with blocking, discussed in a previous section, a direct-sampling SDR receiver does not respond to signals in the same way as an all-analog or hybrid analog-DSP IF receiver. For that reason, 3IMD and Intercept Point measurements only apply to analog and hybrid analog-digital receivers. IMD can be created by analog stages between the RF input and ADC input, such as preamps, preselectors, and any analog stages prior to digitization in the ADC device. However, these sources of IMD are typically much less strong than in legacy receivers. (Passive LC filters used as bandpass filters or preselectors can exhibit passive IMD (PIM) caused by nonlinearities in the components. This can be measured by means of a two-port NPR test.)

Third-order IMD dynamic range (*3IMD_DR*) is the difference between a receiver's MDS and the input level of two interfering signals that create an IMD product on the same frequency and as strong as a desired signal. (The complete test is described in the **Test Measurements and Instruments** chapter and in the book by Allison.) The test is performed with the interfering signals at different spacings (usually 2, 5, and 20 kHz) and with the desired signal at several different levels. The ARRL uses a signal at the MDS, at S5 (–97 dBm) and 0 dBm. For *QST* Product Reports, the test is performed on the 3.5, 14, and 50 MHz bands. All of these conditions must be specified and no single number characterizes 3IMD_DR performance entirely.

This dynamic range is particularly important to the contest and DX community since they often need to copy very weak signals with very strong signals on an adjacent channel or just a few kHz away. Combined with reciprocal mixing, band noise, and spurious emissions from transmitters, third-order IMD products can make for very difficult reception. Note that SDR receivers do not specify this dynamic range because their circuitry does not behave the same as analog superheterodyne receivers as discussed the sections below.

Intercept point describes the IMD performance of an individual stage or a complete receiver. For example, in an analog heterodyne receiver, third-order IM products increase at the rate of 3 dB for every 1-dB increase in the level of each of the interfering input signals (ideally, but not always exactly true). As the input levels increase, the distortion products seen at the output on a spectrum analyzer could catch up to, and equal, the level of the two desired signals if the receiver did not begin to exhibit blocking as discussed earlier. Remember that SDR equipment will behave differently.

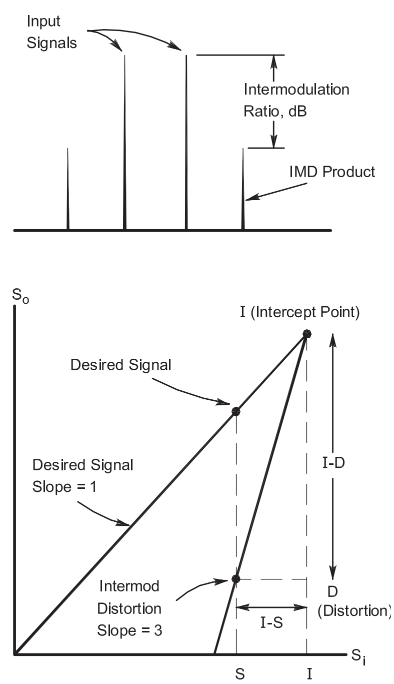
The input level at which this occurs is the *input intercept point*. **Figure 12.3** shows the concept graphically, and also derives from the geometry an equation that relates signal level, distortion and intercept point. The intercept point of the most interest in receiver evaluation is that for third-order IM products and is called the *third-order intercept point* or IP_3. A similar process is used to get a second-order intercept point for second-order IMD. A higher IP_3 means that third-order IM products will be weaker for specific input signal strengths and the operator will experience less interference from IM products from strong adjacent signals.

These formulas are very useful in designing radio systems and circuits. If the input intercept point (dBm) and the gain of the stage (dB) are added the result is an output intercept point (dBm). Receivers are specified by input intercept point, referring distortion back to the receive antenna input. Intercept point is a major performance limitation of receivers used in high density contest or DX operations. Keep in mind that we have been discussing this as an effect of two signals, one that is Δ away from our operating frequency and another at twice Δ. In real life, we may be trying to copy a weak signal at f_0, and have other signals at $f_0 \pm 500$, 750, 1000, 1250…5000 Hz. There will be many combinations that produce products at or near our weak signal's frequency.

Note that the products don't need to end up exactly on top of the desired signal to cause a problem; they just need to be within the operating bandwidth. So far we have been talking about steady carriers, such as would be encountered during CW operation with interference from nearby CW stations. SSB or other wider bandwidth modes with spectrum distributed across a few kHz will have signal components that go in and out of a relationship that results in on-channel interference from IMD. This manifests itself as a time-varying synthetic noise floor, composed of all the resulting products across the channel. The difference in this low level "noise" can be dramatic between different receivers, especially when added to phase noise received from other stations and reciprocal mixing inside the receiver!

SPURIOUS-FREE DYNAMIC RANGE (SFDR)

IM products increase with the amplitude of the interfering signals that cause them and at some point become detectable above the receiver's noise floor. The ratio of the strength of the interfering signals to the noise floor, in dB, is the receiver's *spurious-free dynamic range* or *SFDR*. This is the range of signal strengths over which the receiver does not produce any detectable spurious products. SFDR can be specified for a specific order of IM products; for example, SFDR3 is the SFDR measured for third-order IM products only. The bandwidth for which the receiver's noise floor is measured must also be specified, since smaller bandwidths will result in a lower noise floor.



From the geometry of the figure

$$\frac{I-D}{I-S} = 3$$

Solve for I to get
I = S + 0.5 (S/D)
where S/D is signal to distortion ratio in dB,
S is in dBm and I is in dBm.

For second-order intermodulation,
I = S + (S/D)

HBK05_14-006

Figure 12.3 — Graphical representation of the third-order intercept concept.

INTERFERENCE FREE SIGNAL STRENGTH (IFSS)

To address the behavior of SDR receivers for which the IP_3 specification is irrelevant new performance metrics have been proposed. One such measurement is the *interference-free signal strength* or *IFSS*. This measures the largest input signal for which no interference products are produced above the receiver noise floor. The noise floor is specified using the Rec. ITU P.372.7 band-noise levels. IFSS is similar to SFDR but does not specify a type of intermodulation products.

As a consequence of using the ITU band-noise levels noise levels (see the **RF Techniques** chapter), the IFSS for a given receiver can be much higher on the lower HF bands, where external band noise is high and will mask any interfering products. On the upper HF and VHF bands where band noise is lower, the IFSS can also be lower. Similarly, there can be a difference between the rural, suburban, and urban environments. As such, IFSS measurements ignore distortion products below the noise floor.

IFSS measurements on current receivers show a great deal of variation in the performance over a wide range of signal levels. The best performers show a smooth increase in the distortion product level beginning at the input signal level where the interference products are equal to the noise floor. These radios exhibit predictable performance over a wide range of inputs. Other models show large swings in distortion product levels, even decreasing with increased input over some ranges of signal levels. This indicates a need for better understanding of SDR behavior and design. Nevertheless, IFSS is a useful tool in assessing SDR performance.

RECIPROCAL MIXING DYNAMIC RANGE (RMDR)

All oscillators produce some noise. When the oscillator is used to control a receiver or transmitter, the noise is transferred to signals, causing extra noise in the receiver passband and in the transmitter output signal. In an analog heterodyne receiver, the noise is transferred to the signals in the mixing process from the local oscillator(s). While all of the LOs contribute some noise, in practice the first mixer is the dominant source. In an SDR receiver, noise is transferred primarily from the sampling clock for the input ADC and the DDS signal sources in the receiver. Some noise is contributed by jitter in the digitizing ADC, as well.

RMDR is measured by increasing a test input signal until the noise in the receive channel increases by a specified amount, usually 3 dB. The difference between the input signal level and the receiver's noise floor (MDS) is the RMDR.

How Phase Noise Affects a Receiver

By Steve Hicks, N5AC

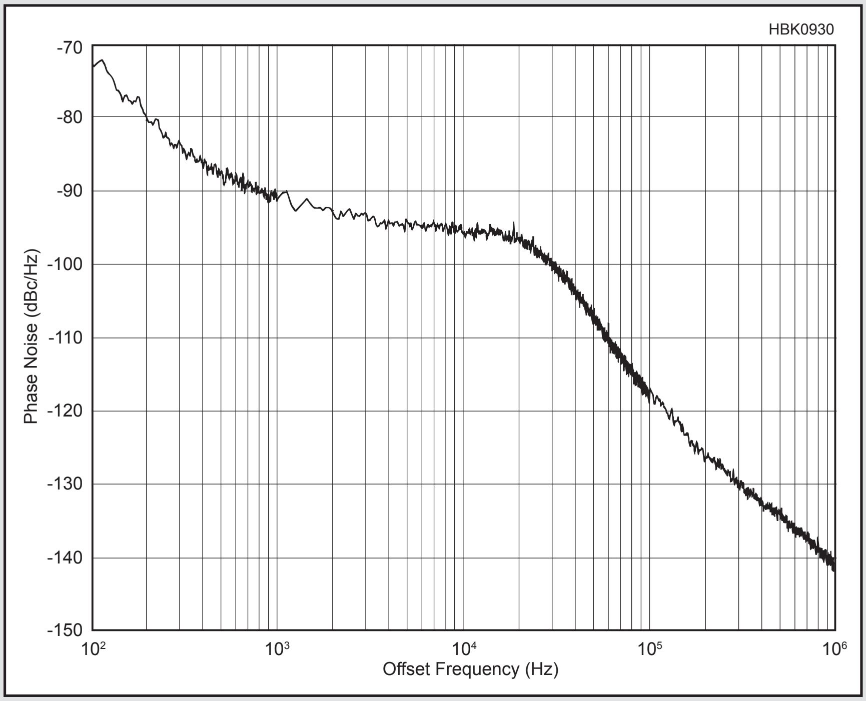


Figure 12.A2 — Plot of phase noise versus frequency separation from the carrier.

Today's receivers are generally built from one of two primary technologies: mixing or sampling. Some receivers will have components of both. In either case, an oscillator sets the frequency of the mixing or the sampling rate. Let's take a look at how phase noise from an oscillator can affect performance of our receiver, looking first at mixing. *(The effects of reciprocal mixing are produced by both amplitude and phase noise, a combination referred to as "composite noise."– Ed)*

Figure 12.A1 has a block diagram of a typical mixer circuit. The oscillator used with a mixer is called a local oscillator (LO) and the mixing process is also known as heterodyning. Let's assume for the moment that our radio designer selected an off-the-shelf synthesizer IC to use as a local oscillator. The phase noise chart in **Figure 12.A2** represents the phase noise provided by the manufacturer of the synthesizer integrated circuit. Assuming that the phase noise in our design matches the manufacturer's data, we can use the phase noise plot to determine how phase noise will affect our receiver. Phase noise will generally change as the oscillator frequency is moved and, in the case of a synthesizer can also change with the parameters used to lock to a reference signal (An overview of PLL operation can be found in a number of good references — for example: **www.ti.com/lit/an/swra029/swra029.pdf** and in **www.analog.com/media/en/training-seminars/tutorials/MT-086.pdf**). For now, let's assume this plot was taken at 14 MHz and this is where our receiver will operate.

The two numbers we will use in our example are at an offset of 2 kHz and 10 kHz. Reading the plot, we see that the 2 kHz phase noise is –93 dBc/Hz and the 10 kHz phase noise is –96 dBc/Hz. The measurement units of dBc/Hz indicates the noise that would be present in a 1 Hz bandwidth receiver, the number of dB below the carrier. For example, –93 dBc/Hz indicates that the noise floor would be 93 dB below a carrier signal measured using a 1 Hz bandwidth receiver. To convert this to a 100 Hz bandwidth we might use in a CW filter we must add 10 log (100 Hz/1 Hz) or 20 dB to our phase noise numbers. So at a 2 kHz offset from our carrier, we have –93 dBc + 20 dB = –73 dBc and at 10 kHz, –96 dBc + 20 dB = –76 dBc. In a mixer, this noise will be added to all signals present in the receiver.

Looking at **Figure 12.A3**, we show a strong S8 CW signal at 14.000 MHz. S8 corresponds to –79 dBm as shown in the graph. Two kilohertz away is the signal we are trying to receive, an S3 (–109 dBm) CW signal using a 100 Hz CW filter. Using our phase noise calculation, the S8 signal at –79 dBm would have phase noise at a 2 kHz offset of –79 dBm – 73 dB = –152 dBm. This is well below our atmospheric noise floor and will not interfere with our ability to listen to the signal at 14.002 MHz.

Figure 12.A4 shows a signal in the same location that is S9+40 dB or –33 dBm. Using the same math, at 14.002 MHz we will experience phase noise of –33 dBm – 73 dB = –106 dBm. Since the signal we are trying to receive is S3 (–109 dBm), our noise floor in 100 Hz is 3 dB above this level, causing us interference while trying to copy the signal. The amount of actual interference received also depends on the required signal-to-noise ratio required to successfully copy a signal.

Although it varies by type of oscillator, phase noise generally decreases the further we are from the carrier, meaning that weak signals that are closer to strong signals are more likely to be interfered with than weak signals farther away from a strong carrier. Stations operating in a rural area with modest antennas and no strong neighbors are much less likely to suffer from the effects of phase noise than are multi-multi contest stations with large antennas and local transmitters in the same band. Field Day is another situation where many are familiar with the effects of phase noise. If you've ever operated the sideband station at your local Field Day and heard the noise floor modulated to the sound of the CW station next to you (or vice versa), you've experienced the effects of interference from phase noise.

The effects of phase noise will also vary by mode. A CW signal may be copied with a SNR in the –10 to –20 dB range while a sideband signal may require an SNR of >+10 dB. Since sideband signals require a larger filter, we must also run our calculations with this larger filter, recog-

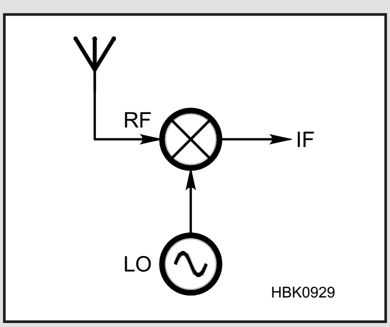


Figure 12.A1 — Typical mixer block diagram.

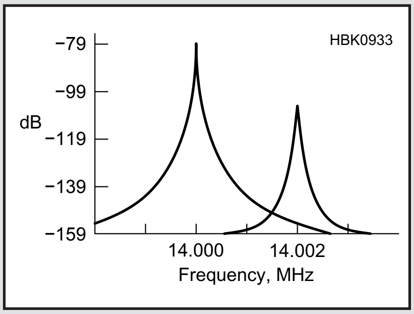


Figure 12.A3 — Spectrum showing adjacent CW signals and the phase noise from each.

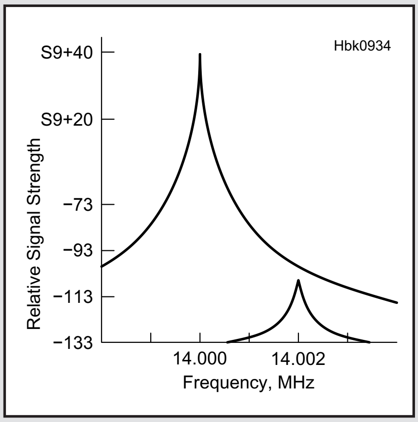


Figure 12.A4 — Same spectrum as Figure 12.A3 but with the left-hand signal replaced with one that is much stronger.

nizing that more noise will be present. Both the additional noise and the larger SNR required to copy a sideband signal make it more susceptible to phase noise interference than a CW transmission.

Division is Good

In most local oscillators for superheterodyne receivers, the phase noise of the oscillator will benefit from division. This means that the more the oscillator is reduced in frequency through division, the better the resulting phase noise of the oscillator. Typically, we will see a 6 dB improvement in phase noise with each division by two of the original oscillator frequency. So a radio that might have good oscillator performance on 20 meters could lose 11 dB or more performance when operating on 6 meters. The superheterodyne receiver designer can take advantage of this by using a good oscillator design and dividing down the oscillator as much as possible, thereby achieving better phase noise.

The raised background noise will mask weak signals, in effect raising the receiver's noise floor when a nearby strong signal is present to transfer the oscillator noise. For example, if a receiver's noise floor or MDS is –127 dBm and an input signal of –37 dBm causes the 3 dB increase in noise, the receiver's RMDR = –37 – (–127) = 90 dB.

Since oscillator noise is lower farther from the oscillator frequency, the transferred noise will have a similar profile. (The discussion of noise in the **Oscillators and Synthesizers** chapter shows the noise profile of typical oscillators.) This requires RMDR to be measured with the strong signal a specified distance from the receiver channel in which noise is being measured. Typical separations are 2, 5, and 20 kHz.

RMDR tends to be the most limiting dynamic range at the close signal spacing of 2 and 5 kHz. Since close-spaced signals are typical of the most demanding receiver environments for contesting and DXing, this is an important measurement to compare when purchasing a transceiver.

NOISE POWER RATIO (NPR)

Another test method applicable to both SDR and heterodyne receivers, *noise power ratio* tests were originally developed for multiplexed telephone systems. These early system combined many individual channels on one transmission system, cable or radio. Nonlinearities in the system would generate distortion products and interfere with the channels. To simulate the effect of many active channels, wideband noise was transmitted with one channel removed by a narrow notch filter. (See **Figure 12.4**).

The noise simulated the presence of complex speech waveforms in the other channels. Any distortion products from the noise would appear as a raised noise floor in the notched channel. As the noise amplitude was raised, the in-channel noise would rise, as well.

This is a close analogue to a busy band with multiple channels of communication. Nonlinearities in the receiver generate distortion products in the monitored channel. The ratio of the out-of-channel to in-channel noise is the noise power ratio. The more linear the receiver, the more noise power can be injected for a given amount of increase in the in-channel noise. (NPR is explained in the March/April 2015 *QEX* article, "Noise Power Ratio (NPR) Testing on HF Receivers" by Adam Farson, VA7OJ, and in the Analog Devices tutorial MT-005, "Noise Power Ratio (NPR)" — see the References section.

ADC OVERLOAD

When overloaded, A/D converters behave similarly to amplifiers driven into saturation or clipping. In this condition, the ADC cannot distinguish between an input signal exactly at the limits of its range and a signal that exceeds the input range. In fact, the ADC cannot tell if the input range has been slightly exceeded or grossly exceeded. This is an example of hard clipping.

Most ADCs used in receivers have an output signal to indicate the input range has been exceeded and the converter is in overload or overflow/underflow mode. This normally activates an overflow indicator (typ. OVF) and can result in the receiver reducing the gain of an amplifier or increasing the attenuation of any stages ahead of the ADC. Alternately,

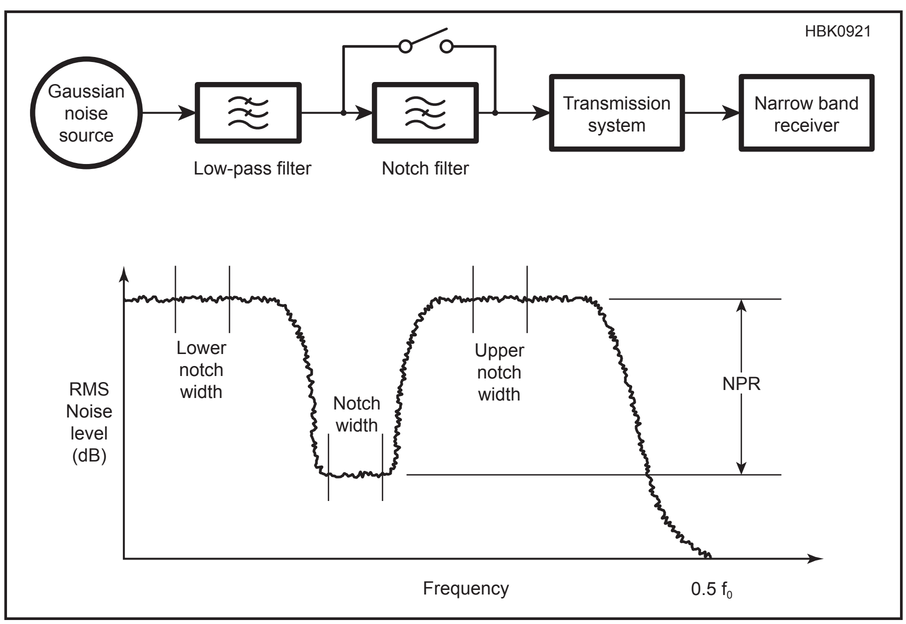


Figure 12.4 — Noise power measurement test set and typical spectrum (after Analog Devices MT-005, see References).

the operator can reduce gain or turn on an attenuator to keep the input signals within range of the ADC.

Overload generally occurs only when a strong out-of-channel signal is present. In-channel signals stimulate AGC action which attenuates the signal at the ADC input. The usual test method is to offset the receiver from the test signal before increasing test signal level until ADC overflow occurs.

LATENCY

With respect to a single receiver, latency is the time interval between arrival of an event at the receiver's RF input (usually the antenna input) and the corresponding output event at the receiver's demodulated output. The receiver output can be either an audio signal or some other type of baseband information, such as I/Q signals. The usual RF test signal is a pulse-modulated or repetitive burst signal that occurs at a well-defined time. A dual-channel oscilloscope is used to measure the time interval between the input and output events.

Latency changes with various detection and filter bandwidth settings. If the receiver includes a waterfall or spectrum display, latency can include the time to display a change after the input changes. A single receiver may show different latencies for audio, data, and visual outputs. This definition applies to only one piece of equipment and does not include network delays for receivers that are operated over a remote-control link.

Latency in a receiver is caused by the ADC that digitizes the incoming signal, either at the RF input or in the IF. It is also caused by the DAC that converts the received signal back to analog form. This converter latency is generally consistent and doesn't depend on band/mode or filtering. Radio group delay is more variable, created by the various DSP processes in the receiver, such as demodulation, frequency shifting, and filtering. Group delay depends on mode and filter settings.

Whether latency is considered excessive depends on the mode being used. For example, some digital modes that use ARQ protocols (see the **Digital Protocols and Modes** chapter) have a maximum response time before a data frame is assumed to have been lost. The resulting re-transmissions can slow the system or cause the link to fail completely. Excessive latency when monitoring one's own speech signal can become quite confusing, as well. Hybrid receivers with analog front ends and DSP IF and audio filtering can also exhibit excessive latency if the DSP is slow or underpowered. Latency in analog receivers is mostly insignificant.

TYPICAL PERFORMANCE

Product reviews of receivers and the receiver sections of transceivers published in *QST* now provide the measured dynamic range in the presence of interfering signals with spacings of 2, 5, and 20 kHz. (Details of the test procedures used are given in the **Test Equipment and Measurements** chapter and in the book by Allison.) At 20 kHz spacing, the interfering signal is usually outside of the roofing filter bandwidth of any of the above architectures. Spacing of 2 and 5 kHz represents likely conditions on a crowded band.

A caveat about transceiver testing, particularly receiver performance, is warranted. While the dynamic range of a receiver is certainly important, most modern receivers perform quite adequately. By managing receiver gain (RF gain, attenuation, preamp, or RF amplifier on or off) even a modest receiver will perform well, even under crowded conditions. Evaluation of a receiver should include a number of parameters that will affect performance according to the operator's personal interests. THD (distortion), audio hiss, filter adjustment and selection, and noise performance are just a few of the features that matter to an operator. While dynamic range values get a lot of attention, don't buy (or replace) a receiver based on one or two optimized figures. Try to look at the whole picture and consider how you are likely to use the receiver first.

A look at recent top-performing receiver measurements by the ARRL Lab and Sherwood Engineering (**www.sherweng.com/table.html**) indicates that receivers have IMD (or distortion-product) dynamic range with 2 kHz spacing results in the following ranges:

- Direct-sampling SDR: 80 to 110 dB
- Up-converting with VHF IF: 80 to 105 dB
- Down-converting with HF IF: 75 to 110 dB
- Hybrid superhet/DSP architecture: Omni VII, 82 dB (2007); Icom IC-756PRO3, 75 dB (2009)

(Note that SDR distortion products are not necessarily IMD. Receiver block diagrams are provided in the following sections. Data is taken from ARRL Lab tests and the long-term test program conducted by Rob Sherwood, NCØB, at **www.sherweng.com/table.html**.)

As we would expect, the blocking and IMD dynamic range (IMD DR) performance of a heterodyne receiver will depend on a combination of the early-stage filtering, the linearity of the mixers and amplifiers, and the dynamic range of any ADC used for DSP at the IF or audio stages.

Let's take an example of what this would mean. If we are listening to a signal at S3, for signals to generate a third-order IMD product at the same level in a receiver with a dynamic range of 60 dB, the $f_0 + \Delta$ and $f_0 + 2\Delta$ signals would have to have a combined power equal to S9 +27 dB, or each at S9 +24 dB. This is not unusual on today's amateur bands. On the other hand, if we had an IMD dynamic range of 102 dB, the interfering signals would have to be S9 +66 dB, much less likely. How much dynamic range you need depends in large measure on the kind of operating you do, how much gain your receiving antennas have and the closeness of the nearest station that operates on the same bands as you. Given that transmitters generate both phase noise and IMD products of their own, the dynamic range of most receivers on the market today is quite adequate for even the most demanding situations. Receivers may have other performance issues to consider such as filter design, AGC reaction, audio distortion, and so forth.

Keep in mind also that it is often difficult to tell whether or not you are hearing internal receiver-generated IMD — it just sounds like there are many more signals than are really present. A good test to assess the source of interference is first switch off any preamplifiers and noise-blankers or noise-reduction systems that affect the receiver's linearity. Observe the level of the interference (if it's still there) and then switch in some attenuation at the front-end of the receiver. If the level of the interference goes down by *more* than the level of attenuation (estimate 6 dB per S unit), then the interference is being generated (or at least aggravated) by non-linearity inside the receiver. Continue to increase attenuation until the interference either goes away or goes down at the same rate as the attenuation is increased. You might be surprised at how much better the band "sounds" when your receiver is operating in its linear region!

12.2 Heterodyne Receivers

The *heterodyne* receiver combines the input signal with a signal from a *local oscillator* (LO) in a mixer as discussed previously to generate the sum and difference frequencies as shown in **Figure 12.5**. The receiver may be designed so the output signal is anything from dc (a so-called *direct conversion* receiver) to any frequency above or below either of the two frequencies. The major benefit is that most of the gain, bandwidth setting and processing are performed at a single frequency, simplifying the design dramatically.

By changing the frequency of the LO, the operator shifts a signal at the input frequency the output, along with all its modulated information. In most receivers the mixer output frequency is designed to be an RF signal, either the sum or difference — the other being filtered out at this point. This output frequency is called an *intermediate frequency* or *IF*. The IF amplifier system can be designed to provide the selectivity and other desired characteristics centered at a single fixed frequency.

When more than one frequency conversion process is used, the receiver becomes a *superheterodyne* or *superhet*. A block diagram of a typical superhet receiver is shown in **Figure 12.6**. In traditional form, the RF filter is used to limit the input frequency range to those frequencies that include only the desired sum or difference but not the other — the so-called *image* frequency. The dotted line represents the fact that in receivers with a wide tuning range, such as a simple AM broadcast receiver that tunes from 500 to 1700 kHz, a more than 3:1 range, the input RF amplifier and filter is often tracked along with the local oscillator. The IF filter is used to establish selectivity — the operating bandwidth required by the information. Circuits from the antenna input through and including the mixer (the first mixer if more than one mixing stage is used) are generally referred to as the receiver's *front-end*.

For reception of suppressed carrier single-sideband voice (SSB) or on-off or frequency-shift keyed (FSK) signals, a second *beat frequency oscillator* or *BFO* is employed to provide an audible voice, an audio tone or tones at the output for operator or FSK processing. This is the same as a heterodyne mixer with an output centered at dc, although the IF filter is usually designed to remove one of the output products.

Recent superhet receivers convert the incoming signal to digital form at one of the intermediate frequencies. DSP techniques are then used to control operating bandwidth and demodulate the input signal. See the **DSP and SDR Fundamentals** chapter for more information on these techniques.

12.2.1 The Direct Conversion Receiver

The heterodyne process can occur at a number of different points in the receiver. The simplest form of heterodyne receiver is called a *direct conversion* or *DC* receiver because it performs the translation directly from the signal frequency to the audio output. It is, in effect, just the BFO and detector of the general superhet shown in Figure 12.6. In this case, the detector is often preceded by an RF amplifier with a typical complete receiver shown in **Figure 12.7**. Such a receiver can be very simple to construct, yet can be quite effective — especially for the ultra-compact low-power consumer-oriented portable station known as the mobile telephone! In fact, given the direct conversion receiver's use in the mobile telephone, it is the most widely used of all receivers.

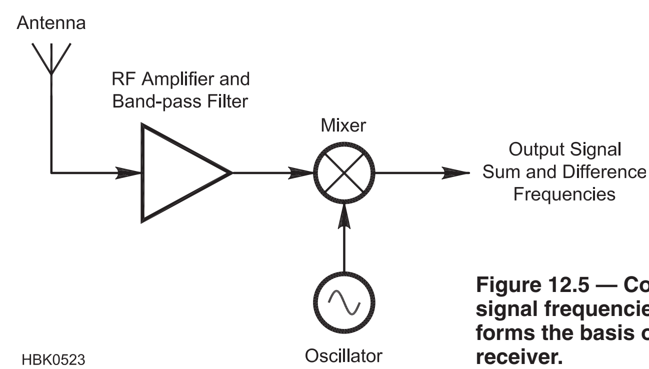


Figure 12.5 — Converting received signal frequencies using a mixer forms the basis of the heterodyne receiver.

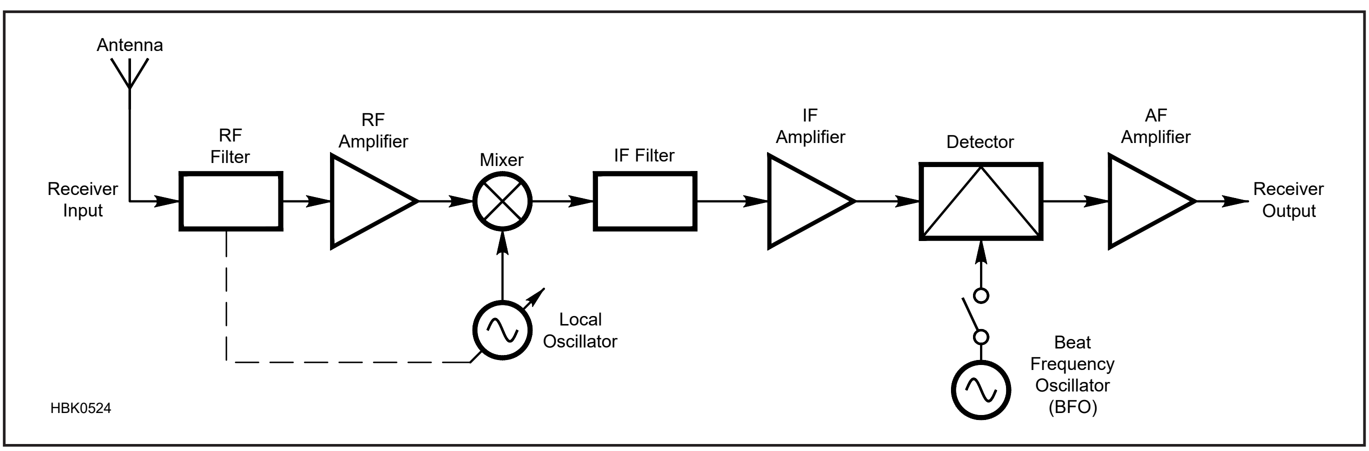


Figure 12.6 — Elements of an analog superheterodyne radio receiver.

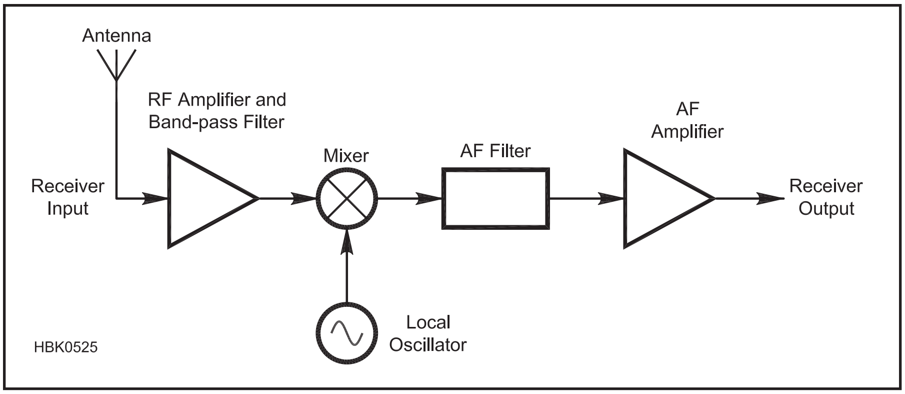


Figure 12.7 — Block diagram of a direct conversion receiver.

The basic function of a mixer is to multiply two sinusoidal signals and generate two new signals with frequencies that are the sum and difference of the two original signals. This function can be performed by a linear multiplier, a switch that turns one input signal on and off at the frequency of the other input signal, or a nonlinear circuit such as a diode. (The output of a nonlinear circuit is made up of an infinite series of products, all different combinations of the two input signals, as described in the sidebar on Nonlinear Signal Combinations.) Much more information about the theory, operation and application of mixers may be found in this chapter's section on them.

Figure 12.8 shows the progression of the spectrum of an on-off keyed CW signal through such a receiver based on the relationships described above. (This example is based on the "Rock-Bending Receiver for 7 MHz" which is included in the online projects.) In 12.8C, we include an undesired image signal on the other side of the local oscillator that also shows up in the output of the receiver. Note each of the desired and undesired responses that occur as outputs of the mixer.

Some mixers are designed to be *balanced* in order to cancel one of the input signals at the output while a *double-balanced* mixer cancels both. A double-balanced mixer simplifies the output filtering job as shown in Figure 12.8D.

Products generated by nonlinearities in the mixing process (see the previous sidebar on Nonlinear Signal Combinations) are heard as intermodulation distortion signals that we will discuss later. Note that the nonlinearities also allow mixing with unwanted signals near multiples of local oscillator frequency. These signals, such as those from TV or FM broadcast stations, must be eliminated in the filtering before the mixer since their audio output will be right in the desired passband on the output of the mixer.

12.2.2 Superheterodyne Receivers

In many instances, it is not possible to achieve all the receiver design goals with a single-conversion receiver and multiple conversion steps are used, creating the superheterodyne architecture. Traditionally, the first conversion is tasked with removing the RF image signals, while the second generates the IF signal where the signal is amplified and filtered.

The superheterodyne concept was introduced by Major Edwin Armstrong, a US Army artillery officer, just as WW I was coming to a close. He is the same Armstrong who invented frequency modulation (FM) some years later and who held many radio patents between WW I and WW II.

In a superhet, a local oscillator and mixer are used to translate the received signal to an intermediate frequency rather than directly to audio. This provides an opportunity for additional amplification and processing. Then a second mixer is used as in the DC receiver to detect the IF signal, translating it to audio. The configuration was shown previously in Figure 12.6.

An example will illustrate how this works. Let's pick a common IF frequency used in a simple AM broadcast radio, 455 kHz. If we want to listen to a 600 kHz broadcast station, the RF stage would be set to amplify the 600 kHz signal and the LO should be set to 600 + 455 kHz or 1055 kHz. The 600 kHz signal, along with any audio information it contains, is translated to the IF frequency and is amplified and then detected.

Note that we could have also set the local oscillator to 600 – 455 kHz or 145 kHz. By setting it to the sum, we reduce the relative range that the oscillator must tune. To cover the 500 to 1700 kHz with the difference, our LO would have to cover from 45 to 1245, a 28:1 range. Using the sum requires LO coverage from 955 to 2155 kHz, a range of about 2.5:1 — much easier to implement.

Note that to detect standard AM signals, the receiver's second oscillator, the beat frequency oscillator (BFO), is turned off since the AM station provides its own carrier signal over the air. Receivers designed only for standard AM reception generally don't have a BFO at all.

It's not clear yet that we've gained anything by doing this; so let's look at another example. If we decide to change from listening to the station at 600 kHz and want to listen to another station at, say, 1560 kHz, we can tune the single dial of our superhet to 1560 kHz. The RF stage is tuned to 1560 kHz, the LO is set to 1560 + 455 or 2010 kHz, and now the desired station is translated to our 455 kHz IF where the bulk of our amplification can take place. Note also that with the superheterodyne configuration, selectivity (the ability to separate stations) occurs primarily in the intermediate-frequency (IF) stages and is thus the same no matter what frequencies we choose to listen to. This simplifies the design of each stage considerably.

12.2.3 Superheterodyne Bandwidth

Now we will discuss the bandwidth requirements of different operating modes and how that affects superhet design. One advantage of a superhet is that the operating bandwidth can be established by the IF stages, and further limited by the audio system. It is thus independent of the RF frequency to which the receiver is tuned. It should not be surprising that the detailed design of a superhet receiver is dependent on the nature of the signal being received. We will briefly discuss the most commonly received modulation types and the bandwidth implications of each below. The typical operating modes expected to be encountered by an HF communications receiver are tabulated in **Table 12.3**. (Each modulation type is discussed in more detail in the **Modulation** or **Digital Protocols and Modes** chapters.) The same concerns for receiver bandwidth apply to SDR receivers although the filters are implemented using DSP techniques.

Table 12.3

Typical Communications Bandwidths for Various Operating Modes

Mode	*Bandwidth (kHz)*
FM Voice	15
AM Broadcast	10
AM Voice	4-6.6
SSB Voice	1.8-3
Digital Voice	1.0-1.2
RTTY (170-850 Hz shift)	0.3-1.5
CW	0.1-0.5

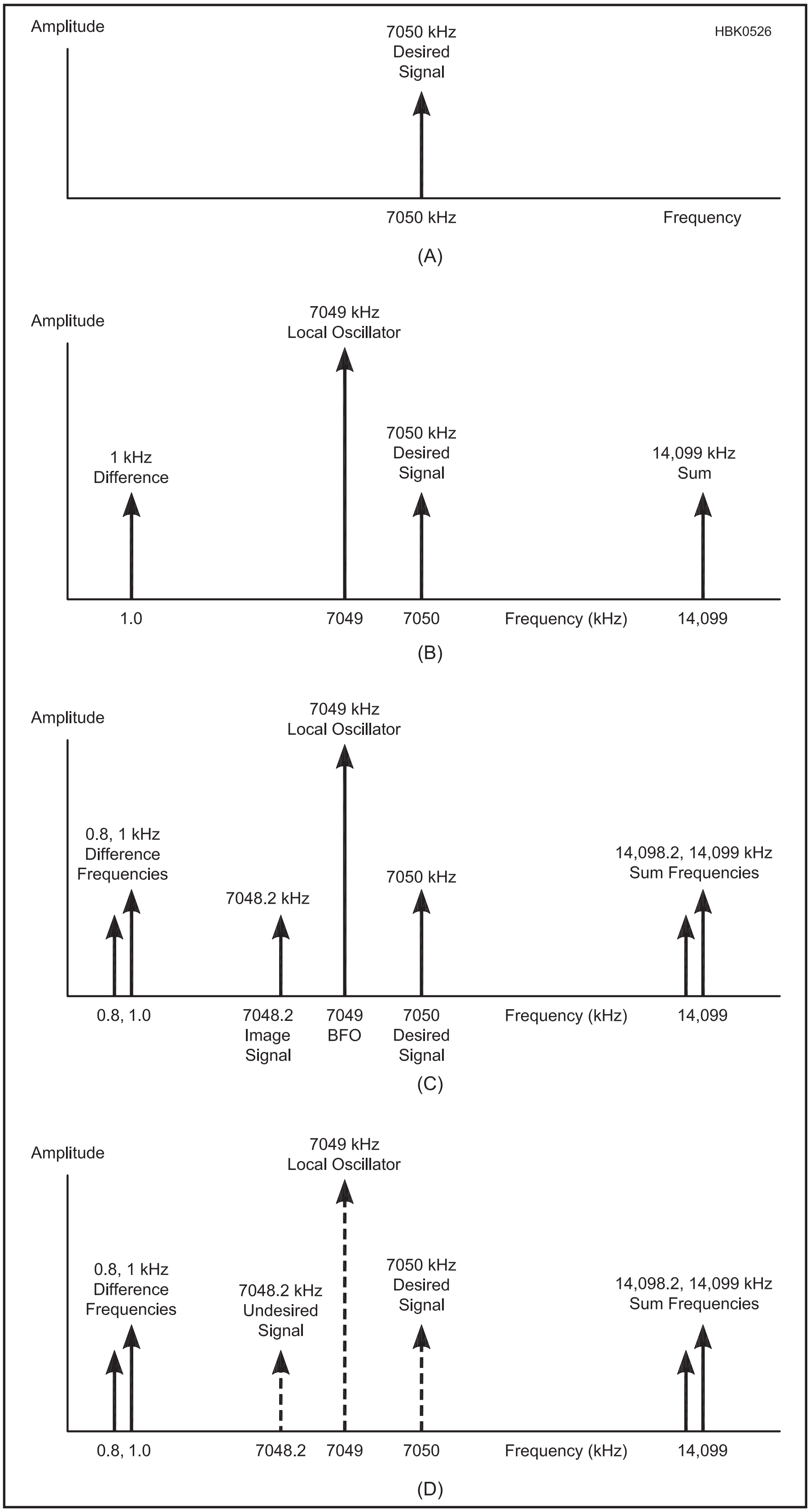


Figure 12.8 — Frequency relationships in a direct conversion receiver. At (A), the desired receive signal from antenna, a 7050 kHz on-off keyed carrier. At (B), the internal local oscillator (LO) and receive frequency relationships. At (C), the frequency relationships of mixer/detector products (not to scale). At (D), the sum and difference outputs from a double balanced mixer (not to scale). Note the mixer inputs that are balanced out so that they cancel in the output (dashed lines).

AMPLITUDE MODULATION AM)

As shown in Figure 12.8, multiplying (in other words, modulating) a carrier with a single tone results in the tone being translated to frequencies of the sum and difference of the two. Thus, if a transmitter were to multiply a 600 Hz tone and a 600 kHz carrier signal, we would generate additional new frequencies at 599.4 and 600.6 kHz. If instead we were to modulate the 600 kHz carrier signal with a band of frequencies corresponding to human speech of 300 to 3300 Hz (the usual range of communication quality voice signals), we would have a pair of information-carrying sidebands extending from 596.7 to 603.3 kHz, as shown in **Figure 12.9**.

Note that the total bandwidth of this AM voice signal is twice the highest modulation signal frequency, or 6600 Hz. If we choose to transmit speech and limited-range music, we might allow modulating frequencies up to 5000 Hz, resulting in a bandwidth of 10,000 Hz or 10 kHz. This is the standard channel spacing that commercial AM broadcasters use in the US. (9 kHz is used in Europe) In actual use, transmitters on adjacent channels are generally geographically separated, so broadcasters can extend some energy into the next channels for improved fidelity. We would refer to this as a *narrow-bandwidth* mode.

What does this say about the bandwidth needed for our receiver? If we want to receive the full information content transmitted by a US AM broadcast station, then we need to set the bandwidth to at least 10 kHz. What if our receiver has a narrower bandwidth? Well, we will lose the higher frequency components of the transmitted signal — perhaps ending up with a radio suitable for voice but not very good at reproducing music.

On the other hand, what is the impact of having too wide a bandwidth in our receiver? In that case, we will be able to receive the full transmitted spectrum but we will also receive some of the adjacent channel information. This will sound like interference and reduce the quality of what we are receiving. If there are no adjacent channel stations, we will get any additional noise from the additional bandwidth and minimal additional information. The general rule is that the received bandwidth should be matched to the bandwidth of the signal we are trying to receive to maximize SNR and minimize interference.

As the receiver bandwidth is reduced, intelligibility suffers, although the SNR is improved. With the carrier centered in the receiver bandwidth, most voices are difficult to understand at bandwidths less than around 4 kHz. In cases of heavy interference, full carrier AM can be received as if it were SSB, as described below, with the carrier inserted

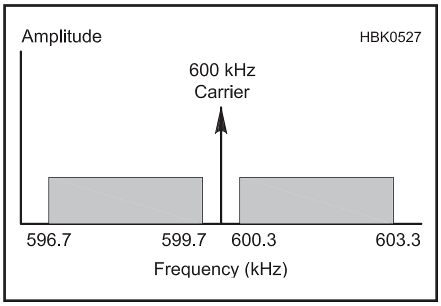


Figure 12.9 — Spectrum of sidebands of an AM voice signal with a carrier frequency of 600 kHz.

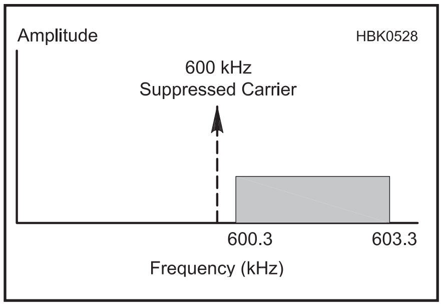


Figure 12.10 — Spectrum of single sideband AM voice signal with the suppressed carrier frequency of 600 kHz.

at the receiver, and the receiver tuned to whichever sideband has the least interference.

SINGLE-SIDEBAND (SSB)

A single-sideband signal is contains just one of the AM signal's sideband and no carrier, as shown in **Figure 12.10**. To receive the SSB signal, the receiver uses a BFO (beat frequency oscillator) in the receiver to provide a substitute carrier. The BFO oscillator signal is multiplied with the sideband in order to provide demodulated audio output. The implications in the receiver are that the bandwidth can be slightly less than half that required for double sideband AM (DSB). The tradeoff is that the BFO signal must be at exactly the right frequency. If the frequency is improperly set, the frequency of the demodulated audio will be offset from the original signal. The effect is quite audible even for small frequency errors of a few tens of Hz. (The *QST* article "About SSB" by the editor illustrates this effect, including a video on tuning in an SSB signal — see the References section.)

This results in a requirement for a much more stable receiver design with a much finer tuning system — a more expensive proposition than the DC receiver. An alternate is to transmit a reduced level carrier and have the receiver lock on to the weak carrier, usually called a *pilot* carrier. Note that the pilot carrier need not be of sufficient amplitude to demodulate the signal, just enough to allow a BFO to lock to it. These alternatives are effective, but tend to make SSB receivers expensive, complex and most appropriate for the case in where a small number of receivers are listening to a single transmitter, as is the case of two-way amateur communication.

Note that the bandwidth required to effectively demodulate an SSB signal is actually less than half that required for the AM signal because the range centered on the AM carrier need not be received. Thus the communications-quality range of 300 to 3300 Hz can be received in a bandwidth of 3000 not 3300 Hz. Early SSB receivers typically used a bandwidth of around 3 kHz, but with the heavy interference frequently found in the amateur bands, it is more common for amateurs to use bandwidths of 1.8 to 2.4 kHz with the corresponding loss of some of the higher- and lower-frequency speech sounds.

RADIOTELEGRAPHY (CW)

We have described radiotelegraphy as being transmitted by "on-off keying of a carrier." You might think that since a carrier takes up just a single frequency, the receive bandwidth needed should be almost zero. This is only true if the carrier is never turned on and off. In the case of CW, it will be turned on and off quite rapidly. The rise and fall of the carrier results in sidebands extending on either side of the carrier, and they must be received in order to reconstruct the signal in the receiver.

A rule of thumb is to consider the rise and fall time as about 10% of the pulse width and the bandwidth as the reciprocal of the quickest of rise or fall time. This results in a bandwidth requirement of about 50 to 200 Hz for the usual CW transmission rates. Another way to visualize this is with the bandwidth being set by a high-Q tuned circuit. Such a circuit will continue to "ring" after the input pulse is gone. Thus, too narrow a bandwidth will actually "fill in" between the code elements and act like a "no bandwidth" full period carrier and this is exactly what is heard if a very narrow crystal filter is used when receiving CW.

DIGITAL MODULATION

(This is a short overview to establish receiver requirements for the digital modes. See the **Modulation** and **Digital Protocols and Modes** chapters and the **Digital Communications** chapter in the *Handbook's* online content for more in-depth treatment.)

The Baudot code (used for teletype communications) and ASCII code — two popular digital communications codes used by amateurs — are constructed with sequences of elements or bits. The state of each bit — ON or OFF — is represented by a signal at one of two distinct frequencies: one designated *mark* and one designated *space*. This is referred to as *frequency shift keying* (FSK). The transmitter frequency shifts back and forth with each character's individual elements.

Amateur operators typically use a 170 Hz separation between the mark and space frequencies for the most popular FSK mode, radioteletype or RTTY, depending on the data rate and local convention, although 850 Hz is sometimes used. The minimum bandwidth required to recover the data is approximately twice the spacing between the tones. Note that the tones can be generated by directly shifting the carrier frequency (*direct FSK*), or by using a pair of 170 Hz spaced audio tones applied to the audio input of an SSB transmitter (*audio FSK* or *AFSK*). Direct FSK and AFSK are indistinguishable to a receiver.

Note that if the standard audio tones of 2125 Hz (mark) and 2295 Hz (space) are used, they fall within the bandwidth of a voice channel and thus a voice transmitter and receiver can be employed without any additional processing needed outside the radio equipment.

If the receiver can shift its BFO frequency appropriately, the two tones can be received through a filter designed for CW reception with a bandwidth of about 300 Hz or wider. Some receivers provide such a narrow filter with the center frequency shifted midway between the tones (2210 Hz) to avoid the need for retuning. The most advanced receivers provide a separate filter for mark and space frequencies, thus maximizing interference rejection and signal-to-noise ratio (SNR). Using a pair of tones for FSK or AFSK results in a maximum data rate of about 1200 bit/s over a high-quality voice channel.

Phase shift keying (PSK) can also be used to transmit bit sequences, requiring good frequency stability to maintain the required time synchronization to detect shifts in phase. If the channel has a high SNR, as is often the case at VHF and higher, telephone network data-modem techniques such as Bell 102 and Bell 202 can be used. (FCC §97.307(f) specifies a maximum transmitted symbol rate for each band.)

At HF, the signal is subjected to phase and amplitude distortion as it travels. Noise is also substantially higher on the HF bands. Under these conditions, modulation and demodulation techniques designed for "wireline" connections become unusable at bit rates of more than a few hundred baud. As a result, amateurs have begun adopting and developing state of the art digital modulation techniques. These include the use of multiple carriers (MFSK, Clover, PACTOR III, etc.), multiple amplitudes and phase shifts (QAM and QPSK techniques), and advanced error detection and correction methods to achieve a net data throughput as high as 3600 bits per second (bps) over a voice-bandwidth channel. Newly developed coding methods for digital voice using QPSK modulation result in a signal bandwidth of less than 1200 Hz. (See **freedv.org** for more information.) Spread-spectrum techniques are also being adopted on the amateur UHF bands, but are beyond the scope of this discussion.

The bandwidth required for data communications can be as low as 100 Hz for PSK31 to 1 kHz or more for the faster speeds of PACTOR III and Clover. Beyond having sufficient bandwidth for the data signal, the primary requirements for receivers used for data communications are linear amplitude and phase response over the bandwidth of the data signal to avoid distorting these critical signal characteristics. The receiver must also have excellent frequency stability to avoid drift and frequency resolution to enable the receiver filters to be set on frequency.

FREQUENCY MODULATION (FM)

Another popular voice mode is *frequency modulation* or *FM*. FM can be found in a number of variations depending on purpose. In Amateur Radio and commercial mobile communication use on the shortwave bands, it is universally *narrow band FM* or *NBFM*. In NBFM, the frequency deviation is limited to around the maximum modulating frequency, typically 3 kHz. The bandwidth requirements at the receiver can be approximated by Carson's Rule of BW = 2 × (D + M), where D is the deviation and M is the maximum modulating frequency. Thus 3 kHz deviation and a maximum voice frequency of 3 kHz results in a bandwidth of 12 kHz, not far beyond the requirements for broadcast AM. (Additional signal components extend beyond this bandwidth, but are not required for voice communications.)

In contrast, broadcast or *wideband FM* or *WBFM* occupies a channel width of 150 kHz. Originally, this provided for a higher modulation index, even with 15 kHz audio that resulted in an improved SNR. However, with multiple channel stereo and sub-channels all in the same allocated bandwidth the deviation is around the maximum transmitted signal bandwidth.

In the US, FCC amateur rules limit wideband FM use to frequencies above 29 MHz. Some, but not all, HF communication receivers provide for FM reception. For proper FM reception, two changes are required in the receiver architecture as shown within the dashed line in **Figure 12.11**. The fundamental change is that the detector must recover information from the frequency variations of the input signal. The most common such detector is called a *discriminator*. The discriminator does not require a BFO, so that is turned off, or eliminated in a dedicated FM receiver. Since amplitude variations convey no information in FM, they are generally eliminated by a *limiter*. The limiter is a high-gain IF amplifier stage that clips the positive and negative peaks of signals above a certain threshold. Since most noise of natural origins is amplitude modulated, the limiting process also strips away noise from the signal.

12.2.4 Superheterodyne FM Receivers

Narrow-band frequency modulation (NBFM) is the most common mode used on VHF and UHF. **Figure 12.12** is a block diagram of an FM receiver for the VHF/UHF amateur bands. Many FM transmitters are actually phase-modulated (PM) but aside from frequency response of the transmitted audio, the two types of signals can be received with the same equipment. This section's references to FM include PM signals unless noted otherwise.

FRONT END

A low-noise front end is desirable because of the decreasing atmospheric noise level at these frequencies and also because portable gear often uses short rod antennas at ground level. Nonetheless, the possibilities for gain compression and harmonic IMD, multi-tone IMD and cross modulation are also substantial. Therefore dynamic range is an important design consideration, especially if large, high-gain antennas are used. FM limiting should not occur until after the crystal filter. Because of the high occupancy of the VHF/UHF spectrum by powerful broadcast transmitters and nearby two-way radio services, front-end preselection is desirable, so that a low noise figure can be achieved economically within the amateur band. (See the section on Preselectors elsewhere in this chapter.)

DOWN-CONVERSION

Down-conversion to the final IF can occur in one or two stages. Favorite IFs are in the 5 to 10 MHz region (10.7 MHz is common), but at the higher frequencies rejection of the image 10 to 20 MHz away can be difficult, requiring considerable preselection as discussed in the sections below on IF selection and image rejection. At the higher frequencies an intermediate IF in the 30 to 50 MHz region is a better choice. Figure 12.12 shows dual down-conversion.

IF FILTERS

The customary peak frequency deviation in amateur FM on frequencies above 29 MHz is about 5 kHz and the audio speech band extends to 3 kHz. This defines a maximum modulation index (defined as the deviation ratio) of 5/3 = 1.67. An inspection of the Bessel functions that describe the resulting FM signal shows that this condition confines most of the 300 to 3000 Hz speech information sidebands within a 15 kHz or so bandwidth. Using filters of this bandwidth, channel separations of 20 or 25 kHz are achievable.

Many amateur FM transceivers are channelized in steps that can vary from 1 to 25 kHz. For low distortion of the audio output

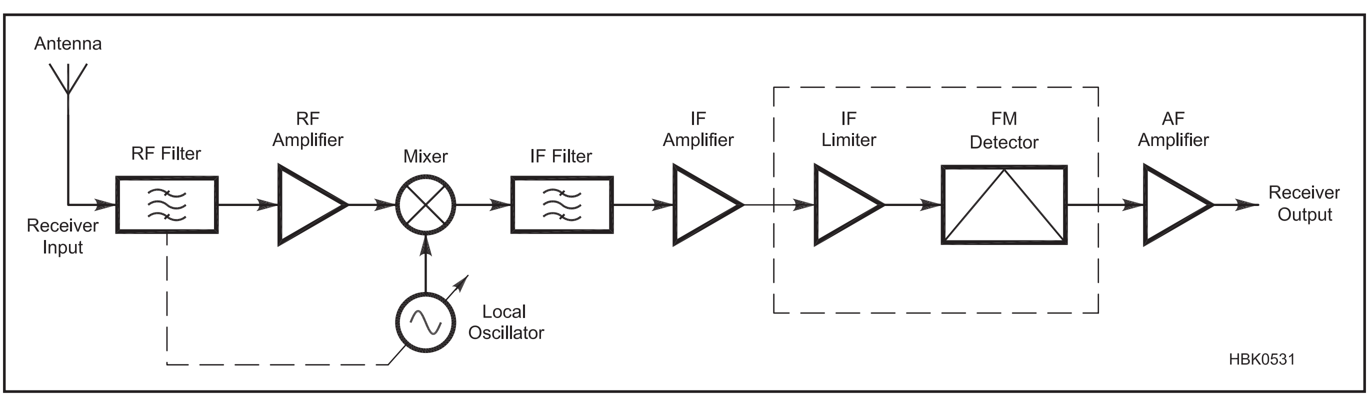


Figure 12.11 — Block diagram of an FM superheterodyne receiver. Changes from an AM/SSB receiver are enclosed by the dashed line.

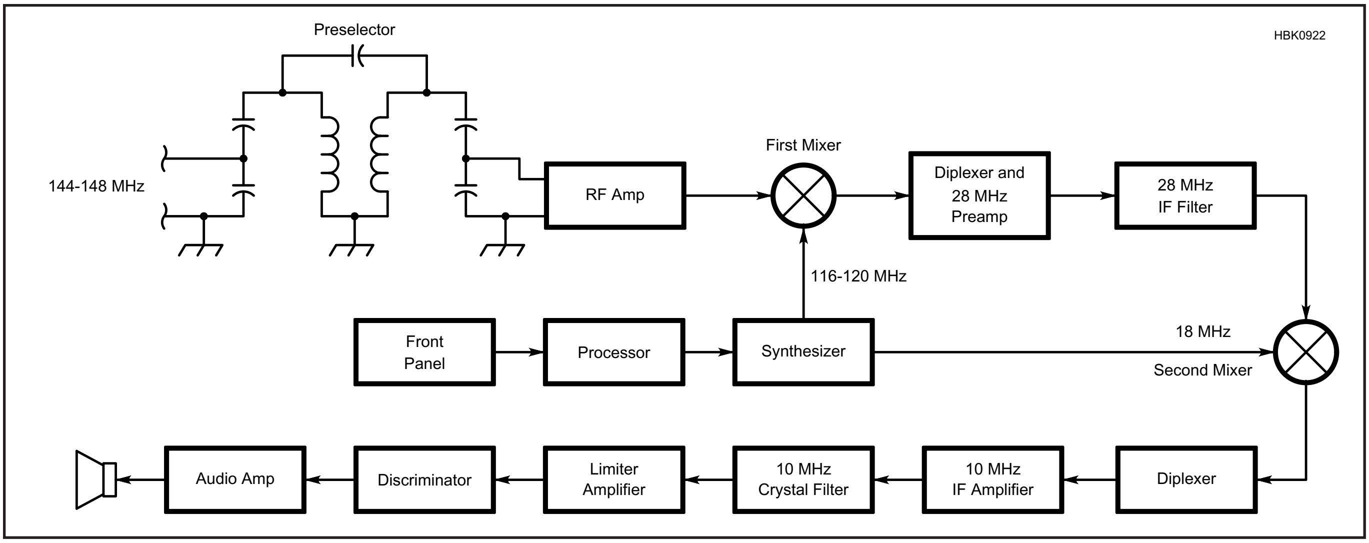


Figure 12.12 — Block diagram of a typical VHF FM receiver using dual down-conversion.

(after FM detection), this filter should have good phase linearity across the bandwidth. This would seem to preclude filters with very steep descent outside the passband, which tend to have very nonlinear phase near the band edges. But since the amount of energy in the higher speech frequencies is naturally less, the actual distortion due to this effect may be acceptable for speech purposes. The normal practice is to apply pre-emphasis to the higher speech frequencies at the transmitter and de-emphasis compensates at the receiver.

LIMITING AND DEMODULATION

After the filter, hard limiting of the IF is needed to remove any amplitude modulation components. In a high-quality receiver, special attention is given to any nonlinear phase shift that might result from the limiter circuit design. This is especially important in data receivers in which phase response must be controlled. In amateur receivers for speech it may be less important. Also, the *ratio detector* largely eliminates the need for a limiter stage, although the limiter approach is probably still preferred. FM demodulation is described in the section on Demodulation and Detection.

FM PERFORMANCE FOR WEAK SIGNALS

The noise bandwidth of the IF filter is not much greater than twice the audio bandwidth of the speech modulation, less than it would be in wideband FM. Therefore such things as capture effect, the threshold effect and the noise quieting effect so familiar to wideband FM are still operational, but somewhat less so, in FM. For FM receivers, sensitivity is specified in terms of a SINAD (see the **Test Equipment and Measurements** chapter) ratio of 12 dB. Typical values are –110 to –125 dBm, depending on the low-noise RF pre-amplification that often can be selected or deselected (in strong signal environments).

EFFECT OF LO PHASE NOISE

In an FM receiver, LO phase noise is superimposed on phase modulation, and therefore creates frequency modulation of the desired signal. This reduces the ultimate signal-to-noise ratio within the passband. This effect is called *incidental FM (IFM)*. The power density of IFM (W/Hz) is proportional to the phase noise power density (W/Hz) multiplied by the square of the modulating frequency (the familiar parabolic effect in FM). If the receiver uses high-frequency de-emphasis at the audio output (–6 dB per octave from 300 to 3000 Hz, a common practice), the IFM level at higher audio frequencies can be reduced. Ordinarily, as the signal increases the noise would be "quieted" (that is, "captured") in an FM receiver, but in this case the signal and the phase noise riding "piggy back" on the signal increase in the same proportion as described in this and the **Oscillators and Synthesizers** chapter's discussion of reciprocal mixing. IFM is not a significant problem in modern FM radios, but phase noise can become a concern for adjacent-channel interference.

As the signal becomes large the signal-to-noise ratio therefore approaches some final value. A similar ultimate SNR effect occurs in SSB receivers. On the other hand, a perfect AM receiver tends to suppress LO phase noise. (See the reference entry for Sabin.)

FM RECEIVER ICs

A wide variety of special ICs for communications-bandwidth FM receivers are available. Many of these were designed for "cordless" or mobile telephone applications and are widely used. One is an RF amplifier chip (NE/SA5204A) for 50 Ω input to 50 Ω output with 20 dB of gain. The second chip (NE/SA602A) is a front-end device with an RF amplifier, mixer and LO. The third is an IF amplifier, limiter and quadrature FM detector (NE/SA604A) that also has a very useful RSSI (logarithmic Received Signal Strength Indicator) output and also a "mute" function. The fourth is the LM386, a widely used audio-amplifier chip. Another FM receiver chip, complete in one package, is the MC3371P.

The MC13135 features double conversion and two IF amplifier frequencies. This allows more gain on a single chip with less of the cross coupling that can degrade stability. This desirable feature of multiple down-conversion was mentioned previously in this chapter.

Design details and specific parts values can be learned from a careful study of the data sheets and application notes provided by the IC vendors. Amateur designers should learn how to use these data sheets and other information such as application notes available online from the manufacturers.

12.2.5 Superheterodyne Image Rejection

Now that we have established the range of bandwidths that our receiver will need to pass, we are in a position to discuss the selection of the IF frequency at which those bandwidths will be established.

In addition, selection of the first and any following IF frequencies is important in receiving weak signals on bands where strong signals are present. If the desired signal is near the noise, a signal at an image frequency could easily be 100 dB stronger, and thus to avoid interference, an image rejection of 110 dB would be needed. While some receivers meet that target, the receiver sections of most current amateur transceivers are in the 70 to 100 dB range.

IF IMAGE RESPONSE

As noted earlier, a superhet with a single local oscillator or LO and specified IF can receive two frequencies, selected by the tuning of the RF stage. For example, using a receiver with an IF of 455 kHz to listen to a desired signal at 7000 kHz can use an LO of 7455 kHz. However, the receiver will also receive a signal at 455 kHz above the LO frequency, or 7910 kHz. This undesired signal frequency, located at twice the IF frequency from the desired signal, is called an *image*.

Images will be separated from the desired frequency by twice the IF and the filter ahead of the associated mixer must reduce the image signal by the amount of the required *image rejection*. For a given IF, this gets more difficult as the received frequency is increased. For example, with a 455 kHz single conversion system tuned to 1 MHz, the image will be at 1.91 MHz, almost a 2:1 frequency ratio and relatively easy to reject with a filter. The same receiver tuned to 30 MHz, would have an image at 30.91 MHz, a much more difficult filtering problem

While an image that falls on an occupied channel is obviously a problem — it's rarely desirable to receive two signals at the same time — problems occur even if the image frequency is clear of signals. This is because the atmospheric noise in the image bandwidth is added to the noise of the desired channel, as well as any internally-generated noise in the RF amplifier stage. If the image response is at the same level as the desired signal response, there will be a 3 dB reduction in SNR.

REDUCING IMAGES BY INCREASING IF

An obvious solution to the RF image response is to raise the IF frequency high enough so that signals at twice the IF frequency from the desired signal are sufficiently attenuated by filters ahead of the mixer. This can easily be done with IF stages operating at 5 to 10% of the highest receiving frequency (1.5 to 3 MHz for a receiver that covers the 3-30 MHz HF band). The concept is used at higher frequencies as well. The FM broad-cast band (150 kHz wide channels over 87.9 to 108 MHz in the US) is generally received on superhet receivers with an IF of 10.7 MHz, which places all image frequencies outside the FM band, eliminating interference from other FM stations.

The use of higher-frequency tuned circuits for IF selectivity works well for the 150 kHz wide FM broadcast channels, but not so well for the relatively narrow channels encountered on HF or lower, or even for many V/UHF narrowband services. Fortunately, there are three solutions that were commonly used to resolve this problem.

The first, *double conversion*, converted the desired signal to a relatively high IF followed by a second conversion to a lower IF to set the selectivity. This was a popular technique in the 1950s. Improvements on the double conversion technique led to *triple conversion* with a very low, highly selective third IF. The *Collins system* of moving a single-range VFO to the second mixer and using switchable crystal oscillators for the first mixer also became popular. The *pre-mixed* arrangement, a third approach to double conversion was a combination of the two, used a single variable oscillator range, as with the Collins, but mixes the VFO and LO before applying them to the first mixer – outside of the signal path.

These methods are obsolete for current receivers but are commonly encountered in vintage equipment. They are discussed in previous editions of the *Handbook*.

HIGH-FREQUENCY CRYSTAL LATTICE FILTERS

Commercial quartz-crystal filters with bandwidths appropriate for CW and SSB became available in the 1970s with center frequencies into the 10 MHz range. This allowed a single-conversion receiver (see Figure 12.2) with an IF in the HF range to provide both high image rejection and needed channel selectivity. This single-conversion architecture remains popular among designers of portable and low-power equipment. Crystal filter design is discussed in the **Analog and Digital Filtering** chapter as well as in the online content.

As an example **Figure 12.13** shows the IF section of a single-conversion superhet using simple filters centered 1500 kHz. While the filters shown are actually buildable by amateurs at low cost, multiple-section filters with much better performance can be pur-

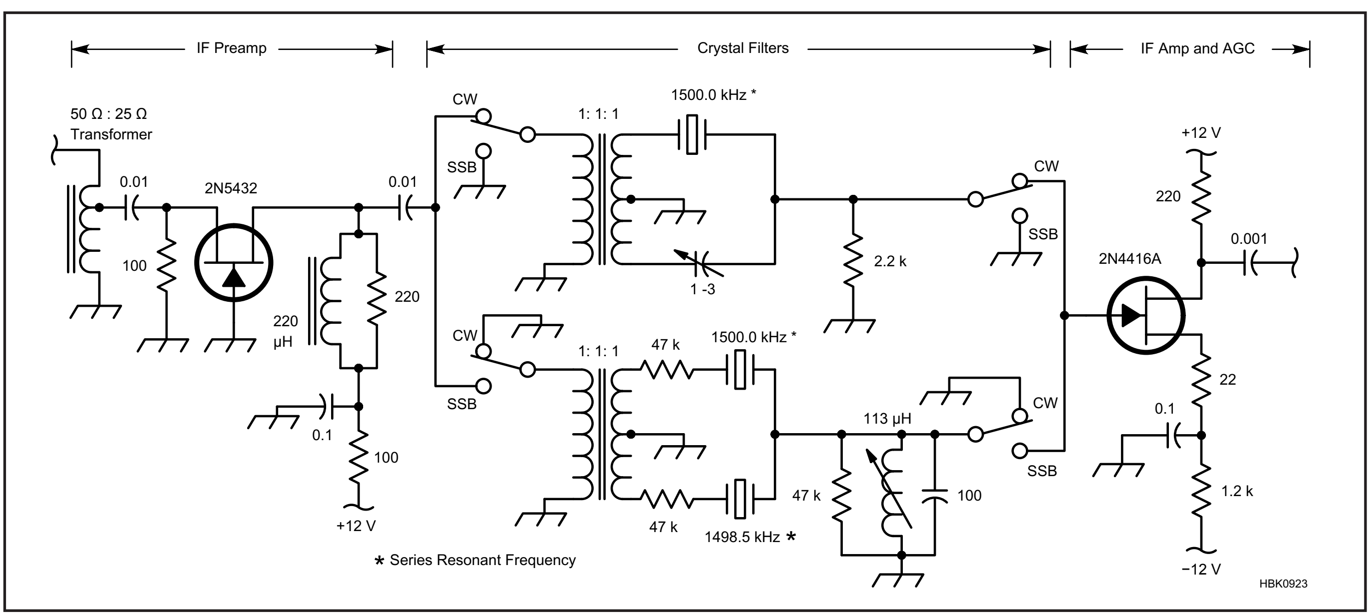


Figure 12.13 — The IF section of a superhet using crystal filters to establish receiver bandwidth.

chased or constructed. Other IF frequencies can be used, depending on crystal or filter availability.

The circuit shown demonstrates the concepts involved and can be reproduced at low cost. Remaining receiver functional blocks such as the AGC circuitry, detectors and BFO, and audio amplifiers and filters can be found elsewhere in the book.

DSP IF FILTERS

Digital signal processing provides a level of filter performance not practical with other technologies. (See the **Analog and Digital Filtering** chapter) While much better than most low frequency IF LC bandwidth filters, the very good crystal or mechanical bandwidth filters in amateur gear are not very close to the rectangular shaped frequency response of an ideal filter, but rather have skirts with a 6 to 60 dB response of perhaps 1.4 to 1. That means if we select an SSB filter with a nominal (6 dB) bandwidth of 2400 Hz, the width at 60 dB down will typically be 2400 × 1.4 or 3360 Hz. Thus a signal in the next channel that is 60 dB stronger than the signal we are trying to copy (as often happens) will have energy just as strong as our desired signal.

DSP filtering approaches the ideal response. **Figure 12.14** shows the ARRL Lab measured response of a DSP bandwidth filter with a 6 dB bandwidth of 2400 Hz. Note how rapidly the skirts drop to the noise level. In addition, while analog filtering generally requires a separate filter assembly for each desired bandwidth, DSP filtering is adjustable — often in steps as narrow as 50 Hz — in both bandwidth and center frequency. In addition to bandwidth filtering, the same DSP can often provide digital noise reduction and digital notch filtering to remove interference from fixed frequency carriers.

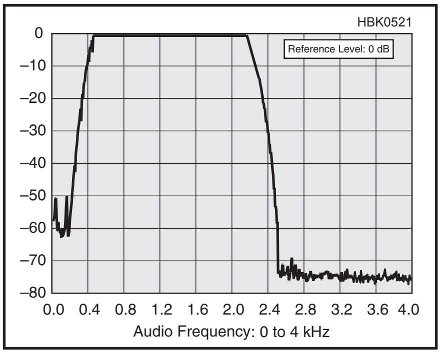


Figure 12.14 — ARRL Lab measured response of an aftermarket 2400 Hz DSP bandpass filter.

UP-CONVERSION AND DOWN-CONVERSION

Current crystal filter technology allows *down-conversion* HF receivers to use an IF in the 4-10 MHz range. With a 10 MHz IF and an LO above the signal frequency, a 30 MHz signal would have an image at 50 MHz. This makes image-rejection filtering relatively straightforward, although many receiver IF frequencies tend to be at the lower end of the above range. Still, as will be discussed in the next section, they have other advantages.

Many current HF receivers (or receiver sections of transceivers) have elected to employ an *up-conversion* architecture. They typically have an IF in the VHF range, perhaps 60 to 70 MHz, making HF image rejection easy. A 30 MHz signal with a 60 MHz IF will have an image at 150 MHz. Not only is it five times the signal frequency, but signals in this range (other than perhaps the occasional taxicab) tend to be weaker than some undesired HF signals. Receivers with this architecture have image responses at the upper end of the above range, often with the image rejected by a relatively simple low-pass filter with a cut off at the top of the receiver range.

Another advantage of this architecture is that the local oscillator can cover a wide continuous range, making it convenient for a general coverage receiver. For example, with a 60 MHz IF, a receiver designed for LF through HF would need an LO covering 60.03 to 90 MHz, a 1.5 to 1 range, easily provided by a number of synthesizer technologies, as described in the **Oscillators and Synthesizers** chapter. The article "A High Performance 45 MHz IF Amplifier" by Colin Horrabin, G3SBI, has been included in the online content as an example of circuits suitable for up-conversion receivers.

The typical up-converting receiver uses multiple conversions to move signals to frequencies at which operating bandwidth can be established. While crystal filters in the VHF range used by receivers with upconverting IFs have become available with bandwidths as narrow as around 3 kHz, they do not yet achieve the shape factor of similar bandwidth filters at MF and HF. Thus, these are commonly used as *roofing filters*, discussed in the following section on Superhet Design for Dynamic Range. They are inserted in the IF chain, prior to a conversion to one or more lower IF frequencies at which the operating bandwidth is established. **Figure 12.15** is a block diagram of a typical upconverting receiver using DSP for setting the operating bandwidth.

THE IMAGE-REJECTING MIXER

Another technique for reduction of image response in receivers is not as commonly encountered in HF receivers as the preceding designs, but it deserves mention because it has some very significant applications. The *image-rejecting* mixer requires phase-shift networks, as shown in **Figure 12.16**. Frequency F_1 represents the input frequency while F_2 is that of the local oscillator. Note that the two 90° phase shifts are applied at different frequencies. The phase shift network following Mixer 1 is at a fixed center frequency corresponding to the IF, while the phase shift network at F_2 must provide the required phase shift as the local oscillator tunes across the band.

If the local oscillator is required to tune over a limited fractional frequency range, this is a very feasible approach. On the other hand, maintaining a 90° phase shift over a wide range can be tricky. The good news is that this approach provides image reduction that is independent of, and in addition to, any other mechanisms such as filters that are employed toward that end.

Additionally, with the ability of DSP components to operate at higher and higher frequencies, the necessary operations seen in Figure 12.16 can be performed in software which does not depend on precision hardware design to maintain nearly exact phase relationships.

The image-rejecting mixing process has several attractive features:

• It is the only way to provide "single signal" reception with a direct conversion receiver, effectively reducing the audio image. This can make the DC receiver a very good performer, although the added complexity is not always warranted in typical amateur DC applications.

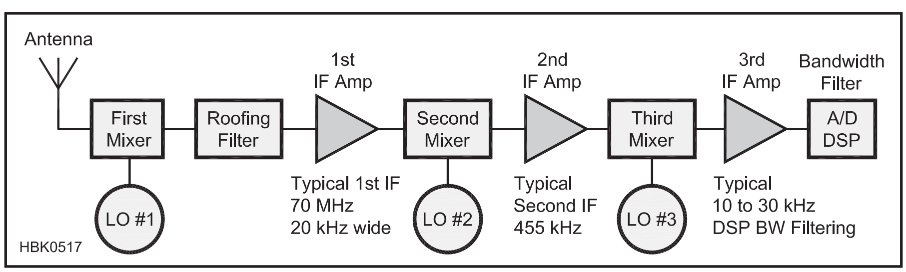


Figure 12.15 — Block diagram of a typical upconverting receiver using DSP for the 3rd IF filter which sets operating bandwidth. Hybrid heterodyne/SDR receivers applying DSP filters at IF frequencies are common.

• This option is frequently found in microwave receivers in which sufficiently selective RF filtering can be difficult to obtain. Since they often operate on fixed frequencies, maintaining the required phase shift can be straightforward.

• It is found in advanced receivers that are trying to achieve optimum performance. Even with a high first IF frequency, additional image rejection can be provided.

• In transmitters, the same system is called the *phasing method* of SSB generation. The same blocks run "backwards" — one of the phase shift networks can be applied to the speech band and used to cancel one sideband. This is discussed in the **Transmitting** chapter.

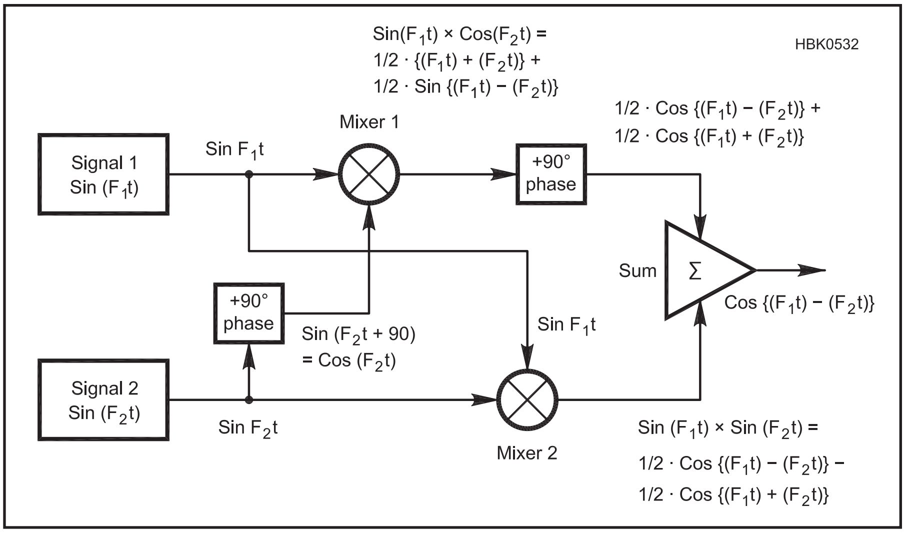


Figure 12.16 — Image rejecting mixer, block diagram, and signal relationships.

12.2.6 Preselectors

Placing a filter at the receiver's antenna input, called a *preselector*, is a technique that improves both image rejection and overload from out-of-band signals. You can see such a filter at the front-end of several of the receiver block diagrams in this chapter. Preselection to avoid overload is often required with the inexpensive USB "dongle" style SDR receivers that are designed for DTV reception and not for strong-signal performance.

There are three types of preselectors:

• Manually-tuned — a tuned circuit or tuned input transformer that is adjusted by the operator for maximum signal level. The tuned circuit usually has several ranges that are selected by the receiver band switch. These were once common, especially on general-coverage receivers. Preselectors were also very popular in European equipment where extremely strong signals from high-power shortwave broadcast (SWBC) stations often caused receiver overload unless attenuated with a filter.

• Switched band-pass — a bank of band-pass filters selected by the band switch or under control of the receiver's controlling microprocessor. Most of the top-performing receivers feature switched band-pass preselectors that do not require manual adjustment.

• Tracking — a continuously-variable tuned circuit or band-pass filter that is controlled by software to have peak response at the frequency of operation. This feature is relatively expensive to implement so it is only available on high-end receivers.

The online content includes a pair of preselector designs. The manual general-coverage preselector from *Ham Radio* magazine is by George Hirshfield, W5OZF. It covers from 0.5 to 30 MHz with a tuning capacitor augmented by fixed capacitors and tapped inductors. A software-controlled switched band-pass preselector covering 1.8 to 30 MHz from *QEX* is included as a project. Designed by Juan Onate, MØWWA and Xavier de Fortuny, the preselector includes a number of auxiliary features (input protection, attenuators, LNA, and so on) and is controlled by a PIC microcontroller. All software is available from the ARRL as described in the article.

12.2.7 Superhet Design for Dynamic Range

In the past, the receivers with the best close-in third order intermodulation distortion and maximum blocking dynamic range were amateur-band-only receivers, such as the primary receiver in the TEN-TEC Orion and Omni series, and the receivers in the Elecraft K2 and K3. A look at a typical block diagram, as shown in **Figure 12.17**, makes it easy to see why. The problems resulting from strong unwanted signals near a desired one are minimized if the unwanted signals are kept out of the places in the receiver where they can be amplified even more and cause the nonlinear effects that we try to avoid.

Note that in Figure 12.17, the only place where the desired and undesired signals all coexist is before the first mixer. If the first mixer and any RF preamp stages have sufficient strong-signal handling capability, the undesired signals will be eliminated in the filter immediately behind the first mixer. This HF crystal filter is generally switchable to support desired bandwidths as narrow as 200 Hz. The later amplifier, mixer and DSP circuits only have to deal with the signal we want. For additional discussion of these issues, see the reference entry for Hallas' Product Review of the FT-1000MP.

Careful attention to gain distribution among the stages between the filters maintains desired sensitivity, but not so high that the undesired products have a chance to become a serious problem. With bandwidths of 20, 6 and 2.5 kHz supplied, and 500 and 300 Hz as accessories, the undesired close-in signals are eliminated before they have an opportunity to cause serious trouble in the DSP stages that follow.

Another variation is found in the Kenwood TS-590S and SG models which switch between down-conversion on the more crowded "contest bands" (160, 80, 40, 20 and 15 meters) and up-conversion on the remaining bands. In effect, this trades sensitivity for dynamic range.

ROOFING FILTERS

Now look at a typical modern general-coverage receiver as shown previously in Figure 12.15. In this arrangement, a single

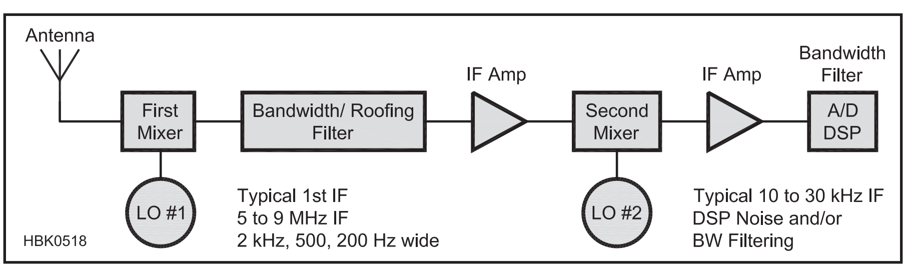


Figure 12.17 — Block diagram of a down-converting amateur band receiver with roofing filters and operating bandwidth filter to improve performance in the presence of strong in-band signals.

digital synthesizer, perhaps covering from 70 to 100 MHz, shifts the incoming signal(s) to a VHF IF, often near 70 MHz. A roofing filter at 70 MHz follows the first mixer. The name "roofing filter" comes from the filter's wide bandwidth extending over the narrower operating bandwidth filter like a house's roof. This arrangement offers simplified local oscillator (LO) design and the possibility of excellent image rejection. Unfortunately, crystal filter technology has only recently been able to produce narrow filters for 70 MHz, and so far they have much wider skirts than the crystal filters used in Figure 12.17. As technology is rapidly shifting to SDR designs, the current models are likely exhibiting the peak performance level for upconverting receivers.

Many receivers and transceivers set the roofing filter bandwidth wider than any operating bandwidth and use DSP filtering much later in the signal chain to set the final operating bandwidth. For a receiver that will receive FM and AM, as well as SSB and CW, that usually means a roofing filter with a bandwidth of around 20 kHz. With this arrangement, all signals in that 20 kHz bandwidth pass all the way through IF amplifiers and mixers and into the A/D converter before we attempt to eliminate them using DSP filters. By that time they have had an opportunity to generate intermodulation products and cause the blocking and IMD problems that we are trying to eliminate. However, top of the line heterodyne radios feature both general coverage at HF and VHF roofing filters, such as the Icom IC-7851 and Yaesu FTDX9000 transceivers.

12.3 SDR Receivers

The SDR was introduced in the **DSP and SDR Fundamentals** chapter. Repeated here are the sections of that chapter describing SDR architecture and performance issues. This section presents and compares several block-diagram-level concepts for software-defined radio. For the basic elements of SDR and DSP, see the chapter referenced above.

Sampling Rates

Digital audio equipment, that is audio equipment that uses digital sampling at some point in the device, has been around since the 1970s. There are two primary fundamental sampling rates that are used in all digital audio equipment: 44.1 ksps and 48 ksps. From these two rates, all other rates are derived by using factors of these two numbers. In amateur radio, the latter number, 48 kHz, is most often used and is why all our signal chains typically operate at a multiple or sub-multiple of these numbers. The most common processing speeds are 24 ksps, 36 ksps, 48 ksps, 96 ksps and 192 ksps. 48 ksps is by far the most common, but it varies by the manufacturer and radio design.

12.3.1 Digitizing at IF

The first generation of radios to make use of DSP techniques at RF performed the analog-digital conversion on an IF signal. **Figure 12.18** shows such a design. In such a receiver, placing the A/D converter after a crystal IF filter improves the blocking dynamic range (BDR) for interfering signals that fall outside the crystal filter bandwidth. As shown, the down-conversion to I/Q format still uses lower-speed A/D converters, but often the signal is actually at a low IF, say, 15 kHz or so. This allows an SSB-bandwidth signal to be contained within the 20 kHz bandwidth of a typical audio codec and avoids errors due to dc offsets in the signal path. With careful design, a receiver with such an architecture can achieve 140 dB or more of BDR (if there are no other limiting factors such as LO phase noise). The third-order dynamic range is similar to that achieved with a conventional analog architecture since the circuitry up to the crystal filter, including amplifiers and mixer(s) is the same.

Another advantage of the IF-based approach compared to directly sampling the RF frequency is that the ADC does not have to run at such a high sample rate. In fact, because the crystal filter acts as a high-performance, narrow-bandwidth anti-aliasing filter, *undersampling* is possible if the A/D converter has sufficient sampling bandwidth (ADCs intended for audio applications generally do not). With bandwidths of a few kHz or less, sample rates in the tens of kHz can be used even though the center frequency of the IF signal is much higher, so long as the ADC's sample-and-hold circuit has sufficient bandwidth.

12.3.2 Direct RF Digitizing

The ultimate SDR architecture is to convert between the analog and digital domains right at the frequency to be transmitted or received, or convert a wide range of frequencies and do all filtering in the digital domain. The receive path of such a design is shown in **Figure 12.19**. In this receiver, the only remaining analog components in the signal chain are a wide-band anti-aliasing filter and an amplifier to improve the noise figure of the ADC if necessary. A preselector stage may be used to prevent overload from strong signals far from the selected frequency or frequency band. The local oscillator, mixer, IF filters, AGC, demodulators and other circuitry are all replaced by digital hardware and software. The digital/software implementations of these functions are perfectly stable with time, temperature, and need no adjustments.

It has only been recently that low-cost high-

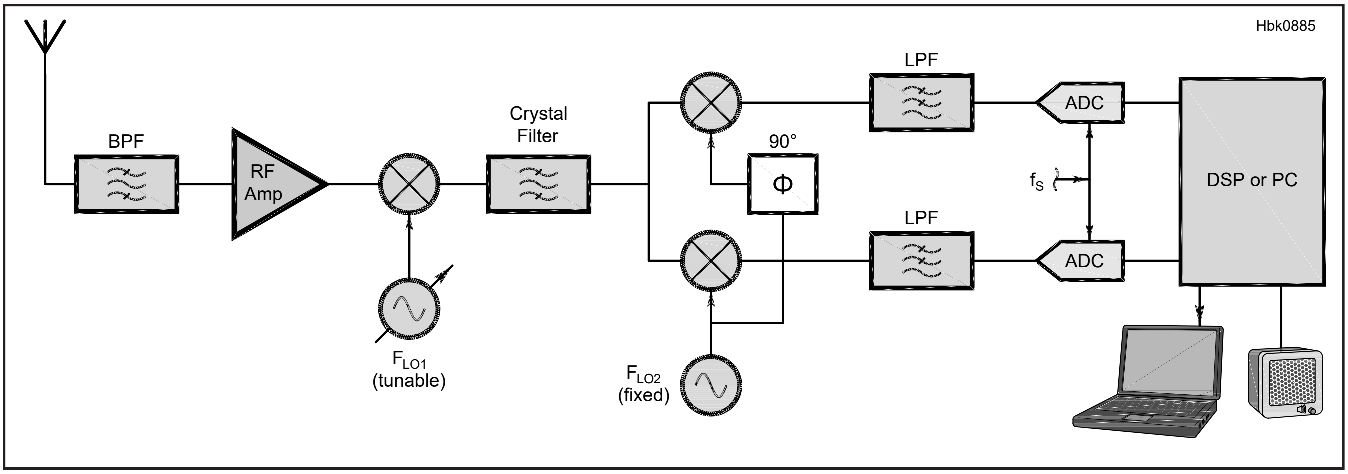


Figure 12.18 — Hybrid superhet/DSP SDR receiver architecture.

speed ADCs have become available with specifications good enough to allow reasonable performance in an RF-sampling communications receiver. Today it is possible to achieve blocking dynamic range of 130 dB. That is not quite as good as the best analog or hybrid radios, but with every new generation of A/D converter that becomes available for use in SDR, the performance gap becomes narrower.

It is worth noting here that while huge BDR numbers can be measured in the laboratory, the performance achieved in a real-world environment with a receiver connected to an antenna is quite different. Often local noise sources raise the noise floor such that the receiver's full BDR cannot be utilized and other specifications become more important. In many of these specifications (dynamic range for close-spaced signals, etc.), RF-sampling SDRs can provide performance comparable or superior to conventional all-analog and IF-sampling receivers.

Third-order dynamic range (3IMD_DR) is not a meaningful specification for this type of radio because it is based on the behavior of analog circuits. In addition, calculation of IP_3 assumes that distortion products increase 3 dB for each 1 dB increase in signal level, which is not always true for an ADC. The level of the distortion products in an ADC tends to be more-or-less independent of signal level until the signal peak exceeds the ADC's full scale input, at which point the distortion increases dramatically. It is important to read the data sheet carefully and note the test conditions for the distortion measurements.

There are definite advantages to sampling at RF. For one thing, it saves a lot of analog circuitry. Even though a high-speed ADC is more expensive than an audio converter, the radio may be end up being cheaper to build because of the reduced component count and fewer adjustments. Performance is improved in some areas. For example, image rejection

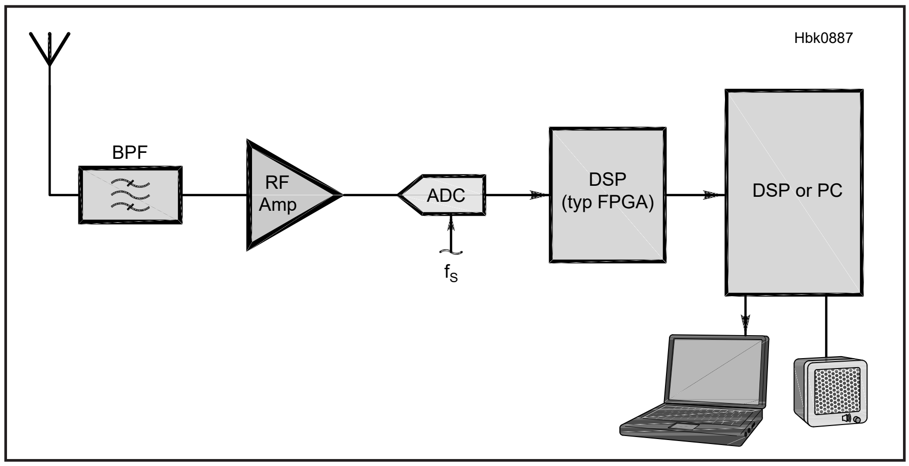


Figure 12.19 — Direct RF-sampling DSP SDR receiver architecture.

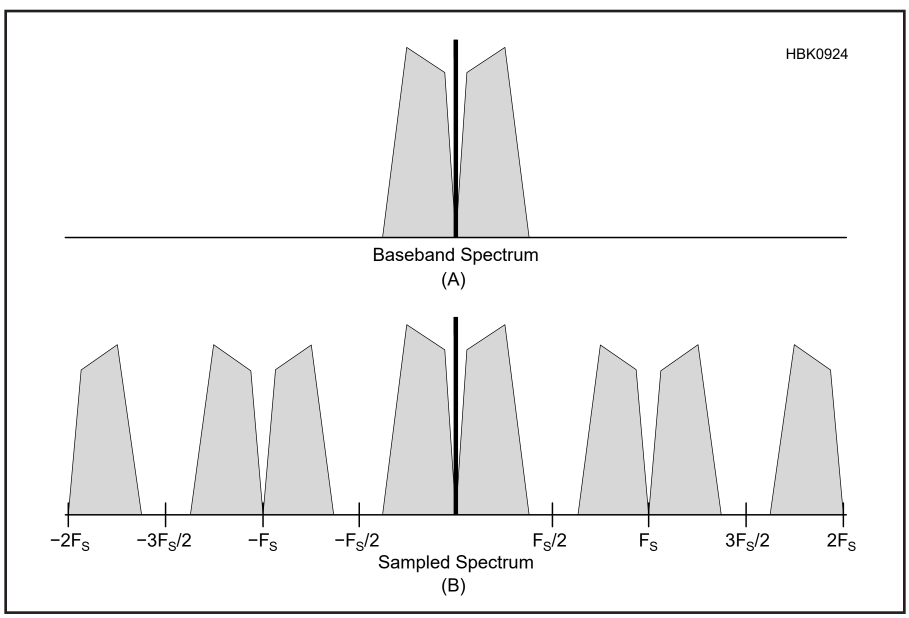


Figure 12.20 — The baseband spectrum of an analog signal before sampling is given at A, showing both positive and negative frequencies. Part B shows the spectrum of the baseband signal after it is sampled by an ideal sampling waveform. Note that there is an upper and lower sideband version of the baseband for each harmonic of the sampling frequency, just as if it were a double sideband suppressed carrier signal.)

is no longer a worry, as long as the anti-aliasing filter is doing its job. (As well as any preselector filters that may be present.) The dynamic range of an SDR theoretically does not depend on signal spacing — close-in dynamic range is often better than with a conventional architecture that uses a wide IF filter. With no crystal filters in the signal chain, the entire system has a completely linear phase response, which can improve the quality of both analog and digital signals after demodulation.

The biggest challenge with RF sampling is what to do with the torrent of high-speed data coming out of the receiver's ADC. To cover 0-54 MHz without aliasing requires a sample rate of at least 120 or 130 MHz, and commercial products typically operate the ADC at sample rates well over 200 MHz. That is much faster than a typical microprocessor or programmable DSP can handle. The local oscillator, mixer, and decimator or interpolator must be implemented in digital hardware so that the DSP can send and receive data at a more-reasonable sample rate. *Digital down-converters* (DDC) perform those functions and output a lower-sample-rate digital I/Q signal to the DSP. Stand-alone DDC ICs were available in the past, but the function is now usually integrated with the A/D converter. It is also possible to implement a DDC in a *field-programmable gate array* or FPGA. (See the "Hands-On SDR" *QEX* columns provided in the online content.) *Digital upconverters* (DUC) do the same conversion in reverse for the transmitter and are available integrated with the D/A converter or can be implemented in an FPGA. Some commercial integrated DDC/DAC products even include the capability to encode several digital modulation formats such as GMSK, QPSK and π/4 DQPSK. In an attempt to simplify the interface to the digital domain, many high-speed converters now use a standardized serial interface specification called JESD204B, capable of handling up to 12 Gb/s. Code to implement this interface on the digital FPGA is readily available.

Some designers have been successful in repurposing a graphics processor (GPU) for this purpose, and some GPU manufacturers now offer FFT libraries to assist in the design process.

12.3.3 Sample-Rate Down-Conversion

SDR receivers generally do not use analog mixers to convert RF frequencies to an IF or to baseband. Down-conversion is achieved by the process of *decimation* as described in the **DSP and SDR Fundamentals** chapter. It describes how the frequency of a sampled signal can be divided by N by removing every Nth sample from the digital signal data. This actually creates an alias of the original signal as if it had been sampled at a rate lower than the Nyquist rate. As shown in **Figure 12.20**, this process creates a replica of the original signal or original spectrum at harmonics of the sampling rate. One of the replica spectra can then be selected with a filter and operated on just as if it had been processed by an analog mixer.

This is one example *multirate signal processing* which is not nearly as complicated as it sounds. The topic we are interested in now is sample-rate down conversion (decimation). (*Sample-rate up conversion* is another multirate signal processing technique that performs frequency multiplication in much the same way as down-conversion performs frequency division.)

If we want to receive directly at the 40 meter band for instance, we would need to sample above 14.6 MHz in order to satisfy the Nyquist criterion. We do not need a sample rate that fast to actually manipulate the

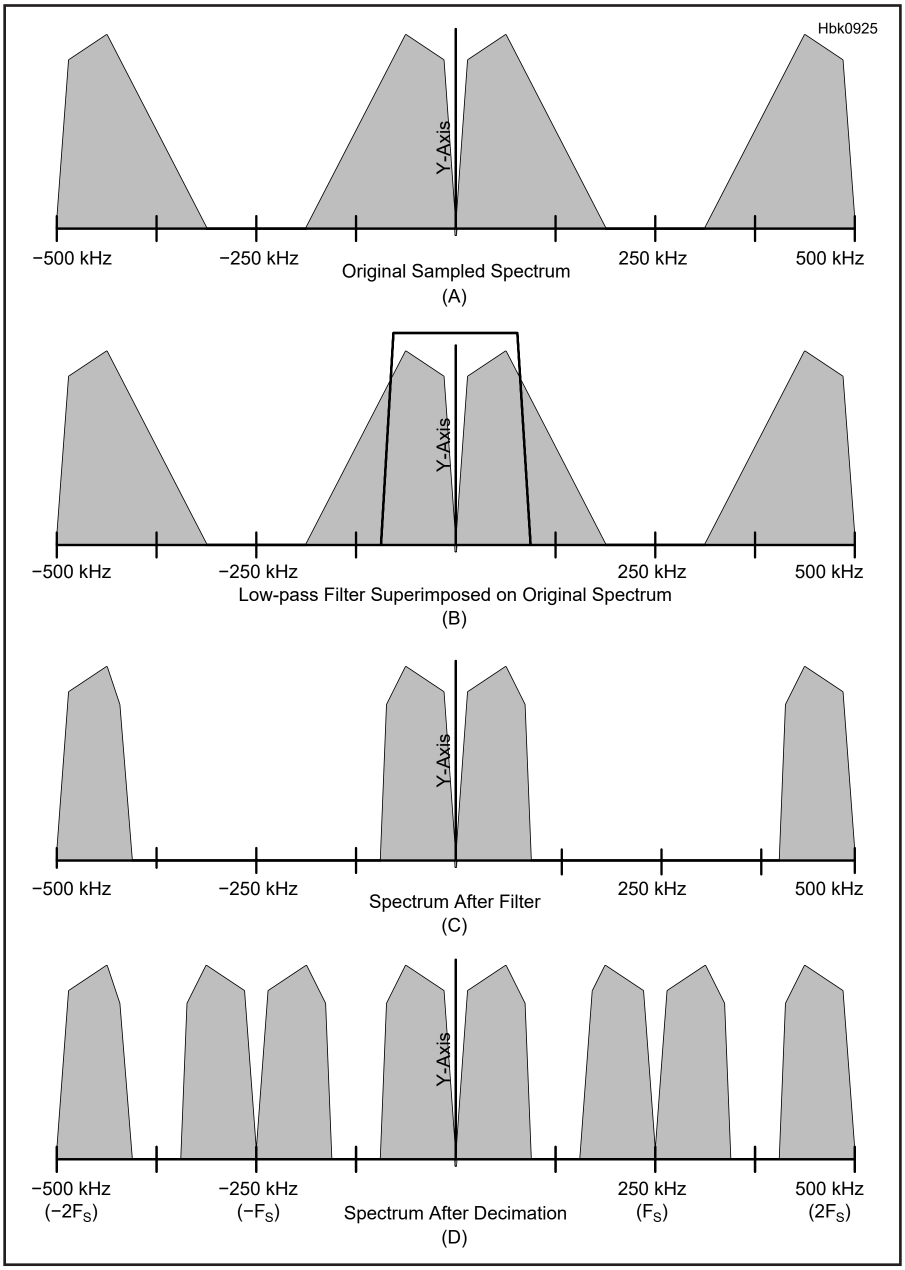


Figure 12.21 — Part A shows the spectrum of the sampled signal. Note that the frequency only extends from –500 kHz to +500 kHz since that is the extent that the math will manage. Part B shows the digital low-pass filter response superimposed on the sampled spectrum. The spectrum of the filtered signal, which is still sampled at 500 kHz, is shown at C. Part D shows the spectrum of the resulting sampled waveform after decimation by 2 (every other sample discarded). This is the spectrum of the signal when it is sampled at 250 kHz.)

information on receive, though, since the bandwidth of the widest signal will be on the order of 7 kHz or less. Even if we want to look at the entire band and generate a spectrum display, we only need about 650 kHz for the sample rate since the band is only 300 kHz wide. Any higher sample rate on receive is a waste of processor resources.

There are two main reasons why we would want to match the sample rate closely to our intended bandwidth. The first reason to lower the sample rate is that the transition band and ripple of our filters are dependent on the ratio of the filter length (the number of taps) to the sample rate. The second reason is to allow more CPU cycles for processing each sample.

An AM radio is a good example for sample rate conversion. Let's say we want to receive an band of signals through a tuned-circuit input filter with a bandwidth of 80 kHz centered at 590 kHz. If we undersample the input at a rate of 100 kHz, we create a range of signals centered at 90 kHz so that the energy is all within the range of dc to 125 kHz. This is *low-pass filter decimating*. (See the *QEX* SDR: Simplified column for July/August 2009 for a more complete discussion of this process.)

The first step in the decimation process is to filter the input signal so that there is no energy above one fourth the sample frequency. A DSP low-pass filter with cutoff at one fourth of the sample frequency has a very easy set of coefficients and can be implemented with a small number of taps. We sampled our AM radio at 500 kHz, so we need 125 kHz cutoff for the low pass filter. We now have a signal that has no energy above 125 kHz and is sampled at 500 kHz. We can throw away every other sample at this point to create a signal that is sampled at 250 kHz, with energy up to 125 kHz. The Nyquist criterion is satisfied with this new signal. **Figure 12.21** shows the spectrum of the process.

The signal at 90 kHz is only about 40 kHz wide (70 kHz to 110 kHz), so it would be nice if we could drop the sample frequency even further. It turns out that our signal will fit into the band from dc to 62.5 kHz if it were translated down in frequency. That would only require a 125 kHz sample rate. This is *integer band-pass decimating*. If we throw away every 3 samples of the original data, we can accomplish both frequency translation and sample rate reduction. The energy now spans from 15 kHz to 55 kHz. The signal spectrum is also inverted. In a superhet AM radio, you would need to band-pass filtered the signal before sampling. It is possible to do the band pass filtering digitally with a small number of taps, and then do the decimation.

We succeeded in reducing the sample rate by a factor of 4. We cannot further reduce the sample rate using straight decimation. Each integer sample rate reduction requires that the band limited data fit completely within one of the Nyquist regions for the new sample rates. The next integer sample rate would reduce the sample rate by 5. The new sample rate would be 100 kHz. The data is contained in both the k = 1 (50 kHz to 100 kHz) and k = 2 (100 kHz to 150 kHz) Nyquist bands. **Figure 12.22** shows the overlap. The overlap into the third Nyquist zone prevents further rate reduction. The requirement that all of the energy fits into one Nyquist zone limits the usefulness of this technique.

Integer sample rate reduction is a very useful tool because the filters are easily realized in hardware such as a field programmable gate array (FPGA) or other dedicated hardware.

12.3.4 Decimation and Dynamic Range

When we talk about dynamic range, we are discussing the range of largest to smallest signals that can represented simultaneously. What does dynamic range buy us? Having a high dynamic range allows us to represent or hear weak signals in the presence of large signals. With a lower dynamic range, our receiver may overload in the presence of a large signal or we may be forced to shift what dynamic range we have to accommodate the large signals. For example, if we only had 80 dB of dynamic range in our receiver with the weakest signal possible of –120 dBm and the largest of –40 dBm, this means the strongest signal we can hear is S9+33 dB (–50 dBm). If an S9+50 dB (–23 dBm) signal is present in our receiver, we would overload.

To allow the reception of large and small signal, receivers can implement automatic gain control (AGC) in the RF section. The AGC will add attenuation to the receiver, shifting the available dynamic range up. For example, it the AGC adds 20 dB of attenua-

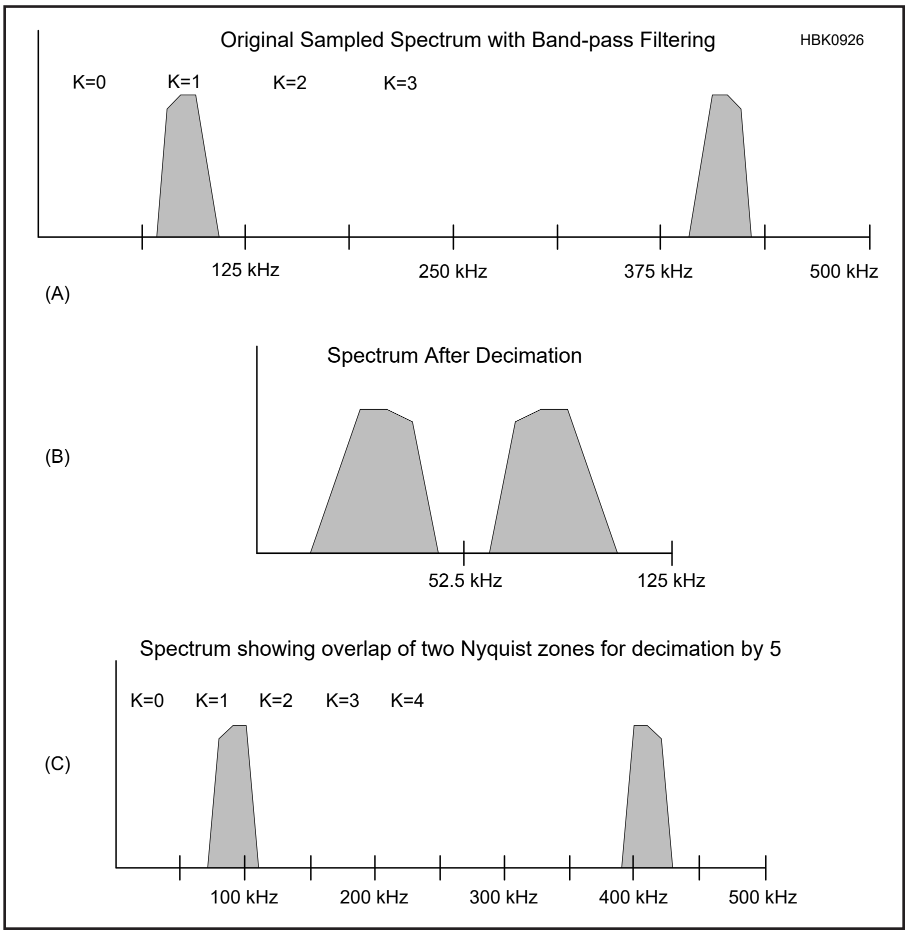


Figure 12.22 — Part A shows the spectrum of a 90 kHz signal that was band-pass filtered and then sampled at 500 kHz (only one-half of the spectrum is shown for clarity). The resulting spectrum when the signal is decimated by 4 (3 of every 4 samples discarded) is shown in Part B. Notice that this causes the signal to be aliased and the signal in the first Nyquist zone (k=0) has its frequencies inverted just as a lower sideband signal is inverted. The signal is now sampled at 125 kHz. The initial sampled signal, showing the Nyquist zones that would occur if a decimation by 5 were attempted, is shown at C. Note that input energy exists in both the second and third Nyquist zones.

tion, we can now hear signals from –100 dBm to –20 dBm. Note that we have just reduced the sensitivity of our receiver by 20 dB and if we were trying to listen to a weak CW signal with a power of –110 dBm, we will now be unable to hear it. RF AGC is necessary in some receivers where the instantaneous dynamic range is not sufficient to handle both very strong and very weak signals simultaneously. This should not be confused with audio-derived AGC which is discussed in the section on AGC below.

A common question about direct sampling receivers centers around available dynamic range. We are taught that for each bit in an ADC, we can have 6 dB of dynamic range. Often, someone will attempt to look at wideband sampled converter (ADC), multiply the bits by 6 dB and state that this is the available dynamic range for narrowband operation. As it turns out, this is an incorrect if not uncommon error. In a wideband sampled system where the receiver outputs are narrowband, we use a process called decimation to reduce the bandwidth for our narrowband receiver and increase dynamic range. How does this work?

Decimation is a two-step process: we first apply a digital filter to the samples and then we discard a portion of the samples. In decimation by two, we apply a filter to cut the RF bandwidth in half and then we discard one half of the samples. In doing so, the digital filter combines multiple samples to produce the output sample stream and in this combination, we pick up additional bits of resolution in the output samples. This increase in dynamic range is called *processing gain*. Specifically, in decimation by two we typically pickup about ½-bit of resolution for each decimation by two. Imagining for a moment a sampled system with 16-bits of resolution at 196.608 Msps, if we decimate down to 48 ksps, this represents twelve decimate-by-two operations. This means that we pick up about six bits per sample, increasing our dynamic range by about 36 dB. As a practical matter, the dynamic range of an ADC depends not only on the number of samples in the output, but also the level of the spurs. ADC manufacturers consider any level below the spurs to be tarnished by the spurs and therefore not included in the specified dynamic range. As a first order approximation for the dynamic range that can be achieved, we can take the dynamic range specified by the ADC manufacturer and add the processing gain that we've computed.

Finally, in amateur radio we generally talk about dynamic range in terms of a 500 Hz bandwidth. So for the purposes of dynamic range in the amateur world, we don't use the entire 48 ksps bandwidth, we look just at the 500 Hz bandwidth as if we had a 500 Hz filter. To determine the total processing gain we would achieve through decimation and filtering, we use can use the formula $3\log_2$(sampling rate/final bandwidth) to get an approximation of the processing gain from our original sampling rate to a 500 Hz bandwidth receiver. Using a logarithmic identity, it may be easier to compute on a calculator as $3\log_{10}$(sampling rate/final bandwidth)/$\log_{10}(2)$.

12.3.5 Phase Noise in Sampled Systems

Just like with mixers, analog to digital converters that sample signals and produce a digital output are also susceptible to phase noise. Many of today's older transceivers use superheterodyne receivers which use mixers and terminate in an ADC operating at baseband. For receiver systems where the ADC is operating in the tens to hundreds of ksps, phase noise introduced by the ADC is rarely an issue. Since good phase noise is easily achievable in this frequency range, most designers ensure that phase noise will not impact the design in the ADC. It's much more likely in such a radio that the LO feeding the mixers nearer the antenna would be at fault. With today's technology, the high we go in oscillator frequency, the more difficult it is to design a low phase noise oscillator.

In an RF sampled system where the ADC is placed either right on the antenna or just behind an amplifier and/or preselectors, we will be sampling at a significantly higher frequency. Because of the Nyquist sampling theorem, if we want to design an HF through 6 meter receiver, we must sample at least 2x the highest frequency, 54 MHz. This means we will be sampling above 108 MHz, with 122.88 MHz being a common frequency because it divides evenly down to standard audio sampling rates (see the sidebar on sampling rates). In these systems, the local oscillator may be called a clock oscillator and phase noise may be referred to as jitter. While phase noise and jitter are roughly interchangeable, jitter benefits the digital designer's view of the world and discusses a time-domain effect while phase noise benefits the RF designer and discusses a frequency domain effect. Most ADCs that will be used for RF will specify both. In case you're wondering, there is no absolute conversion from jitter directly to phase noise, but there are estimates that make assumptions about the characteristics of the oscillator.

Unlike the superheterodyne system, phase noise in a direct sampling receiver is imparted on the signal during the sampling process at the sampling frequency. What does this mean? While a superheterodyne system may rely on the division of the clock to produce an excellent phase noise signal on the band of interest, signals in a direct sampling system inherit the phase noise of the ADC clock at the sampling frequency. So if our phase noise is –110 dBc/Hz at 10 kHz for a 122.88 MHz clock, this is the phase noise we will have on all sampled frequencies. This makes clock (or LO) selection for the direct sampling receiver all the more important. We must ensure that the oscillator's phase noise characteristics will not be an issue at any frequency of interest.

More detailed discussions about the effects of phase noise in direct-sampled receivers can be found in Analog Devices application notes AN-741 and AN-756. (See the reference section.)

12.4 Mixing and Mixers

This section examines mixers which are used for frequency shifting or conversion in heterodyne receivers and transmitters. Mixers are often used as modulators and demodulators because they translate information to an RF signal and back again. These translation processes can be thought of as forms of frequency translation or frequency shifting — the function traditionally ascribed to mixers. We'll therefore begin our investigation by examining what a mixer is (and isn't), and what a mixer does.

MULTIPLYING VERSUS ADDING

Mixer is the term for a circuit that shifts one signal's frequency up or down by combining it with another signal. The word *mixer* is also the name of a device used to blend multiple audio inputs together for recording, broadcast or sound reinforcement. A radio mixer makes new frequencies from its input signals and an audio mixer does not. In their most basic, ideal forms, both devices have two inputs and one output.

The audio mixer is a *combiner* that simply *adds* the instantaneous voltages of the two signals together to produce the output at each point in time (**Figure 12.23**). The radio mixer, on the other hand, *multiplies* the instantaneous voltages of the two signals together to produce its output signal from instant to instant (**Figure 12.24**). Comparing the output spectra of the combiner and mixer, we see that the com-

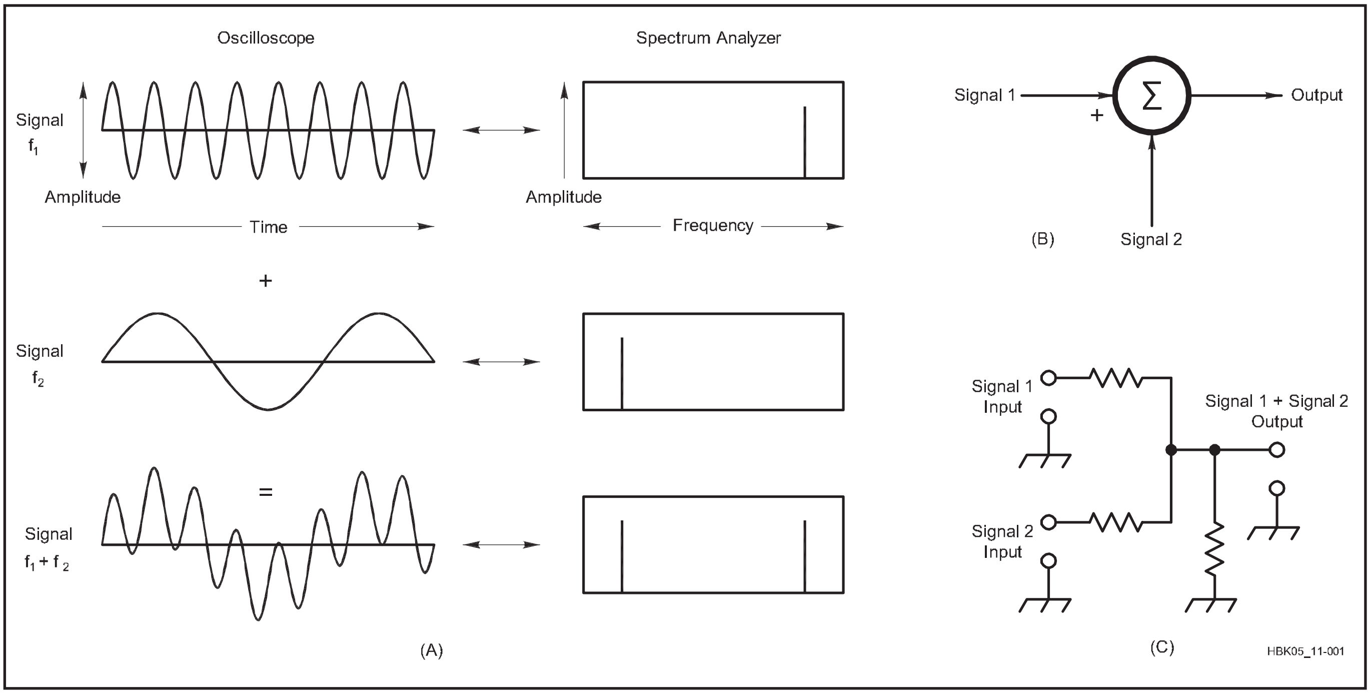


Figure 12.23 — *Adding or summing* two sine waves of different frequencies (f_1 and f_2) combines their amplitudes without affecting their frequencies. Viewed with an *oscilloscope* (a real-time graph of amplitude versus time), adding two signals appears as a simple superimposition of one signal on the other. Viewed with a spectrum analyzer (a real-time graph of signal amplitude versus frequency), adding two signals just sums their spectra. The signals merely coexist on a single cable or wire. All frequencies that go into the adder come out of the adder, and no new signals are generated. Drawing B, a block diagram of a summing circuit, emphasizes the stage's mathematical operation rather than showing circuit components. Drawing C shows a simple summing circuit, such as might be used to combine signals from two microphones. In audio work, a circuit like this is often called a mixer — but it does not perform the same function as an RF mixer.

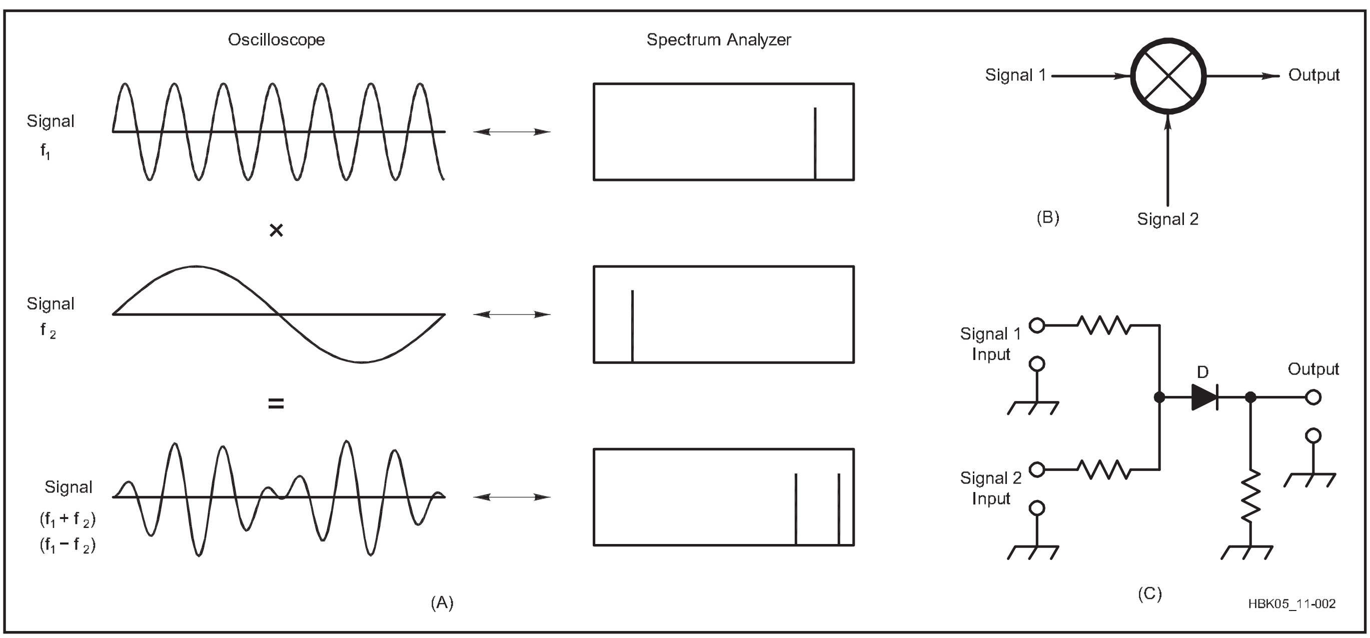


Figure 12.24 — *Multiplying* two sine waves of different frequencies produces a new output spectrum. Viewed with an oscilloscope, the result of multiplying two signals is a composite wave that seems to have little in common with its components. A spectrum-analyzer view of the same wave reveals why: The original signals disappear entirely and are replaced by two new signals — at the *sum* and *difference* of the original signals' frequencies. Drawing B diagrams a multiplier, known in radio work as a mixer. The circled X emphasizes the stage's mathematical operation. (The circled X is only one of several symbols you may see used to represent mixers in block diagrams, as Figure 12.25 explains.) Drawing C shows a very simple multiplier circuit. The diode, D, does the mixing. Because this circuit does other mathematical functions and adds them to the sum and difference products, its output is more complex than $f_1 + f_2$ and $f_1 – f_2$, but these can be extracted from the output by filtering.

Mixer Math: Mixing as Multiplication

Since a mixer works by means of multiplication, a bit of math can show us how they work. To begin with, we need to represent the two signals we'll mix, A and B, mathematically. Signal A's instantaneous amplitude equals

$$A_a \sin 2\pi f_a t$$

in which A is peak amplitude, f is frequency, and t is time. Likewise, B's instantaneous amplitude equals

$$A_b = A \sin(2\pi f_b t)$$

Since our goal is to show that multiplying two signals generates sum and difference frequencies, we can simplify these signal definitions by assuming that the peak amplitude of each is 1. The equation for Signal A then becomes

$$a(t) = A \sin(2\pi f_a t)$$

and the equation for Signal B becomes

$$b(t) = B \sin(2\pi f_b t)$$

Each of these equations represents a sine wave and includes a subscript letter to help us keep track of where the signals go.

Merely combining Signal A and Signal B by letting them travel on the same wire develops nothing new:

$$a(t) + b(t) = A \sin(2\pi f_a t) + B \sin(2\pi f_b t)$$

As simple as that equation may seem, we include it to highlight the fact that multiplying two signals is a quite different story. From trigonometry, we know that multiplying the sines of two variables can be expanded according to the relationship

$$\sin x \sin y = \frac{1}{2}\left[\cos(x - y) - \cos(x + y)\right]$$

Conveniently, Signals A and B are both sinusoidal, so we can use equation 6 to determine what happens when we multiply Signal A by Signal B. In our case, $x = 2\pi f_a t$ and $y = 2\pi f_b t$, so plugging them into equation 6 gives us

$$a(t) \times b(t) = \frac{AB}{2}\cos\left(2\pi\left[f_a - f_b\right]t\right) - \frac{AB}{2}\cos\left(2\pi\left[f_a + f_b\right]t\right)$$

Now we see two momentous results: a sine wave at the frequency *difference* between Signal A and Signal B $2\pi(f_a - f_b)t$, and a sine wave at the frequency *sum* of Signal A and Signal B $2\pi(f_a + f_b)t$. (The products are cosine waves, but since equivalent sine and cosine waves differ only by a phase shift of 90°, both are called *sine waves* by convention.)

This is the basic process by which we translate information into radio form and translate it back again. If we want to transmit a 1-kHz audio tone by radio, we can feed it into one of our mixer's inputs and feed an RF signal — say, 5995 kHz — into the mixer's other input. The result is two radio signals: one at 5994 kHz (5995 – 1) and another at 5996 kHz (5995 + 1). We have performed modulation.

Converting these two radio signals back to audio is just as straightforward. All we do is feed them into one input of another mixer, and feed a 5995-kHz signal into the mixer's other input. Result: a 1-kHz tone. We have performed demodulation.

biner's output contains only the frequencies of the two inputs, and nothing else, while the mixer's output contains *new* frequencies. The process is called *heterodyning* as used in the heterodyne receivers described in the preceding sections. The sidebar, "Mixer Math: Mixing as Multiplication," describes this process mathematically. Use of the word "mixer" in this book should be assumed to mean the radio mixer.

The key principle of a radio mixer is that in mixing multiple signal voltages together, *it adds and subtracts their frequencies to produce new frequencies.* (In the field of signal processing, this process, *multiplication in the time domain,* is recognized as equivalent to the process of *convolution in the frequency domain.* Those interested in this alternative approach to describing the generation of new frequencies through mixing can find more information about it in the many textbooks available on this subject.)

The difference between the mixer we've been describing and any mixer, modulator or demodulator that you'll ever use is that it's ideal. We put in two signals and got just two signals out. *Real* mixers, modulators and demodulators, on the other hand, also produce *distortion* products that make their output spectra "dirtier" or "less clean," as well as putting out some energy at input-signal frequencies and their harmonics. Much of the art and science of making good use of multiplication in mixing, modulation and demodulation goes into minimizing these unwanted multiplication products (or their effects) and making multipliers perform frequency translation as efficiently as possible.

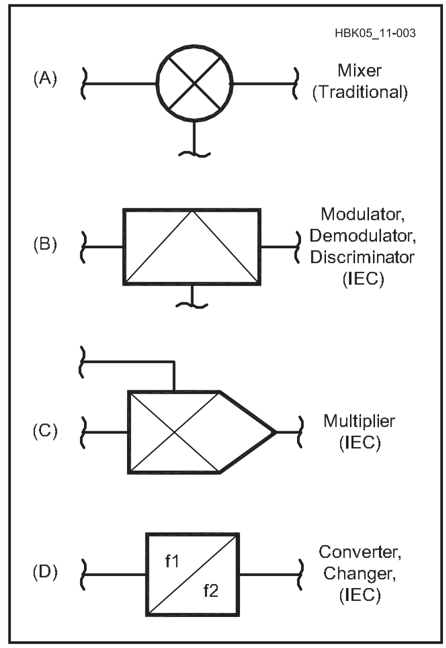


Figure 12.25 — The traditional symbol for a mixer is a circled X (A) although current standards allocate this symbol to a lamp. Current practice is to use one of the three IEC symbols shown at B, C, or D. For the frequency converter or changer symbol at D, a third connection can be included for the local oscillator. (IEC stands for *International Electrotechnical Commission* and the symbols are published in the IEEE 315A standard.)

12.4.1 Mixers and Distortion

This radio-amateur-oriented discussion of mixers, modulators and demodulators will begin with a look at their common underlying mechanism before discussing practical mixer, modulator and demodulator circuits. This will make it easier to understand the functions of those circuits. **Figure 12.25** shows the block symbol for a traditional mixer along with several IEC symbols for other functions mixers may perform.

NONLINEAR DISTORTION

The mechanism underlying multiplication, mixing, modulation and demodulation is a pretty straightforward thing: Any circuit structure that *nonlinearly distorts* ac waveforms acts as a multiplier to some degree.

The phrase *nonlinear distortion* sounds redundant, but isn't. Distortion, an externally imposed change in a waveform, can be linear; that is, it can occur independently of signal amplitude. Consider a radio receiver front-end filter that passes only signals between 6 and 8 MHz. It does this by *linearly distorting* the single complex waveform corresponding to the wide RF spectrum present at the radio's antenna terminals, reducing the amplitudes of frequency components below 6 MHz and above 8 MHz relative to those between 6 and 8 MHz. (Considering multiple signals on a wire as one complex waveform is just as valid, and sometimes handier, than considering them as separate signals. In this case, it's a bit easier to think of distortion as something that happens to a waveform rather than something that happens to separate signals relative to each other. It would be just as valid — and certainly more in keeping with the consensus view — to say merely that the filter attenuates signals at frequencies below 6 MHz and above 8 MHz.) The filter's output waveform certainly differs from its input waveform; the waveform has been distorted. But because this distortion occurs independently of signal level or polarity, the distortion is linear. No new frequency components are created; only the amplitude relationships among the wave's existing frequency components are altered. This is *amplitude* or *frequency* distortion, and all filters do it or they wouldn't be filters.

Phase or *delay distortion*, also linear, causes a complex signal's various component frequencies to be delayed by different amounts of time, depending on their frequency but independently of their amplitude. No new frequency components occur, and amplitude relationships among existing frequency components are not altered. Phase distortion occurs to some degree in all real filters.

The waveform of a non-sinusoidal signal can be changed by passing it through a circuit that has only linear distortion, but only *nonlinear distortion* can change the waveform of a simple sine wave. It can also produce an output signal whose output waveform changes as a function of the input amplitude, something not possible with linear distortion. Nonlinear circuits often distort excessively with overly strong signals, but the distortion can be a complex function of the input level.

Nonlinear distortion may take the form of *harmonic distortion*, in which integer multiples of input frequencies occur, or intermodulation distortion (IMD), in which different components multiply to make new ones, as described in previous sections.

Any departure from absolute linearity results in some form of nonlinear distortion, and this distortion can work for us or against us. Any amplifier, including a so-called linear amplifier, distorts nonlinearly to some degree; any device or circuit that distorts nonlinearly can work as a mixer, modulator, demodulator or frequency multiplier. An amplifier optimized for linear operation will nonetheless mix, but inefficiently; an amplifier biased for nonlinear amplification may be practically linear over a given tiny portion of its input-signal range. The trick is to use careful design and component selection to maximize nonlinear distortion when we want it (as in a mixer), and minimize it when we don't. Once we've decided to maximize nonlinear distortion, the trick is to minimize the distortion products we don't want, and maximize the products we want.

MINIMIZING UNWANTED DISTORTION PRODUCTS

Ideally, a mixer multiplies the signal at one of its inputs by the signal at its other input, but does not multiply a signal at the same input by itself, or multiple signals at the same input by themselves or by each other. (Multiplying a signal by itself — squaring it — generates harmonic distortion [specifically, *second-harmonic* distortion] by adding the signal's frequency to itself. Simultaneously squaring two or more signals generates simultaneous harmonic and intermodulation distortion.)

Consider what happens when a mixer must handle signals at two different frequencies (f_1 and f_2) applied to its first input, and a signal at a third frequency (f_3) applied to its other input. Ideally, a mixer multiplies f_1 by f_3 and f_2 by f_3, but does not multiply f_1 and f_2 by each other. This produces output at the sum and difference of f_1 and f_3, and the sum and difference of f_2 and f_3, but *not* the sum and difference of f_1 and f_2. **Figure 12.26** shows that feeding two signals into one input of a mixer results in the same output as if f_1 and f_2 are each first mixed with f_3 in two separate mixers, and the outputs of these mixers are combined. This shows that a mixer, even though constructed with nonlinearly distorting components, actually behaves as a *linear frequency shifter.* Traditionally, we refer to this process as mixing and to its outputs as *mixing products*, but we may also call it *frequency conversion*, referring to a device or circuit that does it as a *converter*, and to its outputs as *conversion products*. If a mixer produces an output frequency that is higher than the input frequency, it is called an up-converter; if the output frequency is lower than the input, a down-converter.

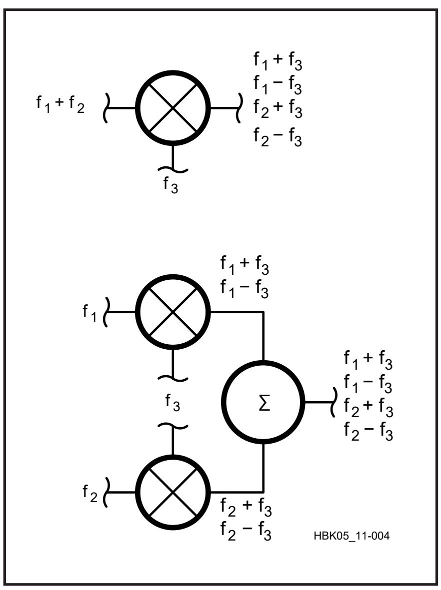


Figure 12.26 — Feeding two signals into one input of a mixer results in the same output as if f_1 and f_2 are each first mixed with f_3 in two separate mixers, and the outputs of these mixers are combined.

Real mixers, however, at best act only as reasonably linear frequency shifters, generating some unwanted IMD products — spurious signals, or *spurs* — as they go. Receivers are especially sensitive to unwanted mixer IMD because the signal-level range over which they must operate without generating unwanted IMD is often 90 dB or more, and includes infinitesimally weak signals. In a receiver, IMD products so weak that you'd never notice them in a transmitted signal can easily obliterate weak signals. This is why receiver designers apply so much effort to achieving "high dynamic range."

The degree to which a given mixer, modulator or demodulator circuit produces unwanted IMD is often the reason why we use it, or don't use it, instead of another circuit that does its wanted-IMD job as well or even better.

MISCELLANEOUS MIXING PRODUCTS

In addition to desired sum-and-difference products and unwanted IMD products, real mixers also put out some energy at their input frequencies. Some mixer implementations may *suppress* these outputs — that is, reduce one or both of their input signals by a factor of 100 to 1,000,000, or 20 to 60 dB. This is good because it helps keep input signals at the desired mixer-output sum or difference frequency from showing up at the IF terminal — an effect reflected in a receiver's *IF rejection* specification. Some mixer types, especially those used in the vacuum-tube era, suppress their input-signal outputs very little or not at all.

Input-signal suppression is part of an overall picture called *port-to-port isolation*. Mixer input and output connections are traditionally called *ports*. By tradition, the port to which we apply the shifting signal is the *local-oscillator (LO)* port. By convention, the signal or signals to be frequency-shifted are applied to the *RF (radio frequency)* port, and the frequency-shifted (product) signal or signals emerge at the *IF (intermediate frequency)* port. This illustrates the function of a mixer in a heterodyne receiver: Since it is often impractical to achieve the desired gain and filtering at the incoming signal's frequency (at RF), a mixer is used to translate the incoming RF signal to an intermediate frequency (the IF), where gain and filtering can be applied. The IF maybe be either lower or higher than the incoming RF signal. In a transmitter, the modulated signal may be created at an IF, and then translated in frequency by a mixer to the operating frequency.

A mixer may be used in an SDR to convert a range of signals into the range the SDR can process. This is very common when using SDR equipment designed for VHF and higher frequencies to receive HF signals. In that case an up-converting mixer is required.

Some mixers are *bilateral*; that is, their RF and IF ports can be interchanged, depending on the application. Diode-based mixers are usually bilateral. Many mixers are not bilateral (*unilateral*); the popular SA602/612 Gilbert cell IC mixer is an example of this.

It's generally a good idea to keep a mixer's input signals from appearing at its output port because they represent energy that we'd rather not pass to subsequent circuitry. It therefore follows that it's usually a good idea to keep a mixer's LO-port energy from appearing at its RF port, or its RF-port energy from making it through to the IF port. But there are some notable exceptions.

12.4.2 Switching Mixers

Depending on the application, mixers may vary from the extremely simple to the complex. For example, a simple half-wave rectifier (a signal diode, such as a 1N34 [germanium] or a 1N914 [silicon]) can do the job. This is an example of a *switching mixer*, in which mixing occurs as one signal — in this case, the carrier, which in effect turns the diode on and off as its polarity reverses — interrupts the transmission of another (in demodulation of full-carrier AM, the sidebands).

A switch can be thought of as an amplifier toggled between two gain states, off and on, by a control signal. It turns out that a binary amplifier is not necessary; any device that can be gain-varied in accordance with the amplitude of a control signal can serve as a mixer.

Most modern radio mixers act more like fast analog switches than analog multipliers. In using a mixer as a fast switching device, we apply a square wave to its LO input with a square wave rather than a sine wave, and feed sine waves, audio, or other complex signals to the mixer's RF input. The RF port serves as the mixer's "linear" input, and therefore must preferably exhibit low intermodulation and harmonic distortion. Feeding a ±1-V square wave into the LO input alternately multiplies the linear input by +1 or –1.

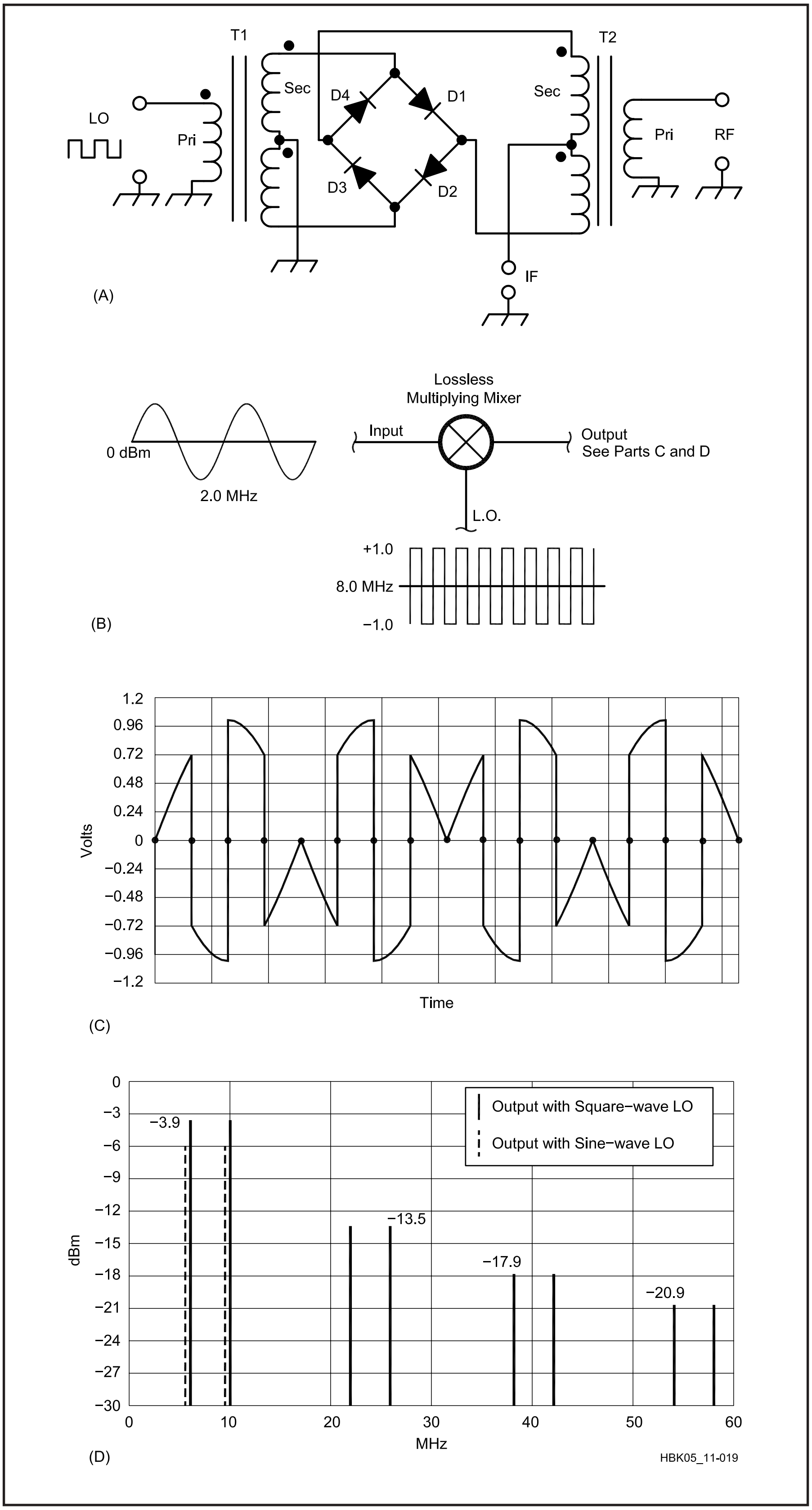


Figure 12.27 — Part (A) shows a general-purpose diode *reversing-switch* mixer. This mixer uses a square-wave LO and a sine-wave input signal. The text describes its action. Part (B) is an ideal *multiplier* mixer. The square-wave LO and a sine-wave input signal produce the output waveform shown in part (C). The solid lines of part (D) show the output spectrum with the square-wave LO. The dashed lines show the output spectrum with a sine-wave LO.

Multiplying the RF-port signal by +1 just transfers it to the output with no change. Multiplying the RF-port signal by –1 does the same thing, except that the signal inverts (flips 180° in phase). The LO port need not exhibit low intermodulation and harmonic distortion; all it has to do is preserve the fast rise and fall times of the switching signal.

REVERSING-SWITCH MIXERS

We can multiply a signal by a square wave without using an analog multiplier at all. All we need is a pair of balun transformers and four diodes (**Figure 12.27A**).

With no LO energy applied to the circuit, none of its diodes conduct. RF-port energy (1) can't make it to the LO port because there's no direct connection between the secondaries of T1 and T2, and (2) doesn't produce IF output because T2's secondary balance results in energy cancellation at its center tap, and because no complete IF-energy circuit exists through T2's secondary with both of its ends disconnected from ground.

Applying a square wave to the LO port biases the diodes so that, 50% of the time, D1 and D2 are on and D3 and D4 are reverse-biased off. This unbalances T2's secondary by leaving its upper wire floating and connecting its lower wire to ground through T1's secondary and center tap. With T2's secondary unbalanced, RF-port energy emerges from the IF port.

The other 50% of the time, D3 and D4 are on and D1 and D2 are reverse-biased off. This unbalances T2's secondary by leaving its lower wire floating, and connects its upper wire to ground through T1's secondary and center tap. With T2's secondary unbalanced, RF-port energy again emerges from the IF port — shifted 180° relative to the first case because T2's active secondary wires are now, in effect, transposed relative to its primary.

A reversing switch mixer's output spectrum is the same as the output spectrum of a multiplier fed with a square wave. This can be analyzed by thinking of the square wave in terms of its Fourier series equivalent, which consists of the sum of sine waves at the square wave frequency and all of its odd harmonics. The amplitude of the equivalent series' fundamental sine wave is $4/\pi$ times (2.1 dB greater than) the amplitude of the square wave. The amplitude of each harmonic is inversely proportional to its harmonic number, so the third harmonic is only ⅓ as strong as the fundamental (9.5 dB below the fundamental), the 5th harmonic is only ⅕ as strong (14 dB below the fundamental) and so on. The input signal mixes with each harmonic separately from the others, as if each harmonic were driving its own separate mixer, just as we illustrated with two sine waves in Figure 12.26. Normally, the harmonic outputs are so widely removed from the desired output frequency that they are easily filtered out, so a reversing-switch mixer is just as good as a sine-wave-driven analog multiplier for most practical purposes, and usually better — for radio purposes — in terms of dynamic range and noise.

An additional difference between multiplier and switching mixers is that the signal flow in a switching mixer is reversible (that is, bilateral). It really only has one dedicated input (the LO input). The other terminals can be thought of as I/O (input/output) ports, since either one can be the input as long as the other is the output.

CONVERSION LOSS IN SWITCHING MIXERS

Figure 12.27B shows a perfect *multiplier* mixer. That is, the output is the product of the input signal and the LO. The LO is a perfect square wave. Its peak amplitude is ±1.0 V and its frequency is 8 MHz. Figure 12.27C shows the output waveform (the product of two inputs) for an input signal whose value is 0 dBm and whose frequency is 2 MHz. Notice that for each transition of the square-wave LO, the sine-wave output waveform polarity reverses. There are 16 transitions during the interval shown, at each zero-crossing point of the output waveform. Figure 12.27D shows the mixer output spectrum. The principle components are at 6 MHz and 10 MHz, which are the sum and difference of the signal and LO frequencies. The amplitude of each of these is –3.9 dBm. Numerous other pairs of output frequencies occur that are also spaced 4 MHz apart and centered at 24 MHz, 40 MHz and 56 MHz and higher odd harmonics of 8 MHz. The ones shown are at –13.5 dBm, –17.9 dBm and –20.9 dBm. Because the mixer is lossless, the sum of all of the outputs must be exactly equal to the value of the input signal. As explained previously, this output spectrum can also be understood in terms of each of the odd-harmonic components of the square-wave LO operating independently.

If the mixer switched without losses, such as in Figure 12.27A, with diodes that are perfect switches, the results would be mathematically identical to the above example. The diodes would commutate the input signal exactly as shown in Figure 12.27C.

Now consider the perfect multiplier mixer of Figure 12.27B with an LO that is a perfect sine wave with a peak amplitude of ±1.0 V. In this case the dashed lines of Figure 12.27D show that only two output frequencies are present, at 6 MHz and 10 MHz (see also Figure 12.24). Each component now has a –6 dBm level. The product of the 0 dBm sine-wave input at one frequency and the ±1.0V sine-wave LO at another frequency (see equation 6 in this chapter) is the –3 dBm total output.

These examples illustrate the difference between the square-wave LO and the sine-wave LO, for a perfect multiplier. For the same peak value of both LO waves, the square-wave LO delivers 2.1 dB more output at 6 MHz and 10 MHz than the sine-wave LO. An actual diode mixer such as Figure 12.27A behaves more like a switching mixer. Its sine-wave LO waveform is considerably flattened by interaction between the diodes and the LO generator, so that it looks somewhat like a square wave. The diodes have nonlinearities, junction voltages, capacitances, resistances and imperfect parameter matching. (See the **RF Techniques** chapter.) Also, "re-mixing" of a diode mixer's output with the LO and the input is a complicated possibility. The practical end result is that diode double-balanced mixers have a conversion loss, from input to each of the two major output frequencies, in the neighborhood of 5 to 6 dB. (Conversion loss is discussed in a later section.)

12.4.3 The Diode Double-Balanced Mixer (DBM)

The diode *double-balanced mixer* (DBM) is standard in many commercial, military and amateur applications because of its excellent balance and high dynamic range. DBMs can serve as mixers (including image-reject types), modulators (including single- and double-sideband, phase, biphase, and quadrature-phase types) and demodulators, limiters, attenuators, switches, phase detectors and frequency doublers. In some of these applications, they work in conjunction with power dividers, combiners and hybrids.

THE BASIC DBM

We have already seen the basic diode DBM circuit (Figure 12.27A). In its simplest form, a DBM contains two or more unbalanced-to-balanced transformers and a Schottky-diode ring consisting of $4 \times n$ diodes, where n is the number of diodes in each leg of the ring. Each leg commonly consists of up to four diodes.

As we've seen, the degree to which a mixer is *balanced* depends on whether either, neither or both of its input signals (RF and LO) emerge from the IF port along with mixing products. An unbalanced mixer suppresses neither its RF nor its LO; both are present at its IF port. A single-balanced mixer suppresses its RF or LO, but not both. A double-balanced mixer suppresses its RF *and* LO inputs. Diode and transformer uniformity in the Figure 12.27 circuit results in equal LO potentials at the center taps of T1 and T2. The LO potential at T1's secondary center tap is zero (ground); therefore, the LO potential at the IF port is zero.

Balance in T2's secondary likewise results in an RF null at the IF port. The RF potential between the IF port and ground is therefore zero — except when the DBM's switching diodes operate!

The Figure 12.27 circuit normally also affords high RF-IF isolation because its balanced diode switching precludes direct connections between T1 and T2. A diode DBM can be used as a current-controlled switch or attenuator by applying dc to its IF port, albeit with some distortion. This causes opposing diodes (D2 and D4, for instance) to conduct to a degree that depends on the current magnitude, connecting T1 to T2.

TRIPLE-BALANCED MIXERS

The triple-balanced mixer shown in **Figure 12.28** (sometimes called a "double double-balanced mixer") is an extension of the single-diode-ring mixer. The diode rings are fed by power-splitting baluns at the RF and LO ports. An additional balun is added at the IF output. The circuit's primary advantage is that the IF output signal is balanced and isolated from the RF and LO ports over a large bandwidth — commercial mixer IF ranges of 0.5 to 10 GHz are typical.

It has higher signal-handling capability and dynamic range (a 1-dB compression point within 3 to 4 dB below LO signal levels) and lower intermodulation levels (by 10 dB or more) than a single-ring mixer. The triple-balanced mixer is used when a very wide IF range is required.

Adding the balancing transformer in the IF output path increases IF-to-LO and IF-to-RF isolation. This makes the conversion process much less sensitive to IF impedance mismatches. Since the IF port is isolated from the RF and LO ports, the three frequency ranges (RF, LO and IF) can overlap. A disadvantage of IF transformer coupling is that a dc (or low-frequency) IF output is not available, so the triple-balanced mixer cannot be used for direct-conversion receivers.

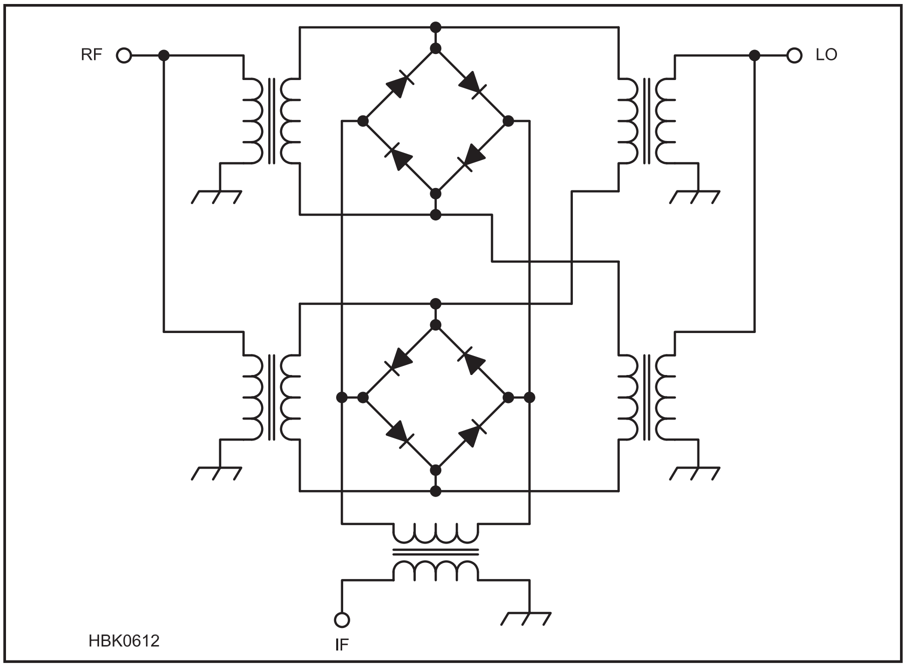


Figure 12.28 — The triple-balanced mixer uses a pair of diode rings and adds an additional balancing transformer to the IF port.

DIODE DBM COMPONENTS

Commercially manufactured diode DBMs generally consist of a supporting base, a diode ring, two or more ferrite-core transformers commonly wound with two or three twisted-pair wires, encapsulating material, an enclosure.

Diodes

Hot-carrier (Schottky) diodes are the devices of choice for diode-DBM rings because of their low ON resistance, although ham-built DBMs for non-critical MF/HF use commonly use switching diodes like the 1N914 or 1N4148. The forward voltage drop, V_f, across each diode in the ring determines the mixer's optimum local-oscillator drive level. Depending on the forward voltage drop of each of its diodes and the number of diodes in each ring leg, a diode DBM will often be specified by the optimum LO drive level in dBm (typical values are 0, 3, 7, 10, 13, 17, 23 or 27). As a rule of thumb, the LO signal must be 20 dB stronger than the RF and IF signals for proper operation. This ensures that the LO signal, rather than the RF or IF signals, switches the mixer's diodes on and off — a critical factor in minimizing IMD and maximizing dynamic range.

Transformers

From the DBM schematic shown in Figure 12.27, it's clear that the LO and RF transformers are unbalanced on the input side and balanced on the diode side. The diode ends of the balanced ports are 180° out of phase throughout the frequency range of interest. This property causes signal cancellations that result in higher port-to-port isolation. **Figure 12.29A** plots LO-RF and LO-IF isolation versus frequency for Synergy Microwave's CLP-403 DBM, which is specified for +7 dBm LO drive level. Isolations on the order of 70 dB occur at the lower end of the band as a direct result of the balance among the four diode-ring legs and the RF phasing of the balanced ports.

As we learned in our discussion of generic switching mixers, transformer efficiency plays an important role in determining a mixer's conversion loss and drive-level requirement. Core loss, copper loss and impedance mismatch all contribute to transformer losses. Ferrite in toroidal, bead, balun (multi-hole)or rod form can serve as DBM transformer cores. Radio amateurs commonly use Fair-Rite Mix 43 ferrite (μ = 950) in HF and VHF applications.

RF transformers combine lumped and distributed capacitance and inductance. The interwinding capacitance and characteristic impedance of a transformer's twisted wires sets the transformer's high-frequency response. The core's μ and size, and the number of winding turns, determine the transformer's lower frequency limit. Covering a specific frequency range requires a compromise in the number of turns used with a given core. Increasing a transformer's core size and number of turns improves its low-frequency response. Cores may be stacked to meet low-frequency performance specs.

Inexpensive mixers operating up to 2 GHz most commonly use twisted trifilar (three-wire) windings made of a wire size between #36 and #32. The number of twists per unit length of wire determines a winding's characteristic impedance. Twisted wires are analogous to transmission lines. The transmission-line effect predominates at the higher end of a transformer's frequency range.

PRACTICAL DIODE DBMS

Important DBM specifications include conversion loss and amplitude flatness across the required IF bandwidth; variation of conversion loss with input frequency; variation of conversion loss with LO drive, 1-dB compression point; LO-RF, LO-IF and RF-IF isolation; intermodulation products; noise figure (usually within 1 dB of conversion loss); port SWR; and dc offset, which is directly related to isolation among the RF, LO and IF ports. Most of these parameters also apply to other mixer types.

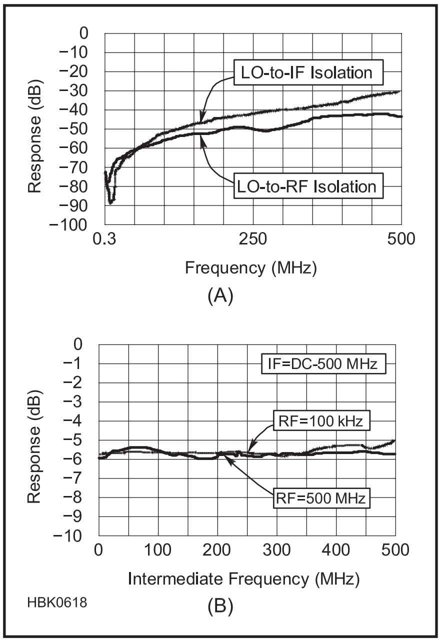


Figure 12.29 — The port-to-port isolation of a diode DBM depends on how well its diodes match and how well its transformers are balanced. (A) shows LO-IF and LO-RF isolation versus frequency and (B) shows conversion loss for a typical diode DBM, the Synergy Microwave CLP-403 mixer. In (B), LO driver level is +7 dBm.

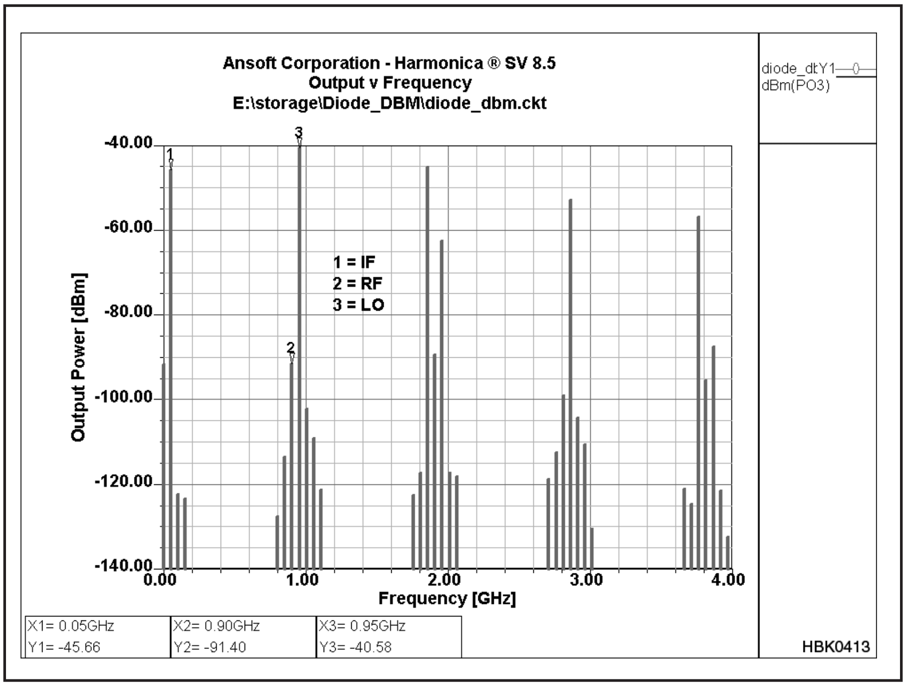


Figure 12.30 — Simulated diode-DBM output spectrum with four LO harmonics evaluated. Note that the desired output products (the highest two products, RF – LO and RF + LO) emerge at a level 5 to 6 dB below the mixer's RF input (–40 dBm). This indicates a mixer conversion loss of 5 to 6 dB. (*Serenade SV8.5* simulation.)

Conversion Loss of Diode DBMs

Figure 12.29B shows conversion loss versus intermediate frequency in a typical DBM. The curves show conversion loss for two fixed RF-port signals, one at 100 kHz and the another at 500 MHz, while varying the LO frequency from 100 kHz to 500 MHz.

Figure 12.30 graphs a diode DBM's simulated output spectrum. Note that the RF input (900 MHz) is –40 dBm and the desired IF output (51 MHz, the frequency difference between the RF and LO signals) is –46 dBm, implying a conversion loss of 6 dB. Very nearly the same value (5 dB) applies to the sum of both signals (RF + LO). We minimize a diode DBM's conversion loss, noise figure and intermodulation by keeping its LO drive high enough to switch its diodes on fully and rapidly. Increasing a mixer's LO level beyond that sufficient to turn its switching devices all the way on merely makes them dissipate more LO power without further improving performance.

Insufficient LO drive results in increased noise figure and conversion loss. IMD also increases because RF-port signals have a greater chance to control the mixer diodes when the LO level is too low.

APPLYING DIODE DBMS

At first glance, applying a diode DBM is easy: We feed the signal(s) we want to frequency-shift (at or below the maximum level called for in the mixer's specifications, such as –10 dBm for the Mini-Circuits SBL-1 and TUF-3, and Synergy Microwave S-1, popular 7 dBm LO power parts) to the DBM's RF port, feed the frequency-shifting signal (at the proper level) to the LO port, and extract the sum and difference products from the mixer's IF port.

There's more to it than that, however, because diode DBMs (along with most other modern mixer types) are *termination-sensitive*. That is, their ports — particularly their IF (output) ports — must be resistively terminated with the proper impedance (commonly 50 Ω, resistive). A wideband, resistive output termination is particularly critical if a mixer is to achieve its maximum dynamic range in receiving applications. Such a load can be achieved by:

• Terminating the mixer in a 50-Ω resistor or attenuator pad (a technique usually avoided in receiving applications because it directly degrades system noise figure);

• Terminating the mixer with a low-noise, high-dynamic-range *post-mixer amplifier* designed to exhibit a wideband resistive input impedance; or

• Terminating the mixer in a *diplexer*, a frequency-sensitive signal splitter that appears as a two-terminal resistive load at its input while resistively dissipating unwanted outputs and passing desired outputs through to subsequent circuitry.

Termination-insensitive mixers are available, but this label can be misleading. Some termination-insensitive mixers are nothing more than a termination-sensitive mixer packaged with an integral post-mixer amplifier. True termination-insensitive mixers are less common and considerably more elaborate. Amateur builders will more likely use one of the many excellent termination-sensitive mixers available in connection with a diplexer, post-mixer amplifier or both.

Figure 12.31 shows one diplexer implementation. In this approach, L1 and C1 form a series-tuned circuit, resonant at the desired IF, that presents low impedance between the diplexer's input and output terminals at the IF. The high-impedance parallel-tuned circuit formed by L2 and C2 also resonates at the desired IF, keeping desired energy out of the diplexer's 50-Ω load resistor, R1.

The preceding example is called a *band-pass diplexer*. **Figure 12.32** shows another type: a *high-pass/low-pass diplexer* in which each inductor and capacitor has a reactance of 70.7 Ω at the 3-dB cutoff frequency. It can be used after a "difference" mixer (a mixer in which the IF is the difference between the signal frequency and LO) if the desired IF and its image frequency are far enough apart so that the image power is "dumped" into the network's 51-Ω resistor. (For a "summing" mixer — a mixer in which the IF is the sum of the desired signal and LO — interchange the 50-Ω idler load resistor and the diplexer's "50-Ω Amplifier" connection.)

Figure 12.33 shows a BJT post-mixer

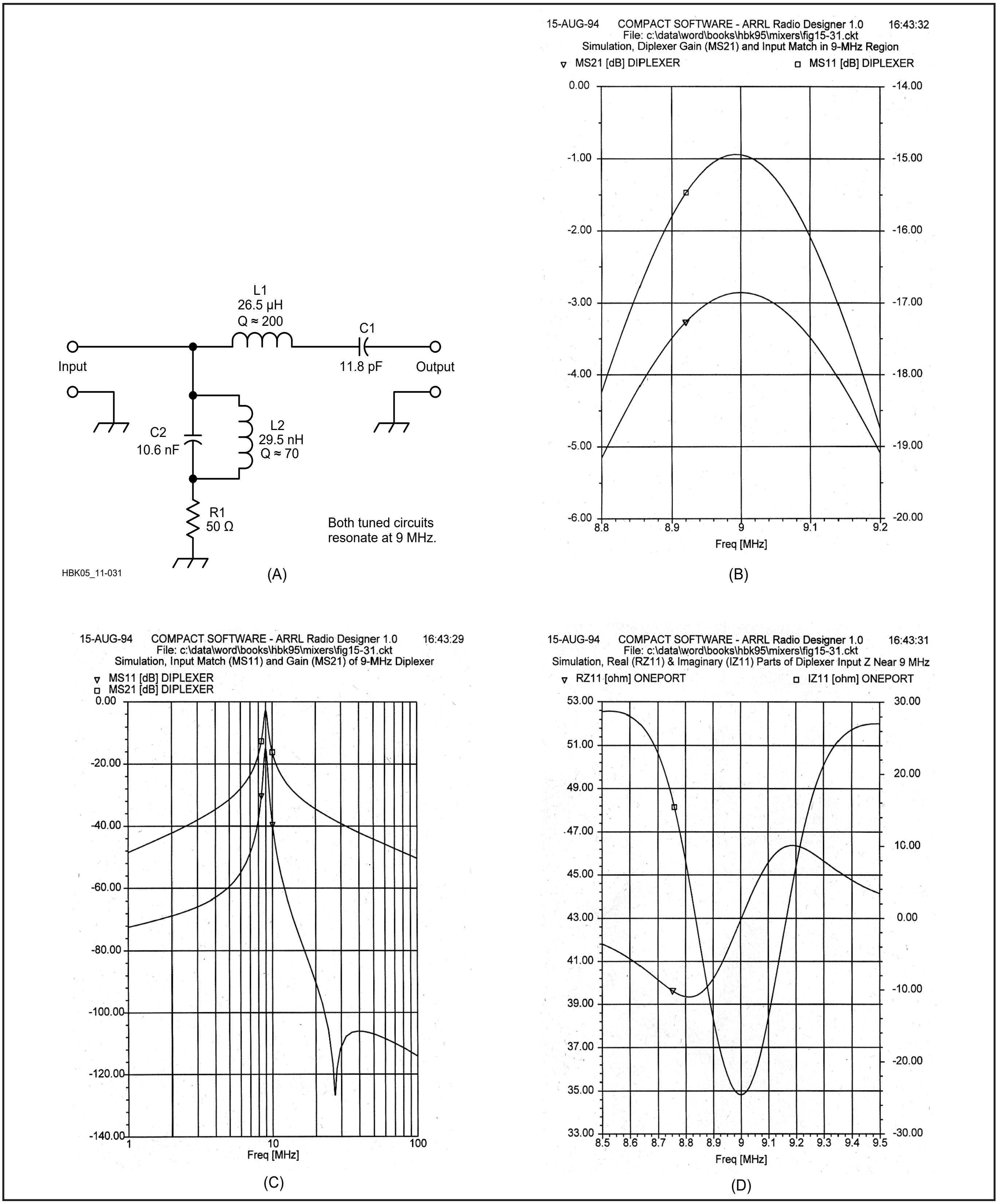


Figure 12.31 — A diplexer resistively terminates energy at unwanted frequencies while passing energy at desired frequencies. This band-pass diplexer (A) uses a series-tuned circuit as a selective pass element, while a high-C parallel-tuned circuit keeps the network's terminating resistor R1 from dissipating desired-frequency energy. Computer simulation of the diplexer's response with *ARRL Radio Designer 1.0* characterizes the diplexer's insertion loss and good input match from 8.8 to 9.2 MHz (B) and from 1 to 100 MHz (C); and the real and imaginary components of the diplexer's input impedance from 8.8 to 9.2 MHz with a 50-Ω load at the diplexer's output terminal (D). The high-C, low-L nature of the L2-C2 circuit requires that C2 be minimally inductive; a 10,000-pF chip capacitor is recommended. This diplexer was described by Rohde and Bucher in *Communications Receivers: Principles and Design, 3rd Edition*.

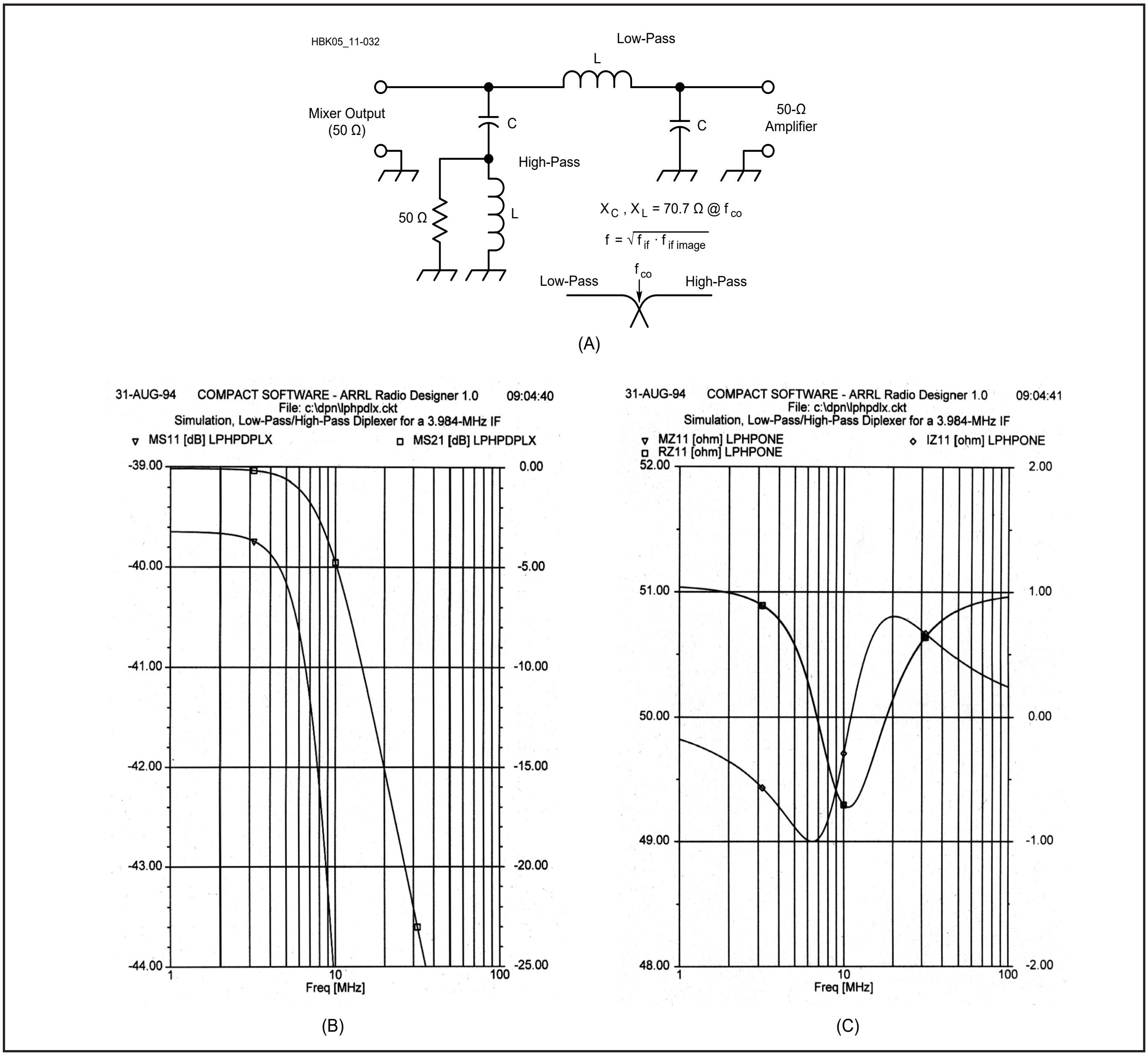


Figure 12.32 — All of the inductors and capacitors in this high-pass/low-pass diplexer (A) exhibit a reactance of 70.7 Ω at its tuned circuits' 3-dB cutoff frequency (the geometric mean of the IF and IF image). B and C show *ARRL Radio Designer* simulations of this circuit configured for use in a receiver that converts 7 MHz to 3.984 MHz using a 10.984-MHz LO. The IF image is at 17.984 MHz, giving a 3-dB cutoff frequency of 8.465 MHz. The inductor values used in the simulation were therefore 1.33 µH (Q = 200 at 25.2 MHz); the capacitors, 265 pF (Q = 1000). This drawing shows idler load and "50-Ω Amplifier" connections suitable for a receiver in which the IF image falls at a frequency *above* the desired IF. For applications in which the IF image falls below the desired IF, interchange the 50-Ω idler load resistor and the diplexer's "50-Ω Amplifier" connection so the idler load terminates the diplexer low-pass filter and the 50-Ω amplifier terminates the high-pass filter.

amplifier design made popular by Wes Hayward, W7ZOI, and John Lawson, K5IRK. RF feedback (via the 1-kΩ resistor) and emitter degeneration (the ac-coupled 5.6-Ω emitter resistor) work together to keep the stage's input impedance near 50 Ω and uniformly resistive across a wide bandwidth. Performance comparable to the Figure 10.28 circuit can be obtained at MF and HF by using paralleled 2N3904s as shown in **Figure 12.34**.

Phase Detection with a DBM

As we saw in our exploration of quadrature detection, applying two signals of equal frequency to a DBM's LO and RF ports produces an IF-port dc output proportional to the cosine of the signals' phase difference (**Figure 12.35**). This assumes that the DBM has a dc-coupled IF port, of course. If it doesn't — and some DBMs don't — phase-detector operation is out. Any dc output offset introduces error into this process, so critical phase-detection applications use low-offset DBMs optimized for this service.

12.4.4 Active Mixers

We've covered diode DBMs in depth because their ease of use in homebrew projects, high performance, and suitability for direct connection into 50-Ω systems makes

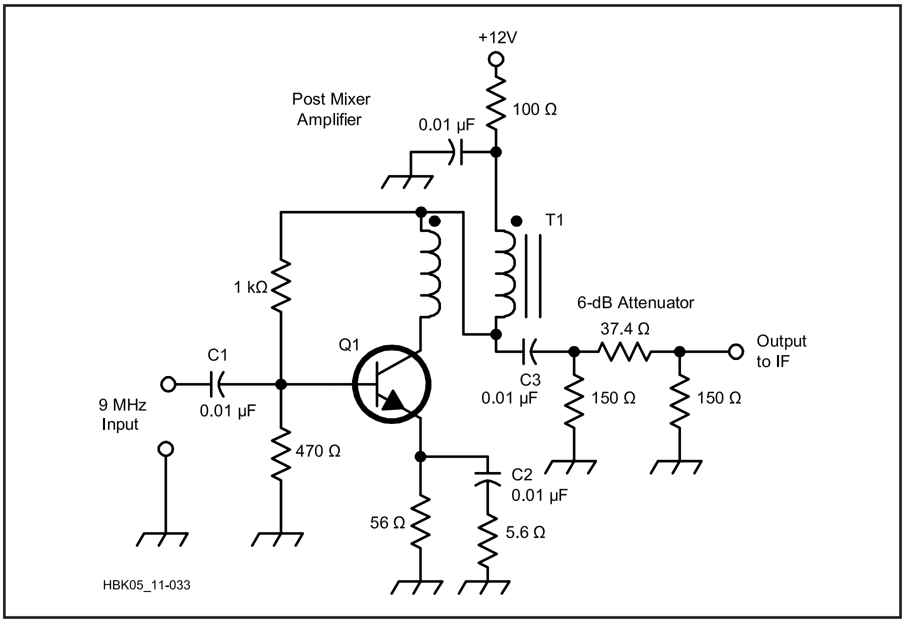


Figure 12.33 — The post-mixer amplifier from Hayward and Lawson's "Progressive Communications Receiver" (November 1981 *QST*). This amplifier's gain, including the 6-dB loss of the attenuator pad, is about 16 dB; its noise figure, 4 to 5 dB; its output intercept, 30 dBm. The 6-dB attenuator is essential if a crystal filter follows the amplifier; the pad isolates the amplifier from the filter's highly reactive input impedance. This circuit's input match to 50 Ω below 4 MHz can be improved by replacing 0.01-µF capacitors C1, C2 and C3 with low-inductance 0.1-µF units (chip capacitors are preferable). Q1 is a TO-39 CATV-type bipolar transistor, f_T = 1 GHz or greater (2N3866, 2N5109, 2SC1252, 2SC1365 or MRF586 suitable.) Use a small heat sink on this transistor. T1 is a broadband ferrite transformer, ≈42 µH per winding: 10 bifilar turns of #28 enameled wire on an FT 37-43 core.

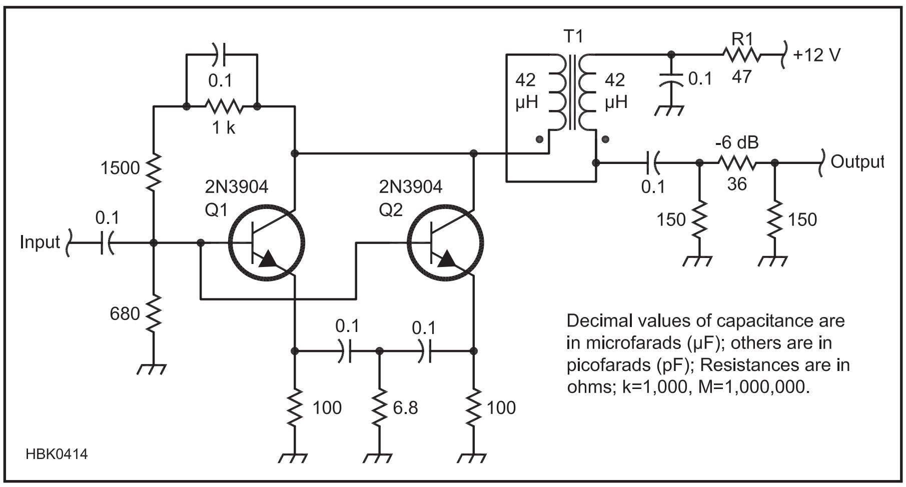


Figure 12.34 — At MF and HF, paralleled 2N3904 BJTs can provide performance comparable to that of the Figure 12.33 circuit with sufficient attention paid to device standing current, here set at »30 mA for the pair. The value of decoupling resistor R1 is critical in that small changes in its value cause a relatively large change in the 2N3904s' bias point. This circuit is part of the "EZ-90 Receiver," described by Hayward, Campbell and Larkin in *Experimental Methods in RF Design.*

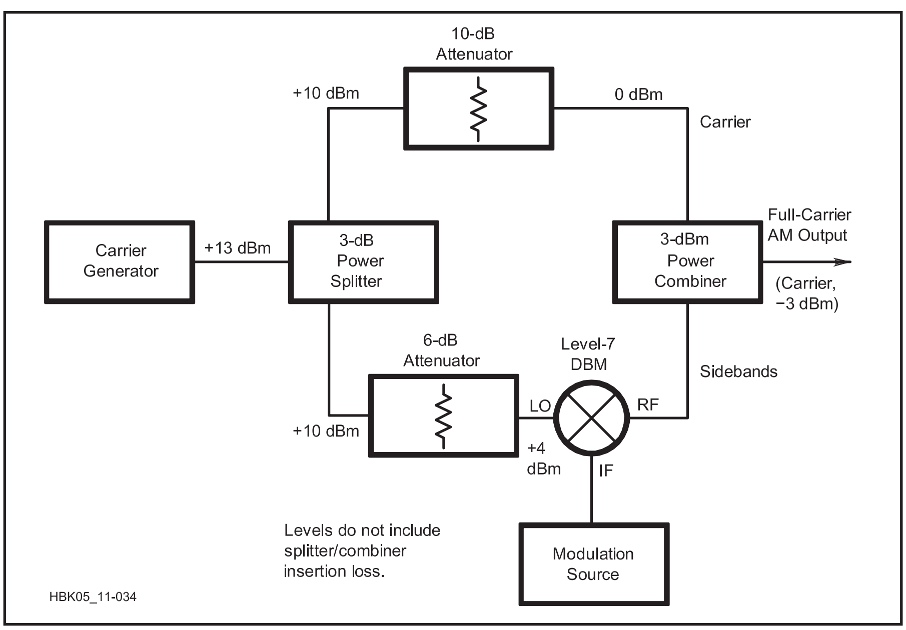


Figure 12.35 — Generating full-carrier AM with a diode DBM. A practical modulator using this technique is described in *Experimental Methods in RF Design*.

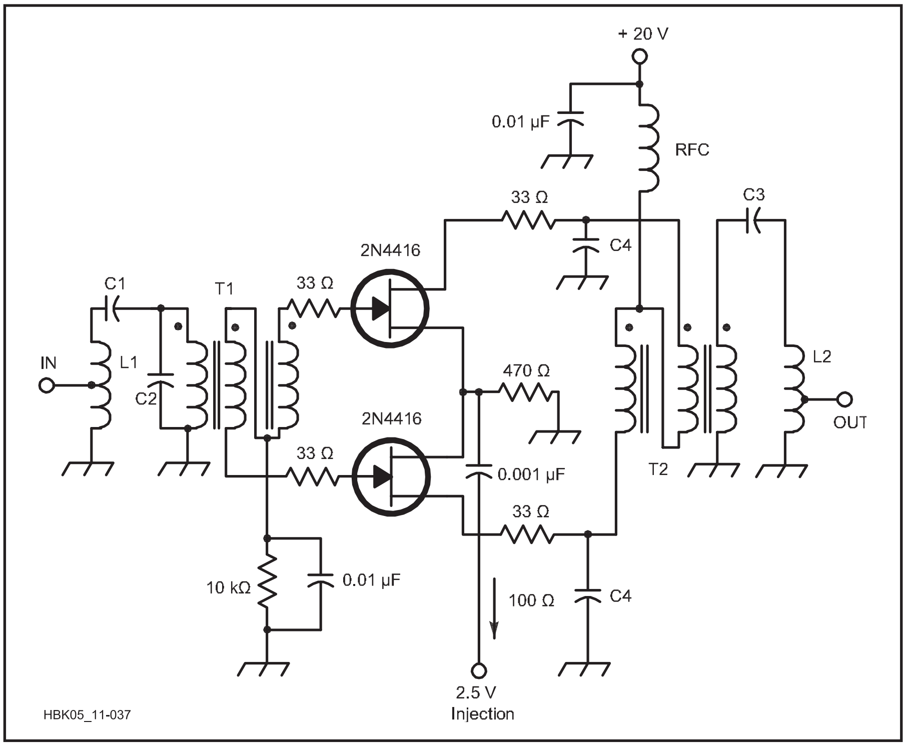


Figure 12.36 — Two 2N4416 JFETs provide highdynamic range in this mixer circuit from Sabin, *QST*, July 1970. L1, C1 and C2 form the input tuned circuit; L2, C3, and C4 tune the mixer output to the IF. The trifilar input and output transformers are broadband transmission-line types.

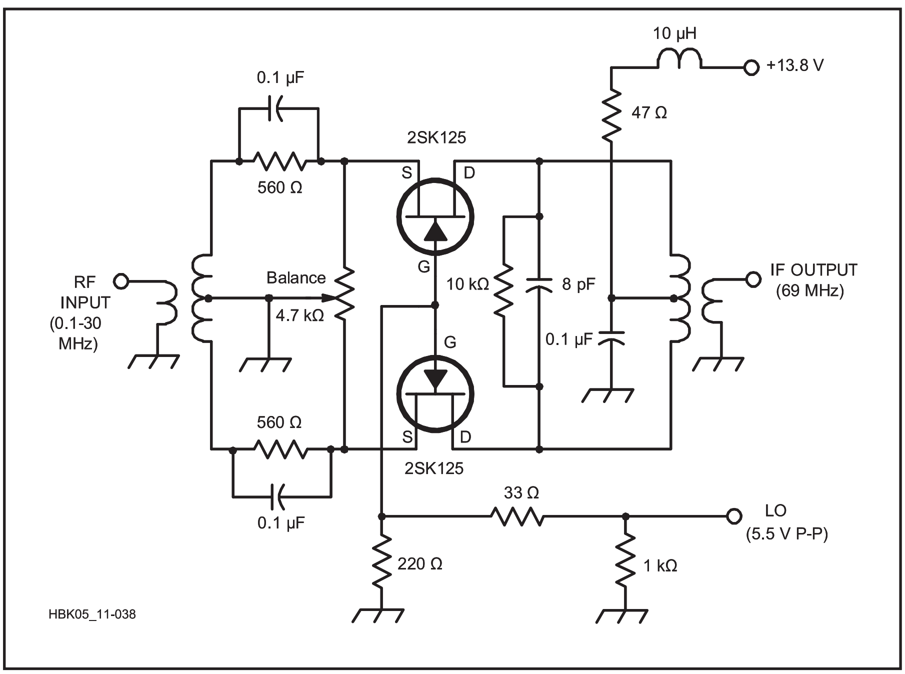


Figure 12.37 — The ICOM IC-765's single-balanced 2SK125 mixer achieves a high dynamic range (per *QST* Product Review, an IP_3 of 10.5 dBm at 14 MHz with preamp off). The first receive mixer in many commercial amateur radio transceiver designs of the 1980s and 1990s used a pair of 2SK125s or similar JFETs in much this way.

them attractive to amateur builders. The abundant availability of high-quality manufactured diode mixers at reasonable prices makes them excellent candidates for home construction projects. Although diode DBMs are common in telecommunications as a whole, their conversion loss and relatively high LO power requirement have usually driven the manufacturers of high-performance MF/HF Amateur Radio receivers and transceivers to other solutions. Those solutions have generally involved single- or double-balanced FET mixers — MOSFETs in the late 1970s and early 1980s, JFETs from the early 1980s to date. A comprehensive paper that explores the differences between various forms of active mixers, "Performance Capabilities of Active Mixers," by Ulrich Rohde, N1UL, is included in the online content accompanying this *Handbook*.

Many of the JFET designs are variations of a single-balanced mixer circuit introduced to *QST* readers in 1970. **Figure 12.36** shows the circuit as it was presented by William Sabin in "The Solid-State Receiver," *QST*, July 1970. Two 2N4416 JFETs operate in a common-source configuration, with push-pull RF input and parallel LO drive. **Figure 12.37** shows a similar circuit as implemented in the ICOM IC-765 transceiver. In this version, the JFETs (2SK125s) operate in common-gate, with the LO applied across a 220-Ω resistor between the gates and ground.

Current state of the art for active mixers in the HF through GHz range replaces discrete device designs with integrated designs such as the Analog Devices AD8342 (**www.analog.com**). Using an IC greatly improves matching of the active devices, improving circuit balance. The AD8342 has a conversion gain of 3.7 dB, a noise figure of 12.2 dB, and an input IP_3 of 22.7 dBm. The device operates with a single-voltage power supply and is well-suited to interface with digital hardware, such as for SDR applications. Reference circuits for applications at HF and VHF/UHF are provided in the device's datasheet.

12.4.5 The Tayloe Mixer

[The following description of the Tayloe Product Detector (a.k.a. — the Tayloe Mixer) is adapted from the July 2002 QEX article, "Software-Defined Radio For the Masses, Part 1" by Gerald Youngblood AC5OG, now K5SDR. — Ed.]

The beauty of the Tayloe detector (see reference listings for Tayloe) is found in both its design elegance and its exceptional performance. In its simplest form, you can build a complete quadrature downconverter with only three or four ICs (less the local oscillator) at a cost of less than $10.

Figure 12.38 illustrates a single-balanced version of the Tayloe detector. It can be visualized as a four-position rotary switch revolving at a rate equal to the carrier frequency. The 50-Ω antenna impedance is connected to the rotor and each of the four switch positions is connected to a *sampling capacitor*. Since the switch rotor is turning at exactly the RF carrier frequency, each capacitor will track the carrier's amplitude for exactly one-quarter of the cycle and will then hold its value for the remainder of the cycle. The rotating switch will therefore sample the signal at 0°, 90°, 180° and 270°, respectively.

As shown in **Figure 12.39**, the 50-Ω impedance of the antenna and the sampling capacitors form an R-C low-pass filter during the period when each respective switch is turned on. Therefore, each sample represents the integral or average voltage of the signal during its respective one-quarter cycle. When the switch is off, each sampling capacitor will hold its value until the next revolution. If the RF carrier and the rotating frequency were exactly in phase, the output of each capacitor will be a dc level equal to the average value of the sample.

If we differentially sum outputs of the 0° and 180° sampling capacitors with an op amp (see Figure 12.38), the output would be a dc voltage equal to two times the value of the individually sampled values when the switch rotation frequency equals the carrier frequency. Imagine, 6 dB of noise-free gain! The same would be true for the 90° and 270° capacitors as well. The 0°/180° summation forms the *I* channel and the 90°/270° summation forms the *Q* channel of a quadrature downconversion. (See the **Modulation** chapter for more information on I/Q modulation.)

As we shift the frequency of the carrier away from the sampling frequency, the values of the inverting phases will no longer be dc levels. The output frequency will vary according to the "beat" or difference frequency between the carrier and the switch-rotation frequency to provide an accurate representation of all the signal components converted to baseband.

Figure 12.40 is the schematic for a simple, single-balanced Tayloe detector. It consists of a PI5V331, 1:4 FET demultiplexer (an analog

Testing Mixer Performance

In order to make proper tests on mixers using signal generators, a hybrid coupler with at least 40 dB of isolation between the two input ports and an attenuator are required. The test set-up provided by DeMaw and Collins in *QST*, January 1981, shown in **Figure 12.C1** is ideal for this.* Two signal generators operating near 14 MHz are combined in the hybrid coupler, then isolated from the mixer under test (MUT) by a variable attenuator. The LO is supplied by a VFO covering 5.0 – 5.5 MHz and applied to the MUT through another variable attenuator. The output is isolated with another attenuator, amplified and applied to a spectrum analyzer for analysis.

Attenuation should be sufficient to provide isolation (minimum of 6 to 10 dB required) and to result in signal levels to the mixer under test (MUT) appropriate for the required testing and as suitable for the particular mixer device.

The 2N5109 amplifier shown may not be sufficient for extremely high intercept point tests as this stage may no longer be transparent (operate linearly) at high signal levels. For stability tests, it is recommended to have a reactive network at the output of the mixer for the sole purpose of checking whether the mixer can become unstable.

The two 14 MHz oscillators must have extremely low harmonic content and very low noise sidebands. A convenient oscillator circuit is provided in **Figure 12.C2**, based on a 1975 *Electronic Design* article.** — *Dr Ulrich L. Rohde, N1UL*

*DeMaw, W1FB, and Collins, ADØW, "Modern Receiver Mixers for High Dynamic Range," QST, Jan 1981, p 19.

**U. Rohde, "Crystal Oscillator Provides Low Noise," *Electronic Design*, Oct 11, 1975.

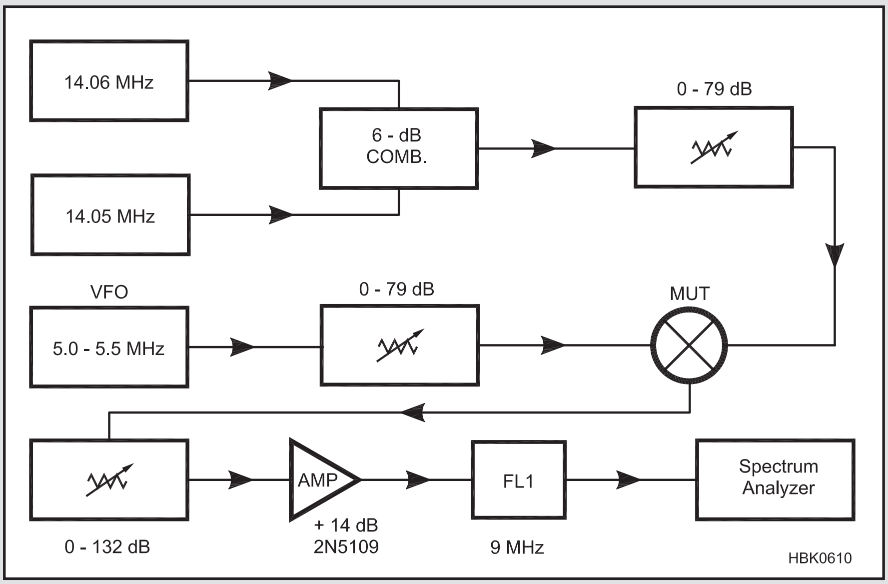


Figure 12.C1 — The equipment setup for measuring mixer performance at HF.

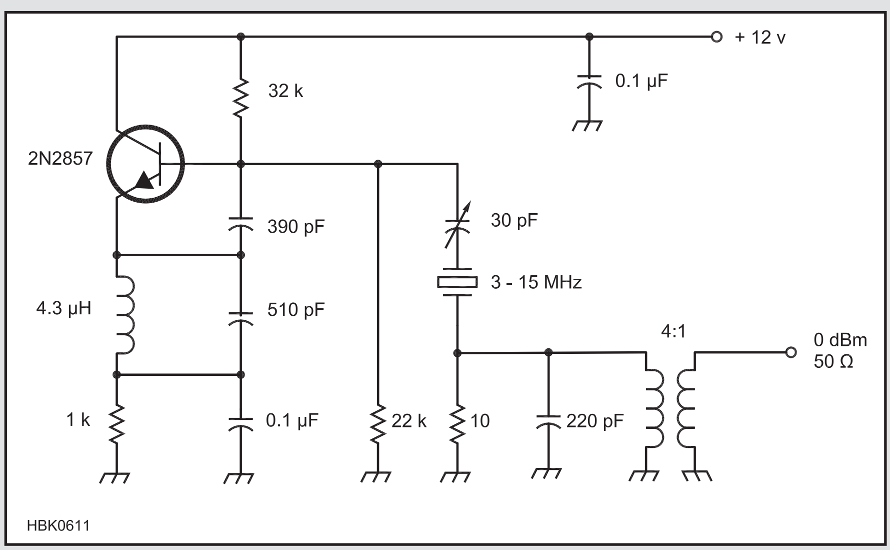


Figure 12.C2 — A low-noise VXO circuit for driving the LO port of the mixer under test in Figure 12.C1.

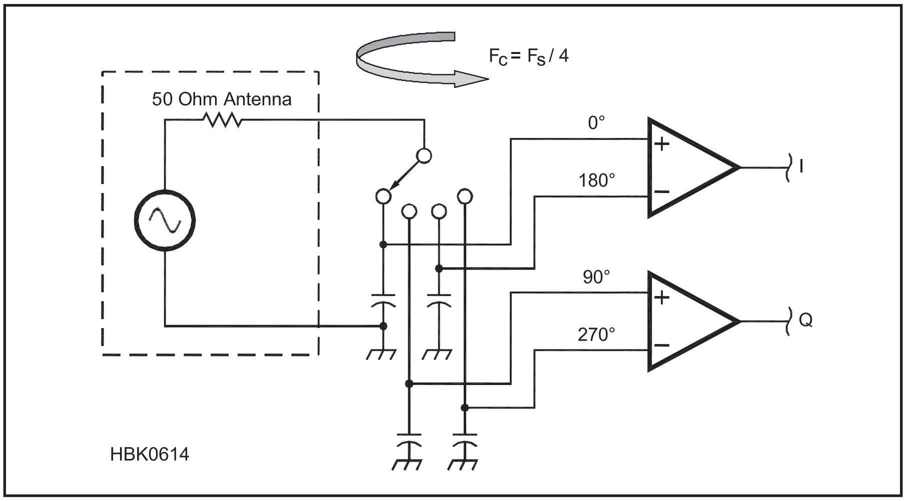


Figure 12.38 — Tayloe detector: The switch rotates at the carrier frequency so that each capacitor samples the signal once each revolution. The 0° and 180° capacitors differentially sum to provide the in-phase (I) signal and the 90° and 270° capacitors sum to provide the quadrature (Q) signal.

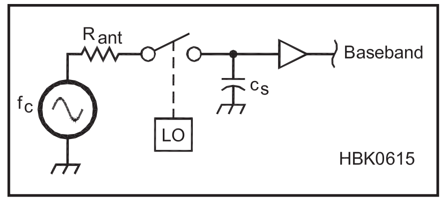


Figure 12.39 — Track-and-hold sampling circuit: Each of the four sampling capacitors in the Tayloe detector form an RC track-and-hold circuit. When the switch is on, the capacitor will charge to the average value of the carrier during its respective one-quarter cycle. During the remaining three-quarters cycle, it will hold its charge. The local-oscillator frequency is equal to the carrier frequency so that the output will be at baseband.

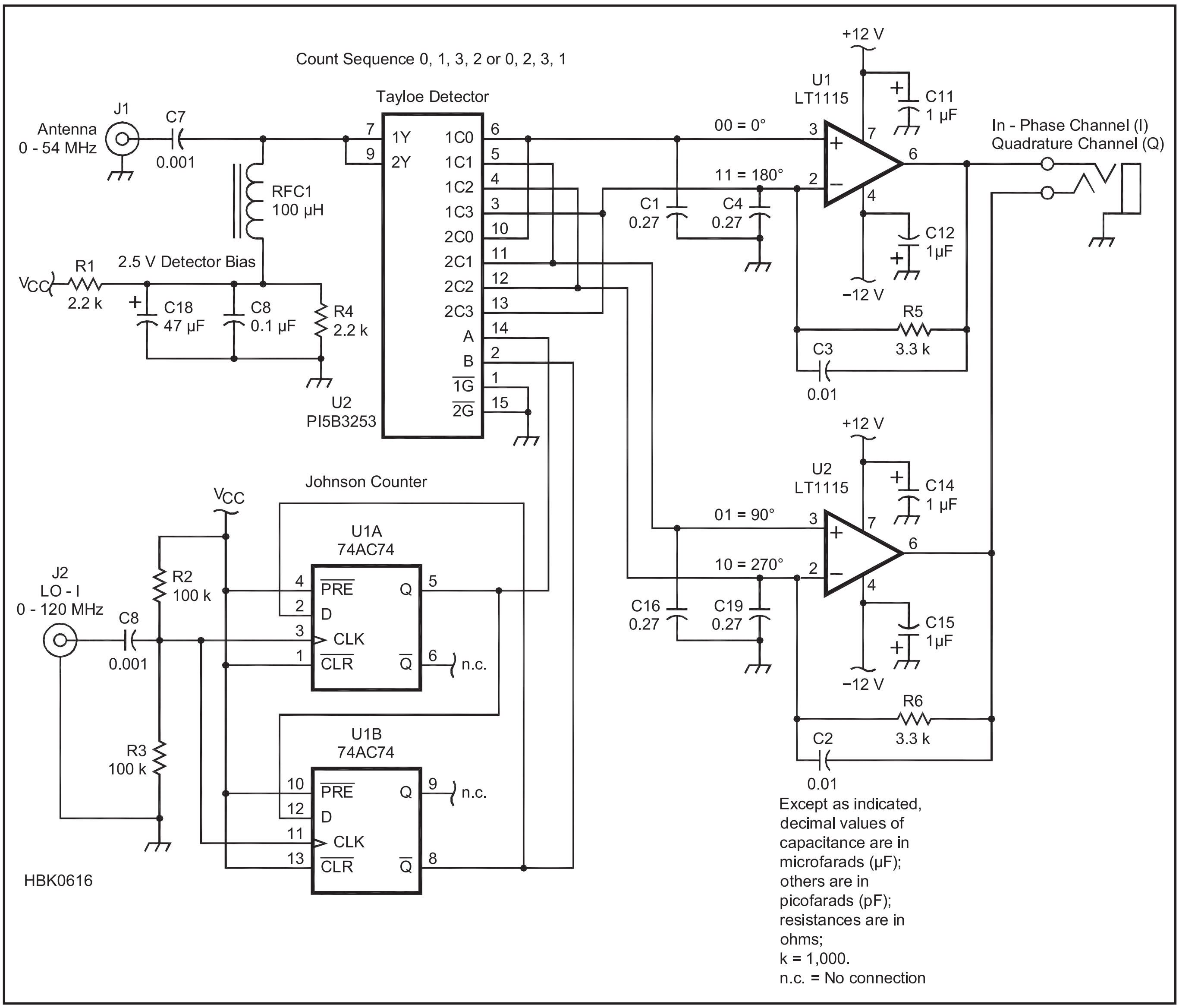


Figure 12.40 — Single-balanced Tayloe detector.

switch) that switches the signal to each of the four sampling capacitors. The 74AC74 dual flip-flop is connected as a divide-by-four Johnson counter to provide the two-phase clock to the demultiplexer chip. The outputs of the sampling capacitors are differentially summed through the two LT1115 ultra-low-noise op amps to form the *I* and *Q* outputs, respectively.

Note that the impedance of the antenna forms the input resistance for the op-amp gain as shown in the equation for gain below. This impedance may vary significantly with the actual antenna. In a practical receiver, a buffer amplifier should be used to stabilize and control the impedance presented to the mixer.

Since the duty cycle of each switch is 25%, the effective resistance in the RC network is the antenna impedance multiplied by four in the op-amp gain formula:

$$G = \frac{R_f}{4R_{ant}}$$

For example, with a feedback resistance, R_f, of 3.3 kΩ and antenna impedance, R_{ant}, of 50 Ω, the resulting gain of the input stage is:

$$G = \frac{3300}{4 \times 50} = 16.5$$

The Tayloe detector may also be analyzed as a *digital commutating filter* (see reference by Kossor). This means that it operates as a very-high-Q tracking filter, where the following equation determines the bandwidth:

$$BW_{det} = \frac{1}{\pi\, n\, R_{ant}\, C_s}$$

where n is the number of sampling capacitors, R_{ant} is the antenna impedance and C_s is the value of the individual sampling capacitors. Q_{det} of the filter is:

$$Q_{det} = \frac{f_c}{BW_{det}}$$

where f_c is the center frequency and BW_{det} is the bandwidth of the filter.

By example, if we assume the sampling capacitor to be 0.27 µF and the antenna impedance to be 50 Ω, then BW and Q at an operating frequency of 14.001 MHz are computed as follows:

$$BW_{det} = \frac{1}{\pi \times 4 \times 50 \times (2.7 \times 10^{-7})} = 5895 \text{ Hz}$$

$$Q_{det} = \frac{14.001 \times 10^{-6}}{5895} = 2375$$

The real payoff in the Tayloe detector is its performance. It has been stated that the *ideal* commutating mixer has a minimum conversion loss (which determines noise figure — see the **RF Techniques** chapter) of 3.9 dB. Typical high-level diode mixers have a conversion loss of 6-7 dB and noise figures 1 dB higher than the loss. The Tayloe detector has less than 1 dB of conversion loss, remarkably. How can this be? The reason is that it is not really a mixer but a sampling detector in the form of a quadrature track-and-hold circuit. This means that the design adheres to discrete-time sampling theory, which, while similar to mixing, has its own unique characteristics. Because a track and hold actually holds the signal value between samples, the signal output never goes to zero. (See the **DSP and SDR Fundamentals** chapter for more on sampling theory.)

This is where aliasing can actually be used to our benefit. Since each switch and capacitor in the Tayloe detector actually samples the RF signal once each cycle, it will respond to alias frequencies as well as those within the Nyquist frequency range. In a traditional direct-conversion receiver, the local-oscillator frequency is set to the carrier frequency so that the difference frequency, or IF, is at 0 Hz and the sum frequency is at two times the carrier frequency. We normally remove the sum frequency through low-pass filtering, resulting in conversion loss and a corresponding increase in noise figure. In the Tayloe detector, the sum frequency resides at the first alias frequency as shown in **Figure 12.41**. Remember that an alias is a real signal and

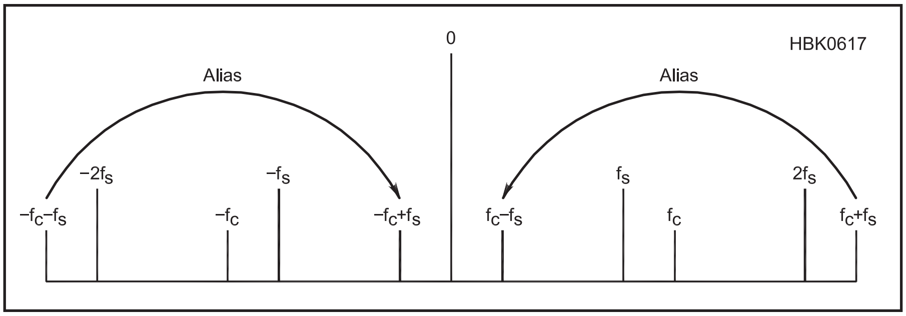


Figure 12.41 — Alias summing on Tayloe detector output: Since the Tayloe detector samples the signal, the sum frequency ($f_c + f_s$) and its image ($-f_c - f_s$) are located at the first alias frequency. The alias signals sum with the baseband signals to eliminate the mixing product loss associated with traditional mixers. In a typical mixer, the sum frequency energy is lost through filtering thereby increasing the noise figure of the device.

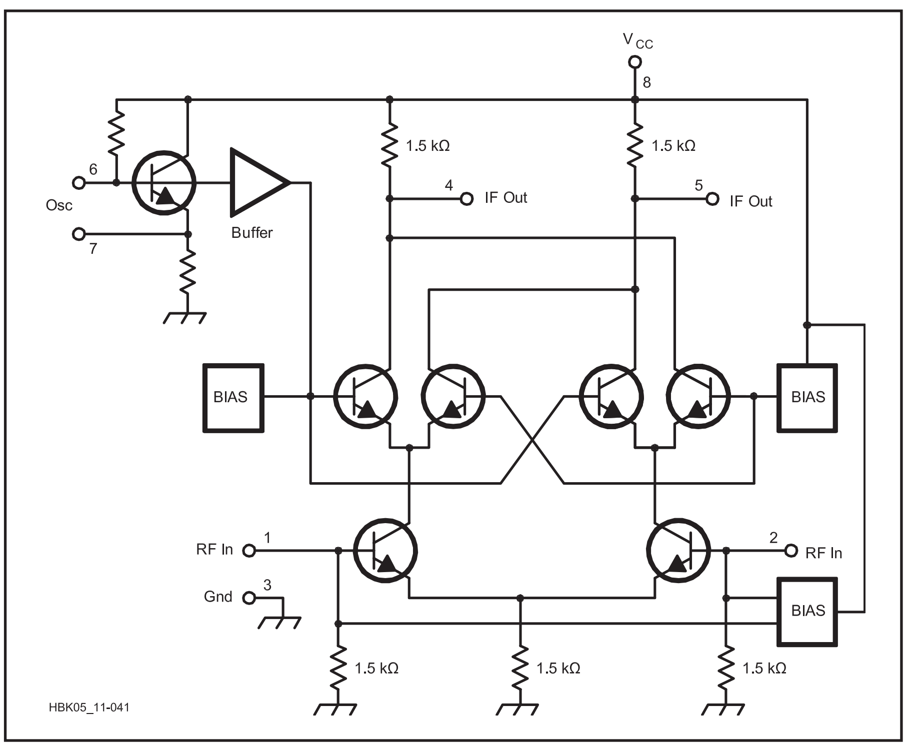


Figure 12.42 — The SA602/612's equivalent circuit reveals its Gilbert-cell origins.

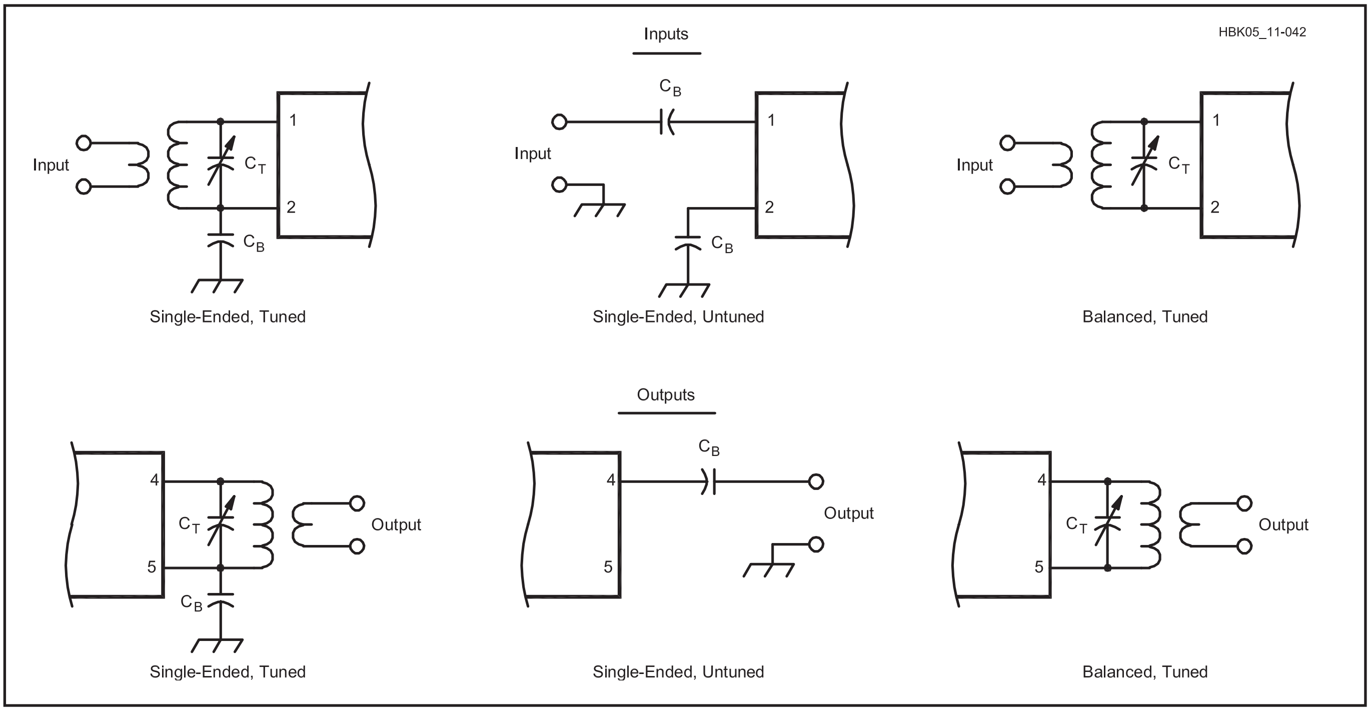


Figure 12.43 — The SA602/612's inputs and outputs can be single- or double-ended (balanced). The balanced configurations minimize second-order IMD and harmonic distortion, and unwanted envelope detection in direct-conversion service. C_T tunes its inductor to resonance; C_B is a bypass or dc-blocking capacitor. The arrangements pictured don't show all the possible input/output configurations; for instance, a center-tapped broadband transformer can be used to achieve a balanced, untuned input or output.

will appear in the output as if it were a baseband signal. Therefore, the alias adds to the baseband signal for a theoretically lossless detector. In real life, there is a slight loss, usually less than 1 dB, due to the resistance of the switch and aperture loss due to imperfect switching times.

12.4.6 The NE602/SA602/SA612 Gilbert Cell Mixer

Introduced as the Philips NE602 in the mid-1980s, the NXP SA602/SA612 mixer-oscillator IC has become greatly popular with amateur experimenters for receive mixers, transmit mixers and balanced modulators. The SA602/612's mixer is a *Gilbert cell* multiplier. **Figure 12.42** shows its equivalent circuit. A Gilbert cell consists of two differential transistor pairs whose bias current is controlled by one of the input signals. The other signal drives the differential pairs' bases, but only after being "predistorted" in a diode circuit. (This circuit distorts the signal equally and oppositely to the inherent distortion of the differential pair.) The resulting output signal is an accurate multiplication of the input voltages.

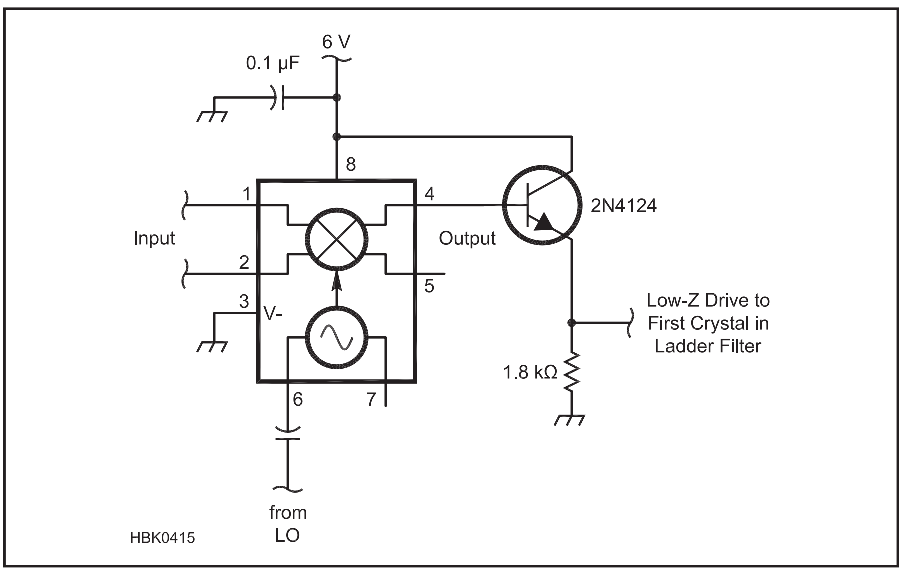


Figure 12.44 — An NPN transistor at the output of an SA602/612 mixer provides power gain and low-impedance drive for a 4.914-MHz crystal filter. A low-reactance coupling capacitor can be added between the emitter and the circuitry it drives if dc blocking is necessary. [Circuit from the Elecraft KX1 transceiver courtesy of Wayne Burdick, N6KR]

SA602/612A VARIANTS

The SA602/612 began life as the NE602/SA602. SA-prefixed 602/612 parts are specified for use over a wider temperature range than their NE-prefixed equivalents. Parts without the A suffix have a slightly lower IP_3 specification than their A counterparts. The pinout-identical NE612A and SA612A cost less than their 602 equivalents as a result of wider tolerances. All variants of this popular part should work satisfactorily in most "NE602" experimenter projects. The same mixer/oscillator topology, modified for slightly higher dynamic range at the expense of somewhat less mixer gain, is also available in the mixer/oscillator/FM IF chips NE/SA605 (input IP_3 typically –10 dBm) and NE/SA615 (input IP_3 typically –13 dBm).

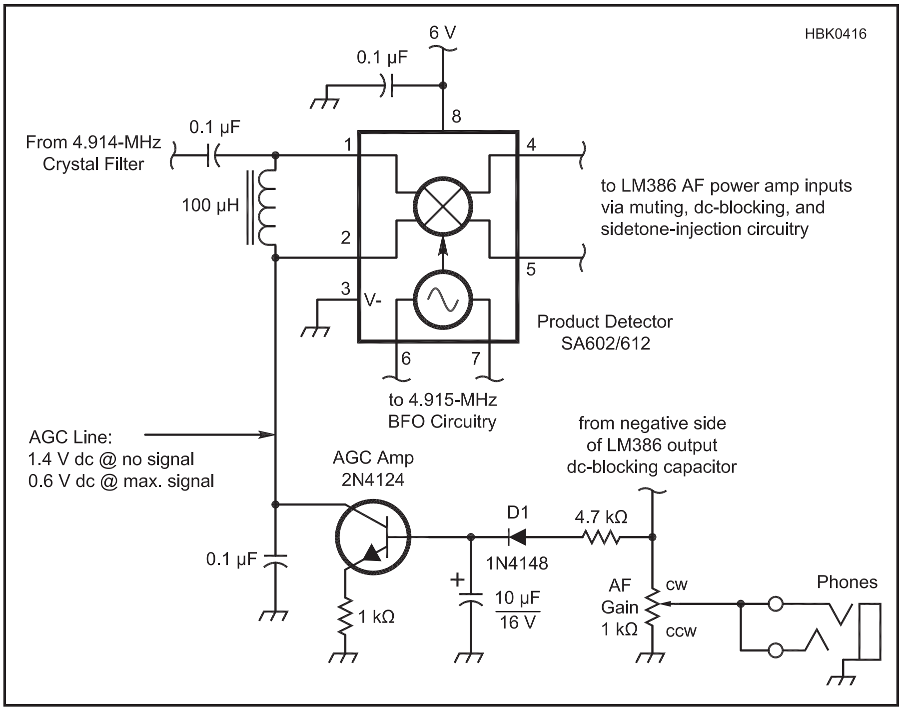


Figure 12.45 — SA602/612 product detector AGC from the Elecraft KX1 transceiver. Designed by Wayne Burdick, N6KR, this circuit first appeared in the Wilderness Radio SST transceiver with an LED used at D1 for simultaneous signal indication and rectification. The selectivity provided by the crystal filter preceding the detector works to mitigate the effects of increasing detector distortion with gain reduction. [Circuit courtesy of Wayne Burdick, N6KR, and Bob Dyer, K6KK]

SA602/612 USAGE NOTES

The SA602/612's typical current drain is 2.4 mA; its supply voltage range is 4.5 to 8.0 V. Its inputs (RF) and outputs (IF) can be single- or double-ended (balanced) according to design requirements (**Figure 12.43**). The equivalent ac impedance of each input is approximately 1.5 kΩ in parallel with 3 pF; each output's resistance is 1.5 kΩ. **Figure 12.44** shows the use of an NPN transistor at the SA602/612 output to obtain low-impedance drive for a crystal filter; **Figure 12.45** shows how AGC can be applied to an SA602/612.

The SA602/612 mixer can typically handle signals up to 500 MHz. At 45 MHz, its noise figure is typically 5.0 dB; its typical conversion gain, 18 dB. Note that in contrast to the diode-based mixers described earlier, which have conversion *loss*, most Gilbert-cell mixers have conversion *gain*. Considering the SA602/612's low current drain, its input IP_3 (measured at 45 MHz with 60-kHz spacing) is usefully good at –15 dBm. Factoring in the mixer's conversion gain results in an equivalent output IP_3 of about 3 dBm.

The SA602/612's on-board oscillator can operate up to 200 MHz in LC and crystal-controlled configurations (**Figure 12.46** shows three possibilities). Alternatively, energy from an external LO can be applied to the chip's pin 6 via a dc blocking capacitor. At least 200 mV P-P of external LO drive is required for proper mixer operation.

The SA602/612 was intended to be used as the second mixer in double-conversion FM cellular radios, in which the first IF is typically 45 MHz, and the second IF is typically 455 kHz. Such a receiver's second mixer can be relatively weak in terms of dynamic range because of the adjacent-signal protection afforded by the high selectivity of the first-IF filter preceding it. When used as a first mixer, the SA602/612 can provide a two-tone third-order dynamic range between 80 and 90 dB, but this figure is greatly diminished if a preamplifier is used ahead of the SA602/612 to improve the system's noise figure.

When the SA602/612 is used as a second mixer, the sum of the gains preceding it should not exceed about 10 dB. An SA602/612 can serve as low-distortion (THD <1%) product detector if overload is avoided through the use of AGC and appropriate attenuation between the '602/612 and the IF strip that drives it.

The SA602/612 is generally not a good choice for VHF and higher-frequency mixers because of its input noise and diminishing IMD performance at high frequencies. There are applications, however, where 6-dB noise figure and 60- to 70-dB dynamic range performance is adequate. If your target specifications exceed these numbers, you should consider other mixers at VHF and up.

Figure 12.47 shows the schematic of a complete 7-MHz direct-conversion receiver based on the SA602/612 and the widely used LM386 AF power amplifier IC. Such simple product-detector-based receivers sometimes suffer from incidental envelope detection, which causes audio from strong, full-carrier-AM shortwave or medium-wave broadcast stations to be audible regardless of where the receiver LO is tuned. RF attenuation and/or band-limiting the receiver input with a double- or triple-tuned-circuit filter can usually reduce this effect to inaudibility.

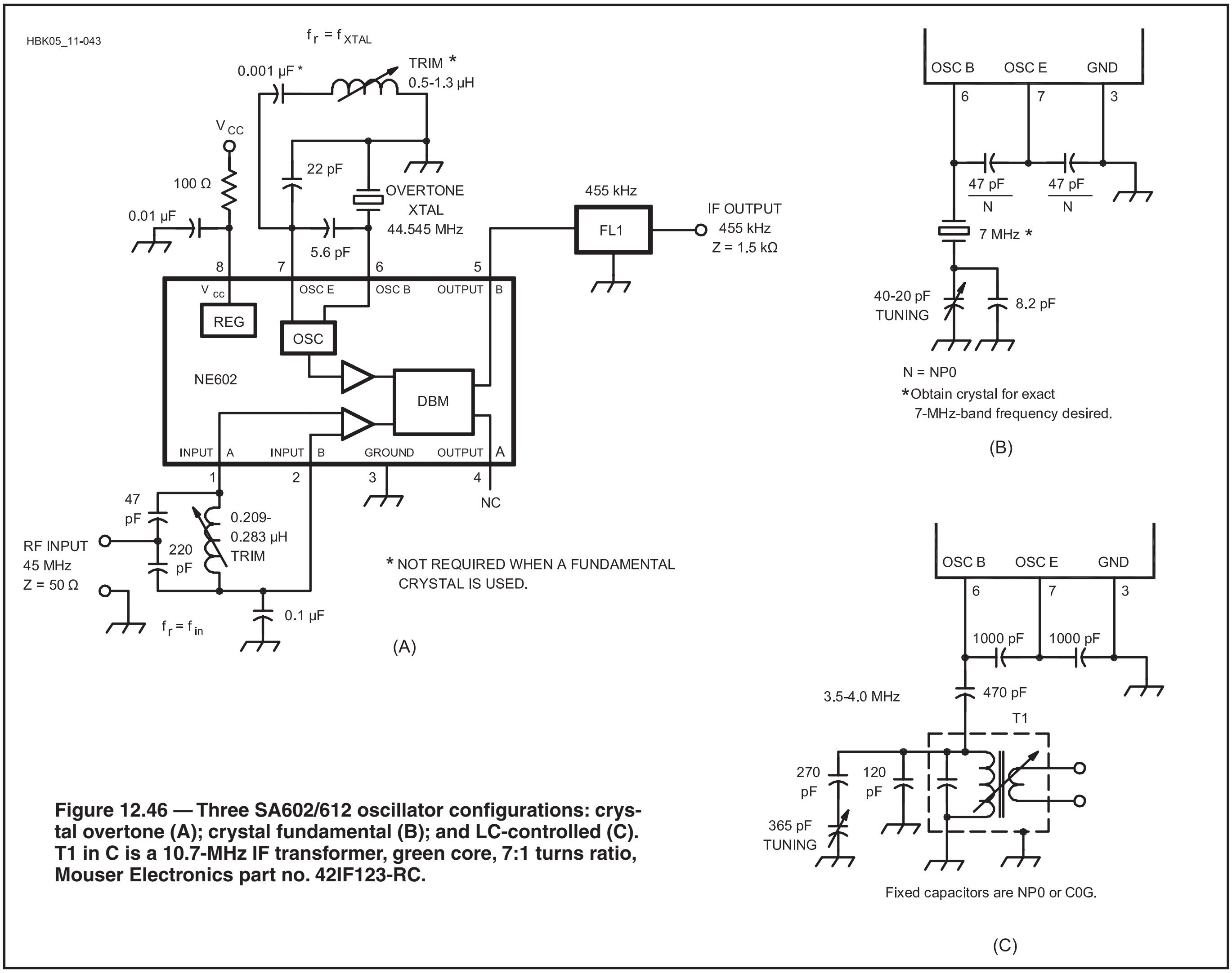


Figure 12.46 — Three SA602/612 oscillator configurations: crystal overtone (A); crystal fundamental (B); and LC-controlled (C). T1 in C is a 10.7-MHz IF transformer, green core, 7:1 turns ratio, Mouser Electronics part no. 42IF123-RC.

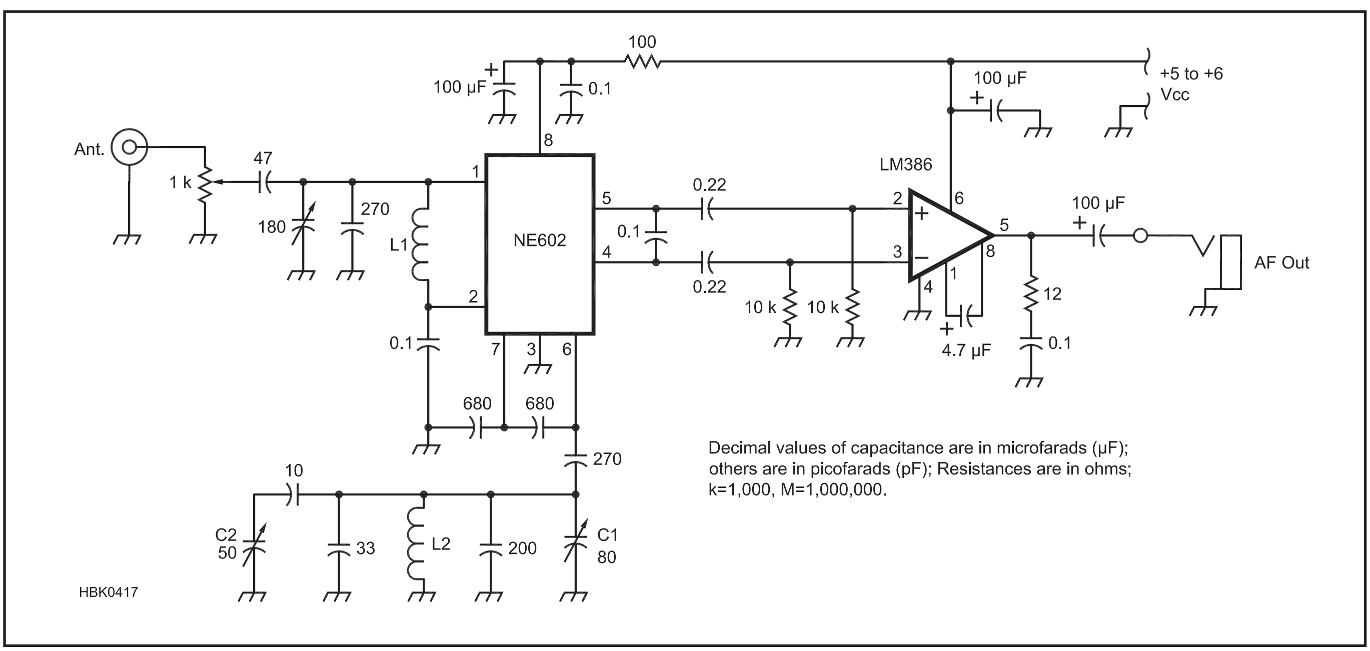


Figure 12.47 — A 7-MHz direct-conversion receiver based on the NE602/SA602/612. Equipped with a stage or two of audio filtering and a means of muting during transmit periods, such a receiver is entirely sufficient for basic Amateur Radio communication at MF and HF. L1 and L2 are 1.2 µH. This receiver is described in greater detail in *Experimental Methods in RF Design.*

12.5 Demodulation and Detection

Translating information from radio form back into its original form — demodulation — is also traditionally called *detection*. If the information signal we want to detect consists merely of a baseband signal frequency-shifted into the radio realm, almost any low-distortion frequency-shifter that works according to the sidebar "Mixer Math: Mixing as Multiplication" can do the job acceptably well.

Sometimes we recover a radio signal's information by shifting the signal back to its original form with no intermediate frequency shifts. This is direct conversion. More commonly, we first convert a received signal to an intermediate frequency so we can amplify, filter and level-control it prior to detection. This is superheterodyne reception.

Whatever the receiver type, however, the received signal ultimately makes its way to one last mixer or demodulator (analog or digital) that completes the final translation of information back into audio, video, or into a signal form suitable for device control or computer processing.

In a heterodyne receiver's last translation, the incoming signal is converted back to recovered-information form by mixing it with one last RF signal. In heterodyne or *product detection*, that final frequency-shifting signal comes from a BFO. The incoming-signal energy goes into one mixer input port, BFO energy goes into the other, and audio (or whatever form the desired information takes) results.

In SDR receivers, the process may involve the digital equivalent of the analog process or some other mathematical process may be used.

12.5.1 Envelope Detection and Full-Carrier AM

If the incoming signal is full-carrier AM and we don't need to hear the carrier as a tone, we can modify this process somewhat, if we want. We can use the carrier itself to provide the heterodyning energy in a process called *envelope detection*.

A full-carrier AM signal's *modulation envelope* corresponds to the shape of the modulating wave. If we can derive from the modulated signal a voltage that varies accord-

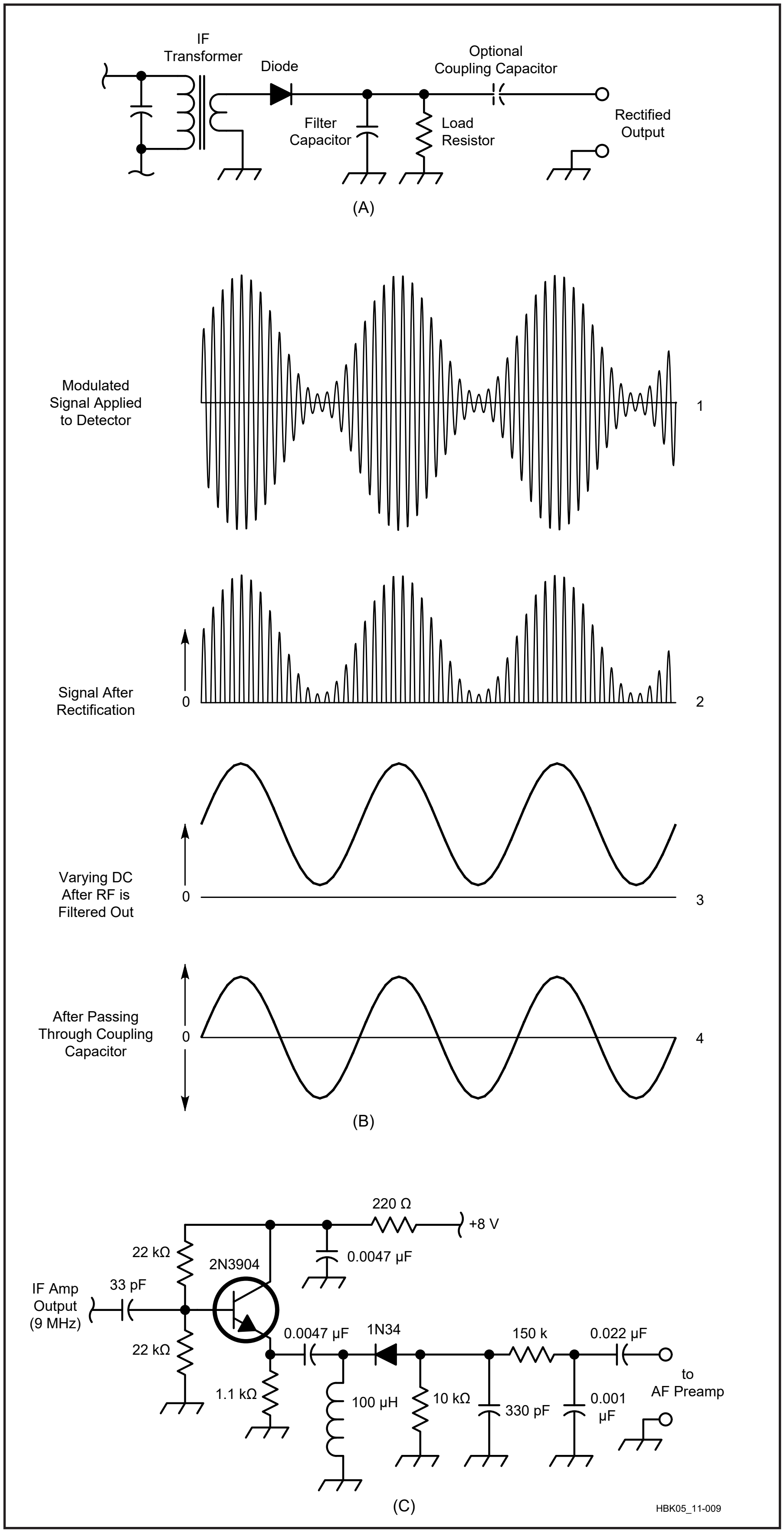


Figure 12.48 — Radio's simplest demodulator, the diode rectifier (A), demodulates an AM signal by acting as a switch that multiplies the carrier and sidebands to produce frequency sums and differences, two of which sum into a replica of the original modulation (B). Modern receivers often use an emitter follower to provide low-impedance drive for their diode detectors (C).

ing to the modulation envelope, we will have successfully recovered the information present in the sidebands. This process is called *envelope detection*, and we can achieve it by doing nothing more complicated than half-wave-rectifying the modulated signal with a diode (**Figure 12.48**).

That a diode demodulates an AM signal by allowing its carrier to multiply with its sidebands may jar those long accustomed to seeing diode detection ascribed merely to "rectification." But a diode is certainly nonlinear. It passes current in only one direction, and its output voltage is (within limits) proportional to the square of its input voltage. These nonlinearities allow it to multiply.

Exploring this mathematically is tedious with full-carrier AM because the process squares three summed components (carrier, lower sideband and upper sideband). Rather than fill the better part of a page with algebra, we'll instead characterize the outcome verbally: In "just rectifying" a DSB, full-carrier AM signal, a diode detector produces

• Direct current (the result of rectifying the carrier);

• A second harmonic of the carrier;

• A second harmonic of the lower sideband;

• A second harmonic of the upper sideband;

• Two difference-frequency outputs (upper sideband minus carrier, carrier minus lower sideband), each of which is equivalent to the modulating wave-form's frequency, and both of which sum to produce the recovered information signal; and

• A second harmonic of the modulating waveform (the frequency difference between the two sidebands).

Three of these products are RF signals. Low-pass filtering, sometimes little more than a simple RC network, can remove the RF products from the detector output. A capacitor in series with the detector output line can block the carrier-derived dc component. That done, only two signals remain: the recovered modulation and, at a lower level, its second harmonic — in other words, second-harmonic distortion of the desired information signal.

12.5.2 Detecting AM Signals

Note that the shape of the envelope of the modulated RF signal matches the shape of the modulating signal. (This is only true for full AM including the carrier and both sidebands. The envelope of a single-sideband SSB signal is not an accurate reproduction of the modulating signal which is why it cannot be recovered by an envelope detector.) That suggests a possible demodulation method. A diode detector puts out a signal proportional to the envelope of the RF signal, recovering the original modulation. See **Figure 12.49**. The capacitor should be large enough that it filters out most of the RF ripple but not so large that it attenuates the higher audio frequencies.

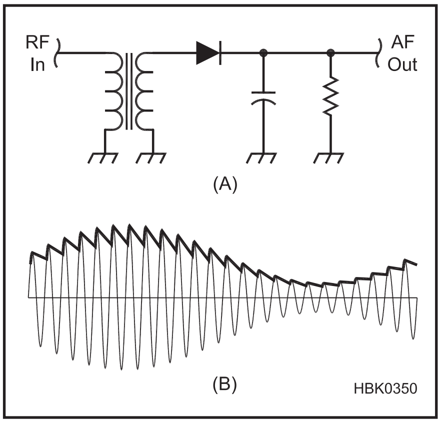


Figure 12.49 — A simple diode-type AM detector, also known as an *envelope detector* (A). The demodulated output waveform has the same shape as the envelope of the RF signal (B). In (B) the thin line is the RF signal modulated with a sine wave and the darker line is the demodulated audio frequency with some residual RF ripple.

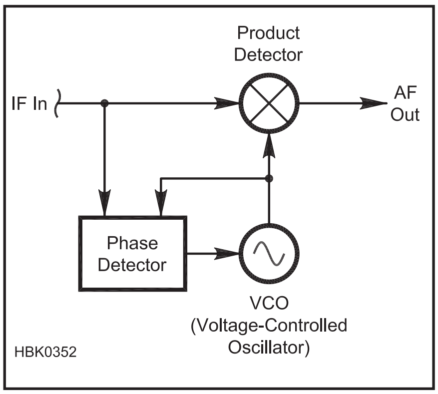


Figure 12.50 — Block diagram of a synchronous detector. The voltage-controlled oscillator (VCO) is part of a phase-locked loop that locks the oscillator to the carrier frequency of the incoming AM or DSBRC signal.

One problem with AM is that if the amplitude of the carrier becomes attenuated for any reason, then the modulation is distorted, especially the negative-going portion near the 100%-modulation (zero power) point. This can happen due to a propagation phenomenon called *selective fading*. It occurs when the signal arrives simultaneously at the receive antenna via two or more paths, such as ground wave and sky wave. If the difference in the distance of the two paths is an odd number of half-wavelengths, then the two signals are out of phase. If the amplitudes are nearly the same, they cancel and a deep fade results. Since wavelength depends on frequency, the fading is frequency-selective. On the lower-frequency amateur bands it is possible for an AM carrier to be faded while the two sidebands are still audible.

A solution is to regenerate the carrier in the receiver. Since the carrier itself carries no information about the modulation, it is not necessary for demodulation. The transmitted signal may be a standard full-carrier AM signal or the carrier may be suppressed, resulting in double sideband, suppressed carrier (DSB-SC). An AM detector that regenerates the carrier from the signal is known as a *synchronous detector*. Often the regenerated carrier oscillator is part of a phase-locked loop (PLL) that locks onto the incoming carrier. See **Figure 12.50**. Synchronous detectors not only reduce the effects of selective fading but also are usually more linear than diode detectors so they have less distortion. Some commercial short-wave broadcasts include a reduced, but not suppressed, carrier (DSB-RC) to allow operation with PLL-type synchronous detectors.

Since the advent of single sideband in the 1960s, full-carrier double-sideband AM has become less popular in Amateur Radio. It does retain several advantages however. We have already mentioned the simplicity of the circuitry. Another advantage is that, because of the presence of the carrier, the automatic gain control system in the receiver remains engaged at all times, ensuring a constant audio level. Unlike with SSB, there is no rush of noise during every pause in speech. Also tuning is less critical than with SSB. There is no "Donald Duck" sound if the receiver is slightly mistuned. Finally, the audio quality of an AM signal is usually better than that of SSB because of the lack of a crystal filter in the transmit path and the wider filter in the receiver.

SDR AM SQUARE-LAW DETECTOR

(*This section is taken from the Jul/Aug 2009 QEX column SDR: Simplified by W5IFS.*) We need to convert the full-carrier double sideband signal into the baseband information. One of the classic AM demodulation methods is a *square-law detector*. Very few devices have a true square law response over a large range of signals. The JFET comes close, but has limited range. With DSP we can implement a square law detector with range that is limited only by the size of the data that the DSP can handle. If we use 16-bit data for the input, our multiplier must have a 32-bit result. This is a very simple and fast operation for the DSP.

The square-law response to a DSB-AM signal with a carrier, sin(y), and two sidebands, cos(x-y) and cos(x+y) is as follows:

$$V_{out} = (1\sin(y) + \tfrac{1}{2}\cos(x-y) - \tfrac{1}{2}\cos(x+y))^2$$

Applying trig identities and working through the math:

$V_{out} = \sin(y)^2 + \sin(y)\cos(x–y) – \sin(y)\cos(x+y) + ¼ \cos(x+y)^2 + ¼ \cos(x–y)^2$

$V_{out} = ¾ – ½ \sin(2y) + ½ \sin(x) + ½ \sin(2y–x) – ½ \sin(–x) – ½ \sin(2y+x) + ⅛ \cos(2y+2x) + ⅛ \cos(2y–2x)$

Finally:

$V_{out} = ¾ + 1 \sin(x) – ½ \sin(2y) + ½ \sin(2y–x) – ½ \sin(2y+x) + ⅛ \cos(2y+2x)+ ⅛ \cos(2y–2x)$

(This result will have 32 bits, but our filter math should only be 16 bits. We can divide the result by 65536 to get the signal back into range for further operations. In software, this just means we throw away the bottom 16 bits of data before we store the result.)

The signal we want is sin(x), but we have a significant dc component (¾) and a whole bunch of unwanted signals at or close to twice the carrier frequency, 2y. The only real problem is the dc component, since it is really close to the desired audio signal. We need to use a band-pass filter to remove the dc and the RF components. Since all but the dc are far removed from the desired audio, we can use a relatively simple band-pass filter.

12.5.3 Demodulating SSB Signals

A complete discussion of the three main techniques used to receive and demodulate SSB signals — filter, phasing, and Weaver — can be found in the **Modulation** chapter. Examples of practical circuits that implement these methods are included in the receiver projects provided in the online content.

HETERODYNE RECEIVERS

As with an AM synchronous detector, the carrier is regenerated in a superheterodyne receiver but since only one sideband is present, the synchronous detector's phase-locked loop (described above) is not possible. Instead, the detector uses a free-running *beat-frequency oscillator* (BFO). The detector itself is called a *product detector* because its output is the mathematical product of the BFO and the SSB signal. The BFO must be tuned to the same frequency as the suppressed carrier to prevent distortion of the recovered audio. That is done by carefully tuning the local oscillator (the main tuning dial) such that after conversion to the intermediate frequency, the suppressed carrier aligns with the BFO frequency. Demodulation of SSB signals in an analog heterodyne receiver must be done using a product detector or equivalent technique. Envelope detection will not work because the waveform does not have the shape of the modulating waveform.

SDR RECEIVERS

SDR receivers use I/Q quadrature demodulator architecture for SSB and many other modes. This technique is discussed in the **Modulation** chapter. A block diagram of an SDR demodulator for SSB is shown in **Figure 12.51.** This particular system shows the phasing method of SSB generation in reverse. The excellent tutorial, "Understanding the 'Phasing Method' of Single Sideband Demodulation" by Rick Lyons (**www.dsprelated.com/showarticle/176.php**) explains the technique in detail. The filter method and the Weaver method can also be implemented with an I/Q architecture although this is not the usual method used by designers.

A quadrature I/Q demodulator generates the I_1 and Q_1 signals which are low-pass filtered to remove high-frequency components, then down-converted using decimation. The I and Q channels are band-pass filtered, with a Hilbert transformer applied to the Q channel as described in the **DSP and SDR Fundamentals** chapter. The resulting streams are then combined so that the Q channel is subtracted from the I channel. This leaves the USB signal, which is converted to analog audio, and low-pass filtered.

If a LSB signal is being received, change the summing symbol immediately preceding the DAC so as to add the I and Q channels together. This produces the LSB signal which is then converted to audio and filtered as before. In practice, both the LSB and USB signals are available at any time and the digital audio stream can be routed to the DAC and low-pass filter under software control.

12.5.4 Demodulating FM and PM

ANALOG DEMODULATORS

Although angle modulation does generate an infinite number of sidebands, demodulating angle modulation requires little more than

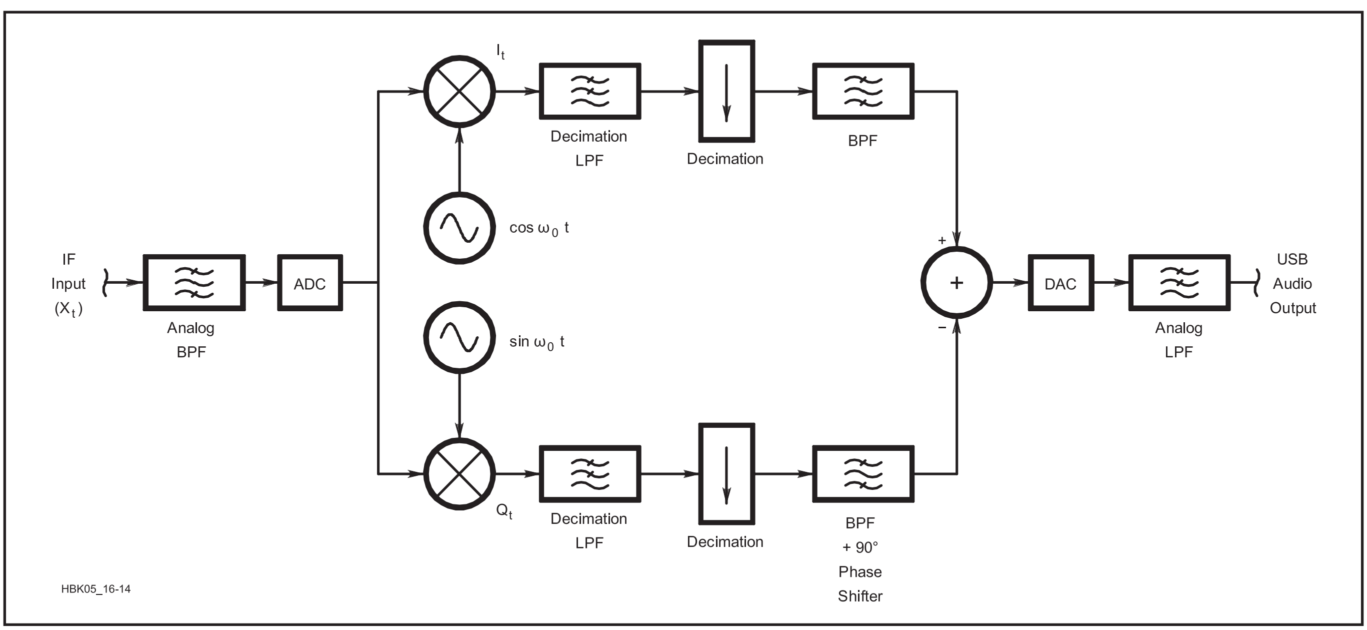
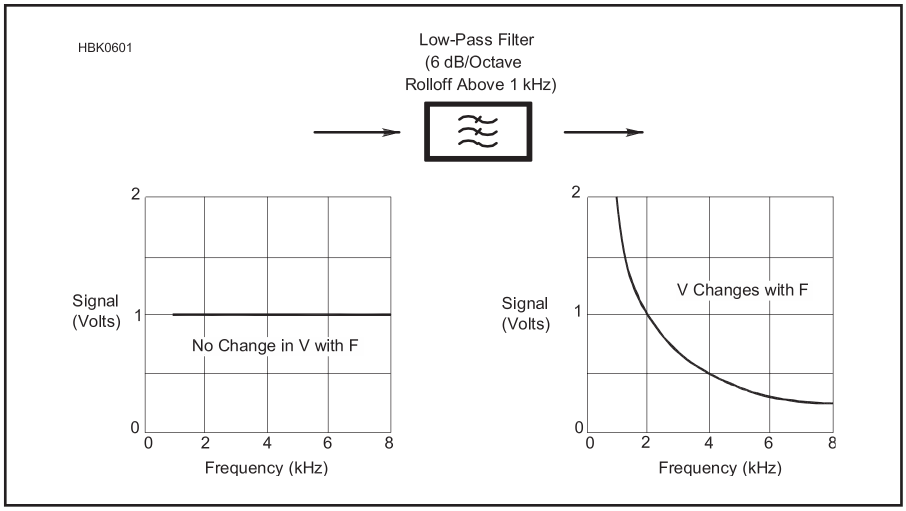


Figure 12.51 — Block diagram of a digital SSB demodulator.

Figure 12.52 — Frequency-sweeping a constant-amplitude signal and passing it through a low-pass filter results in an output signal that varies in amplitude with frequency. This is the principle behind the angle-demodulation process called *frequency discrimination.*



turning it into AM and then envelope- or product-detecting it! But this is what happens in many of our FM receivers and transceivers, and we can get a handle on this process by realizing that a form of angle-modulation-to-AM conversion is created from linear distortion of the modulation by amplitude-linear circuitry. This happens to angle-modulated signals in any linear circuit that doesn't have an amplitude-versus-frequency response that's utterly flat out to infinity.

Think of what happens, for example, when we sweep a constant-amplitude signal up in frequency — say, from 1 kHz to 8 kHz — and pass it through a 6-dB-per-octave filter (**Figure 12.52**). The filter's rolloff causes the output signal's amplitude to decrease as frequency increases. Now imagine that we linearly sweep our constant-amplitude signal *back and forth* between 1 kHz and 8 kHz at a constant rate of 3 kHz per second. The filter's output *amplitude* now varies cyclically over time as the input signal's *frequency* varies cyclically over time. Right before our eyes, a frequency change turns into an amplitude change. The process of converting angle modulation to amplitude modulation has begun.

This is what happens whenever an angle-modulated signal passes through circuitry with an amplitude-versus-frequency response that isn't flat out to infinity. As the signal deviates across the frequency-response curves of whatever circuitry passes it, its angle modulation is, to some degree, converted to AM — a form of crosstalk between the two modulation types, if we wish to look at it that way. (Variations in system phase linearity also cause distortion and FM-to-AM conversion, because the sidebands do not have the proper phase relationship with respect to each other

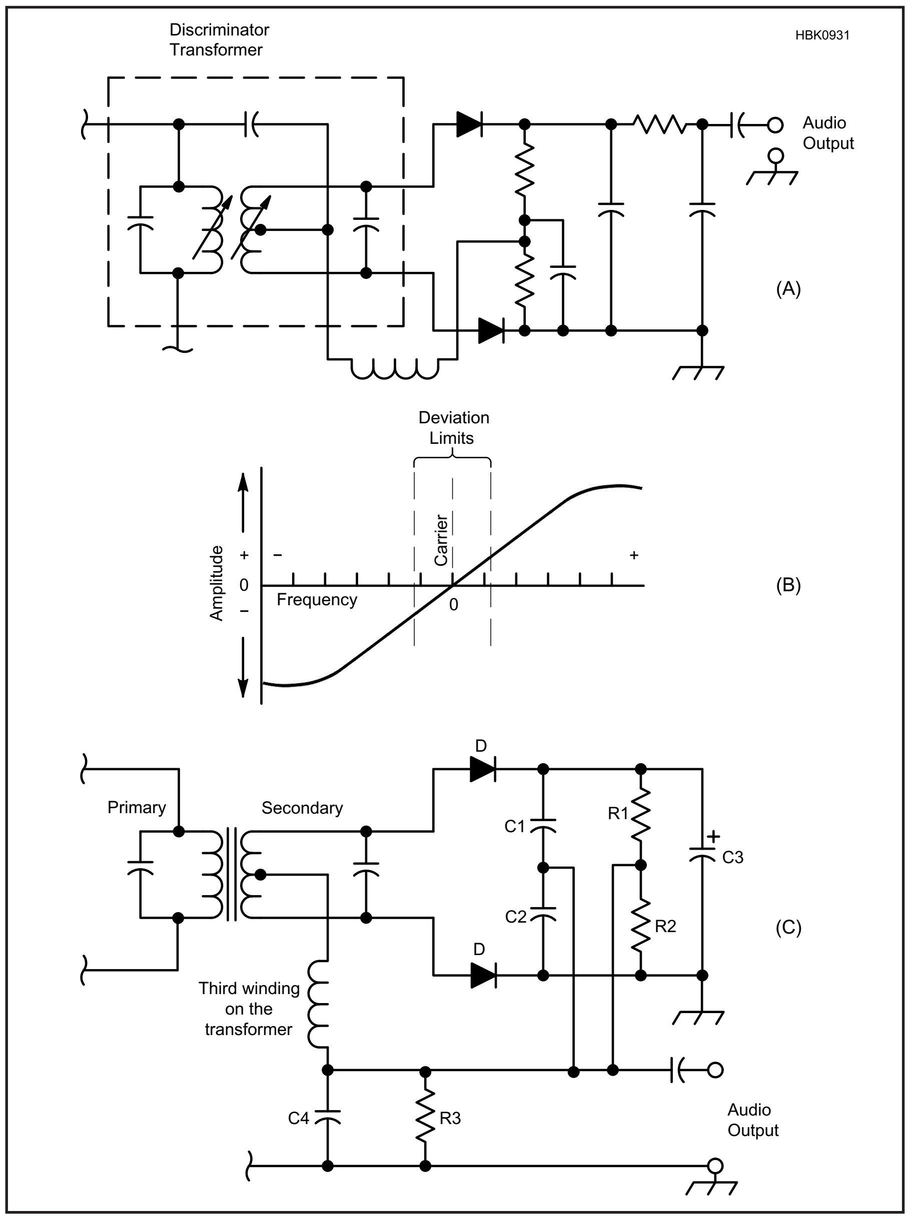


Figure 12.53 — A *discriminator* (A) converts an angle-modulated signal's deviation into an amplitude variation (B) and envelope-detects the resulting AM signal. For undistorted demodulation, the discriminator's amplitude-versus-frequency characteristic must be linear across the input signal's deviation. A *crystal discriminator* uses a crystal as part of its frequency-selective circuitry. The *ratio detector* at (C) operates similarly but has an improved rejection of amplitude modulated noise.

and with respect to the carrier.)

All we need to do to put this effect to practical use is develop a circuit that does this frequency-to-amplitude conversion linearly across the frequency span of the modulated signal's deviation. Then we envelope-demodulate the resulting AM, and we have achieved angle demodulation.

Figure 12.53 shows such a circuit — a Foster-Seeley *discriminator* — and the sort of amplitude-versus-frequency response we expect from it. (**www.electronics-notes.com/articles/radio/modulation/fm-frequency-demodulation-foster-seeley-detector-discriminator.php**) It's actually possible to use an AM receiver to recover understandable audio from a narrow angle-modulated signal by "off-tuning" the signal so its deviation rides up and down on one side of the receiver's IF selectivity curve. This *slope detection* process served as an early, suboptimal form of frequency discrimination in receivers not designed for FM It is always worth trying as a last-resort-class means of receiving narrowband FM with an AM receiver.

The ratio detector in Figure 12.53C is a variation on the discriminator which is better at rejecting AM noise mixed with the FM signal. (**www.electronics-notes.com/articles/radio/modulation/fm-frequency-demodulation-ratio-detector-discriminator.php**) Note that the diodes are in series in this circuit. A similar phase shift occurs in the third transformer winding as the signal moves away from the frequency of the tuned transformer winding. This causes an imbalance in the transformer secondary and current flows in the third winding. That signal is then filtered by C_4 and R_3 before being passed to the audio stages of the receiver.

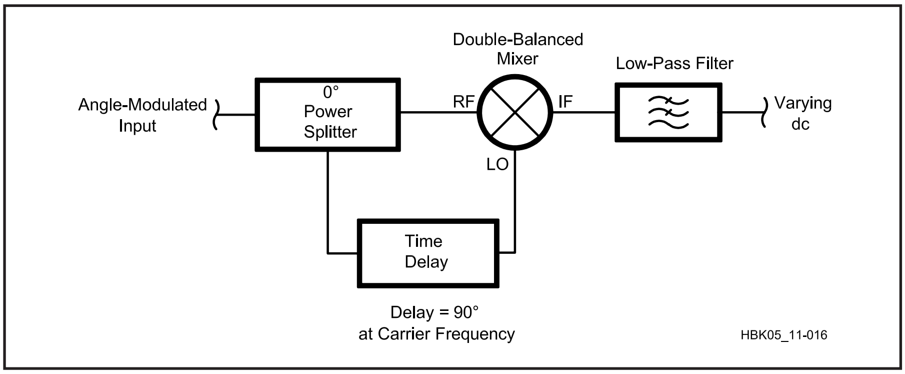


Figure 12.54 — In *quadrature detection*, an angle-modulated signal multiplies with a time-delayed copy of itself to produce a dc voltage that varies with the amplitude and polarity of its phase or frequency excursions away from the carrier frequency. A practical quadrature detector can be as simple as a 0° power splitter (that is, a power splitter with in-phase outputs), a diode double-balanced mixer, a length of coaxial cable ¼-λ (electrical) long at the carrier frequency, and a bit of low-pass filtering to remove the detector output's RF components. IC quadrature detectors achieve their time delay with one or more resistor-loaded tuned circuits (Figure 12.55).

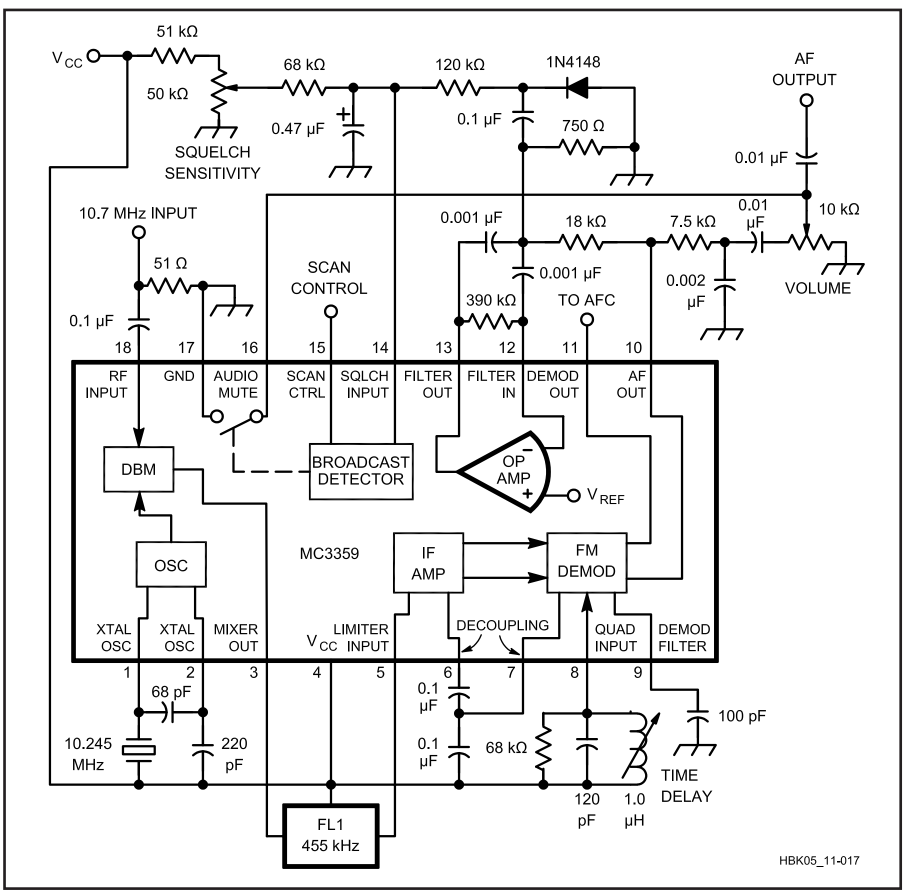


Figure 12.55 — The Freescale MC3359/NTE680 is one of many FM subsystem ICs that include limiter and quadrature-detection circuitry. The TIME DELAY coil is adjusted for minimum recovered-audio distortion.

QUADRATURE DETECTORS

It's also possible to demodulate an angle-modulated signal merely by multiplying it with a time-delayed copy of itself in a double-balanced mixer as shown in **Figure 12.54**; the sidebar, "Mixer Math: Quadrature Demodulation," explains the process numerically.

An ideal quadrature detector puts out 0 V dc when no modulation is present (with the carrier at f_c). The output of a real quadrature detector may include a small *dc offset* that requires compensation. If we need the detector's response all the way down to dc, we've got it; if not, we can put a suitable blocking capacitor in the output line for ac-only coupling.

Quadrature detection is more common than frequency discrimination in current receivers because it doesn't require a special discriminator transformer or resonator, and because the necessary balanced-detector circuitry can easily be implemented in IC structures along with limiters and other receiver circuitry. The NXP Semiconductor SA604A FM IF IC is one example of this; **Figure 12.55** shows another, the Freescale Semiconductor (formerly Motorola) MC3359 (equivalent, NTE860). Quadrature detection is also simple to perform in SDR designs.

AMPLITUDE LIMITING

FM radio communication systems are superior to AM in their ability to suppress and ignore static, manmade electrical noise and (through a characteristic called *capture effect*) co-channel signals sufficiently weaker than the desired signal. AM-noise immunity and capture effect are not intrinsic to angle modulation, however; they must be designed into the angle-modulation receiver in the form of signal amplitude *limiting*.

The amplitude of a quadrature detector's input signal affects the amplitude of a quadrature detector's three output signals. A quadrature detector therefore responds to AM, and so does a frequency discriminator. To achieve FM's excellent amplitude noise immunity, then, these angle demodulators must be preceded by *limiter* circuitry that removes all amplitude variations from the incoming signal.

Mixer Math: Quadrature Demodulation

Demodulating an angle-modulated signal merely by multiplying it with a time-delayed copy of itself in a double-balanced mixer results in quadrature demodulation (Figure 12.54). To illustrate this mathematically, for simplicity's sake, we'll represent the mixer's RF input signal as just a sine wave with an amplitude, *A*:

$$A \sin(2\pi ft)$$

and its time-delayed twin, fed to the mixer's LO input, as a sine wave with an amplitude, *A*, and a time delay of d:

$$A \sin[2\pi f(t + d)]$$

Setting this special mixing arrangement into motion, we see

$$A \sin(2\pi ft) \times A \sin(2\pi ft + d)$$

$$= \frac{A^2}{2}\cos(2\pi fd) - \frac{A^2}{2}\cos(2\pi fd)\cos(2 \times 2\pi ft) + \frac{A^2}{2}\sin(2\pi fd)\sin(2 \times 2\pi ft)$$

Two of the three outputs — the second and third terms — emerge at twice the input frequency; in practice, we're not interested in these, and filter them out. The remaining term — the one we're after — varies in amplitude and sign according to how far and in what direction the carrier shifts away from its resting or center frequency (at which the time delay, d, causes the mixer's RF and LO inputs to be exactly 90° out of phase — in *quadrature* — with each other). We can examine this effect by replacing f in the RF input and LO input sinusoids with the sum term $f_c + f_s$, where f_c is the center frequency and f_s is the frequency shift. A 90° time delay is the same as a quarter cycle of f_c, so we can restate d as

$$d = \frac{1}{4f_c}$$

The first term of the detector's output then becomes

$$\frac{A^2}{2}\cos\left(2\pi(f_c + f_s)d\right)$$

$$= \frac{A^2}{2}\cos\left(2\pi(f_c + f_s)\frac{1}{4f_c}\right)$$

$$= \frac{A^2}{2}\cos\left(\frac{\pi}{2} + \frac{\pi f_s}{2f_c}\right)$$

When f_s is zero (that is, when the carrier is at its center frequency), this reduces to

$$\frac{A^2}{2}\cos\left(\frac{\pi}{2}\right) = 0$$

As the input signal shifts higher in frequency than f_c, the detector puts out a positive dc voltage that increases with the shift. When the input signal shifts lower in frequency than fc, the detector puts out a negative dc voltage that increases with the shift. The detector therefore recovers the input signal's frequency or phase modulation as an amplitude-varying dc voltage that shifts in sign as f_s varies around f_c — in other words, as ac. We have demodulated FM by means of quadrature detection.

SDR ANGLE DEMODULATION

A popular technique for angle demodulation starts with a phase detector as shown in **Figure 12.56**. The angle-modulated signal is down-converted to a convenient frequency and input to a pair of multipliers where it is mixed with a pair of constant frequency LO signals in quadrature. Low-pass filters then remove signals at the sum frequency, leaving only the I and Q baseband signals.

In PM demodulation, the phase comparison is an absolute comparison against the carrier phase. The phase angle of the input signal, relative to the LO center frequency, can be determined from the I and Q signals as:

$$\varphi = \tan^{-1}\left(\frac{V_Q}{V_I}\right)$$

The arctan function can be computed either by polynomial approximations or a lookup table, depending on system resources. If the signal is a PM signal, the phase signal contains the original modulation and can be filtered and output as audio.

FM demodulation requires additional steps as a special case of PM. The change in angle from the unmodulated carrier is constantly increasing or decreasing and exceeds ±90° as a normal part of operation. For that reason, a simple phase comparator using the recovered carrier will not demodulate FM. FM requires that we measure the rate-of-change of phase and sum those incremental phase changes.

Since frequency is defined as the rate-of-change of phase, it is necessary to differentiate the phase signal to recover the FM modulation signal. Implementing differentiation in DSP involves subtracting each sample from the previous sample. It is necessary to account for the point at which phase passes through 360º and resets to 0º. This can be done through special routines or by scaling of fixed-point data in the computations so that the phase rollover and integer rollover (65535 to 0, for example, in 16-bit math) coincide.

The FM signal will also need a de-emphasis filter (see the **Modulation** and **Transmitting** chapters) to cancel the high-frequency gain that was added during the FM process by the modulator.

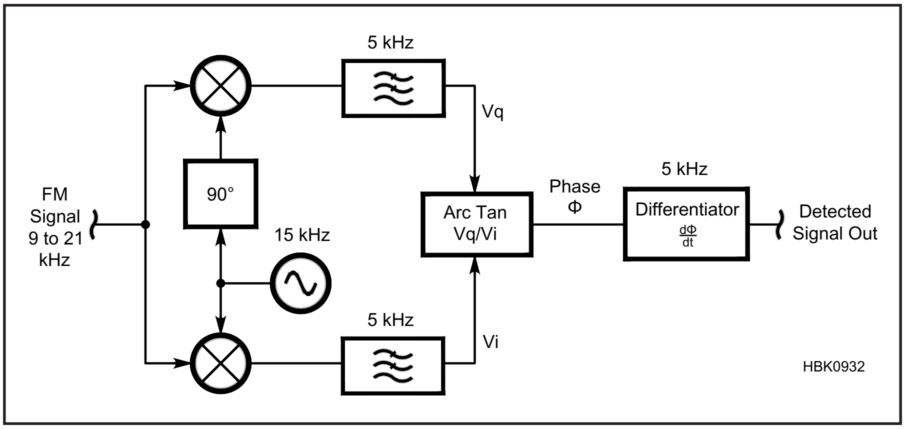


Figure 12.56 — An FM detector using an arctangent phase detector and a differentiator.

12.6 Automatic Gain Control (AGC)

The amplitude of the desired signal at each point in the receiver is generally controlled by the AGC system, although manual control is usually provided as well. Each stage has a distortion versus signal level characteristic that must be known, and the stage input level must not become excessive. The signal being received has a certain signal-to-distortion ratio that must not be degraded too much by the receiver. For example, if an SSB signal has –30 dB distortion products the receiver should create additional distortion no greater than –40 dB with respect to the desired signal. The correct AGC design ensures that each stage gets the right input level. It is often necessary to redesign some stages in order to accomplish this task.

While this chapter deals mostly with AGC in the guise of analog circuits, the same function is also implemented digitally in DSP and SDR receivers. The goal of both is the same — to maintain a signal level at all stages of the receiver that is neither too large nor too small so that the various processing systems operate properly. Whether or not the AGC offset and time constant are implemented by an analog component or by a microprocessor output is immaterial. The point is to manage the RF amplifier gain so that the overall receiver behavior is satisfactory.

AGC in the receive signal chain of an SDR should not be confused with AGC in an analog receiver. AGC in the analog receiver will adjust the fixed dynamic range of the receiver up or down, altering the weak and strong signal performance of the radio dynamically.

The effects of an improperly operating AGC system can be quite subtle or nearly disabling to a receiver and vary with how the AGC system is constructed. This chapter attempts to describe the requirements for proper operation and provides some examples of implementation and common AGC failures in terms of analog circuitry which is somewhat easier to describe than software algorithms, noting that similar behaviors exist even in purely software receivers. The interested student should consider studying the AGC systems of commercial receivers to understand how professional design teams deal with the problem of managing so much gain with such stringent requirements for linearity and distortion.

THE AGC LOOP

Figure 12.57A shows a typical AGC loop that is often used in amateur superhet receivers. The AGC is applied to the stages through RF decoupling circuits that prevent the stages from interacting with each other. The AGC amplifier helps to provide enough AGC loop gain so that the gain-control characteristic of Figure 12.57B is achieved. If effect, the AGC system causes the receiver to act as a compression amplifier with lower overall gain for stronger signals.

The AGC action does not begin until a certain level, called the AGC *threshold*, is reached. The THRESHOLD VOLTS input in Figure 12.57A serves this purpose. After that level is exceeded, the audio level increases more slowly than for weaker signals. The audio rise beyond the threshold value is usually in the 5 to 10 dB range. Too much or too little audio rise are both undesirable for most operators.

As an option, the AGC signal to the RF amplifier is offset by the 0.6 V forward drop of the diode so that the RF gain does not start to decrease until larger signals appear. This prevents a premature increase of the receiver noise figure. Also, a time constant of one or two seconds after this diode helps keep the RF gain steady for the short term.

Figure 12.58 is a typical plot of the signal levels at the various stages of a certain ham band receiver using analog circuitry. Each stage has the proper level and a 115 dB change in input level produces a 10 dB change in audio level. A manual gain control could produce the same effect.

AGC PUMPING

AGC *pumping* can occur in receivers in which the AGC measurement point is located ahead of the stages that determine operating bandwidth, such as when an audio filter is added to a receiver externally and outside the reach of the AGC system. If the weak signal is the only signal within the first IF passband, the AGC will cause the receiver to be at maximum gain and optimum SNR. If an interfering signal is within the first IF passband, but outside the audio DSP filter's passband, we won't hear the interfering signal, but it will enter the AGC system and reduce the gain so we might not hear our desired weak signal. AGC pumping is audible as sudden reductions in signal strength without a strong signal in the passband of the receiver.

AGC pumping is generally not as much of a problem in SDR receivers as it is in heterodyne receivers but the phenomena still exists. The severity depends on the algorithm employed by the SDR and the operating configuration controlled by the user.

AGC TIME CONSTANTS

There are two primary AGC time constants. AGC *attack time* describes the time it takes the AGC system to respond to the presence

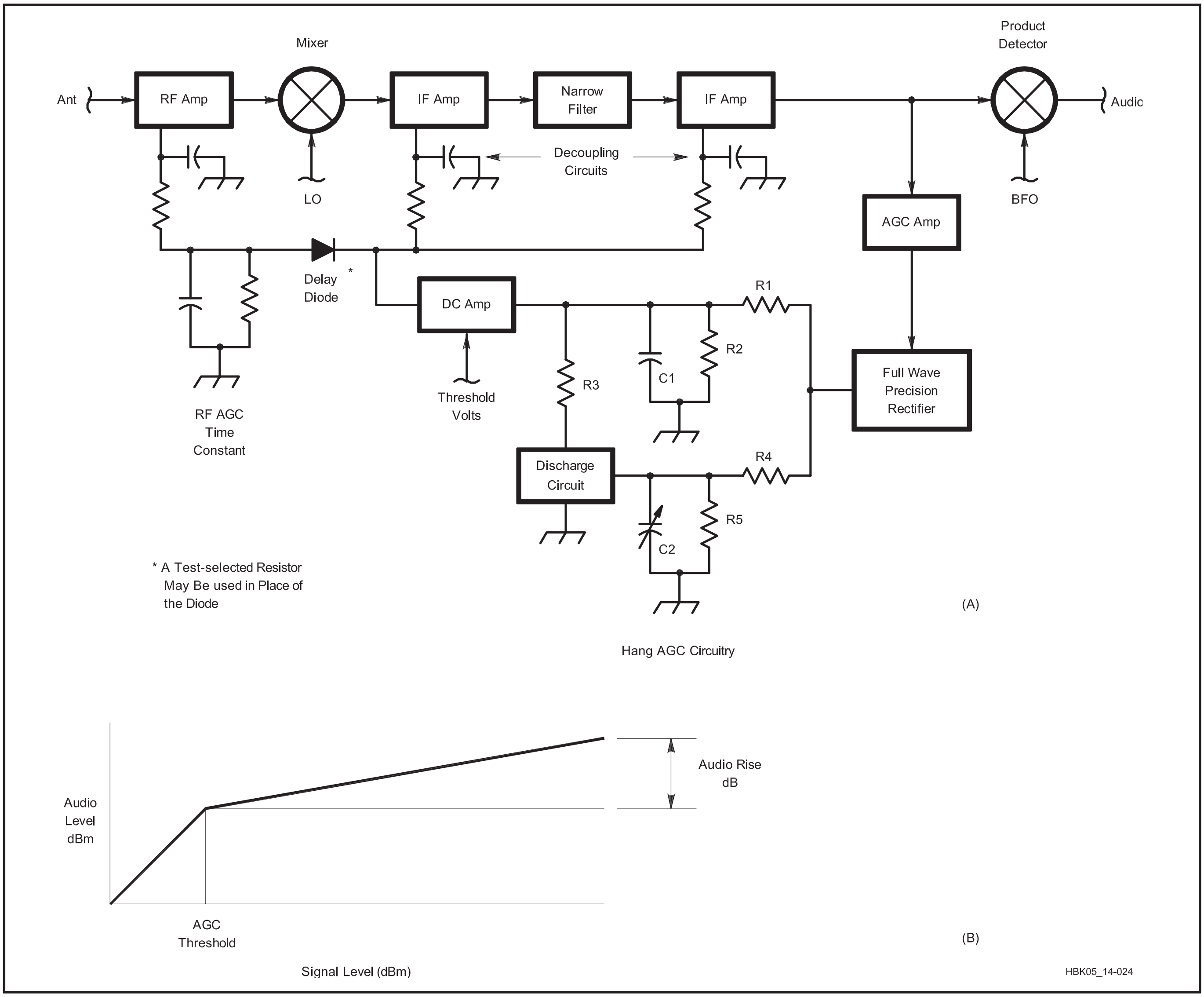


Figure 12.57 — AGC principles. At A: typical superhet receiver with AGC applied to multiple RF and IF stages. At B: audio output level as a function of antenna signal level.

of a signal. AGC *decay time* describes the response of the AGC system to changes in a signal that is present. The optimum time constants for the AGC system depends on the type of signal being received, the type of operation being conducted, and the operator's preference.

The operator usually has a control that allows for setting the time that it takes the AGC to recover or decay. If we are listening to two relatively loud stations converse, we may set the AGC to slow. Then as each station stops transmitting, the noise floor is slow to rise and we often hear the next station before the noise floor is again heard. In contrast, we may set the AGC decay time to fast which allows us to hear a weak station immediately after the strong station.

In Figure 12.57, following the precision rectifier, R1 and C1 set an attack time, to prevent excessively fast application of AGC. One or two milliseconds is a good value for the R1 × C1 product. If the antenna signal suddenly disappears, the AGC loop is opened because the precision rectifier stops conducting. C1 then discharges through R2 and the C1 × R2 product can be in the range of 100 to 200 ms. At some point the rectifier again becomes active, and the loop is closed again.

An optional modification of this behavior is the *hang AGC* circuit. If we make R2 × C1 much longer, say 3 seconds or more, the AGC voltage remains almost constant until the R5, C2 circuit decays with a switch selectable time constant of 100 to 1000 ms. At that time R3 quickly discharges C1 and full receiver gain is quickly restored. This type of control is appreciated by many operators because of the lack of AGC pumping due to modulation, rapid fading and other sudden signal level changes.

In an SDR receiver, the typical AGC algorithm has a fast-attack and slow-decay that is adjustable. The AGC algorithm has a gain value that it applies to the receiver buffers which persists from one buffer to the next. If a loud signal appears while the gain is set high, it immediately lowers the gain to prevent a loud sound in our headphones. Once the sound has subsided, the gain is allowed to slowly increase back to its previous value.

AGC LOOP RESPONSE PROBLEMS

If the various stages have the property that each 1 V change in AGC voltage changes the gain by a constant amount (in dB), the AGC loop is said to be *log-linear* and regular feedback principles can be used to analyze and design the loop. But there are some difficulties that complicate this textbook model. One has

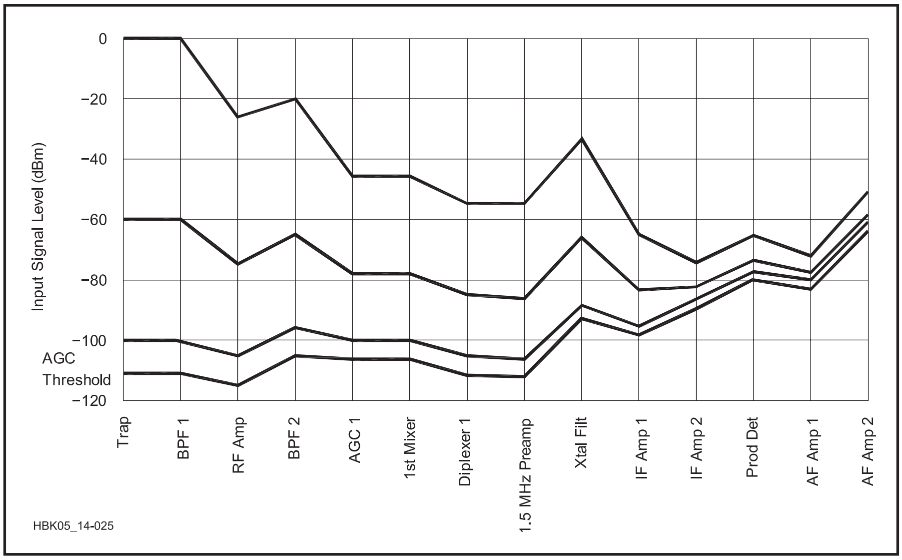


Figure 12.58 — Gain control of a ham-band receiver using AGC. A manual gain control could produce the same result.

already been mentioned, that when the signal is rapidly decreasing the loop becomes open and the various capacitors discharge in an open loop manner. As the signal is increasing beyond the threshold, or if it is decreasing slowly enough, the feedback theory applies more accurately.

In SSB and CW receivers rapid changes are the rule and not the exception. It is important that the AGC loop not overshoot or ring when the signal level rises past the threshold. The idea is to design the ALC loop to be stable when the loop is closed. It obviously won't oscillate when open (during decay time). But the loop must have smooth and consistent transient response when the loop goes from open to closed state.

Another problem involves the narrow band-pass analog IF filters. The group delay of analog filters constitutes a time lag in the loop that can make loop stabilization difficult. Moreover, these filters nearly always have much greater group delay at the edges of the passband, so that loop problems are aggravated at these frequencies. Overshoots and undershoots, called *gulping*, are very common. Compensation networks that advance the phase of the feedback help to offset these group delays. The design problem arises because some of the AGC is applied before the filter and some after the filter. It is a good idea to put as much fast AGC as possible after the filter and use a slower decaying AGC ahead of the filter. The delay diode and RC in Figure 12.57A are helpful in that respect. Complex AGC designs using two or more compensated loops are also in the literature. If a second cascaded narrow filter is used in the IF it is usually a lot easier to leave the second or *downstream* filter out of the AGC loop at the risk of allowing AGC pumping as described in the preceding section.

Another problem is that the control characteristic is often not log-linear. For example, dual-gate MOSFETs tend to have much larger dB/V at large values of gain reduction. Many IC amplifiers have the same problem. The result is that large signals cause instability because of excessive loop gain. Variable gain op amps and other similar ICs are available that are intended for gain control loops.

Audio frequency components on the AGC bus can cause problems because the amplifier gains are modulated by the audio and distort the desired signal. A hang AGC circuit (essentially a low-pass filter) can reduce or eliminate this problem.

Finally, if we try to reduce the change in audio levels to a very low value, the required loop gain becomes very large, and stability problems become very difficult. It is much better to accept a 5 to 10 dB variation of audio output.

Because many parameters are involved and many of them are not strictly log-linear, it is best to achieve good AGC performance through an initial design effort and finalize the design experimentally. Use a signal generator, attenuator and a signal pulser (2 ms rise and fall times, adjustable pulse rate and duration) at the antenna and a synchronized oscilloscope to look at the IF envelope. Tweak the time constants and AGC distribution by means of resistor and capacitor decade substitution boxes. Be sure to test throughout the passband of each filter. The final result should be a smooth and pleasant sounding SSB/CW response, even with maximum RF gain and strong signals. Patience and experience are helpful.

12.6.1 Audio-Derived AGC

Some receivers, especially direct-conversion types, use audio-derived AGC. There are problems with this approach as well. At low audio frequencies the AGC control action can be slow to develop. That is, low-frequency audio sine waves take longer to reach their peaks than the AGC time constants. During this time the RF/IF/AF stages can be overdriven. If the RF and IF gains are kept at a low level this problem can be reduced. Also, attenuating low audio frequencies prior to the first audio amplifier should help. With audio AGC, it is important to avoid so-called "charge pump" rectifiers or other slow-responding circuits that require multiple cycles of audio to pump up the AGC voltage. Instead, use a peak-detecting circuit that responds accurately on the first positive or negative half-cycle.

12.6.2 AGC Circuits

Figure 12.59 shows some gain-controllable circuits. Figure 12.59A shows a two-stage 455-kHz IF amplifier with PIN diode gain control. This circuit is a simplified adaptation from a production receiver, the Collins 651S. The IF amplifier section shown is preceded and followed by selectivity circuits and additional gain stages with AGC. The 1.0 µF capacitors aid in loop compensation. The favorable thing about this approach is that the transistors remain biased at their optimum operating point. Right at the point at which the diodes start to conduct, a small increase in IMD may be noticed, but that goes away as diode current increases slightly. Two or more diodes can be used in series, if this is a problem (it very seldom is). Another solution is to use a PIN diode that is more suitable for such a low-frequency IF. Look for a device with $\tau > 10 / (2\pi f)$ where τ is the minority carrier lifetime in ms and f is the frequency in MHz.

Figure 12.59B is an audio derived AGC circuit using a full-wave rectifier that responds to positive or negative excursions of the audio signal. The RC circuit follows the audio closely.

Figure 12.59C shows a typical circuit for the MC1350P RF/IF amplifier. The graph of gain control versus AGC voltage shows the change in dB/V. If the control is limited to the first 20 dB of gain reduction this chip should be favorable for good AGC transient response and good IMD performance. Use multiple low-gain stages rather than a single high-gain stage for these reasons. The gain control within the MC1350P is accomplished by diverting signal current from the first amplifier stage into a *current sink*. This is also known as the *Gilbert cell multiplier* architecture. Another chip of this type is the NE/

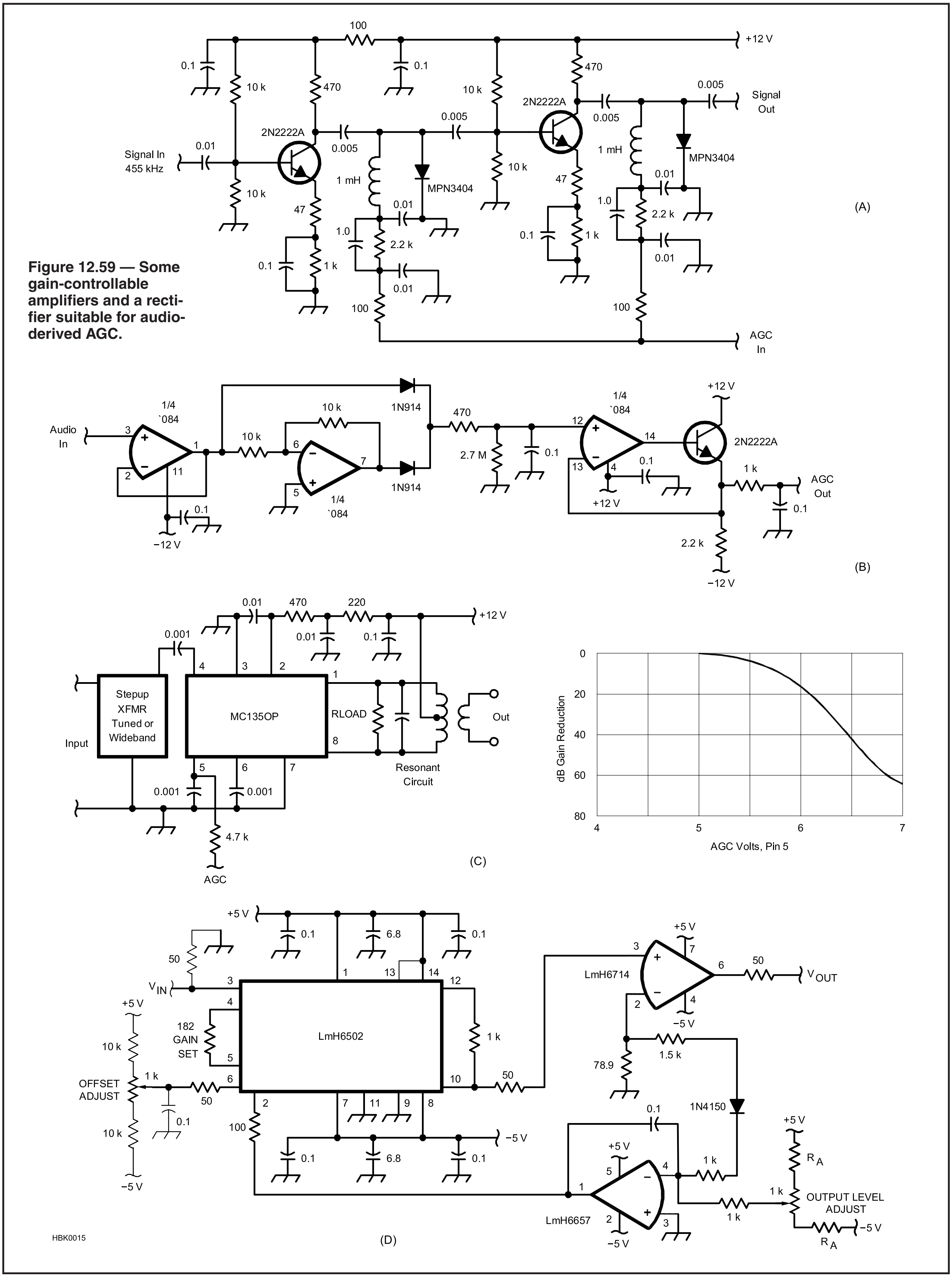


Figure 12.59 — Some gain-controllable amplifiers and a rectifier suitable for audio-derived AGC.

SA5209. This type of approach is simpler to implement than discrete circuit approaches, such as dual-gate MOSFETs that are now being replaced by IC designs.

Figure 12.59D shows the high performance National Semiconductor LMH6502MA (14-pin DIP plastic package) voltage controlled amplifier. It is specially designed for accurate log-linear AGC from 0 to 40 dB with respect to a preset maximum voltage gain from 6 to 40 dB. Its ±3 dB bandwidth is 130 MHz. It is an excellent IF amplifier for high performance receiver or transmitter projects.

Additional info on voltage-controlled amplifier ICs can be found on the Analog Devices web site (**www.analog.com**). Search the site for Tutorial MT-073, which describes the operation of various types of gain-controlled amplifiers with numerous product examples.

12.7 Noise Management

A major problem for those listening to receivers has historically been local impulse noise. For HF and VHF receivers it is often from the sparks of internal combustion engine spark plugs, electric fence chargers, light dimmers, faulty power-line insulators and many other similar devices that put out short duration wide band signals. In the UHF and microwave region, radar systems can cause similar problems.

Additional sources of noise are atmospheric and man-made noise with a variety of different profiles — static crashes, power-line buzz, and the ever-increasing white noise and spurious signals from consumer and industrial electronics, particularly switch-mode devices. The capabilities of DSP can be used to combat these diverse types of noise.

Finally, noise canceling by subtracting it from the incoming signals is available as a station accessory using external antennas for sensing and beam-steering. True diversity reception is also available, pioneered for the amateur station, using the spatial characteristics of the arriving noise signals to discriminate between them and the desired signals. All of these techniques provide formidable tools for noise management that were simply unavailable to amateurs of earlier eras.

12.7.1 The Noise Limiter

The first device used in an early (1930s) attempt to limit impulse noise was called a *noise limiter* or *clipper* circuit as originally described by H. Robinson, W3LW. This circuit would *clip* or limit noise (or signal) peaks that exceeded a preset limit. The idea was to have the limit set to about as loud as you wanted to hear anything and nothing louder would get through. This was helpful in eliminating the loudest part of impulse noise or even nearby lightning crashes, but it had two problems. First it didn't eliminate the noise, it just reduced the peak loudness; second, it also reduced the loudness of loud non-noise signals and in the process distorted them considerably.

The second problem was fixed shortly thereafter, with the advent of the *automatic noise limiter* or ANL as described by J. Dickert. The ANL automatically set the clipping threshold to that of a loud signal. It thus would adjust itself as the loudness of signals you listened to changed with time. An ANL was fairly easy to implement and became standard equipment on amateur receivers from the late 1930s on. While ANL circuits are no longer common, simple receivers used today do sometimes incorporate passive clipping circuits to account for their limited AGC ability.

12.7.2 The Noise Blanker

It turned out that improvements in receiver selectivity over the 1950s and beyond, while improving the ability to reduce random noise, actually made receiver response to impulse noise worse. The reason for this is that a very short duration pulse will actually be lengthened while going through a narrow filter. This is due to the filter's different delay times for the pulse's wide spectrum of components, resulting in the components arriving at the filter output at different times. You can demonstrate this in your superhet receiver if it has a narrow crystal filter. Find a frequency with heavy impulse noise and switch between wide and narrow filters. If your narrow filter is 500 Hz or less, the noise pulses will likely be more prominent with the narrow filter. DSP filters with their superior group delay performance exhibit less smearing than their analog counterparts.

The noise limiters described previously were all connected at the output of the IF amplifiers and thus the noise had passed most of the selectivity before the limiter and had been widened by the receiver filters. In SSB receivers, since signals vary in strength as someone talks, the usual AGC responds quickly to reduce the gain of a strong signal and then slowly increases it if the signal is no longer there. This means that a strong noise pulse may reduce the receiver gain for much longer than it lasts.

The solution — a *noise blanker*. An analog noise blanker is almost a separate wideband receiver. It takes its input from an early stage in the heterodyne receiver before much selectivity or AGC has been applied. It amplifies the wideband signal and detects the narrow noise pulses without lengthening them. The still-narrow noise pulses are used to shut off or "blank" the receiver at a point ahead of the selectivity and AGC, thus keeping the noise from getting to the parts of the receiver at which the pulses would be extended. In other words, the receiver is shut off or *gated* during the noise pulse.

In addition to an ON/OFF switch, noise blankers include a control labeled THRESHOLD. The THRESHOLD control adjusts the level of noise that will blank the receiver. If it is set for too low a level, it will blank on signal peaks as well as noise, resulting in distortion of the signal.

An SDR noise blanker can implement the same detect-and-gate technique as an analog receiver. Once the noise impulse is detected, the samples making up the impulse are replaced by an interpolation between the "normal" samples just before and just after the impulse. It is as is the noise impulse was never received.

An alternative technique that also handles longer noise crashes and other large-amplitude, long-duration noise waveforms is described in **Figure 12.60**. The basic idea is to create a filtered average amplitude and watch for incoming signals that exceed the average multiplied by a weighting factor which is how the SDR noise blanker threshold is controlled. When the incoming signal is below this value, it is simply passed along to the rest of the receiver. Above this level, the blanker generates a ramp factor that limits the rate at which the signal can increase. This turns short, sharp impulses into relatively slowly increasing and decreasing ramps, leaving normal signals unaffected.

A well-designed noise blanker can be very effective. Instead of just keeping the noise at the level of the signal as a noise limiter does, the noise blanker can actually *eliminate* the noise. If the pulses are narrow enough, the loss of desired signal during the time the receiver is disabled is not noticeable and the noise may seem to disappear entirely.

Noise blankers can also create problems, particularly in heterodyne receivers. The wide-band circuit that detects noise pulses also detects any signals in that bandwidth. If such a signal is strong and has sharp peaks (as voice and CW signals do), the noise

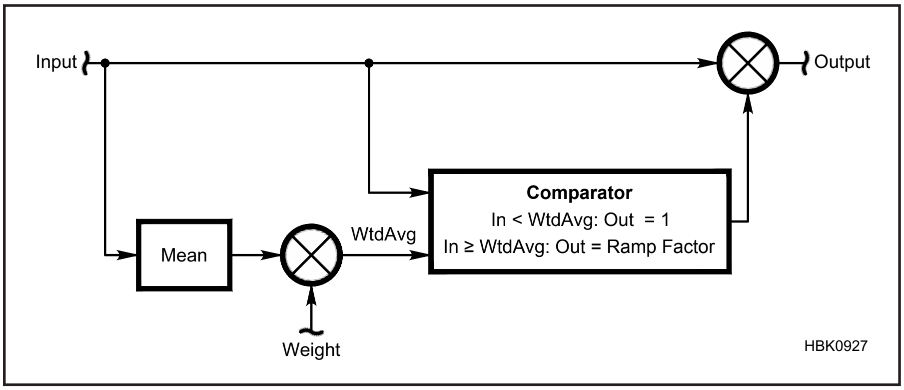


Figure 12.60 — Block diagram of a noise blanker. The input signal is compared to a weighted average. If exceeded, the input signal rise and fall times are limited by a ramp factor. [After *Communications Receivers, 4th ed*]

blanker will treat them as noise pulses and shut down the receiver accordingly. This causes tremendous distortion and can make it sound as if the strong signal to which the noise blanker is responding is generating spurious signals that cause the distortion. (This is less of a problem in SDR receivers.) Before you assume that the strong signal is causing problems, turn the noise blanker on and off to check. When the band is full of strong signals, a noise blanker may cause more problems than it solves.

12.7.3 Operating Noise Limiters and Blankers

Many current receivers include both a noise limiter (often labeled NL) and a noise blanker (labeled NB). If your receiver has both, they will have separate controls and it is worthwhile to try them both. There are times at which one will work better than the other, and other times when it goes the other way, depending on the characteristics of the noise. There are other times when both work better than either. In any case, they can make listening a lot more pleasant — just remember to turn them off when you don't need them since either type can cause some distortion, especially on strong signals that should otherwise be easy to listen to.

Recognizing that it is difficult for a single noise blanker to work properly with the wide variations of noise pulses, it is common for current receivers to have two noise blankers with different characteristics that are optimized for the different pulse types. One noise blanker is typically optimized for very short pulses and the other for longer pulses. The operator can switch between the blankers to see which works best on the noise at hand.

The usual approach to operating the noise blanker is to activate it, then adjust the THRESHOLD control until the noise is just blanked. You will probably need to make occasional adjustments as the noise impulse characteristics change. Don't forget to turn the noise blanker off when the noise goes away.

The previous techniques represent the most commonly available techniques to reduce impulse noise. There are a few other solutions as well. Note that we haven't been talking about reducing interference here. By interference, we mean another intended signal encroaching on the channel to which we want to listen. There are a number of techniques to reduce interference, and some also can help with impulse noise.

Many times impulse noise is coming from a particular direction. If so, by using a directional antenna, we can adjust the direction for minimum noise. When we think about directional antennas, the giant HF Yagi springs to mind. For receiving purposes, especially on the lower bands such as 160, 80 and 40 meters (where the impulse noise often seems the worst), a small indoor or outdoor receiving loop antenna as described in the *ARRL Antenna Book* can be very effective at eliminating either interfering stations or noise (both if they happen to be in the same direction).

12.7.4 Noise Canceling

Noise cancellers work by combining signals from our main antenna with the signal from a "sense" antenna and feeding that combination to the receiver. Adjustable gain stages and phasing networks within the unit must then be carefully adjusted so that the two noise signals are equal in level and 180 degrees out of phase. This adjustment is frequency sensitive, so it must be readjusted each time we change frequency. It must also be readjusted for every noise source. Signals coming from the same direction as the noise source will also be canceled, but this effect is minimized by placing the sense antenna as close as practical to the noise source. An active noise canceller can greatly reduce a single noise source, but it won't help with more than one source at a time. **Figure 12.61** shows the block diagram of a typical noise canceller and antennas, in this case the DX Engineering NCC-2.

Be careful when using any unit in line with

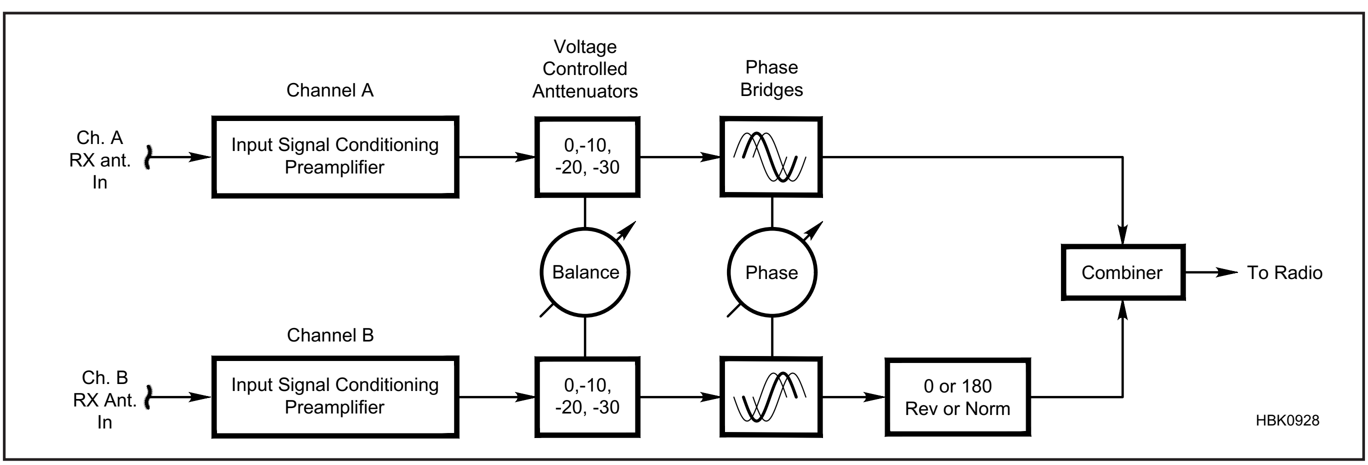


Figure 12.61 — Block diagram of a noise canceling system following the DX Engineering NCC-2. [Courtesy DX Engineering, used by permission]

the transceiver output; the carrier detect-driven relay that switches from receive to transmit (that is, bypassing the unit in transmit mode) is known to generate key clicks (transients) in some products. The MFJ 1026 and the DX Engineering NCC-1 and NCC-2 are generally good performers in this regard.

Some transceivers provide connectors where an external preamp or receive antenna can be connected. The click problem can be avoided by connecting the noise canceller at this point. Some SDRs have multiple antenna inputs and the ability to feed them to separate software-defined "receivers", and some software has been written to allow those separate receivers to provide the phasing networks needed to provide noise canceling. In some software, this function is improperly labeled "Diversity Receive." (See the Diversity Reception description below.)

BEAM STEERING

Noise cancellers can also be used to vary the phase of the signals from two receive antennas to vary their directivity. Two vertical antennas, connected to a receiver single receiver by individual feed lines, form a directional antenna. The directional pattern depends on the phase shift provided by the two feed lines (as well as the physical spacing and spatial orientation of the antennas), and it can be adjusted by varying the length of one or both of the feed lines. Used with two omnidirectional verticals, for example, a noise canceller can be used to vary the phase relationship between the two antennas, accomplishing the same result.

SIGNAL FADING

Most periodic (repetitive) fading is the result of the cancellation of signals from the same source taking slightly different paths, so that one is delayed with respect to the other. The direct and reflected signals cancel when they are nearly equal and nearly 180 degrees out of phase with each other. The frequency(ies) at which this 180-degree phase relationship exists is a function of the time difference; the fading interval is longer for lower frequencies. This fading mechanism is heard as very slow fading on the AM broadcast band and 160 meters, and as "picket fencing" at VHF and UHF.

12.7.5 Diversity Reception

Diversity reception has been used since the earliest days of radio to reduce the effects of signal fading. A receiving diversity system consists of two receivers, each connected to its own antenna and to its own loudspeaker (or opposite ears of stereo headphones). When the two antennas are widely separated as a fraction of a wavelength, the probability of cancellation occurring at both antennas at the same time is low, and the operator copies from the receiver providing the best signal.

Diversity reception is also widely used to listen using different receive antennas having different directivity in the horizontal or vertical plane, or aimed in different directions. This use is quite common on the lower HF bands and 160 meters. An operator may listen to the transmit antenna in one ear and a Beverage in the other, or to a loop and a Beverage.

Diversity reception is widely used in consumer FM receivers in vehicles, and in the wireless microphone systems used in live sound and broadcasting. In both of these applications, a circuit called a "voter" chooses the signal from the receiver having the best quality and switches it to the output. Many VHF and UHF repeater systems have receivers at multiple sites that are relayed back to the main transmit site, where a voter chooses between the best signal. Diversity is also used with receive antennas having different wave polarization; cross-polarization between a receive antenna and the wavefront typically results in a 20 dB loss of gain.

12.7.6 DSP Noise Reduction

DSP noise reduction (often labeled NR on the receiver controls) can actually look at the statistics of the signal and noise and figure out which is which and then reduce the noise significantly. These *adaptive filters* can't quite eliminate the noise, and need enough of the desired signal to figure out what's happening, so they won't work if the signal is far below the noise. (See the **DSP and SDR Fundamentals** and **Analog and Digital Filtering** chapters for information on adaptive filters.) Many DSP systems "color" the resulting audio to a degree. Nonetheless, they do improve the SNR of a signal in random or impulse noise.

Noise reduction is designed to reduce non-correlated noise such as atmospheric noise and thermal noise. Most noise reduction blocks rely on the knowledge that noise is much more random in nature than the signals we are attempting to demodulate. A typical NR will use a Least Mean Squares (LMS) or similar algorithm to detect correlated signals and reduce the amplitude of anything else in the passband.

How does this work? The algorithm looks for signals that are highly correlated, meaning that a copy of the signal shifted in time resembles the same signal at a different time. These are the signals that are preserved while the other signals in the passband are judged to be noise and are deemphasized or reduced in volume. Because there are a number of methods for achieving noise reduction in different classes of noise, some vendors often more than one type of noise reduction. Each manufacturer's algorithms for noise reduction are generally proprietary although the *GNU Radio* community's (see the **Transceiver Design Topics** chapter) open-source design approach offers some guidance in how noise reduction is achieved.

As with noise blankers, receivers frequently offer two or more noise reduction settings that apply different noise reduction algorithms optimized for different conditions. Different combinations and types of noise require the operator to select and adjust the noise reduction system for best performance.

12.7.7 DSP Notch Filtering

An Automatic Notch Filter (ANF) is really the opposite of a NR filter. The auto-notch filter looks for the same correlated signals, but instead or preserving these signals, the algorithm attempts to remove these signals. The classic example for which an ANF is useful is when there is a birdie (spur) or undesired CW signal in the passband of a sideband signal. Because the CW signal or spur are much more correlated than the speech we are listening to, these signals are judged to be undesired and the filter deemphasizes them. In this way, the NR and ANF filters are cousins, doing the opposite of each other in support of their specific noise reduction goals.

12.8 References and Bibliography

BOOKS

Allison, B., WB1GCM, *Amateur Radio Transceiver Performance Testing* (ARRL, 2013).

Hallas, J., W1ZR, "International Radio Roofing Filters for the Yaesu FT-1000 MP Series Transceivers," Product Reviews, *QST*, Feb. 2005, pp. 76.

Hayward, W., W7ZOI*, Introduction to Radio Frequency Design* (ARRL, 1994).

Hayward, W., Campbell, R., and Larkin, B., *Experimental Methods in RF Design* (ARRL, 2009).

Horowitz, P., Hill, W., *The Art of Electronics*, 2nd Edition (Cambridge University Press, 1989).

Larson, L., *RF and Microwave Circuit Design for Wireless Communication* (Artech House, 1996).

Maas, S., *The RF and Microwave Circuit Design Cookbook* (Artech House, 1998).

Maas, S., *Nonlinear Microwave and RF Circuits*, 2nd Edition (Artech House, 2003).

Maas, S., *Noise in Linear and Nonlinear Circuits* (Artech House, 2005).

McClanning, K., and Vito T., *Radio Receiver Design* (Noble Publishing Corporation, 2001).

RF/IF Designer's Handbook (Scientific Components, 1992).

RF Communications Handbook (Philips Components-Signetics, 1992).

Rohde, U., Newkirk, D., *RF/Microwave Circuit Design for Wireless Applications* (John Wiley & Sons, 2000). (Large number of useful references in the mixer chapter.)

Rohde, U., Whitaker, J., and Zahnd, H., *Communications Receivers*, 4th Edition (McGraw-Hill, 2017). (Large number of useful references in each chapter, as well.)

Synergy Microwave Corporation Product Handbook (Synergy Microwave Corporation, 2007).

Terman, F., *Radio Engineers' Handbook* (McGraw-Hill, 1943).

ARTICLES ABOUT MIXERS

Henderson, B., "Mixers in Microwave Systems (Part 1)", Watkins-Johnson Tech Note, Vol. 17 No. 1, 1990, Watkins-Johnson.

Henderson, B., "Mixers in Microwave Systems (Part 2)", Watkins-Johnson Tech Note, Vol. 17 No. 2, 1990, Watkins-Johnson.

Gilbert, B., "Demystifying the Mixer," (self-published monograph, 1994).

Joshi, S., "Taking the Mystery Out of Double-Balanced Mixers," *QST,* Dec. 1993, pp. 32 – 36.

Kazdan D., "What's a Mixer?" *QST,* Aug. 1992, pp. 39 – 42.

Kearney F., and Frizelle, D., "Complex RF Mixers, Zero-IF Architecture, and Advanced Algorithms," *Analog Dialogue*, 51-02, Analog Devices.

Kossor, M., "A Digital Commutating Filter," *QEX*, May/June 1999, pp. 3 – 8.

Rohde, U., "Key Components of Modern Receiver Design," Part 1, *QST,* May 1994; Part 2, June 1994; Part 3, Jul. 1994.

Rohde, U., "Testing and Calculating Intermodulation Distortion in Receivers," *QEX,* July 1994, pp. 3 – 4.

Rohde, U., and Poddar, A., "High Intercept Point Broadband, Cost Effective and Power Efficient Passive Reflection FET DBM," *EuMIC Symposium*, 10 – 15 Sep. 2006, UK.

Rohde U., and Poddar, A., "A Unified Method of Designing Ultra-Wideband, Power-Efficient, and High IP3 Reconfigurable Passive FET Mixers," *IEEE International Conference on Ultra-Wide Band*, Sep. 24 – 27, 2006, MA, USA.

Rohde U., and Poddar, A., "Low Cost, Power-Efficient Reconfigurable Passive FET Mixers," 20th *IEEE Canadian Conference on Electrical and Computer Engineering* 2007, 22 – 26 Apr. 2007, British Columbia, Canada.

Tayloe, D., "Letters to the Editor, Notes on 'Ideal' Commutating Mixers (Nov/Dec 1999)," *QEX*, Mar./Apr. 2001, p. 61.

Tayloe, D., "A Low-noise, High-performance Zero IF Quadrature Detector/Preamplifier", Mar. 2003, *RF Design*.

Tayloe, D., US Patent 6,230,000, "A Product Detector and Method Therefore."

Vermusvuori, J., "A Synchronous Detector for AM Transmissions," *QST*, pp. 28 – 33. Uses SA602/612 mixers as phase-locked and quasi-synchronous product detectors.

Wiers, P., "Mirror, Mirror on the Wall — Understanding Image Rejection and Its Impact on Desired Signals," *Analog Dialogue*, 51-08, Analog Devices.

Zavrel, R., "Using the NE602," Technical Correspondence, *QST,* May 1990, pp. 38 – 39.

ARTICLES ABOUT RECEIVERS

ARRL Lab, "Test Procedures Manual," **www.arrl.org/product-review**.

Brannon, B., "Sampled Systems and the Effects of Clock Phase Noise and Jitter," Analog Devices, AN-756, **https://www.analog.com/media/en/technical-documentation/application-notes/AN-756.pdf**.

Brannon, B., "Some Recent Development in the Art of Receiver Technology," *Analog Dialogue*, 52-08, Analog Devices.

Farson, A., "Noise Power Ratio (NPR) Testing on HF Receivers," *QEX*, Mar. 2015, pp. 20 – 27.

Harris, J., "What's Up With Digital Downconverters — Part 1 and Part 2," *Analog Dialogue*, 50-07 and 50-11, Analog Devices.

Henderson, R., "A Rock Bending Receiver for 7 MHz," *QST*, Aug. 1995, pp. 22 – 25.

Hirshfield, G., "Ham Notebook: General-Coverage Preselector," *Ham Radio*, Oct. 1970, p. 77.

Kester, W., "Understand SINAD, ENOB, SNR, THD, THD+N, and SFDR so You Don't Get Lost in the Noise Floor," Analog Devices, MT-003 Tutorial, 2008, **www.analog.com/media/en/training-seminars/tutorials/MT-003.pdf**.

Kester, W., "Noise Power Ratio (NPR)," Analog Devices, MT-005 Tutorial, 2008, **www.analog.com/media/en/training-seminars/tutorials/MT-005.pdf**.

Lyons, R., "Understanding the 'Phasing Method' of Single Sideband Demodulation," **www.dsprelated.com/showarticle/176.php**.

Onate, J., Fortuny, X., "A Software-Controlled Preselector," *QEX*, May/June 2008, pp. 11 – 18.

Sabin, W., "Envelope Detection and AM Noise-Figure Measurement," *RF Design,* Nov. 1988, p. 29.

Shuch, P., N6TX, "Quiet! Preamp at Work," *ham radio*, Nov. 1984, pp. 14 – 20, **www.setileague.org/articles/ham/preamp.pdf**.

Silver, W., NØAX, "About FM," *QST*, Jul. 2004, pp. 38 – 42.

Silver, W., "About SSB," *QST*, Jan. 2016, pp. 51 – 54.

Smith, D., "Improved Dynamic Range Testing," *QEX*, Jul./Aug. 2002, pp. 46 – 52.

Smith, P., "Little-Known Characteristics of Phase Noise," Analog Devices, AN-741, **www.analog.com/media/en/technical-documentation/application-notes/58932485574881240369444855770343421704525427531621533025450669924401 6AN741_0.pdf**.