Lecture 17: Modulation Techniques for Mobile Radio

Types of Digital Modulation

Quadrature Amplitude Modulation (QAM)

Often, we need to transmit more bits per symbol that what simple digital modulation techniques described in previous lecture can practically do. We have seen that M-ary PSK can be used for transmitting 3 bits per symbol or higher bits per symbol but once the number of bits per symbol increases, the efficiency (probability of error) of the modulation increases significantly because the points of the constellation are confined to be a circle of a specific radius. Redistributing the constellation points in a different type of modulation gives better performance in terms of probability of error at the expense of a more complicated system and a no longer constant envelope as it is the case with all PSK modulations.

Consider the two basis pulses given below

$$\phi_1(t) = \sqrt{\frac{2}{T_s}} \cdot \cos\left(2\pi f_c t\right)$$

$$\phi_2(t) = \sqrt{\frac{2}{T_s}} \cdot \sin\left(2\pi f_c t\right)$$

We can generate M signals such that each carries n bits ($n = \log_2 M$) by having a scaled combination of these two basis pulses as follows. Consider the signals obtained by

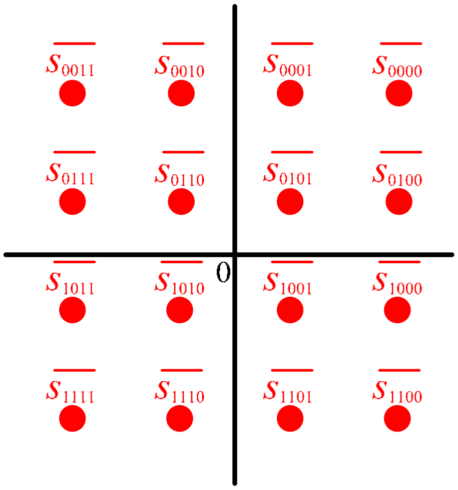
$$\begin{aligned} s_i(t) &= \sqrt{\frac{2E_{\min}}{T_s}} \cdot a_i \cdot \cos\left(2\pi f_c t\right) + \sqrt{\frac{2E_{\min}}{T_s}} \cdot b_i \cdot \cos\left(2\pi f_c t\right) \qquad \mathit{for} \qquad 0 \le t \le T_s \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, 2, \ldots, M \\ &= \sqrt{E_{\min}} \cdot a_i \cdot \phi_1(t) + \sqrt{E_{\min}} \cdot b_i \cdot \phi_2(t) \end{aligned}$$

where $E_{\min}$ is the minimum energy of any of the pulses and $\left\{a_i, b_i\right\}$ is a pair of coefficients that are given by a matrix of pairs depending on the desired M-ary QAM modulation scheme (or sometimes called M-QAM for short). For example, 16-QAM will have a matrix for $\left\{a_i, b_i\right\}$ given by

$$\left\{a_i, b_i\right\} = \begin{bmatrix} (-3,3) & (-1,3) & (1,3) & (3,3) \\ (-3,1) & (-1,1) & (1,1) & (3,1) \\ (-3,-1) & (-1,-1) & (1,-1) & (3,-1) \\ (-3,-3) & (-1,-3) & (1,-3) & (3,-3) \end{bmatrix}$$

So in 16-QAM, each 4 bits are combined and based on the pattern of these 4 bits (16 possibilities), the corresponding pair $\{a_i, b_i\}$ is selected from the above matrix and the pulse the corresponding pulse for these 4 bits is generated and transmitted.

The constellation of above 16-QAM is shown below:



More complicated constellations are obtained for higher levels of QAM. Once the pulses have been transmitted, the received pulses can be projected along the two dimensions to obtain a point on the constellation of each of the received pulses. The pulse assumed to have been transmitted is the one with the point in the constellation that is closest to that of the received pulse. QAM is widely used in digital communication systems including the V.92 56k bps modem.

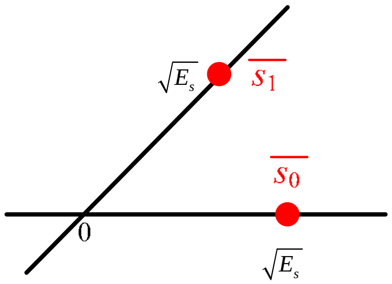
Binary Frequency Shift Keying (BFSK)

The third major type of digital modulation is digital frequency modulation. Here, different pulses will have different frequency. Since signals have different frequencies that may be of any value, a pulse cannot be represented in terms of another one, yet the pulses may not be orthogonal depending on the frequencies and pulse durations. In this digital modulation technique, we transmit one of two pulses for each bit:

1) for logic “0”, we transmit $s_0(t) = \sqrt{\frac{2E_s}{T_s}} \cos(2\pi f_0 t) \qquad \text{for} \qquad 0 \le t \le T_S$

2) for logic “1”, we transmit $s_1(t) = \sqrt{\frac{2E_s}{T_s}} \cos(2\pi f_1 t) \qquad \text{for} \qquad 0 \le t \le T_S$

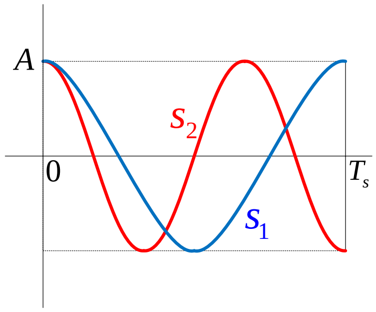
This configuration requires two dimensions to represent the transmitted pulses since non of the pulses is a linear multiple of the other.



The two symbols are generally not orthogonal. The angle between the two pulses depends on the two frequencies of the two pulses. If the difference between the two frequencies satisfies a specific criterion, the two pulses become orthogonal. BFSK signals can be generated by having two oscillators running at the desired frequencies and the output signal is switched between them depending on the bit we would like to transmit at a specific time. This results in a signal with phase discontinuities whenever we transmit a “0” after a “1” or vice versa.

Continuous-Phase Frequency Shift Keying (CP-FSK)

This is a special case of the FSK modulation where we try to make the phase of the FSK signal continuous by enforcing the orthogonality of the transmitted pulses. The continuity of the phase reduces the tails of the spectrum of the signal and therefore reduces the bandwidth requirements significantly. There is a minimum frequency different between the two frequencies of FSK pulses to insure that the two signals are orthogonal and guarantee continuous phase. Consider the following two pulses defined over the time period $0 \le t \le T_S$. One of the pulses makes an integer number of periods (the S_1 pulse makes 1 period) over the pulse period T_S while the other makes an integer number plus one half periods (S_2 pulse makes 1.5 periods) in the same period T_S.



In this case, the two signals can be written as:

$$s_1(t) = A\cos\left(2\pi \frac{1}{T_s} t\right)$$

$$s_2(t) = A\cos\left(2\pi \frac{3}{2T_s} t\right)$$

We can easily verify that the two pulses are orthogonal over the period $0 \le t \le T_S$. It is clear that the frequency difference between the two signals is $\Delta f = f_2 - f_1 = \frac{1}{2T_s}$.

In general, two pulses that have M periods and $M + \frac{n}{2}$ periods for any integer n over the pulse period T_S are orthogonal. Having, In fact, any pulses with frequency differences. This means that the minimum frequency difference $\Delta f_{\min}$ between orthogonal pulses is

$$\Delta f_{\min} = \frac{1}{2T_s}$$

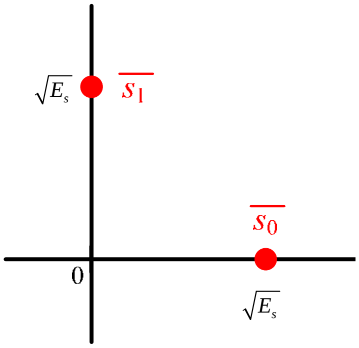
and frequency differences that are multiples of this minimum frequency difference also result in orthogonal pulses.

Minimum Shift Keying (MSK) [Binary Orthogonal Frequency Shift Keying]

In this FSK modulation technique, we transmit one of the following pulses for each bit:

1) for logic "0", we transmit $s_0(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi\left[f_c - \frac{1}{4T_s}\right]t\right)$ for $0 \le t \le T_S$
2) for logic "1", we transmit $s_1(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi\left[f_c + \frac{1}{4T_s}\right]t\right)$ for $0 \le t \le T_S$

This insures that the frequency difference between the two pulses satisfies the minimum frequency difference of $\Delta f_{\min} = \frac{1}{2T_s}$. This configuration requires two dimensions to represent the constellation of the modulation.



The basis pulses for this modulation are

$$\phi_0(t) = \sqrt{\frac{2}{T_s}} \cos\left(2\pi\left[f_c - \frac{1}{4T_s}\right]t\right) \qquad \text{for} \qquad 0 \le t \le T_S$$

$$\phi_1(t) = \sqrt{\frac{2}{T_s}} \cos\left(2\pi\left[f_c + \frac{1}{4T_s}\right]t\right) \qquad \text{for} \qquad 0 \le t \le T_S$$

For more complicated MSK modulations (when a higher power of two number of pulses are used), the modulation requires higher bandwidth and each pulse requires an independent orthogonal dimension to represent it.

Exercise 1:

Consider the MSK modulation given above with $T_s = 2s$, $E_s = 1$, $f_c = 5\,\text{Hz}$, and the frequency difference between the two pulses is equal to $\Delta f_{\min} = \frac{1}{2T_s}$. Clearly,

$$\phi_0(t) = \cos\left(2\pi\left[5 - \frac{1}{8}\right]t\right) \qquad \text{for} \qquad 0 \le t \le 2$$

$$\phi_1(t) = \cos\left(2\pi\left[5 + \frac{1}{8}\right]t\right) \qquad \text{for} \qquad 0 \le t \le 2$$

For each of the following signals that were detected over the period $0 < t \le 2$, determine the value of the bit that was most likely transmitted:

i) $g(t) = \cos(2\pi 5t)$

ii) $g(t) = \cos(2\pi(5.5)t)$

iii) $g(t) = e^{-t}$

iv) $g(t) = \sin(2\pi(5.125)t)$

Hint: Determine the projection of each signal on the two dimensions using the basis functions. Plot the point that corresponds to the projection along the two axis and determine which transmitted pulse is closer to the received pulse in each case.

Gaussian Minimum Shift Keying (GMSK)

This is a modification to the Minimum Shift Keying (MSK). The fact that the phase in MSK is continuous from one pulse to the other significantly reduces the bandwidth requirements compared to a non-continuous phase modulation. However, the frequency changes abruptly from one pulse to another in MSK. Forcing the frequency to change smoothly from one pulse to another even improves the frequency characteristics of the modulation technique.

The smoothing of the frequency is done by smoothing the digital message signal. Consider a binary digital signal which is used in a MSK modulation system such that the frequency of the transmitted pulse is proportional to the message signal. Although the phase is continuous for this modulation, the frequency either has a low value (for logic "0", for example) or a high value (for logic "1"). To make the frequency also continuous, the message signal is passed through a linear Gaussian filter. Passing a binary signal through a smooth filter like one with a Gaussian-shaped impulse response smoothes the edges of the pulse since the output of the filter is the convolution of the binary signal with a the Gaussian pulse. The wider the Gaussian pulse the smoother the output of the filter becomes at the expense of having a shorter period of the digital signal remaining with the same frequency. The impulse response of a Gaussian filter would be

$$h_G(t) = \frac{\sqrt{\pi}}{\alpha} e^{-\frac{\pi^2}{\alpha^2}t^2}$$

where α is a parameter that is related to the spread of the spectrum of the GMSK modulated signal. That is, the smaller the value of α is, the closer the shape of the impulse response of the filter becomes a delta function, which means that less smoothing of the frequency variations occurs, and the closer the GMSK becomes to the regular MSK. Increasing the value of α provides more frequency smoothing and the spectrum of the GMSK becomes narrower.

In GMSK, the MSK signal is generated using the output of the Gaussian filter as the message signal rather than the original message signal. So, the frequency of the generated FM modulated signal changes smoothly in proportion to the output of the Gaussian filter. The GMSK is exactly the modulation used in some wireless systems including the Global System for Mobile telecommunications (GSM).

Difference in Spectrum of Different Digital Modulation Techniques

This point can be summarized as follows:

- The spectrum of a specific digital modulation is spread over a range of frequencies that theoretically extends from $-\infty$ to ∞. However, if we consider the bandwidth of the main lobe as the bandwidth of the signal, signals with discontinuities (ASK, PSK, and general types of FSK) have relatively wide bandwidths as a result of discontinuities.
- MSK is efficient in the sense that the phase is continuous, however the minimum frequency difference must be satisfied which requires a significant bandwidth especially for fast data transmission.
- Increasing the number of symbols of a modulation algorithm without increasing the number of dimensions required to geometrically represent the different transmitted pulses generally does not increase the bandwidth of the modulation algorithm. For example, QPSK and 8-ary PSK require the same bandwidth. So, effectively, the required bandwidth for transmitting each bit per second is reduced (lower Hz/bps).

Lecture 17: Modulation Techniques for Mobile Radio

Spread Spectrum Modulation

An important type of modulation that is gaining more importance is called spread spectrum modulation. The main concept of spread spectrum is to start with a narrowband message signal (for example a voice signal with bandwidth 5 kHz) and spread its spectrum such that it will effectively occupy 1 MHz of bandwidth. This modulation practically does not waste bandwidth as more than 1 voice signal can be transmitted in the same 1 MHz without interference between the different signals. There are to flavors of spread spectrum, which are described briefly next.

Frequency Hopping Spectrum Modulation (FH-SSM)

In this type of spread spectrum, a simple process is used to spread the spectrum of a narrowband signal. Consider for example a message signal that requires 1 MHz of bandwidth to be transmitted as a PSK modulated signal. In FH-SSM, the PSK signal would be generated at a specific carrier frequency for 1 ms, for example. The carrier frequency would be changed then another value for another 1 ms, and the process of changing the carrier frequency continues over a large number of frequencies (1000 different carrier frequencies, for example) in what appears to be a random order. This process is called frequency hopping because the carrier frequency keeps hoping from one value to another every short period of time which spreads the spectrum of the original modulated signal over a very wideband range (around 1000 MHz in this example). The receiver must know the sequence of hoping over frequencies to be able to synchronize and demodulate the signal. If multiple transmitters and multiple receivers exist in the same geographical region, they can use the same frequency channel without interference by using different frequency hopping sequences such that no two transmitters hop to the same carrier frequency at the same time. All the receivers need to have is the hopping sequence of the desired signal.

Some of the advantages of this modulation method are:

1. Higher security and lower possibility of detection compared to regular non-spread spectrum modulations
2. Immunity to jamming (it is more difficult for an enemy to jam the signal and prevent the receiver from receiving it)

Direct-Sequence Spread Spectrum Modulation (DS-SSM)

In this spread spectrum modulation technique, the transmitted signal is spread in a completely different way. Consider for example binary data that requires 1 MHz to be modulated using one of the type of digital modulation techniques studied before and transmitted, dividing each bit into L smaller bits (called chips in SSM terminology) will result in the new signal requiring L MHz of bandwidth. The concept of the DS-SSM is to spread multiple digital sequences using orthogonal spreading sequences. Each spreading sequence divides a bit of data into L chips (so the L chips all carry one bit of information) and the different spreading sequences are orthogonal over the period of one bit.

Since the different spreading sequences are all orthogonal, it can easily be shown that multiplying the combined signal containing all received signals by one of the spreading sequences and integrating over the period of each bit produces the original bits of the signal that were spread using that spreading sequence and bits of the other signals are eliminated. Because the spreading sequence effectively increases the transitions in the message signal by the number of chips L in the spreading sequence, the bandwidth of the transmitted signal is L times the bandwidth of the original message signal. For example, consider the following set of signals (Column 1) containing 5 bits each. These signals are spread using the three orthogonal spreading sequences (Column 2). Note that the duration of each spreading sequence is the duration of one bit. The first spreading sequence happens to be constant for the whole period of the bit. The spread signals are shown afterwards (Column 3) where generally they will have wider bandwidth than the original signals. The spread signals are then combined as they are transmitted and received by an antenna (Column 4). Upon multiplying each bit of the combined received signal by each of the spreading sequences and integrating over the bit period, which is the de-spreading process, we get the original bit sequences (Column 5).