12. DIGITAL MODULATION SCHEMES

12.1. Introduction

Digital modulation systems can be divided into two big groups. The baseband signals are transmitted in channels with lowpass character while in the case of modulated signals the channel has a bandpass character. A general block diagram of the digital modulation systems is shown in Fig. 12.1.

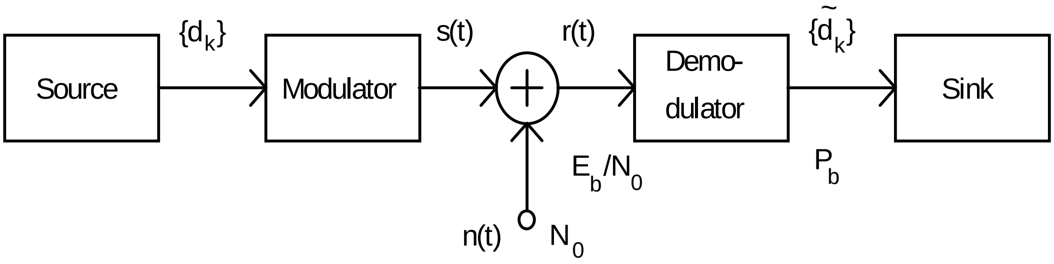

Fig. 12.1 General block diagram of the digital modulation systems

Symbols d_k transmitted by the source at time T feed the modulator which generates the modulated signal $s(t)$. In the channel the signal is exposed to various disturbing effects (additive noise, linear and nonlinear distortion, etc.). The distorted signal $r(t)$ comes into the demodulator which produces a series of estimated symbols $\tilde{d}_k$.

The system quality can be characterized by the bit-error ratio P_b related to the series of estimated symbols. If only Gaussian white noise is added to the signal in the channel, the bit-error ratio P_b depends firstly on the signal-to-noise ratio related to the signal $r(t)$ which in fact is determined by the ratio E_b/N_o where E_b is the signal energy carried by one bit and N_o is the single side power density of the additive Gaussian white noise.

In the following the possibilities of the modulation by a digital signal will be examined similarly as it was done for the analog modulating signal in the previous chapter. It has to be kept in mind, however, that true-to-form transmission of digital signals is not as important as the possible smallest probability of error of reproducing the original digital data from the transmitted signal. Anyway, if the ratio of mistaken decisions can be kept low then the shape of the transmitted signals is indifferent.

12.2. Baseband Modulation

In the case of digital *baseband modulation*, the information can be encoded in the amplitude, in the duration or in the position of the impulse. Accordingly, they are called the Pulse Amplitude Modulation (PAM), Pulse Duration Modulation (PAM) and Pulse Position Modulation (PPM) systems. PAM is the system with the most efficient use of the power and bandwidth thus we will concentrate to this kind of pulse modulation.

A general block diagram of the baseband PAM system is shown in Fig. 12.2.

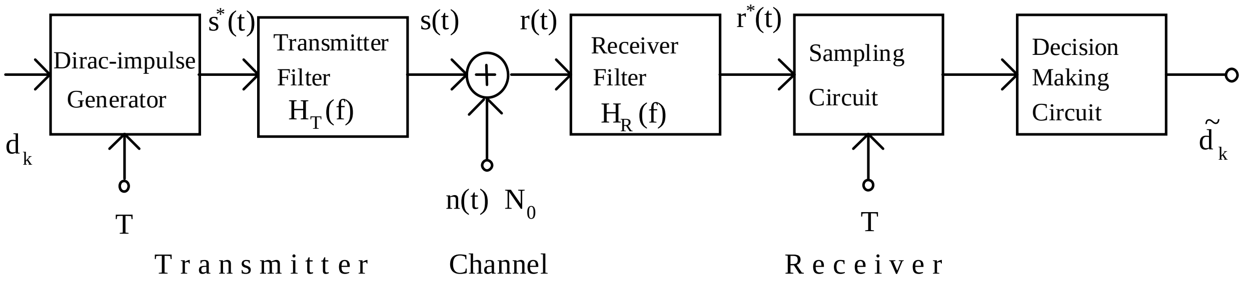

Fig. 12. 2. Block Diagram of Baseband Digital Transmission Systems

Source symbols $\{d_k\}$ control the Dirac-impulse generator at each time T. Suppose that the symbols can have M different values within the set $\{-(M-1),...,-3,-1,1,3,...,(M-1)\}$ and that the output signal of the Dirac-impulse generator is as follows:

$$s^*(t) = \sqrt{PT} \sum_{k=-\infty}^{+\infty} d_k d\left(t - kT\right) \tag{12.1}$$

where P is power of the signal. This signal controls the transmitting filter at the output of which the modulated signal is obtained in the following form:

$$s(t) = \sqrt{PT} \sum_{k=-\infty}^{+\infty} d_k h_T\left(t - kT\right) \tag{12.2}$$

where $h_T(t)$ is the impulse response of the transmitting filter with the frequency response $H_T(f)$. Additive Gaussian white noise added to the signal results in

$$r(t) = s(t) + n(t) \tag{12.3}$$

appearing at the input of the receiver. It is important to note that the noise $n(t)$ has infinite power thus the signal $r(t)$ has to be filtered in the receiver. This is carried out by the receiving filter with frequency response $H_R(t)$ which simultaneously shapes the signal passed to the sampling and decision circuits.

The signal $r(t)$ appearing at the output of the receiving filter has to fulfil two following requirements:

- the power of the filtered additive noise ($n^*(t)$) has to be as low as possible,
- the signal samples appearing periodically every time T has to be dependent on only one input symbol.

The first condition can be satisfied by matching the transmitting and the receiving filters i.e. by choosing their transfer functions so that $H_R(f) = H_T^*(f)$. (Matched filters best suppress the noise without significant distortion of the useful signal.) To satisfy the second condition, the transfer function $H(f)$ of the entire transmission chain given by

$$H(f) = H_T(f) \cdot H_R(f) \tag{12.4}$$

must be chosen in such a way that the impulse response $h(t)$ has special properties, namely: among the samples of $h(t)$ taken every time T only one sample may differ from zero, all the others must be zero. Such a system is said to have zero *Inter Symbol Interference* (ISI). The selection of the baseband signal shape is discussed in the next chapter.

12.2.1. Baseband Pulse Shaping

Let us examine the case when the channel is ideal, i.e. free of noises so that the output signal $r^*(t)$ is determined only by the transfer functions of the filters. Let $h(t)$ be the impulse response corresponding to the transfer function $H(t)$ and let us determine the signal at the output of the receiving filter in the absence of noise:

$$r^*(t) = \sqrt{PT}\sum_{k=-\infty}^{+\infty} d_k h_T(t-kT). \tag{12.5}$$

According to the so-called Nyquist criterium, the ISI-free condition can be satisfied with impulse responses satisfying the above condition:

$$\sum_{k=-\infty}^{+\infty} H\left(f+\frac{1}{T}\right) = T, if|f| \le \frac{1}{2T} \tag{12.6}$$

where in the case of matched filters:

$$H(f)=H_R(f)\cdot H_T(f)=|H_R(f)|^2=|H_T(f)|^2$$

The Nyquist criterion defined by equation (12.6.) can also be expressed in graphic form as shown in the upper part of Fig. 12.3. As can be seen the frequency response of a Nyquist filter is specified at frequency $1/2T$ as 50 per cent of the maximum value, furthermore it is point-symmetrical to the so-called Nyquist point. Since the Nyquist criterium specifies only one point and the symmetry of the frequency response, the number of possible responses is practically infinite.

To make the choice easier, the Fig. 12.3 b) shows the shapes of the pulses in the time domain corresponding to the same parameter α (a parameter characteristic for the slope of the spectrum rounding). It can be noticed that the greater is the slope of the rounding, the greater are the ripples, the 'overshots' of the time functions. It can also be seen that if the bandwidth is less than $1/2T$ then the ISI cannot be eliminated even theoretically since the spectrum cannot be shaped as point-symmetrical. On the other hand it is not necessary to have the bandwidth greater than $1/T$ since -because of the point-symmetry- the spectrum does not exceed this value.

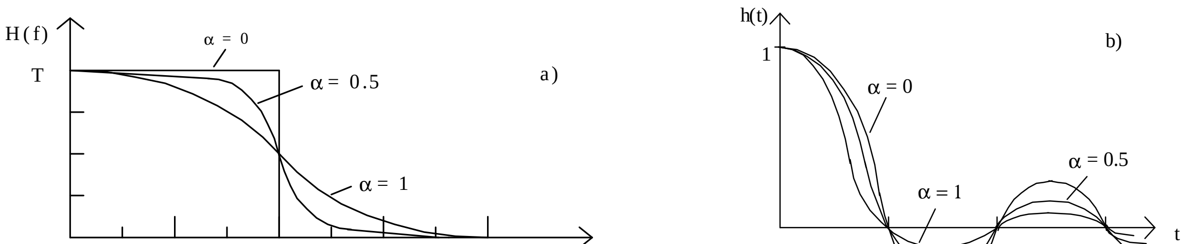

Fig. 12.3 Frequency (a) and Time-Domain (b) Representations of the Nyquist Criterium

To sum up the conclusions made for the received pulses: A minimum of $1/2T$ bandwidth is required for the zero-ISI data transmission. If a greater bandwidth is available then data can be transmitted with a considerably greater reliability by proper shaping of the impulse $h(t)$, i.e. by using the band up to $1/T$. In practice the so-called 'raised cosine' function with soft transition of $H(t)$ from T to zero is used (α=1) since it is 'gently sloping' and it is an acceptable compromise between the bandwidth increase (with respect to $1/2T$) and the time function free of overshots. (Overshots are dangerous since the probability of wrong decisions may increase if the timing is not proper in the receiver.)

12.2.2. The Error Ratio

In the case of PAM the error ratio is determined by the amount of the noise power within the resulting signal $r^*(t)$. Let us suppose that the signal is binary (d_k may be +1 or -1) and that the $H(f)$ satisfies the Nyquist criterium. In the absence of noise the signal appearing at times nT at the output of the receiving filter might have the value $\pm\sqrt{PT}$ (see Fig. 12.3). The Gaussian white noise $n(t)$ filtered by the receiving filter is denoted as $n^*(t)$. The expected value of the filtered noise is zero, the standard deviation is (using also Fig. 12.3.) as follows:

$$s_n^2 = \int_{-\frac{1}{T}}^{\frac{1}{T}} \frac{N_0}{2} \left|H_R(f)\right|^2 df = \int_{-\frac{1}{T}}^{\frac{1}{T}} \frac{N_0}{2} H(f) df = \frac{N_0}{2} \int_{-\frac{1}{T}}^{\frac{1}{T}} H(f) df = \frac{N_0}{2} \qquad (12.7)$$

Probability density functions of the received binary symbols disturbed by the noise are shown in Fig 12.4.

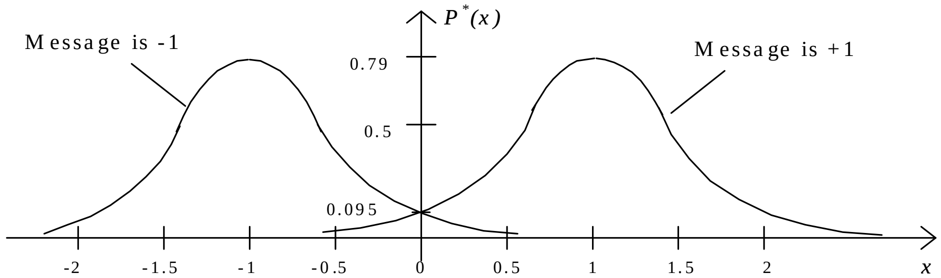

Fig. 12.4 Probability Density Function of the Signal at the Output of the Receiver Filter if $s^2 = 0.25$ and $\sqrt{PT} = 1$

As it can be seen, instead of ideal values (+1, -1) any signal may practically appear with a certain chance (distribution) at the output of the receiving filter. To decide whether the sent signal was a +1 or a -1, the sign of the sampled signal $r^*(t)$ has obviously to be examined. The probability of error can then be given as

$$P_b = \mathrm{P}\{r^*(nT) \geq 0 \mid d_n = -1\} \qquad (12.8)$$

which can be calculated from the following expression:

$$\begin{aligned} P_b &= \int_o^{\infty} \frac{1}{s\sqrt{2p}} \exp\left(\frac{\left(x+\sqrt{PT}\right)^2}{2s^2}\right) dx = \int_{\sqrt{PT}}^{\infty} \frac{1}{s\sqrt{2p}} \exp\left(-\frac{y^2}{2s^2}\right) dy = \\ &= \int_{\frac{\sqrt{PT}}{s}}^{\infty} \frac{1}{\sqrt{2p}} \exp\left(\frac{z^2}{2}\right) dz = Q\left(\frac{\sqrt{PT}}{s}\right) = Q\left(\sqrt{\frac{2PT}{N_o}}\right) \end{aligned} \qquad (12.9)$$

where Q is the so-called Gaussian error function defined as follows:

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2p}} \exp\left(-\frac{y^2}{2}\right) dy \qquad (12.10)$$

Since $P \cdot T = E_b$ is the energy of the useful signal per one bit, equation (12.9) is usually given as

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_o}}\right) \qquad (12.11)$$

The above function is shown in double logarithmic scale in Fig. 12.5.

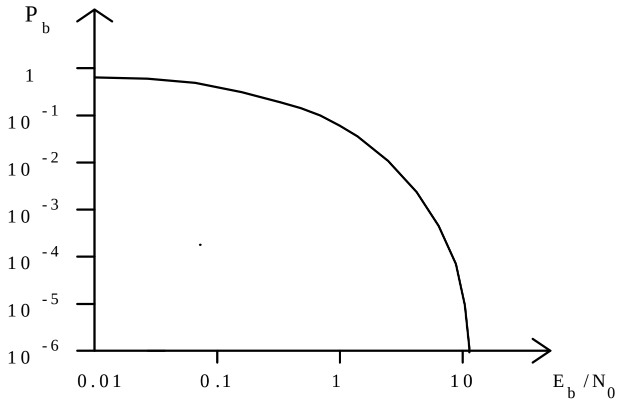

Fig. 12.5 Probability of Bit Error of the PAM System
as the function of the Signal-to-Noise Ratio

12.3. Carrier Modulations

Four important *carrier modulation* methods suitable for the transmission of binary data through a bandpass channel are shown in Fig. 12.6.(a)...(d). Figure (a) shows the discrete amplitude modulation with the corresponding binary data shown below. As it can be seen, in this modulation the carrier is switched on if the data are equal to 1 and switched off if the data are equal to 0. This modulation is called the *amplitude shift keying* and is abbreviated as ASK.

Similarly, the carrier frequency can also be switched between two different values corresponding to the binary data. This modulation shown in Figure (b) is called the *frequency shift keying* and is abbreviated as FSK. As it can be seen, it is the frequency of the FSK signal which changes in the rhythm of the binary data.

Figure (c) shows a signal both the amplitude and the frequency of which are constant; it is only the phase which changes according to the modulation. Therefore it is quite logical to call this modulation as *phase shift keying* and abbreviate it as PSK.

Finally, Figure (d) shows the case when the sinusoidal carrier is modulated in amplitude by such a discrete PAM signal which was previously `smoothed' i.e. filtered as shown in Chapter 12.2. Among the four procedures presented, this modulation -which in fact is AM-DSB- requires the minimum bandwidth although the equipment generating, transmitting, and demodulating such signals are very complex. On the contrary, ASK, FSK and PSK can be implemented with simpler and much cheaper devices. The price we have to pay for it is the greater bandwidth and the greater required transmitting power. If the bandwidth is not the main aspect of the design then the digital procedures can be well used since they have relatively good parameters, e.g. high noise immunity against different interferences.

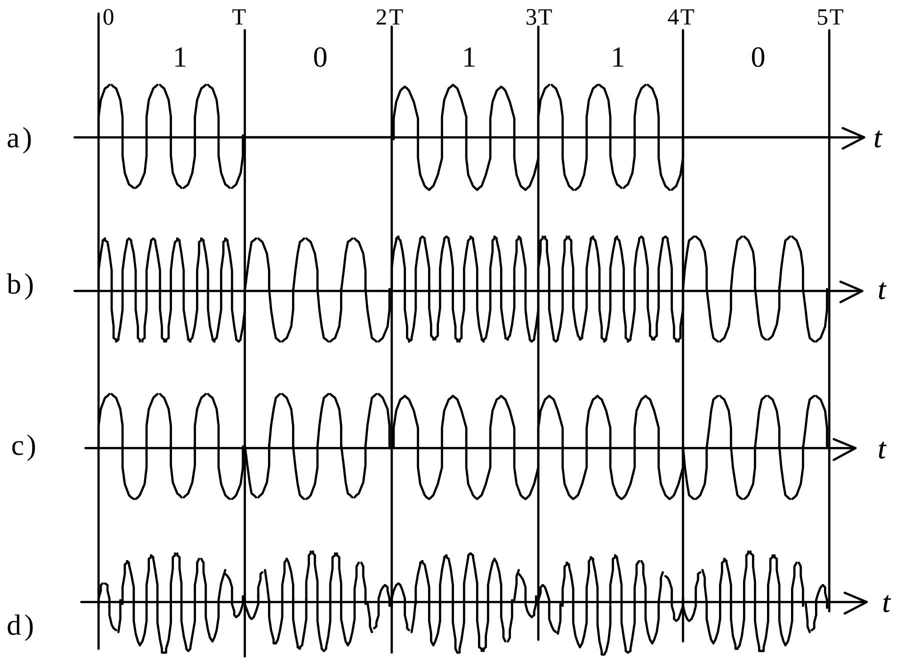

Fig. 12.6. Waveshapes of Digitally Modulated Carrier

12.3.1. Structure of Binary Modulation Systems

The essential task of the ASK, FSK and PSK receivers is to recognize binary data, i.e. to be able to make difference between the $s_1(t)$ and the $s_2(t)$. The quality of the receiver is determined by the probability of error and the structure of the receiver is considered optimal if the probability of error is minimal. In this chapter the structure of an optimal receiver suitable to receive ASK, FSK and PSK signals will be presented.

If the input noise of a receiver is a Gaussian white noise then it can be shown that the most important part of the receiver is the matched filter. It can be also shown that this filter can be realized by a correlator consisting of a multiplier and an integrator. The receiver is synchronized to the input signal, i.e. its local oscillator generates a sinewave whose frequency and phase are exactly the same as those of the input signal (This feature is called *coherence*).

Binary ASK, FSK and PSK signals can be demodulated also by non coherent methods. Although their quality is not optimal, noncoherent receivers are much simpler and thus widely used in low-speed data transmission systems.

The general block diagram of a binary data transmission system using digital modulation is shown in Fig. 12.7.

The input signal is a sequence of bits d_k, the bitrate (the speed) of which is $1/T$ and the duration of the bits -the time-slot- is T. At the time-slot k the output signal of the demodulator depends on the value of the d_k. The signal $s(t)$ generated by the modulator in the time-slot k is one of the two possible waveforms, $s_1(t)$ or $s_2(t)$ shifted to the time of the k_{th} bit. Thus the $s(t)$ is a stochastic process determined as follows:

$$s(t)=\begin{cases} s_1(t-kT), & if\, d_k=-1 \\ s_2(t-kT), & if\, d_k=1 \end{cases} \tag{12.12}$$

provided that $k \cdot T \le t < (k+1) \cdot T$.

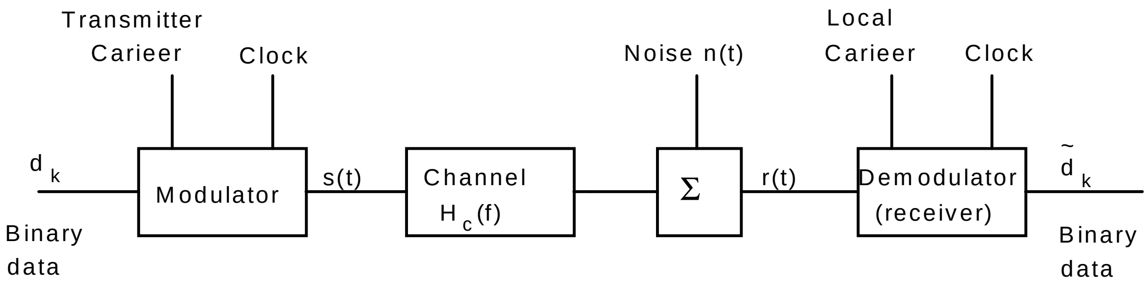

Fig. 12.7 Block diagram of Modulated Binary Data Transmission System

The duration of both signals is T and both are finite energy signals since both $s_1(t)$ and $s_2(t) \equiv 0$ if t lies outside the period 0 ... T; inside the period, however, the time integral of the square of both functions is finite.

Table 12.1. Signaling Waveforms of ASK, PSK and FSK

$s_1(t); 0 \le t \le T$	$s_2(t); 0 \le t \le T$	Modulation
0	$A \cdot \cos(w_c \cdot t)$ $or A \cdot \sin(w_c \cdot t)$	Amplitude shift-keying (ASK)
$-A \cdot \cos(w_c \cdot t)$ $or -A \cdot \sin(w_c \cdot t)$	$A \cdot \cos(w_c \cdot t)$ $or A \cdot \sin(w_c \cdot t)$	Phase shift-keying (PSK)
$A \cdot \cos(w_c - w_d)t$ $or A \cdot \sin(w_c - w_d)t$	$A \cdot \cos(w_c + w_d)t$ $or A \cdot \sin(w_c + w_d)t$	Frequency shift-keying (FSK)

The signal shape depends on the actual modulation as summarized in Table 12.1. The output signal of the modulator goes through a bandpass channel the transfer function of which is $H_c(f)$. Suppose the channel is ideal, i.e. the transmission is free of distortions, except a finite time delay and a zero-mean, stationary Gaussian noise with known double-side spectral power density ($N_o/2$). The received signal can thus be written as follows:

$$r(t)=\begin{cases} s_1(t-kT-t)+n(t), if\, d_k=-1, \\ s_2(t-kT-t)+n(t), if\, d_k=1, \end{cases} \tag{12.13}$$

provided that $k \cdot T \le t < (k+1) \cdot T$, where τ is the time delay of the transmission which can be considered to be zero without restricting the generality.

The block diagram of the receiver is shown in Fig. 12.8. The task of the receiver is to decide which one of the $s_1(t)$ and $s_2(t)$ functions is present at the input. The actual receiver consists of a filter, a sampler and a threshold detector (comparator). First, the signal $r(t)$ goes through a filter and is sampled at the end of each bit time-slot. The sample is then compared with a previously determined threshold and the generated bit is decoded as 1 or -1 depending on the relation between the $r^*(t)$ and the threshold.

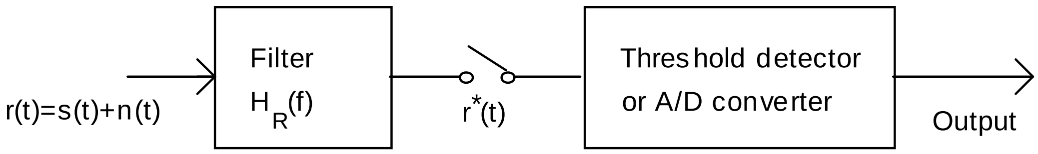

Fig. 12.8. Receiver of the Binary Data Transmission System

Owing to the noise, the receiver makes sometimes false decisions. The probability of errors depends on the signal power and spectral power density of the noise at the receiver's input, on the signalling frequency (bitrate) and on the parameters of the receiver, such as the transfer function $H_R(t)$ of the filter and the threshold value (similarly to the PAM type transmission).

(a) Binary ASK Modulation

Being very simple to realize, the binary ASK was used in wireless telegraph communication (spark telegraph) at the beginning of this century. Although ASK has been almost entirely replaced by the much more effective FSK and PSK, it is instructive to get acquainted with the most important parameters of the ASK since the ASK is a very clear model of the binary signalling systems. The actual form of the ASK signal -defined generally by equation (12.12)- is $s_2(t) = A\cdot\cos(\omega_c t)$, if $0 \le t \le T$ and $s_1 \equiv 0$. Suppose that the 2π multiple of the carrier frequency is $\omega_c = 2\cdot\pi\cdot n/T$ where n is an integer. The time function of the modulated signal can be written as

$$s(t) = d(t)\cdot[A\cdot\cos(\omega_c t)] \qquad (12.14)$$

where $d(t)$ represents a baseband pulse train. Suppose that $d(t)$ is a random squarewave with period T. It follows from equation (12.14) that the ASK signal can be generated by a multiplier which means that the $d(t)$ signal can be used to switch the carrier on and off. The relation between the spectral power density of the modulated signal and that of the pulse train $d(t)$ is as follows:

$$G_s(f) = \frac{A^2}{4}\cdot\left(G_d(f-f_c)+G_d(f+f_c)\right). \qquad (12.15)$$

The time function $d(t)$ is a random binary signal with only two levels: 0 and 1. Using the autocorrelation function it can be shown that the spectrum of the modulated signal is as follows:

$$G_s(f) = \frac{A^2}{16}\left[d(f-f_c)+d(f+f_c)+\frac{\sin^2(pT(f-f_c))}{p^2T^2(f-f_c)^2}+\frac{\sin^2(pT(f+f_c))}{p^2T^2(f+f_c)^2}\right]. \qquad (12.16)$$

The graphic form of the above spectrum is shown in Fig. 12.9. The Dirac functions at frequencies f_c and $-f_c$ represent the carrier while the $(\sin x)/x$ components are the sidebands. As can be seen, the bandwidth is theoretically infinite but the spectrum decreases rapidly with the distance from the carrier frequency; neglecting the small components a finite bandwidth can be specified. It can be shown that if the spectrum is limited so that less than 5 per cent of the whole signal energy is that of the neglected part, the bandwidth of the ASK signal is about $3/T$.

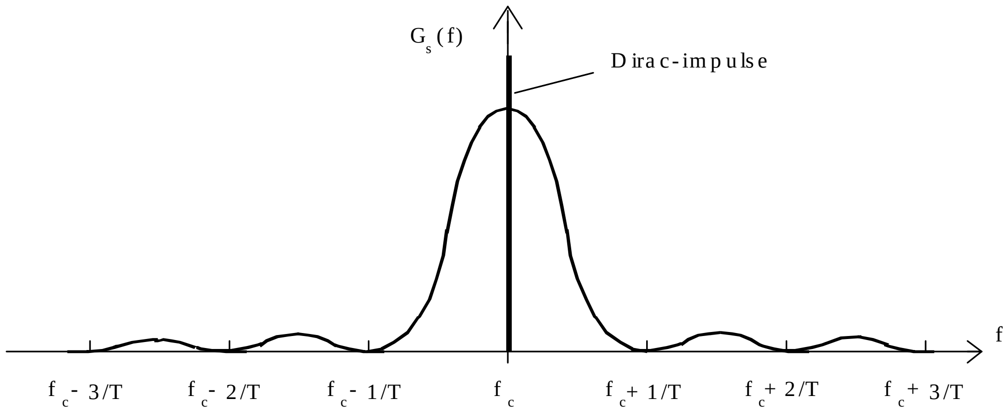

Fig. 12.9 Spectral Power Density Function of a Random Binary ASK Signal

The bandwidth can be reduced even more if the modulating signal is not a square wave but the $d(t)$ is smoothed before the modulation. So the bandwidth decreases to about $2/T$ thus for such an ASK signal a channel with bandwidth $2/T...3/T$ is appropriate.

The ASK signal can be demodulated both in a coherent and a non-coherent way. The coherent method requires the information about frequency and the phase of the input signal and in principle consists of an integration and a decision while the non-coherent one uses an envelope detector.

(b) Binary PSK Modulation

PSK is a discrete phase modulation in which two signals of opposite phases: $s_1(t) = -A\cdot\cos(\omega_c t)$ and $s_2(t) = A\cdot\cos(\omega_c t)$ are assigned to the binary data -1 and +1. As usually, the modulated PSK signal $s(t)$ can be written as

$$s(t) = d(t)\cdot[A\cdot\cos(\omega_c t)] \qquad (12.17)$$

where $d(t)$ is a random binary sequence of -1's and 1's with period T. It is interesting to notice that the only difference between the ASK and the PSK signal is that if the data are -1, the ASK multiplies the carrier by zero while the PSK multiplies it by -1. It can be shown that the spectral power density of the PSK signal is

$$G_s(f) = \frac{A^2}{4}\left(G_d(f-f_c) + G_d(f+f_c)\right), \qquad (12.18)$$

where

$$G_d(f) = \frac{\sin^2(p \cdot f \cdot T)}{p^2 f^2 T^2}, \qquad (12.19)$$

Comparing equations (12.19) and (12.16), the spectral power densities seem to be similar. The only difference is that PSK does not contain the Dirac functions which means that no discrete spectral component is present at the carrier frequency. It is also a logical conclusion that the bandwidth of the PSK is the same as for the ASK.

The PSK signal, however, can be demodulated only by a coherent demodulator, i.e. the carrier generated in the receiver has to be synchronized in frequency and in phase to the carrier of the input signal. As can be exactly shown, PSK requires half of the power (-3 dB) required by the (also coherent) ASK for the same probability of error.

(c) Binary FSK Modulation

FSK is used mainly in low-speed data transmission systems. The FSK receiver can be realized without coherent demodulation since two different frequencies can be detected by quite simple circuits. The generation of the FSK signal is also simple. Nevertheless it has to be mentioned that as far as the power requirement and the bandwidth are concerned, the efficiency of the FSK is not as good as that of the PSK.

Two shapes of the binary FSK signal corresponding to +1 and -1 data are as follows:

$$s_1(t) = A\cdot\cos(\omega_c t - \omega_d t), \text{ and } \quad s_2(t) = A\cdot\cos(\omega_c t + \omega_d t), \qquad (12.20)$$

where ω_d is the frequency deviation. Thus the information transmitted by an FSK signal is given by its actual frequency.

The binary FSK is an FM signal with constant amplitude and continuous phase. One of the possible mathematical descriptions of such a signal is given by the following formula:

$$s(t) = A\cdot\cos\left(\omega_c t + \omega_d \int_{-\infty}^{i} d(\tau)d\tau + \varphi\right), \qquad (12.21)$$

where $d_k(t)$ is a random binary pulse train with the value +1 if $d_k = 1$ and -1 if $d_k = -1$. The instantaneous frequency of the binary FSK signal is -according to the definition- the time derivate of the phase of the signal $s(t)$, i.e.:

$$f_p = \frac{d}{dt}\left(\omega_c t + \omega_d \int_{-\infty}^{i} d(\tau)d\tau + \varphi\right) = \omega_c + \omega_d\cdot d(t). \qquad (12.22)$$

Since $d(t) = \pm 1$, the instantaneous frequency has also only two values: $\omega_c \pm \omega_{d.}$

The analysis of the digital FM signal is rather complex. Instead of a detailed discussion only the results of the calculations are given in Fig. 12.10. If the frequency deviation is small (less than 25 per cent in comparison with the carrier frequency) then the main part of the spectrum is concentrated around the carrier frequency (which is the reciproque value of the bitrate T). The bandwidth is in the order of $2/T$ Hz which corresponds to the value of the PSK. If the deviation, however, increases above 25 per cent, then two peaks are shifted towards the signalling frequencies ($f_c + f_d$) and ($f_c - f_d$) thus the bandwidth is greater than $2/T$ Hz. If the deviation is even greater yet (above 75 per cent), then the spectrum looks like two interlaced ASK spectra with two carrier frequencies at about ($f_c + f_d$) and ($f_c - f_d$).

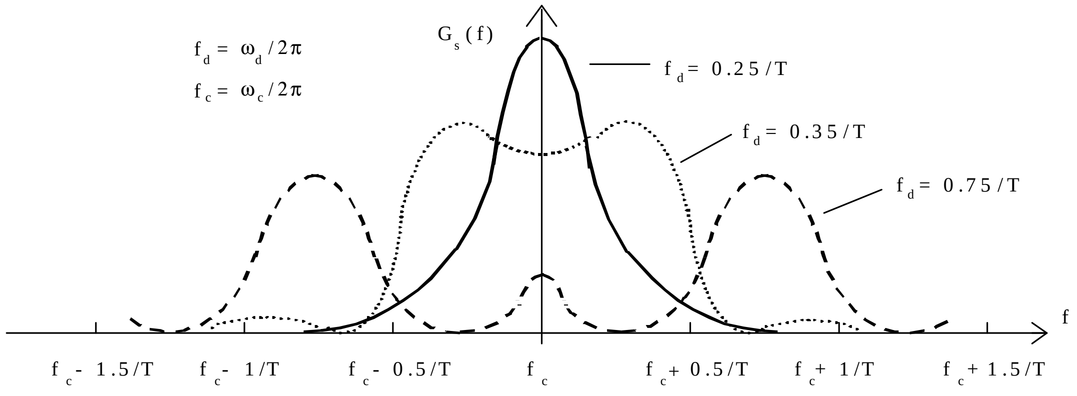

Fig. 12.10. a) Spectral Power Density Function of the Binary FSK Signal

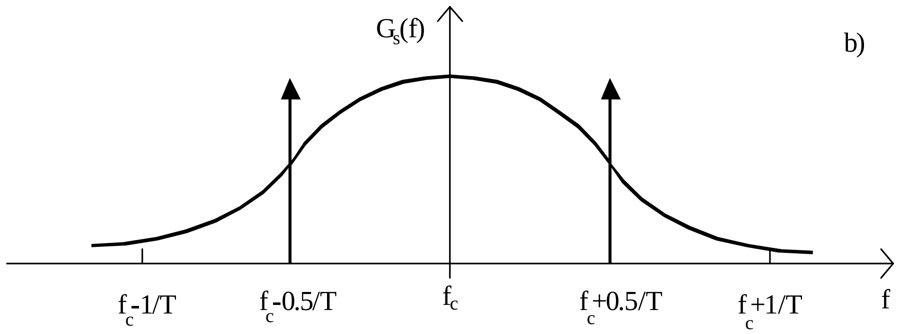

Fig. 12.10. b) Spectrum of binary FSK signal when $2\,f_d = 1/T$

Further, if the frequency deviation is such that $2 \cdot f_d = m/T$ where m is an integer, then the spectrum contains also two discrete spectrum lines as shown in Fig. 12.10.b. It can be said generally that the bandwidth of an r signal is greater than either that of the ASK or of the PSK.

As it was mentioned above the FSK signal given by equation (12.21) is continuous in phase, i.e. the signal actually keyed is of the same phase as it was before keying.

Control Questions

1. What does the general structure of the baseband digital modulation systems look like?. Define the term of matched filtering in a channel with Gaussian white noise.
2. What is the condition of ISI-free transmission in a baseband digital modulation system?
3. How can the probability of error of a baseband binary digital modulation system be determined?
4. What are the characteristic types of digital carrier modulating schemes? Draw the specific waveforms of these schemes.
5. Draw the general block diagram of digital carrier modulation systems.
6. Give the spectral power density of the output signals of binary ASK, PSK and FSK systems.

Exercises

1. Determine the frequency response of the receiving filter for the baseband binary digital modulation to have an ISI-free transmission, if

$$H_T(f) = \begin{cases} T, & \text{if} \left| f \right| < \dfrac{1}{T} \\ 0 & \text{otherwise.} \end{cases}$$

2. Determine the error ratio of a binary baseband digital transmission if $H_T(f) = H_R(f)$; $P = 10^{-9}$ W; $T = 10^{-3}$ s; $N_o = 4 \cdot 10^{-11}\ \frac{W}{Hz}$ and

$$H(f) = \begin{cases} T, & \text{if} \left| f \right| < \dfrac{1}{2T} \\ 0 & \text{otherwise.} \end{cases}$$

3. Draw the spectral power density of the output signal of a binary PSK system if $A = 1$ V ; $T = 10^{-3}$ s and the reference resistance value is 1 Ω.

References

[1] Proakis, J.G.: Digital Communication, McGraw-Hill, New York, 1983.
[2] Viterbi, A.J.: Principles of Coherent Communication, McGraw-Hill, New York, 1966.
[3] Lucky,R.W.-Salz,J.-Weldon,E.J.: Adatátvitel, Mûszaki Könyvkiadó, Budapest, 1973.

Abbreviations

ASK Amplitude Shift Keying
BER Bit Error Ratio
FSK Frequency Shift Keying
ISI Intersymbol Interference
PAM Pulse Amplitude Modulation
PDM Pulse Duration Modulation
PPM Pulse Position Modulation