Contents

Chapter 8

Modulation

Radio amateurs use a wide variety of modulation types to convey information. This chapter, written by Alan Bloom, N1AL, explores analog, digital and image modulation techniques, as well as impairments that affect our ability to communicate. A glossary of terms and suggestions for additional reading round out the chapter.

8.1 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, "Am I transmitting or not?" One can think of it as a single data bit with two states, *on* or *off*. However, in reality even a beacon station must periodically identify with the station call sign, so we can say that modulation is an essential ingredient of Amateur Radio.

To represent the information being sent, the radio signal must periodically change its 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 analog US 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 Modes** chapter covers digital transmission protocols and the **Digital Communications** supplement on the *Handbook* CD covers the practical aspects of operating the various digital modes.

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 802.11n, the latest generation of "WiFi" wireless LAN, uses up to 40 MHz for data transmission.

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 sidebar, "Emission Classifications." 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.

For example, the designator for a CW signal might be 150H0A1A, which means 150 Hz

Emission Classifications

Emissions are designated according to their classification and their necessary bandwidth. A minimum of three symbols is used to describe the basic characteristics of radio waves. Emissions are classified and symbolized according to the following characteristics:

I. First symbol—Type of modulation of the main carrier
II. Second symbol—Nature of signal(s) modulating the main carrier
III. Third symbol—Type of information to be transmitted
Note: A fourth and fifth symbol are provided for in the ITU Radio Regulations. Use of the fourth and fifth symbol is optional.
IV. Details of signal(s)
V. Nature of multiplexing

First symbol—type of modulation of the main carrier

(1) Emission of an unmodulated carrier .... N
(2) Emission in which the main carrier is amplitude-modulated (including cases where subcarriers are angle-modulated):
- Double sideband .... A
- Single sideband, full carrier .... H
- Single sideband, reduced or variable level carrier .... R
- Single sideband, suppressed carrier .... J
- Independent sidebands .... B
- Vestigial sideband C

(3) Emission in which the main carrier is angle-modulated:
- Frequency modulation .... F
- Phase modulation .... G

Note: Whenever frequency modulation (F) is indicated, phase modulation (G) is also acceptable.

(4) Emission in which the main carrier is amplitude and angle-modulated either simultaneously or in a pre-established sequence .... D
(5) Emission of pulses[1]
- Sequence of unmodulated pulses .... P
- A sequence of pulses:
- Modulated in amplitude .... K
- Modulated in width/duration L
- Modulated in position/phase .... M
- In which the carrier is angle-modulated during the period of the pulse .... Q
- Which is a combination of the foregoing or is produced by other means V

(6) Cases not covered above, in which an emission consists of the main carrier modulated, either simultaneously or in a pre-established sequence in a combination of two or more of the following modes: amplitude, angle, pulse .... W
(7) Cases not otherwise covered X

Second symbol—nature of signal(s) modulating the main carrier

(1) No modulating signal 0
(2) A single channel containing quantized or digital information without the use of a modulating subcarrier, excluding time-division multiplex 1
(3) A single channel containing quantized or digital information with the use of a modulating subcarrier, excluding time-division multiplex 2
(4) A single channel containing analog information .. 3
(5) Two or more channels containing quantized or digital information 7
(6) Two or more channels containing analog information 8
(7) Composite system with one or more channels containing quantized or digital information, together with one or more channels containing analog information 9
(8) Cases not otherwise covered X

Third symbol—type of information to be transmitted[2]

(1) No information transmitted .... N
(2) Telegraphy, for aural reception .... A
(3) Telegraphy, for automatic reception .... B
(4) Facsimile C
(5) Data transmission, telemetry, telecommand .... D
(6) Telephony (including sound broadcasting) .... E
(7) Television (video) .... F
(8) Combination of the above .... W
(9) Cases not otherwise covered X

Where the fourth or fifth symbol is used it shall be used as indicated below. Where the fourth or the fifth symbol is not used this should be indicated by a dash where each symbol would otherwise appear.

Fourth symbol—Details of signal(s)

(1) Two-condition code with elements of differing numbers and/or durations .... A
(2) Two-condition code with elements of the same number and duration without error-correction.... B
(3) Two-condition code with elements of the same number and duration with error-correction C
(4) Four-condition code in which each condition represents a signal element (of one or more bits) .... D
(5) Multi-condition code in which each condition represents a signal element (of one or more bits) .... E
(6) Multi-condition code in which each condition or combination of conditions represents a character .... F
(7) Sound of broadcasting quality (monophonic) ..... G
(8) Sound of broadcasting quality (stereophonic or quadraphonic) .... H
(9) Sound of commercial quality (excluding categories given in (10) and (11) below .... J
(10) Sound of commercial quality with frequency inversion or band-splitting .... K
(11) Sound of commercial quality with separate frequency-modulated signals to control the level of demodulated signal L
(12) Monochrome .... M
(13) Color .... N
(14) Combination of the above .... W
(15) Cases not otherwise covered X

Fifth symbol—Nature of multiplexing

(1) None .... N
(2) Code-division multiplex[3] C
(3) Frequency-division multiplex .... F
(4) Time-division multiplex .... T
(5) Combination of frequency-division and time-division multiplex .... W
(6) Other types of multiplexing X

[1] Emissions where the main carrier is directly modulated by a signal which has been coded into quantized form (eg, pulse code modulation) should be designated under (2) or (3).
[2] In this context the word “information” does not include information of a constant unvarying nature such as is provided by standard frequency emissions, continuous wave and pulse radars, etc.
[3] This includes bandwidth expansion techniques.

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.

8.2 Analog Modulation

While the newer digital modes are all the rage today, traditional analog modulation is still quite popular on the amateur bands. There are many reasons for that, some technical and some not. Analog equipment tends to be simpler; no computer is required. For example, a low-power CW transceiver is easy to construct and simple to understand. CW and SSB are both narrowband modes with spectral efficiency that is quite competitive with their digital equivalents. While heated arguments have been made about the relative power efficiency of analog vs digital modes, it is clear that in the hands of skilled operators CW and SSB are at least in the same ballpark with, if perhaps not better than, many digital modes.

When copying under difficult conditions, the human brain is capable of higher-level signal processing that is difficult to duplicate on a computer. For example, in a DX pileup, an operator may realize that the signal he wants is the weak one and mentally filter out all the stronger stations based on signal strength alone. Other aspects of the signal may be used for filtering as well, such as the operator's accent (on voice) or "fist" (on CW), propagation effects, the operator's speed and operating pattern, CW hum and chirp, and other factors. Different portions of the other station's call sign may be copied at different times and stitched together in the operator's brain. Missing portions of a transmission can be filled in based on context. Many operators enjoy the challenge of using their ears and brain to decipher a weak signal, an aspect of operating that is lost with the digital modes.

8.2.1 Amplitude-Modulated Analog Signals

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. Since the RF output voltage of a class-C amplifier is approximately proportional to the power supply voltage, an AM modulator needs only to vary the power supply voltage in step with the audio signal, as shown in **Fig 8.1**.

To prevent overmodulation and the resulting distortion, the modulating signal should not try to reduce the power supply voltage to less than zero volts or more than twice the average voltage. The *modulation index* of an AM signal is the ratio of the peak modulation to those limits. The resulting RF signal might look like the waveform in **Fig 8.2C**.

Note that the shape of the envelope of the modulated RF signal matches the shape of the modulating signal. That suggests a possible demodulation method. A diode detector puts out a signal proportional to the envelope of the RF signal, recovering the original modulation. See **Fig 8.3**. The capacitor should be large enough that it filters out most of the RF ripple but not so large that it attenuates the higher audio frequencies.

So far we have been discussing signals as a function of time — all the graphs have time as the horizontal axis. It is equally valid

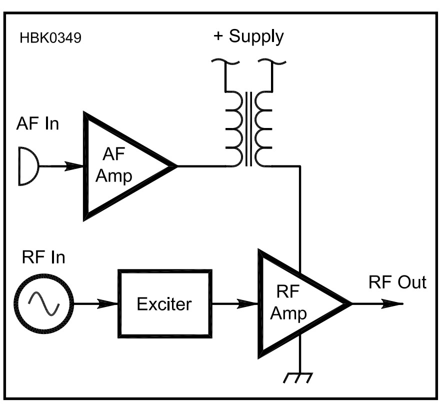

Fig 8.1 — An AM modulator. With a vacuum-tube RF power amplifier this method is known as *plate modulation* since the modulated power supply voltage is applied to the vacuum tube plate.

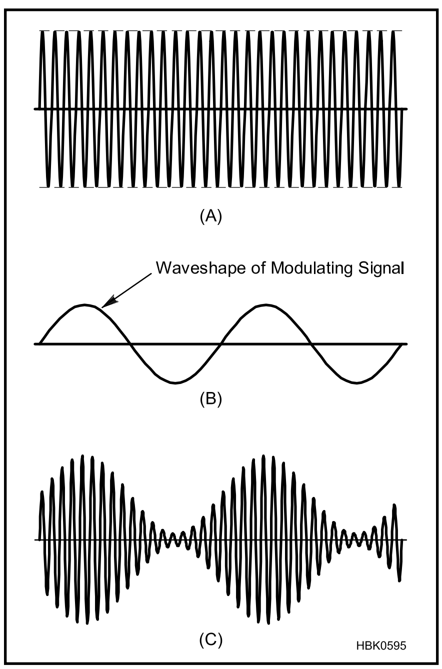

Fig 8.2 — Graphical representation of amplitude modulation. In the unmodulated carrier (A) each RF cycle has the same amplitude. When the modulating signal (B) is applied, the RF amplitude is increased or decreased according to the amplitude of the modulating signal (C). A modulation index of approximately 75% is shown. With 100% modulation the RF power would just reach zero on negative peaks of the modulating signal.

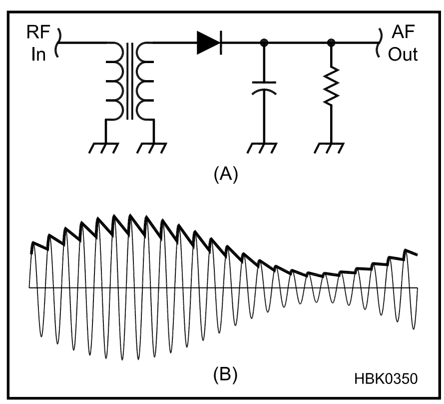

Fig 8.3 — A simple diode-type AM detector, also known as an *envelope detector* (A). The demodulated output waveform has the same shape as the envelope of the RF signal (B). In (B) the thin line is the RF signal modulated with a sine wave and the darker line is the demodulated audio frequency with some residual RF ripple.

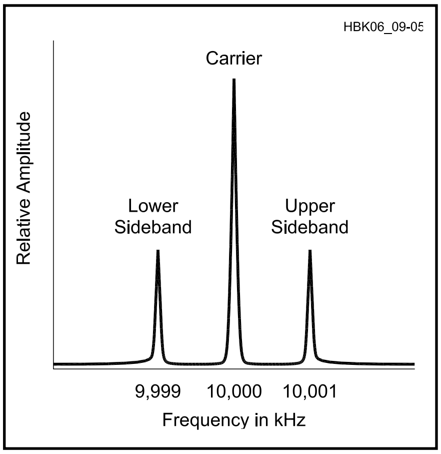

Fig 8.4 — A 10-MHz carrier that is AM-modulated by a 1-kHz sine wave.

to look in the frequency domain. It turns out that the modulation spectrum consists of a pair of sidebands above and below the RF carrier. **Fig 8.4** shows 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.

To see why that happens, refer to **Fig 8.5**. Simply adding the first three sine waves shown, representing a carrier and upper and lower sidebands, results in the 100%-modulated AM waveform at the bottom. Note that each sideband has half the amplitude of the carrier, which corresponds to one-quarter the power or –6 dB. Each sideband in an AM signal that is 100% modulated with a sine wave is therefore –6 dB with respect to the carrier.

If a more complex modulating signal is used, each frequency component of the modulation f_m appears at $f_c + f_m$ and $f_c - f_m$, where f_c is the carrier frequency. An AM modulator is really nothing more than a type of mixer that generates new output frequencies at the sum and difference of the carrier and modulation frequencies. The upper sideband is simply a frequency-shifted version of the modulation spectrum and the lower sideband is both frequency-shifted and inverted. For communications purposes, the audio frequency response typically extends up to about 2.5-3 kHz, resulting in a 5-6 kHz RF bandwidth.

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

A solution is to regenerate the carrier in the receiver. Since the carrier itself carries no information about the modulation, it is not necessary for demodulation. The transmitted signal may be a standard full-carrier AM signal or the carrier may be suppressed, resulting in *double sideband, suppressed carrier* (DSBSC). An AM detector that regenerates the carrier from the signal is known as a *synchronous detector*. Often the regenerated carrier oscillator is part of a phase-locked loop (PLL) that locks onto the incoming carrier. See **Fig 8.6**. Synchronous detectors not only reduce the effects of selective fading but also are usually more linear than diode detectors so they have less distortion. Some commercial short-wave broadcasts include a reduced, but not suppressed, carrier (DSBRC) to allow operation with PLL-type synchronous detectors.

Since the advent of single sideband in the 1960s, full-carrier double-sideband AM has become something of a niche activity in Amateur Radio. It does retain several advantages however. We have already mentioned the simplicity of the circuitry. Another advantage is that, because of the presence of the

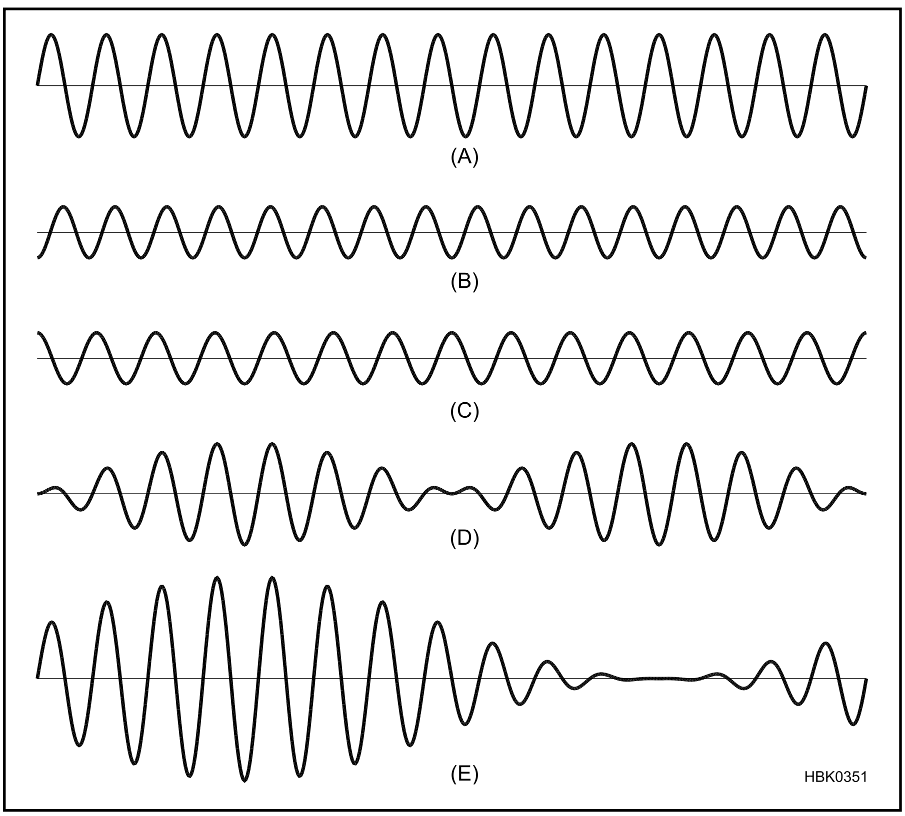

Fig 8.5 — At A is an unmodulated carrier. If the upper (B) and lower (C) sidebands are added together a double-sideband suppressed carrier (DSBSC) signal results (D). If each sideband has half the amplitude of the carrier, then the combination of the carrier with the two sidebands results in a 100%-modulated 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.

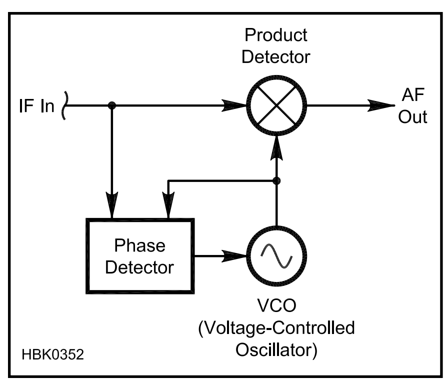

Fig 8.6 — Block diagram of a synchronous detector. The voltage-controlled oscillator (VCO) is part of a phase-locked loop that locks the oscillator to the carrier frequency of the incoming AM or DSBRC signal.

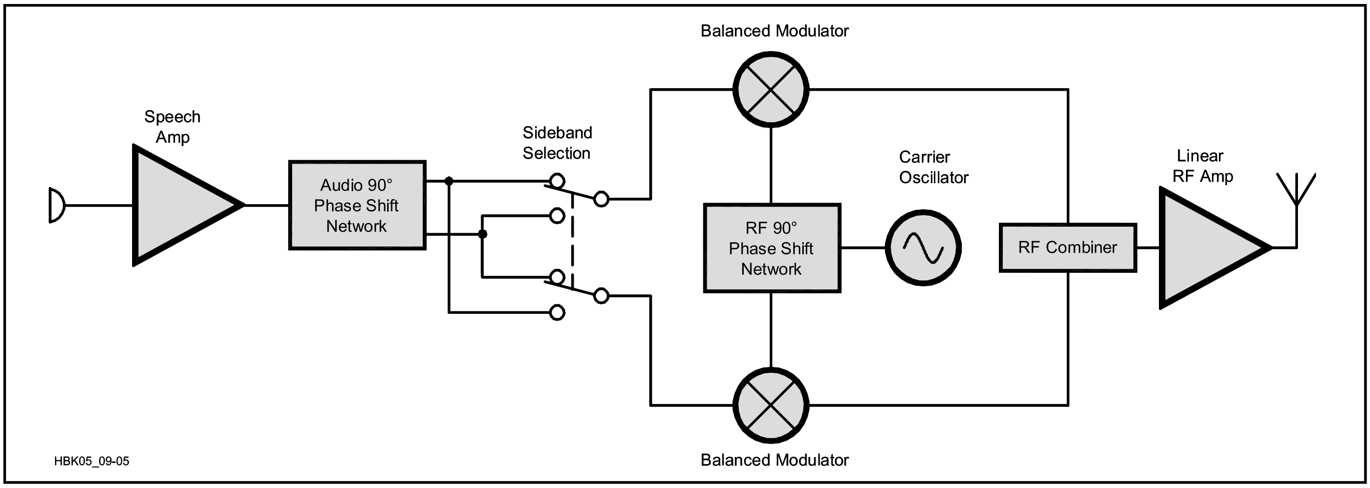

Fig 8.7 — Block diagram of a phasing-type SSB generator.

carrier, the automatic gain control system in the receiver remains engaged at all times, ensuring a constant audio level. Unlike with SSB, there is no rush of noise during every pause in speech. Also tuning is less critical than with SSB. There is no "Donald Duck" sound if the receiver is slightly mistuned. Finally, the audio quality of an AM signal is usually better than that of SSB because of the lack of a crystal filter in the transmit path and the wider filter in the receiver.

SINGLE SIDEBAND

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. In addition, 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* (SSBSC), 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). The bandwidth is a little less than half that of an equivalent AM signal. A typical SSB transmitter might have an overall audio response from 300 Hz to 2.8 kHz, resulting in 2.5 kHz bandwidth.

As with an AM synchronous detector, the carrier is regenerated in the receiver but since only one sideband is present, the synchronous detector's phase-locked loop is not possible. Instead, the detector uses a free-running *beat-frequency oscillator* (BFO). The detector itself is called a *product detector* because its output is the mathematical product of the BFO and the SSB signal. The BFO must be tuned to the same frequency as the suppressed carrier to prevent distortion of the recovered audio. In a superheterodyne receiver, that is done by carefully tuning the local oscillator (the main tuning dial) such that after conversion to the intermediate frequency, the suppressed carrier aligns with the BFO frequency.

In the transmitter, the SSB modulator must suppress both the carrier and the unwanted sideband. Carrier suppression is normally accomplished with a *balanced modulator*, a type of mixer whose output contains the sum and difference frequencies of the two input signals (the modulating signal and the carrier) but not the input signals themselves. There are several ways to eliminate the unwanted sideband, but the most common is to pass the output of the balanced mixer through a crystal filter that passes the wanted sideband while filtering out the unwanted one. This is convenient in a transceiver since the same filter can be used in the receiver by means of a transmit-receive switch.

Another method to generate single sideband is called the *phasing method*. See **Fig 8.7**. Using trigonometry, it can be shown mathematically that the sum of the signals from two balanced modulators, each fed with audio signals and RF carriers that are 90° out of phase, consists of one sideband only. The other sideband is suppressed. The output can be switched between LSB and USB simply by reversing the polarity of one of the inputs, which changes the phase by 180°. The phasing method eliminates the need for an expensive crystal filter following the modulator or, in the case of a transceiver, the need to switch the crystal filter between the transmitter and receiver sections. In addition, the audio quality is generally better because it eliminates the poor phase dispersion that is characteristic of most crystal filters. In the old days, designing and building an audio phase-shift network that accurately maintained a 90° differential over a decade-wide frequency band (300 Hz to 3000 Hz) was rather complicated. With modern DSP techniques, the task is much easier.

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. Back in the days when AM ruled the HF amateur voice bands, operators had to put up with a cacophony of heterodynes emanating from their headphones. 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.

8.2.2 Angle-Modulated Analog Signals

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. Phase modulation is one type of angle modulation, but so is frequency 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 **Fig 8.8**.

Since any change in frequency results in a change in phase and vice versa, *frequency*

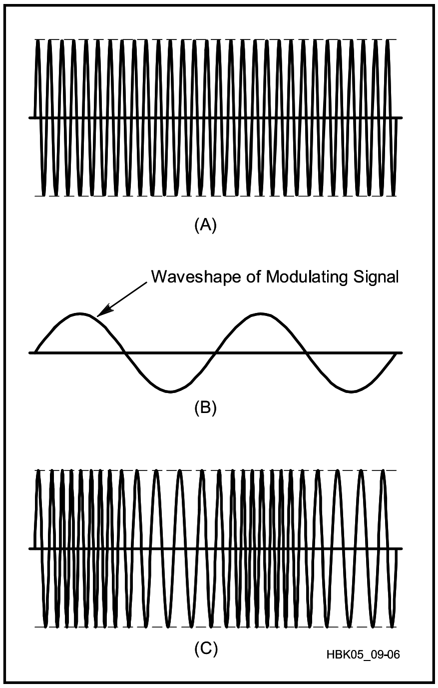

Fig 8.8 — Graphical representation of frequency modulation. In the unmodulated carrier (A) each RF cycle occupies the same amount of time. When the modulating signal (B) is applied, the radio frequency is increased or decreased according to the amplitude and polarity of the modulating signal (C).

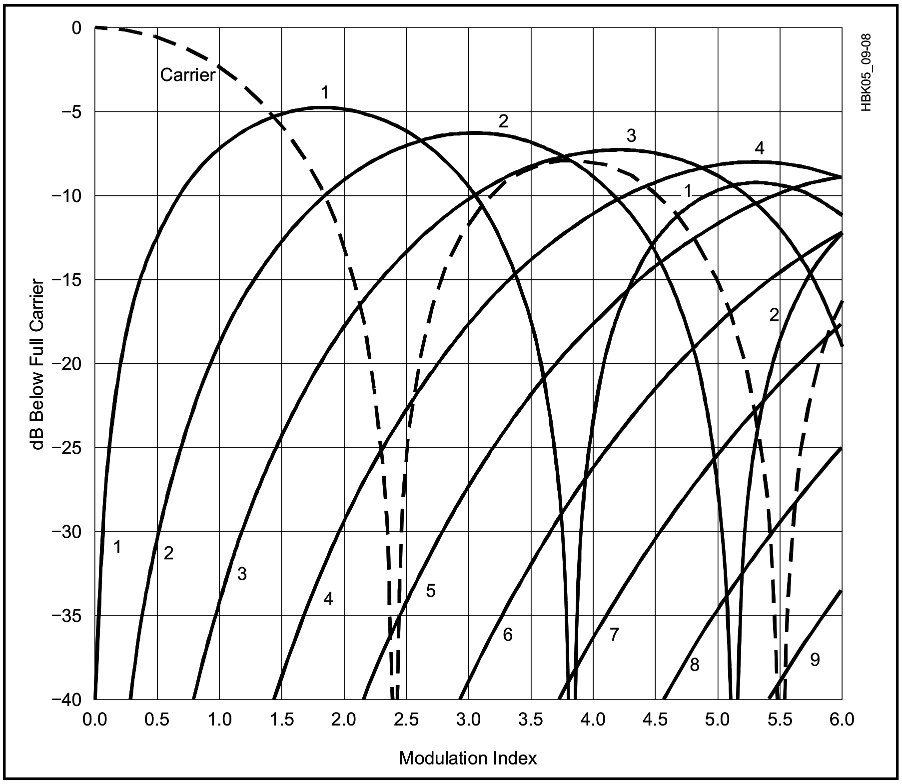

Fig 8.9 — Amplitude of the FM carrier and the first nine sidebands versus modulation index. This is a graphical representation of mathematical functions developed by F. W. Bessel. Note that the carrier completely disappears at modulation indices of 2.405 and 5.52.

modulation (FM) and *phase modulation* (PM) are fundamentally the same. The only difference 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.

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.

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. 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.

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.

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. And it is even worse than that. 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*:

$$BW = 2(f_d + f_m) \quad (1)$$

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.

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 within the bandwidth predicted by Carson's rule. 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.

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 **Fig 8.9**. 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.

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 "0.25 µV for 20 dB of quieting."

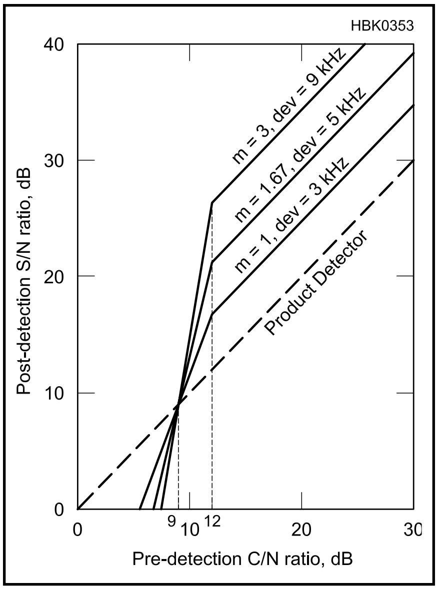

Fig 8.10 — A plot of post-detection signal-to-noise ratio (S/N) versus input carrier-to-noise ratio (C/N) for an FM detector at various modulation indices, m. For each modulation index the deviation is also noted assuming a maximum modulating frequency of 3 kHz. For comparison, the performance of an SSB product detector is also shown.

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 **Fig 8.10**.

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.

8.3 Digital Modulation

The first digital modulation used in Amateur Radio was on-off keying using Morse code. However, the term "digital" is commonly understood to refer to modes where the sending and receiving is done by machine. By that definition, the first popular amateur digital mode (called radioteleprinter, or RTTY) used frequency-shift keying with all the coding and decoding done by mechanical teleprinters. You still hear RTTY on the bands today, although most operators are using computers these days rather than Teletype machines. More-modern digital modes use more efficient modulation formats as well as error detection and correction to ensure reliable communications. Data can be sent much more rapidly than with Morse code or voice, with less effort, and with fewer errors. In some systems, stations operate unattended 24 hours per day, transferring data with little or no operator intervention. For point-to-point transmission of large quantities of text or other data, these modern digital modes are the obvious choice.

8.3.1 Amplitude-Modulated Digital Signals

We start this section by discussing *on-off keying* (OOK), which is a special case of *amplitude-shift keying* (ASK) and is normally sent using 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. While most amateurs think of CW as being an analog mode, we include it here because it has technical characteristics that are more similar to other digital modes than to analog.

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 **Fig 8.11**. 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

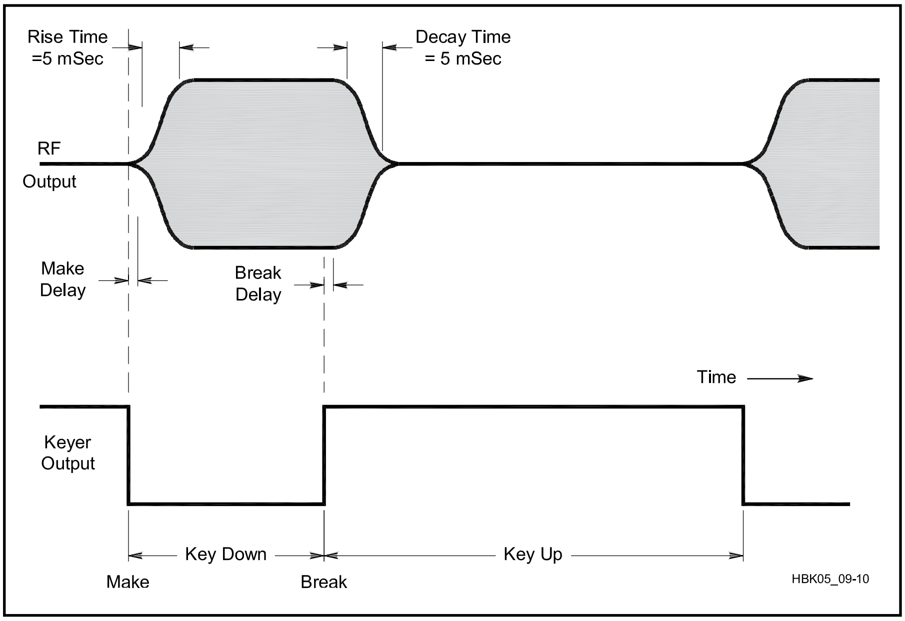

Fig 8.11 — A filtered CW keying waveform. The on-off transitions of the RF envelope should be as smooth as possible while transitioning as quickly as possible. The raised-cosine shape shown is nearly optimum in that respect.

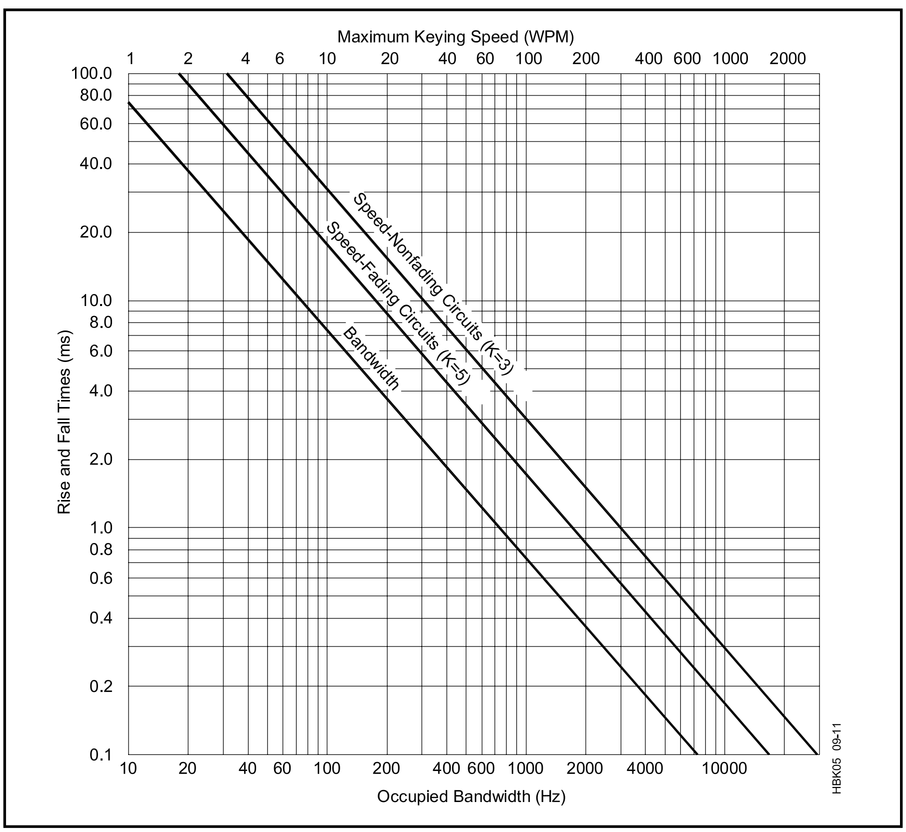

Fig 8.12 — Keying speed vs rise and fall times vs bandwidth for fading and non-fading communications circuits. For example, to optimize transmitter timing for 20 WPM on a non-fading circuit, draw a vertical line from the WPM axis to the K = 3 line. From there draw a horizontal line to the rise/fall axis (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.

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. **Fig 8.12** 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 *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 **Fig 8.13**. 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:

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

$$= 1.2 \times \text{bauds} = 2.4 \times \text{dits / sec} \qquad (2)$$

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.

PULSE MODULATION

Another type of digital amplitude modulation comes under the general category of *pulse modulation*. The RF signal is broken

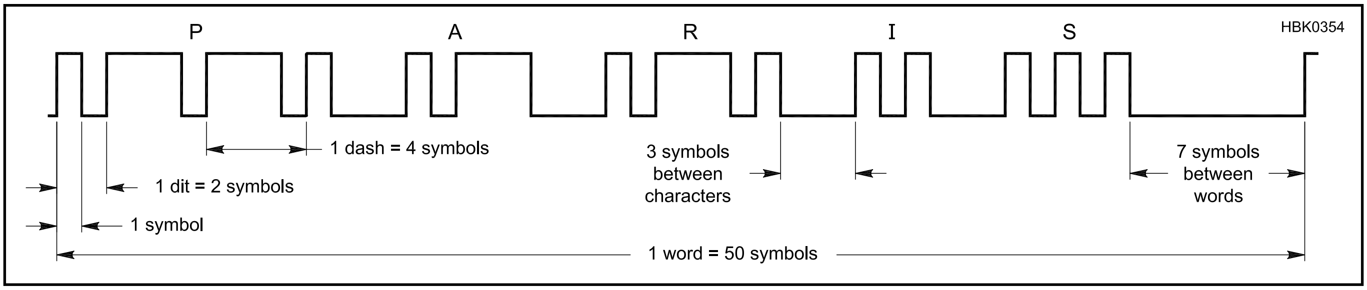

Fig 8.13 — CW timing of the word PARIS, which happens to have a length equal to the standard 50 symbols per word. By programming it multiple times into a memory keyer, the speed may be calibrated simply by counting the number of times the word is completed in one minute.

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 **Fig 8.14B**. 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 Fig 8.14B, 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 DSBSC.

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.

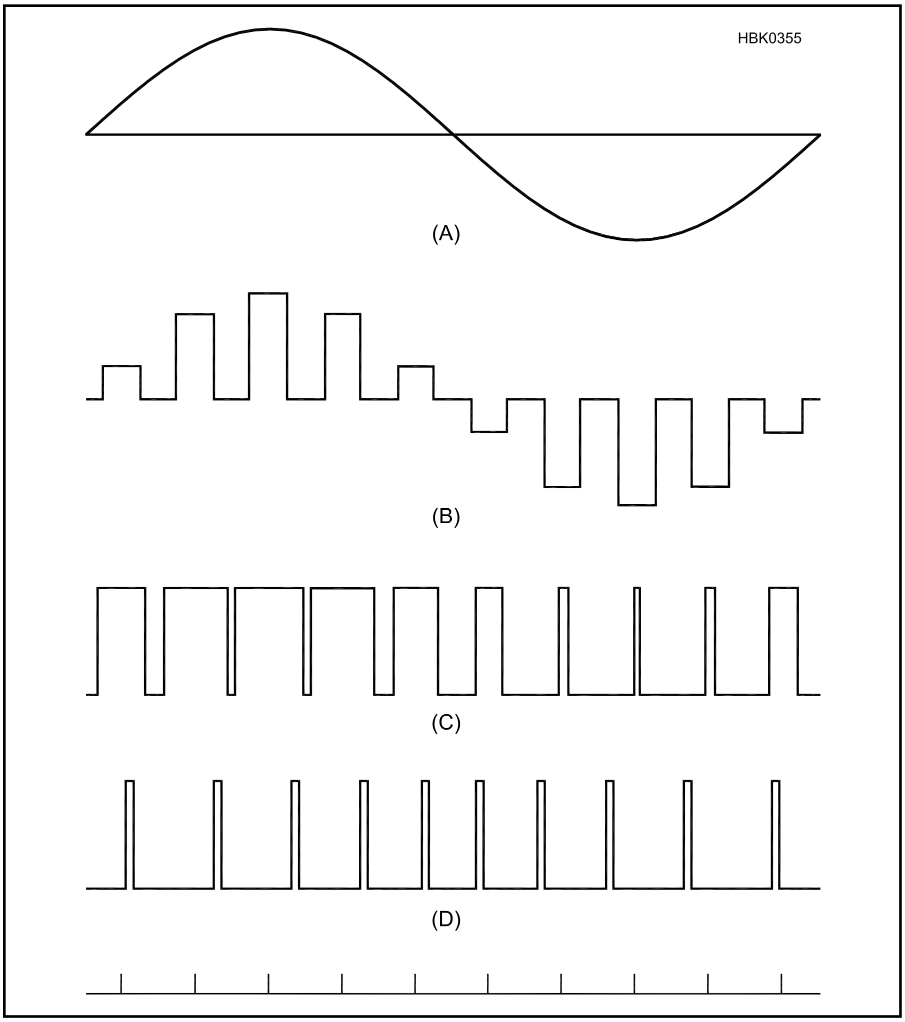

Fig 8.14 — Three types of pulse-code modulation. A sine-wave modulating signal (A) is shown at the top. Pulse amplitude modulation (B) varies the amplitude of the pulses, pulse width modulation (C) varies the pulse width, and pulse-position modulation (D) varies the pulse position, proportional to the 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.

Pulse-width modulation (PWM) is a series of pulses whose width varies in proportion to the amplitude of the modulating signal. Fig 8.14C 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),

Fig 8.14D, 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.

8.3.2 Angle-Modulated Digital Signals

Just as with analog modulation, angle-modulated digital signals have constant amplitude and vary the phase and frequency to represent the modulating signal. They share some of the same advantages as their analog counterparts. The receiver can be made relatively insensitive to amplitude noise and on-channel interference. The constant-amplitude signal simplifies the job of the receiver's automatic-gain-control system. The transmitter does not require a linear amplifier, so a more-efficient class-C amplifier may be used. However digital angle modulation also shares analog's nonlinearity so that the bandwidth is not as well-constrained as it is with linear modulation systems.

Frequency-shift keying (FSK) was the first digital angle-modulated format to come into common use. FSK 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) must normally be 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*. See **Fig 8.15.** 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 bauds.

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, 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 8FSK than with binary FSK and the receiver is theoretically seven times (16.9 dB) more susceptible to noise and interference. However the IF limiter stage removes 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.

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.

Minimum-shift keying (MSK) 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 baud rate or, using the common definition of frequency shift, a difference between the two tones of 0.5 times the baud 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 causes the phase to change by either +90 or –90° from one symbol to the next.

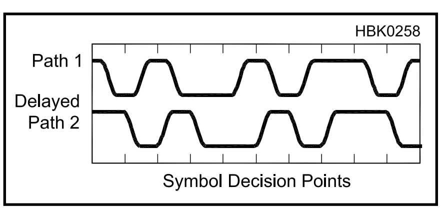

Fig 8.15 — Multipath propagation can cause inter-symbol interference (ISI). At the symbol decision points, which is where the receiver decoder samples the signal, the path 2 data is often opposing the data on path 1.

Gaussian minimum-shift keying (GMSK) refers to MSK where the modulating signal has been filtered with a low-pass filter that has 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. It turns out that a Gaussian filter, with its gradual transition from passband to stopband, has the optimum shape for an angle-modulated digital system. However, a Gaussian filter also has a gradual transition from symbol-to-symbol in the time domain as well. The transition is not totally completed by the time of the next symbol, which means that there is some inter-symbol interference 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.

Binary phase-shift keying (BPSK), often referred to simply as phase-shift keying (PSK), is included in this section because the name suggests that it is true constant-amplitude angle modulation. It is possible to implement it that way. An example is MSK which, as described above, can be considered to be differential BPSK with a ±90° phase shift. However the term BPSK is normally understood 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.

8.3.3 Quadrature Modulation

Quadrature modulation encodes digital signals using a combination of amplitude

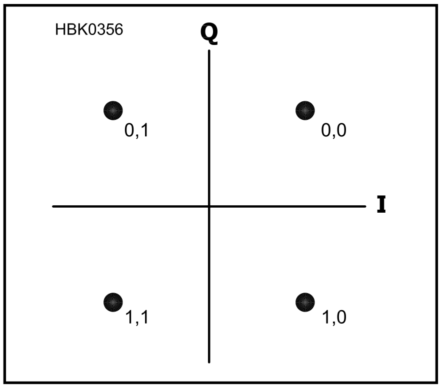

Fig 8.16 — Constellation diagram of a 4-QAM signal, also known as QPSK. The four symbol locations all have the same amplitude and have phase angles of 45°, 135°, –135° and –45°. 4-QAM is a two-bits-per-symbol format. The four symbols are selected by the four possible states of the two data bits, which 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.

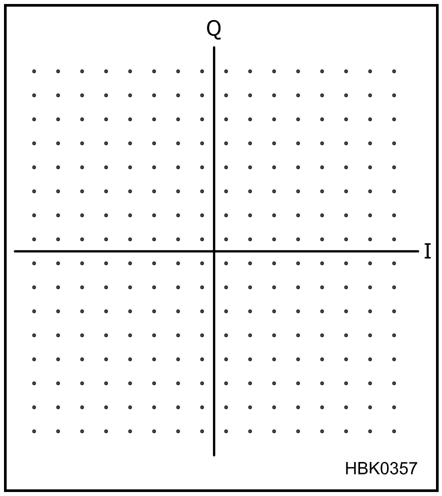

Fig 8.17 — Constellation diagram of a 256-QAM signal. Since an 8-bit number has 256 possible states, each symbol represents 8 bits of data.

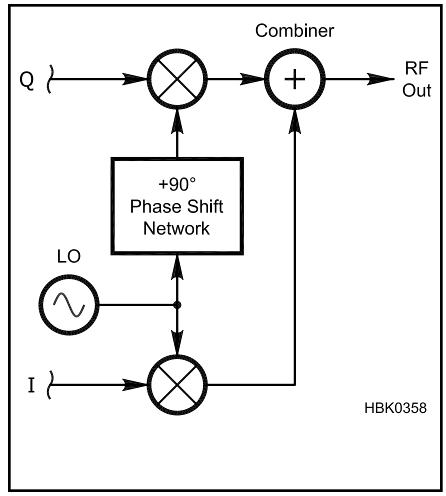

Fig 8.18 — Block diagram of an I/Q modulator. Note the similarity to Fig 8.7. By connecting an audio 90° phase-shift network to the I and Q inputs, an I/Q modulator can generate an analog SSB signal by the phasing method.

and phase modulation. With two types of modulation to work with, it is possible to cram 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.

Since quadrature-modulated symbols are defined by both amplitude and phase, the most common way to represent symbol states is with a polar plot, called a *constellation diagram*. See **Fig 8.16**, which 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. **Fig 8.17** 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.

There are at least two ways to generate a QAM signal. One is to combine a phase modulator and an amplitude modulator. The phase modulator places the symbol location at the correct phase angle and the amplitude modulator adjusts the amplitude so the symbol is the correct distance from the origin. That method is seldom used, both because of the circuit complexity and because the nonlinear phase modulator makes the signal difficult to filter so as to limit the bandwidth.

A much more common method of generating QAM is with an *I/Q modulator*. See **Fig 8.18**. By using two modulators (mixers) 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.

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. 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 **Fig 8.19**. 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.

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.

The block diagram of an I/Q demodulator looks like an I/Q modulator drawn backward. See **Fig 8.20**. If the local oscillator is not tuned to exactly the same frequency as the one in the transmitter (after downconversion to the IF, assuming 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.

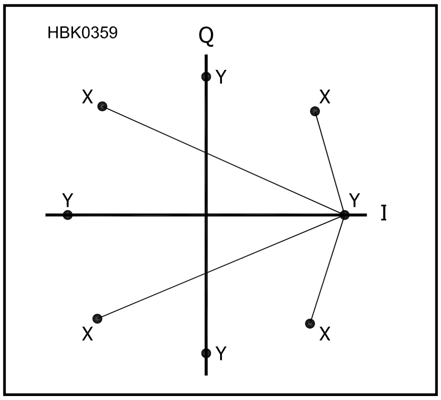

Fig 8.19 — Constellation diagram of a π/4 DQPSK signal. There are eight possible symbol locations. If the current symbol location is at one of the four positions labeled "X" then the next symbol will be at one of the four locations labeled "Y" and vice versa. That guarantees that no possible symbol trajectories (the lines between symbol locations) can ever pass through the origin.

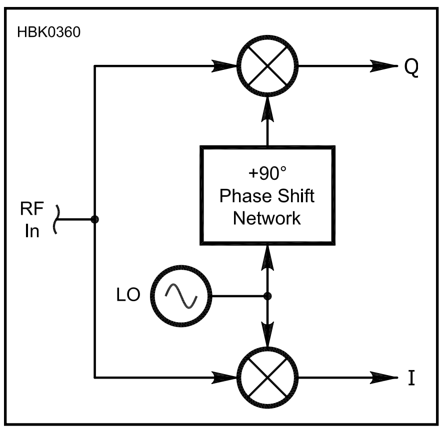

Fig 8.20 — Block diagram of an I/Q demodulator.

8.3.4 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-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 following section on filtering and bandwidth.) 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). 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 **Fig 8.21B**. 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 Fig 8.21A.

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

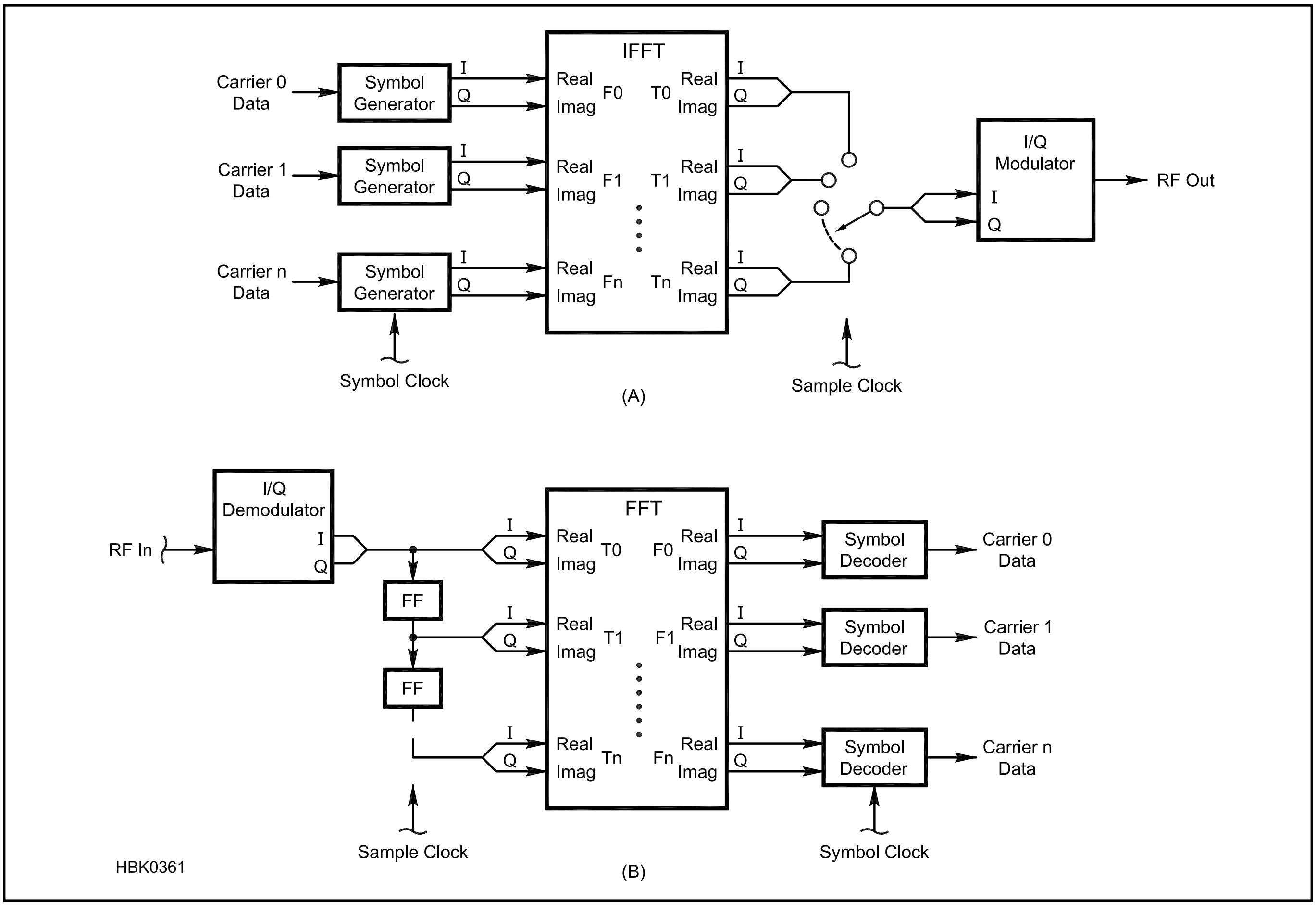

Fig 8.21 — Block diagram of an OFDM modulator (A) and demodulator (B) using the FFT/IFFT technique. The number of carriers is n, and the sample rate is n times the symbol rate. Once per symbol, all the symbol generators in the modulator are loaded with new data and the inverse fast-Fourier transform (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.

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.

8.3.5 AFSK and Other Audio-Based Modulation

Audio frequency-shift keying (AFSK) is the generation of radio-frequency FSK using an audio-frequency FSK signal fed into the microphone input of an SSB transmitter. 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 modulation type 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 *slow-scan television* (SSTV), which uses frequency modulation for the video signal. In addition, nearly all narrowband digital signals today are generated in this manner. The audio is generated and received either from a dedicated hardware modulator/demodulator (modem) or using the sound card that is found in virtually all personal computers.

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.

8.3.6 Spread Spectrum

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.

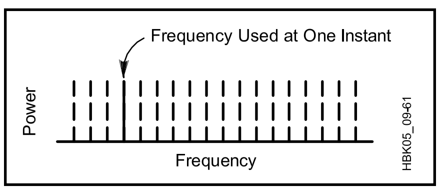

Fig 8.22 — Power versus frequency for a frequency-hopping spread spectrum signal. Emissions jump around to discrete frequencies in pseudo-random fashion. Normally the spacing of the frequencies is approximately equal to the bandwidth of the unspread signal so that the average spectrum is approximately flat.

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:

$$\text{Processing Gain} = 10 \times \log\left(\frac{\text{spread bandwidth}}{\text{unspread bandwidth}}\right) \text{ dB} \qquad (3)$$

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.

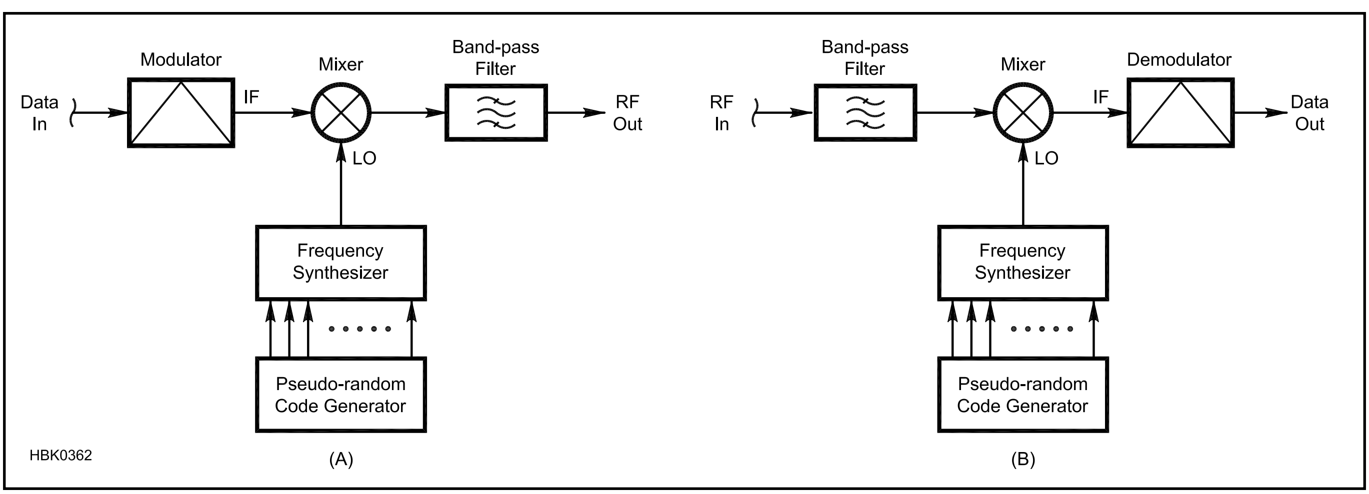

Fig 8.23 — Block diagram of an FHSS transmitter (A) and receiver (B). The receiver may be thought of as a conventional superhet with a local oscillator (LO) that is continually hopping its frequency in response to a pseudo-random code generator. The transmitter has a similar architecture to up-convert a conventionally-modulated intermediate frequency (IF) to a frequency-hopped radio frequency (RF).

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 **Fig 8.22**.

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.

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 then 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 **Fig 8.23**. 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.

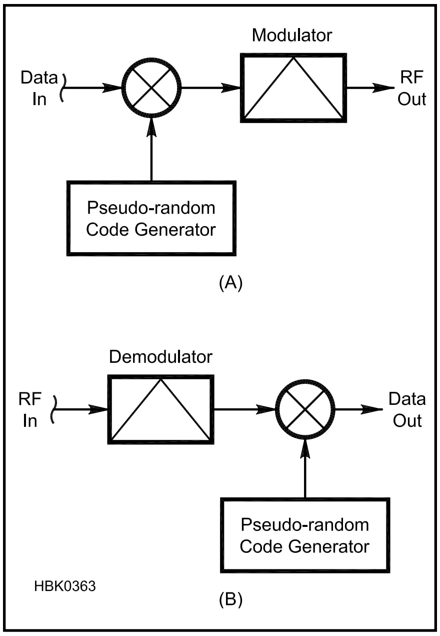

Fig 8.24 — Block diagram of a DSSS transmitter (A) and receiver (B). For BPSK modulation, the modulation has the same format as the bipolar data (±1), so the modulator and demodulator could be moved to the other side of the multiplier if desired.

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 **Fig 8.24**. 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:

$$\text{Processing Gain} = 10 \times \log\left(\frac{\text{chip rate}}{\text{bit rate}}\right) \text{ dB} \quad (4)$$

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 Fig 8.24 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 **Fig 8.25A**, 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. Fig 8.25B shows the DSSS signal in the presence of noise at the input to the receiver, and Fig 8.25C illustrates the improvement in signal-to-noise ratio of the de-spread narrowband signal.

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

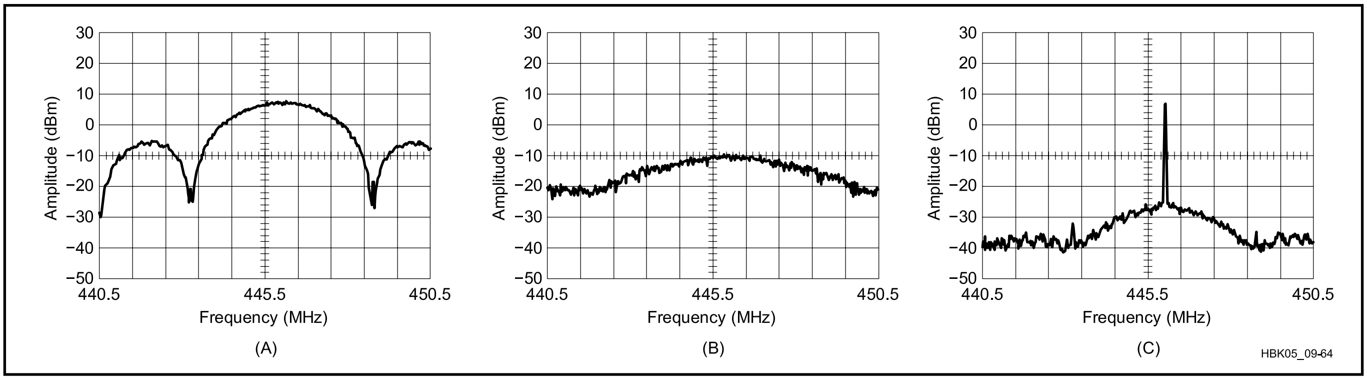

Fig 8.25 — (A) The frequency spectrum of an actual unfiltered biphase-modulated spread spectrum signal as viewed on a spectrum analyzer. In this practical system, band-pass filtering is used to confine the spread spectrum signal to the amateur band. (B) At the receiver end of the line, the filtered spread 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.)

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.

8.3.7 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.

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

$$\operatorname{sinc}(f / f_s) = \frac{\sin(\pi f / f_s)}{\pi f / f_s} \qquad (5)$$

where f_s is the symbol rate. See **Fig 8.26**. 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.

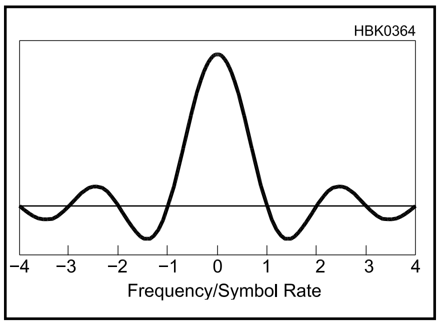

Fig 8.26 — The sinc function, which is the spectrum of an unfiltered BPSK modulating signal with random data, plotted with a linear vertical scale. The center (zero) point corresponds to the RF carrier frequency, To see what the double-sided RF spectrum looks like on a logarithmic (dB) scale, see Fig 8.25A.

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.

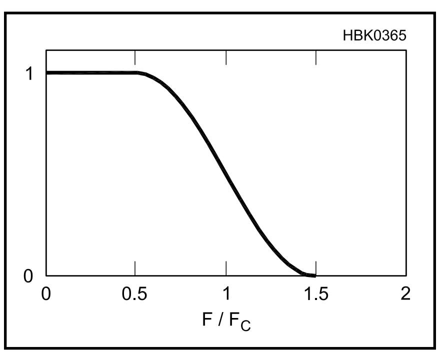

Fig 8.27 — Amplitude versus frequency for a 0.5-alpha, raised-cosine filter. The vertical scale is linear, not logarithmic as would be seen on a spectrum analyzer. The amplitude is 0.5 (–6 dB) at the cutoff frequency, Fc. The amplitude is 1.0 for frequencies less than Fc × (1 – alpha) and is 0.0 for frequencies above Fc × (1 + alpha).

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 Fig 8.27**. 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 Software Radio Design** chapter.

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 intersymbol 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.

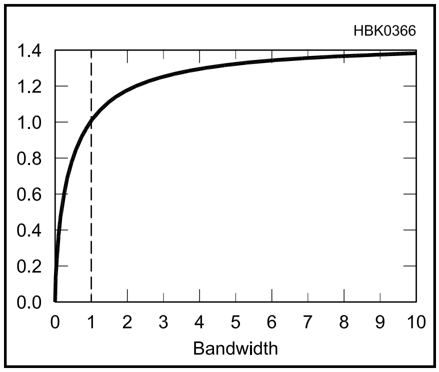

Fig 8.28 — Plot of channel capacity versus bandwidth, calculated by the Shannon-Hartley theorem. The S/N ratio has been selected to be unity at Bandwidth = 1.

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:

$$C = B\log_2\left(1+\frac{S}{N}\right) \text{ bits/s} \qquad (6)$$

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. **Fig 8.28** 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.

8.4 Image Modulation

The following section covers the modulation aspects of amateur television and facsimile communications. More detailed information on protocols and operating standards can be found in the **Image Communications** supplement on the *Handbook* CD.

8.4.1 Fast-Scan Television

Amateur fast-scan television (ATV) is a wideband mode used in the amateur bands above 420 MHz. It is called "fast scan" to differentiate it from slow-scan TV.

The most popular ATV mode is based on NTSC, the same technical standard used by commercial analog television stations in the United States before most switched to digital TV in 2009. **Fig 8.29** shows the spectrum of an NTSC analog TV channel. It's basically a full-carrier, double-sideband AM signal, with filtering to partially remove the lower sideband. The partially-filtered *vestigial sideband* (VSB) extends 1.25 MHz below the carrier frequency. The channel is 6 MHz wide to accommodate the composite video

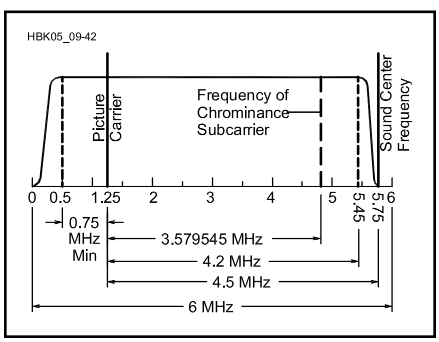

Fig 8.29 — An analog NTSC 6-MHz video channel with the video carrier 1.25 MHz up from the lower edge. The color subcarrier is at 3.58 MHz and the sound subcarrier at 4.5 MHz above the video subcarrier.

and two subcarriers, one at 3.58 MHz for the color burst and the other an FM-modulated signal at 4.5 MHz for the sound. For simplicity, amateur stations often transmit unfiltered full-DSB AM, but the normally-removed portion of the lower sideband is unused by the TV receiver. Since less than 5% of the video energy appears more than 1 MHz below the carrier, little of the transmitter power is wasted. If needed, to reduce interference to other band users, a VSB filter in the antenna line can attenuate the lower sideband color and sound subcarrier frequencies by 20-30 dB.

The video signal includes pulses to synchronize the vertical and horizontal scan oscillators in the receiver. See **Fig 8.30**. The sync pulses and the "front porch" and "back porch" areas that bracket them are "blacker than black" so that the signal is blanked during retrace. 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 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.

Given NTSC's 525 horizontal lines and its 30 frames per second scan rate, the resulting horizontal resolution bandwidth is 80 lines

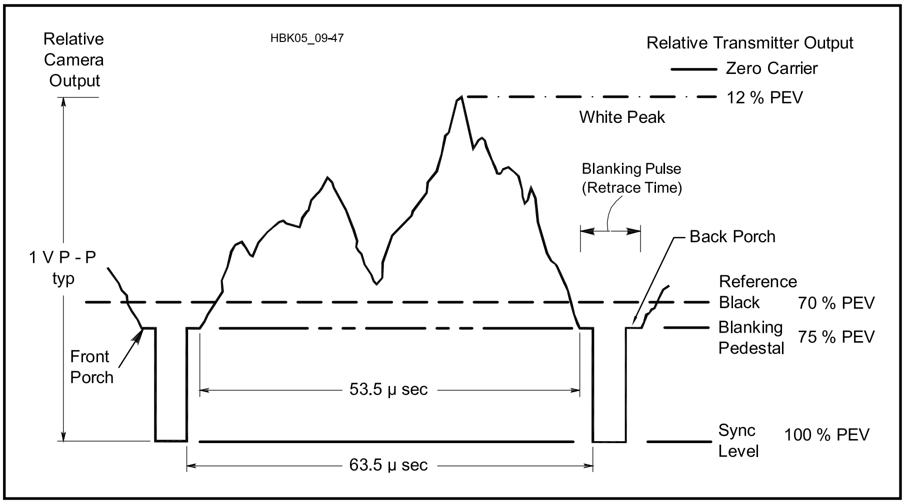

Fig 8.30 — An ATV waveform, showing the relative camera output as well as the transmitter output RF power during one horizontal line scan for black-and-white TV. (A color camera would generate a "burst" of 8 cycles at 3.58 MHz on the back porch of the blanking pedestal.) Note that "black" corresponds to a higher transmitter power than "white."

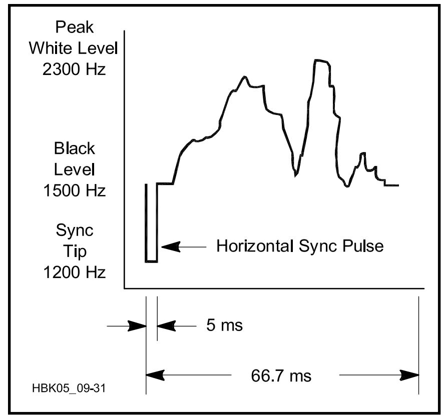

Fig 8.31 — The basic 8-second black and white transmission format developed by early SSTV experimenters. The sync pulses are "blacker than black" to blank the signal during retrace. A complete frame has 120 lines (8 seconds at 15 lines per second). Horizontal sync pulses occur at the beginning of every line; a 30-ms vertical sync pulse precedes each frame.

per MHz. Therefore, with the typical TV set's 3-dB rolloff at 3 MHz (primarily in the IF filter), up to 240 vertical black lines can be seen, corresponding to 480 pixels per line. Color bandwidth in a TV set is less than that, resulting in up to 100 color lines. (Lines of resolution should not be confused with the number of horizontal scan lines per frame.)

The PAL analog TV system in Europe used frequency modulation rather than AM. 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. 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. Most available FM ATV equipment is made for the 1.2, 2.4 and 10.25-GHz bands.

Digital fast-scan TV has been explored by amateurs in Europe and the US using the commercial DVB-S satellite digital video broadcasting standard. It uses MPEG2 audio and video data compression and QPSK modulation with symbol rates up to 20 Mbauds. Much of work has been on the 23 cm band since inexpensive set-top boxes are available that cover that frequency range.

In the United States and some other countries, commercial over-the-air digital television uses a modulation type called *8-level vestigial sideband* (8-VSB), which is basically 8-level amplitude-shift keying with the lower sideband filtered out. The eight levels result in three bits per symbol, resulting in approximately 32 Mbits/second raw data rate with a 10.76 Mbaud symbol rate. The net data rate with error correction and other overhead is 19.39 Mbits/second. The DTV signal fits into the same 6 MHz channel as analog TV.

8.4.2 Facsimile

Amateur *facsimile* (fax) is a method for sending high-resolution still pictures by radio. Standard 120-line-per-minute fax uses a frequency-modulated tone between 1500 Hz (black) and 2300 Hz (white), which may be sent and received as an audio signal via any voice transmitter and receiver. Commercially, fax has been used for many years to send newspaper photos over the landline and to receive weather images from satellites.

The resolution of typical fax images greatly exceeds what can be obtained using SSTV or even conventional television. Typical images are made up of 480 to 1600 scanning lines, compared to the 240 lines in a typical SSTV image. This high resolution is achieved by slowing down the rate at which the lines are transmitted, resulting in image transmission times typically in the range of 4 to 13 minutes. For color, transmission times are longer since the image is sent three times to include the red, green and blue components.

Besides the increased resolution, the main difference with SSTV is that many fax systems use no sync pulses. Instead, they rely on the extreme accuracy of external time bases at both the transmitting and receiving ends of the circuit to maintain precise synchronization.

Modern personal computers have virtually eliminated bulky mechanical fax machines from most amateur installations. Now the incoming image can be stored in computer memory and viewed on a high-resolution computer graphics display. The use of a color display system makes it entirely practical to transmit color fax images when band conditions permit the increased transmission time.

8.4.3 Slow-Scan Television

Despite its name, so-called *slow-scan television* (SSTV) is actually a method for sending still images, like facsimile. The original monochrome analog SSTV format illustrated in **Fig 8.31** 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. That has the disadvantage that you have to 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) and thus spoil 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 in order 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.

8.5 Modulation 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.

8.5.1 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 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 **Fig 8.32**. The third-order products that occur near the desired signal are at 2F1 – F2 and 2F2 – F1. The fifth-order products occur at 3F1 – 2F2 and 3F2 – 2F1. If, for example, the two tones are spaced 1 kHz apart, then the two third-order products show up at 1 kHz above the high tone and 1 kHz below the low tone. The fifth-order products are at 2 kHz above and below the high and low tones respectively.

8.5.2 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

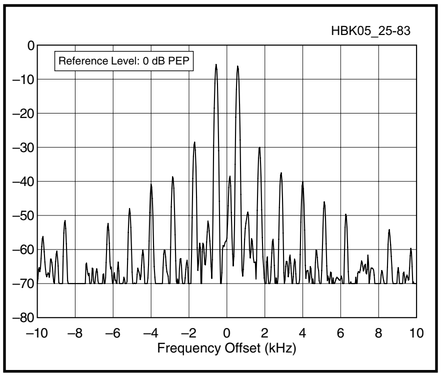

Fig 8.32 — Intermodulation products from an SSB transmitter. The two signals at the center are the desired frequencies. The frequencies of the third-order products are separated from the frequencies of the desired tones by the tone spacing. The fifth-order frequencies are separated by two times the tone spacing, and so on for higher-order products.

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.

8.5.3 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

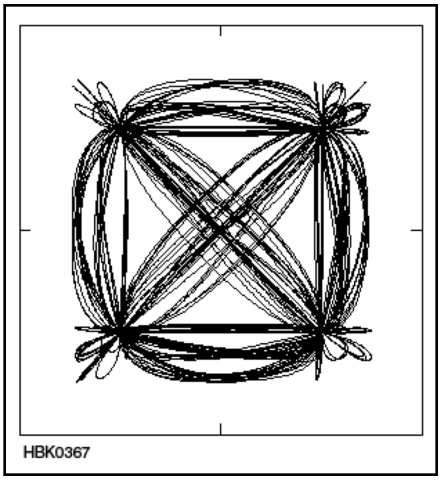

Fig 8.33 — Simulated I/Q constellation display of the trajectory of a QPSK signal over an extended period. A 0.5-alpha raised-cosine filter was used. Because it is a Nyquist filter, the trajectories pass exactly through each symbol location.

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 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. **Fig 8.33** 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 **Fig. 8.34**. 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."

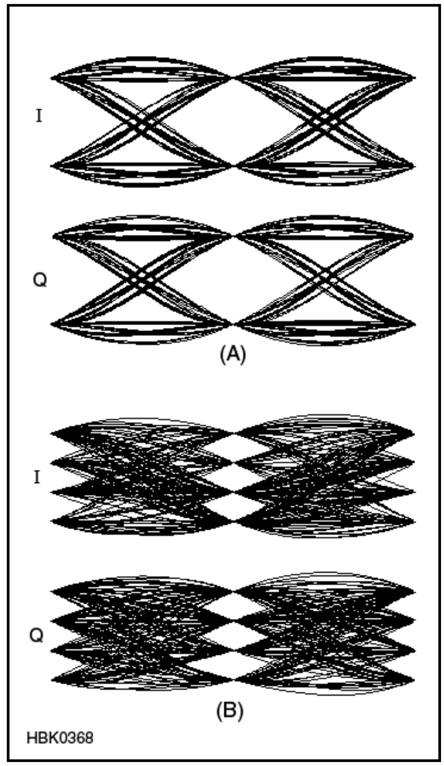

Fig 8.34 — Eye diagrams of the I and Q outputs of a QPSK generator (A) and a 16-QAM generator (B). The "eye" is the empty area between the symbol decision points, visible as the points where all the symbol trajectories come together. The bigger the "eye" the easier the signal is to decode.

8.6 Modulation Glossary

GENERAL

AM — Amplitude modulation.

Angle modulation — Modulation of an RF carrier by varying the phase or frequency.

Emission designator — Official ITU code to specify the bandwidth and modulation type of a radio transmission.

Frequency diversity — The use of a wideband signal to compensate for selective fading. While one band of frequencies is faded, the data can be reconstructed from other frequencies that are not faded.

FM — Frequency modulation.

IMD — Intermodulation distortion. Unwanted frequencies that occur at the sum and difference of the desired frequencies and their harmonics.

ITU — International Telecommunications Union. An agency of the United Nations that coordinates and recommends technical standards for electronic communications.

Modulation — The periodic alteration of some parameter of a carrier wave in order to transmit information.

PM — Phase modulation.

Protocol — A formal set of rules and procedures for the exchange of information within a communications network.

Selective fading — A propagation phenomenon in which closely-spaced frequencies experience markedly different fading at the same time.

Telemetry — The use of telecommunication for automatically indicating or recording measurements at a distance from the measuring instrument.

Telephony — A form of telecommunication primarily intended for the exchange of information in the form of speech.

Telegraphy — A form of telecommunication in which the transmitted information is intended to be recorded on arrival as a graphic document; the transmitted information may sometimes be presented in an alternative form or may be stored for subsequent use.

Television — A form of telecommunication for the transmission of transient images of fixed or moving objects.

ANALOG MODULATION (INCLUDING IMAGE)

ATV — Amateur fast-scan television.

BFO — Beat frequency oscillator. In an SSB or DSBSC receiver, the intermediate-frequency oscillator in the receiver that re-inserts the suppressed carrier.

Carson's rule — A rule of thumb to calculate FM bandwidth that says that 98% of the energy is typically contained within a bandwidth equal to two times the sum of the peak frequency deviation and the highest modulating frequency.

Capture effect — The tendency of the strongest signal to suppress other signals in an FM detector, which improves the signal-to-noise and signal-to-interference ratio.

Deviation ratio — The ratio of the maximum permitted peak deviation of an angle-modulated signal to the maximum permitted modulating frequency.

DSBSC — Double sideband, suppressed carrier. An AM signal in which the carrier has been removed but both sidebands remain.

Fax — Facsimile. The sending of still images by wire or by radio.

Frame — In television, one complete scanned image. On systems with interlaced scanning, there are two vertical scans per frame.

Frequency deviation — The amount the RF frequency of an angle-modulated signal deviates from the center (carrier) frequency in response to the modulating signal. The term is often understood to mean the maximum deviation available in a given system.

Limiter — A high-gain amplifier in an FM receiver that limits the peak amplitude of the signal in order to eliminate any AM component.

LSB — Lower sideband. An SSB signal with the upper sideband removed.

Martin — A series of analog SSTV formats, especially popular in Europe.

Modulation index (AM) — The ratio of the peak value of the modulation of an AM signal to the value that just causes the modulation envelope of the RF signal to reach zero on negative peaks and twice the average value on positive peaks.

Modulation index (FM) — The ratio of the peak deviation of an angle-modulated signal to the highest modulating frequency.

Narrowband FM — An FM signal with a modulation index less than or equal to 1.0.

NTSC — National Television System Committee. The analog television standard used in the US, Japan and several other countries.

PAL — Phase alteration line. The analog television standard used in many parts of Europe.

Phasing method — A method of generating an SSBSC signal that does not require a filter to remove the unwanted sideband.

Pixel — Picture element. The dots that make up images on a computer's monitor.

Product detector — A detector that multiplies a BFO signal with the received signal, typically SSB or DSBSC-modulated.

RGB — Red, green and blue. The three primary colors required to transmit a full-color image in many television and facsimile systems.

Scottie — A series of analog SSTV formats, especially popular in the US.

SSB — Single sideband. An AM signal in which one sideband has been removed. The term is usually understood to mean SSBSC.

SSBSC — Single sideband, suppressed carrier. An AM signal in which one sideband and the carrier have been removed.

SSTV — Slow scan TV, a system for sending and receiving still images.

Sync — Modulation pulses used in ATV and SSTV to synchronize the horizontal and/or vertical scanning.

Synchronous detector — A type of AM detector in which the carrier is regenerated in the receiver.

Two-tone test — A procedure for testing the IMD of an SSB transmitter. Two equal-amplitude tones are fed to the microphone input and the transmitter RF output is examined with a spectrum analyzer to determine the amplitudes of the IMD products.

USB — Upper sideband. An SSB signal with the lower sideband removed.

Vestigial sideband — The filtering of all but the bottom MHz or so of the lower sideband of an ATV or NTSC television signal.

VIS — Vertical interval signaling. Digital encoding of the transmission mode during the vertical sync interval of an SSTV frame.

Wideband FM — An FM signal with a modulation index greater than 1.0.

DIGITAL MODULATION

AFSK — Audio FSK. The use of an SSB transceiver to transmit and receive FSK using an audio-frequency modem.

ASK — Amplitude-shift keying. Digital amplitude modulation in which the amplitude depends on the modulating code.

Baud rate — The rate at which a digital signal transitions between symbol states. Symbol rate.

Bauds — Symbols per second. Unit of baud rate or symbol rate.

Bit rate — The total number of physically transferred bits per second over a communication link. Bit rate can be used interchangeably with **baud rate** only when each modulation transition carries exactly one bit of data.

BPSK — Binary PSK. PSK with only two possible states. The term is usually understood to mean a non-constant-envelope signal in which the two states differ by 180°.

Chirp — Incidental frequency modulation of a carrier as a result of oscillator instability during keying.

CW — Continuous wave. The term used for on-off keying using Morse code.

Constellation diagram — A diagram showing the constellation of possible symbol locations on a polar plot of modulation amplitude and phase.

DBPSK — Differential BPSK.

Decision point — The point, typically in the center of a symbol time, at which the receiver decides which symbol is being sent.

Differential modulation — A modulation technique that encodes the information in the difference between subsequent symbols, rather than in the symbols themselves.

Equalization — Correction for variations in amplitude and/or phase versus frequency across a communications channel.

DQPSK — Differential QPSK.

Eye diagram — An oscilloscope measurement of a digital modulating signal with the horizontal sweep synchronized to the symbol times. With Nyquist filtering, there should be a clear separation, or “eye”, in the trajectories at the symbol decision points.

EVM — Error vector magnitude. A measure of the RMS error in the symbol locations at the symbol decision points in the constellation plot of a digital signal.

FFT — Fast Fourier transform. The Fourier transform is a mathematical function that calculates the frequency spectrum of a signal. The FFT is a software algorithm that does the calculations very efficiently.

FSK — Frequency-shift keying. A form of digital frequency modulation in which the frequency deviation depends on the modulating data.

GMSK — Gaussian MSK. MSK with a Gaussian modulation filter.

I/Q modulation — Quadrature modulation implemented with an I/Q modulator, one that uses in-phase (I) and quadrature (Q) modulating signals to generate the zero-degree and 90° components of the RF signal.

ISI — Inter-symbol interference. Interference of a signal with itself, caused when energy from one symbol is delayed long enough to interfere with a subsequent symbol.

MFSK — Multi-level FSK. FSK with more than two states represented by different frequency deviations.

Modem — Modulator/demodulator. A device that generates and demodulates digital modulation signals, usually at audio frequencies. It connects between the data terminal (usually a computer) and the radio.

MSK — Minimum-shift keying. A form of FSK with a frequency shift equal to one-half the symbol rate.

Nyquist criterion — A principle that states that the sampling frequency must be greater than twice the highest frequency in the sampled signal.

Nyquist frequency — One-half the sampling frequency.

OFDM — Orthogonal frequency-division multiplexing. A transmission mode that uses multiple carriers, spaced such that modulation on each carrier is orthogonal with the others.

OOK — On-off keying. A type of ASK with only two states, on and off.

OQPSK — Offset QPSK. By offsetting in time the symbol transitions of the I and Q channels, symbol trajectories through the origin are eliminated.

Orthogonal — Refers to streams of data that are uncorrelated with each other such that there is no mutual interference.

PAM — Pulse amplitude modulation. A type of pulse modulation where the modulating signal is encoded in the pulse amplitude.

π/4 DQPSK — PI-over-four differential QPSK. A form of differential 8PSK in which the only allowed transitions are ±45 and ±135°, resulting in four allowed states per symbol.

PPM — Pulse position modulation. A type of pulse modulation where the modulating signal is encoded in the pulse position.

PSK — Phase-shift keying. A form of digital phase modulation in which the phase of the RF signal depends on the modulating code. The term often is understood to refer to BPSK.

Pulse modulation — The modulating signal is sampled at regular intervals to generate a series of modulation symbols in the form of pulses.

PWM — Pulse width modulation. A type of pulse modulation where the modulating signal is encoded in the pulse width.

QAM — Quadrature amplitude modulation. A digital modulation type in which both amplitude and phase are varied. The number that precedes it, for example 64QAM, is the number of different possible states of the amplitude and phase.

QPSK — PSK with four possible states. The term is usually understood to be equivalent to four-level QAM, in which the four states have the same amplitude and differ in phase by 90°.

Quadrature modulation — Refers to modulation using two RF carriers in phase quadrature, that is, 90° out of phase.

Symbol rate — The rate at which a digital signal transitions between different states. Baud rate.

Varicode — A coding method in which the length of each character code depends on its frequency of occurrence. It is used to optimize the ratio of characters per second to baud rate.

SPREAD SPECTRUM

CDMA — Code-division multiple access. A method of allowing several stations to use the same frequency band simultaneously by assigning each station a different orthogonal spreading code.

Chip — The shortest-duration modulation state in a SS system.

De-spreading — Conversion of a SS signal back to its narrowband equivalent by convoluting it with the spreading code.

DSSS — Direct-sequence SS. Spreading of a signal by multiplying the data stream by a higher-rate pseudo-random digital sequence.

FHSS — Frequency-hopping SS. Spreading of a signal by means of pseudo-random frequency-hopping of the unspread signal.

Processing gain — The increase in signal-to-noise ratio that occurs when an SS signal is de-spread.

SS — Spread spectrum. A system that intentionally increases the bandwidth of a digital signal by means of a special spreading code.

FILTERING AND BANDWIDTH

ACP — Adjacent channel power. The amount of transmitted power that falls into a communications channel immediately adjacent to the desired one.

Alpha — A design parameter for a Nyquist or root Nyquist filter. The smaller the alpha, the sharper the passband-to-stopband transition at the cutoff frequency.

Cutoff frequency — The frequency at which a filter response changes from the passband to the stopband.

Gaussian filter — A modulation filter with a Gaussian frequency response. The transition between passband and stopband is more gradual than with most Nyquist filters.

Key clicks — Out-of-channel interference from an OOK signal caused by too-sharp transitions between the on and off states.

Nyquist filter — A filter that causes no inter-symbol interference (ISI). It is so-called because the cutoff frequency is at one-half the symbol rate, the Nyquist frequency.

Nyquist criterion — The rule that states that the sample rate must be greater than twice the highest frequency to be sampled.

Occupied bandwidth — The bandwidth within which a specified percentage of the total power occurs, typically 99%.

Raised-cosine filter — A type of Nyquist filter whose passband-to-stopband transition region has the shape of the first half-cycle of a cosine raised so that the negative peak is at zero.

Root Nyquist filter — A filter that, when cascaded with another identical filter, forms a Nyquist filter. The frequency response is the square root of that of the Nyquist filter.

Root raised-cosine filter — A root Nyquist filter with a frequency response that is the square root of a raised-cosine filter.

Shannon-Hartley theorem — A formula that predicts the maximum channel capacity that is theoretically possible over a channel of given bandwidth and signal-to-noise ratio.

Sinc function — The spectrum of a pulse or random series of pulses, equal to sin(x)/x.

Splatter — Out-of-channel interference from an amplitude-modulated signal such as SSB or QPSK caused by distortion, typically in the power amplifier stages.

8.7 References and Further Reading

Agilent Technologies, "Digital Modulation in Communications Systems — An Introduction" Application Note 1298. **cp.literature.agilent.com/litweb/pdf/5965-7160E.pdf**

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

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

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**

McConnell, Bodson, and Urban, *FAX: Facsimile Technology and Systems* (Artech House 1999)

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, **www.baycom.org/~tom/ham/dcc2001/datv.pdf**

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

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*, chapters 23, 24 and 25, 9th Ed. (Newnes, 2001)