Lecture 8: Digital Modulation in Wireless Communication Systems

8.1. Digital Signal Modulation: Main Definitions

As was mentioned in Lecture 7, *modulation* is the process where the baseband message information is added to the bandpass carrier. In *digital modulation* the digital beam stream is transmitted as a *message*, and then is converted into the analog signal of type $a(t)\cos(\omega t + \theta)$ that modulates the digital bit stream into a RF carrier. The analog signal has amplitude *a*(*t*), frequency $f = \omega / 2\pi$, and phase $\theta = \omega t$. Changing these three characteristics, we can formulate three kinds of digital modulation. They are:

Amplitude shift keying (ASK) for phase and frequency kept constant;
Frequency shift keying (FSK) for amplitude and phase kept constant;
Phase shift keying (PSK) for amplitude and frequency kept constant.

In so-called hybrid modulation methods we can use combinations of these three kinds of modulation. Namely, for frequency is constant, but amplitude and phase are not constat, we have quadrarure amplitude modulation (QAM).

Some modulation methods are linear, as binary phase shift keying (BPSK), quadrarure phase shift keying (QPSK), including $\pi/4-$QPSK, DQPSK, including $\pi/4-$DQPSK, and so on.

At the same time, FSK, as well as minimum shift keying (MSK) and Gaussian minimum shift keying (GMSK) are non-linear modulation methods.

We will give some examples later (full information you can find in [1, 4, 10]) but now will put a question what afvantages of digital modulation method compared with analog modulation method.

Because digital modulation offers many advantages over analog modulation, it often used in modern wireless communication systems. Some advantages include greater noise immunity and robustness to channel impairments, easier multiplexing of various forms of information (such as voice, data, and video), and greater security. Moreover, digital transmissions accommodate digital error-control codes, which detect and correct transmission errors, and support complex signal processing techniques such as source coding, encryption etc.

In digital wireless communication systems, the modulating signal, i.e., the message, can be represented as a time sequence of symbols or pulses, where each symbol has *m* finite states. Each symbol represents *n* bits of information, where $n = \log_2 m$ bits/symbol. Many digital modulation schemes are used in modern wireless communication systems, and many are in a stage to be introduced. Despite differences between techniques, each technique belonds to a family of related modulation methods.

<u>Advantages to use digital modulation</u>:

- Cost Effective - VLSI and DSP technologies
- Greater noise immunity
- Robustness to channel impairments
- Easier multiplexing of information (voice, data, video)

- Error detection and correction codes
- Signal encryption
- Equalization to improve overall performance.

What influences the choice of digital modulation?

Requirements:

- low BER at low received SNR
- good performance in multipath and fading conditions (see below)
- low bandwidth (BW)
- implementation
- cost effectiveness.

The power efficiency and bandwidth efficiency. The efficiency of each digital modulation technique depends on many factors. The main goal of such a modulation is to obtain in various multipath and fading conditions low bit error rate (BER) at low received SNRs, minimum of bandwidth occupation and so on. The performance of a modulation scheme is often measured in terms of its *power efficiency* and *bandwidth efficiency*. First term describes the ability of a modulation technique to preserve the fidelity of the digital message at low power levels. In a digital communication system, in order to increase noise immunity, it is necessary to increase the signal power. However, the amount by which the signal power should be increased to obtain a certain level of fidelity, i.e., an acceptable bit error probability, depends on the particular type of modulation employed.

The *power efficiency*, η_p , of a digital modulation scheme is a measure of how favorably this tradeoff between fidelity and signal power is made, and is often expressed as a *ratio of the signal energy per bit to noise power spectral density*, E_b / N_0 , required at the receiver input for a certain probability of error (say 10^{-5}).

The *bandwidth efficiency*, η_B , describes the ability of a modulation scheme to accomodate data within a limited bandwidth. In general, increasing the data rate implies decreasing the pulse width of a digital symbol, which increases the bandwidth of the signal. Bandwidth efficiency reflects how efficiently the allocated bandwidth is utilized and is defined as the ratio of the *throughput data rate (*bits per second*, bps) per Hertz* in given bandwidth. If R is the data rate in bits per second, and B is the bandwidth occupied by the modulated RF signal, then bandwidth efficiency is expressed as

$$\eta_B = \frac{R}{B} \text{ bps / Hz} \qquad (8.1)$$

The system capacity of a digital communication system is directly related to the bandwidth efficiency of the modulation scheme, since a modulation with a greater value of η_B will transmit more data in a given spectrum allocation. There is a fundamental upper bound on achievable bandwidth efficiency, which can be defined from well-known Shannon's channel coding theorem [1]. This theorem states that for an arbitrary snall probability of error, the maximum possible bandwidth efficiency is limited by the noise in the channel, and is given by the channel capacity formula:

$$\eta_{B\max} = \frac{C}{B} = \log_2\left(1 + \frac{S}{N}\right) = \log_2\left(1 + \frac{S}{N_0 B}\right) \tag{8.2}$$

where *C* is the channel capacity (in bps), *B* is the RF bandwidth, and S/N is the SNR.

Bandwidth and power spectral density of digital signals. The definition of *signal bandwidth* varies with context, and there is no single definition which covers all applications [2]. All definitions, however, are based on some measure of the *power spectral density* (PSD) of the signal. The PSD of a random signal $x(t)$ is defined as [3]

$$P_x(t) = \lim_{T\to\infty}\left(\frac{\left\langle |X_T(f)|^2 \right\rangle}{T}\right) \tag{8.3}$$

where $\langle \bullet \rangle$ denotes an ensemble average, $X_T(f)$ is the Fourier transform of $x_T(t)$, which is the truncated version of the signal $x(t)$, defined as

$$x_T(t) = \begin{cases} x(t) & -T/2 < t < T/2 \\ 0 & t \ \text{ elsewhere} \end{cases} \tag{8.4}$$

The power spectral density of a modulated (bandpass) signal is related to the power spectral density of its baseband complex envelope. If a bandpass signal $s(t)$ is represented as

$$s(t) = \mathrm{Re}[g(t)\exp(j2\pi f_c t)] \tag{8.5}$$

where $g(t)$ is the complex baseband envelope. Then the PSD of the bandpass signal is given by

$$P_s(f) = \frac{1}{4}\left[P_g(f + f_c) + P_g(-f - f_c)\right] \tag{8.6}$$

where $P_g(f)$ is the PSD of $g(t)$.

The *absolute bandwidth* of a signal is defined as the range of frequencies over which the signal has non-zero power spectral density. For symbols represented as rectangular baseband pulses, the PSD has a $(\sin f)^2 / f^2$ profile which extends over an infinite range of frequencies, and has an absolute bandwidth of infinity. A simpler and more widely accepted measure of bandwidth is the first *null-to-null bandwidth*. The null-to-null bandwidth is equal to the width of the main spectral lobe. A very often used measure of bandwidth which measures the dispersion of the power spectrum is the *half-power bandwidth*, which is defined as the interval between frequencies at which the PSD has dropped to half power, or 3 dB below the peak value. Half-power bandwidth is also called the *3 dB bandwidth*. According to definition adopted by the Federal Communication Committee, the occupied bandwidth is the band which consists 99% of the signal power: 0.5 percent of the signal power leaves above the upper band limit and 0.5 percent of the signal power leaves below the lower band limit.

Line coding. Digital baseband signals often use line codes. The most common codes for wireless communication are *return-to-zero* (RZ), *non-return-to-zero* (NRZ), and so-called the *Manchester* codes (see Fig. 8.1). All of these codes may either be *unipolar* (with voltage levels being either 0 or *V*) or *bipolar* (with voltage levels being either -*V* or *V*) (see Fig. 8.2). RZ code implies that the pulse returns to zero within every bit period. This leads to spectral widening, but improves timing synchronization. NRZ codes do not return to zero during a bit period; the signal stays at constant levels throughout a bit period. NRZ codes are more spectrally efficient than RZ codes, but offer poorer synchronization capabilities. The Manchester code is a special type of NRZ line code that is ideally suited for signaling that must pass through phone lines and other DC blocking circuits, as it has no dc component and offers simple synchronization. This code use two pulses to represent each binary symbol, and thereby provide easy clock recoverty since zero-crossings are guaranteed in every bit period.

Example: The RF bandwidth of the wireless network is $B = 30\ kHz$ and SNR within the link is 20 *dB*.
Find: The maximum theoretical data rate C that can be transmitted through such a link.
Compare it with USA Digital Cellular (USDC) standard of 48.6 *kbps*.

Solution: According to (8.2) and taking into account that S/N=20 dB=100, we get:

$$C = B \cdot \log_2\left(1 + \frac{S}{N}\right) = 30000\ \ Hz \cdot \log_2(1 + 100) = 199.5\ \ kbps$$

This rate is four time faster than in standard USDC for the given limit of S/N=20 *dB*.

8.2. Pulse Shaping Technique

During their pass through a band-limited communication channel, rectangular pulses, which are presented in form of the message of symbols or bits, will spread in time, and the pulse of each symbol will spread into the time interval of neighbour symbols. This effect causes so-called the *intersymbol interference* (ISI) and finally leads to an increased probability of the receiver making an error in detecting a symbol. One way to minimize the ISI is to increase the channel bandwids. However, wireless communication systems often operate with minimal bandwidth, and techniques that reduce the modulation bandwidth and suppress out-of-band radiation, while reducing ISI, are highly desirable. Since it is difficult to directly control the transmitter spectrum at RF frequencies, *spectral shaping* is done through baseband or intermediate frequency (IF) processing. There are some effective pulse-shaping techniques are used to simultaneously reduce the ISI effect and the spectral width of a modulated digital signal.

8.2.1. Nyquist criterion for ISI cancellation

Nyquist was the first who observed that the effect of ISI could be completely cancelled if the overall response of the communication system, including transmitter, receiver, and channel, is designed so that at every sampling instant at the receiver, the response due to all symbols except the current symbol is equal to zero [1]. If $h_{eff}(t)$ is the impulse response of the system, the Nyquist condition can be state as

$$h_{eff}(nT_s) = \begin{cases} K & n = 0 \\ 0 & n \neq 0 \end{cases} \quad (8.7)$$

where T_s is the symbol period, n is an integer, and K is a non-zero constant. The effective transfer function of the system can be represented by the δ-function, the pulse shape of a symbol, $p(t)$, the channel impulse response, $h_c(t)$, and the receiver impulse response, $h_r(t)$, as

$$h_{eff}(t) = \delta(t) * p(t) * h_c(t) * h_r(t) \quad (8.8)$$

Nyquist derived transfer functions $H_{eff}(f)$ which satisfy the conditions of (8.7). He based on the two assumptions: first, that $h_{eff}(t)$ should have a fast decay with a small magnitude near the sample values for $n \neq 0$; second, the channel is ideal, $h_c(t) = \delta(t)$, and it is possible to realize shaping filters at both the transmitter and receiver to produce the desired $H_{eff}(f)$. Considering the impulse response as

$$h_{eff}(t) = \frac{\sin(\pi t / T_s)}{(\pi t) / T_s} \quad (8.9)$$

it is clear seen that this response satisfies the Nyquist condition (8.7) for ISI cancelation. Nyquist ideal pulse shape for zero ISI is shown in Fig. 8.3.

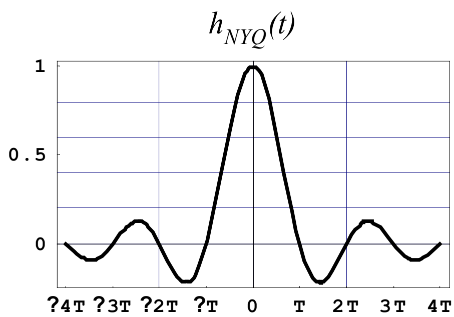

Fig. 8.3

Hence, if the overall communication system can be modeled as a filter with the impulse response described by (8.9), it is possible to completely eliminate the effects of ISI. The transfer function for such a filter can be obtained from (8.9) by taking the Fourier transform of the impulse response, that is,

$$H_{eff}(f) = \frac{1}{f_s} \Pi\left(\frac{f}{f_s}\right) \quad (8.10)$$

This transfer function corresponds to a rectangular "brick-wall" filter with absolute bandwidth $f_s / 2$, where f_s is the symbol rate. While this transfer function satisfies the zero ISI criterion with a minimum of bandwidth, there are practical difficulties in implementing it, since it corresponds to a noncausal system ($h_{eff}(t)$ exists for t<0) and is thus difficult to approximate.

Nyquist also provide that any filter with a transfer function having a rectangular filter of bandwidth $f_0 \geq 1/2T_s$, convolved with any arbitrary even function *Z(f)* with zero magnitude outside the passband of the rectangular filter, satisfies the zero ISI condition (8.9). Such a filter with zero ISI conditions can be expressed as

$$H_{eff}(f) = \Pi\left(\frac{f}{f_s}\right) \otimes Z(f) \tag{8.11}$$

where *Z(f)*=*Z(-f)*, and *Z(f)*=0 for $|f| \geq f_0 \geq 1/2T_s$. Expressed in terms of the impulse response, the Nyquist criterion states that any filter with an impulse response

$$h_{eff}(t) = \frac{\sin(\pi t / T_s)}{\pi t / T_s} z(t) \tag{8.12}$$

can achieve ISI cancellation. Filter which satisfies the Nyquist criterion is called Nyquist filter, the transfer function of which is shown in Fig. 8.4.

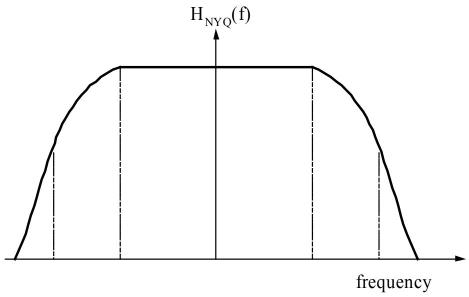

Fig. 8.4

8.2.2. Raised cosine filter

The often used pulse shaping filter, which satisfies the Nyquist criterion, is the raised cosine filter, the transfer function of which is given as [1]

$$H_{RC}(f) = \begin{cases} T_s & 0 \leq |f| \leq (1-\alpha)/2T_s \\ \dfrac{T_s}{2}\left[1+\cos\left(\dfrac{\pi|f| - \frac{1}{2T_s} + \alpha}{2\alpha}\right)\right] & (1-\alpha)/2T_s |<|f| \leq (1+\alpha)/2T_s| \\ 0 & |f| > (1+\alpha)/2T_s \end{cases} \tag{8.13}$$

where α is the rolloff factior which ranges between 0 and 1. When α=0, the raised cosine rolloff filter corresponds to a rectangular filter of minimum bandwidth (see Fig. 8.5). The corresponding impulse response of the filter can be obtained by taking the inverse Fourier transform of the transfer function, that is,

$$h_{RC}(t) = \frac{\sin(\pi t / T_s)}{(\pi t / T_s)}\left[\frac{\cos 2\pi\alpha t}{1-(4\alpha t)^2}\right] \tag{8.14}$$

As follows from formulas presented, as the rolloff factor α increases, the bandwidth of the filter also increases, and the time sidelobe levels decrease in adjacent symbol slots. This implies that increasing α increases the occupated bandwidth from $R/2$ to R for α=0 to α=1 (see Fig. 8.6). The symbol rate that can passed through a baseband raised cosine rolloff filter is given by [1]

$$R_s = \frac{1}{T_s} = \frac{2B}{1+\alpha} \tag{8.15}$$

where B is the absolute filter bandwidth. For RF systems, for example, the RF passband bandwidth doubles and

$$R_s = \frac{B}{1+\alpha} \tag{8.15a}$$

The cosine rolloff transfer function can be achieved by using identical $\sqrt{H_{RC}(f)}$ filters at the transmitter and receiver, while providing a matched filter for optimum performance in a flat fading channel. To implement the filter responses, as a rule, pulse shaping filters are used on the baseband data. Moreover, these filters are typically implemented for $\pm 6T_s$ about the t=0 point for each symbol. As an example, assume binary baseband pulses transmitted using a raised cosine rolloff filter with α=1/2. If the modulator stores 3 bits at a time, then there are 8 possible waveform states that may be produced at random for the group (see Fig. 8.7). If $\pm 6T_s$ is used to represent time span for each symbol (a symbol is the same as a bit in this case), then the time span of the discrete-time waveform will be $14T_s$.If, for example, the RF time waveform corresponds to the data sequence 1, 0, 1, then the optimal bit decision points occur at $4T_s$, $5T_s$, and $6T_s$, and the time dispersive nature of pulse shaping can be seen.

8.2.3. Gaussian pulse-shaping filter

It is also possible to use non-Nyquist techniques for pulse shaping. A Gaussian pulse-shaping filter is often used for such techniques for special type of digital modulation, known as minimum shift keying modulation (MSK) and is applicable for power non-linear amplifiers. This is very important question, because there is a dilemma for designers of digital communication systems that reducing the bandwidth of pulse shaping filter according to Nyquist requires linear amplifiers which are not power efficient. Gaussian pulse-shaping filters are well suited for power nonlinear amplifiers and resolve this problem. Unlike Nyquist filters which have zero-crossings at adjacent symbol peaks and a truncated transfer function, the Gaussian filter has a smooth transfer function with no zero-crossings. The Gaussian lowpass filter has a transfer function given by

$$H_G(f) = \exp\left(-\alpha^2 f^2\right) \tag{8.16}$$

The parameter is related to B, the 3-dB bandwidth of the baseband shaping filter (the impulese response of Gaussian filter gives rise to a transfer function which is highly dependent upon the 3-dB bandwidth),

$$\alpha = \frac{\sqrt{2\ln 2}}{B} = \frac{1.1774}{B} \tag{8.17}$$

As α increases, the spectral accupancy of the Gaussian filter decreases and time dispersion of the applied signal increases (see Fig. 8.8). The impulse responce of the Gaussian filter is given by

$$h_G(t) = \frac{\sqrt{\pi}}{\alpha} \exp\left(-\frac{\pi^2}{\alpha^2} t^2 \right) \tag{8.18}$$

The Gaussian filter has a narrow absolute bandwidth (although not as narrow as a raised cosine rolloff filter), and has sharp cut-off, low overshoot, and pulse area preservation properties which make it very attractive for use in modulation techniwues that use nonlinear RF amplifiers and do not accurately preserve the transmitted pulse shape.

8.3. Geometric Representation of Digital Modulation Signals

8.3.1. Vector-space presentation of the digital modulated signal

For digital modulation schemes, the complex envelope $r(t)$ during the kth baud $kT \le t \le (k+1)T$ belonds to a finite set of waveforms $r_i(t)$, i=1, 2,..., M. In other words, the procedure of digital modulation involves choosing a particular signal waveform $r_i(t)$ from a finite set of possible signal waveforms (or symbols) based on the information bits applied to the modulator. If there are a total of M possible waveforms, the modulation signal set can be represented in the vector form, that is as the M points in a *vector space*:

$$\mathbf{r}_i = \{ r_{i1}, r_{i2}, r_{i3}, \ldots\ldots, r_{iN} \}, \quad i=1, 2, \ldots, M \tag{8.19}$$

Consequently, the waveforms $\{r_i\}|_{i=1}^{M}$ can be represented in terms of a set of basis functions $\{\phi_i\}|_{i=1}^{M}$ which are defined on the interval $[0, T]$ and are orthogonal to one another, that is,

$$\int_0^T \phi_i(t) \phi_j(t) dt = \delta_{ij} \tag{8.20}$$

where δ_{ij} is the Dirac delta function, $\delta_{ij} = 0,\ i \ne j;\ \delta_{ij} = \infty,\ i = j$. The parameter N is the dimensionality of the vector-space that is needed to represent the finite set of waveforms. This finite set of discrete signals represents the modulation signals $r_i(t)$ on a vector space as a linear combination of the basis signals $\{\phi_i\}|_{i=1}^{M}$ such that

$$r_i(t) = \sum_{j=1}^{N} r_{ij} \phi_j(t), \quad i = 1,\ 2, \ldots,\ M \tag{8.21}$$

where

$$r_{ij} = \int_0^T r_i(t)\phi_j^*(t)dt \tag{8.22}$$

Here $\phi_j^*(t)$ is complex conjugate to $\phi_j(t)$. Each of the basis signals is normalized to have unit energy, that is,

$$W = \int_0^T \phi_i^2(t)dt = 1 \tag{8.23}$$

The basis signals form a coordinate system for the vector space. As was shown in [1, 4, 8-10], the Gram-schmidt orthonormalization procedure provides a systematic way of obtaining the basis signals for a given set of waveforms. Let us present, first of all, an example of linear modulation.

Linear Modulation is modulation where amplitude of the transmitted signal varies linearly with the modulating digital signal *m*(*t*) according to the following low:

$$\begin{aligned} s(t) &= \mathrm{Re}\left[A\,m(t)\exp(j2\pi f_c t)\right] \\ &= A\left[m_R(t)\cos(2\pi f_c t) - m_I(t)\sin(2\pi f_c t)\right] \\ m(t) &= m_R(t) + jm_I(t) \end{aligned}$$

This kind of modulation has a good spectral efficiency, but linear amplifiers have *poor* power efficiency. At the same time, nonlinear amplifiers (which are used) have *good* power efficiency. Sidelobes are generated, increasing adjacent channel interference and
cancel the benefits of linear modulation.

a) First, let us consider the set of Amplitude Shift Keying (ASK) signals, where keying (or switching) the carrier sinusoid *on* if the input bit is “1” and *off* if “0” (so-called On-Off-Keying-OOK). This kind of modulation is shown in Fig. 8.9:

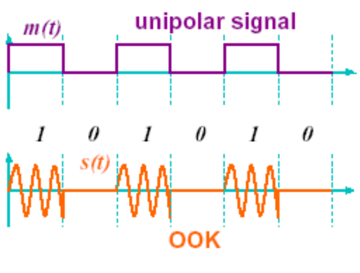

Fig. 8.9

b) Secondly, let us consider the set of Binary Phase Shift Keying (BPSK) signals (the meaning of this abbreviation see later) $r_1(t)$ and $r_2(t)$ given by

$$r_1(t) = \sqrt{\frac{2W_b}{T_b}}\cos(2\pi f_c t), \qquad 0 \le t \le T_b \tag{8.24a}$$

and

$$r_2(t) = -\sqrt{\frac{2W_b}{T_b}}\cos(2\pi f_c t), \qquad 0 \le t \le T_b \qquad (8.24b)$$

where W_b is the energy per bit, T_b is the bit period, and a rectangular pulse shape $p(t) = \Pi\left((t - T_b/2)/T_b\right)$ is assumed. Basis signals ϕ_i for this signal set in 2D-vector-space simply consists of a single wave form ϕ_1 where

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi f_c t), \qquad 0 \le t \le T \qquad (8.25)$$

Result of such kind of modulation is presented in Fig. 8.10.

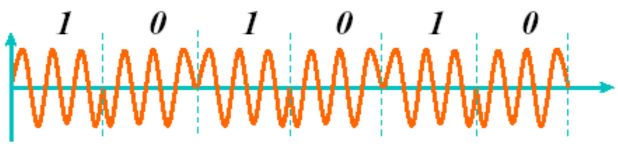

Fig. 8.10

Using this basis signal, the BPSK signal set can be represented as

$$\mathbf{r}_{iBPSK} = \left\{\sqrt{W_b}\,\phi_1(t), -\sqrt{W_b}\,\phi_1(t)\right\} \qquad (8.26)$$

Such a representation is called a *constellation diagram* which provides a graphical representation of the complex envelope of each possible symbol state. The *x*-axis of this diagram represents the in-phase component *I* of the complex envelope, and the *y*-axis represents the quadrature component *Q* of the complex envelope. The distance between signals on a constellation diagram relates to how different the modulation waveforms are, and how well a receiver can differentiate between all possible symbols when random noise is present. It should be noted that the number of basis signals will always be less than or equal to the number of signals in the set. The number of basis signals required to represent the complete modulation signal set is called the *dimension* of the vector space (in our example above it is two-dimensional (2D) vector space) . If there are many basis signals in the modulation signal set, then all of them must be orthogonal according to (8.20).

c) Thirdly, let us consider Differential Phase Shift Keying (DPSK) modulated signals:

$$d_k = \overline{m_k \oplus d_{k-1}}$$

Table 8.1

{mk}		1	0	0	1	0	1	1	0
{dk-1}		1	1	0	1	1	0	0	0
{dk}	1	1	0	1	1	0	0	0	1

The probability of DPSK modulated bit with energy E_b is

$$P_{e,DPSK} = \frac{1}{2}\exp\left(\frac{E_b}{N_0}\right)$$

Noncoherent PSK (no need of reference signal), easy to build, cheap, energy efficiency is inferior to PSK (3 dB less).

d) Forth example is Quadrature Phase Shift Keying (QPSK) signal. Its advantage is that it has twice the bandwidth efficiency or two bits at a time:

$$s_{QPSK}(t) = \sqrt{\frac{2E_s}{T_s}}\cos\left[2\pi f_c t + i\frac{\pi}{2}\right] \qquad 0 \le t \le T_s \qquad i = 0,1,2,3$$
$$= \sqrt{\frac{2E_s}{T_s}}\cos\left(i\frac{\pi}{2}\right)\cos(2\pi f_c t) - \sqrt{\frac{2E_s}{T_s}}\sin\left(i\frac{\pi}{2}\right)\sin(2\pi f_c t)$$

This signal set is shown geometrically in Fig. 8.11, where left diagram is for pure QPSK and the right one for $\pi/4-$QPSK modulation:

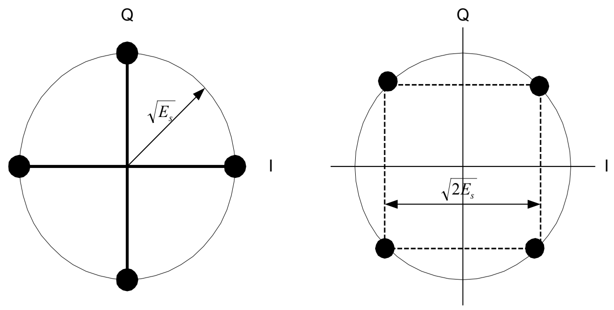

Fig. 8.11

The probability of QPSK modulated bit with energy E_b is

$$P_{e,QPSK} = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

e) Fifth example is a set of Frequency Shift Keying (FSK) signals, where keying (switching) the carrier sinusoid frequency into

$$f_c - \Delta f$$

if the input bit is “0”

and into

$$f_c + \Delta f$$

if input bit is “1”. Results of modulation are presented in Fig. 8.12.

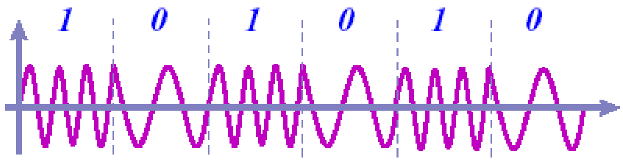

Fig. 8.12

Now we will consider a question on how different kinds of digital modulation affects different communication channels.

8.3.2. Digital modulation in AWGN channel

The simplest practical case of the wireless communication channel is an additive white Gaussian noise (AWGN) channel. When the modulated signal is transmitted over such a channel, the signal arriving at the demodulator is perturbed only by the addition of some noise. This channel applies only in the static case, where the terminal antennas and the obstructions are not in motion. Such a noise is white, that is, with a constant power spectral density (PSD) and Gaussian, i.e., has a normal distribution. The received signal in time t, $s(t)$, is then given by

$$s(t) = Ag(t) + n(t) \qquad (8.27)$$

where $n(t)$ is the noise waveform, $g(t)$ is the modulated signal and A is overall path loss, assumed not to vary in time. Equation (8.27) is a complex baseband signals (both modulated signal and noise) representation. If so, both the real and imaginary parts of noise $n(t)$ are zero mean, independent, real Gaussian processes, each of with a standard deviation of σ_n [9]. For digital signals, which consist of symbols with an individual energy W_s and a finite duration T_s. Then $W_s = A^2 T_s / 2$. Similarly, if the noise is contains within a bandwidth $B = 1/T_s$, and has power spectral density $PSD \equiv N_0$, then the mean noise power [1, 4, 8-10]

$$P_n = \frac{1}{2}\left\langle n(t) n^*(t) \right\rangle \equiv \sigma_n^2 = BN_0 = N_0 / T_s \qquad (8.28)$$

The signal-to-noise ratio (SNR) at the input of the demodulator is then (see above all definitions):

$$\gamma \equiv SNR = \frac{\left\langle A^2 g^2(t) \right\rangle}{2P_n} = \frac{A^2 \left\langle g^2(t) \right\rangle}{2\sigma_n^2} = \frac{W_s}{N_0} = \frac{A^2 T_s}{2N_0} \qquad (8.29)$$

It is usual to express the error rate performance of a digital system in terms of this parameter or in terms of the corresponding SNR per bit:

$$\gamma_b = \frac{\gamma}{m} = \frac{W_b}{N_0} \qquad (8.30)$$

where m is the number of bits per symbol. The SNR is the key parameter in calculating the digital modulation system performance in the AWGN channel. Let us, as above in an example presented above, calculate the bit nerror rate (BER) performance of binary phase shift keying (BPSK) signals in AWGN channel. We, first of all will consider the two-dimensional case of BPSK signals, that is, two signals which correspond to a binary 1 and 0. Their complex baseband presentation is

$$g_1 = \sqrt{\frac{2W_s}{T_s}}, \quad g_0 = -\sqrt{\frac{2W_s}{T_s}} \qquad (8.31)$$

where the duration of each symbol is T_s, the energy of of each symbol is W_s , and A=1. These signals consist of segments of carrier of duration T_s and a phase difference of 180^0, and can be represented in the BPSK constellation diagram (see Fig. 8.14). For 2D-vector space presentation it can be shown according to [1, 8-10] that the error rate performance of digital modulation scheme in AWGN channel with PSD of N_0 depends on the *Euclidean distance d* between the transmitted waveforms, corresponding to different transmitted bits according to [1, 8-10], and is determined by the probability of error

$$P_e = \int_0^\infty P_e(\gamma)p(\gamma)d\gamma \qquad (8.32)$$

where $P_e(\gamma)$ is the probability of error for a digital modulation at a specific value of SNR, γ , where $\gamma = A^2W_s / N_0$ and $p(\gamma)$ is a probability density function (PDF) due to AWGN channel. In [1, 4, 8-10] this probability of error was estimated through the Q-function:

$$P_e = Q\left(\sqrt{\frac{A^2d^2}{2N_0}}\right) \equiv Q\left(\sqrt{\frac{d^2}{2N_0}}\right) \qquad (8.33)$$

It is clear from Fig. 8.14 that in our example $d = 2\sqrt{W_s}$ and

$$P_r = Q\left(\sqrt{\frac{d^2}{2N_0}}\right) = Q\left(\sqrt{\frac{4W_b}{2N_0}}\right) = Q\left(\sqrt{2\gamma}\right) \qquad (8.34)$$

This result is illustrated in Fig. 8.15. The rapid decrease in bit error rate as SNR increases is main specific feature of an AWGN channel. This decrease is the fastest, which could take place for digital modulation channels, so the AWGN channel is a “best case” channel. For high SNRs the bit error rate decreases by a factor of approximately 10, that is, in decade for every 1 dB increase in SNR.

In the M-dimensional vector space the probability of bit error is also proportional to the Euclidean distance between the closest points in the constellation diagram. The same upper bound for the probability od symbol error, as in 2D-case for AWGN channel, can be obtained for the M-dimensional case following to derivations carried out in [1, 4, 8]. According to [1, 4, 8], the average probability of error for a particular modulation signal, $P_r(\varepsilon|r_i)$

$$P_r(\varepsilon|r_i) \le \sum_{\substack{j=1 \\ j\ne i}} Q\left(\frac{d_{ij}}{\sqrt{2N_0}}\right) \tag{8.35}$$

where d_{ij} is the Euclidean distance between the *i*th and *j*th signal points in the constellation diagram, *Q*-function, we will repeat its definition once more

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{x^2}{2}\right) dx \tag{8.36}$$

If all the *M* modulation waveforms are equally likely to be transmitted, then the average density probability of error for a modulation can be presented as

$$P_r(\varepsilon) = P_r(\varepsilon|r_i) P_r(r_i) = \frac{1}{M} \sum_{i=1}^{M} P_r(\varepsilon|r_i) \tag{8.37}$$

For symmetric canstellations, the distance between all constellation points are equivalent, and the conditional error probability $P_r(\varepsilon|r_i)$ is the same for all *i*. Hence (8.35) gives the average probability of symbol error for a particular constellation set.

8.3.3. Performance of digital modulation in flat fading channels

As was discussed [1, 2], flat fading channels cause a multiplicative (gain) variations in the transmitted signal *s*(*t*) or in its envelope *r*(*t*). The AWGN digital channel described above, is a typical slow flat fading channel, for which the probability of error is described by *Q*-function in (8.34) versus SNR, γ .

Let us now consider the case of *narrowband flat digital channel* in which fading affects all frequencies in the modulated signal equally, so it can be modeled as a single multiplicative process. In this channel the fading process differs from AWGN considered above. We will show this for Rayleigh fading channel and will compare both kinds of channels. Since the fading varies with time, the SNR at the demodulator input also varies with time. It is necessary, in contrast with AWGN case, to distinguish between the instantaneous SNR, γ , and the mean SNR, Γ . Since fading in the channel changes much slow than the applied modulation message *m*(*t*), it can be assumed that the attenuation and phase shift of the signal is constant over at least one symbol interval. Therefore the received signal *r*(*t*) may be expressed as

$$r(t) = A\alpha(t)g(t) + n(t) \tag{8.38}$$

where $\alpha(t)$ is the complex fading coefficient at time *t*. Equation (8.38) mathematically describes the narrowband fading channel which is presented in Fig. 8.16 by the corresponding block-scheme. If the fading is assumed constant over the transmitted symbol duration, then is also constant over a symbol and is given

$$\gamma(t) = \frac{A^2 |\alpha(t)|^2 \left\langle |g(t)|^2 \right\rangle}{2P_n} = \frac{A^2 |\alpha(t)|^2}{2P_n} \tag{8.39}$$

and

$$\Gamma = \langle \gamma(t) \rangle \tag{8.40}$$

Usually we consider the flat fading, having unit variance, and lump any change in mean signal power into the path loss, so

$$\Gamma = \langle \gamma(t) \rangle = \frac{A^2}{2P_n} \tag{8.41}$$

Two things are therefore needed in order to find the performance of a system in the narrowband slow fading channel: the mean SNR and a description of the way the fading causes the instantaneous SNR to vary relative to this mean as shown in Fig. 8.17.

Let us now consider the distribution of SNR for a Rayleigh fading flat narrowband channel. We will use now expression (8.32) for the probability of error in a slow, flat fading channel, where now $\gamma = \alpha^2 W_s / N_0$ (all parameters for digital modulated signals (symbols) are defined above). for Rayleigh distribution the PDF of error is now [1, 4, 8, 10]

$$p(\gamma) = \frac{1}{\Gamma}\exp\left(-\frac{\gamma}{\Gamma}\right) \tag{8.42}$$

where $\Gamma = (W_s / N_0)\langle \alpha^2 \rangle$ is the average value of the signal-to-noise ratio. If so, for the comulative density function CDF we have

$$CDF(\gamma) \equiv \Pr(\gamma < \gamma_s) = \int_0^{\gamma_s} p(\gamma)d\gamma = 1 - \exp\left(-\frac{\gamma_s}{\Gamma}\right) \tag{8.43}$$

Here γ_s is the constant threshold of SNR in the Rayleigh fading channel. The result of (13.43) is illustrated in Fig. 8.18. This result can be used to calculate the mean SNR required to obtain an SNR above some threshold for an acceptable percentage of the time. But the same PDF from (8.42) can be successfully used to predict the error rate performance of digital modulation schemes in a Rayleigh channel by assuming that the SNR is constant over one symbol duration. In this case the Rayleigh error rate performance can be predicted directly from the AWGN channel case. Thus for binary phase shift keying (BPSK) signals, as above in previous sections, using AWGN error rate performance from (8.34) and the Rayleigh statistics according to (8.42) for the instantaneous SNR, γ, according to illustration in Fig. 8.18.

Following (8.42) and this illustration, the average bit error probability P_e can be defined as

$$\begin{aligned} P_{eBPSK} &= \langle P_e(\gamma) \rangle = \int_0^{\infty} P_e(\gamma)p(\gamma)d\gamma \\ &= \int_0^{\infty} Q\left(\sqrt{2\gamma}\right)\frac{1}{\Gamma}\exp\left(-\frac{\gamma}{\Gamma}\right)d\gamma = \frac{1}{2}\left[1 - \sqrt{\frac{\Gamma}{1+\Gamma}}\right] \end{aligned} \tag{8.44}$$

This result, as is an approximation, which leads from the case of large Γ, $P_e \approx 1/(4\Gamma)$ (see Fig. 8.19). This inverse proportionality is characteristic of decoded modulation in a Rayleigh channel leading to a reduction of bit-error-rate (BER) by one decade for every 10 dB increase in SNR. This contrasts sharply with ~1 dB per decade variation in the AWGN channel (see Fig. 8.15). The performance of other digital modulation schemes in Rayleigh fading channel can be analyzed in a similar way following to [1, 4, 8, 10]. Because BPSK is an example of linear digital modulation technique, where the amplitude of the transmitting signal *s*(*t*) varies linearly with the modulating signal *m*(*t*), let us present the same average bit error probability P_e for coherent binary frequency shift keying (FPSK) signal as an example of nonlinear modulation methods following to [1, 10]

$$P_{eFPSK} = \frac{1}{2}\left[1 - \sqrt{\frac{\Gamma}{2+\Gamma}}\right] \tag{8.45}$$

It also can be shown according to [1, 4, 10] that the average error probability for differential BPSK, as linear modulation method, and for incoherent binary FPSK, as nonlinear modulation method, in a slow, flat Rayleigh fading channel are given, respectively, by

$$P_{eDPSK} = \frac{1}{2(1+\Gamma)} \tag{8.46}$$

and

$$P_{eICFPSK} = \frac{1}{(2+\Gamma)} \tag{8.47}$$

Figure 8.20 illustrates how the BER for various modulations changes as a function of W_s / N_0 with respect to that for AWGN slow, flat channel. The same slow rate of reduction in BER obtained for BPSK signals as compared to a typical performance curve in AWGN is observed.

The rate of reduction in BER may be increased with increase of *Ricean factor*, *K*, introduced in [1-4], as a ratio between LOS component and NLOS component of the total multipath signal. The Ricean flat slow-fading channel is usually used whenever one path field component exceeds or at the same level to the other multipath components due to multi-scattering, multi-reflection and multi-diffraction. The bit error rate (BER) for BPSK modulated signals and for a Ricean fading channel, depending on *K* value, is intermediate between the AWGN and Rayleigh channel cases, as shown in Fig. 8. 21. It is clear that the Ricean channel behaves like AWGN channel in the limit as $k \to \infty$ without any multipath or non-line-of-sight components.

Bibliography

[1] Rappaport, T. S., *Wireless Communications*, New York: Prentice Hall PRT, 1996.
[2] Amorozo, F., "The bandwidth of digital data signals", *IEEE Communications Magazine*, Nov. 1980, pp. 13-24.

[3] Couch, L. W., *Digital and Analog Communication Systems*, Macmillan, New York, 1993.
[4] Stuber, G. L*., Principles of Mobile Communication*, Kluwert Academic Publishers, Boston-London, 1996.
[5] Gosling, W., J. P. McGeehan, and P. G. Holland, "Receivers for the Wolfson SSB/VHF land mobile radio system*", Proc. IEEE Conf. on Radio Receivers and Associated Systems*, Southampton, England, pp. 169-178, July 1978.
[6] Lusignan, B. B., "Single-sideband transmission for land mobile radio", *IEEE Spectrum*, July 1978, pp. 33-37.
[7] Wells, R., "SSB for VHF mobile radio at 5 kHz channel spacing*", Proc. IERE Conf. on Radio Receivers and Associated Systems*, Southampton, England, pp. 169-178, July 1978.
[8] Ziemer, R. E., and R. L. Peterson*, Introduction to Digital Communications*, Macmillan Publishing Co., 1992.
[9] Saunders, S. R., *Antennas and Propagation for Wireless Communication Systems*, John Wiley & Sons, New York, 1999.
[10] Proakis, J. G., *Digital Communications*, McGraw-Hill, New York, 1989.
[11] Blaunstein, N., and J. B. Andersen, *Multipath Phenomena in Cellular Networks*, Artech House, Boston-London, 2002.