Digital Encoding And Decoding

Dr. George W Benthien

August 13, 2007

Revised March 30, 2010

E-mail: george@gbenthien.net

1 Introduction

Many electronic communication devices today process and transfer information digitally. Examples are cable televisions, cell phones, cable/DSL modems, and wireless routers. Digital information is specified by a sequence of zeroes and ones such as 1101010001. Each zero or one is called a binary digit or bit. Digital information is usually embedded in an electric or optical signal. This article describes a number of methods for encoding digital information in an electric or optical signal and later extracting this information. I have divided these methods into two types. The first type (sometimes referred to as baseband) involves the encoding of digital information in digital signals consisting of a sequence of rectangular pulses. These methods are used primarily in local area Ethernet networks. The second type often referred to as broadband) involves modulation of one or more sinusoidal carrier waves. These methods are used in cable/DSL broadband networks, satellite communication, cell phones, digital TV, and wireless networks. Some of the specific techniques that will be discussed are Manchester encoding, 4B/5B block encoding, MLT-3 encoding, Frequency Shift Keying (FSK), Quadrature Phase Shift Keying (QPSK), Quadrature Amplitude Modulation (QAM), Orthogonal Frequency Division Multiplexing (OFDM), Frequency Hopping Spread Spectrum (FHSS), and Direct-Sequence Spread Spectrum (DSSS).

In digital encoding time is broken up into a succession of equal time intervals of length T. In each time interval either one or a group of bits of the given sequence are encoded in the signal. T is called the *symbol period* and the group of bits encoded during a time interval T is called a *symbol*. The rate at which symbols are encoded is called the *baud rate*. The rate at which bits are encoded is called the *bit rate*.

2 Encoding Digital Information in Digital Signals

A digital signal consists of a sequence of equal width rectangular pulses. The timing of the pulses is controlled by an internal clock whose output is an alternating sequence of high and low valued rectangular pulses. A high/low pair is called a clock cycle.

2.1 NRZ encoding

The simplest digital signal representing a bit sequence uses just two voltage levels and represents a 1 by the higher voltage and a zero by the lower voltage. This type of encoding is called NRZ (Non-Return to Zero). An example of an NRZ encoded signal is shown in Figure 1.

Although simple, this method for encoding digital information has some serious drawbacks and is seldom used. First, it is difficult to keep the clocks of the source and receiver synchronized if there happen to be long sequences of ones or zeros. The receiver uses transitions in level to determine clock cycle boundaries. Second, it is impossible to distinguish between a long sequence of zeros and the absence of a signal. Third, a long series of zeros or ones causes the average signal value,

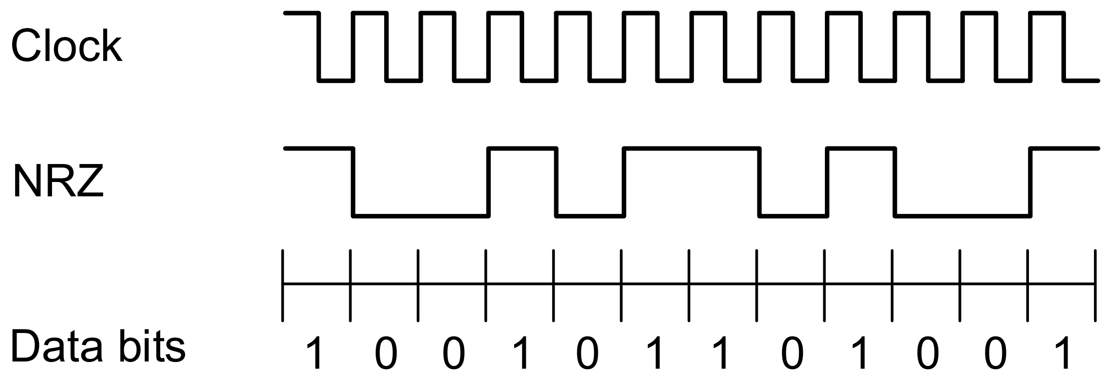

Figure 1: Example of NRZ encoding.

which is used to distinguish between high and low values, to drift. Thus, for many reasons, it is desirable to have frequent transitions between the high and low values.

2.2 NRZI encoding

Another simple encoding method, called NRZI (Non-Return to Zero Inverted), changes level for a 1 bit and stays at the same level for a 0 bit. An example of an NRZI encoded signal is shown in Figure 2.

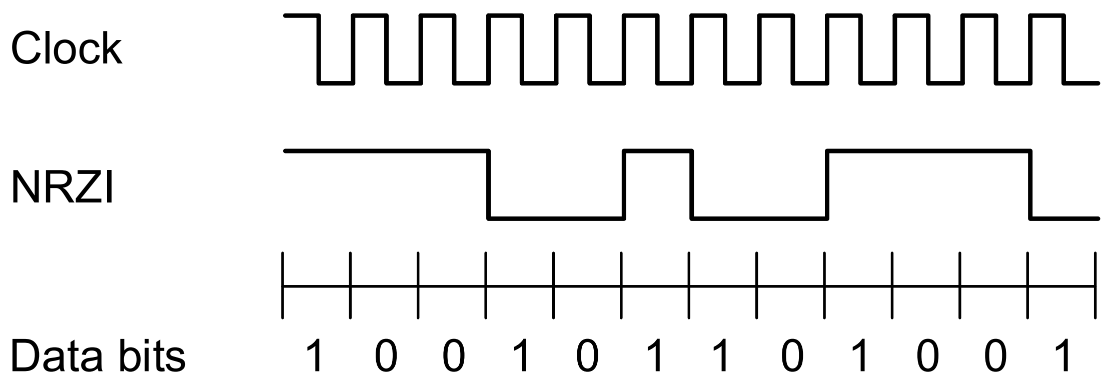

Figure 2: Example of NRZI encoding.

This method gets rid of the problems associated with long strings of ones, but does nothing about long strings of zeros. We will see later that the NRZI method can be effectively used in combination with other methods that don't produce long strings of zeros.

2.3 Manchester encoding

In Manchester encoding 0 and 1 bits are represented in a clock cycle by the signals shown in Figure 3.

Figure 3: Manchester encoding of 0 and 1 bits.

Here the signal transition occurs in the middle of the cycle. An example of a Manchester encoded sequence is shown in Figure 4.

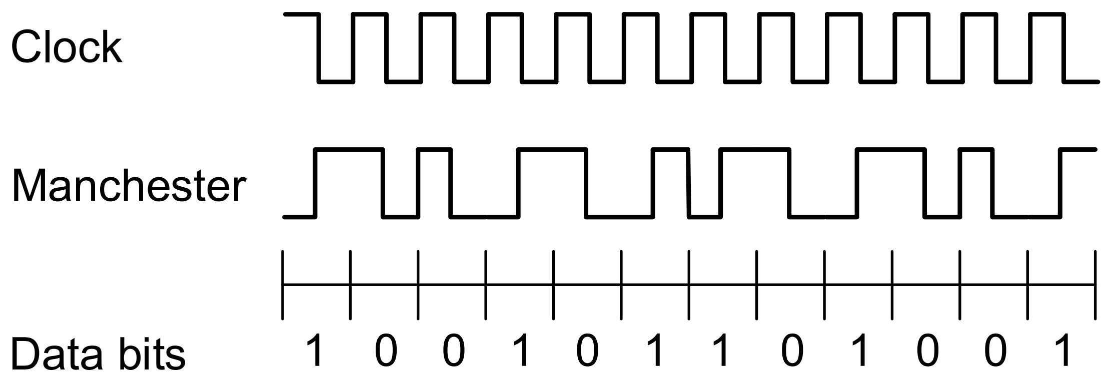

Figure 4: An example of a Manchester encoded signal.

This encoding method is used in 10 Mbs (Megabits per second) 10BaseT Ethernet networks. Manchester encoding solves the problems mentioned previously in connection with NRZ encoding. However, since the signal alternates level every clock cycle, Manchester encoding has a broader frequency spectrum than NRZ. For 100 Mbs and higher Ethernet networks, the spectrum of a Manchester encoded signal extends past the high frequency limit for unshielded twisted pair Ethernet cables. Thus, other encoding schemes are used for high speed Ethernet.

2.4 4B/5B encoding

4B/5B is a block encoding scheme designed to break up long strings of ones and zeros without increasing the frequency bandwidth. In this scheme the bit sequence is broken up into four bit blocks. Each block of four bits is replaced with a five bit block according to Table 1.

Table 1: Four bit to five bit conversion table.

Four bit data	Five bit code	Four bit data	Five bit code
0000	11110	1000	10010
0001	01001	1001	10011
0010	10100	1010	10110
0011	10101	1011	10111
0100	01010	1100	11010
0101	01011	1101	11011
0110	01110	1110	11100
0111	01111	1111	11101

The five bit codes were selected so that there is no more than one leading zero and no more than two trailing zeros. Thus, when the codes are strung together, there can be no more than three consecutive zeros. The string of bits after the replacement are transmitted using NRZI. As we saw before, NRZI effectively handles strings of ones. An example of a 4B/5B encoded signal is shown in Figure 5.

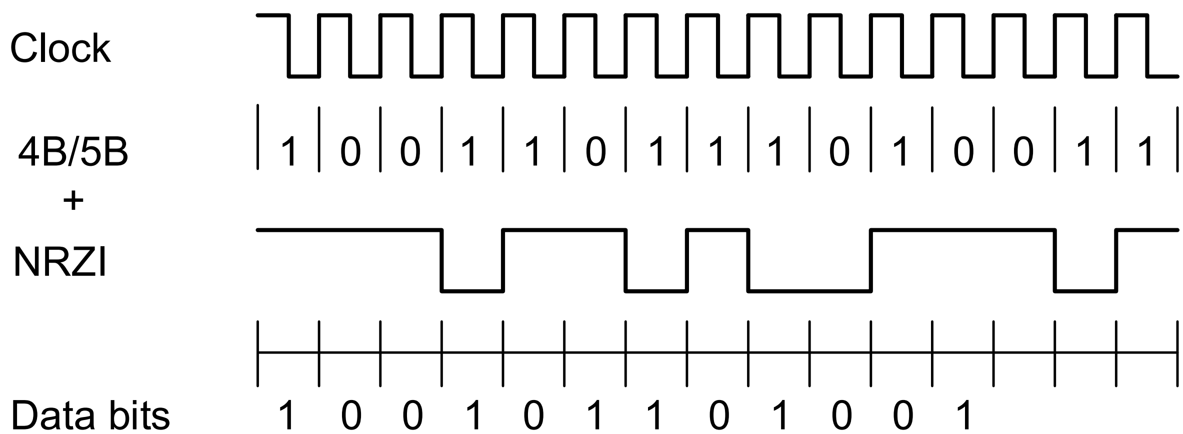

Figure 5: An example of 4B/5B encoding.

4B/5B encoding followed by NRZI is used in 100BaseTX (Fast Ethernet) networks in conjunction with the multilevel encoding scheme MLT-3 described in the next section. Since only half of the five bit codes are used in Table 1, the remaining codes can be used for other purposes.

2.5 MLT-3 encoding

MLT (Multi Level Threshold) encoding is used to decrease the high frequency content of the signal. MLT-3 uses three levels denoted by -1, 0, and 1. The process cycles through the four values

-1, 0,+1, 0. It moves to the next of the four states in a cyclical manner to transmit a 1 bit, and stays in the same state to transmit a 0 bit. An example of MLT-3 encoding is shown in Figure 6. The fastest an MLT-3 signal can go through a complete cycle is four clock cycles. Thus, the high frequency limit of an MLT-3 signal will be about one-fourth that of a Manchester encoded signal. In Fast Ethernet (100 Mbps) MLT-3 is applied to the signal generated by 4B/5B and NRZI. Higher order block codes such as 8B/10B and higher order multi-level methods such as MLT-5 are used in higher speed Ethernet networks.

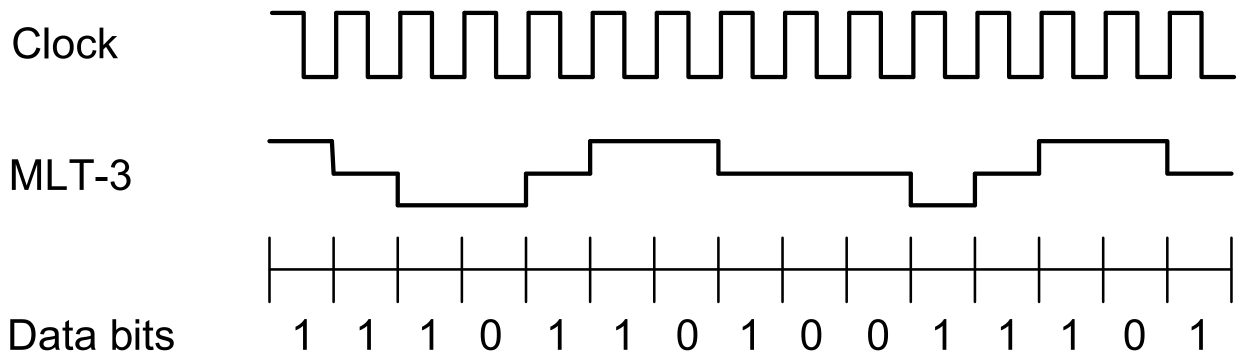

Figure 6: An example of MLT-3 encoding.

3 Encoding Digital Information by Modulating Analog Signals

Another class of methods encodes digital data by modulating one or more sinusoidal carrier waves. The process of backing out the digital information from the modulated wave is called demodulation. A device for performing modulation and demodulation is called a modem (**mod**ulator-**dem**odulator). The three basic methods for modulation are amplitude modulation, frequency modulation, and phase modulation. Sometimes a combination of one or more of these basic methods is used.

We will sometimes refer to a modulation/demodulation technique as coherent or incoherent. A coherent modulation technique is one in which the phase of the signal is controlled, and an incoherent modulation technique is one in which the phase of the signal is not controlled. A coherent demodulation scheme is one in which use is made of the phase of the signal. An incoherent demodulation scheme makes no use of the phase of the signal.

Another factor that is important in all modulation schemes is the rate at which the frequency spectrum of the modulated signal decays as frequency increases. It turns out that the rate of decay depends on the smoothness of the modulated signal. Smoother time functions have frequency spectra that approach zero faster as frequency increases. Since transmitted signals are often restricted to a certain frequency band, slow decay of the frequency spectrum can lead to leakage of the signal into other bands. Generally modulated signals with jump discontinuities decay as one over the frequency, continuous modulated signals decay as one over the frequency squared, modulated signals with a continuous derivative decay as one over the frequency cubed, and so on. It is particularly difficult to obtain smoothness of the modulated signal at the transitions between symbol periods.

3.1 Amplitude modulation

In amplitude modulation the amplitude of the carrier wave is allowed to vary from interval to interval in order to specify the digital information. If only two different amplitudes are used, then one amplitude would correspond to a 0 bit and the other would correspond to a 1 bit. If four amplitudes are used, then we could encode two bits in each interval. The four amplitudes would correspond to the bit patterns 00, 01, 10, and 11. Similarly, we can define higher order encoding schemes where 8 amplitudes could encode 3 bits per interval, 16 amplitudes could encode 4 bits per interval, etc. An amplitude modulated signal could be written as follows

$$x(t) = a(t)\sin(2\pi f_c t) \tag{1}$$

where $a(t)$ is constant for each symbol period and f_c is the frequency of the carrier wave. Amplitude modulation is more sensitive to noise than the other techniques we will discuss. It is seldom used by itself, but it is used in conjunction with other modulation techniques. An example of an amplitude modulated signal involving two amplitudes is shown in Figure 7.

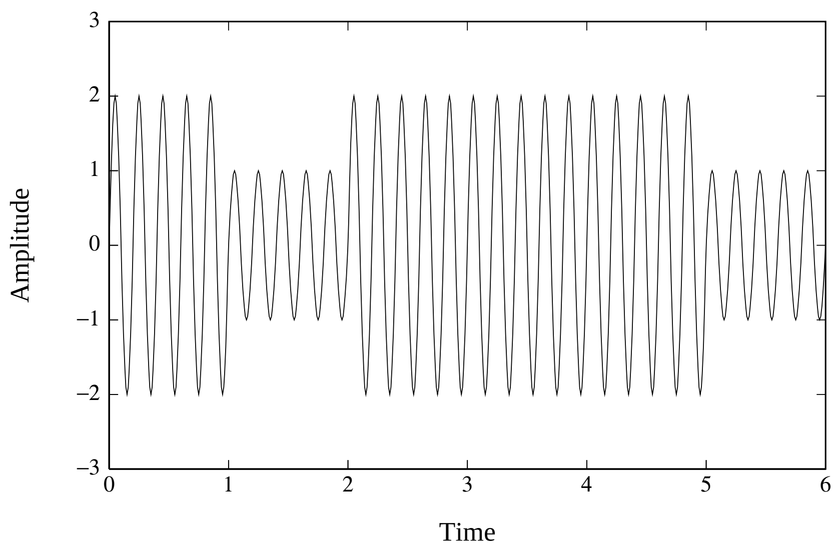

Figure 7: Amplitude modulated signal corresponding to the bit sequence (101110).

Amplitude modulated signals are usually not continuous at the switching times nT, $n = 1, 2, \ldots$. An amplitude modulated sine wave as in equation (1) will be continuous if the carrier wave is zero at the edges of each symbol period. This will be true if the carrier frequency f_c is selected so that there are an integral number of half-cycles in a symbol period, i.e., if

$$f_c = \frac{p}{2T} \qquad \text{for some integer } p.$$

3.2 Frequency modulation

In frequency modulation the frequency of the signal is allowed to vary from interval to interval. The most common type of frequency modulation for digital encoding is *Frequency Shift Keying* (FSK) in which only two frequencies are used. In FSK one frequency corresponds to a zero bit and the other frequency corresponds to a one bit. An FSK signal can be written as follows

$$x(t) = A\cos\Big(2\pi\big(f_c + m(t)\Delta f\big)t\Big) \tag{2}$$

where A is a fixed amplitude, f_c is a center frequency, Δf is a frequency increment, and $m(t)$ is a digital function that is either plus or minus one on each symbol interval. Here the two frequencies are $f_c - \Delta f$ and $f_c + \Delta f$. An example of an FSK signal is shown in Figure 8.

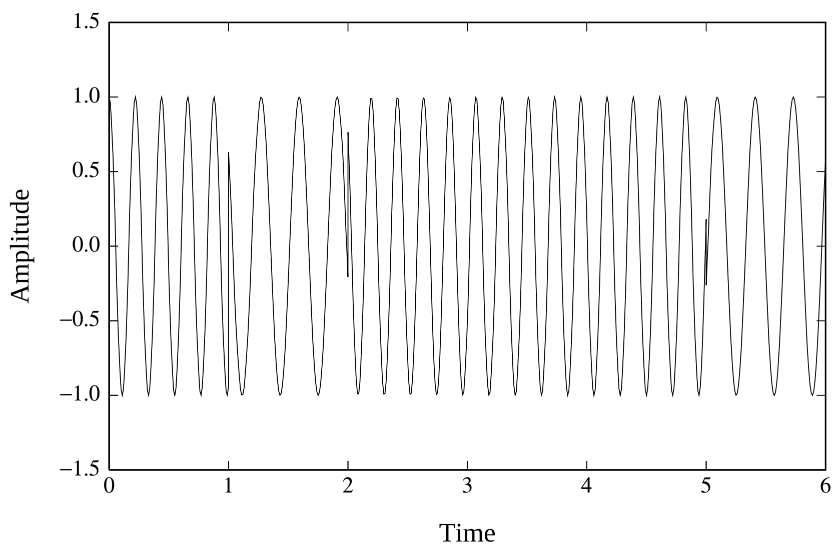

Figure 8: Frequency modulated signal corresponding to the bit sequence (101110).

FSK signals are usually not smooth at the boundary between intervals. In the next section we will see how FSK signals can be made continuous by adding a variable phase angle. The usual method for demodulating an FSK signal is to pass the signal through two band pass filters centered at the two transmission frequencies. The one having the largest output over a symbol period corresponds to the frequency used in that symbol period.

3.3 Continuous Phase FSK Modulation (CPFSK)

A continuous phase FSK signal has the form

$$x(t) = A\cos\Big(2\pi f_c t + 2\pi\Delta f \int_0^t m(\tau)\,d\tau\Big). \tag{3}$$

where m is a digital message signal that is constant on each time interval of width T. The argument of the cosine is continuous and hence $x(t)$ is continuous. If t is in the interval $nT \le t \le (n+1)T$, then

$$\begin{aligned}\int_0^t m(\tau)\,d\tau &= T\sum_{k=0}^{n-1} m_k + (t-nT)m_n \\ &= m_n t + \Big(T\sum_{k=0}^{n-1} m_k - nTm_n\Big) \\ &= m_n t + \phi_n/(2\pi\Delta f)\end{aligned} \tag{4}$$

where m_k is the value of $m(t)$ in the interval $\big[kT, (k+1)T\big]$ and

$$\phi_n = 2\pi\Delta f\Big(T\sum_{k=0}^{n-1} m_k - nTm_n\Big).$$

Substituting equation (4) into equation (3), we get

$$x(t) = A\cos(2\pi f_c t + 2\pi m_n \Delta f\ t + \phi_n) \qquad \text{for } nT \le t \le (n+1)T. \tag{5}$$

Thus, $x(t)$ is a continuous FSK signal with an added phase term in each interval. In a later section we will look at Minimum Shift Keying (MSK) which is an important subclass of CPFSK signals.

3.4 Phase modulation

Phase modulation uses signals of the form

$$x(t) = A\cos\big(2\pi f_c t + \phi(t)\big) \tag{6}$$

where $\phi(t)$ can take on one of a finite set of values in each symbol period. The signal in equation (6) can be written in the alternate form

$$x(t) = I(t)\cos(2\pi f_c t) + Q(t)\sin(2\pi f_c t) \tag{7}$$

where $I(t) = A\cos\phi(t)$ and $Q(t) = A\sin\phi(t)$. A two-dimensional plot of the possible pairs (I, Q) is called a *constellation diagram*.

Phase modulated signals are usually demodulated using a coherent scheme. Multiplying the signal $x(t)$ in equation (7) by $\cos(2\pi f_c t)$ and using trigonometric formulas for double angles, we get

$$\begin{aligned} x(t)\cos(2\pi f_c t) &= I(t)\cos^2(2\pi f_c t) + Q(t)sin(2\pi f_c t)cos(2\pi f_c t) \\ &= \tfrac{1}{2}I(t)[1+\cos(4\pi f_c t)] + \tfrac{1}{2}Q(t)\sin(4\pi f_c t). \end{aligned} \tag{8}$$

Passing $x(t)\cos(2\pi f_c t)$ through a low-pass filter gives $\frac{1}{2}I(t)$. Multiplying the signal $x(t)$ by $\sin(2\pi f_c t)$ and using trigonometric formulas for double angles, we get

$$\begin{aligned} x(t)\sin(2\pi f_c t) &= I(t)sin(2\pi f_c t)cos(2\pi f_c t) + Q(t)\sin^2(2\pi f_c t) \\ &= \tfrac{1}{2}I(t)\sin(4\pi f_c t) + \tfrac{1}{2}Q(t)[1-\cos(4\pi f_c t)]. \end{aligned} \tag{9}$$

Passing $x(t)\sin(2\pi f_c t)$ through a low-pass filter gives $\frac{1}{2}Q(t)$. The closest point to (I, Q) in the constellation diagram is used to obtain the symbol.

One of the most popular forms of phase modulation is *Quadrature Phase Shift Keying* (QPSK). In QPSK the phase $\phi(t)$ takes on one of the four values $\pi/4$, $3\pi/4$, $5\pi/4$, or $7\pi/4$ in each symbol period. The signals corresponding to these four phase angles are $A\big[\cos(2\pi f_c t) - \sin(2\pi f_c t)\big]/\sqrt{2}$, $A\big[-\cos(2\pi f_c t) - \sin(2\pi f_c t)\big]/\sqrt{2}$, $A\big[-\cos(2\pi f_c t) + \sin(2\pi f_c t)\big]/\sqrt{2}$, and $A\big[\cos(2\pi f_c t) + \sin(2\pi f_c t)\big]/\sqrt{2}$. If we let $A = \sqrt{2}$, then the pair (I, Q) takes on the values $(1, -1)$, $(-1, -1)$, $(-1, 1)$, and $(1, 1)$. Each signal $x(t)$ corresponds to a unique pair (I, Q). Figure 9 shows the points corresponding to the pairs (I, Q) for QPSK along with the associated symbols.

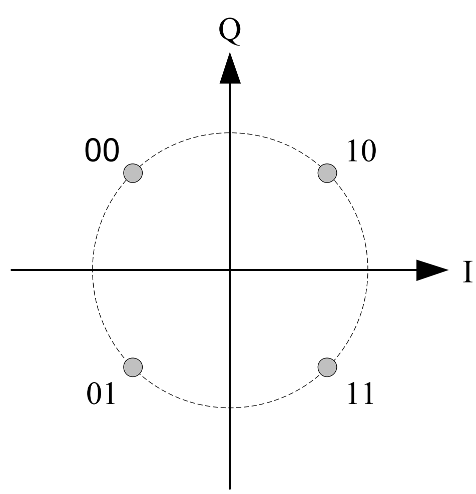

Figure 9: Constellation diagram for QPSK modulation.

The points in this constellation diagram are often associated with points in the complex plane. For phase shift keying the points in the constellation diagram all lie on a circle. QPSK can be easily generated using equation (7). Let $d(t)$ be a digital signal representing the bit sequence to be encoded. Choose $d(t)$ to be -1 on an interval for a zero bit and +1 on an interval for a one bit. Let $d_0, d_1 \ldots$ be the successive values of $d(t)$. Define $d_I(t)$ to be a digital signal with symbol period $2T$ representing the even values $d_0, d_2, \ldots$. Define $d_Q(t)$ to be a digital signal with symbol period $2T$ representing the odd values $d_1, d_3, \ldots$. Figure 10 shows the signals d, d_I, and d_Q corresponding to the bit sequence 11000111. The QPSK signal is obtained by letting $I(t) = Ad_I(t)/\sqrt{2}$ and $Q(t) = Ad_Q(t)/\sqrt{2}$.

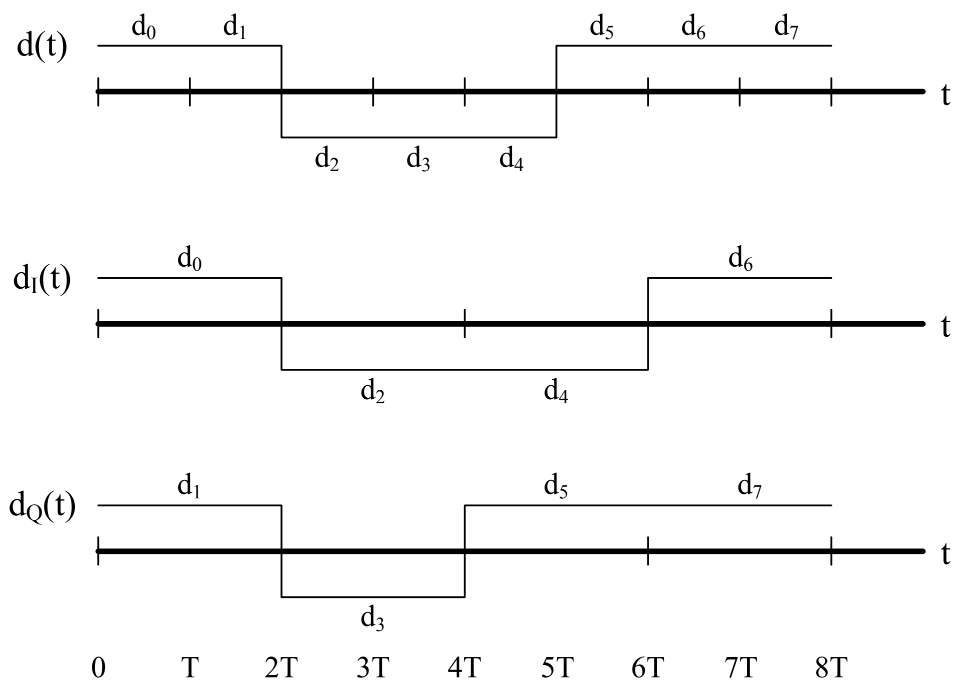

Figure 10: Digital signals used in QPSK modulation.

The form of QPSK we have constructed has the same bit rate as the original, but the symbol period is doubled. Doubling the symbol period has the effect of narrowing the frequency bandwidth of the signal. We could have kept the same symbol period for the QPSK signal and the bit rate would then have doubled.

An alternate form of QPSK uses the four phases 0, $\pi/2$, π, and $3\pi/2$. QPSK is widely used in cable/DSL modems along with the Quadrature Amplitude Methods to be discussed later. Higher order phase shift keying methods can be constructed. Figure 11 shows phases that could be used to represent 3 bits per symbol. This modulation scheme is called 8-PSK.

Some systems use a modification of phase shift keying called *Differential Phase Shift Keying* (DPSK). Instead of using phase angles relative to a fixed standard, DPSK uses phase angles relative to the phase in the preceding symbol period. DPSK is simpler to implement than ordinary PSK, but has larger demodulation errors.

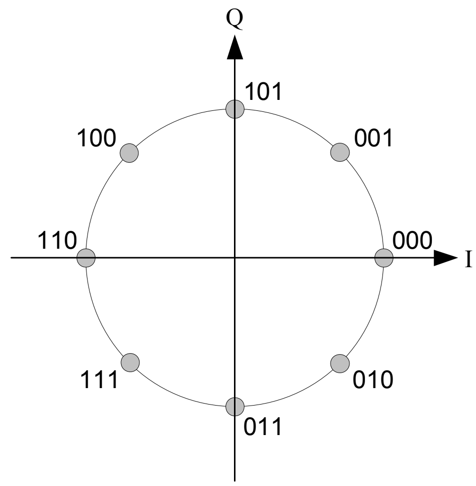

Figure 11: Constellation diagram for 8-PSK modulation.

3.5 Minimum Shift Keying (MSK)

The goal of minimum shift keying is to obtain a smoother signal and thus a faster decaying frequency spectrum. There are several ways to approach MSK. We will consider MSK as a modification of QPSK. In QPSK we modulated the carrier waves $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$ by the digital signals $d_I(t)$ and $d_Q(t)$. Suppose that we replace the square pulses in d_I and d_Q by half cycle sinusoidal pulses as in

$$x(t) = d_I(t)\cos\left(\frac{\pi t}{2T}\right)\cos(2\pi f_c t) + d_Q(t)\sin\left(\frac{\pi t}{2T}\right)\sin(2\pi f_c t) \tag{10}$$

where $d_I(t)$ and $d_Q(t)$ are the same digital signals used previously. The functions d_I and d_Q are continuous at odd multiples of T, and $\sin\left(\frac{\pi t}{2T}\right)$ is zero at even multiples of T. Therefore, the second term on the right-hand-side of equation (10) is continuous. Since $\cos\left(\frac{\pi t}{2T}\right)$ vanishes at odd multiples of T, the signal $x(t)$ will be continuous if we can modify $d_I(t)$ so that it is continuous at even multiples of T. This is easily done. We only need to shift $d_I(t)$ to the left by T as is illustrated in Figure 12. With this modification $x(t)$ is called an MSK signal. We will now show that the MSK signal $x(t)$ is in fact a continuous phase FSK signal.

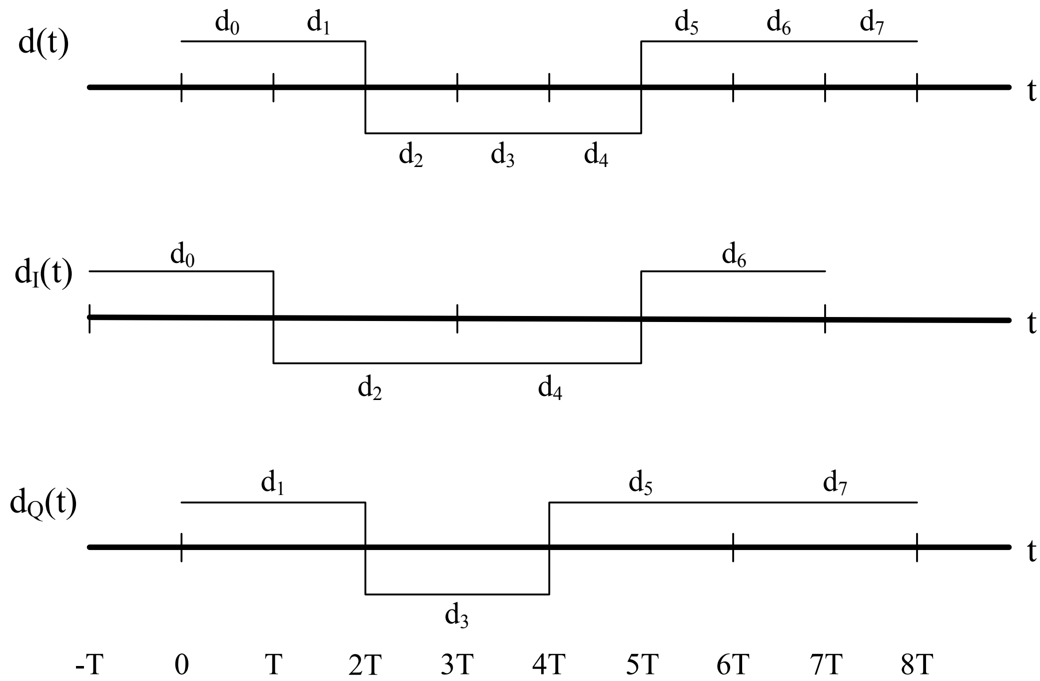

Figure 12: Shifted digital signal d_I

Since d_I and d_Q are either plus or minus one on each interval, applying trigonometric identities for sums and differences of angles to equation (10) gives

$$x(t) = +cos(2\pi f_c t - \frac{\pi t}{2T}) \qquad \text{when } d_I(t) = +1 \text{ and } d_Q(t) = +1 \tag{11a}$$

$$= -cos(2\pi f_c t - \frac{\pi t}{2T}) \qquad \text{when } d_I(t) = -1 \text{ and } d_Q(t) = -1 \tag{11b}$$

$$= +cos(2\pi f_c t + \frac{\pi t}{2T}) \qquad \text{when } d_I(t) = +1 \text{ and } d_Q(t) = -1 \tag{11c}$$

$$= -cos(2\pi f_c t + \frac{\pi t}{2T}) \qquad \text{when } d_I(t) = -1 \text{ and } d_Q(t) = +1. \tag{11d}$$

These equations can be written more concisely as

$$x(t) = d_I(t)\cos\Big(2\pi f_c t - d_I(t) d_Q(t) \frac{\pi t}{2T}\Big) = \cos\Big(2\pi f_c t - d_I(t) d_Q(t) \frac{2\pi t}{4T} + \phi(t)\Big). \tag{12}$$

where $\phi(t) = \big[1 - d_I(t)\big]\pi/2$. Thus, $x(t)$ is an FSK signal with $\Delta f = 1/4T$ and a phase on each interval of 0 or π. The name MSK is applied to this signal since Δf is the minimum frequency increment that will allow the signals corresponding to $f_c + \Delta f$ and $f_c - \Delta f$ to be orthogonal over a symbol period.

Let us now look at the derivative of $x(t)$. Differentiating equation (10), we obtain

$$\begin{aligned}\dot{x}(t) &= d_I(t)\Big[-\frac{\pi}{2T}\sin\Big(\frac{\pi t}{2T}\Big)\cos(2\pi f_c t) - 2\pi f_c\cos\Big(\frac{\pi t}{2T}\Big)\sin(2\pi f_c t)\Big] \\ &\quad + d_Q(t)\Big[-\frac{\pi}{2T}\cos\Big(\frac{\pi t}{2T}\Big)\sin(2\pi f_c t) + 2\pi f_c\sin\Big(\frac{\pi t}{2T}\Big)\cos(2\pi f_c t)\Big] \\ &= \Big[-d_I(t)\frac{\pi}{2T} + d_Q(t)2\pi f_c\Big]\sin\Big(\frac{\pi t}{2T}\Big)\cos(2\pi f_c t) \\ &\quad + \Big[-d_I(t)2\pi f_c + d_Q(t)\frac{\pi}{2T}\Big]\cos\Big(\frac{\pi t}{2T}\Big)\sin(2\pi f_c t). \end{aligned} \tag{13}$$

Since $\sin\Big(\frac{\pi t}{2T}\Big)$ is zero for t an even multiple of T and $\cos\Big(\frac{\pi t}{2T}\Big)$ is zero for t an odd multiple of T, $\dot{x}(t)$ will be zero at all multiples of T if f_c is chosen so that

$$\begin{aligned}\cos(2\pi f_c t) &= 0 \qquad \text{for } t \text{ an odd multiple of } T \\ \sin(2\pi f_c t) &= 0 \qquad \text{for } t \text{ an even multiple of } T.\end{aligned}$$

These conditions will hold if f_c is an odd multiple of $\Delta f = 1/4T$. If f_c is chosen in this way, then the MSK signal will not only be continuous, but will have a continuous derivative.

We mentioned earlier that the Δf used in MSK was the smallest increment that allowed the signals corresponding to $f_c + \Delta f$ and $f_c - \Delta f$ to be orthogonal over every symbol period. Let us look now at the conditions necessary for orthogonality. Using trigonometric identities for the sum and difference of two angles we obtain

$$\begin{aligned}\int_{nT}^{(n+1)T}\cos\big(2\pi(f_c+\Delta f)t\big)\cos\big(2\pi(f_c-\Delta f)t\big)\,dt &= \tfrac{1}{2}\int_{nT}^{(n+1)T}\big[\cos(4\pi f_c t) + \cos(4\pi\Delta f\, t)\big]\,dt \\ &= \tfrac{1}{2}\left[\frac{\sin(4\pi f_c t)}{4\pi f_c} + \frac{\sin(4\pi\Delta f\, t)}{4\pi\Delta f}\right]_{nT}^{(n+1)T}. \end{aligned} \tag{14}$$

For orthogonality we need this integral to vanish over each symbol period. It can be seen from equation (14) that the integral will vanish if f_c and Δf satisfy the conditions

$$f_c = \frac{p}{4T} \qquad \text{for some integer } p \tag{15a}$$

$$\Delta f = \frac{q}{4T} \qquad \text{for some integer } q < p. \tag{15b}$$

The smallest Δf satisfying the condition (15b) is the one used in MSK.

The process described for the MSK signal is not the only way to obtain a smooth FSK signal. Consider an FSK signal $x(t)$ of the form

$$x(t) = r(t)\cos\big(2\pi f_c t + s(t)2\pi\Delta f t\big) \tag{16}$$

where $r(t)$ and $s(t)$ are ± 1 on each symbol interval. The only way we can hope to match slopes at the boundary between each pair of intervals is for the cosine term in equation (16) to have zero slope at each boundary point nT, $n = 1, 2, \ldots$. The value of the cosine at the zero slope points is ± 1. If we can match the zero slopes, then the function $r(t)$ can be chosen so as to match the values at this boundary. Therefore, we want the conditions

$$\cos\big(2\pi(f_c + \Delta f)nT\big) = \pm 1 \tag{17a}$$
$$\cos\big(2\pi(f_c - \Delta f)nT\big) = \pm 1 \tag{17b}$$

to hold for all n. These conditions will hold if f_c and Δf satisfy the following relations

$$2(f_c + \Delta f)T = p \qquad \text{for some integer } p \tag{18a}$$
$$2(f_c - \Delta f)T = q \qquad \text{for some integer } q < p. \tag{18b}$$

By adding and subtracting equations (18a) and (18b) we obtain

$$f_c = \frac{p+q}{4T} \tag{19}$$
$$\Delta f = \frac{p-q}{4T}. \tag{20}$$

Figure 13 shows a smooth FSK signal corresponding to the parameters $p = 8$, $q = 4$, and $T = 1$. MSK corresponds to the cases where $p = q + 1$. comparing the conditions (19) and (20) with the conditions (15a) and (15b) we see that the two frequency signals will be orthogonal on every symbol period. The orthogonality can be used to separate the two frequencies in the demodulation process.

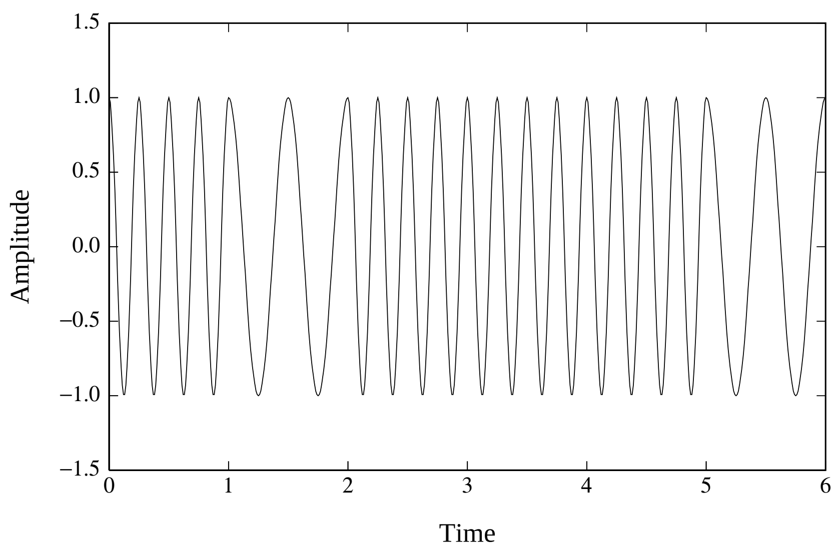

Figure 13: Smooth FSK signal corresponding to the bit sequence (101110).

Let us now look at the demodulation of an MSK signal. An MSK signal can be demodulated using a coherent scheme. Multiplying equation (10) by $cos(2\pi f_c t)$ and using trigonometric identities for double angles, we get

$$x(t)\cos(2\pi f_c t) = \tfrac{1}{2}d_I(t)\cos\left(\frac{\pi t}{2T}\right)[1+\cos(4\pi f_c t)] + \tfrac{1}{2}d_Q(t)\sin\left(\frac{\pi t}{2T}\right)\sin(4\pi f_c t). \tag{21}$$

Passing $x(t)\cos(2\pi f_c t)$ through a low pass filter, we obtain the signal $x_c(t)$ defined by

$$x_c(t) = \tfrac{1}{2}d_I(t)\cos\left(\frac{\pi t}{2T}\right). \tag{22}$$

If we now integrate $x_c(t)$ over an interval $[(2k-1)T, (2k+1)T]$, we obtain

$$\int_{(2k-1)T}^{(2k+1)T} x_c(t)\,dt = \tfrac{1}{2}d_I(2kT)\frac{2T}{\pi}\sin\left(\frac{\pi t}{2T}\right)\Bigg|_{(2k-1)T}^{(2k+1)T} = 2d_I(2kT)\frac{T}{\pi}(-1)^k. \tag{23}$$

From this result we can obtain $d_I(2kT)$.

Similarly, multiplication of equation (10) by $\sin(2\pi f_c t)$ yields

$$x(t)\sin(2\pi f_c t) = \tfrac{1}{2}d_I(t)\cos\left(\frac{\pi t}{2T}\right)\sin(4\pi f_c t) + \tfrac{1}{2}d_Q(t)\sin\left(\frac{\pi t}{2T}\right)[1+\cos(4\pi f_c t)]. \tag{24}$$

Passing $x(t)\sin(2\pi f_c t)$ through a low pass filter, we obtain the signal $x_s(t)$ defined by

$$x_s(t) = \tfrac{1}{2}d_Q(t)\sin\left(\frac{\pi t}{2T}\right). \tag{25}$$

If we now integrate $x_s(t)$ over an interval $[2kT, (2k+2)T]$, we obtain

$$\int_{2kT}^{(2k+2)T} x_s(t)\,dt = -\tfrac{1}{2}d_Q\big((2k+1)T\big)\frac{2T}{\pi}\cos\left(\frac{\pi t}{2T}\right)\Bigg|_{2kT}^{(2k+2)T} = 2d_Q\big((2k+1)T\big)\frac{T}{\pi}(-1)^k. \tag{26}$$

From this result we can obtain $d_Q\big((2k+1)T\big)$. The original digital sequence can be reconstructed from the d_I and d_Q values.

There is a modification of MSK called Gaussian Minimum Shift Keying (GMSK) that is used in a number of systems. GMSK follows the same process as MSK except that the modulating digital signals are smoothed with a Gaussian filter. The process is pictured in Figure 14.

The smoothing decreases the bandwidth and the inter-channel interference, but it increases the inter-symbol interference. The demodulation of a GMSK signal is similar to that used for an MSK signal. GMSK is used in a number of cellular devices as well as Bluetooth devices for short range connectivity.

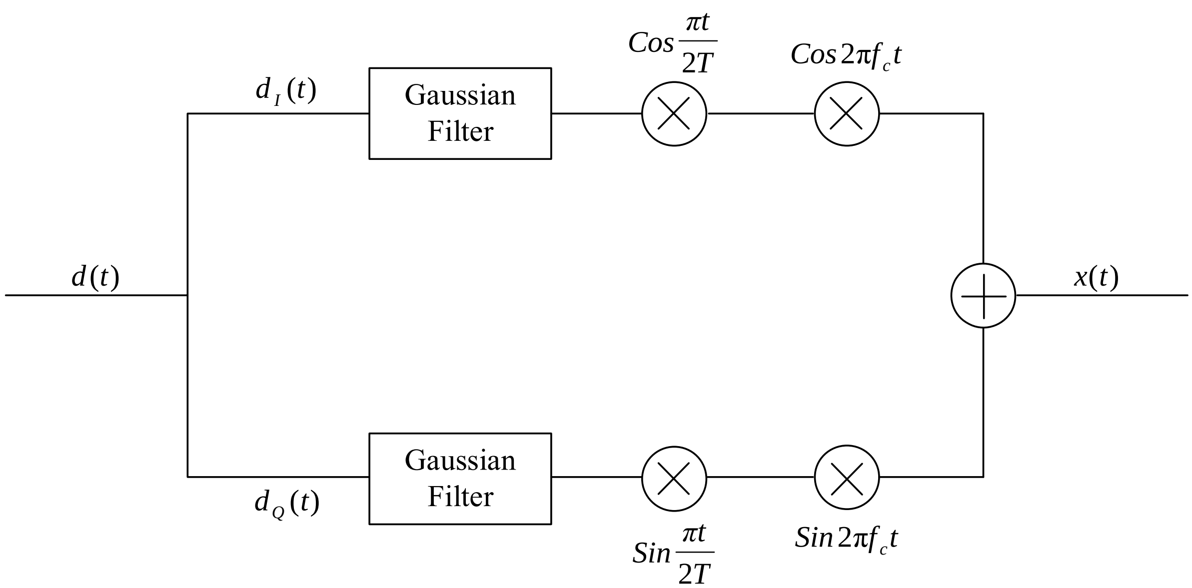

Figure 14: Generation of a GMSK signal

3.6 Quadrature Amplitude Modulation (QAM)

Amplitude and phase modulation are often combined. In Quadrature Amplitude Modulation (QAM) the signal has the form

$$s(t) = I(t)\cos(2\pi f_c t) + Q(t)\sin(2\pi f_c t) \tag{27}$$

where $I(t)$ and $Q(t)$ are constant on each symbol period. The signal given in equation (27) can be written in the alternate form

$$s(t) = a(t)\cos\bigl(2\pi f_c t - \phi(t)\bigr) \tag{28}$$

where $a(t) = \sqrt{I^2(t) + Q^2(t)}$, $\cos\phi = I(t)/a(t)$, and $\sin\phi = Q(t)/a(t)$. In this form we see that QAM can be considered as an amplitude and phase modulation scheme. Phase modulation schemes such as QPSK are special cases of QAM. The I-Q representation in equation (27) is usually preferred over the amplitude-phase representation since the signal is easier to generate in this form. The I and Q values corresponding to different symbols are usually represented by a rectangular grid of points in the I-Q plane. The QAM scheme based on the diagram in Figure 15 consists of 16 points and is called 16-QAM. This modulation scheme encodes 4 bits per symbol.

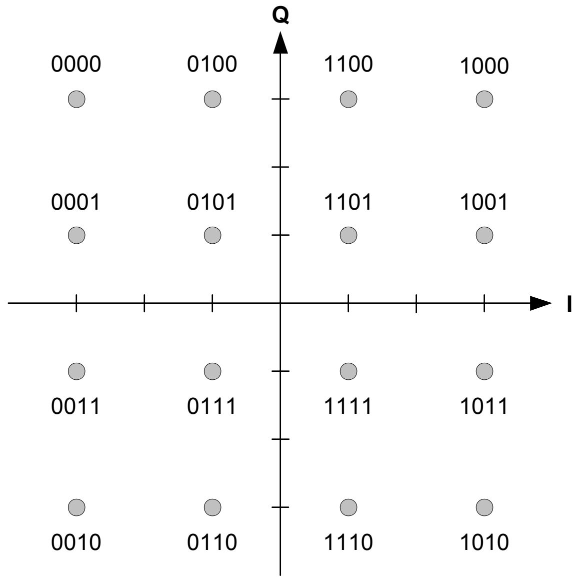

Figure 15: Constellation diagram for 16-QAM modulation.

QAM signals can be demodulated using a coherent scheme like that used for phase shift keying. Multiplying the signal $s(t)$ by $\cos(2\pi f_c t)$ and using trigonometric formulas for double angles, we get

$$\begin{aligned} s(t)\cos(2\pi f_c t) &= I(t)\cos^2(2\pi f_c t) + Q(t)\sin(2\pi f_c t)\cos(2\pi f_c t) \\ &= \tfrac{1}{2}I(t)[1+\cos(4\pi f_c t)] + \tfrac{1}{2}Q(t)\sin(4\pi f_c t). \end{aligned} \tag{29}$$

Passing $s(t)\cos(2\pi f_c t)$ through a low pass filter, we get $\frac{1}{2}I(t)$. Similarly, multiplication of the signal by $\sin(2\pi f_c t)$ gives

$$\begin{aligned} s(t)\sin(2\pi f_c t) &= I(t)\cos(2\pi f_c t)\sin(2\pi f_c t) + Q(t)\sin^2(2\pi f_c t) \\ &= \tfrac{1}{2}I(t)\sin(4\pi f_c t) + \tfrac{1}{2}Q(t)[1-\cos(4\pi f_c t)]. \end{aligned} \tag{30}$$

Passing $s(t)\sin(2\pi f_c t)$ through a low pass filter, we get $\frac{1}{2}Q(t)$. The symbol of the closest point to (I, Q) in the constellation diagram is taken as the decoded symbol.

The modulation schemes QPSK, 16-QAM, 64-QAM (8×8 grid), and 256-QAM (16×16 grid) are widely used in cable/DSL modems. The higher order QAM modulation schemes provide higher bit rates, but require higher signal-to-noise ratios in order to work correctly. Often the order of the modulation scheme is chosen adaptively depending on the quality of the transmission channel.

3.7 Orthogonal Frequency Division Multiplexing (OFDM)

Orthogonal Frequency Division Multiplexing (OFDM) is probably the most complicated of the methods described in this paper, but it is widely used in wireless devices. OFDM is based on the Discrete Fourier Transform (DFT). For a sequence of complex values $x_0, x_1, \ldots, x_{N-1}$ the Discrete Fourier Transform $X_0, X_1, \ldots, X_{N-1}$ of this sequence is defined by

$$X_n = \sum_{m=0}^{N-1} x_m e^{-i2\pi mn/N}. \tag{31}$$

It can be shown that

$$x_m = \frac{1}{N} \sum_{n=0}^{N-1} X_n e^{i2\pi mn/N}. \tag{32}$$

This relation is called the inverse DFT. The DFT and inverse DFT can be computed rapidly using Fast Fourier Transform (FFT) algorithms or devices. The sequence $\{x_m\}$ is usually considered to be in the time domain, and the sequence $\{X_n\}$ is usually considered to be in the frequency domain. If T is the symbol period, we can write equations (31) and (32) as follows

$$x(t_m) = \frac{1}{N} \sum_{n=0}^{N-1} X(f_n) e^{i2\pi f_n t_m} = x_m \tag{33a}$$

$$X(f_n) = \sum_{m=0}^{N-1} x(t_m) e^{-i2\pi f_n t_m} = X_n \tag{33b}$$

where $t_m = m\Delta t$, $f_n = n\Delta f$, $\Delta t = T/N$, and $\Delta f = 1/T$. Thus, the $x(t_m)$ values can be considered as sampled values of the time function $x(t)$ given by

$$x(t) = \frac{1}{N} \sum_{n=0}^{N-1} X(f_n) e^{i2\pi f_n t}. \tag{34}$$

It can be shown that the functions $\{e^{i2\pi f_n t}\}$ are orthogonal over each symbol period, i.e.,

$$\int_{kT}^{(k+1)T} e^{i2\pi f_m t} e^{-i2\pi f_n t}\, dt = 0 \qquad m \neq n. \tag{35}$$

In the OFDM method the symbols are represented by complex values that are then used for the frequency components X_n in equation (32). The constellation diagrams introduced in connection with the PSK and QAM modulation schemes can be looked upon as defining mappings between symbols and complex values. Using one of the constellation diagrams we assign the complex number corresponding to the first symbol to X_0, the complex number corresponding to the second symbol to X_1, and so on until the complex number corresponding to the N-th symbol is assigned to X_{N-1}. We then obtain a sequence of time values $\{x_m\}$ by means of the inverse DFT defined in equation (32). This sequence is in general complex. The real and imaginary parts of this sequence can be converted to continuous functions of time on the interval T using a digital-to-analog converter (DAC). The real and imaginary functions of time on each symbol period are often used to modulate carrier signals $cos(2\pi f_c t)$ and $sin(2\pi f_c t)$ respectively. A diagram of the process is shown in Figure 16.

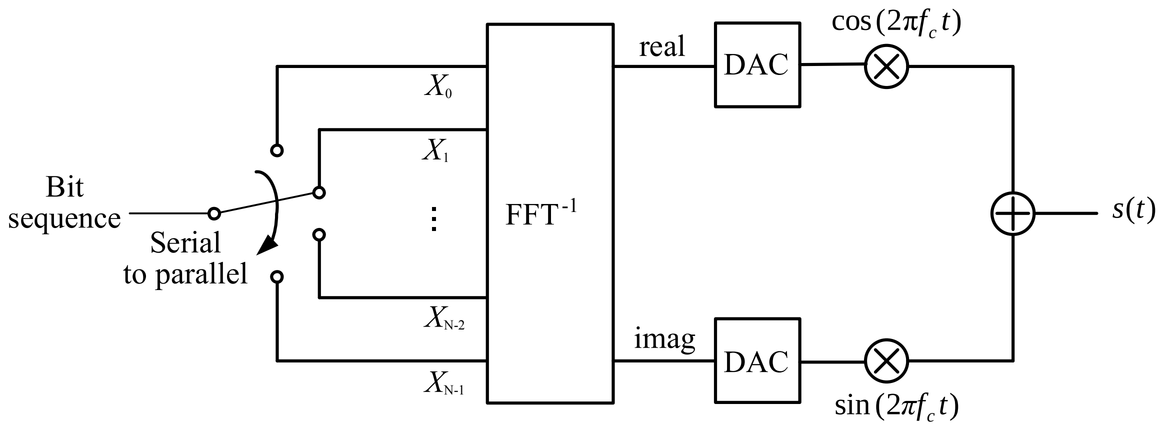

Figure 16: Diagram showing the steps in Orthogonal Frequency Division Multiplexing.

We can think of the frequency components X_n as representing N parallel channels. The first N symbols produce a time signal on the first symbol period, the next N symbols produce a time signal on the second symbol period, ... Thus, the first step in the OFDM process amounts to converting a linear sequence of bits into N parallel sequences. Since this parallelism allows a large number of bits to be handled at each step, the symbol period used is often taken to be larger than is used in other methods.

To demodulate an OFDM signal we first shift the signal down to baseband using the same technique used for QPSK. The signal $s(t)$ is given by

$$s(t) = u(t)\cos(2\pi f_c t) + v(t)\sin(2\pi f_c t) \tag{36}$$

where $u(t)$ and $v(t)$ are the real and imaginary parts of the complex signal $x(t)$. Multiplying $s(t)$ by $\cos(2\pi f_c t)$, we obtain

$$\begin{aligned} s(t)\cos(2\pi f_c t) &= u(t)\cos^2(2\pi f_c t) + v(t)\sin(2\pi f_c t)\cos(2\pi f_c t) \\ &= \tfrac{1}{2}u(t)[1 + \cos(2\pi 2 f_c t)] + \tfrac{1}{2}v(t)\sin(2\pi 2 f_c t). \end{aligned} \tag{37}$$

Thus, passing $s(t)\cos(2\pi f_c t)$ through a low-pass filter gives us $\frac{1}{2}u(t)$. Similarly,

$$s(t)\sin(2\pi f_c t) = \tfrac{1}{2}u(t)\sin(2\pi 2 f_c t) + \tfrac{1}{2}v(t)[1 - \cos(2\pi 2 f_c t)]. \tag{38}$$

Thus, passing $s(t)\sin(2\pi f_c t)$ through a low-pass filter gives us $\frac{1}{2}v(t)$. Having $u(t)$ and $v(t)$, we can form $x(t) = u(t) + iv(t)$. Sampling $x(t)$ at the times $0, T/N, \ldots, (N-1)T/N$, we obtain $x_0, x_1, \ldots, x_{N-1}$. Using the DFT defined by equation (31), we can compute $X_0, X_1, \ldots, X_{N-1}$. The mapping represented by the constellation diagram can then be used to decode the symbols. Finally, the symbols can be strung together to obtain the transmitted bit sequence.

OFDM is widely used in wireless communication devices. For example, wireless routers conforming to the specifications IEEE 802.11a, IEEE 802.11g, and the proposed IEEE 802.11n make use of OFDM. Some of the advantages of the OFDM method are that it makes very efficient use of the frequency spectrum, it has very little interference between symbols, and it has greater resistance to multipath distortion than most other methods.

3.8 Frequency Hopping Spread Spectrum (FHSS)

Frequency Hopping Spread Spectrum (FHSS) is a technique used in many wireless devices in order to spread the transmitted information over a wider frequency band. There are several reasons why this spreading is desirable. (1) Spread-spectrum signals are highly resistant to narrow-band noise and jamming. (2) Spread-spectrum signals are difficult to intercept. They tend to look like background noise. (3) Spread-spectrum signals can share a frequency band with many types of conventional transmissions with minimal interference.

In FHSS the carrier frequency hops over a predetermined but random-like sequence of frequencies. The pattern of frequency hops must be known by both the transmitter and the receiver. If the period of the frequency hopping is shorter than the symbol period (fast hopping), then there is a built in redundancy since each symbol will occur in more than one carrier. This redundancy is useful in reducing the effects of narrow band noise and jamming. The main disadvantage to fast hopping is that coherent detection is difficult and seldom used. If the period of the frequency hopping is greater than the symbol period (slow hopping), then coherent detection schemes are feasible and the implementation is easier. The disadvantage of slow hopping is that narrow-band noise can destroy one or more bits of information making error correcting codes almost a necessity. Any of the techniques discussed previously can be used to modulate the carrier signals. FHSS allows multiple devices with different hopping codes to operate simultaneously. Since the hopping pattern is random like, this technique offers some security against eavesdropping. Frequency Hopping Spread Spectrum (FHSS) is used in a number of wireless devices including Bluetooth devices.

3.9 Direct-Sequence Spread Spectrum (DSSS)

Direct Sequence Spread Spectrum (DSSS) is another technique for spreading the digital information over a wider frequency band. Suppose the bit sequence is encoded in a signal $m(t)$ consisting of a sequence of square pulses having period T and amplitudes ± 1. Let $d(t)$ be another signal consisting of ± 1 amplitude square pulses but having a much smaller period T/N for some large N. Typically N is ten or more. The predetermined sequence of ± 1 values in $d(t)$ is chosen to have a random-like pattern. The signal $s(t)$ is obtained by multiplying $m(t)$ and $d(t)$, i.e.,

$$s(t) = m(t)d(t). \tag{39}$$

The signal $d(t)$ must be known by both the transmitter and receiver. Often the pattern of plus and minus ones in $d(t)$ is generated by an algorithm based on some shared seed. The signal $s(t)$ has a noise-like appearance and thus provides some protection against eavesdropping. Figure 17 shows an example of the various signals that make up a DSSS signal.

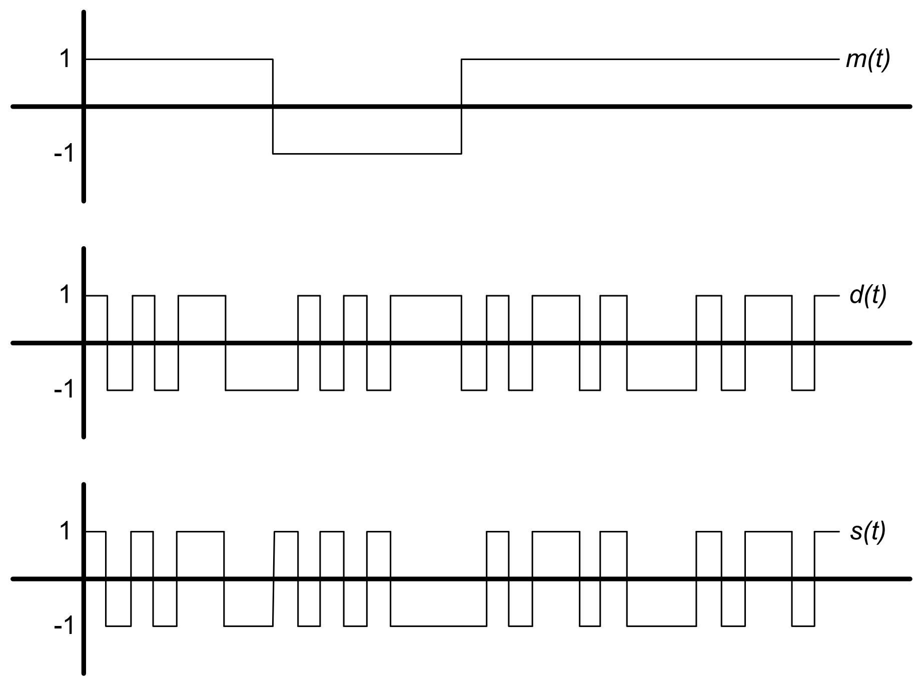

Figure 17: Example of Direct Sequence Spread Spectrum.

Since $d(t)$ only takes on the values ± 1, multiplying d by itself produces a function that is identically one. Therefore, multiplication of equation (39) by $d(t)$ gives $m(t) = s(t)d(t)$. Thus, decoding of this process is very simple. There is an inherent redundancy in DSSS that is useful in overcoming narrow band noise and jamming. Since the resulting signal in DSSS is also digital, this method can also be combined with any of the modulation techniques discussed previously. DSSS is used in a number of wireless applications. For example, wireless routers conforming to the specification IEEE 802.11b use DSSS and wireless routers conforming to the specification IEEE 802.11g use DSSS and OFDM.

3.9.1 Code Division Multiple Access (CDMA)

Analog cell phones divide the frequency band up into a number of smaller sub bands called channels and assign each user a different channel. Digital cell phones commonly use a technique called Code Division Multiple Access (CDMA) that allows each user to use the whole frequency band. CDMA is a form of DSSS in which each user is assigned a different pseudo random signal to multiply the original digital signal. The pseudo random signals assigned to different users are chosen so that the correlation between them is small. The DSSS signals are used to modulate a high frequency carrier. QPSK is often used for this modulation.