Contents

Chapter 11 — Online Content

Articles

- About FM by Ward Silver, NØAX
- About SSB by Ward Silver, NØAX
- Emissions Designator Table
- Modulation Glossary
- SDR Simplified — Introduction to I and Q by Ray Mack, W5IFS

Chapter 11
Modulation

Radio amateurs use a wide variety of modulations to convey information. This chapter, updated from Alan Bloom, N1AL's original material by Doug Grant, K1DG, explores various characteristics of modulation commonly used by amateurs. Traditional modulation types used for analog signals are discussed, as well as techniques suited for digital transmissions. Image modulations, updated for fast-scan ATV by Jim Andrews, KH6HTV, are also included, as well as pulse modulation as used by amateurs. Practical methods of using these modulations are presented in the chapters on **Receiving** and **Transmitting.** The chapter concludes with a discussion of modulation impairments and suggestions for additional reading.

11.1 Introduction

The purpose of amateur radio transmissions is to send information via radio. The one possible exception to that is a beacon station that transmits an unmodulated carrier for propagation testing. In that case the only information being sent is, "I am transmitting (or not) from this location." One can think of it as a single data bit with two states, *on* or *off*. However, in reality even a beacon station or test transmission must periodically identify with the station call sign! Modulation is what allows the signal to carry the information, no matter how much or little.

To represent the information being sent, the radio signal must periodically change its characteristics (or *state*) in some way that can be detected by the receiver. In the early days of radio, the only way to do that was on-off keying using Morse code. By alternating the on and off states with the proper sequence and rhythm, a pair of highly-skilled operators can exchange textual data at rates up to perhaps 60 WPM.

Later, engineers figured out how to amplify the signal from a microphone and use it to vary the power of the radio signal continuously. Thus was born amplitude modulation, which allowed transmitting voices at full speaking speeds, that is, up to about 200 WPM. That led to analog modes such as television with even faster information rates. Today's digital radio systems are capable of transferring tens of megabits of information per second, equivalent to tens of millions of words per minute.

Modulation is but one component of any transmission mode. For example AM voice and NTSC television (the old analog TV system) both use amplitude modulation, but the transmission protocol and type of information sent are very different. Digital modes also are generally defined not only by the modulation but also by multiple layers of protocol and data coding. Of the three components of a mode (modulation, protocol and information), this chapter will concentrate on the first. The **Digital Protocols and Modes** chapter covers digital transmission protocols and the **Digital Communications** operating chapter with the online content covers practical aspects of operating using the various digital modes.

Types of modulation are often referred to as "analog" or "digital." There is really no distinction between the two. What is usually meant is whether the modulating signal itself is an analog (continuously varying) or digital (some number of fixed states) signal.

The type of modulation is the choice of the radio system designer. The combination of modulation and a protocol create a mode. Modes can be analog or digital, again referring to whether they carry a continuously varying signal or not.

As an example of the difference between mode and modulation, consider HF and VHF packet radio. Both modes use the same AX.25 protocol to control how the data packets are formed and how the stations establish, conduct, and terminate the contact. Both modes also encode data as a pair of audio tones. On HF, however, the audio tones modulate an SSB transceiver to create an AFSK signal. On VHF, the tones modulate an FM transceiver, creating a very different type of signal. Both transceivers (SSB and FM) don't "know" whether the audio tones represent speech or data. If the audio is speech, the result is an "analog" mode. If the audio tones represent data, the result a "digital" mode.

11.1.1 Emission Designators

We tend to think of a radio signal as being "on" a particular frequency. In reality, any modulated signal occupies a band of frequencies. The bandwidth depends on the type of modulation and the data rate. A Morse code signal can be sent within a bandwidth of a couple hundred Hz at 60 WPM, or less at lower speeds. An AM voice signal requires about 6 kHz. For high-fidelity music, more bandwidth is needed; in the United States, an FM broadcast signal occupies a 200-kHz channel. Television signals need about 6 MHz, while 802.11ad, the latest generation of "WiFi" wireless LAN, uses up to 2 GHz for data transmission at its maximum transfer rate.

The International Telecommunication Union (ITU) has specified a system for designating radio emissions based on the bandwidth, modulation type and information to be transmitted. The emission designator begins with the bandwidth, expressed as a maximum of five numerals and one letter. The letter occupies the position of the decimal point and represents the unit of bandwidth, as follows: H = hertz, K = kilohertz, M = megahertz and G = gigahertz. The bandwidth is followed by three to five emission classification symbols, as defined in the table of emission designators available with the online content. The first three symbols are mandatory; the fourth and fifth symbols are supplemental. These designators are found in Appendix 1 of the ITU Radio Regulations, ITU-R Recommendation SM.1138 and in the FCC rules §2.201. More information on emissions designators is also available online at **fccid.io/Emissions-Designator**.

For example, the designator for a CW signal might be 150H0A1A, which means 150 Hz bandwidth, double sideband, digital information without subcarrier, and telegraphy for aural reception. SSB would be 2K5J3E, or 2.5 kHz bandwidth, single sideband with suppressed carrier, analog information, and telephony. The designator for a PSK31 digital signal is 60H0J2B, which means 60 Hz bandwidth, single sideband with suppressed carrier, digital information using a modulating subcarrier, and telegraphy for automatic reception.

Authorized modulation modes for amateur radio operators depend on frequency, license class, and geographical location, as specified in the FCC regulations §97.305. Technical standards for amateur emissions are specified in §97.307. Among other things, they require that no amateur station transmission shall occupy more bandwidth than necessary for the information rate and emission type being transmitted, in accordance with good amateur practice. We will discuss the necessary bandwidth for each type of modulation as it is covered in the following sections.

11.1.2 Bandwidth Definition

The general definition of bandwidth for a tuned circuit or filter are not used for the legal definition of signal bandwidth. The FCC regulations are more concerned with the amount of spectrum a signal consumes. This leads to the concept of *occupied bandwidth*. This is the range of frequencies within which a specified percentage of the total power occurs. A common percentage used is 99%. This means that the total signal power outside the occupied bandwidth must be less than 20 dB less than the total signal power. For a properly-adjusted, low-distortion transmitting system, the occupied bandwidth is determined mainly by the modulation type and filtering and, in the case of digital modulation, the symbol rate.

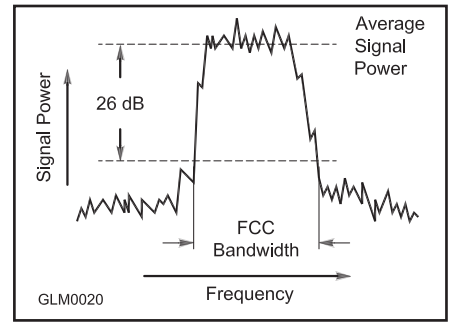

Figure 11.1 — The FCC definition of bandwidth (see text).

FCC Rule §97.3(a)(8) provides the legal definition of a signal's bandwidth as "The width of a frequency band outside of which the mean power of the transmitted signal is attenuated at least 26 dB below the mean power of the transmitted signal with the band." The 26 dB limit is equivalent to 1/400th of the signal's power. **Figure 11.1** illustrates this relationship.

11.2 Amplitude Modulation (AM)

Of the various properties of a signal that can be modulated to transmit voice information, amplitude was the first to be used. Not only are modulation and demodulation of AM signals simple in concept, but they are simple to implement as well.

11.2.1 Double-Sideband, Full-Carrier AM

An AM signal is created from two signals; the RF signal that can be transmitted and the modulating signal that will be combined with the RF signal. The RF signal is called the *carrier*, c(t), which is a single-frequency sinusoid at a frequency of f_C.

$$c(t) = C \sin (2\pi f_C t)$$

The modulating signal is represented by m(t) and may be a sine wave or a complex signal like speech. The modulating signal is also referred to as *baseband modulation*.

$$m(t) = M \cos (2\pi f_M t)$$

(The use of sine and cosine are to help identify which signal is which — there is no requirement for the carrier and modulating signal to have a specific phase relationship.)

Amplitude modulation is performed when c(t) is multiplied by m(t). Mathematically, the process of amplitude modulation is easiest to envision if the modulating signal, m(t) is

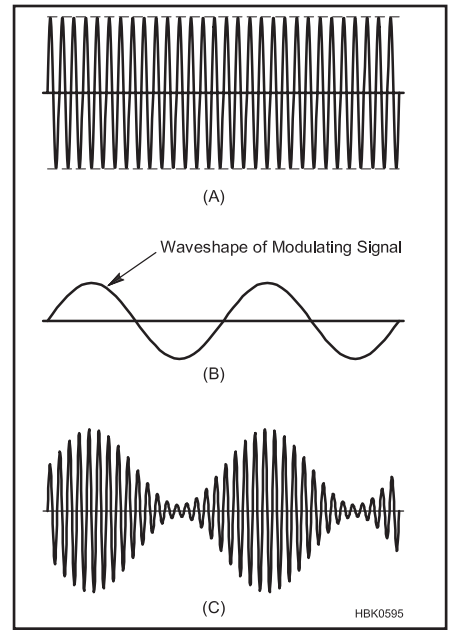

Figure 11.2 — Graphical representation of amplitude modulation. In the unmodulated carrier (A) each RF cycle has the same amplitude. When the modulating signal (B) is applied, the RF amplitude is increased or decreased according to the amplitude of the modulating signal (C). A modulation index of approximately 75% is shown. With 100% modulation the RF power would just reach zero on negative peaks of the modulating signal.

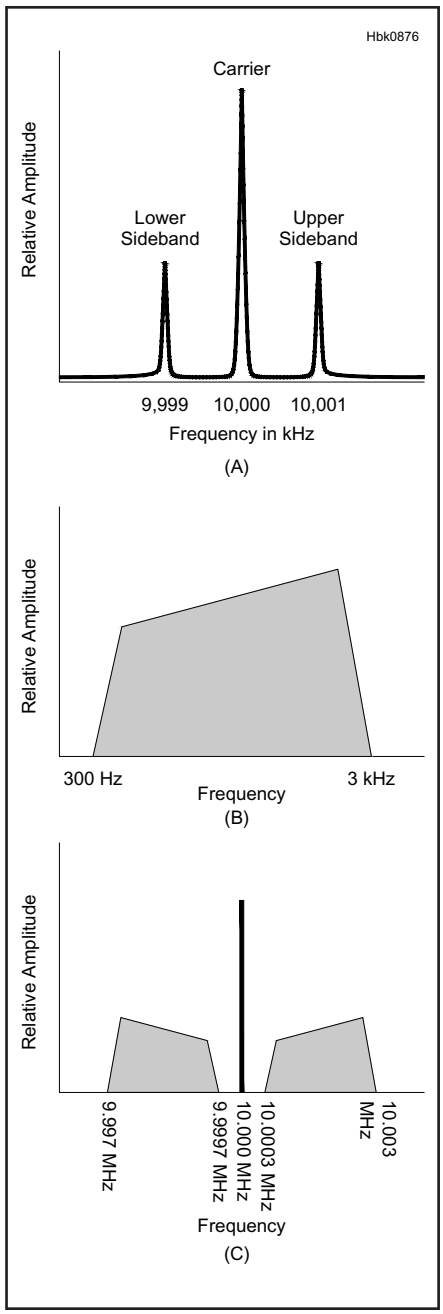

Figure 11.3 — AM with a single 1 kHz tone modulating a 10 MHz carrier (A) and a speech waveform (B) modulating the same carrier in (C).

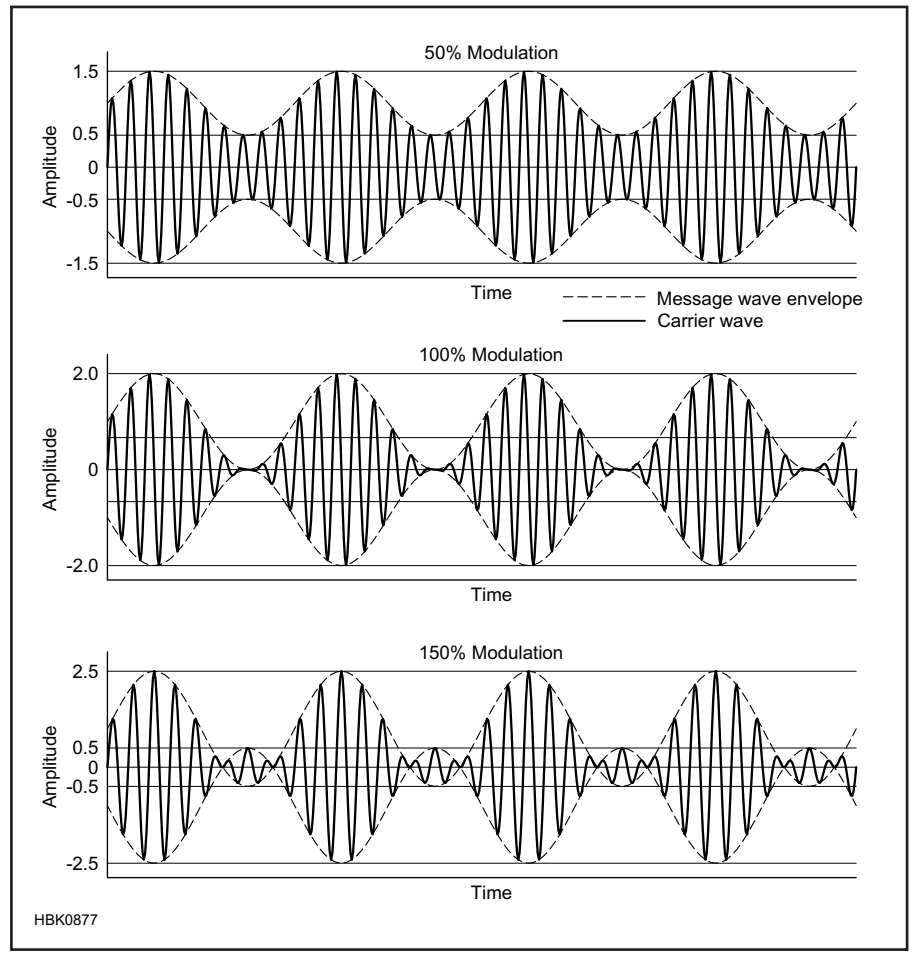

Figure 11.4 — AM waveforms showing 50% modulation (A) and 100% modulation (B). Overmodulation (C) causes distortion of the envelope and reversal of the RF carrier's phase, creating spurious emissions and interference on adjacent channels.

a single audio tone with a frequency of f_M.

$$c(t) \times [1 + m(t)] = [1+M \cos (2\pi f_M t)] \times C \sin (2\pi f_C t)$$

The constant 1 represents a dc component which is necessary to allow the *envelope* of the AM signal to both increase and decrease as in **Figure 11.2**. Note that if the modulating signal is zero (such as when there is no speech or tone) then the result is just the original carrier, c(t). If M=1, the value of 1 + m(t) can vary from zero to 2 and the signal would just go to zero on the peaks of the modulating signal. Figure 11.2C shows a case very similar to this where the signal's envelope almost goes to zero but not quite.

Mathematically, the resulting AM signal is:

$$C\sin(2\pi f_C t) + \frac{C \times M}{2}\sin(2\pi(f_C + f_M)t) + \frac{C \times M}{2}\sin(2\pi(f_C - f_M)t)$$

This expression describes a signal with three components represented by the three terms: first, the carrier; second, an *upper sideband* with a frequency of $f_C + f_M$; and third, a *lower sideband* with a frequency of $f_C - f_M$. This is *double-sideband, full-carrier* (DSB-FC) AM.

If M=1, each sideband's amplitude (C/2) has half the amplitude of the carrier (C). This means each sideband has one-quarter of the carrier's power level or is 6 dB weaker than the carrier. Adding the two sidebands together means that ⅓ of the total signal power is in the sidebands and 2/3 in the carrier. For example, if the carrier power is 100 watts, each sideband will have $100 \times ¼ = 25$ watts. The total signal power is $100 + 2 \times 25 = 150$ watts.

AM SPECTRUM

Figure 11.3A shows the spectrum of a 10-MHz signal modulated with a 1-kHz sine wave. The upper sideband is a single frequency at 10 MHz + 1 kHz = 10.001 MHz. The spectrum of the lower sideband is inverted, so it is at 10 MHz – 1 kHz = 9.999 MHz.

When the modulating signal is speech, instead of a single tone, the result is shown in Figure 11.3B and 11.3C. The speech signal is represented by the shaded region in Figure 11.3B extending from 300 Hz to 3 kHz with the higher frequency speech components having a slightly higher amplitude. When the 10 MHz carrier is modulated with the speech signal, the resulting sidebands are shown in Figure 11.3C. Each speech component behaves like a separate tone, creating its own sideband. The set of components in speech create the set of sidebands shown by the shaded regions. Note that the upper and lower sideband spectra are

"mirror images" of each other with the lower components closest to the carrier.

With both sidebands and the carrier, the AM signal's total bandwidth is twice the bandwidth of the baseband modulation signal. (A single tone's bandwidth is considered to be the frequency of the tone.) The bandwidth of an AM signal modulated by communications-quality speech with a frequency range of 300 Hz to 3 kHz is 2 × 3 kHz = 6 kHz. Note that the AM signal's bandwidth does not depend on the lower limit of the baseband modulation signal.

AM MODULATION INDEX

If the AM signal's envelope is zero on negative peaks of the modulating signal, that corresponds to 100% modulation. For the case where the modulation signal is zero leaving just the steady carrier signal, that is 0% modulation. Expressed as a value from 0 to 1, this is the *modulation index*:

Modulation index = M / C
Modulation percentage = (M / C) × 100%

Figure 11.4A and 11.4B show the AM signal waveform when M = C/2 (50% modulation or a modulation index of 0.5) and when M = C (100% modulation or a modulation index of 1.0.) If M is greater than C, the envelope of the AM signal can go "below zero" as shown in Figure 11.4C, causing the envelope of the signal to be distorted on negative peaks (and possibly positive peaks depending on the design of the modulator circuit). This is the condition of *overmodulation* and the distortion is known as "flat-topping" because of the shape of the envelope which often exhibits a flattening of peaks during overmodution. This also results in spurious emissions the extend beyond the upper and lower sideband, causing interference on adjacent channels.

11.2.2 Double-Sideband, Suppressed Carrier (DSB-SC) AM

Another way of seeing how an AM signal is constructed is illustrated in **Figure 11.5**. Figure 11.5A shows the carrier and the sidebands from a modulating tone are shown in 11.5B and 11.5C. If you look closely, you can see that the waveforms in Figure 11.5B and 11.5C have slightly different frequencies than the carrier. If the two sidebands are added together, the signal of Figure 11.5D is produced. This is a *double-sideband, suppressed carrier* (DSB-SC) signal and its spectrum is shown in **Figure 11.6**, assuming the same 10 MHz carrier and 300-3000 Hz sidebands.

When the carrier signal is added, the full AM signal is produced in Figure 11.5D. When all of the signals are in-phase, the resulting signal has its maximum amplitude. When all of the signals are out of phase, the resulting signal goes to zero. If the carrier's phase is used as our reference, the phase of each sideband can be viewed as slipping behind (lower sideband) or moving ahead (upper sideband) of the carrier. The sidebands are out of phase with each other at the frequency of the tone so the resulting envelope reproduces the modulating tone's sine wave.

11.2.3 Single-Sideband, Suppressed Carrier (SSB-SC)

As we have seen, since the carrier itself contains no modulation, it does not need to be transmitted, which saves at least 67% of the transmitted power. Since the two sidebands carry identical information, one of them may be eliminated as well, saving half the bandwidth. The result is *single sideband, suppressed carrier* (SSB-SC), which is commonly referred to simply as "SSB." If the lower sideband is eliminated, the result is called *upper sideband* (USB). If the upper sideband is eliminated, you're left with *lower sideband* (LSB). **Figure 11.7** shows the result of removing one of the sidebands in Figure 11.6.

The effect of SSB modulation is that the baseband modulation signal is simply frequency-shifted to the RF carrier frequency (whether the carrier is transmitted or not.)

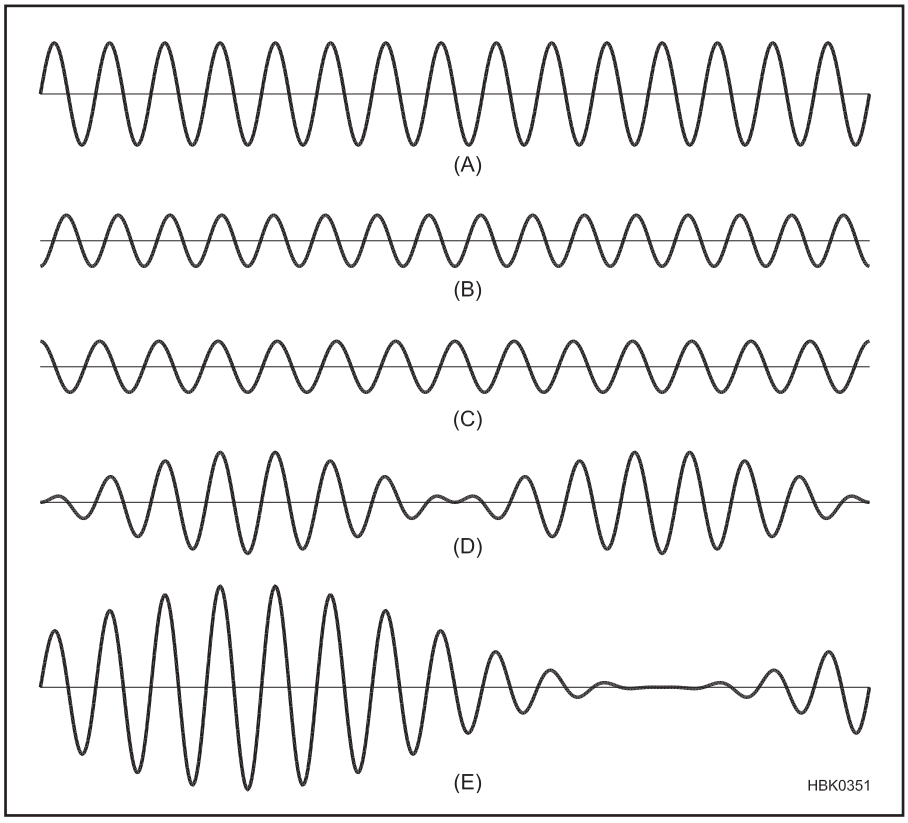

Figure 11.5 — At A is an unmodulated carrier. If the upper (B) and lower (C) sidebands are added together a double-sideband suppressed carrier (DSB-SC) signal results (D). If each sideband has half the amplitude of the carrier, then the combination of the carrier with the two sidebands results in a 100%-modulated AM signal (E). Whenever the two sidebands are out of phase with the carrier, the three signals sum to zero. Whenever the two sidebands are in phase with the carrier, the resulting signal has twice the amplitude of the unmodulated carrier.

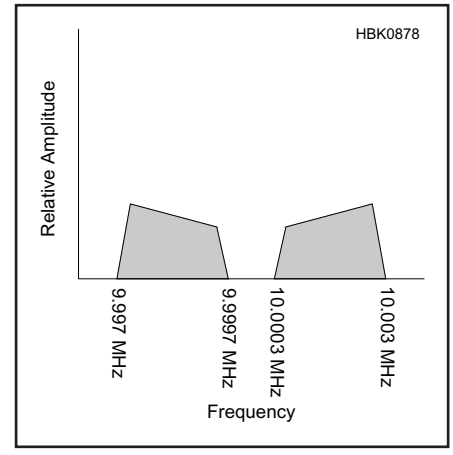

Figure 11.6 — The spectrum of a double-sideband, suppressed-carrier (DSB-SC) signal created from the signal shown in Figure 11.3.

The spectrum of the modulation may be inverted during the shifting process, requiring the demodulation process to re-invert the signal as it is translated back to the baseband frequency range. Aside from that technical consideration, there is no difference between the information in the LSB and USB signals.

The bandwidth of an SSB signal is a little less than half that of an equivalent AM signal. Using the same 300 Hz to 3 kHz speech bandwidth, the SSB signal has a bandwidth of 3000 – 300 = 2700 Hz = 2.7 kHz. In practice, SSB transmitters have an overall audio response from 300 Hz to 2.8 kHz, resulting in 2.5 kHz bandwidth.

Compared to most other analog modulation modes, SSB has excellent power efficiency because the transmitted power is proportional to the modulating signal and no power at all is transmitted during pauses in speech. Another advantage is that the lack of a carrier results in less interference to other stations.

SSB transceivers also are well-suited for the narrowband digital modes. Because an SSB transceiver simply frequency-translates the baseband audio signal to RF in the transmitter and back to audio again in the receiver, digital modulation may be generated and detected at audio frequencies using the sound card in a personal computer.

11.2.4 Amplitude-Modulated On-Off Keying (OOK)

On-off keying (OOK) is a special case of *amplitude-shift keying* (ASK) and is normally used for sending Morse code. For historical reasons dating from the days of spark transmitters, amateurs often refer to this as c*ontinuous wave* (CW) even though the signal is actually keyed on and off.

You can think of OOK as being the same as analog AM that is modulated with a two-level signal that switches between full power and zero power. For example, imagine that the modulation is a 10 Hz square wave, equivalent to sending a series of dits at 24 WPM. A square wave may be decomposed into a sine wave of the same frequency and a theoretically infinite succession of odd harmonics. Recall that the lower and upper sidebands of an AM signal are simply the inverted and non-inverted spectra of the baseband modulation. The RF spectrum therefore contains sidebands at the carrier frequency ±10 Hz, ±30 Hz, ±50 Hz, and so on to plus and minus infinity. This phenomenon is called *key clicks*. Stations listening on nearby frequencies hear a click upon every key closure and opening.

To prevent interference to other stations, the modulation must be low-pass filtered, which slows down the transition times between the on and off states. See **Figure 11.8**. If the transitions are too fast, then excessive key clicks occur. If the transitions are too slow, then at high speeds the previous transition may not have finished before the next one starts, which makes the signal sound mushy and hard to read. Traditionally, filtering was done with a simple resistor-capacitor low-pass filter on the keying line, but using a transition with a raised-cosine shape allows faster transition times without excessive key clicks. Some modern transceivers use DSP techniques to generate such a controlled transition shape.

The optimum transition time, and thus bandwidth, depends on keying speed. It also depends on propagation conditions. When the signal is fading, the transitions must be sharper to allow good copy. **Figure 11.9** gives recommended keying characteristics based on sending speed and propagation. As a compromise, many transmitters use a 5 ms rise and fall time. That limits the bandwidth to approximately 150 Hz while allowing good copy up to 60 WPM on non-fading channels and 35 WPM on fading channels, which covers most requirements.

With any digital system, the number of changes of state per second is called the *baud rate* or the *symbol rate*, measured in *bauds* or

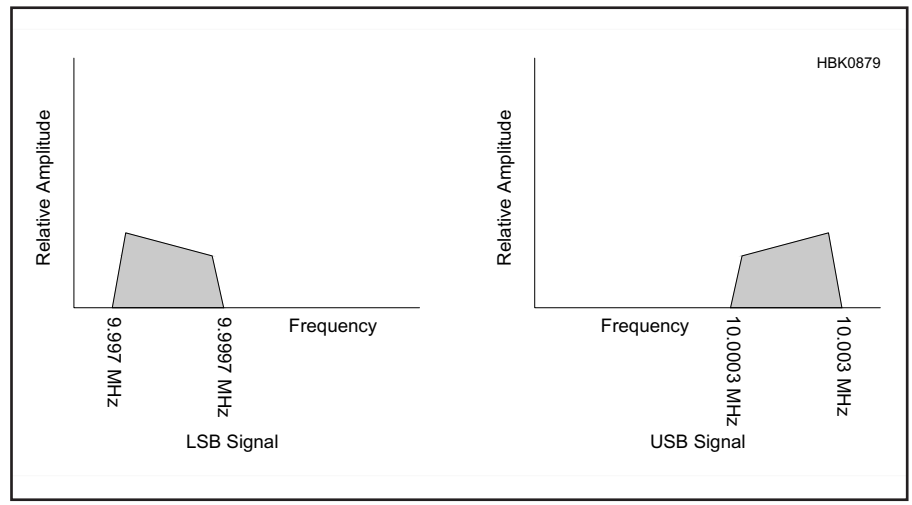

Figure 11.7 — The spectrum of the two single sideband (SSB) signals that could be created from the signal shown in Figure 11.6.

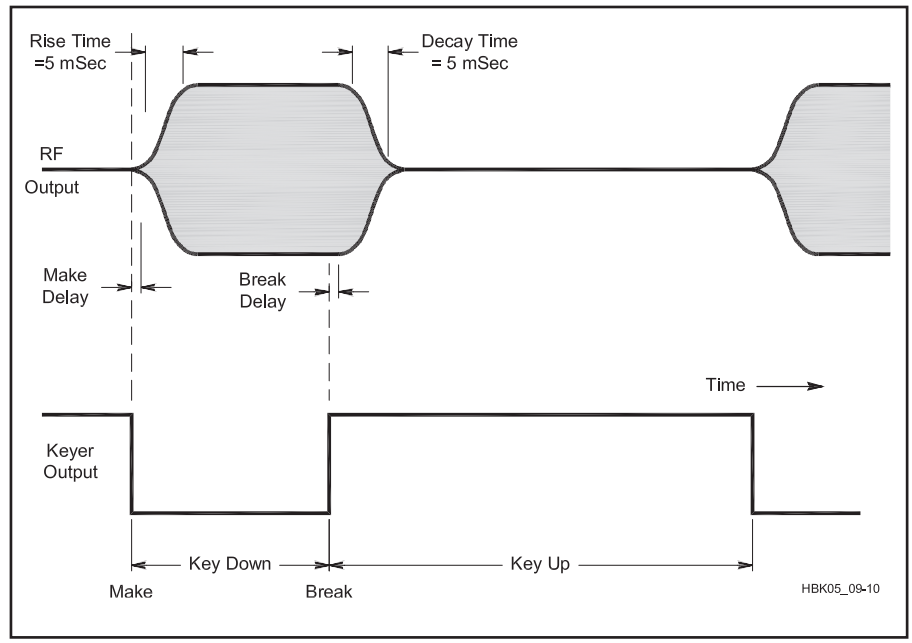

Figure 11.8 — A filtered CW keying waveform. The on-off transitions of the RF envelope should be as smooth as possible while transitioning as quickly as possible. The shape of the RF Output waveform is nearly optimum in that respect.

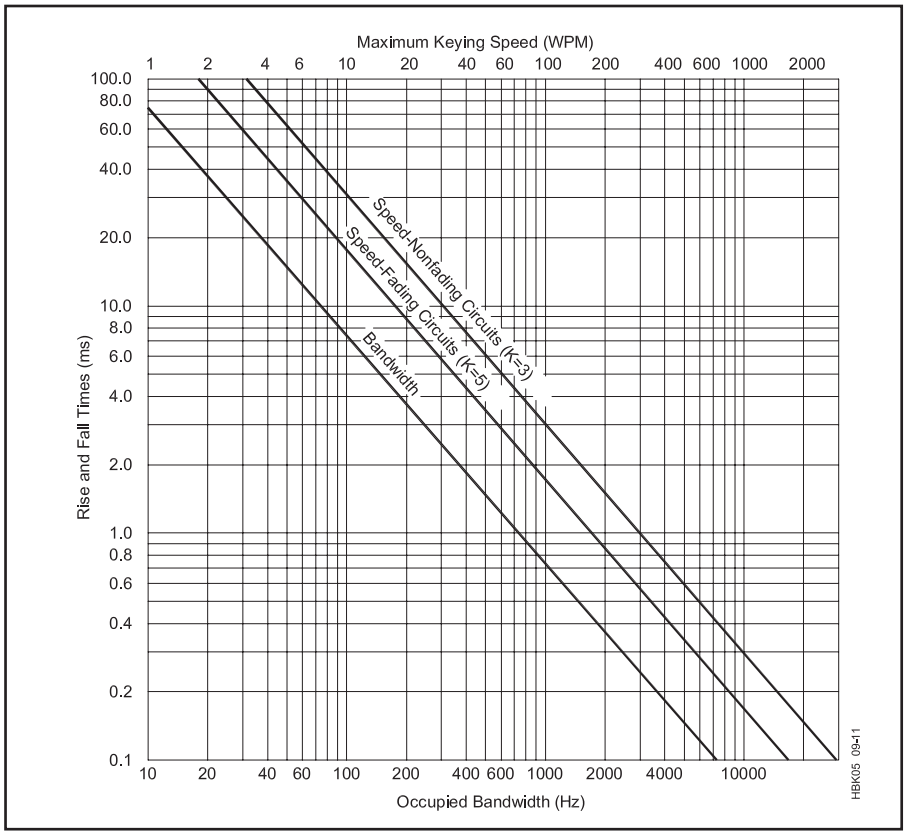

Figure 11.9 — Keying speed vs rise and fall times vs bandwidth for fading and non-fading communications circuits. For example, to optimize transmitter timing for 20 WPM on a non-fading circuit, draw a vertical line from the WPM axis to the K = 3 line. From there draw a horizontal line to the rise/fall axis (approximately 15 ms). Draw a vertical line from where the horizontal line crosses the bandwidth line and see that the bandwidth is about 50 Hz. Harder (more abrupt) keying is required to maintain copying ability in the presence of fading.

symbols per second. Sending a single Morse code dit requires two equal-length states, or symbols: *on* for the length of the dit and *off* for the space between dits. Thus, a string of dits sent at a rate of 10 dits per second has a symbol rate of 20 bauds.

Refer to **Figure 11.10**. A Morse code dash is on for three times the length of a dit. Including one symbol for the off time, the total time to send a dash is four symbols, twice the time to send a dit. For example, at a baud rate of 20 bauds, there are 10 dits per second or 5 dashes per second. The spacing between characters within a word is three symbols, two more than the normal space between dits and dashes. The spacing between words is seven symbols.

For purposes of computing sending speed, a standard word is considered to have five characters, plus the inter-word spacing. On average, that results in 50 symbols per word. From that, the speed in words per minute may be computed from the baud rate:

$$\text{WPM} = \frac{60\ (\text{sec} / \text{min})}{50\ (\text{symbols} / \text{word})} \times \text{bauds}$$
$$= 1.2 \times \text{bauds} = 2.4 \times \text{dits} / \text{sec}$$

Characters in Morse code do not all have the same length. Longer codes are used for characters that are used less frequently while the shortest codes are reserved for the most common characters. For example, the most common letter in the English language, E, is sent as a single dit. In that way, the average character length is reduced, resulting in a faster sending speed for a given baud rate. Such a variable-length code is known as *varicode*, a technique that has been copied in some modern digital modes (see the **Digital Protocols and Modes** chapter).

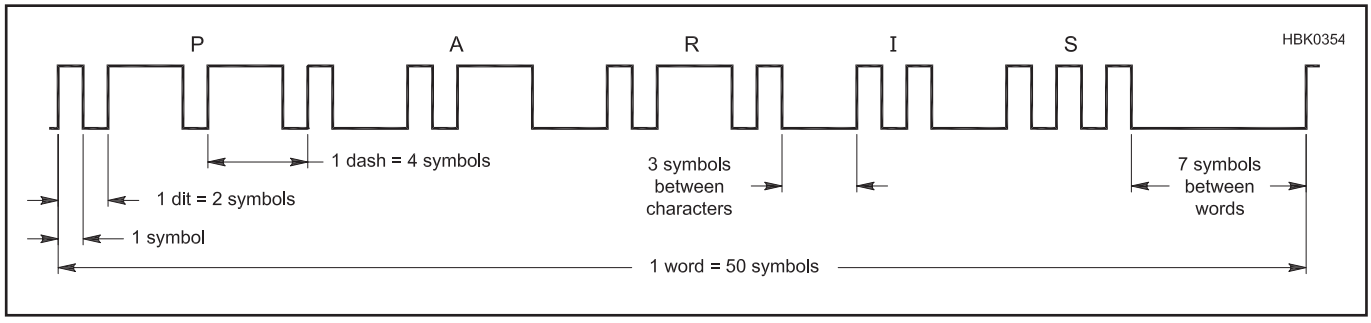

Figure 11.10 — CW timing of the word PARIS, which happens to have a length equal to the standard 50 symbols per word. By programming it multiple times into a memory keyer, the speed may be calibrated simply by counting the number of times the word is completed in one minute.

11.3 Angle Modulation

Angle modulation varies the phase angle of an RF signal in response to the modulating signal. In this context, "phase" means the phase of the modulated RF sine wave with respect to the unmodulated carrier. Angle modulation includes both frequency modulation and phase modulation because any change in frequency results in a change in phase. For example, the way to smoothly ramp the phase from one value to another is to change the frequency and wait. If the frequency is changed by +1 Hz, then after 1 second the phase will have changed by +360°. After 2 seconds, the phase will have changed by +720°, and so on. Change the frequency in the other direction and the phase moves in the opposite direction as well. With sine-wave modulation, the frequency and phase both vary in a sinusoidal fashion. See **Figure 11.11**.

For angle modulation, the modulation signal, m(t) is applied to the frequency or phase of the carrier, not its amplitude:

$$\sin (2\pi f_C t + k_P m(t)) \text{ or } \sin (2\pi f_C t \times f_d m(t))$$

Mathematically, these are equivalent and most texts work with the second form. The constants k_P and f_d are described below in the discussion on modulation index.

Since any change in frequency results in a change in phase and vice versa, *frequency modulation* (FM) and *phase modulation* (PM) are fundamentally the same. The term *frequency deviation* means the amount the RF frequency deviates from the center (carrier) frequency with a given modulating signal. The instantaneous deviation of an FM signal is proportional to the instantaneous amplitude of the modulating signal. The instantaneous deviation of a PM signal is proportional to the instantaneous *rate of change* of the modulating signal. Since the rate of change of a sine wave is proportional to its frequency as well as its amplitude, the deviation of a PM signal is proportional to both the amplitude and the frequency of the modulating signal.

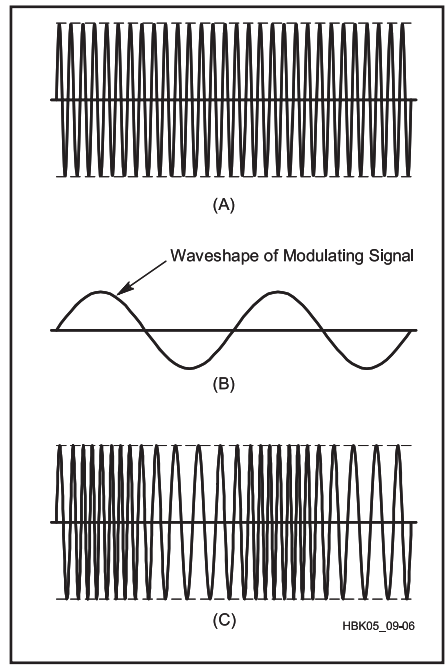

Figure 11.11 — Graphical representation of frequency modulation. In the unmodulated carrier (A) each RF cycle occupies the same amount of time. When the modulating signal (B) is applied, the radio frequency is increased or decreased according to the amplitude and polarity of the modulating signal (C).

11.3.1 Angle-Modulated Modulation Index

The *modulation index* is the ratio of the peak deviation to the highest audio frequency. The *deviation ratio* is the ratio of maximum permitted peak deviation to the maximum permitted modulating frequency. The modulation index is a measure of the maximum amount of phase change (ϕ_{MAX}) the modulation signal can cause. For a single-tone modulating signal:

$$\text{Modulation index} = A \times f_d / f_M \text{ (for FM)}$$

$$\text{Modulation index} = (\phi_{MAX}) = k_P \times A \text{ (for PM)}$$

where the modulation index is calculated in radians and there are 180/π radians in the 360º of one complete sine wave cycle (1 radian ≈ 57.3º). In addition,

- A is the amplitude of the modulation signal in volts
- f_M is the frequency of the message signal in hertz
- f_d is the *frequency deviation constant* that represents the sensitivity of the modulator in hertz of deviation per volt of the modulating signal
- $A \times f_d$ is the *peak deviation*
- ϕ_{MAX} is the maximum value of phase change caused by the modulation signal
- k_P is the *phase deviation constant* and is similar to f_d in that it specifies the sensitivity of the phase modulator in radians of phase change per volt of the modulating signal.

For PM, the modulation index doesn't depend on message frequency at all. For an FM signal, m will be larger if the peak deviation gets larger or if f_M gets smaller. For example, loud low-frequency signals can cause the modulation index to become quite large unless the transmitter limits deviation and microphone gain or frequency response.

FCC regulations limit the modulation index at the highest modulating frequency to a maximum of 1.0 for frequencies below 29 MHz. If the audio is low-pass filtered to 3 kHz, then the deviation may be no more than 3 kHz. For that reason, FM transmitters for frequencies below 29 MHz are usually set for 3 kHz deviation while FM transmitters at higher frequencies are typically set for about 5 kHz deviation. The term *narrowband FM* generally refers to deviation of no more than 3 kHz and *wideband FM* refers to deviation greater than 3 kHz.

11.3.2 Angle Modulation Audio Frequency Response

The only difference between FM and PM is the audio frequency response. An FM transmitter with 6 dB/octave pre-emphasis of the modulating signal is indistinguishable from a PM transmitter. A PM transmitter with 6 dB/octave de-emphasis is indistinguishable from an FM transmitter. The reverse happens at the receiver. A frequency detector followed by a 6 dB/octave de-emphasis network acts like a phase detector. It is interesting to note that most VHF and UHF amateur "FM" transceivers should really be called "PM" transceivers due to the pre-emphasis and de-emphasis networks used in the transmitters and receivers respectively.

Most FM and PM transmitters include some kind of audio compressor before the modulator to limit the maximum deviation. Common usage of the term *deviation* is that it refers to the maximum peak deviation allowed by the audio compressor. If the frequency swings a maximum of 5 kHz above the center frequency and 5 kHz below the center frequency, we say the deviation is 5 kHz.

11.3.3 Angle Modulation Bandwidth

The spectrum of an angle-modulated signal is fairly complex. With sine-wave modulation, the RF frequency spectrum from an angle-modulated transmitter consists of the carrier and a series of sideband pairs, each spaced by the frequency of the modulation. For example, with 3 kHz sine-wave modulation, the spectrum includes tones at ±3 kHz, ±6 kHz, ±9 kHz and so on with respect to the carrier. When the modulation index is much less than 1.0, only the first sideband pair is significant, and the spectrum looks similar to that of an AM signal (although the phases of the sidebands are different).

As the modulation index is increased, more and more sidebands appear. Unlike with AM,

the carrier amplitude also changes with deviation and actually disappears altogether for certain modulation indices. Because the amplitude of an angle-modulated signal does not vary with modulation, the total power of the carrier and all sidebands is constant. **Figure 11.12** shows several graphs of an FM signal with a 1 kHz modulation signal at different modulation indexes. Note how the sidebands increase and decrease as modulation index changes. All of these spectra have the same total power.

The amplitudes of the carrier and the various sidebands are described by a series of mathematical equations called Bessel functions, which are illustrated graphically in **Figure 11.13**. Note that the carrier disappears when the modulation index equals 2.405 and 5.52. That fact can be used to set the deviation of an FM transmitter. For example, to set 5 kHz deviation, connect the microphone input to a sine-wave generator set for a frequency of 5 / 2.405 = 2.079 kHz, listen to the carrier on a narrowband receiver, and adjust the deviation until the carrier disappears.

11.3.4 Carson's Rule

Unlike amplitude modulation, angle modulation is nonlinear. Recall that with amplitude modulation, the shape of the RF spectrum is the same as that of the modulation spectrum, single-sided with SSB and double-sided with DSB. That is not true with angle modulation. A double-sideband AM signal with audio band-limited to 3 kHz has 6 kHz RF bandwidth. It is easy to see that an FM transmitter with the same audio characteristics but with, say, 5 kHz deviation must have a bandwidth of at least 10 kHz. While you might think that the bandwidth equals twice the deviation, in reality the transmitted spectrum theoretically extends to infinity, although it does become vanishingly small beyond a certain point. As a rule of thumb, approximately 98% of the spectral energy is contained within the bandwidth defined by *Carson's rule*:

$$BW = 2(f_d + f_m)$$

where
f_d = the peak deviation, and
f_m = the highest modulating frequency.

For example, if the deviation is 5 kHz and the audio is limited to 3 kHz, the bandwidth is approximately 2 (5 + 3) = 16 kHz.

Be careful not to "over apply" Carson's Rule. It is just a method of estimating the bandwidth in which a certain amount (98%) of the signal power is contained. This is not as strict a definition as the FCC definition cited earlier in this chapter. Limiting the modulation index and stating the bandwidth in the rules defines a single FM or PM "channel" and how much energy is allowed to be outside that channel. **Figure 11.14** shows a typical 2 meter FM repeater output and compares Carson's Rule and the FCC bandwidth limits. The bandwidths derived from the two rules are "close" but not equivalent. The FCC would judge the signal to be wider than what Carson's Rule predicts. Relying on Carson's Rule is not sufficient to guarantee that there will be no interference to adjacent channels at minimum spacing.

11.3.5 AM Noise and FM Signals

Since the amplitude carries no information, FM receivers are designed to be as insensitive to amplitude variations as possible. Because noise tends to be mostly AM in nature, that results in a quieter demodulated signal. Typically the receiver includes a *limiter*, which is a very high-gain amplifier that causes the signal to clip, removing any amplitude variations, before being applied to the detector. Unlike with AM, as an FM signal gets stronger the volume of the demodulated audio stays the same, but the noise is reduced. Receiver sensitivity is often specified by how much the noise is suppressed for a certain input signal level. For example, if a 0.25 µV signal causes the noise to be reduced by 20 dB, then we say the receiver sensitivity is

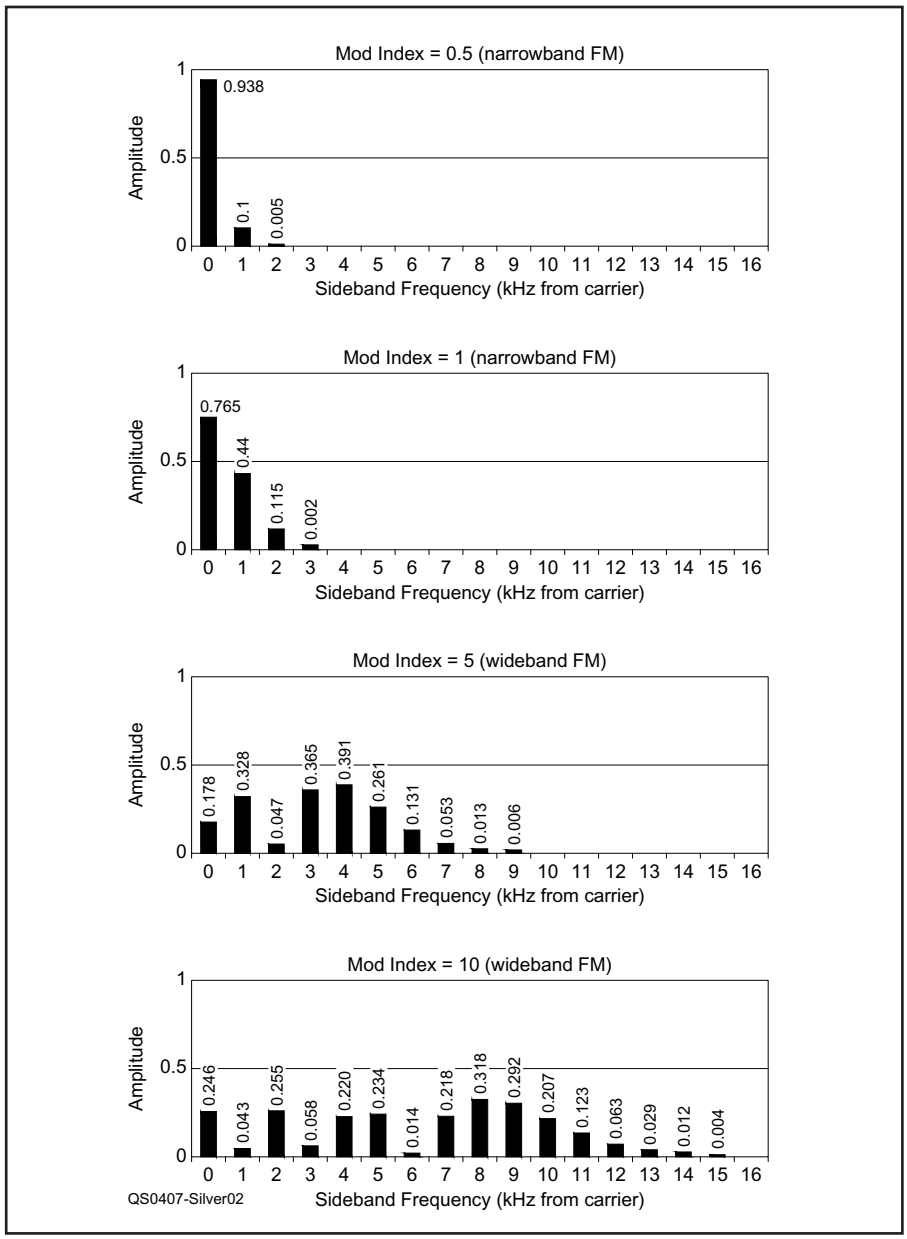

Figure 11.12 — Charts showing one-half of an FM signal's sidebands for a modulating signal frequency of 1 kHz at different modulation indexes.

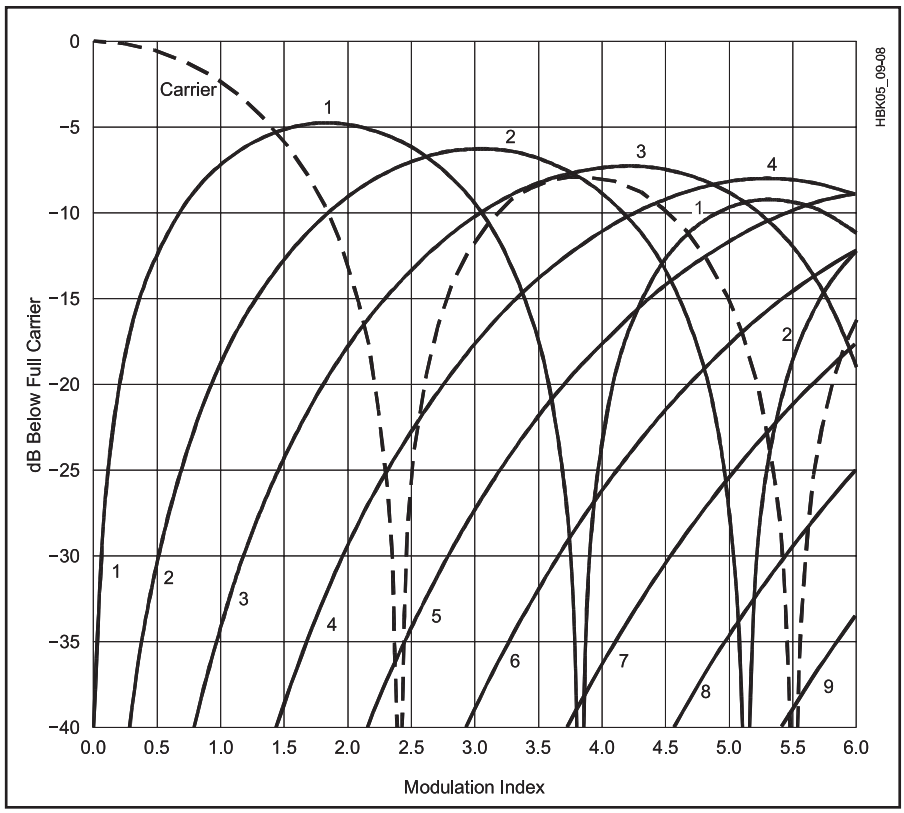

Figure 11.13 — Amplitude of the FM carrier and the first nine sidebands versus modulation index. This is a graphical representation of mathematical functions developed by F. W. Bessel. Note that the carrier completely disappears at modulation indices of 2.405 and 5.52.

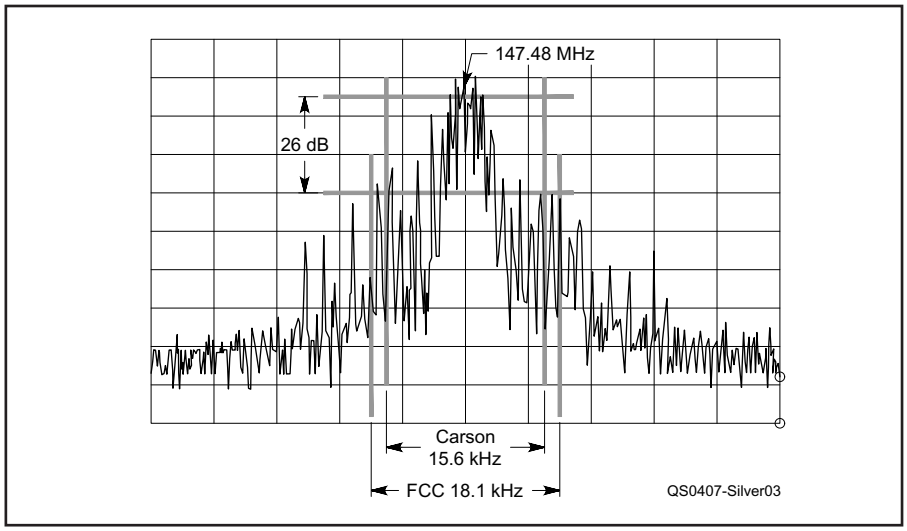

Figure 11.14 — Spectrum analyzer display for a typical 2 meter FM repeater. The bandwidths and amplitudes for Carson's Rule and the FCC's bandwidth definition are overlaid on the spectrum.

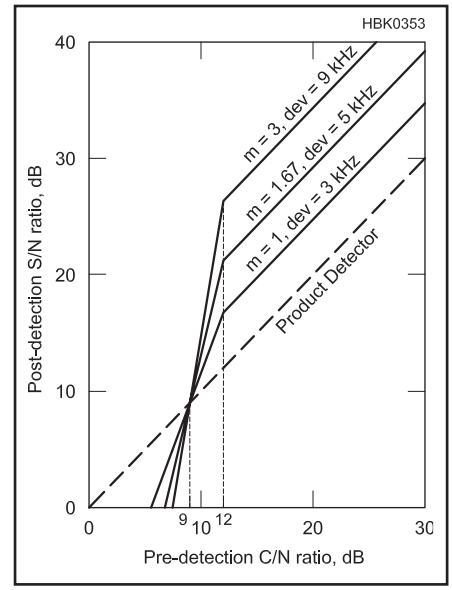

Figure 11.15 — A plot of post-detection signal-to-noise ratio (S/N) versus input carrier-to-noise ratio (C/N) for an FM detector at various modulation indices, m. For each modulation index the deviation is also noted assuming a maximum modulating frequency of 3 kHz. For comparison, the performance of an SSB product detector is also shown.

"0.25 μV for 20 dB of quieting."

The limiter also causes a phenomenon known as *capture effect*. If more than one signal is present at the same time, the limiter tends to reduce the weaker signal relative to the stronger one. We say that the stronger signal "captures" the receiver. The effect is very useful in reducing on-channel interference.

The suppression of both noise and interference is greater the wider the deviation. FM signals with wider deviation do take up more bandwidth and actually have a poorer signal-to-noise (S/N) ratio at the detector output for weak signals but have better S/N ratio and interference rejection for signal levels above a certain threshold. See **Figure 11.15**.

In addition to the noise and interference-reduction advantage, angle-modulated signals share with full-carrier AM the advantages of non-critical frequency accuracy and the continuous presence of a signal, which eases the task of the automatic gain control system in the receiver. In addition, since the signal is constant-amplitude, the transmitter does not need a linear amplifier. Class C amplifiers may be used, which have greater power efficiency.

11.4 FSK and PSK

While they are technically types of angle modulation, *frequency-shift keying* (FSK) and *phase-shift keying* (PSK) are more easily discussed on their own terms. The technique of shifting a signal's frequency or phase to represent the 0 and 1 values of digital data bits is the dominant form of communications outside of amateur radio. Amateurs are making wider use of these techniques as they adopt more digital modes, particularly on the HF and MF bands. (See the **Digital Protocols and Modes** chapter.)

11.4.1 Frequency-Shift Keying (FSK)

FSK was the first digital angle-modulated format to come into common use. It can be thought of as being the same as analog FM that is modulated with a binary (two-level) signal that causes the RF signal to switch between two frequencies. It can also be thought of as equivalent to two on-off-keyed (OOK) signals on two nearby frequencies that are keyed in such a way that whenever one is on the other is off. Just as with OOK, if the transitions between states are instantaneous, then excessive bandwidth occurs — causing interference on nearby channels similar to key clicks. For that reason, the modulating signal must be low-pass filtered to slow down the speed at which the RF signal moves from one frequency to the other.

Although FSK is normally transmitted as a true constant-amplitude signal with only the frequency changing between symbols, it does not have to be received that way. The receiver can treat the signal as two OOK signals and demodulate each one separately. This is an advantage when HF propagation conditions exhibit selective fading — even if one frequency fades out completely the receiver can continue to copy the other, a form of *frequency diversity*. To take advantage of it, wide shift (850 Hz) is normally used. With narrow-shift FSK (170 Hz), the two tones are generally too close to exhibit selective fading.

As previously discussed, selective fading is caused by the signal arriving at the receiver antenna by two or more paths simultaneously. The same phenomenon can cause another signal impairment known as *inter-symbol interference* (ISI) as shown in **Figure 11.16.** If the difference in the two path lengths is great enough, then the signal from one path may arrive delayed by one entire symbol time with respect to the other. The receiver sees two copies of the signal that are time-shifted by one symbol. In effect, the signal interferes with itself. One solution is to slow down the baud rate so that the symbols do not overlap. It is for this reason that symbol rates employed on HF are usually no more than 50 to 100 baud.

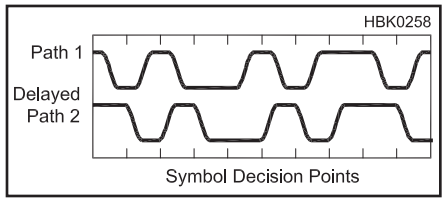

Figure 11.16 — Multipath propagation can cause inter-symbol interference (ISI). At the symbol decision points, which is where the receiver decoder samples the signal, the path 2 data is often opposing the data on path 1.

GAUSSIAN FSK (GFSK)

Gaussian FSK (GFSK) refers to FSK in which the modulating signal has been filtered with a low-pass filter with a Gaussian frequency response. As mentioned before, with any type of angle modulation the spectrum of the modulation is not duplicated at RF but spreads out into an increased bandwidth. For that reason, there is no point in using a modulation filter with a sharp cutoff since the RF spectrum will be wider anyway. A Gaussian filter, with its gradual transition from passband to stopband, has the optimum shape for an angle-modulated digital system.

Figure 11.17 shows band-pass responses that compare the more familiar Butterworth response to Gaussian with an ideal or "brick wall" response as a reference. (See the **Analog and Digital Filtering** chapter for more about filter responses.) While the Butterworth response (with an order n=5) is closer to the ideal filter, the Gaussian filter's smoother transition outside the passband creates fewer products that increase the modulated RF signal's bandwidth.

However, a Gaussian filter also has a gradual transition from symbol-to-symbol in the time domain. The transition is not totally completed by the time of the next symbol, which means that there is some inter-symbol interference (ISI) in the transmitted signal, even in the absence of propagation impairments. With the proper choice of filter bandwidth, however, the ISI is small enough not to seriously affect performance

MULTI-LEVEL FSK

Multi-level FSK (MFSK) is one method to reduce the symbol rate. Unlike with conventional binary FSK, more than two shift frequencies are allowed. For example, with eight frequencies (8-FSK), each symbol can have eight possible states. Since three bits are required to represent eight states, three bits are transmitted per symbol. That means you get three times the data rate without increasing the baud rate. The disadvantage is a reduced signal-to-noise and signal-to-interference ratio. If the maximum deviation is the same, then the frequencies are seven times closer with 8-FSK than with binary FSK and the receiver is theoretically seven times (16.9 dB) more susceptible to noise and interference. However, an IF limiter stage can be used to remove most amplitude variations before the signal arrives at the FSK detector, so the actual increase in susceptibility is less for signal levels above a certain threshold.

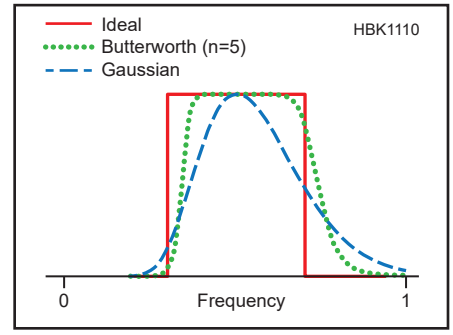

Figure 11.17 — The Gaussian bandpass filter response compared to Butterworth and ideal "brick wall" responses. A Gaussian low-pass filter can be applied to digital data before performing FSK modulation to reduces the resulting RF signal bandwidth.

With any type of FSK, you can theoretically make the shift as narrow as you like. The main disadvantage is that the receiver becomes more susceptible to noise and interference, as explained above. In addition, the bandwidth is not reduced as much as you might expect. Just as with analog FM, you still get sidebands whose extent depends on the symbol rate, no matter how small the deviation.

11.4.2 Phase-Shift Keying (PSK)

Binary phase-shift keying (BPSK), often referred to simply as phase-shift keying (PSK), can be implemented as a true constant-amplitude angle modulation. An example is *minimum-shift keying* (MSK) which is FSK with a deviation that is at the minimum practical level, taking bandwidth and signal-to-noise ratio into account. That turns out to be a frequency shift from the center frequency of 0.25 times the symbol rate or, using the common definition of frequency shift, a difference between the tone frequencies of 0.5 times the symbol rate. If, on a given symbol, the frequency is shifted to the upper tone, then the phase of the RF signal will change by 0.25 cycles, or 90°, during the symbol period. If the lower tone is selected, the phase shift is –90°.

Thus, MSK may be regarded as either FSK with a frequency shift of 0.5 times the symbol rate, or as *differential phase-shift keying* (DPSK) with a phase shift of ±90°. The binary data can cause the phase to change by either +90 or –90° from one symbol to the next. *Gaussian minimum-shift keying* (GMSK) applies a Gaussian filter to the modulating signal for the same reasons as for GFSK which is described in the preceding section.

The term BPSK is normally understood, however, to refer to phase-shift keying with a 180° phase difference between symbols. To transition from one state to the other, the modulation filter smoothly reduces the amplitude to zero where the polarity reverses (phase changes 180°) before smoothly ramping up to full amplitude again. For that reason, BPSK as usually implemented really should not be considered to be PSK, but rather a form of amplitude-shift keying (ASK) with two modulation amplitudes, +1 and –1. Unlike true angle modulation, it is linear so that the spectrum of the modulation filter is duplicated at RF. The transmitter must use a linear amplifier to prevent distortion and excessive bandwidth similar to the splatter that results from an over-driven SSB transmitter.

11.4.3 Audio Frequency-Shift Keying (AFSK)

Audio frequency-shift keying (AFSK) is the generation of radio-frequency FSK using an audio-frequency FSK signal fed into the modulation input of an SSB transmitter, usually the microphone input. Assume the SSB transmitter is tuned to 14.000 MHz, USB. If the audio signal consists of a sine wave that shifts between 2125 Hz and 2295 Hz (170-Hz frequency shift), then the RF signal is a sine wave shifting between 14.002125 and 14.002295 MHz. The frequency shift and spectral characteristics are theoretically unchanged, other than being translated 14 MHz upward in frequency. The RF signal should be indistinguishable from one generated by varying the frequency of an RF oscillator directly.

This technique works not only for FSK but also for nearly any modulating signal with a bandwidth narrow enough to fit within the passband of an SSB transmitter and receiver. The most common non-voice analog modulation type to use this technique on the amateur bands is FT8, part of the *WSJT-X* software suite. Nearly all narrowband digital signals today are generated in this manner. The audio is usually generated and received by a PC sound card or external audio to USB interface. Dedicated hardware modulator/demodulator (modems) are also used, such as for PACTOR-III and PACTOR-IV.

Whatever the method, it is important to ensure that unwanted interference is not caused by audio distortion or by insufficient suppression of the carrier and unwanted sideband. For example, with the AFSK tone frequencies mentioned above, 2125 and 2295 Hz, the tone harmonics cause no trouble because they fall outside of the transmitter's passband. However, some AFSK modems use 1275 and 1445 Hz (to accommodate 850-Hz shift without changing the 1275-Hz mark frequency). In that case, the second harmonics at 2550 and 2890 Hz must be suppressed since those frequencies are not well-attenuated by the transmitter.

With non-constant-envelope modulation types such as QPSK or the various multi-carrier modes, it is important to set the amplitude of the audio input to the SSB transmitter below the level that activates the transmitter's automatic level control (ALC). That is because the ALC circuit itself generates distortion of signals within the bandwidth of its feedback loop.

11.5 I-Q Modulation

11.5.1 I and Q Components

A sinusoidal wave of any arbitrary amplitude and phase with a frequency of $\omega = 2\pi f$ (f is frequency in Hz) can be represented by the weighted sum of a sine and cosine wave:

$$x(t) = I \cos(\omega t) + Q \sin(\omega t)$$

(See the online **Radio Mathematics** supplement for a list of math tutorials, including trigonometry.) I, the abbreviation for "in-phase" should not be confused with the imaginary number *i*. Q is the abbreviation for "quadrature".

For computational purposes, it is convenient to consider the *in-phase (I)* and *quadrature* (*Q*) components separately. The components are usually represented as vectors on a *phasor* diagram with the I and Q axes at a 90 degree angle from each other as in **Figure 11.18**. The I component lies on the horizontal I axis and the Q component on the vertical Q axis. The sum of the components is represented by the vector with a magnitude of A at an angle, φ, from the I axis.

Each component oscillates back and forth along its own axis. For example, if Q = 0, then as time increases the I signal oscillates along the I (horizontal) axis, tracing out the path back and forth between I = +1 and I = –1 in a sinusoidal fashion. Conversely, if I = 0, then the signal oscillates along the Q axis.

What if both I and Q are non-zero, for example I = Q = 1? The cosine, x(t) = cos(ωt), and sine waves x(t) = sin(ωt), are 90° out of phase as shown in **Figure 11.19**. When t = 0, cos(ωt) = 1 and sin (ωt) = 0. A quarter cycle later, cos(ωt) = 0 and sin (ωt) = 1. Comparing Figure 11.23 with Figure 11.24 you can imagine that the signal (the tip of the vector) is tracing out a circle in the counter-clockwise direction. Similarly, you can imagine that changing ω to –ω results in a signal that circles the origin in the opposite, clockwise direction.

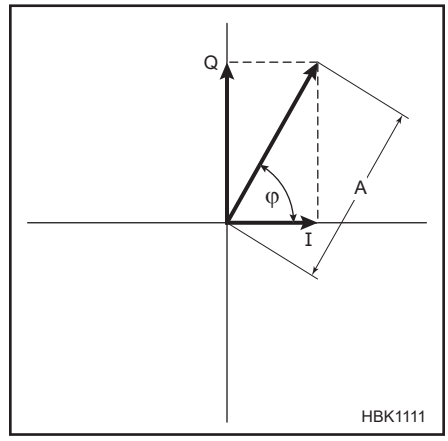

Figure 11.18 — In-phase (I) and quadrature (Q) components of a signal on a phasor diagram.

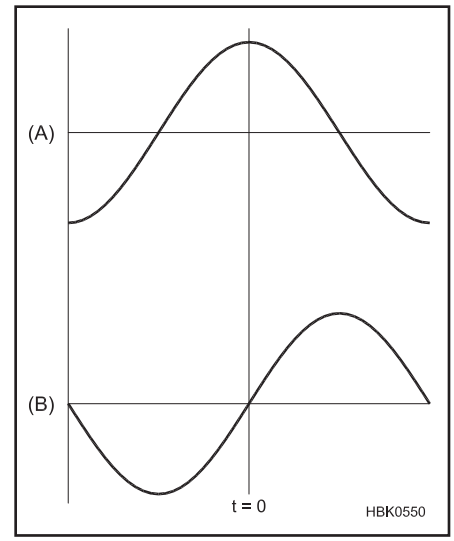

Figure 11.19 — Cosine wave (A) and sine wave (B).

NEGATIVE FREQUENCY

What is negative frequency? Changing the sign of the frequency is equivalent to running time backwards because $(-\omega)t = \omega(-t)$. By examining Figure 11.19A you can see that, because a cosine wave is symmetrical about the time t = 0 point, a negative frequency results in exactly the same signal. That is, $\cos(-\omega t) = \cos(\omega t)$. If, for example, you add a positive-frequency cosine wave to its negative-frequency twin, you get the same signal with twice the amplitude:

$$x(t) = [\cos(\omega t) + \sin(\omega t)] + [\cos(-\omega t) + \sin(-\omega t)] = 2\cos(\omega t)$$

Figure 11.20 illustrates that graphically. Imagine the two vectors rotating in opposite directions, the top one in the counter-clockwise direction of positive frequency. If you mentally add them by placing the tail of one vector on the head of the other, as shown by the dotted line, the result always lies on the I axis and oscillates between +2 and –2 as described by $2\cos(\omega t)$.

That assumes that the phase of the signal is such that it reaches a peak at t = 0. What if instead we had a sine wave, which is zero at t = 0? From Figure 11.24B you can see that running time backwards results in a reversal of polarity, $\sin(-\omega t) = -\sin(\omega t)$. If you add positive and negative-frequency sine waves of the same frequency and amplitude, they cancel, resulting in zero net signal.

That is why we say that a single sinusoidal signal, $\cos(\omega t)$, actually contains two frequencies, $+\omega$ and $-\omega$. It also offers a logical explanation of why a mixer or modulator produces the sum and difference of the frequencies of the two inputs. For example, an AM modulator produces sidebands at the carrier frequency plus and minus the modulating frequency precisely because those positive and negative frequencies are actually already present in the modulating signal.

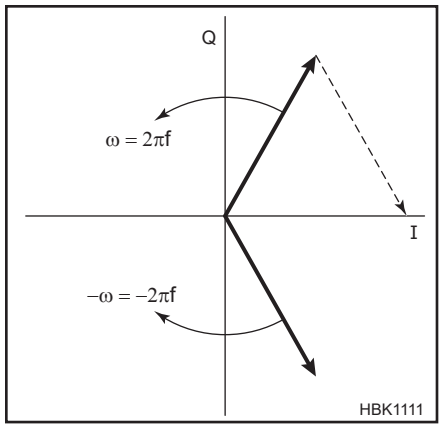

Figure 11.20 — A real signal on the I axis with a single frequency is generated by the sum of I and Q components with positive and negative frequencies.

11.5.2 Analytic Signals

For many purposes, it is useful to separate the portion of the signal that specifies the amplitude and phase (I and Q) from the oscillating part ($\sin(\omega t)$ and $\cos(\omega t)$). For mathematical convenience, the I/Q part is represented by a complex number, x = I + *j*Q. (The **Radio Mathematics** tutorials cover complex numbers.) The oscillating part is also a complex number $e^{-j\omega t} = \cos(\omega t) - j\sin(\omega t)$. (Don't worry if you don't know where that equation comes from — concentrate on the part to the right of the equal sign.) In the equations, $j = \sqrt{}$ of –1. Of course, –1 does not have a real square root (any real number multiplied by itself is positive) so *j*, or any real number multiplied by *j*, is called an *imaginary number*. A number with both real and imaginary parts is called a *complex number*. The total *analytic signal* is a complex number equal to

$$x(t) = xe^{-j\omega t} = (I + jQ)\,(\cos(\omega t) - j\sin(\omega t))$$

In the above equation, the $\cos(\omega t) - \sin(\omega t)$ portion generally represents an RF carrier, with ω being the carrier frequency (a positive or negative value). The I + *j*Q part is the modulation. The *scalar* value of a modulated signal (what you would measure with an oscilloscope) is just the real (Re) part of the analytic signal. Using the fact that $j^2 = -1$,

$$\mathrm{Re}[x(t)] = \mathrm{Re}[(I + jQ)\,(\cos(\omega t) - j\sin(\omega t))]$$

$$\mathrm{Re}[x(t)] = \mathrm{Re}[(I\cos(\omega t) + Q\sin(\omega t)) + j(Q\cos(\omega t) - I\sin(\omega t))]$$

$$\mathrm{Re}[x(t)] = I\cos(\omega t) + Q\sin(\omega t)$$

Note that if the modulation (I and Q) varies with time, the above equation assumes that the modulated signal does not overlap zero Hz. That is, I and Q have no frequency components greater than ω.

An I/Q phasor diagram, such as Figure 11.18, shows only I and Q (the modulation) and not the oscillating part (the RF carrier). The I/Q vector represents the *difference* in phase and amplitude of the RF signal compared to the unmodulated carrier. For example, if the I/Q vector is at 90°, that means the carrier has been phase-shifted by 90° from what it otherwise would have been. If the I/Q vector is rotating counter-clockwise 10 times per second, then the signal has been increased by 10 Hz from the carrier frequency.

It is worth noting that the modulation can be specified either by the in-phase and quadrature (I and Q) values as shown or alternatively by the amplitude and phase. The amplitude is the length of the I/Q vector in the phasor diagram,

$$A = \sqrt{I^2 + Q^2}$$

The phase is the angle of the vector with respect to the +I axis,

$$\varphi = \arctan\left(\frac{Q}{I}\right)$$

An alternative expression for the modulated analytic signal using amplitude and phase is

$$x(t) = Ae^{-j(\varphi+\omega t)} = A\left[\cos(\varphi + \omega t) + j\sin(\varphi + \omega t)\right]$$

and for the scalar signal

$$\mathrm{Re}[x(t)] = A\cos(\varphi + \omega t)$$

One final comment. So far, we have been looking at signals that consist of a single sinusoidal frequency. In any linear system, anything that is true for a single frequency is also true for a combination of many frequencies. Each frequency is affected by the system as though the others were not present. Since any complicated signal can be broken down into a (perhaps large) number of single-frequency sinusoids, all our previous conclusions apply to multi-frequency signals as well.

11.5.3 I/Q Modulation and Demodulation

An *I/Q modulator* is just a device that controls the amplitude and phase of an RF signal directly from the in-phase (I) and quadrature (Q) components. See **Figure 11.21A**. You can think of an I/Q modulator as a device that converts the analytic signal I + *j*Q into a scalar signal at some RF frequency. The spectrum of the I/Q signal, both positive and negative frequencies, is translated upward in frequency so that it is centered on the carrier frequency. Thinking in terms of the phasor diagram, any components of the I/Q signal that are rotating counter-clockwise appear above the carrier frequency and clockwise components appear below.

By using two modulators fed with RF sine waves in quadrature (90° out of phase with each other), any amplitude and phase may be obtained by varying the amplitudes of the two modulation inputs. The input labeled I (for in-phase) moves the symbol location horizontally in the constellation diagram and the one labeled Q (for quadrature) moves it vertically. For example, to obtain an amplitude of 1 and a phase angle of –45°, set I to +0.707 and set Q to –0.707.

It is possible to generate virtually any type of modulation using an I/Q modulator. For example, to generate BPSK or on-off keying, simply disconnect the Q input and apply the modulation to I. For angle modulation, such as FM or PM, a waveform generator applies a varying signal to I and Q in such a manner

to cause the symbol to rotate at constant amplitude with the correct phase and frequency (rotation rate). Even a multi-carrier signal may be generated with a single I/Q modulator by applying the sum of a number of signals, each representing one carrier, to the I and Q inputs. The phase of each signal rotates at a rate equal to the frequency offset of its carrier from the center.

An *I/Q demodulator* is basically an I/Q modulator in reverse. It generates I and Q signals that represent the in-phase and quadrature components of the incoming RF signal. See Figure 11.21B. Assuming the demodulator's local oscillator is on the same frequency and is in phase with the carrier of the signal being received then the I/Q output of the receiver's demodulator is theoretically identical to the I/Q input at the transmitter end.

If the demodulator's local oscillator is not tuned to exactly the same frequency as the one in the transmitter (after down-conversion to the IF in a superhet receiver) then the demodulated signal will rotate in the I/Q plane at a rate equal to the frequency error. Most receivers include a *carrier-recovery* circuit, which phase-locks the local oscillator to the average frequency of the incoming signal to obtain stable demodulation. While that corrects the frequency error, it does not correct the phase, which must be accounted for in some other manner such as by using differential modulation or some kind of symbol-recovery mechanism.

I/Q modulators and demodulators can be built with analog components. The LO could be a transistor oscillator and the 90° phase-shift network could be implemented with coils and capacitors. The circles with the multiplication symbol would be double-balanced mixers. Also required would be trim adjustments to balance the amplitude between the I and Q channels and to adjust the phase shift as close as possible to 90°.

No analog circuit is perfect, however. If the 90° phase-shift is not exactly 90° or the amplitudes of the I and Q channels are not perfectly balanced, you don't get perfect opposite-sideband rejection. The modulator output includes a little bit of signal on the unwanted sideband and the I/Q signal from the demodulator includes a small signal rotating in the wrong direction. If there is a small dc offset in the amplifiers feeding the modulator's I/Q inputs, that shows up as carrier feedthrough. On receive, a dc offset makes the demodulator think there is a small signal at a constant amplitude and phase angle that is always there even when no actual signal is being received. Nor is analog circuitry distortion-free, especially the mixers. Intermodulation distortion shows up as out-of-channel "splatter" on transmit and unwanted out-of-channel responses on receive.

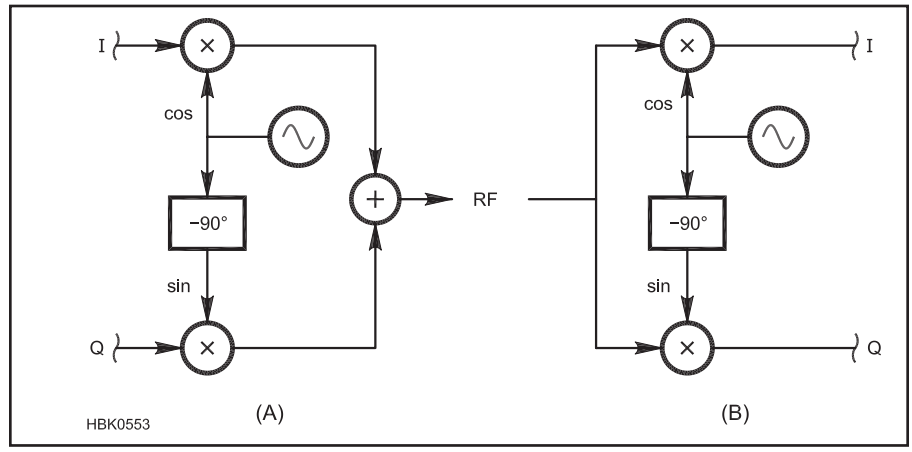

Figure 11.21 — I/Q modulator (A) and demodulator (B). The demodulator is essentially the same circuit as the modulator but operated in the reverse direction.

All those problems can be avoided by performing these operations digitally. If the analog I/Q inputs to the modulator are converted to streams of digital numbers with a pair of ADCs, then the mixers, oscillator, phase-shift and summing functions can all be digital. In many systems, the I and Q signals are also generated digitally, so that the digital output signal has perfect unwanted sideband rejection, no carrier feedthrough and no distortion within the dynamic range afforded by the number of bits in the data words. A similar argument holds for a digital demodulator. If the incoming RF signal is first digitized with an ADC, then the demodulation can be done digitally without any of the artifacts caused by imperfections in the analog circuitry.

I/Q modulators and demodulators can also be used as so-called *imageless mixers*. A normal mixer with inputs at f_1 and f_2 produces outputs at f_1, f_2, $f_1 + f_2$, and $f_1 - f_2$. A balanced mixer eliminates the f_1 and f_2 terms but both the sum and difference terms remain, even though normally only one is desired. By feeding an RF instead of AF signal into the input of an SSB modulator, we can choose the sum or difference frequency in the same way as choosing the upper or lower sideband. If the input signal is a sine wave, the Hilbert transformer can be replaced by a simple 90° phase shifter. Similarly, a mixer with the same architecture as an SSB demodulator can be used to down-convert an RF signal to IF with zero image response. Analog imageless mixers are covered in the **Receiving** chapter. They are sometimes used in microwave receivers and transmitters where it is difficult to build filters narrow enough to reject the image response, but they typically only achieve image rejection in the 20-30 dB range. With a digital imageless mixer, the image rejection is "perfect" within the dynamic range of the bit resolution.

11.6 Applications of I/Q Modulation

11.6.1 Quadrature Modulation

Quadrature modulation encodes digital signals using a combination of amplitude and phase modulation. With two types of modulation to work with, it is possible to pack more data bits into each modulation symbol, which allows more throughput for a given bandwidth. Modulation formats in use today have up to 8 or more bits per symbol which would be impractica1 for ASK, FSK or PSK alone.

The modulation states of the various digital formats map to positions in the phasor diagram, what is called a *constellation diagram*. The transmitter generates the correct modulation states by placing the correct values on the I and Q inputs to the I/Q modulator. In the receiver, the filtered I and Q values are sampled at the symbol decision times to determine which modulation state they most closely match.

Since quadrature-modulated symbols are defined by both amplitude and phase, the most common way to represent symbol states is on the constellation diagram. **Figure 11.22** shows a *four-level quadrature amplitude modulation* (4-QAM) signal, often referred to simply as QAM. The distance of each state from the origin represents the amplitude. The phase angle with respect to the +I axis represents the phase. The four states shown have phase angles of +45°, +135°, –45° and –135°. Normally, the receiver has no absolute phase

information and can only detect phase differences between the states. For that reason, we could just as easily have drawn the four states directly on the I and Q axes, at 0°, +90°, 180° and –90°. Note that each of the four states has the same amplitude, differing only in phase. For that reason, 4-QAM is often referred to as *quadrature phase-shift keying* (QPSK), in the same manner that two-level ASK is normally referred to as BPSK.

Since 4-QAM has four possible states, there are two data bits per symbol. However, QAM is not limited to four states. **Figure 11.23** illustrates a 256-QAM signal. It takes 8 bits to represent 256 states, so 256-QAM packs 8 data bits into each symbol. The disadvantage of using lots of states is that the effect of noise and interference is worse at the receiver. Since, for the same peak power, the states of a 256-QAM signal are 15 times closer together than with 4-QAM, the receiver's decoder has to determine each symbol's location to 15 times greater accuracy. That means the ratio of peak power to noise and interference must be 20 log(15) = 23.5 dB greater for accurate decoding. That is why QAM with a large number of states is normally not used on the HF bands where fading, noise and interference are common. A more common application is digital cable television where the coaxial-cable transmission channel is much cleaner. For example, the European DVB-C standard provides for 16-QAM through 256-QAM, depending on bandwidth and data rate.

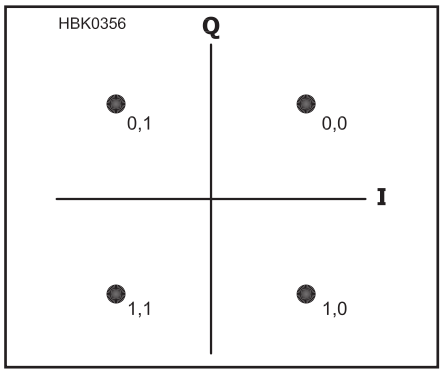

Figure 11.22 — Constellation diagram of a 4-QAM signal, also known as QPSK. The four symbol locations all have the same amplitude and have phase angles of 45°, 135°, –135° and –45°. 4-QAM is a two-bits-per-symbol format. The four symbols are selected by the four possible states of the two data bits, which can be assigned in any order. With the assignment shown, the Q value depends only on the first bit and the I value depends only on the second, an arrangement that can simplify symbol encoding.

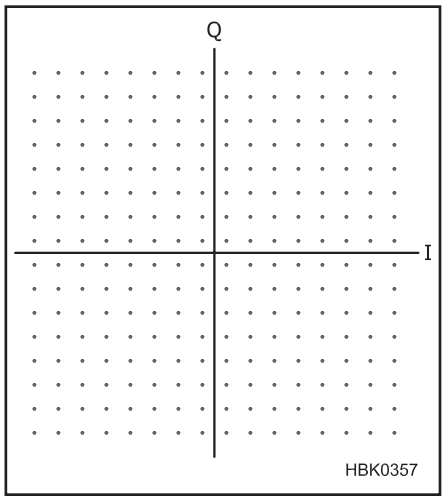

Figure 11.23 — Constellation diagram of a 256-QAM signal. Since an 8-bit number has 256 possible states, each symbol represents 8 bits of data.

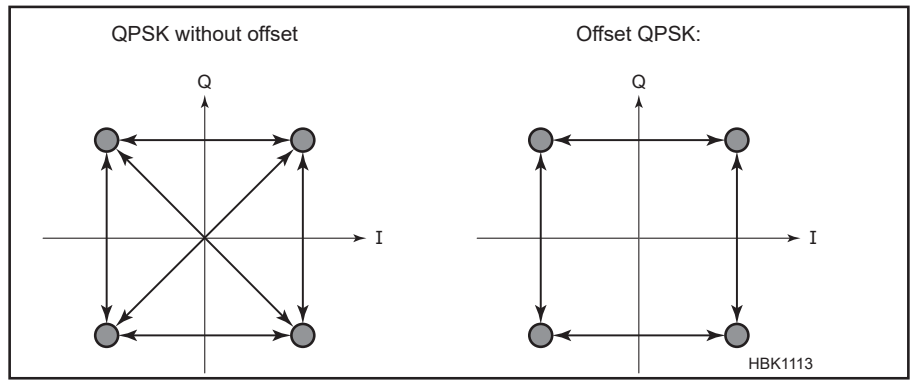

Figure 11.24 — Constellation diagrams of QPSK and OQPSK. Transitions from symbol to symbol for QPSK can pass through the origin where amplifier linearity can be poor. OQPSK avoids this problem with two-step symbol transitions which eliminate direct corner-to-corner transitions.

One problem with QAM is that whenever the signal trajectory between two symbol states passes through the origin, the signal amplitude momentarily goes to zero. That imposes stringent linearity requirements on the RF power amplifier, since many amplifiers exhibit their worst linearity near zero power. One solution is *offset QPSK* (OQPSK) modulation as shown in **Figure 11.24**. In this case, the symbol transitions of the I and Q channels are offset by half a symbol. That is, for each symbol, the I channel changes state first then the Q channel changes half a symbol time later. That allows the symbol trajectory to sidestep around the origin, allowing use of a higher-efficiency or lower-cost power amplifier that has worse linearity.

Another solution to the zero-crossing problem is called *Pi-over-4 differential QPSK* (π/4 DQPSK). See **Figure 11.25**. This is actually a form of 8-PSK, where the eight symbol locations are located every 45° around a constant-amplitude circle. On any given symbol, however, only four of the symbol locations are used. The symbol location always changes by an odd multiple of 45° (π/4 radians). If the current symbol is located on the I or Q axis, then on the next symbol only the four non-axis locations are available and vice versa. As with OQPSK, that avoids transitions that pass through the origin.

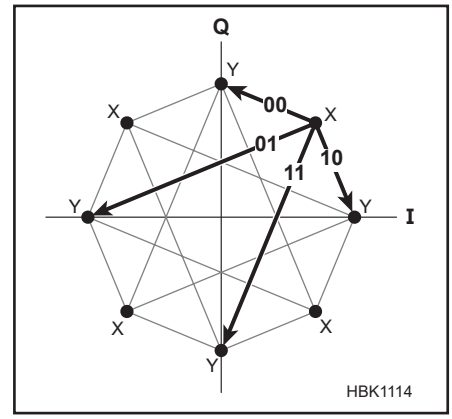

Figure 11.25 — Constellation diagram of a π/4 DQPSK signal. There are eight possible symbol locations. If the current symbol location is at one of the four positions labeled "X" then the next symbol will be at one of the four locations labeled "Y" and vice versa. That guarantees that no possible symbol trajectories (the lines between symbol locations) can ever pass through the origin.

Another advantage of π/4 DQPSK which is shared with other types of differential modulation is that absolute phase doesn't matter. The information is encoded only in the *difference* in the phase of successive symbols. That greatly simplifies the job of the receiver's demodulator.

11.6.2 AM Using I/Q Modulation

How do you make an AM modulator and demodulator using DSP and software? The modulator is easy. Simply add a constant value, representing the carrier, to the audio signal and multiply the result by a sine wave at the carrier frequency, as shown in **Figure 11.26**.

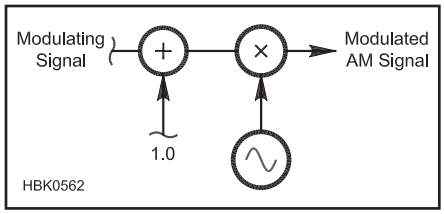

Figure 11.26 — A digital AM modulator

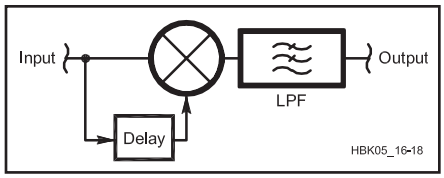

Figure 11.27 — A digital quadrature detector.

Demodulation is almost as easy. We could just simulate a full-wave rectifier by taking the absolute value of the signal, as mentioned previously, and low-pass filter the result to remove the RF energy. If the signal to be demodulated is complex, with I and Q components, then instead of absolute value we take the magnitude

$$A = \sqrt{I^2 + Q^2}$$

The dc bias can be removed by adding a "series blocking capacitor" — a high-pass filter with a suitable cut-off frequency.

A little more elegant way to do it would be to include the AM detector as part of the AGC loop. In the C code snippet shown in **Table 11.1**, the variable "carrier" is the average AM carrier level. It is passed to another subroutine to control the gain.

Note that no "series capacitor" is needed since the audio signal is computed by subtracting the average historical value, carrier, from the magnitude of the current I/Q signal, am. A small fraction of its value is added to the historical value so that the AGC tracks the average AM carrier level. AGC speed is controlled by that fraction. Dividing by 2^{10} = 1024 gives a time constant of about 1024 clock cycles.

Another type of detector we haven't discussed yet is for frequency modulation. For a scalar signal, the *quadrature detector* shown in **Figure 11.27** is one elegant solution. This is the same circuit whose analog equivalent has been used in millions of FM receivers around the world. In the digital implementation, the delay block is a FIFO buffer constructed from a series of shift registers. Multiplying the signal by a delayed version of itself gives an output with a cosinusoidal response versus frequency. The response crosses zero whenever the carrier frequency is $f = N/(4\tau)$, where N is an odd integer and the delay in seconds is τ=n/f_s, where n is the number of samples of delay and f_s is the sample frequency. As the carrier deviates above and below the zero-crossing frequency the output varies above and below zero, just what we want for an FM detector.

Table 11.1
Software AM Detector

```
static long int carrier;
long int am;
int i, q, signal;

/* Code that generates i and q omitted */
am = (long int)sqrt((long int)i*i + (long int)q*q);
signal = am - carrier;
// Divide signal by 2^10:
carrier += signal >> 10;
// Send audio output to DAC:
```

For an I/Q signal, probably the most straightforward FM detector is a phase detector followed by a differentiator to remove the 6 dB per octave rolloff caused by the phase detector. The phase is just

$$\varphi = \arctan\left(\frac{Q}{I}\right)$$

You have to be a little careful since there is a 180° phase ambiguity in the arctangent function. For example,

$$\arctan\left(\frac{1}{1}\right) = \arctan\left(\frac{-1}{-1}\right)$$

Software will have to check which quadrant of the phasor diagram the I/Q signal is in and add 180° when necessary. If there is no arctan function in the library, one can be constructed using a look-up table. Frequency is the derivative of the phase. A differentiator is nothing more than a subtractor that takes the difference between successive samples.

$$f = \frac{\varphi_n - \varphi_{n-1}}{2\pi f_s}$$

where n is the sample number and f_s is the sample rate. It is important to make sure that the difference equation functions properly around 360°. If the phase variable is scaled so that 360° equals the difference between the highest and lowest representable numbers, then standard two's complement subtraction should roll over to the right value at the 360° to 0° transition. Another thing to watch out for is that the derivative of the phase may be a rather small signal, so it might be necessary to carry through all the calculations using long integers or floating-point numbers.

11.6.3 SSB Using I/Q Modulation

This section walks through the process of generating an upper-sideband signal using an I/Q modulator. See **Figure 11.28**. We'll first describe the mathematics in the following paragraph and then give the equivalent explanation using the phasor diagram.

The modulating signal is a sine wave at a frequency of ω_m radians per second ($\omega_m / 2\pi$

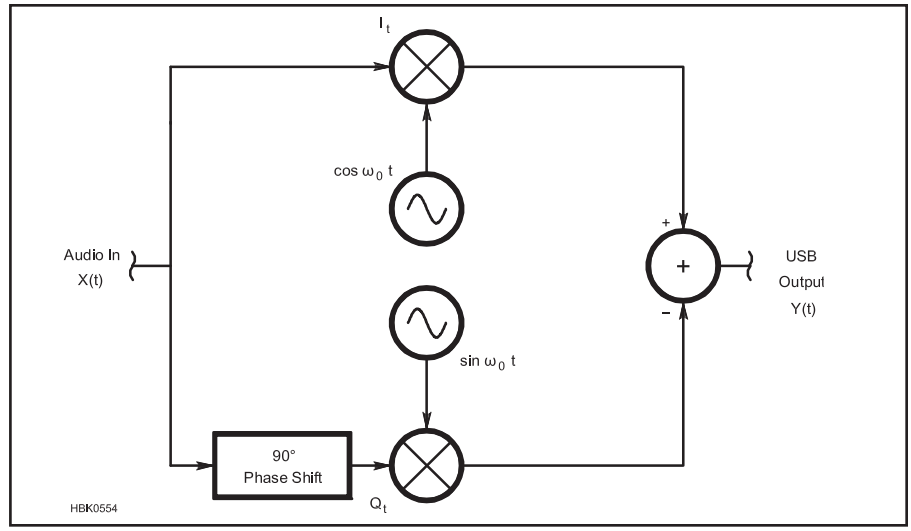

Figure 11.28 — Generating a USB signal with an I/Q modulator.

cycles per second). Because ω_m is a positive frequency the signals applied to the I/Q inputs are $I(t) = \cos(\omega_m t)$ and $Q(t) = \sin(\omega_m t)$. Assume the modulating frequency ω_m is much less than the RF frequency ω. The analytic signal is

$$x(t) = A\cos(\omega_n t) + j\sin(\omega_m t)x\left[\cos(\omega t) + j\sin(\omega t)\right]$$

so that the real, scalar signal that appears at the modulator output is

$$\mathrm{Re}[x(t)] = \cos(\omega_m t)\cos(\omega t) + \sin(\omega_m t)\sin(\omega t)$$

At the moment when t = 0, then $\cos(\omega_m t) = 1$ and $\sin(\omega_m t) = 0$, so the real signal is just $\cos(\omega t)$, the RF signal with zero phase. One quarter of a modulation cycle later $\omega_m t = \pi/2$, so $\cos(\omega_m t) = 0$ and $\sin(\omega_m t) = 1$, and the real signal is now $\sin(\omega t)$, the RF signal with a phase of $+\pi/2$, or +90°. Every quarter cycle of the modulating signal, the RF phase, increases by 90°. That means that the RF phase increases by one full cycle for every cycle of the modulation, which is another way of saying the frequency has shifted by ω_m. This creates an upper sideband at a frequency of $\omega + \omega_m$.

On the phasor diagram, the I/Q signal is rotating counterclockwise at a frequency of ω_m radians per second. As it rotates it is increasing the phase of the RF signal at the same rate, which causes the frequency to increase by ω_m radians per second. To cause the phasor to rotate in the opposite direction, creating a lower sideband, you could change the polarity of either I or Q or you could swap the I and Q inputs.

THE HILBERT TRANSFORMER

For this process to work, the baseband signals applied to the I and Q inputs must be 90° out of phase. That's not hard to do for a single sine wave, but to generate a voice SSB signal, all frequencies in the audio range must be simultaneously phase-shifted by 90° without changing their amplitudes. To do that with analog components requires a broadband phase-shift network consisting of an array of precision resistors and capacitors and a number of operational amplifiers.

To do that with DSP requires a *Hilbert transformer*, an FIR filter with a constant 90° phase shift at all frequencies. Recall that a symmetrical FIR filter has a constant *delay* at all frequencies. That means that the phase shift is not constant — it increases linearly with frequency. It turns out that with an *anti-symmetrical* filter, in which the top half of the coefficients are the negative of the mirror image of the lower half, the phase shift is 90° at all frequencies, which is exactly what we need to generate an SSB signal.

The Hilbert transformer is connected in series with either the I or Q input, depending on whether USB or LSB is desired. Just as with any FIR filter, a Hilbert transformer has a delay equal to half its length, so an equal delay must be included in the other I/Q channel as shown in **Figure 11.29**. It is possible to combine the Hilbert transformer with the normal FIR filter that may be needed anyway to filter the baseband signal. The other I/Q

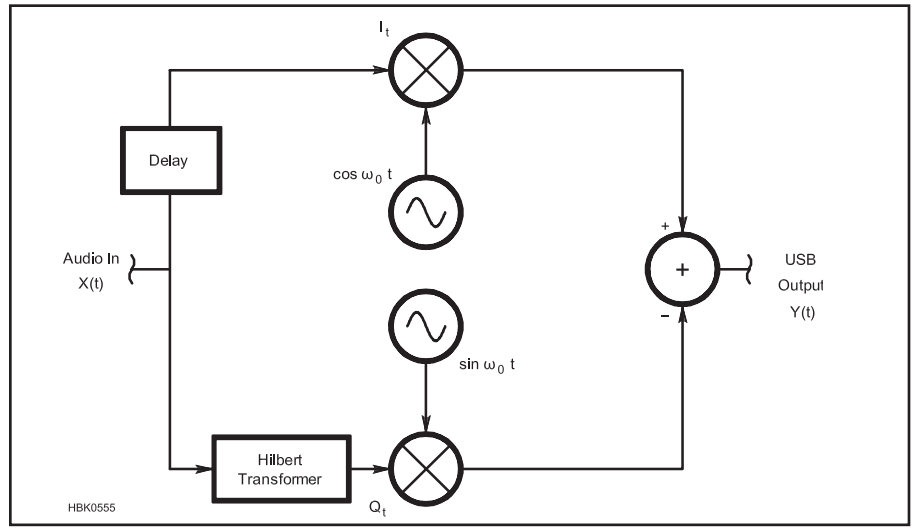

Figure 11.29 — Generating USB for a non-sinusoidal audio or speech with an I/Q modulator.

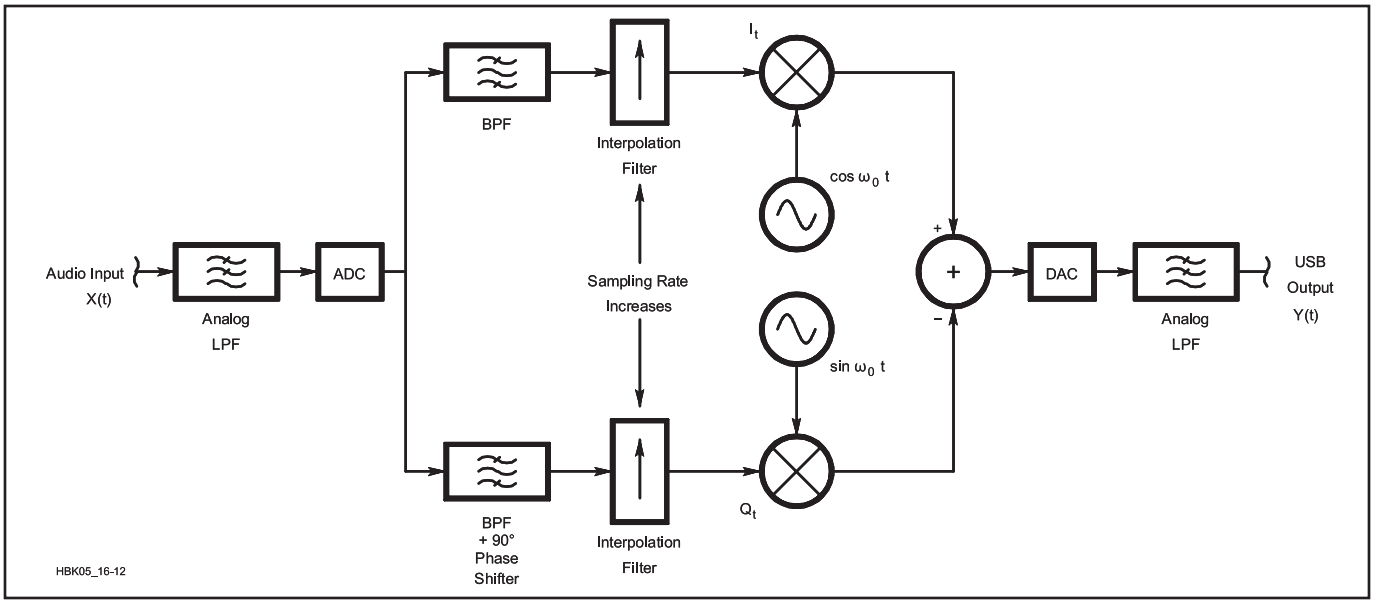

Figure 11.30 — Block diagram of a digital SSB modulator.

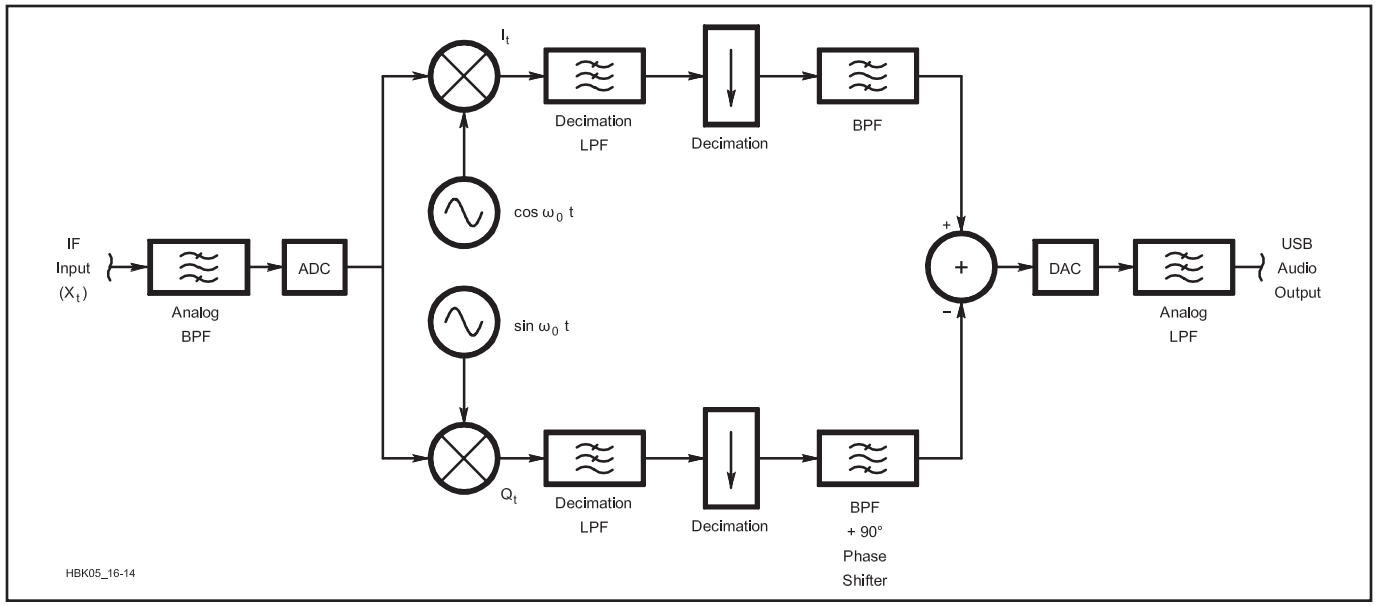

Figure 11.31 — Block diagram of a digital SSB demodulator.

channel then simply uses a similar filter with the same delay but without the 90° phase shift.

Because the RF output of the modulator is normally at a much higher frequency than the audio signal, it is customary to use a higher sample rate for the output signal than for the input. The FIR filters can still run at the lower rate to save processing time, and their output is then up-sampled to a higher rate with an interpolator. It is convenient to use an output sample rate that is exactly four times the carrier frequency because each sample advances the RF phase by exactly 90°. The sequence of values for the sine wave is 0, 1, 0 and –1. To generate the 90° phase shift for the cosine wave, simply start the sequence at the second sample: 1, 0, –1, 0. The complete block diagram is shown in **Figure 11.30**.

This is also known as the "phasing method" of SSB generation. It was popular when SSB first became common on the amateur bands back in the 1950s because suitable crystal filters were expensive or difficult to obtain. The phasing method had the reputation of producing signals with excellent-quality audio, no doubt due to the lack of the phase distortion caused by crystal filters.

A Hilbert transformer may also be used in an SSB demodulator at the receiver end of the communications system. It is basically the same block diagram drawn backwards, as illustrated in **Figure 11.31**.

It is important to note that an ideal Hilbert transformer is impossible to construct because it theoretically has an infinitely-long impulse response. However, with a sufficiently-long impulse response, the accuracy is much better than an analog phase-shift network. Just as with an analog network, the frequency passband must be limited both at the low end as well as the high end. That is, the audio must be bandpass filtered before the 90° phase shift. Actually, the filtering and phase shifting can be combined into one operation using the following method.

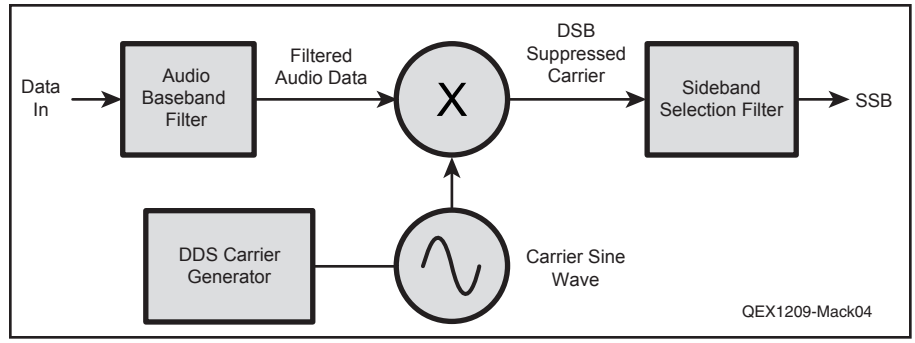

Figure 11.32 — The DSP method of using filters to create a SSB signal is simple and equivalent to the analog filter method.

First design a low-pass FIR filter with a bandwidth one-half the desired audio bandwidth. For example, if the desired passband is 300 to 2700 Hz, the low-pass filter bandwidth should be (2700 – 300)/2 = 1200 Hz. Then multiply the impulse response coefficients with a sine wave of a frequency equal to the center frequency of the desired passband, (2700 + 300)/2 = 1500 Hz in this case. That results in a band-pass filter with the desired 300 – 2700 Hz response. By using sine waves 90° out of phase for the I and Q channels, you end up with two bandpass filters with the same amplitude response and delay but a 90° phase difference at all frequencies. Multiply by a cosine for zero phase and by a sine for a 90° phase shift.

This bears a striking resemblance to the Weaver method, the so-called "third method" of SSB generation (after "filter" and "phasing"). It was originally used back in the late 1950s to eliminate the need for a wide-band audio phase-shift network. It is almost as if there is no such thing as truly new technology, just the same mathematics implemented with different technology!

THE FILTER METHOD IN DSP

As DSP hardware has grown more powerful, it has also become possible to implement the filter method of SSB generation in software. In fact, the DSP filters that can be constructed in software are quite a bit better than the best analog filters. The result is a very simple system as shown in **Figure 11.32** that is equivalent to the analog method. Ray Mack, W5IFS, discusses the various tradeoffs and filter characteristics required of this method in his Sep/Oct 2012 *QEX* SDR Simplified column (included with this book's online content).

Phil Karn, KA9Q, uses a DSP-based filter method for SSB generation because it avoids

the need for a Hilbert transformer function. He performs his filtering by converting the signal to the frequency domain with an FFT (see the **DSP and SDR Fundamentals** chapter), multiplies each frequency component by a constant that defines the desired filter shape, and then converts back to the time domain with the inverse FFT. The windowing functions discussed with the FFT material can be applied to this filtering operation as well, with the Kaiser window recommended as supporting smooth trade-offs for sharp response against stop-band attenuation.

Karn has developed a fairly general purpose package in C that down-converts, filters, and detects a range of modulation/demodulates methods. The modulation/demodulation package and code for a WWV emulator is available as a github repository. For more information, see this book's website at **www.arrl.org/arrl-handbook-reference**. It currently demodulates AM, CAM (coherent AM), LSB, USB, FM, ISB (independent sideband) and IQ (straight IQ pass through). The software should work on processors that support C and floating point operations. It's not yet a polished product for the end user, but rather a set of building blocks for the experimenter.

11.6.4 Multi-Carrier Modulation

An effective method to fit more data bits into each symbol is to use more than one separately-modulated signal at a time, each on its own carrier frequency spaced an appropriate distance from the frequencies of the other carriers. An example is multi-carrier FSK. This is not to be confused with MFSK which also uses multiple frequencies, but only one at a time. With multi-carrier FSK, each carrier is present continuously and is frequency-shifted in response to a separate data stream. The total data rate equals the data rate of one carrier times the number of carriers. A disadvantage of multi-carrier FSK is that the resulting signal is no longer constant-amplitude — a linear amplifier must be used. In general, multi-carrier signals using any modulation type on each carrier tend to have high peak-to-average power ratios.

In the presence of selective fading, one or more of the carriers may disappear while the others are still present. An advantage of multi-carrier modulation is that error-correcting coding can use the unaffected carriers to reconstruct the missing data. Also, since each carrier signal is relatively narrowband, propagation conditions are essentially constant within that bandwidth. That makes it easier for the receiver to correct for other frequency-selective propagation impairments such as phase distortion. If a single-carrier signal of the same total bandwidth had been used instead, the receiver would need an adaptive equalizer to correct for the amplitude and phase variations across the transmission channel.

By using multiple carriers each with multiple-bit-per-symbol modulation it is possible to obtain quite high data rates while maintaining the low symbol rates that are required to combat the effects of multi-path propagation on the HF bands. For example, PACTOR-

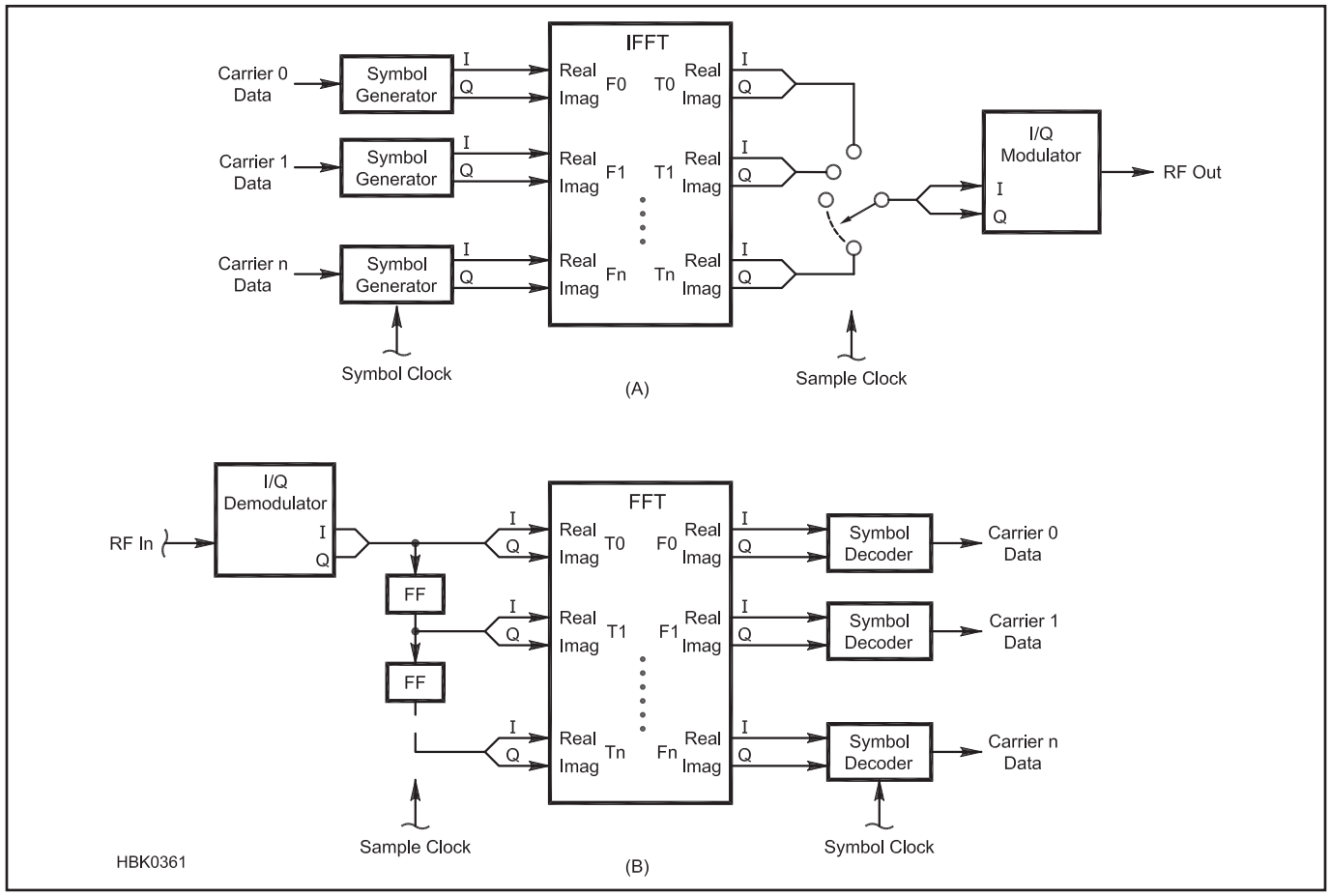

Figure 11.33 — Block diagram of an OFDM modulator (A) and demodulator (B) using the FFT/IFFT technique. The number of carriers is n, and the sample rate is n times the symbol rate. Once per symbol, all the symbol generators in the modulator are loaded with new data and the inverse fast-Fourier transform (IFFT) generates n output samples, which are selected in succession by the switch. In the demodulator, n samples are stored in a shift register (string of flip-flops) for each symbol, then the FFT generates one "frequency" output for each carrier frequency. From the amplitude and phase of each frequency, the symbol decoders can determine the symbol locations and thus the data.

III achieves a raw data rate of 3600 bits per second with a 100-baud symbol rate using 18 carriers of DQPSK. (100 bauds × 2 bits/symbol × 18 carriers = 3600 bits per second.) Similarly, Clover-2000 modulation gets 3000 bits per second with a 62.5-baud symbol rate using eight carriers of 16-DPSK combined with 4-level DASK. (62.5 bauds x (4+2) bits/symbol × 8 carriers = 3000 bits per second.) Decoding is rather fragile using these complex modulation techniques, so PACTOR and Clover include means to automatically switch to simpler, more-robust modulation types as propagation conditions deteriorate.

What is the minimum carrier spacing that can be used without excessive interference between signals on adjacent frequencies? The answer depends on the symbol rate and the filtering. It turns out that it is easy to design the filtering to be insensitive to interference on frequencies that are spaced at integer multiples of the symbol rate. (See the sub-section on filtering and bandwidth later in this chapter.) For that reason, it is common to use a carrier spacing equal to the symbol rate. The carriers are said to be *orthogonal* to each other since each theoretically has zero correlation with the others.

Orthogonal frequency-division multiplexing (OFDM), sometimes called *coded OFDM* (COFDM), refers to the multiplexing of multiple data streams onto a series of such orthogonal carriers. The term usually implies a system with a large number of carriers. In that case, an efficient decoding method is to use a DSP algorithm called the fast Fourier transform (FFT). (See the **DSP and SDR Fundamentals** chapter) The FFT is the software equivalent of a hardware spectrum analyzer. It gathers a series of samples of a signal taken at regular time intervals and outputs another series of samples representing the frequency spectrum of the signal. See **Figure 11.33**. If the length of the series of input samples equals one symbol time and if the sample rate is selected properly, then each frequency sample of the FFT output corresponds to one carrier. Each frequency sample is a complex number (containing a "real" and "imaginary" part) that represents the amplitude and phase of one of the carriers during that symbol period. Knowing the amplitude and phase of a carrier is all the information required to determine the symbol location in the I/Q diagram and thus decode the data. At the transmitter end of the circuit, an inverse FFT (IFFT) can be used to encode the data, that is, to convert the amplitudes and phases of each of the carriers into a series of I and Q time samples to send to the I/Q modulator. See Figure 11.33A.

One advantage of OFDM is high spectral efficiency. The carriers are spaced as closely as theoretically possible and, because of the narrow bandwidth of each carrier, the overall spectrum is very square in shape with a sharp drop-off at the passband edges. One disadvantage is that the receiver must be tuned very accurately to the transmitter's frequency to avoid loss of orthogonality, which causes cross-talk between the carriers.

11.7 Image Modulation

The following section covers the modulation of amateur television communications. Both full-motion (*fast-scan*) and still-frame (*slow-scan*) images can be transmitted. Fast-scan amateur television or ATV is full-motion, commercial broadcast quality, color TV video and audio using very wide channel bandwidths of several MHz. Bandwidth restrictions restrict the use of wide-band fast-scan signals to the 70 cm band and higher frequencies. Amateurs also transmit slow-scan TV (SSTV) on the HF bands below 30 MHz in a SSB voice channel bandwidth.

More detailed information on the operating conventions for amateur television can be found in the **Image Communications** operating chapter in this book's online content. Mesh network video using IP protocols over a wireless network such as AREDN is not considered in this section. See the section on High-Speed Multimedia in the **Digital Protocols and Modes** chapter.

11.7.1 Fast-Scan Television

NTSC BASEBAND VIDEO

NTSC is an abbreviation for the National Television Standards Committee that developed the modulation standard. It was released initially in 1941 (black-and-white only) and again in 1953 (color). It remained the dominant television standard until the late 1990s. The *baseband* portion of the standard describes the video signal, such as is output by a camera or video recorder before it is used to modulate an RF signal. (The current NTSC standard is SMPTE 170M — **www.smpte.org**.)

The complete NTSC video picture consists of 525 *horizontal scan lines* organized in two interlaced *even* and *odd fields* which contain the even-numbered or odd-numbered lines. (*Scanning* refers to moving an electron beam across the screen of a cathode ray tube to create the picture.) The scan lines are organized to display the picture from left to right and top to bottom.

The fields alternate at 60 Hz to display one complete *frame* 30 times per second. (The actual rates are 59.94 fields/sec and 39.97 frames/sec.) Of the 525 lines, only 480 lines are visible. The interlaced horizontal lines are the reason for the video designation for NTSC video as 480i. The complete image shown on a monitor is called a *raster*.

The video signal includes pulses to synchronize the vertical and horizontal scan display circuitry or software in the receiver. See **Figure 11.34A** which defines the basic elements of a single line in an NTSC signal. Figure 11.34B shows one horizontal line of a typical baseband NTSC video waveform in a picture of six vertical color bars. Each line contains six short segments of different colors which are seen as the sequential bursts of color information.

The video *luminance* (brightness) is represented by the signal's amplitude. The waveform's top level of 1 V represents white. The level representing black can be seen as the straight line on the far right of Figure 11.34B following the final color bar segment. Signal amplitudes below this level are termed "blacker than black." The sync pulses and the *front porch* and *back porch* areas that bracket them (see Figure 11.34A) are blacker than black.

The original function of using blacker-than-black signal levels was to turn off or *blank* the electron beam in a cathode ray tube as it returned to the left side or top of the picture without creating any visible lines or artifacts. In computer-type displays there is no electron beam but the pulses are used for timing synchronization.

Each line of video starts with a negative-going, 4.88 µs duration, *horizontal sync pulse* that occurs at a 15.734 kHz rate. The horizontal sync pulse marks the start (left-hand side) of a single horizontal line of video information. A longer, *vertical sync pulse* (not shown) occurs between the interlaced frames of information to reset the display to the top of the screen.

The horizontal sync pulse is followed by an eight-cycle sine wave *color burst* at

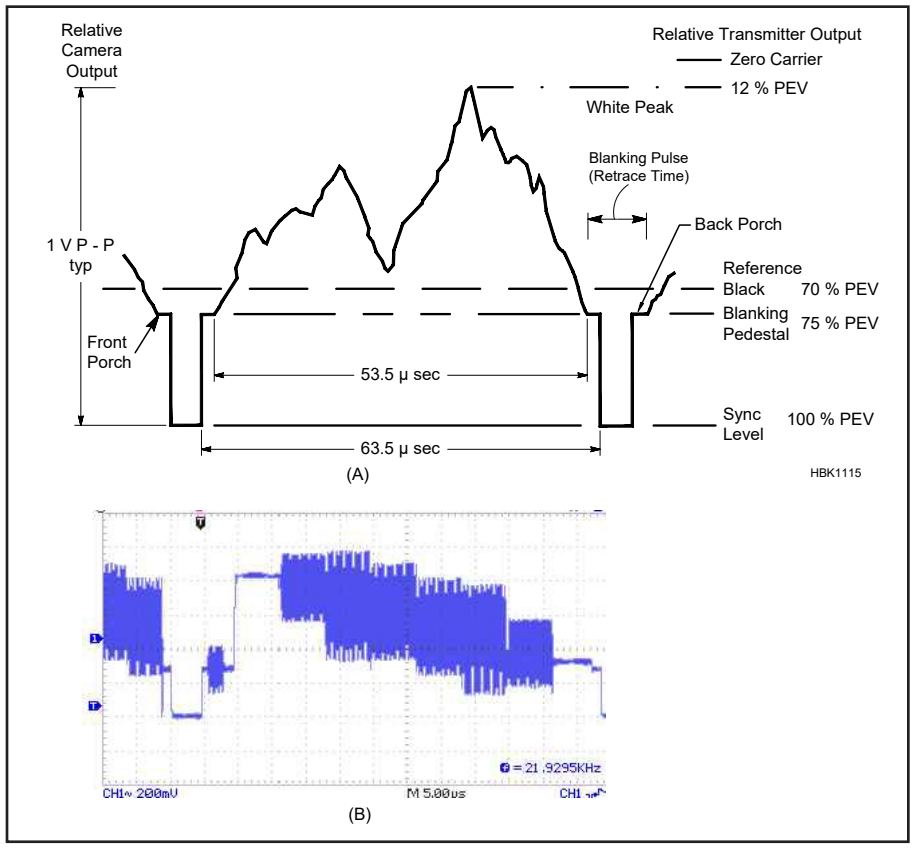

Figure 11.34 — Part A illustrates an ATV waveform, showing the relative camera output as well as the transmitter output RF power during one horizontal line scan for black-and-white TV. (A color camera would generate a "burst" of 8 cycles at 3.58 MHz on the back porch of the blanking pedestal.) Note that "black" corresponds to a higher transmitter power than "white." Part B shows the NTSC video waveform. One horizontal scan line is shown. Vertical scale is 200 mV/div and horizontal scale is 5 µs/div.

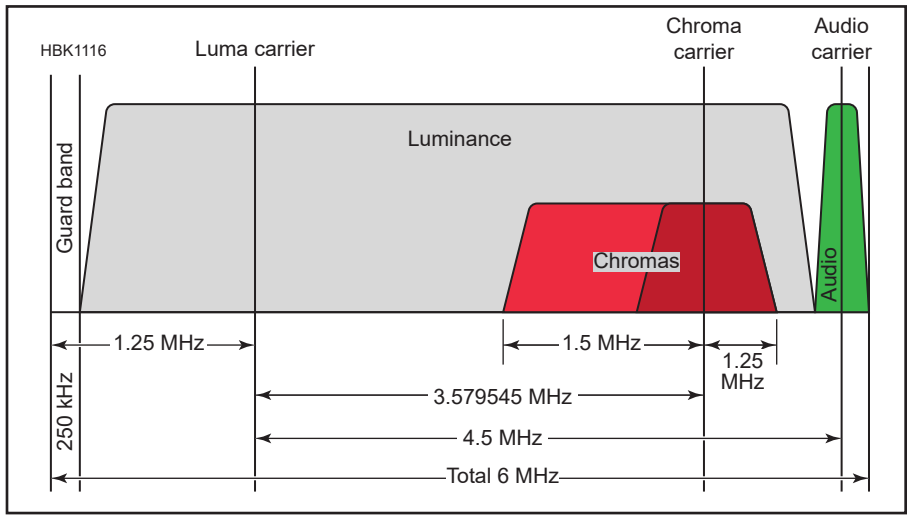

Figure 11.35 — The spectrum of an analog NTSC, VUSB-TV signal.

3.58 MHz (actually 3.579545 MHz) in the middle of the back porch. The color burst is used as a reference frequency for the *chroma* or color information and the internal oscillator used by the receiving display. The video information for this line follows the color burst.

ANALOG FAST-SCAN ATV (ATV)

Although obsolete commercially, there is a lot of ATV activity using the same NTSC modulation standard for US broadcast analog television stations prior to the change to digital TV in 2009. Analog TV receivers can be used to receive and display analog ATV signals.

Figure 11.35 shows the spectrum of an NTSC analog TV full-carrier, double-sideband AM signal filtered to partially remove the lower sideband. The partially-filtered *vestigial sideband* (VSB) extends 1.25 MHz below the carrier frequency. This is called VUSB-TV.

The resulting TV channel is 6 MHz wide to accommodate the composite video and two sub-carriers. One sub-carrier is at 3.58 MHz for the color information and the other is an FM-modulated sub-carrier at 4.5 MHz for the sound. The video bandwidth is limited to 4.2 MHz but the overall TV channel is 6 MHz wide to accommodate both the audio and video information.

In an analog AM-TV transmitter, the RF envelope waveform is inverted from the baseband video waveform. The sync pulses thus become the strongest portion of the envelope and the white levels are the weakest. An AM-TV transmitter is rated like a SSB transmitter with a peak envelope power (PEP) representing the tip of the sync pulses. The video-to-sync ratio must remain constant throughout all of the linear amplifiers in the transmit chain as the video level from the camera changes. To maintain the *sync tips* (bottom of each sync pulse) at 100% of peak power, the modulator usually contains a clamp circuit that also acts as a *sync stretcher* to compensate for amplifier gain compression.

For simplicity in transmitter design, analog ATV stations often transmit an unfiltered full-DSB AM signal, but the normally-removed portion of the lower sideband is unused by the TV receiver. (To conserve spectrum, this practice is discouraged.) The bulk of an NTSC TV signal's energy is contained close to the video carrier and within ±1 MHz of the video carrier frequency. To reduce signal bandwidth, a 6 MHz wide bandpass filter is used after DSB-AM modulation to attenuate the unused lower sideband color and sound subcarrier frequencies by 20-30 dB. This creates a Vestigial Upper Sideband or VUSB-TV signal with a spectrum as shown in Figure 11.35.

With the typical analog TV receiver's 3-dB rolloff at 3 MHz (primarily in the IF filter), up to 480 pixels per horizontal line can be seen. Color bandwidth in a TV set is lower than that, resulting in up to 100 color lines. (This is referred to as *lines of resolution* and should not be confused with the number of horizontal scan lines per frame.)

The PAL analog TV system is an AM system based on NTSC. The SECAM system uses FM color subcarriers. As described previously, FM achieves superior noise and interference suppression for signal levels above a certain threshold, although AM seems to work better for receiver signal levels below about 5 µV.

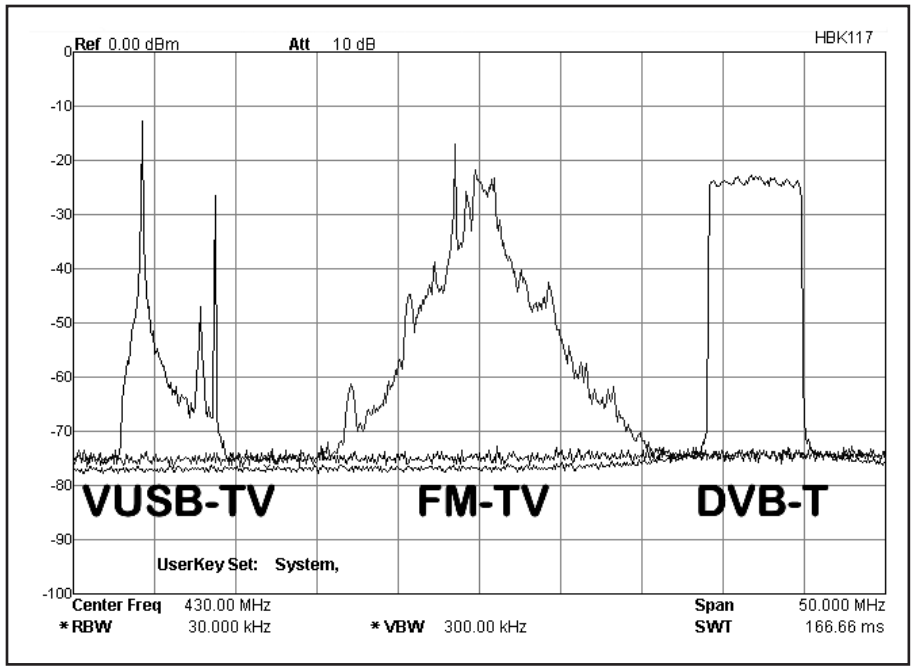

Figure 11.36 — Typical spectra of commonly used TV modulations. Span of 50 MHz shown with vertical scale 10 dB/div and horizontal scale of 5 MHz/div.

FM ATV

On the 33 cm and 23 cm bands and higher frequencies, ATV stations often use FM modulation instead of AM. The main disadvantage is the extremely wide spectrum occupied. **Figure 11.36** shows the spectrum of a typical FM-TV signal with video modulation only. If FM sound sub-carrier(s) are added, the spectrum becomes even broader.

FM ATV in the United States typically uses 4 MHz deviation with NTSC video and a 5.5-MHz sound subcarrier set to 15 dB below the video level. Using Carson's rule, the occupied bandwidth comes out to just under 20 MHz. Because it is so wide, FM-TV is not used on the 70 cm band. Most available FM ATV equipment is made for the 1.2, 2.4 and 10.25-GHz bands.

DIGITAL FAST-SCAN ATV (DATV)

Commercial digital TV (DTV) uses the same 6 MHz-wide channels for analog TV. The 6 MHz-wide over-the-air channels themselves didn't change, just the signals carried in them. Channel 2 is still 54 – 60 MHz, Channel 3 is 60 – 66 MHz, and so forth, up to Channel 51. Channels 2 – 6 have largely been abandoned in the US as broadcast digital TV moved to UHF channels. UHF channels 52 – 69 will be reallocated to other uses.

The US digital broadcast standard is called Advanced Television Standards Committee (ATSC). ATSC uses 8-VSB or 8-level vestigial sideband modulation. This is similar to 256-QAM which means 256-state *quadrature amplitude modulation*. (See the discussion on quadrature modulation earlier in this chapter.) The 256 "states" are combinations of signal phase and amplitude values that represent the 256 different transmitted symbols. This digital format is used in cable TV networks (64-QAM is also used by cable companies). In the case of 8-VSB, the "8" refers to the eight-level baseband DTV signal that amplitude modulates an IF signal. For more information on 8-VSB modulation, see the online article, "What Exactly Is 8-VSB Anyway?" by David Sparano at **www.arrl.org/files/file/Technology/TV_Channels/8_Bit_VSB.pdf** and other references at **www.arrl.org/tv-channels-air-catv**.

The eight levels result in three bits per symbol, resulting in approximately 32 Mb/second raw data rate with a 10.76 Mbaud symbol rate. The net data rate with error correction and other overhead is 19.39 Mb/second.

Digital television signals appear as noise-like signals over their 6 MHz bandwidth. If a digital TV signal interferes with analog radio communications such as FM or SSB, the effect is similar to degraded signal-to-noise ratio. Indeed, a digital TV signal can be thought of as a 6 MHz wide "pile of noise." On a spectrum analyzer, it looks like a "haystack" as shown in Figure 11.36. Of course, the IF bandwidth of the amateur receiver is quite narrow relative to the 6 MHz-wide digital TV channel, so the actual noise level seen by the receiver (assuming no overload or IMD problems) will in part be determined by the receiver's IF bandwidth. Still, the effect on the receiver is what amounts to an elevated noise floor, similar to the effects of wide-spectrum broadband noise.

In Europe and most of the rest of the world, Digital Video Broadcast (DVB) is used. For DVB three basic variations are used depending on what type of signal propagation is available.

DVB-C is used for cable TV where the medium is perfect with strong signal levels and low VSWR, i.e. no reflections. It is not used for DATV as it does not work well when multi-path is present.

DVB-S is for satellite TV where the signals are weak but with little or no multi-path. DVB-S uses MPEG2 audio and video data compression and QPSK modulation with symbol rates up to 20 Mbaud. Most DATV use of DVB-S has been on the 23 cm band since inexpensive set-top demodulation boxes are available for that frequency range.

DVB-T is for terrestrial broadcast TV where the signals could be weak or strong, but where multi-path (what causes "ghost" images on analog TV) is almost always present. Here in the USA, most DATV stations are using DVB-T modulation. In the US, a new modulation standard called ATSC 3.0 is being introduced for broadcast TV. It is an enhanced form of DVB-T and is not backward compatible with ATSC 1.0. DVB-T2 and DVB-S2 are newer versions that will allow 4K high-definition video.

The British Amateur Television Club (BATC) has taken a lead in pushing the bandwidth limits for DTV down far below the commercial standards. They have developed systems in both DVB-S and DVB-T where the bandwidth is less than 1 MHz and still transmit high quality digital video. DATV stations living in major metro areas where the radio spectrum is very crowded are now successfully using narrow-band DVB-T with 2 and 4 MHz bandwidths.

ATV REPEATERS

ATV stations also use repeaters to greatly expand the reach of their signals. Many of the current ATV repeaters are still analog but some are replacing their analog systems or adding digital capability. ATV repeaters are predominately found on the 70 cm and 23 cm bands, with some having capabilities up to the 3 cm (10 GHz) band.

European DATV stations can also use the QO-100 satellite for relaying DATV signals. QO-100 is a stationary satellite covering Europe, Asia, and Africa. The uplink is on the 2.4 GHz band and the downlink on the 10 GHz band. There are some US ATV repeater groups experimenting with DVB-S, ATSC, and ATSC 3.0.

11.7.2 Slow-Scan Television

Despite its name, so-called *slow-scan television* (SSTV) is a method for sending still images, like facsimile. The original mono-

chrome analog SSTV format takes approximately 8 seconds to send one complete frame. The 1500 to 2300 Hz frequency-modulated audio tone resembles that of a fax signal but is sent at a faster rate and includes pulses at 1200 Hz for synchronization.

Color may be sent using any of several methods. The first to be used was the *frame-sequential* method, in which each of the three primary colors (red, green, and blue) is sent sequentially, as a complete frame. A disadvantage of sending complete frames for each color is that you must wait for the third frame to begin before colors start to become correct, and any noise or interference is three times more likely to corrupt the image and risks ruining the image registration (the overlay of the frames), thus spoiling the picture.

In the *line-sequential* method, each line is electronically scanned three times: once each for red, green, and blue. Pictures scan down the screen in full color as they are received, and registration problems are reduced. The Wraase SC-1 modes are examples of early line-sequential color transmission. They have a horizontal sync pulse for each of the color component scans. The major weakness of this method is that if the receiving system gets out of step, it doesn't know which scan represents which color.

Rather than sending color images with the usual RGB (red, green, blue) components, Robot Research decided to use luminance and chrominance signals for their 1200C modes. The first half or two thirds of each scan line contains the luminance information, which is a weighted average of the R, G, and B components. The remainder of each line contains the chrominance signal with the color information. Existing black-and-white equipment could display the B&W-compatible image on the first part of each scan line and the rest would go off the edge of the screen. That compatibility was very beneficial when most people still had only B&W equipment.

The luminance-chrominance encoding makes more efficient use of the transmission time. A 120-line color image can be sent in 12 seconds, rather than the usual 24. Our eyes are more sensitive to details in changes of brightness than color, so the time is used more efficiently by devoting more time to luminance than chrominance. The NTSC and PAL broadcast standards also take advantage of this vision characteristic and use less bandwidth for the color part of the signal. For SSTV, luminance-chrominance encoding offers some benefits, but image quality suffers. It is acceptable for most natural images but looks bad for sharp, high-contrast edges, which are more and more common as images are altered via computer graphics. As a result, all newer modes have returned to RGB encoding.

The 1200C introduced another innovation, called *vertical interval signaling* (VIS). It encodes the transmission mode in the vertical sync interval. By using narrow FSK encoding around the sync frequency, compatibility is maintained. This new signal just looks like an extra-long vertical sync to older equipment.

The Martin and Scottie modes are essentially the same except for the timings. They have a single horizontal sync pulse for each set of RGB scans. Therefore, the receiving end can easily get back in step if synchronization is temporarily lost. Although they have horizontal sync, some implementations ignore them on receive. Instead, they rely on very accurate time bases at the transmitting and receiving stations to keep in step. The advantage of this "synchronous" strategy is that missing or corrupted sync pulses won't disturb the received image. The disadvantage is that even slight timing inaccuracies produce slanted pictures.

In the late 1980s, yet another incompatible mode was introduced. The AVT mode is different from all the rest in that it has *no horizontal sync.* It relies on very accurate oscillators at the sending and receiving stations to maintain synchronization. If the beginning-of-frame sync is missed, it's all over. There is no way to determine where a scan line begins. However, it's much harder to miss the 5-s header than the 300-ms VIS code. Redundant information is encoded 32 times and a more powerful error-detection scheme is used. It's only necessary to receive a small part of the AVT header to achieve synchronization. After this, noise can wipe out parts of the image, but image alignment and colors remain correct.

Digital images may be sent over amateur radio using any of the standard digital modulation formats that support binary file transfer. *Digital SSTV* (DSSTV) is one method of transmitting computer image files, such as JPEG or GIF, as described in an article by Ralph Taggart, WB8DQT, in the Feb 2004 issue of *QST.* This format phase-modulates a total of eight subcarriers (ranging from 590 to 2200 Hz) at intervals of 230 Hz. Each subcarrier has nine possible modulation states. This signal modulation format is known as *redundant digital file transfer* (RDFT) developed by Barry Sanderson, KB9VAK.

Most digital SSTV transmission has switched to using Digital Radio Mondiale (DRM), derived from a system developed for shortwave digital voice broadcasting. The DRM digital SSTV signal occupies the bandwidth between 350 and 2750 Hz. As many as 57 subcarriers may be sent simultaneously, all at the same level. Three pilot carriers are sent at twice the level of the other subcarriers. The subcarriers are modulated using OFDM and QAM, which were described earlier in this chapter. DRM SSTV includes several methods of error correction.

11.8 Spread Spectrum Modulation

A *spread-spectrum* (SS) system is one that intentionally increases the bandwidth of a digital signal beyond that normally required by means of a special spreading code that is independent of the data sequence. There are several reasons for spreading the spectrum in that way.

Spread spectrum was first used in military systems, where the purpose was to encrypt the transmissions to make it harder for the enemy to intercept or jam them. Amateurs are not allowed to encrypt transmissions for the purpose of concealing the information, but reducing interference, intentional or otherwise, is an obvious benefit. The signal is normally spread in such a fashion that it appears like random noise to a receiver not designed to receive it, so other users of the band may not even be aware that an SS signal is present.

Another advantage to spreading the spectrum is that it can make frequency accuracy less critical. In addition, the wide bandwidth means that expensive narrow-bandwidth filters are not required in the receiver. It also provides a measure of frequency diversity. If certain frequencies are unusable because of interference or selective fading, the signal can often be reconstructed using information in the rest of the bandwidth.

There are several ways to spread the spectrum — we will cover the two most common methods below — but they all share certain characteristics. Imagine that the unspread signal occupies a 10 kHz bandwidth and it is spread by a factor of 100. The resulting SS signal is 1000 kHz (1 MHz) wide. Each 10-kHz channel contains 1/100 of the total signal power, or –20 dB. That means that any narrowband stations using one of those 10-kHz channels experience a 20 dB reduction in interference, but also are more likely to be interfered with because of the 100-times greater bandwidth of the SS station's emissions.

How is the spread-spectrum station affected by interference from narrowband stations? In effect, the SS receiver attenuates the signal received on each 10 kHz channel by 20 dB in order to obtain a full-power signal when all 100 channels are added together. That means that the interference from a narrowband station is reduced by 20 dB but, again, the interference is more likely to occur because of the 100-times greater bandwidth of the SS station's receiver.

How is the spread-spectrum station affected by interference from another SS station on the same frequency? It turns out that if the other station is using a different orthogonal spreading code then, once again, the interference reduction is 20 dB for 100-times spreading. That means that many SS stations can share the same channel without interference as long as they are all received at roughly the same signal level. Commercial mobile-telephone SS networks use an elaborate system of power control with real-time feedback to ensure that the signals from all the mobile stations arrive at the base station at approximately the same level.

That scheme works well in a one-to-many (base station to mobile stations) system architecture but would be much more difficult to implement in a typical amateur many-to-many arrangement because of the different distances and thus path losses between each pair of stations in the network. On the HF bands it is not uncommon to see differences in signal levels of 80 to 90 dB or more. (For example, the difference between S1 and 40 dB over S9 is 88 dB, assuming 6 dB per S-unit.) A spread spectrum signal at S9 + 40 dB with a spreading ratio of 100 times would interfere with any other signals below about S9 + 20 dB. It works the same in the other direction as well. The SS signal would experience interference from any other stations that are more than 20 dB louder than the desired signal.

Normally, increasing the bandwidth of a transmission degrades the signal-to-noise (S/N) ratio at the receiver. A 100-times greater bandwidth contains 100 times as much noise, which causes a 20 dB reduction in S/N ratio. However SS receivers benefit from a phenomenon known as processing gain. Just as the receiver is insensitive to other SS signals with different orthogonal spreading codes, so it is insensitive to random noise. The improvement in S/N ratio due to processing gain is:

$$\text{Processing gain} = 10 \times \log\left(\frac{\text{spread bandwidth}}{\text{unspread bandwidth}}\right) \text{dB}$$

That is exactly equal to the reduction in S/N ratio due to the increased bandwidth. The net result is that an SS signal has neither an advantage nor a disadvantage in signal-to-noise ratio compared to the unspread version of the same signal. When someone states that, because of processing gain, an SS receiver can receive signals that are below the noise level (signals that have a negative S/N ratio), that is a true statement. However, it does not imply better S/N performance than could be obtained if the signal were not spread.

11.8.1 Frequency Hopping Spread Spectrum

One simple way to spread the spectrum of a narrowband signal is to repetitively sweep it across the frequency range of the wider spectrum, either continuously or in a series of steps at discrete frequencies. That technique, called *chirp modulation*, can be considered a special case of *frequency-hopping spread spectrum* (FHSS), in which the narrowband signal covers the expanded spectrum by rapidly hopping back and forth from frequency to frequency in a pseudo-random manner. On average, each frequency is used the same percentage of time so that the average spectrum is flat across the bandwidth of the FHSS signal. See **Figure 11.37**.

If the receiver hops in step with the transmitter, using the same pseudo-random sequence synchronized to the one in the transmitter, then the transmitter and receiver are always tuned to the same frequency and the receiver's detector sees a continuous, non-hopped narrowband signal which can be demodulated in the normal way. We say that the signal has been *de-spread*, that is, returned to its normal narrowband form. Synchronization between the receiver and transmitter is one of the challenges in an FHSS system. If the timing of the two sequences differs by even one hop, then the receiver is always tuned to the wrong frequency, unless the same frequency happens to occur twice in succession in the pseudo-random sequence. Any signal with an unsynchronized sequence, or with a different sequence, is reduced in amplitude by the processing gain.

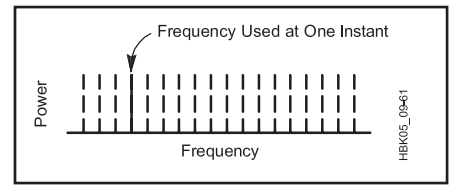

Figure 11.37 — Power versus frequency for a frequency-hopping spread spectrum signal. Emissions jump between discrete frequencies in pseudo-random fashion. Normally the spacing of the frequencies is approximately equal to the bandwidth of the unspread signal so that the average spectrum is approximately flat.

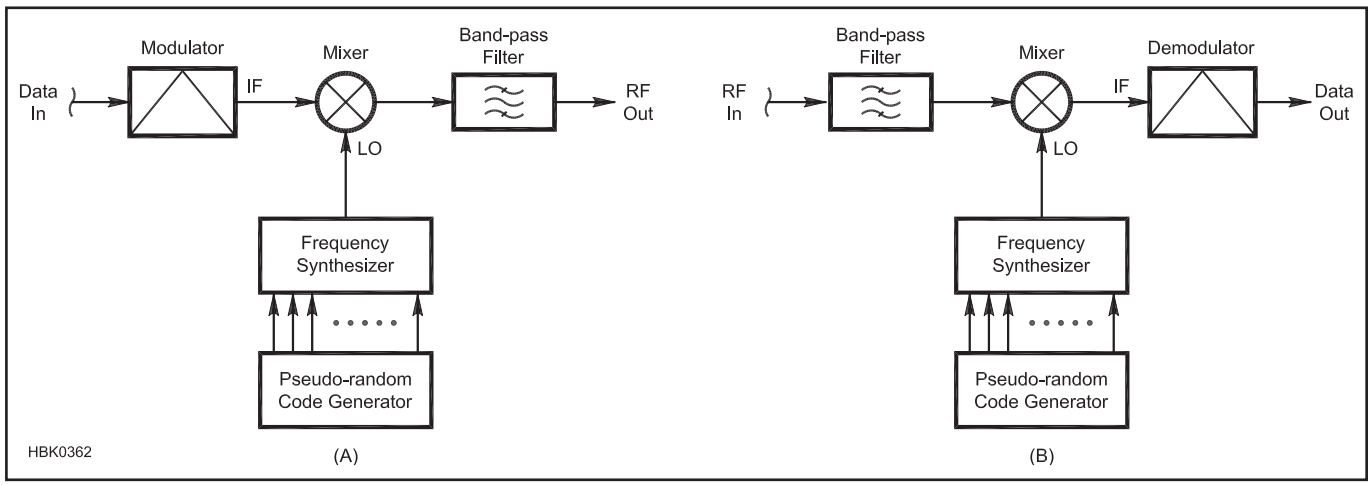

Figure 11.38 — Block diagram of an FHSS transmitter (A) and receiver (B). The receiver may be thought of as a conventional superhet with a local oscillator (LO) that is continually hopping its frequency in response to a pseudo-random code generator. The transmitter has a similar architecture to up-convert a conventionally-modulated intermediate frequency (IF) to a frequency-hopped radio frequency (RF).

There are two types of FHSS based on the rate at which the frequency hops take place. *Slow-frequency hopping* refers to a hop rate slower than the baud rate. Several symbols are sent per hop. With *fast-frequency hopping*, the hop rate is faster than the baud rate. Several hops occur during each symbol. The term *chip* refers to the shortest-duration modulation state in the system. For slow-frequency hopping, that is the baud rate. For fast-frequency hopping it is the hop rate. Fast-frequency hopping can be useful in reducing the effects of multi-path propagation. If the hop period is less than the typical time delay of secondary propagation paths, then those signals are uncorrelated to the main path and are attenuated by the processing gain.

A diagram of a frequency-hopping spread spectrum system is shown in **Figure 11.38**. In both the transmitter and the receiver, a pseudo-random code generator controls a frequency synthesizer to hop between frequencies in the correct order. In this way, the narrowband signal is first spread by the transmitter, then sent over the radio channel, and finally de-spread at the receiver to obtain the original narrowband signal again. One issue with FHSS is that many synthesizers do not maintain phase coherence over successive frequency hops. That means the basic (non-spread) modulation must be a type that does not depend on phase information. That rules out PSK, QPSK and QAM. That is the reason that modulation types that do not depend on phase, such as FSK and MFSK with non-coherent detection, are frequently used as the base modulation type in FHSS systems.

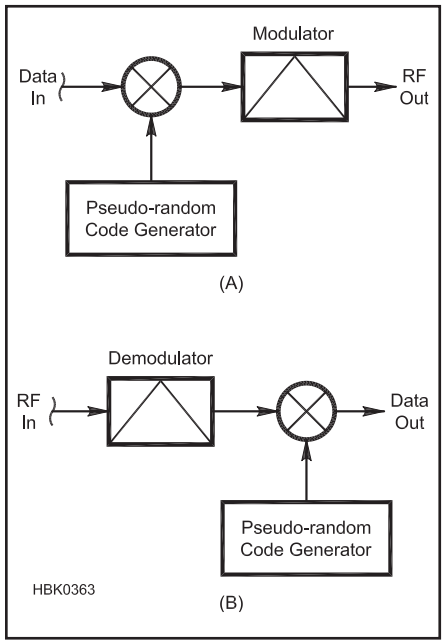

Figure 11.39 — Block diagram of a DSSS transmitter (A) and receiver (B). For BPSK modulation, the modulation has the same format as the bipolar data (±1), so the modulator and demodulator could be moved to the other side of the multiplier if desired.

11.8.2 Direct Sequence Spread Spectrum

Whereas an FHSS system hops through a pseudo-random sequence of frequencies to spread the signal, a *direct-sequence spread spectrum* (DSSS) system applies the pseudo-random sequence directly to the data in order to spread the signal. See **Figure 11.39**. The binary data is considered to be in *polar* form, that is, the two possible states of each bit are represented by –1 and +1. The data bits are multiplied by a higher-bit-rate pseudo-random sequence, also in polar form, which chops the data into smaller time increments called *chips*. The ratio of the chip rate to the data's bit rate equals the ratio of the spread bandwidth to the unspread bandwidth, which is just the processing gain:

$$\text{Processing gain} = 10 \times \log\left(\frac{\text{chip rate}}{\text{bit rate}}\right) \text{dB}$$

Although it doesn't have to, DSSS normally uses a one-bit-per-symbol modulation type such as BPSK. In that case, the modulator and demodulator in Figure 11.39 would consist simply of a mixer, which multiplies the RF local oscillator by the bipolar DSSS modulating signal. Since unfiltered BPSK is constant-envelope, a nonlinear class-C power amplifier may be used for high efficiency. The demodulator in the receiver would also be a mixer, which multiplies the RF signal by a local oscillator to regenerate the DSSS modulating signal. Not shown are additional mixers and filters that would be used in a superheterodyne receiver to convert the received signal to an intermediate frequency before demodulation and decoding.

The unfiltered spectrum has the form of a sinc function. That shows up clearly in the spectrum of an actual DSSS signal in **Figure 11.40A**, which is plotted on a logarithmic scale calibrated in decibels. The humps in the response to the left and right (and additional ones not shown off scale) are not needed for communications and should be filtered out to avoid excessive occupied bandwidth. Figure 11.40B shows the DSSS signal in the presence of noise at the input to the receiver, and Figure 11.40C illustrates the improvement in signal-to-noise ratio of the de-spread narrowband signal.

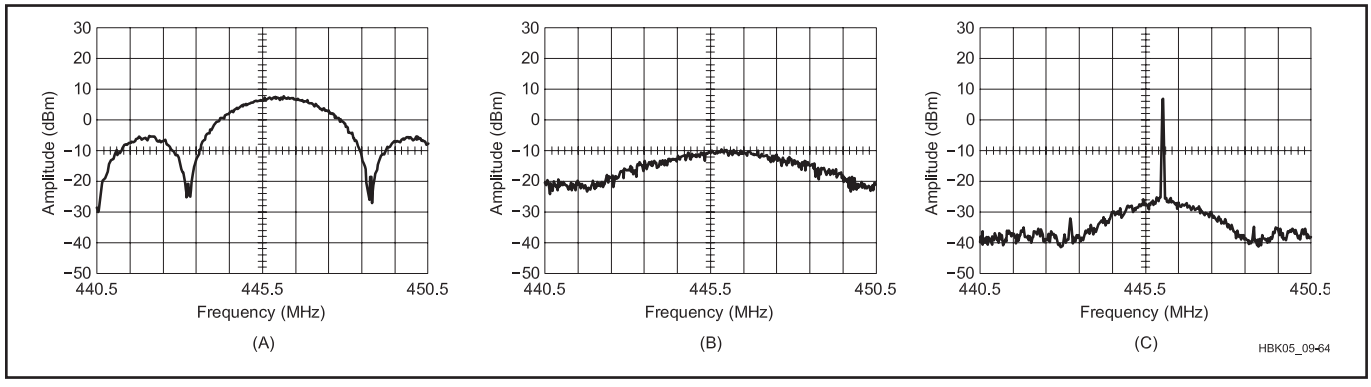

Figure 11.40 — (A) The frequency spectrum of an actual unfiltered biphase-modulated spread spectrum signal as viewed on a spectrum analyzer. In this practical system, band-pass filtering is used to confine the spread spectrum signal to the amateur band. (B) At the receiver end of the line, the filtered spread spectrum signal is apparent only as a 10-dB hump in the noise floor. (C) The signal at the output of the receiver de-spreader. The original carrier — and any modulation components that accompany it — has been recovered. The peak carrier is about 45 dB above the noise floor — more than 30 dB above the hump shown at B. (These spectrograms were made at a sweep rate of 0.1 s/division and an analyzer bandwidth of 30 kHz; the horizontal scale is 1 MHz/division.)

11.8.3 Code-Division Multiple-Access (CDMA)

As mentioned before, to receive an SS signal, the de-spreading sequence in the receiver must match the spreading sequence in the transmitter. The term *orthogonal* refers to two sequences that are coded in such a way that they are completely uncorrelated. The receiver's response to an orthogonal code is the same as to random noise, that is, it is suppressed by a factor equal to the processing gain. One can take advantage of this property to allow multiple SS stations to access the same frequency simultaneously, a technique known as *code-division multiple access* (CDMA). Each transmitter is assigned a different orthogonal code. A receiver can "tune in" any transmitter's signal by selecting the correct code for de-spreading.

If multiple stations want to be able to transmit simultaneously without using spread spectrum, they must resort to either *frequency-division multiple access* (FDMA), where each station transmits on a different frequency channel, or *time-division multiple access* (TDMA), where each transmission is broken up into short time slots which are interleaved with the time slots of the other stations. Compared to TDMA, CDMA has the advantage that it does not require an external synchronization network to make sure that different stations' time slots do not overlap. Compared to both TDMA and FDMA, CDMA has the further advantage that it experiences a gradual degradation in performance as the number of stations on the channel increases. It is relatively easy to add new users to the system. Also CDMA has inherent resistance to interference due to multi-path propagation or intentional jamming. The primary disadvantages of CDMA are the relative complexity and the necessity for accurate power-level control to make sure that unwanted signals do not exceed the level that can be rejected through processing gain.

11.9 Pulse Modulation

Another type of digital amplitude modulation comes under the general category of *pulse modulation*. The RF signal is broken up into a series of pulses, which are usually equally-spaced in time and separated by periods of no signal. We will discuss three types of pulse modulation, PAM, PWM and PPM.

Pulse-amplitude modulation (PAM) consists of a series of pulses of varying amplitude that correspond directly to the amplitude of the modulating signal. See **Figure 11.41B**. The pulses can be positive or negative, depending on the polarity of the signal. The modulating signal can be recovered simply by low-pass filtering the pulse train. The effect of the negative pulses is to reverse the polarity of the RF signal. The result is very similar to a double-sideband, suppressed-carrier signal that is rapidly turned on and off at the pulse rate of the PAM. In other words, the DSBSC signal is periodically sampled at a certain pulse repetition rate.

For the signal to be properly represented, its highest modulating frequency must be less than the *Nyquist frequency*, which is one-half the sample rate. That condition, known as the *Nyquist criterion*, applies not only to PAM but to any digital modulation technique.

There are two variations of PAM that should be mentioned. *Natural sampling* does not hold the amplitude of each pulse constant throughout the pulse as shown in Figure 11.41B, but rather follows the shape of the analog modulation. The *single-polarity method* adds a fixed offset, or pedestal, to the modulating signal before the PAM modulator. As long as the pedestal is greater than or equal to the peak negative modulation, then the RF phase never changes and the signal is equivalent to sampled full-carrier AM rather than DSB-SC.

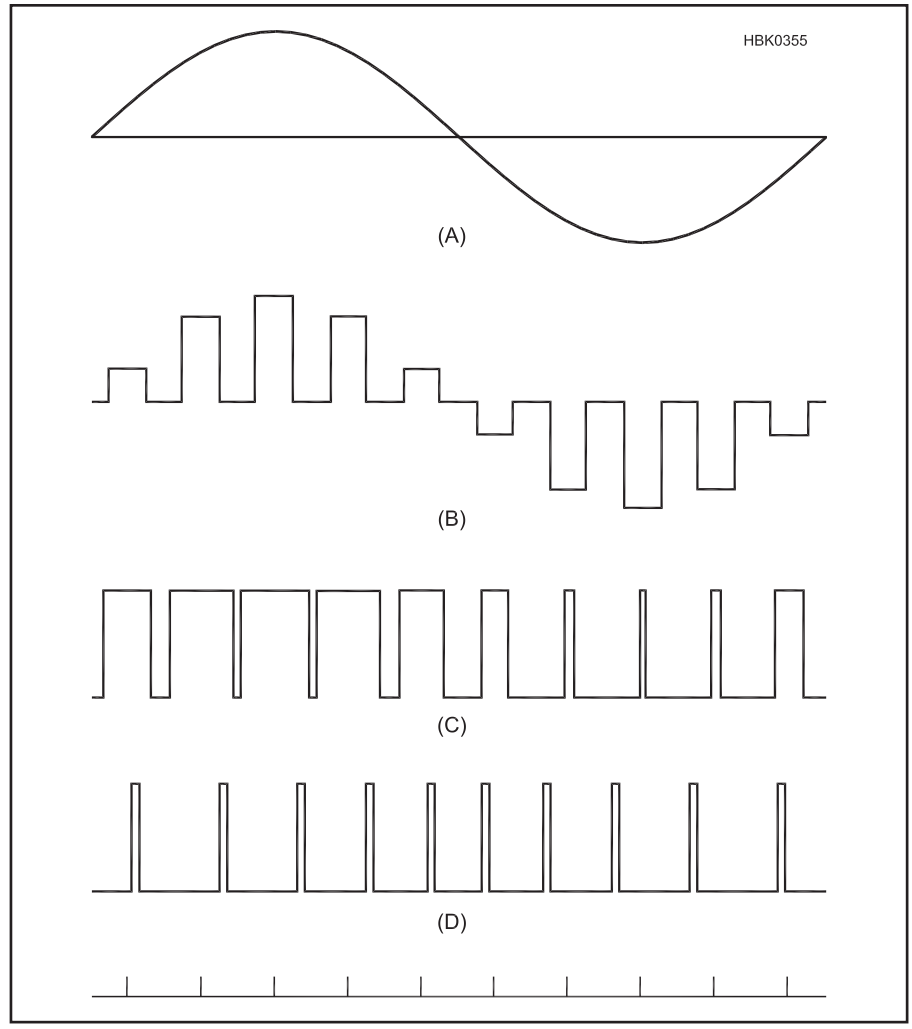

Figure 11.41 — Three types of pulse-code modulation. A sine-wave modulating signal (A) is shown at the top. Pulse-amplitude modulation (B) varies the amplitude of the pulses, pulse-width modulation (C) varies the pulse width, and pulse-position modulation (D) varies the pulse position, proportional to the modulating signal. The tic marks show the nominal pulse times. The RF signal is created by an AM modulator using (B), (C) or (D) as the modulating signal.

PAM is rarely used on the amateur bands because it increases the transmitted bandwidth and adds circuit complexity with no improvement in signal-to-noise ratio for most types of noise and interference. The concept is useful, however, because PAM is similar to the signal generated by the sample-and-hold circuit that is used at the input to an analog-to-digital converter (ADC). An ADC is used in virtually every digital transmitter to convert the analog voice signal to a digital signal suitable for digital signal processing.

Pulse-width modulation (PWM) is a series of pulses whose width varies in proportion to the amplitude of the modulating signal. Figure 11.41C shows the pulses centered on the sample times, but in some systems the sample times may correspond to the leading or trailing pulse edges. With either method, the modulating signal can be recovered by low-pass filtering the pulse train and passing it through a coupling capacitor to remove the dc component.

Pulse-position modulation (PPM), Figure 11.41D, varies the position, or phase, of the pulses in proportion to the amplitude of the modulating signal. With both PWM and PPM, the peak amplitude of the signal is constant. That allows the receiver to be designed to be insensitive to amplitude variations, which can result in a better post-detection signal-to-noise ratio, in a manner similar to analog angle modulation.

11.10 Modulation Bandwidth and Impairments

Most of the previous discussion of the various modulation types has assumed the modulation is perfect. With analog modulation, that means the audio or video modulating signal is perfectly reproduced in the RF waveform without distortion, spurious frequencies or other unwanted artifacts. With digital modulation, the symbol timing and the locations and trajectories in the I/Q constellation are perfectly accurate. In all cases, the RF power amplifier is perfectly linear, if so required by the modulation type, and it introduces no noise or other spurious signals close to the carrier frequency.

In the real world, of course, such perfection can never be achieved. Some modulation impairments are caused by the transmitting system, some by the transmission medium through which the signal propagates, and some in the receiving system. This section will concentrate on impairments caused by the circuitry in the transmitter and, to some extent, in the receiver. Signal impairments due to propagation are covered in detail in the **Propagation of Radio Signals** chapter.

11.10.1 Filtering and Bandwidth of Digital Signals

We have already touched on this topic in previous sections, but let us now cover it a little more systematically. The bandwidth required by a digital signal depends on the filtering of the modulation, the symbol rate and the type of modulation. For linear modulation types such as OOK, BPSK, QPSK and QAM, the bandwidth depends only on the symbol rate and the modulation filter.

As an example, an unfiltered BPSK modulating signal with alternating ones and zeroes for data (10101010...) is a square wave at one-half the symbol rate. Like any square wave, its spectrum can be broken down into a series of sine waves at the fundamental frequency (symbol rate / 2) and all the odd harmonics. If the data consists of alternating pairs of ones and pairs of zeroes (11001100...) then we have a square wave at one-fourth the symbol rate and the spectrum is a series of sine waves at one-fourth the symbol rate and all its odd harmonics. Random data contains energy at all frequencies from zero to half the symbol rate and all the odd harmonics of those frequencies. The harmonics are not needed for proper demodulation of the signal, so they can be filtered out with a low-pass filter with a cutoff frequency of one-half the symbol rate.

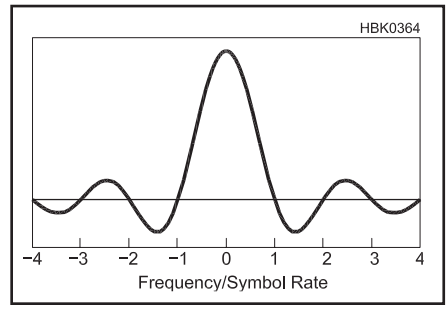

Figure 11.42 — The sinc function, which is the spectrum of an unfiltered BPSK modulating signal with random data, plotted with a linear vertical scale. The center (zero) point corresponds to the RF carrier frequency, To see what the double-sided RF spectrum looks like on a logarithmic (dB) scale, see Figure 11.40A.

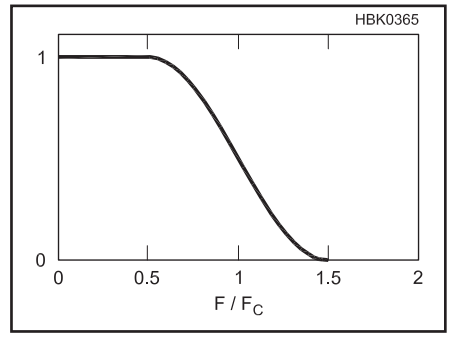

Figure 11.43 — Amplitude versus frequency for a 0.5-alpha, raised-cosine filter. The vertical scale is linear, not logarithmic as would be seen on a spectrum analyzer. The amplitude is 0.5 (–6 dB) at the cutoff frequency, Fc. The amplitude is 1.0 for frequencies less than Fc × (1 – alpha) and is 0.0 for frequencies above Fc × (1 + alpha).

With random data, the shape of the unfiltered spectrum is a sinc function,

$$\mathrm{sinc}(f / f_s) = \frac{\sin(\pi f / f_s)}{\pi f / f_s}$$

where f_s is the symbol rate. See **Figure 11.42**. Note that the response is zero (minus infinity dB) whenever f is an integer multiple of the symbol rate. That is why multi-carrier modulation generally uses a carrier spacing equal to the symbol rate.

The previous discussion applies to the baseband signal, before it modulates the RF carrier. BPSK is a double-sideband-type modulation so the baseband spectrum appears above and below the RF carrier frequency, doubling the bandwidth to 2 × (symbol rate / 2) = symbol rate. In reality, since no practical filter has an infinitely-sharp cutoff, the occupied bandwidth of a BPSK signal must be somewhat greater than the symbol rate. That is also true for all other double-sideband linear modulation types, such as OOK and the various forms of QPSK and QAM.

If the low-pass filtering is not done properly, it may slow down the transition between symbols to the point where one symbol starts to run into another, causing *inter-symbol interference* (ISI). A type of filter that avoids that problem is called a *Nyquist filter*. It ensures that each symbol's contribution to the modulating signal passes through zero at the center of all other symbols, so that no ISI occurs. The most common type of Nyquist filter is the *raised-cosine* filter, so-called because the frequency response (plotted on a linear scale) in the passband-to-stopband transition region follows a raised-cosine curve. See **Figure 11.43**. The sharpness of the frequency cutoff is specified by a parameter called *alpha*. If alpha is 1.0, then the transition from passband to stopband is very gradual — it starts to roll off right at zero hertz and finally reaches zero response

at two times the nominal cutoff frequency. An alpha of 0.0 specifies an ideal "brick-wall" filter that transitions instantaneously from full response to zero right at the cutoff frequency. Values in the range of 0.3 to 0.5 are common in communications systems.

Unfortunately, if any additional filter is placed before or after the Nyquist filter it destroys the anti-ISI property. In order to allow filtering in both the transmitter and the receiver many systems effectively place half the Nyquist filter in each place. Because the frequency response of each filter is the square root of the response of a Nyquist filter, they are called *root-Nyquist* filters. The *root-raised-cosine* filter is an example. While a Nyquist or root-Nyquist response theoretically could be approximated with an analog filter, they are almost always implemented as digital filters. More information on digital filters appears in the **DSP and SDR Fundamentals** and the **Analog and Digital Filtering** chapters.

As mentioned before, filtering is more difficult with angle modulation because it is nonlinear. The RF spectrum is not a linear transposition of the baseband spectrum as it is with linear modes and Nyquist filtering doesn't work. Old-fashioned RTTY transmitters traditionally just used an R-C low-pass filter to slow down the transitions between mark and space. While that does not limit the bandwidth to the minimum value possible, the baud rate is low enough that the resulting bandwidth is acceptable anyway. In more modern systems, there is a tradeoff between making the filter bandwidth as narrow as possible for interference reduction and widening the bandwidth to reduce inter-symbol interference. For example, the GSM (Global System for Mobile communications) standard, used for some cellular telephone systems, uses minimum-shift keying and a Gaussian filter with a BT (bandwidth symbol-time product) of 0.3. A 0.3 Gaussian filter has a 0.3 ratio of 3-dB bandwidth to baud rate, which results in a small but acceptable amount of ISI and a moderate amount of adjacent-channel interference.

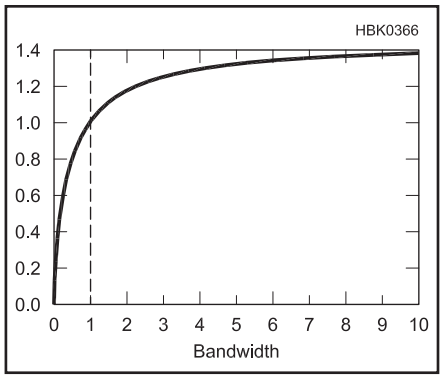

Figure 11.44 — Plot of channel capacity versus bandwidth, calculated by the Shannon-Hartley theorem. The S/N ratio has been selected to be unity at Bandwidth = 1.

CHANNEL CAPACITY

It is possible to increase the quantity of error-free data that can be transmitted over a communications channel by using an error-correcting code. That involves adding additional error-correction bits to the transmitted data. The more bits that are added, the greater the errors that can be corrected. However, the extra bits increase the data rate, which requires additional bandwidth, which increases the amount of noise. For that reason, as you add more and more error-correction bits, requiring more and more bandwidth, you eventually reach a point of limited additional return. In the 1940s, Claude Shannon worked out a formula (called the *Shannon-Hartley theorem*) for the maximum capacity possible over a communications channel, assuming a theoretically-perfect error-correction code:

$$C = B \log_2 \left(1 + \frac{S}{N}\right) \text{bit/s}$$

where

C = the net channel capacity, not including error-correcting bits,
B = the bandwidth in Hz, and
S/N = the signal-to-noise ratio, expressed as a power ratio.

Note that as B increases, N increases in the same proportion. **Figure 11.44** is a plot of channel capacity versus bandwidth based on the formula. As bandwidth is increased, channel capacity increases rapidly until the point where the S/N ratio drops to unity (labeled Bandwidth = 1 in the graph) after which channel capacity increases much more slowly.

11.10.2 Intermodulation Distortion

To minimize distortion of the modulation, each stage in the signal chain must be linear, from the microphone or modem, through all the intermediate amplifiers and processors, to the modulator itself. For the linear modulation types, all the amplifiers and other stages between the modulator and the antenna must be linear as well.

If the modulation consists of a single sine wave, then nonlinearity causes only harmonic distortion, which produces new frequencies at integer multiples of the sine-wave frequency. If multiple frequencies are present in the modulation, however, then *intermodulation distortion* (IMD) products are produced. IMD occurs when a nonlinear amplifier or other device acts as a mixer, producing sum and difference frequencies of all the pairs of frequencies and their harmonics. For example if two frequencies, F1 and F2 > F1, are present, then IMD will cause spurious frequencies to appear at F1 + F2, F2 – F1, 2F1, 2F2, 2F1 – F2, 2F2 – F1, 2F2 – 2F1, 3F1, 3F2, and so on. *Odd-order* products are those that include the original frequencies an odd number of times, such as 3F1, 2F1 + F2, 3F1 – 2F2, and so on. *Even-order* products contain an even number of the original frequencies, such as 2F2, F1 + F2, 3F1 + F2, and so on. If more than two frequencies are present in the undistorted modulation, then the number of unwanted frequencies increases exponentially.

Although intermodulation distortion that occurs before modulation is fundamentally the same as IMD that occurs after the modulator (at the intermediate or radio frequency), the effects are quite different. Consider two frequency components of a modulating signal at, for example, 1000 Hz and 1200 Hz that modulate an SSB transmitter tuned to 14.000 MHz, USB. The desired RF signal has components at 14.001 and 14.0012 MHz. If the IMD occurs before modulation, then the F1 + F2 distortion product occurs at 1000 + 1200 = 2200 Hz. Since that is well within the audio passband of the SSB transmitter, there is no way to filter it out so it shows up at 14.0022 MHz at the RF output. However, if the distortion had occurred after the modulator, the F1 + F2 product would be at 14.001 + 14.0012 = 28.0022 MHz which is easily filtered out by the transmitter's low-pass filter.

That explains why RF speech processors work better than audio processors. A speech processor clips or limits the peak amplitude of the modulating signal to prevent over-modulation in the transmitter. However, the limiting process typically produces considerable intermodulation distortion. If the limiter is located after the modulator rather than before it, then it is an RF signal being clipped, rather than audio. If the RF limiter is followed by a band-pass filter then many of the distortion products are removed, resulting in a less-distorted signal.

If the distortion is perfectly symmetrical (for example equal clipping of positive and negative peaks) then only odd-order products are produced. Most distortion is not symmetrical so that both even and odd-order products appear. However, the odd-order products are of particular interest when measuring the linearity of an RF power amplifier. The reason is that even-order products that occur after the modulator occur only near harmonics of the RF frequency where they are easy to filter out. Odd-order products can fall within the desired channel, where they cause distortion of the modulation, or at nearby frequencies, where they cause interference to other stations.

The informal term for such IMD that interferes with other stations outside the

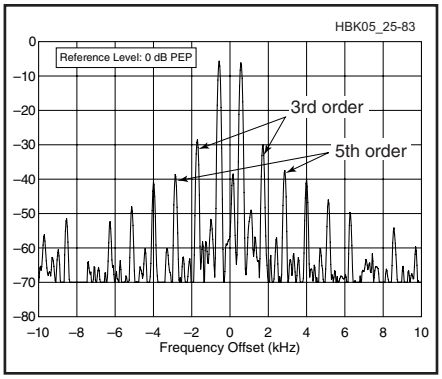

Figure 11.45 — Intermodulation (IMD) products from an SSB transmitter modulated by a pair of audio tones (see text). The two signals at the center are the desired modulation sidebands. The frequencies of the third-order IMD products are separated from the frequencies of the desired sidebands by the tone spacing (1 kHz) and from each other by three times the tone spacing (3 kHz). The fifth-order products are separated from the third-order products by twice the tone spacing (2 kHz) and from each other by five times the tone spacing (5 kHz).

desired channel is *splatter*. It becomes severe if the linear amplifier is over-driven, which causes clipping of the modulation envelope with the resulting odd-order IMD products.

The method commonly used to test the linearity of an SSB transmitter or RF power amplifier is the *two-tone test*. Two equal amplitude audio-frequency tones are fed into the microphone input, the transmitter and/or amplifier is adjusted for the desired power level, and the output signal is observed on a spectrum analyzer. See **Figure 11.45**. If the modulating tones are F1 and F2, the third-order IMD products closest to the modulation sidebands have frequencies of 2F1 – F2 and 2F2 – F1. The difference in frequency between the third-order IMD products is 3(F1 – F2). Similarly, the fifth-order IMD products occur at 3F1 – 2F2 and 3F2 – 2F1, so they are spaced at 5(F1 – F2). As in the figure, if F1 and F2 are 1 kHz apart, the two third-order products are 1 kHz above the upper modulation sideband and 1 kHz below the lower sideband. That is a spacing of 3 × 1 kHz = 3 kHz. The fifth-order products are twice the tone spacing (2 kHz) from the modulation sidebands for a product-to-product spacing of 2 + 1 + 2 kHz = 5 kHz.

11.10.3 Transmitted Bandwidth

We have already discussed the necessary bandwidth for each of the various modulation types. The previous section explained how intermodulation distortion is one phenomenon that can cause unwanted emissions outside of the desired communications channel. Another is failing to properly low-pass filter the modulating signal to the minimum necessary bandwidth. That is especially a concern for linear modulation modes such as SSB, AM, OOK (CW), BPSK and QAM. For angle-modulated modes like FM and FSK, excessive bandwidth can result from simply setting the deviation too high.

There are other modulation impairments that cause emissions outside the desired bandwidth. For example, in an SSB transmitter, if the unwanted sideband is not sufficiently suppressed, the occupied bandwidth is up to twice as large as it should be. Also, an excessively strong suppressed carrier causes particularly-annoying heterodyne interference to stations tuned near that frequency. In some SSB modulators, there is an adjustment provided to optimize carrier suppression.

The carrier suppression may be degraded by a modulating signal that is too low in amplitude. For example, if the signal from the microphone to an SSB transmitter is one-tenth (–20 dB) of the proper amplitude, and if the gain of the RF amplifier stages is increased to compensate, then the carrier suppression is degraded by 20 dB.

The term *adjacent-channel power* (ACP) refers to the amount of transmitted power that falls into an adjacent communications channel above or below the desired channel. Normally the unwanted out-of channel power is worse for the immediately-adjacent channels than for those that are two channels away, the so-called *alternate* channels. ACP is normally specified as a power ratio in dB. It is measured with a spectrum analyzer that can measure the total power within the desired channel and the total power in an adjacent channel, so that the dB difference can be calculated.

The *occupied bandwidth* is the bandwidth within which a specified percentage of the total power occurs. A common percentage used is 99%. For a properly-adjusted, low-distortion transmitting system, the occupied bandwidth is determined mainly by the modulation type and filtering and, in the case of digital modulation, the symbol rate. For example, the IS-54 TDMA format that has been used in some US digital cellular networks has about 30 kHz occupied bandwidth using 24.3-kilosymbols/sec, π/4-DQPSK modulation with a 0.35-alpha root-raised-cosine filter. The GSM cellular standard requires about a 350-kHz occupied bandwidth for its 270.833-kilosymbols/sec, 0.3 Gaussian-filtered MSK signal.

11.10.4 Modulation Accuracy

For analog modes, modulation accuracy is mainly a question of maintaining the proper frequency response across the desired bandwidth with minimal distortion and unwanted signal artifacts. In-band artifacts like noise and spurious signals should not be a problem with any reasonably-well-designed system. Maintaining modulation peaks near 100% for AM signals or the proper deviation for FM signals is facilitated by an audio compressor. It can be either the type that uses a detector and an automatic-gain-control feedback loop to vary the gain in the modulation path or a clipper-type compressor that limits the peak amplitude and then filters the clipped signal to remove the harmonics and intermodulation products that result. SSB transmitters can also use audio speech compression to maintain the proper peak power level although, as explained previously, clipping of the signal before it reaches the modulator can cause unacceptable distortion unless special techniques are used.

For digital signals, there are a number of other possible sources of modulation inaccuracy. For modes that use Nyquist filtering, the cutoff frequency and filter shape must be accurate to ensure no inter-symbol interference (ISI). Fortunately, that is easy to do with digital filters. However any additional filtering in the signal path can degrade the ISI. For example, most HF digital modes use an analog SSB transceiver to up-convert the signal from audio to RF for transmission and to down-convert from RF to audio again at the receiver end. The crystal filters used in the transmitter and receiver can significantly degrade group delay flatness, especially near the edges of the filter passband. That is why most HF digital modes use a bandwidth substantially less than a typical transceiver's passband and attempt to center the signal near the crystal filters' center frequency.

Distortion can also impair the proper decoding of digital signals, especially formats with closely-spaced symbols such as 256-QAM. Any "flat-topping" in the final amplifier causes the symbols at the outermost corners of the constellation to be closer together than they should be. The accuracy of the symbol clock is critical in formats like JT65 that integrate the detection process over a large number of symbols. Perhaps surprisingly, the clock rate on some inexpensive computer sound cards can be off on the order of a percent, which results in a similar error in symbol rate.

The modulation accuracy of an FSK signal is normally characterized by the frequency shift at the center of each symbol time, the point at which the receiver usually makes the decision of which symbol is being received. For other digital formats, the modulation accuracy is normally characterized by the amplitude and phase at the symbol decision points. Amplitude error is typically measured as a percentage of the largest symbol amplitude. The phase

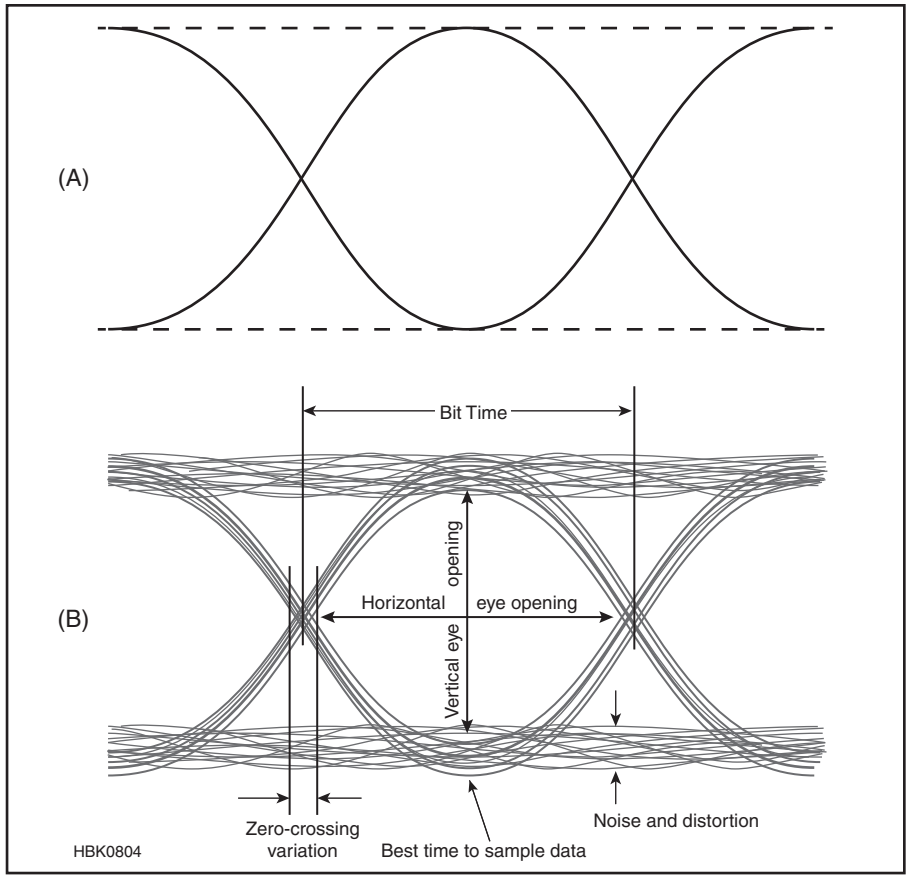

Figure 11.47 — The eye diagram is an oscilloscope-style diagram showing repeated samples of the digital signal as it transitions between points in the constellation. The eye diagram at A is for a perfect, undistorted signal. The eye diagram in B shows a signal with noise, distortion, and timing variations.

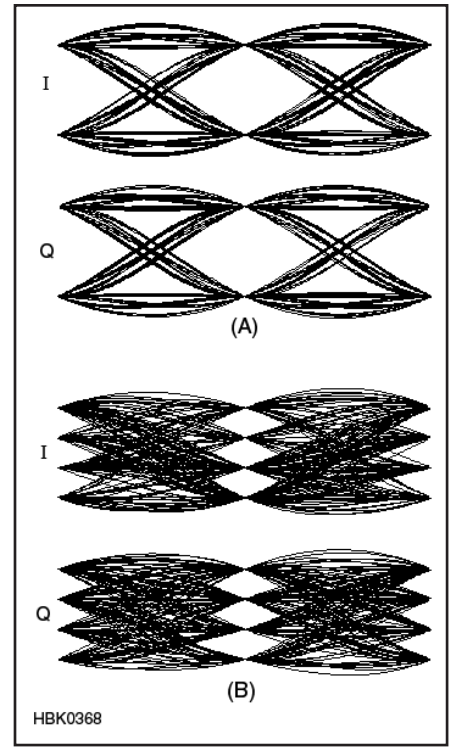

Figure 11.48 — Eye diagrams of the I and Q outputs of a QPSK generator (A) and a 16-QAM generator (B). The "eye" is the empty area between the symbol decision points, visible as the points where all the symbol trajectories come together. The bigger the "eye" the easier the signal is to decode.

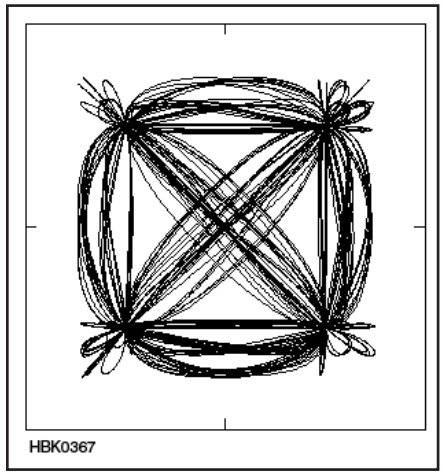

Figure 11.46 — Simulated I/Q constellation display of the trajectory of a QPSK signal over an extended period. A 0.5-alpha raised-cosine filter was used. Because it is a Nyquist filter, the trajectories pass exactly through each symbol location.

error is quoted in degrees. In both cases one can specify either the average RMS value or the peak error that was detected over the measurement period. Amplitude error is the most important consideration for modulation types such as BPSK where the information is encoded as an amplitude. For a constant-envelope format like MSK, phase error is a more important metric.

For formats like QPSK and QAM, both amplitude and phase are important in determining symbol location. A measurement that includes both is called *error vector magnitude* (EVM). It is the RMS average distance between the ideal symbol location in the constellation diagram and the actual signal value at the symbol decision point, expressed as a percentage of either the RMS signal power or the maximum symbol amplitude. This is a measurement that requires specialized test equipment not generally available to home experimenters. However it is possible to estimate the effect on EVM of various design choices using computer simulations.

Previously, we have plotted constellation diagrams with the transitions between symbol locations indicated by straight lines. In an actual system, both the I and Q signals are low-pass filtered which makes the symbol transitions smoother, with no abrupt changes of direction at the symbol locations. **Figure 11.46** shows the actual symbol trajectories for a QPSK signal using a 0.5-alpha raised-cosine filter with random data. Since a raised-cosine is a type of Nyquist filter, the trajectories pass exactly through each symbol location. This is what you would see if you connected the I and Q outputs from a QPSK baseband generator to the X (horizontal) and Y (vertical) inputs of an oscilloscope. It is also what you would see in a receiver at the input to the symbol detector after the carrier and clock-recovery circuits have stabilized the signal. If the symbol trajectories were not accurate, then the dark areas at the symbol points would be less distinct and more spread out. Such a constellation diagram is also a good troubleshooting tool for other purposes, for example to check for amplitude distortion or to see if a faulty symbol encoder is missing

some symbols or placing them at the wrong location.

In a system where a root-Nyquist filter is used in both the transmitter and receiver to obtain a net Nyquist response, the transmitter output will not display sharply defined symbol locations as in the figure. In that case, the measuring instrument must supply the missing root-Nyquist filter that is normally in the receiver, in order to obtain a clean display. Professional modulation analyzers normally include a means of selecting the proper matching filter.

Another way to view modulation accuracy is with an *eye diagram*. See **Figures 11.47** and **11.48** In this case the oscilloscope's horizontal axis is driven by its internal sweep generator, triggered by the symbol clock. The two vertical channels of the oscilloscope are connected to the I and Q outputs of the baseband generator. The "eye" is the empty area at the center of the I and Q traces in between the symbol decision points, where all the traces come together. The eye should be as "open" as possible to make the job of the receiver's symbol decoder as easy as possible. The oscilloscope would typically be set for infinite persistence, so that the worst-case excursions from the ideal symbol trajectory are recorded. QPSK has four symbol locations, one in each quadrant, such that I and Q each only have two possible values and there is only one "eye" for each. 16-QAM has 16 symbol locations with four possible locations for I and Q, which forms three "eyes."

11.11 References

"A Brief Introduction to Sigma-Delta Conversion," Intersil Application Note AN9504. **www.renesas.com/us/en/www/doc/application-note/an9504.pdf**

Andrews, J., KH6HTV, *ATV Handbook*, revised 2021, **www.arrl.org/files/file/Technology/pdf/an-55a-atv-handbook-1.pdf**

Costas, J. P., "Poisson, Shannon, and the Radio Amateur," *Proceedings of the IRE*, vol 47, pp 2058-2068, Dec 1959.

"Digital Modulation in Communications Systems — An Introduction" Agilent Technologies, Application Note 1298. **literature.cdn.keysight.com/litweb/pdf/5965-7160E.pdf?id=1817805**

Ford, S., WB8IMY, *Get On the Air with HF Digital* 3rd Edition (ARRL, 2022).

Ford, S., WB8IMY, *ARRL's VHF Digital Handbook* (ARRL, 2008).

Haykin, S., *Digital Communications* (Wiley, 1988).

Langner, J., WB2OSZ, "SSTV Transmission Modes," **www.comunicacio.net/digigrup/ccdd/sstv.htm**.

Mack, R., W5IFS, SDR: Simplified column, *QEX*, from 2009.

"Manual of Transmission Methods – Reference Document," Rohde & Schwarz, 2014, available at **resources.rohde-schwarz-usa.com/c/manual-of-transmissi-2**

Nagle, J., K4KJ, "Diversity Reception: an Answer to High Frequency Signal Fading," *Ham Radio*, Nov 1979, pp. 48 – 55.

Sabin, W., et al, *Single-Sideband Systems and Circuits* (McGraw Hill, 1987).

Shavkoplyas, A., VE3NEA, "CW Shaping in DSP Software," *QEX*, May/Jun. 2006, pp. 3 – 7.

Seiler, T., HB9JNX/AE4WA, et al, "Digital Amateur TeleVision (D-ATV)," proc. 2001 ARRL/TAPR Digital Communications Conference.

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

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

Smith, D., KF6DX, "Distortion and Noise in OFDM Systems," *QEX*, Mar./Apr. 2005, pp. 57 – 59.

Smith, D., KF6DX, "Digital Voice: The Next New Mode?," *QST*, Jan. 2002, pp. 28 – 32.

Stanley, J., K4ERO, "Observing Selective Fading in Real Time with Dream Software," *QEX*, Jan./Feb. 2007, pp. 18 – 22.

Taggart, R., WB8DQT, "An Introduction to Amateur Television," *QST*, Apr., May., and Jun. 1993.

Taggart, R., WB8DQT, *Image Communications Handbook* (ARRL, 2002).

Taggart, R., WB8DQT, "Digital Slow-Scan Television," *QST*, Feb. 2004, pp. 47 – 51.

Van Valkenburg, M., et al, *Reference Data for Engineers: Radio, Electronics, Computer and Communications*, 9th Edition (Newnes, 2001). Chapters 23, 24 and 25.