CHAPTER

22 Digital modulation techniques

Digital modulation is both the newest and the oldest radio technique. Morse code transmissions were strictly binary, with "key down" and "key up" equivalent to multiplying the carrier by one or zero. Many modern systems also use binary keying, but the zero state is usually signaled by reversing the polarity of the signal (*binary phase-shift keying*, BPSK) or by shifting the frequency (*binary frequency-shift keying*, BFSK). This improves the probability of distinguishing zeros from ones in the presence of noise.

In this chapter we look first at some of the methods used for binary and "*m*-ary" modulation. We then see how specially shaped pulses can be used with these methods in order to avoid intersymbol interference when the pulses, dispersed in time, partially overlap at the receiver. The "8-VSB" system used for digital television in the U.S. (see Chapter 19) provides an example of pulse amplitude modulation (PAM). Finally, we discuss two newer digital modulation systems: multicarrier and spread spectrum. A glossary is provided at the end of this chapter, listing the many common abbreviations used (BPSK, BFSK, 8-VSB, PAM, etc.).

22.1 Digital modulators

Digital modulation differs from analog modulation in that only a discrete set of states (in the space of amplitudes, phases, and frequencies) is used, and that the time devoted to any state is always an integral multiple of a basic time-step. The state during this time period constitutes a transmitted "symbol," and the symbol rate is one of the parameters defining a modulation system. Figure 22.1 shows the simple and widely used binary phase shift keying (BPSK) modulation technique. Instead of turning off the carrier to indicate a zero, the carrier phase is flipped 180°. This is equivalent to multiplying the carrier signal by minus one, as shown in the figure. Binary values of 0 or 1 are handled by sending −1 or +1, respectively, to the multiplier. As shown in Chapter 8, the multiplier (mixer) produces a double-sideband suppressed-carrier RF signal.

Figure 22.1. BPSK modulator.

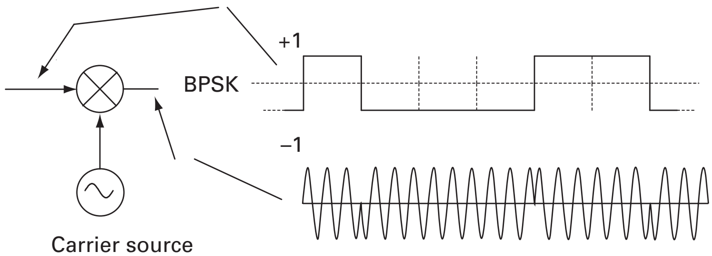
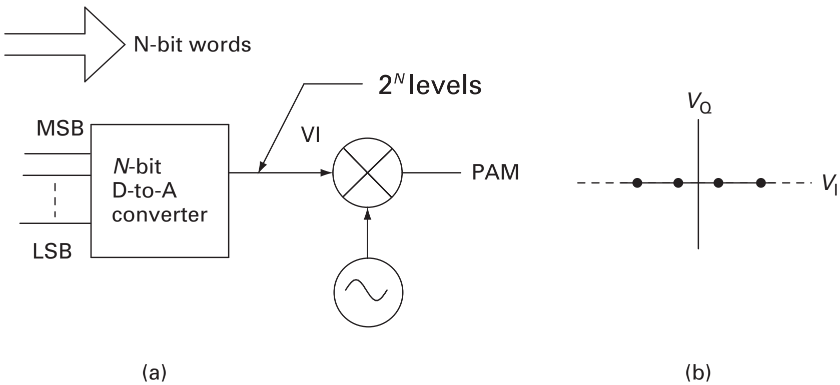


Figure 22.2. *N*-level pulse amplitude modulation (PAM).



Since the BPSK signal has a nominally constant envelope, the receiver needs a phase-sensitive or "coherent" detector. In contrast, on–off keying (OOK) or frequency shift keying (FSK) can be detected incoherently. The advantage in using coherent systems is one of sensitivity; for a given situation, coherent detection provides a higher signal-to-noise ratio at the detector output. The one-bit BPSK system can be generalized to more than two amplitude levels. We saw in Chapter 21 that the U.S. ATSC digital television system uses eight (2^3) modulation levels in order to send three bits per symbol. Figure 22.2(a) shows an N-bit (2^N-level) pulse amplitude (PAM) modulator. The "constellation" shown in (b) shows the signed amplitudes as discrete points on a linear scale. Note that, because the negative amplitudes are produced by changing the phase 180°, PAM requires coherent demodulation.

We saw in Chapter 8 that quadrature AM (QAM) systems can transmit independent information, modulated on a Q-carrier which is in quadrature, i.e., 90° out of phase, with respect to the I ("in-phase") carrier. This "phase multiplexing" method is commonly used in digital communications. Figure 22.3 shows a "4-QAM" modulator.

The binary (two-level) I and Q signals result in an IQ constellation with four points. In this case, the points on the constellation have four possible phase values, but the amplitudes are equal, since each of the four points lies at an equal distance from the origin. You can see why this scheme is also known as QPSK (quadrature phase shift keying) and 4-PSK (four-state phase-shift keying). Figure 22.4 shows a QAM modulator with four-level modulation on both I and Q channels, producing a total of 16 states.

The constellation of output signals for this 16-QAM modulator is shown in (b). The 16 binary numbers, 0000 through 1111, could be assigned arbitrarily to the constellation points, but the assignments shown in the figure have the property that nearest neighbors differ by only one bit. As a result, in an environment with only modest noise, errors in transmission will nearly all be single-bit errors. Note that next-nearest neighboring points on the constellation differ by

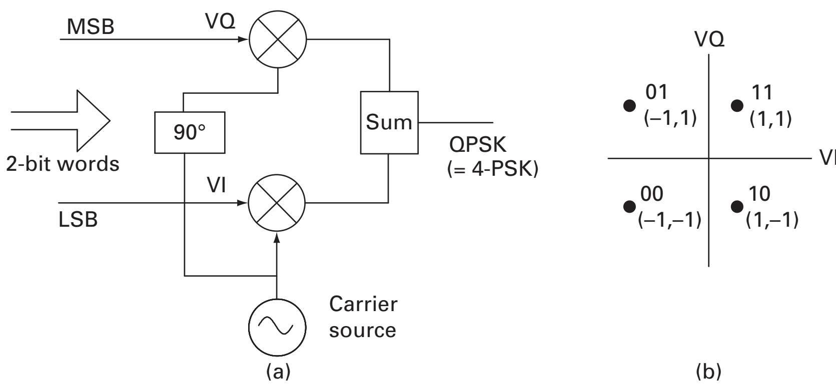


Figure 22.3. (a) A 4-QAM (QPSK) digital modulator; (b) QPSK constellation.

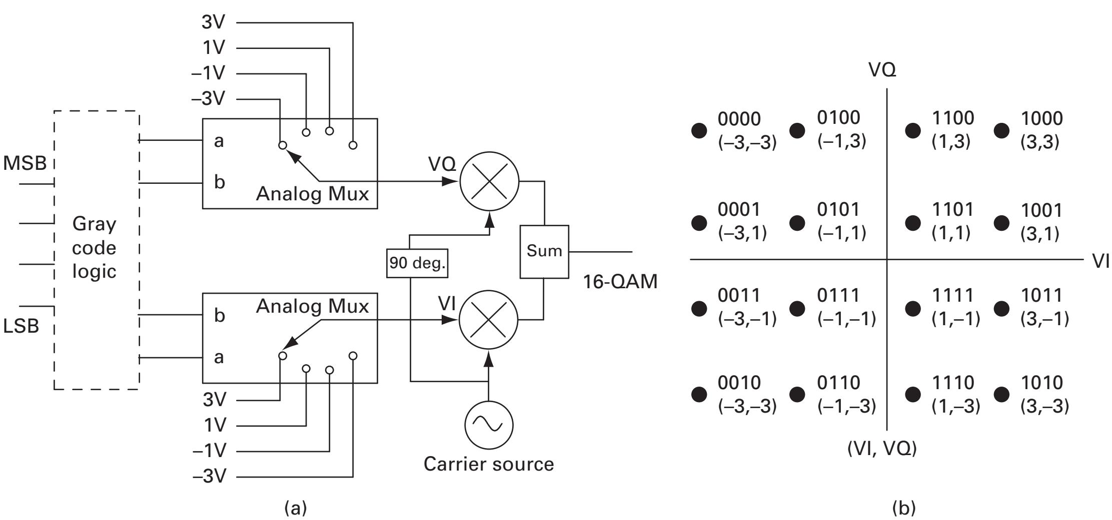


Figure 22.4. A 16-QAM digital modulator.

just two bits. QAM systems can also have constellations in which the points are arranged in polar fashion; a constellation with eight points arrayed on a circle in the *I-Q* plane is known as 8-PSK. Note that BPSK could just as well be called 2-PSK.

22.2 Pulse shaping

The abrupt reversals in the BPSK waveform of Figure 22.1 result in a signal with an objectionably large bandwidth. The signal is a continuous stream of rectangular RF pulses whose polarity is determined by the data. If the data bits are random, with equal probabilities of being a one or a zero, these rectangular pulses produce a power spectrum proportional to $[\sin(\pi f T)/(\pi f T)]^2$, where T is the duration of a symbol, the "baud" length. This spectrum is plotted in Figure 22.5.[1] On this graph, frequency is in units of $1/T$, and the zero is at the nominal carrier frequency.

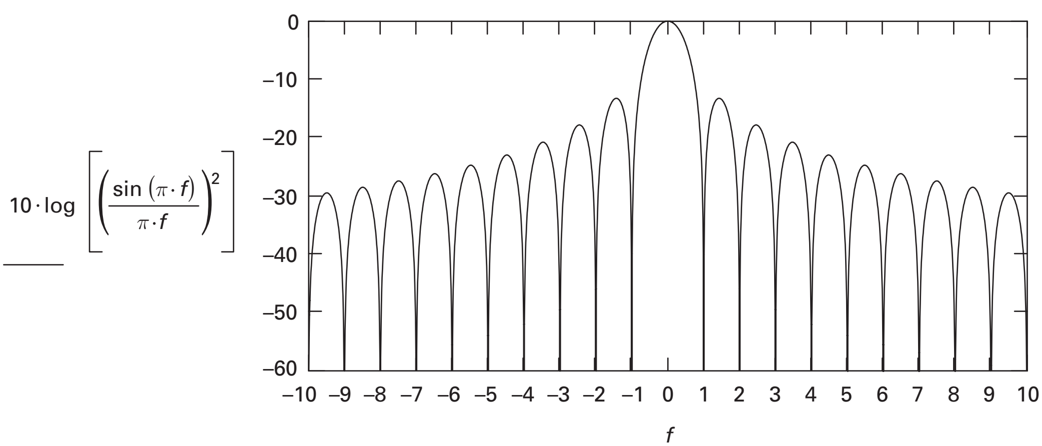


Figure 22.5. Power spectrum of a BPSK signal with random data and rectangular pulses. Frequency is in units of $1/T$, and the zero is at the nominal carrier frequency.

This wide spectrum applies to all the digital modulators discussed above, and is clearly unacceptable if other users are to use nearby frequency channels. Bandpass filtering at the transmitter, or otherwise transmitting a pulse shape whose spectrum is band-limited, solves this problem, but there is a complication – limiting the bandwith causes a widening of each pulse in the sequence so that, at the receiver, a currently arriving pulse will be contaminated by remnants or precursors of nearby pulses. This effect is known as *intersymbol interference* (ISI). However, there exists a class of pulse shapes, *Nyquist pulses*, for which (a) the bandwidth is limited and (b) there is no ISI, despite the fact that the pulses do spread into each other. One such pulse shape is $V(t) = \sin(\pi t/T)/(\pi t/T) = \mathrm{sinc}(t/T)$, the normalized sinc function. This pulse shape is the impulse response of a

[1] The power spectrum is calculated from the signal's average autocorrelation function (see Chapter 27).

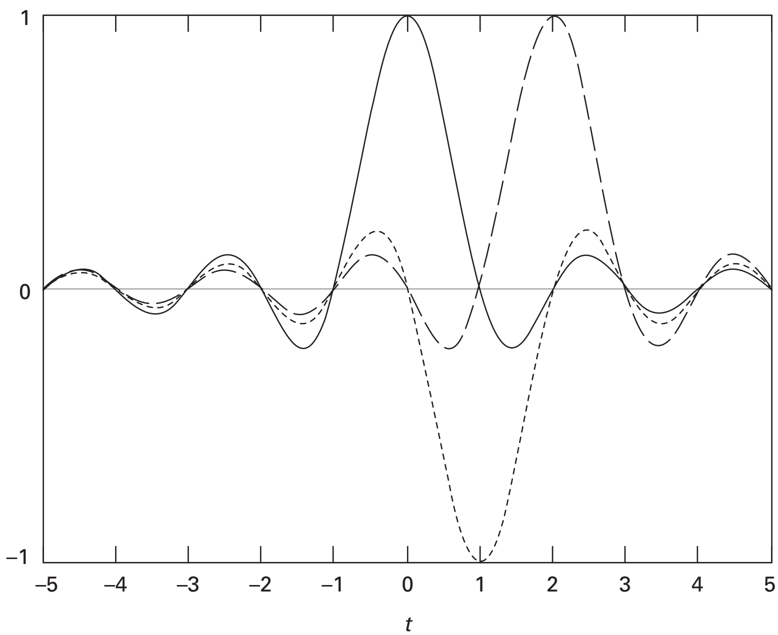


Figure 22.6. Successive $\sin(\pi t)/(\pi t)$ signaling pulses.

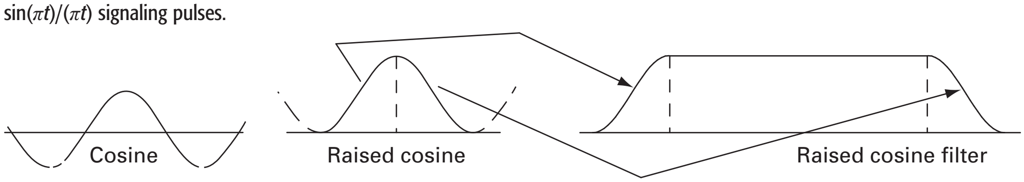


Figure 22.7. Construction of the raised cosine filter; $H(\omega)$ is the voltage response.

rectangular lowpass filter with a cutoff frequency of 1/(2T)Hz. This filter confines all the power to a bandwidth of 1/(2T). Successive output pulses for a 101 data string are plotted in Figure 22.6. The third pulse reaches a maximum at $t = 2$, the instant at which we sample it. Neither of the two previous pulses has yet died out but, at $t = 2$, both are crossing zero and thus contribute no interference to the sampled value of the desired pulse.

The same holds true for the contributions from all the other pulses, of any amplitude and either polarity, allowing us to transmit data pulses at a rate 1/T without intersymbol interference.

Pulses with this property are said to be *orthogonal* to one another. While they do invade each other's space, they do not create intersymbol interference. The $\sin(\pi t/T)/(\pi t/T)$ pulse is an extreme case; bandwidth is minimized, but the pulse is widely spread out in time. In practice, other orthogonal pulse shapes are chosen, which have somewhat greater bandwidth but are only modestly spread in time. One frequently used pulse shape is the impulse response of the *raised-cosine* filter. Figure 22.7 shows how this filter shape is produced by replacing the vertical sides of the rectangular filter with smooth roll-offs formed from quarter cycles of the cosine function, raised by 1/2 in the y-direction.

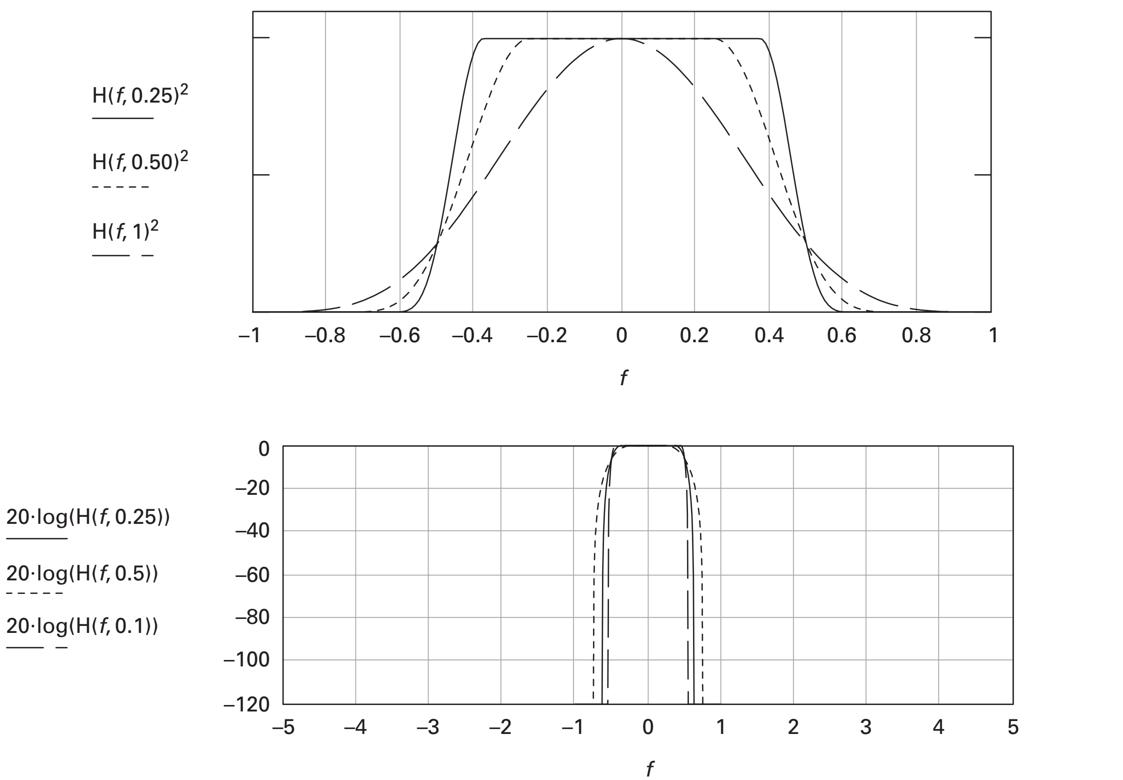


Figure 22.8. Power response, |$H(\omega)$|2, of raised cosine filter for $\alpha = 0.25$, 0.5, and 1.0.

We can express this filter shape as

$$H(f,\alpha) = \begin{cases} T: & 2|f|T \leq 1-\alpha \\ \dfrac{T}{2}\left[1+\cos\left(\dfrac{\pi T}{\alpha}\left(|f| - \dfrac{1-\alpha}{2T}\right)\right)\right]: & 1-\alpha \leq 2|f|T \leq 1+\alpha \\ 0: & 2|f|T 1+\alpha, \end{cases} \tag{22.1}$$

where α, a roll-off factor between zero and one, determines the width of the raised cosine edge sections. Note that the $\alpha = 0$ case is just the rectangular filter discussed above. This filter's power response, $|H(f,\alpha)|^2$, for $\alpha = 0.25$, $\alpha = 0.5$ and $\alpha = 1$ (pure raised cosine with no flat center section) is plotted in Figure 22.8(a) with a linear power scale and Figure 22.8(b) with a dB power scale.

The impulse response of this filter, i.e., the Fourier transform of $H(f,\alpha)$, often called by extension a "raised-cosine pulse," is given by

$$h(t,\alpha) = \frac{\sin(\pi t/T)}{\pi t/T} \cdot \frac{\cos(\pi\alpha t/T)}{1-(2\alpha t/T)^2}. \tag{22.2}$$

The zeros of the first term, $\sin(\pi t/T)$ /$(\pi t/T)$, give this pulse shape the same orthogonality property as $\sin(\pi t/T)$ /$(\pi t/T)$ alone. Figure 22.9 shows three

https://doi.org/10.1017/CBO9780511626951.023 Published online by Cambridge University Press

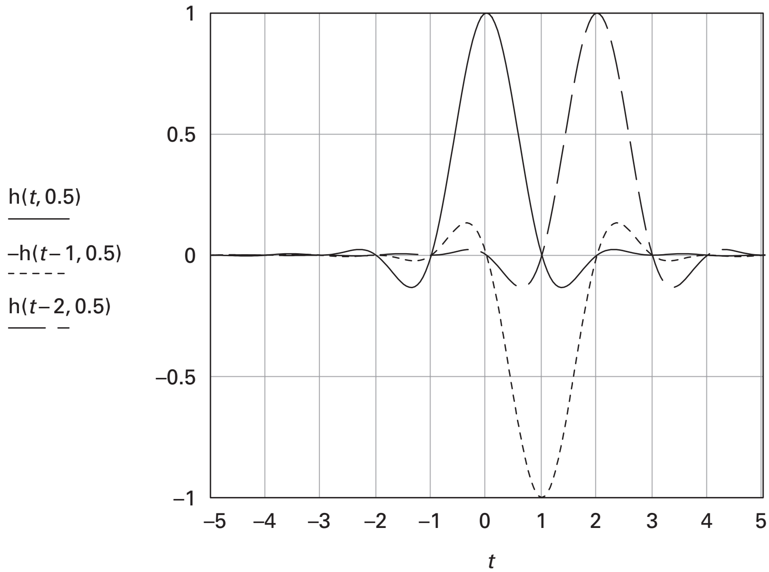


Figure 22.9. Successive raised-cosine signaling pulses.

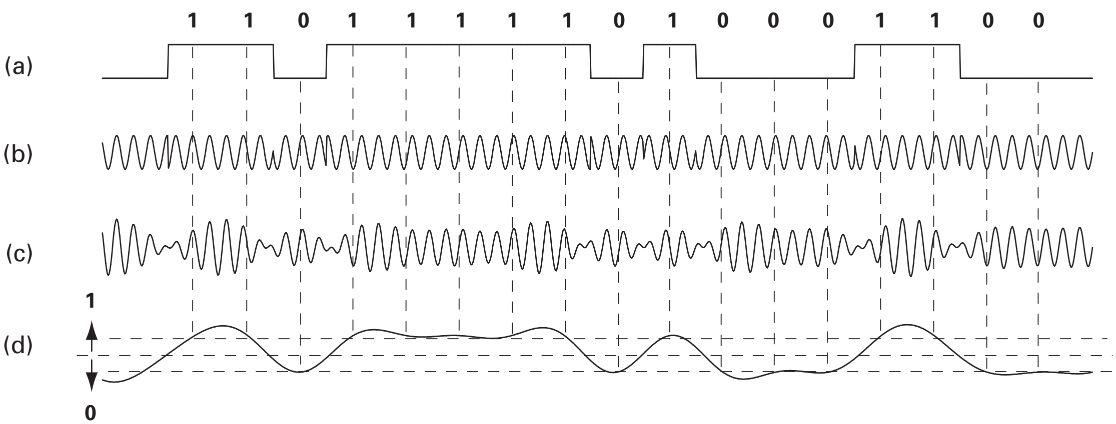


Figure 22.10. BPSK waveforms: (a) input data; (b) bi-phase modulated carrier; (c) modulated signal after raised-cosine filtering ($\alpha = 0.22$); (d) coherently detected signal.

consecutive $\alpha = 0.5$ raised cosine pulses. Note that these pulses, like the sin $(\pi t/T)/(\pi t/T)$ pulses of Figure 22.6, produce no intersymbol interference, but unlike those pulses, these are only modestly spread out in time.

Despite the bandwidth advantage, there are two penalties associated with choosing a low value for α. First, the impulse response duration of the filter is longer, requiring more computations per symbol when the filter is implemented digitally. Second, the increased ringing means that the received signal must be sampled with higher timing accuracy to avoid ISI.

The successive processes of BPSK modulation, pulse shaping, and detection are shown in Figure 22.10. The data string to be transmitted is shown in (a). In (b), the sinusoidal carrier has been multiplied by the data, producing phase

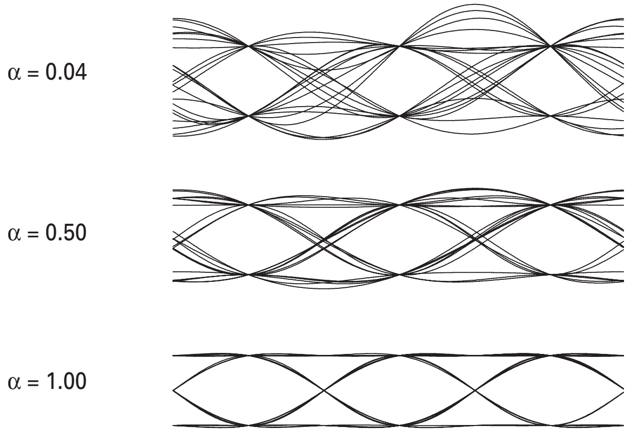


Figure 22.11. BPSK eye diagram.

reversals. In (c), the modulated signal has been band-limited by a raised-cosine filter of $\alpha = 0.22$. In this simulation the modulation rate is high, causing obvious variations in the envelope of the filtered signal. Finally, in (d), the filtered signal has been coherently detected, i.e., multiplied by a replica of the original sine wave carrier and then lowpass filtered to remove the components at twice the carrier frequency. Note that, at the sampling instants, the sampled values are exactly the same 0's and 1's of the original data, i.e., there is no intersymbol interference. Note: for clarity in alignment with the original data, no filter delay has been added to the waveforms in (c) and (d), so they appear noncausal.

If the detected data is displayed versus time for many sweeps on a storage oscilloscope that is synchronized to the data, we obtain the well-known "eye" diagram, shown in Figure 22.11. You can see that, at the sampling instants, the value of the waveform is either zero or one; there is no intersymbol interference. Eye diagrams are presented for three values of α. You can see that, for smaller values of α (narrower bandwidths), the sampling instants must be increasingly accurate to obtain only the values zero and one, whereas larger values of α make the system more tolerant of sampling "jitter."

22.3 Root raised-cosine filter

Any receiver must contain a bandpass filter to reject unwanted signals (i.e., signals as well as noise on nearby frequencies). If the receiver filter has a rectangular shape whose bandwidth is equal to or greater than the signal bandwidth, Nyquist pulses will pass through undistorted, maintaining their orthogonality, i.e., their freedom from intersymbol interference. However, a basic result from signal processing theory (see Chapter 23) is that the overall signal-to-noise ratio of a communication link is maximized when the receiver uses a *matched* *filter*, which is a filter whose impulse response is proportional to the time-reversed waveform of the incoming signal pulse. Let us apply this consideration to the two pulse shapes discussed above. For the $\sin(\pi t/T)/(\pi t/T)$ signaling pulse, the matched filter is just the rectangular filter which might have been used to produce the pulse. The cascade of two rectangular filters is equivalent to a single rectangular filter, and the pulse shape is preserved. For a raised-cosine signaling pulse, however, the matched filter is the raised-cosine filter of Equation 22.1. If we put this filter at the input of the receiver, the overall response will be the *square* of the raised-cosine shape. The signaling pulses will take on the shape of a raised-cosine squared filter – a shape that is not free of intersymbol interference. To avoid this, it is common to transmit "root raised-cosine" signaling pulses, i.e., pulses whose shape is the impulse response of a filter whose amplitude response is the square root of the raised-cosine filter. A root raised-cosine filter is used at the input of the receiver as a matched filter. The pulses exiting the receiver filter will therefore have the "raised-cosine pulse" shape shown in Figure 22.9, and there will be no intersymbol interference. The frequency response of the root raised-cosine filter is given by Figure 22.8(b), if you relabel the y-scale to go from 0 to – 60 dB instead of 0 to −120 dB.

Let us summarize this section. So-called Nyquist pulses can be transmitted without intersymbol interference up to a maximum rate of B pulses per second using double-sideband AM modulation within an RF bandwidth of at least B Hz. A second carrier, at the same frequency, but phase shifted by 90°, can be used simultaneously to transmit independent pulses. When the pulses are identical except for 0° or 180° phase changes, the system is called BPSK (one carrier only) or QPSK (pulses on both I and Q carriers). Synchronous detection, i.e., phase-sensitive detection, makes this 0° and 180° phase modulation equivalent to amplitude modulation in which the sign of the signal is preserved. If the I and Q components of a QPSK signal are each allowed to take on four values, the modulation forms a 4×4 constellation and is referred to as 16-QAM. (It would be convenient if the term "AM" always referred to this kind of sign and magnitude modulation. BPSK could be called "sign modulation." Traditional AM could be called "magnitude modulation".)

Note that in the above analysis we have not needed to bother with including a phase factor in the filter amplitude response functions. These particular filter functions are purely real, except for this omitted factor, $e^{-j\omega T_{\text{delay}}}$, where T_{delay} is a frequency-independent time delay. Neglecting this factor is equivalent to setting the delay equal to zero, which causes the impulse responses to begin at non-causal negative times.

22.4 8-VSB and GMSK modulation

The U.S. digital television system, discussed in Chapter 19, modulates an I-carrier with three-bit digital data in the form of eight evenly spaced amplitude levels whose nominal values are 7, 5, 3, 1, −1, −3, −5, and −7. A bias of 1.5 is added to these levels to create a small pilot carrier (a phase reference for the receiver) resulting in five transmitted levels that are positive and three that are negative. This is a PAM system (see Figure 22.2). The resulting double-sideband suppressed carrier signal is bandpass filtered to eliminate all but a vestige of the lower sideband, reducing the bandwidth by factor of almost 2. This modulation is called 8-VSB (eight levels, vestigial sideband). Because the signal had no Q-component before filtering, the lower sideband was the mirror image of the upper sideband and contained redundant information. With a general QAM signal (using both I and Q), the lower and upper sidebands are different and both are needed. But, as a result of filtering away the lower sideband, a transmitted VSB signal acquires a Q component, making the "Q-space" unavailable for the transmission of a second television signal.

The modulation method used in GSM cellular telephones is known as *GMSK* (*Gaussian minimum shift keying*), and works as follows. A standard QPSK modulator is used, summing the signal from an I mixer with the signal from a Q mixer. But the input signals to these mixers are such as to cause the sum signal to rotate smoothly from one point on the constellation to an adjacent point during the baud time. The rotation is 90°, clockwise or counterclockwise, depending on the binary data bit. The rotation takes place at a uniform rate. Since frequency is the time derivative of phase, this modulation is actually BFSK. In addition, smoothing, derived from a Gaussian filter shape, is applied to the phase command, so that the time derivative of the frequency is continuous during the transitions between the two nominal frequencies. This system is designed to minimize bandwidth, allowing closer channel spacing to accommodate more users. Bandwidth reduction comes from using both I and Q, from changing the phase slowly and smoothly, and from allowing some intersymbol interference (which is almost all from the immediately preceding baud and can be compensated for by the receiver).

22.5 Demodulation

Synchronous demodulation is required for the BPSK and QAM systems discussed above. This requires that the receiver be able to synthesize a replica of the carrier. This carrier recovery, which takes place in the presence of data, can be done, for example, using a squaring PLL (see Chapter 12). A standard PLL suffices if a pilot carrier is transmitted. The receiver contains a matched filter, followed by a mixer whose L.O. signal is the recovered carrier. The baseband output of the mixer is sampled at the signal baud rate. The signal may include a periodic synchronizing pattern to assist the receiver in establishing the correct phase for the sampler. A digital matched filter is often used. When realized as a weighted sum of signals from a tapped delay line (a transversal FIR filter), such a filter illustrates how a matched filter is equivalent to a correlator, finding the best alignment (match) of the signal to a replica of the signaling pulse.

Sometimes a free-running L.O. will be close enough to the carrier frequency that the dephasing over a data word or packet is significantly less than $\pi/2$. In this situation, differential modulation can be used whereby, for example, a "one" is signaled by flipping the phase of a BPSK signal, while a "zero" is signaled by leaving the phase at its prior value. If both I and Q channels are implemented, the phase changes can be reliably sensed and the data can be reconstructed. In a QPSK system, both I and Q channels are always implemented and the receiver can sense differential changes between points on the modulation constellation. Differential modulation, however, results in a higher error rate than coherent demodulation.

22.6 Orthogonal frequency-division multiplexing – OFDM

OFDM is a relatively new modulation technique in which the digital data stream to be transmitted is divided into N separate parallel data substreams, each having a data rate that is $1/N$ times the data rate of the original stream. Each substream independently modulates a separate RF carrier. The modulation scheme can be any of the schemes described above, except for FSK and MSK, where the phase does not remain constant throughout the duration of a symbol. At the receiver, the N substreams are demodulated in parallel and then combined to produce the original data stream. Applications of OFDM include 802.11*a* , *g*, and *n* Wi-Fi modems, DAB (digital audio broadcasting) in Europe and Canada, Digital Radio Mondiale short-wave broadcasting, HD (hybrid digital) radio broadcasting on the U.S. AM and FM bands, and European DVB-T (terrestrial digital television broadcasting). The number of carriers used in these systems varies from tens to thousands.

22.6.1 Advantages of OFDM

The system is bandwidth neutral, occupying about the same bandwidth as the traditional modulation systems it replaces. Its main advantage comes from its ability to deal with multipath propagation and channel equalization. Consider the problem of equalization. The propagation channel for a digital signal should have a specific amplitude and phase response, such as the raised-cosine shape discussed above. There are several propagation effects that can distort the shape. Receiving antennas are not always flat in frequency and linear in phase, so they can modify the channel response. Constructive and destructive interference of multipath components are frequency dependent and modify the channel response. Television receivers for 8-VSB use time correlation to detect multipath components and then subtract or realign them. This process, while essential, is not simple, and is particularly difficult in a mobile environment where the multipath situation is constantly changing. At fast baud rates, new multipath components are still arriving at the time the baud ends. With the slow data rates on the individual OFDM carriers, the baud can extend far beyond the time needed for all the multipath components to arrive. Dispersion is also a classic problem in wired data links.[2] Fast DSL (*digital subscriber line*) data links over ordinary telephone circuits have to deal with cable dispersion, poorly controlled impedances, and multipath echos caused by haphazard terminations and discontinuities. An early application of OFDM was in "discrete multi-tone" DSL modems of the early 1990s. While the principles of OFDM date back to the 1960s, practical implementations had to await the advent of fast digital signal processors. Most OFDM system incorporate coding and decoding for forward error correction and are known as COFDM (coded OFDM). As in digital TV transmission, the coding usually encloses the inner modulation/demodulation "physical layer."

22.6.2 Single-frequency broadcasting networks

The essentially perfect ability of OFDM to deal with multipath propagation can be exploited in single-frequency networks, in which broadcasting transmitters, all on the same frequency, serve overlapping areas of coverage. When the identical transmitted signals are synchronized in time and frequency (which can be done using GPS satellite signals), the signals arriving at any given site are indistinguishable from multipath signals originating from a single transmitter. At any given site, the receiving antenna is aimed at the strongest signal. The weaker signals, depending on their phase, increase or decrease the signal strength. They also shift the phases of the recovered Fourier coefficients, but this is not a problem because OFDM channels can be individually equalized via an occasional "training" sequence sent by the transmitter or by using a set of pilot carriers. Single-frequency networks allow efficient use of the spectrum, since the same television or radio channels can be used in adjacent areas – something like solving the classic map coloring problem with a single color. It is assumed that 8-VSB broadcasting could also use single-frequency networking, given echo cancellers of sufficiently high performance. Same-frequency fill-in or "booster" stations with well-synchronized frequencies have been used occasionally for conventional analog AM, FM, and television, but, without echo cancellation, they produce video echos ("ghosts").

[2] The earliest trans-Atlantic telegraph cable, over its great length, accumulated so much dispersion that baud rate had to be slowed down to where the received signal was traced out by a chart recorder. Likewise, early overland telegraph lines were too dispersive to support baseband telephony. The first solution, series loading coils, was a product of the then recent theory of waves on transmission lines, which showed that the reactance per unit length had to be increased, relative to the resistance per unit length.

22.6.3 Implementation of OFDM

The key to practical implementation of OFDM is the discrete Fourier transform (DFT) and its inverse (IDFT), which, in practice are implemented using FFT algorithms. At the transmitter, the data is arranged as a stream of data blocks, each containing N pairs of digital numbers. Each pair of numbers can be considered as the real and imaginary parts of a complex number. Let the complex data numbers in a given block be denoted as D_n, for $n = 0$ to $N-1$.

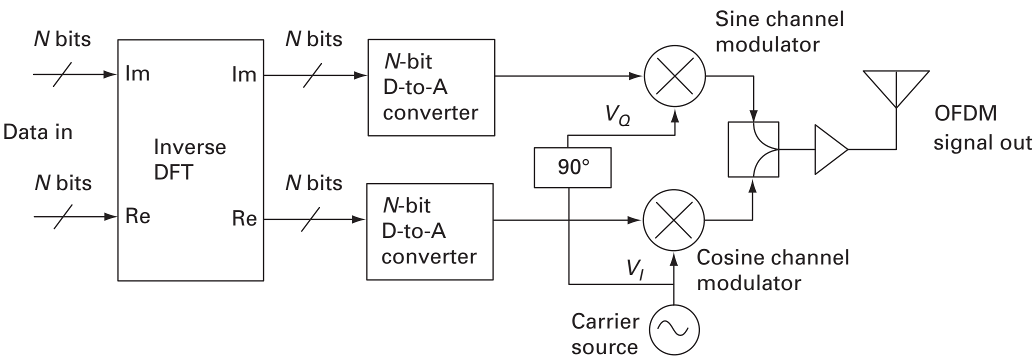


Figure 22.12. OFDM transmitter block diagram.

Figure 22.12, a block diagram for an OFDM transmitter, shows that these numbers are first inverse Fourier-transformed as follows:

$$V_m = \sum_{n=0}^{N-1} D_n e^{j2\pi mn/N}. \tag{22.3}$$

The resulting set of N complex numbers, V_m, for $m = 0$ to $N-1$, is read out serially at a rate N/T, where T must be no greater than the spacing between blocks of input data. If V_0 appears at $t = 0$, we can express the output numbers as a function of time:

$$V(t_m) = \sum_{n=0}^{N-1} D_n e^{j2\pi n t_m/T} = \sum_{n=0}^{N-1} D_n e^{j\omega_n t_m}, \tag{22.4}$$

where ω_n is $2\pi n/T$. Note that this is just the sum of N carriers at equally spaced frequencies, ω_n, and that each carrier is multiplied (amplitude modulated) by a coefficient, D_n.

This set of N modulated carriers is converted from baseband to a QAM signal at RF by using a "cosine mixer" for the real part and a "sine mixer" for the imaginary part. Of course the real and imaginary outputs of the IDFT must be converted to analog voltages before they are applied to the mixers. These (suppressed) carriers are often called subcarriers, although they are not part of the sideband structure of some main carrier. Figure 22.13 shows an OFDM

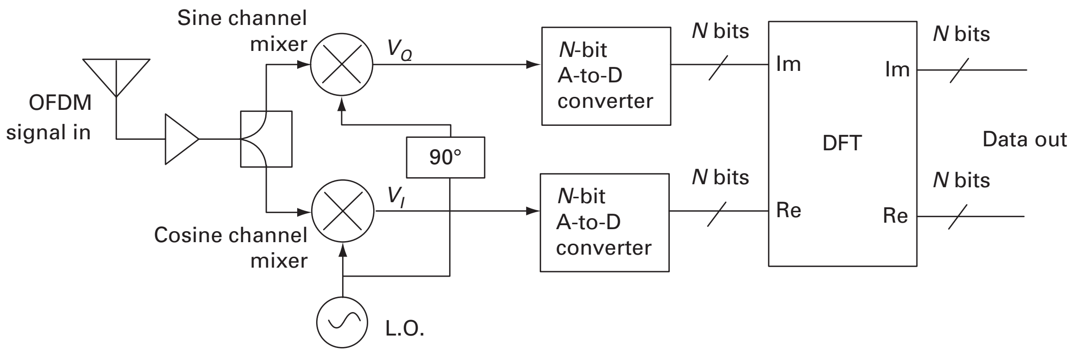


Figure 22.13. OFDM receiver block diagram.

receiver. Note that it is a mirror image of the transmitter. The operations done by the transmitter are undone in the reverse order by the receiver. I and Q samples from the cosine and sine mixers are digitized and combined to from the sequence of complex numbers, V_m, which is Fourier transformed to recover the date (the complex amplitudes of the subcarriers),

$$\frac{1}{N}\sum_{m=0}^{N-1} V_m \, \mathrm{e}^{-\mathrm{j}2\pi nm/N} = D_n, \tag{22.5}$$

and we see that the data has indeed been recovered from the separate carriers. Thus the inverse Fourier transform effectively acts as N carrier oscillators and N mixers to create N modulated carriers while the Fourier transform at the receiver acts as N L.O.'s and N mixers to detect the (complex) amplitude of each carrier. During the interval T, the n-th carrier has constant amplitude; it is an RF pulse with a rectangular envelope. The Fourier transform of this pulse has the shape $\sin(\Delta\omega T/2)/(\Delta\omega T/2)$, where $\Delta\omega$ is the offset from ω_n. The shape plotted in Figure 22.5. We see that the spectrum for any one of the carriers has considerable overlap with the spectra of nearby carriers. Nevertheless, at the nominal frequency of any one of the carriers, the power spectrum of any other carrier is crossing through zero. For this reason, the carriers are said to be orthogonal in the frequency domain, analogous to Nyquist pulses in the time domain, whose maximum value occurs at an instant when the voltage from any earlier pulse is crossing zero.

22.6.4 OFDM guard interval

The key to dealing with multipath in OFDM is the use of a guard interval after every T-second transmission of the group of multicarrier signals. To see that this is necessary, consider the situation shown in Figure 22.14, where no guard interval is used. In the figure, the data is advancing in time toward the right. Ideally we would receive N signal samples from a

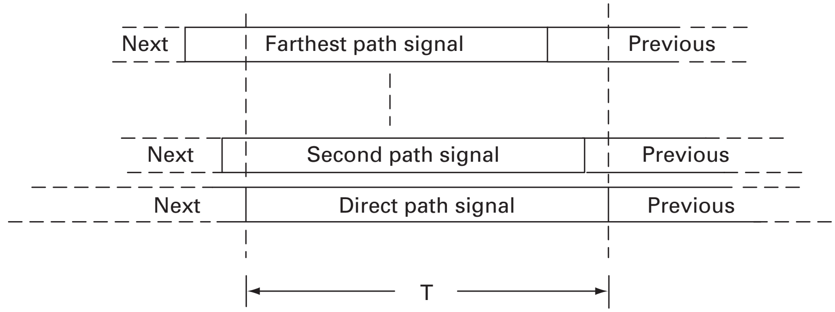


Figure 22.14. OFDM without a guard interval.

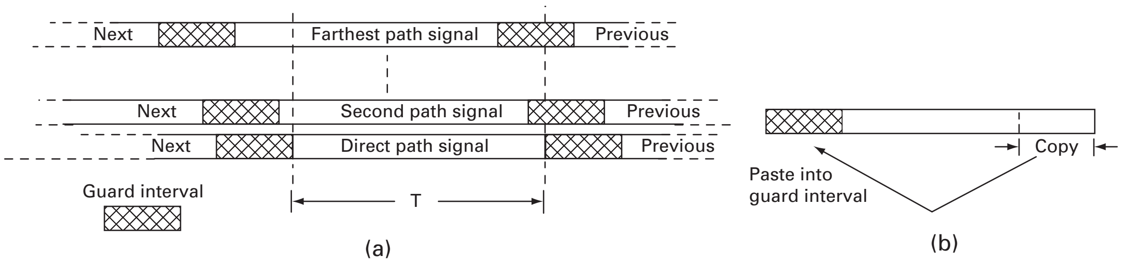


Figure 22.15. OFDM with a guard interval.

direct signal during the time interval 0 to t and then apply the Fourier transform to this set of numbers. But signals arriving on paths other than the direct path will arrive somewhat later, as shown in the figure, where the signal from the farthest path arrives late by about $0.2T$. A set of voltage samples taken from $t = 0$ to T will be the sum of the signal voltages from the various paths so the first 20% of the samples will be contaminated by data from the previous signal interval, arriving via the farthest path and the first 5% or so will be contaminated by the previous signal arriving over all the delayed paths.

Now let us add a guard interval of length αT, as shown in Figure 22.15(a). The contributions from the delayed paths during the interval $t = 0$ to T will now be from whatever has been placed in the guard intervals. You might think of zeros as a possible choice; at least they contain no data from the previous signal. However, they would compromise the orthogonality property. A much better scheme is to extend the data cyclically into the guard intervals, as shown in Figure 22.15(b). Data from the first part of the sequence is pasted into the guard interval so that when the guard interval advances into the sampler, the data sequence for the data interval begins over again. The result is that each multi-path signal is properly demodulated over a complete interval and orthogonality is maintained. The overall system works as follows. Blocks of $2N$ real numbers are considered as blocks of N complex numbers. Each arriving block is subjected to an inverse DFT, whose output is also a block of N complex numbers. These output blocks are cyclically extended from length N to length $N(1+\alpha)$.

The numbers are converted to analog voltages and fed to a QAM modulator at a rate $N(1+\alpha)/T(1+\alpha) = N/T$. The receiver takes a burst of N samples every $T(1+\alpha)$ seconds, nominally aligned with the frame of data arriving over the direct path. Sampling bursts are separated by a time interval αT. The frames of N samples are Fourier-transformed to yield the complex amplitude of each carrier. This set of amplitudes is the original data. Many methods are possible to get the proper framing at the receiver and to regenerate the carrier for coherent detection. DAB audio broadcasting uses differential PSK modulation in order to tolerate carrier phase errors. DVB-T television, however, uses PSK with coherent detection.

22.6.5 Disadvantages of OFDM

One disadvantage of OFDM is that the signal has a high dynamic range, requiring that the transmitter have excellent linearity. (The same is true for any transmitter that transmits the sum of many independent signals, e.g., a cellular telephone base station transmitter.) Another potential problem is use in a mobile environment where different multipath signal components will generally have different Doppler shifts, as when the receiver is moving away from the transmitter but is moving toward a reflective object (maybe a building or tower) which provides a strong additional signal. The spacing between OFDM subcarriers would have to be considerably greater than the Doppler spread of the multipath signals. But, if there are appreciable differences in the time delays of the multipath components, the carriers must be closely spaced. This conflict is less severe at low radio frequencies, since Doppler shifts are proportional to frequency.

22.7 Spread-spectrum and CDMA

Best known now for its application in CDMA cellular telephony, spread spectrum was originally developed to provide secure communications links by using signals spread out in frequency and, therefore, reduced in spectral amplitude below the noise level of common intercept receivers. Besides being hard to detect, the signals cannot be demodulated without finding the long code word used in the spreading process. In addition, spreading makes the signals hard to jam.

22.7.1 Direct sequence spreading

Figure 22.16 shows how a rapidly changing binary pseudorandom code sequence multiplies a data stream to produce a signal whose bandwidth is determined by the "chip"[3] length of the code, rather than by the relatively slow data baud.

[3] Code elements, unlike signal elements, carry no information and are called "chips" rather than "bauds."

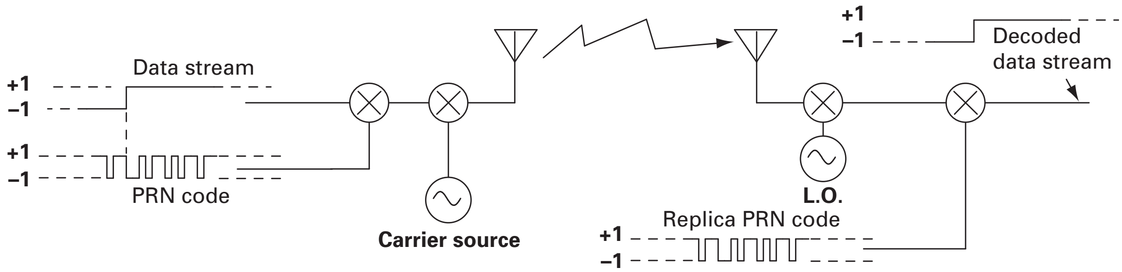


Figure 22.16. Pseudonoise spread spectrum.

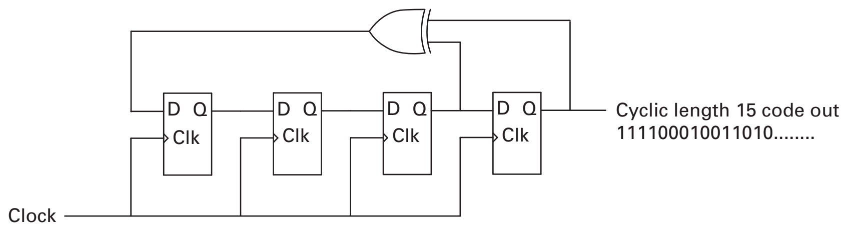


Figure 22.17. Four-stage maximum length shift register code generator.

This system is known as a *direct-sequence* spread spectrum (DSSS). We saw the same circuitry in Chapter 21, where its purpose was to produce a wide radar pulse with modest power that could be decoded (decompressed) to form a narrow high-power pulse. The purpose there was to get around the peak power limitation of the transmitter, but the system also provides the radar with a signal that is harder to intercept and jam. We will encounter the circuit again in Chapter 25, where its use in GPS satellites provides the high range resolution of pulse compression radar as well as multi-access – allowing all the satellites to transmit on the same frequency.

The pseudonoise or "direct sequence" spreading used in Figure 22.16 often uses *maximum length shift register sequences*, which are pseudorandom codes of length $2^N - 1$ generated by a circuit made of an N-stage shift register and a two (or more)-input modulus 2 adder. Figure 22.17 shows a four-stage generator and its 15-chip sequence. For this four-stage generator, only two feedback taps are needed, so the modulus-2 adder is a simple exclusive-OR gate. You can see that, for this kind of code generator, the shift register will never contain all zeros, as that state would then repeat endlessly. All other states do occur, so the length of the sequence is 2^N-1. A sequence of length $2^{32}-1$ can be generated by a 32-stage shift register with taps at stages 32, 31, 30, and 10. These taps can be added with a cascade of three ex-OR gates. With a 1-MHz clock, this generator would run for more than an hour before the sequence starts to repeat. For a given value of N, there are usually several sets of taps, each of which will yield a different maximum length sequence. The key property of the maximum length sequences is that they are noise-like. There is essentially zero correlation between different codes or between a code and a time-shifted version of itself.

The power spectrum of the signal transmitted by this system has a $[\sin(\pi f T)/(\pi f T)]^2$ shape, where ω is the offset from the carrier frequency and T is the chip length. Thus, the bandwidth is about $1/T$, independent of the code length. Longer codes, however, provide better security against interception. This system uses coherent detection; there must be provision for the receiver L.O. to acquire the correct phase (and, therefore, frequency) and for the replica code generator to acquire the phase for alignment with the incoming signal.

Wide frequency spreading requires a short chip time which, in turn, requires a high-precision time alignment of the replica code. Another spreading technique, *frequency hopping*, requires less alignment precision.

22.7.2 Frequency-hopping spread spectrum (FHSS)

In this system, the transmitter carrier frequency "hops" pseudorandomly from one frequency to another throughout the band. The timing of the hops is usually regular, though it, too, could be pseudorandom. An advantage of the hopping scheme is that the bandwidth is determined primarily by the assignment of hopping frequencies and not by the chip time – the time between hops. A convenient scheme is to hop to a new frequency for each data baud. These systems usually use incoherent detection since, until the advent of the direct digital synthesizer (DDS), it was difficult to achieve phase-coherent changes in frequency. One modulation scheme uses frequency shift keying. At the transmitter, the FSK signal is up-converted to the RF band by a frequency hopping carrier generator. At the receiver, a synchronized hopping L.O. converts the signal down to a constant IF frequency. The FSK IF signal can then be demodulated by an ordinary incoherent FM discriminator.

22.7.3 CDMA in cellular telephone systems

By itself, a direct sequence spread spectrum link is as power-efficient as a narrowband link, but deliberately has low spectral efficiency (data rate/RF bandwidth). However, the low cross-correlation property of the codes makes it possible for many users to have spread spectrum links operating in the same band. To each user, the signals from the others appear as random noise. This is known as *code-division multiple access* (CDMA). One advantage of CDMA is that no synchronization is required between users, in contrast to conventional time division and frequency multiplex systems where users must occupy scheduled frequencies and time slots. Another advantage is that the CDMA system quality degrades gracefully; as users are added, each user sees the background noise increase somewhat and experiences a lower signal-to-noise ratio. In time division and frequency division multiplexing, the maximum number of users is strictly limited by the number of frequency and time slots. For the same reason, in a CDMA cellular telephone system, adjacent cells can use the same band. Signals from adjacent cells will be, on average, weaker, and contribute only modestly to the background noise. Moreover, since a mobile user moving from one cell to another stays on the same frequency, there is a "soft handoff" from one base station to the next, rather than a momentary disconnection while new time or frequency slots are allocated.

Problems

Problem 22.1. Show that the impulse response of the filter function of Equation (22.1) is indeed the pulse shape of Equation (22.2).

Problem 22.2. Show that the filter whose frequency response is the *square* of that of the raised-cosine filter will produce intersymbol interference, i.e., show that such a filter, when excited by delta function at $t=0$, will have an impulse response that is not identically zero at $t=T$.

Problem 22.3. Consider the simple single-frequency "network" for traditional AM broadcasting in which a single "booster" transmitter site is located at the fringe of the coverage area of the main transmitter. Describe what a listener, midway between the sites, could experience if (a) the carrier frequencies differ by 400 Hz; (b) the carrier frequencies differ by 0.1 Hz; (c) the carriers are locked to the same frequency.

Problem 22.4. Write a program to simulate the pseudonoise code generator of Figure 22.15 but use 10 stages, rather than four. One input to the ex-OR gate is the output of the tenth stage. Find which stage should be connected to the other input of this gate to produce a sequence of length 1023.

Hint: Start with the shift register in the all-ones state. Count the clock pulses needed to bring it back to this state. Do this for different taps until you find one that requires 1023 pulses.

Glossary

ATSC	Advanced Television Systems Committee: developed the ATV standard for US digital television broadcasting.
BER	Bit Error Rate: probability that a received "one" is deemed a zero or vice versa
BFSK	Binary Frequency-Shift Keying.
BPSK	Binary Phase-Shift Keying.

CDMA	Code Division Multiple Access: spread spectrum modulation that allows multiple users to share the same nominal frequency.
COFDM	Coded Orthogonal Frequency Division Multiplexing: multicarrier modulation system suited to multipath propagation environments. Includes coding for forward error correction.
DAB	Digital Audio Broadcasting standard used in Europe and Canada.
DDS	Direct Digital Synthesizer: frequency synthesizer with phase-continuous frequency transitions (see Chapter 12).
DFT	Discrete Fourier Transform, equivalent to a matrix multiplication that converts a set of N complex numbers to another set of N complex numbers.
DSL	Digital Subscriber Line: Telephone company high-speed digital service
DTV	Digital Television
DSSS	Direct Sequence Spread Spectrum: frequency spreading done via multiplication with pseudorandom ± 1 sequence.
DVB-T	Digital Video Broadcasting – Terrestrial: European over-the-air digital television.
FEC	Forward Error Correction
FDM	Frequency Division Multiplexing: users are assigned different frequencies
FHSS	Frequency Hopping Spread Spectrum: transmitted signal “hops” rapidly between pseudorandomly chosen frequencies.
FIR	Finite Impulse Response: FIR filters are often based on a weighted sum of the signals from a tapped delay line.
GMSK	Gaussian Minimum Shift Keying: a particular frequency shift keying with smooth transitions between ones and zeros.
GPS	Global Positioning System: satellite navigation system.
GSM	Global System for Mobile Communications: the dominant cellular phone system.
HD radio	Hybrid Digital radio: a transition U.S. radio broadcasting standard in which COFDM digital signals share frequency channel assignments with AM and FM signals.
IDFT	Inverse Discrete Fourier Transform: equivalent to a matrix multiplication that inverts (undoes) a discrete Fourier transform.
ISI	Intersymbol Interference: self-interference caused by spreading of nearby pulses.
PAM	Pulse amplitude modulation
OFDM	Orthogonal Frequency Division Multiplexing: multi-carrier modulation system suited to multipath propagation environments.
OOK	On–Off Keying
QAM	Quadrature Amplitude Modulation: system in which two independent baseband signals are transmitted at the same frequency. One is up-converted to an “in-phase” carrier $\cos(2\pi f_{RF}\, t)$, while the other is up-converted to a “quadrature” carrier $\sin(2\pi f_{RF}\, t)$.
QPSK	Quadrature Phase Shift Keying: QAM using BPSK on both I and Q carriers.
TDM	Time Division Multiplexing: users on a single frequency channel have assigned time slots.
VSB	Vestigial Sideband: The remaining vestige of the lower sideband of an AM signal after asymmetric bandpass filtering.

References

[1] Federal Communications Commission *DTV Report on COFDM and 8-VSB Performance*, Office of Engineering and Technology, September 30, 1999.

[2] IEEE Std 802.11a-1999(R2003) (Supplement to IEEE Std 802.11–1999) [Adopted by ISO/IEC and redesignated as ISO/IEC 8802-11:1999/Amd 1:2000(E)]. This standard describes the Wi-Fi COFDM format.

[3] Lahti, B. P., *Modern Digital and Analog Communication Systems*, 3rd edn, Oxford: Oxford University Press, 1998.

[4] Proakis, J. B., *Digital Communications*, 4th edn, New York: McGraw-Hill, 2000.

[5] Schulze, H. and Lüders, C., *Theory and Applications of OFDM and CDMA Wideband Wireless Communications*, New York: John Wiley, 2005.

https://doi.org/10.1017/CBO9780511626951.023 Published online by Cambridge University Press