Ministry of Education and Science of Ukraine
O.S. Popov Odessa National Academy of Telecommunications

A.G. Zuko Telecommunication Theory Department

Ivaschenko P., Rozenvasser D.

TELECOMMUNICATION THEORY

Textbook

Part 1

Odessa 2016

UDK 621.391 Plane of Education Materials Publication in 2016

Ivaschenko P. Telecommunication theory: textbook / P. Ivaschenko, D. Rozenvasser. – Odessa: O.S. Popov ONAT. – Part 1, 2016. – 124 p.

Editor of English language: Golub N.

The Telecommunication Theory textbook consists of two parts:

part 1 – «Communication Signals» and «Theory of Noise Immunity of Telecommunication Signals Reception»;

part 2 – «Transmission of Information in Telecommunication Systems» and «Bases of the Error-control Codes Theory».

This textbook (part 1) contains main theoretical positions, questions for examination of knowledge and short English-Russian and Russian-English dictionaries.

The textbook is intended for students training on the direction 172 – Telecommunications and radio technology studying the Telecommunication Theory.

APPROVED on A.G. Zuko Telecommunication Theory Department meeting.
Protocol № 3, October 20, 2016

APPROVED on Methodical Council O.S. Popov ONAT.
Protocol № 6, February 23, 2016

CONTENTS

1 GENERAL CONCEPTS OF TELECOMMUNICATION SYSTEMS

The branch of science “telecommunication theory” studies the common relationships of information transfer on distance. The object of studying is telecommunication system.

Telecommunication system provides transfer of information on distance by electric signals. The problem of information transfer on distance is formulated as: there is a source of the information (a person, a computer, etc.), having some information that is necessary to transfer to the remote recipient. This information should be transferred with the given level of fidelity and with an allowable delay. To discuss this problem we shall define the basic concepts: information, message, signal and communication channel.

1.1 Definition of basic concepts

Information is a collection of knowledge on any process, events, object. This knowledge reduces uncertainty for the recipient before it obtains knowledge. For transfer or to store information it is used different signs (or symbols), which allow to present it in some form.

Message is a material form of representation of information. First of all it is a set of signs that represent information. Message transfer on a distance is carried out by a signal.

Signal is a physical process in which message is represented and which is used to transfer information on distance. Signal can be electric, sound or light. In the telecommunication theory (by default) signal is an electric current or a voltage that represented the transferred message.

Information system is a system, which functions on the basis of information usage. Information takes place outside and/or inside considered system. Special case of information system is a telecommunication system.

Telecommunication system provides message transfer with a certain quality from a message source to a message recipient. Telecommunication system can provide one-way message transfer (broadcasting) or two-way message transfer (communication). In the first case simplex transfer takes place, in the second case duplex transfer takes place: full duplex– when a system provides simultaneously reception and transfer of messages; half duplex – when a system provides reception and transfer of messages one-by-one.

The generalized block diagram of telecommunication system for one-way message transfer is shown on fig. 1. Here $a(t)$ is transmitting message; $\hat{a}(t)$ is receiving message; $b(t)$ is transmitting baseband signal; $\hat{b}(t)$ is receiving baseband signal. In the case of duplex transfer it is necessary two sets of the units shown on fig. 1.

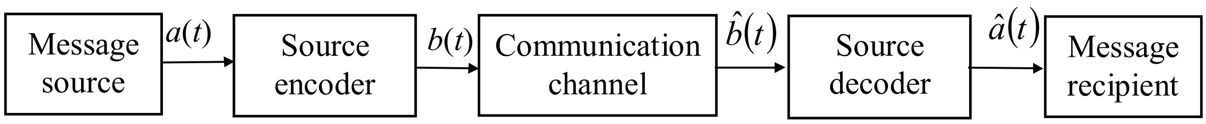


Figure 1 – Generalized block diagram of telecommunication system

A source of information gives out information by messages. The construction of the equipment is mainly defined by characteristics of transferred messages; therefore we speak about “message source”, instead of “information source”; in a similar way we speak about “message recipient”, instead of “information recipient”.

1.2 Messages and baseband signals

All messages are divided into continuous and discrete messages.

Discrete message consists of sequence of separate signs, which quantity is finite. These signs form an **alphabet of a source**. An example of a typical discrete message is the text. Transformation of discrete messages to electric signals consists in their encoding which is carried out by a source encoder. Encoding of a message is carried out on the basis of a *code*. **Code** is a rule or a table on the basis of which each sign of a message is converted into the code combination (a set of binary symbols) (fig. 2). As a result of encoding we receive **digital baseband signal** (fig. 3, *a*). Inverse transformation of digital baseband signals to messages consists in decoding, which is carried out in the source decoder (fig. 1). The basic characteristic of a digital signal is its rate R, bit per second (bit is the short name of binary symbols). On fig. 3, *a* it is shown $T_b = 1/R$, T_b is bit duration.

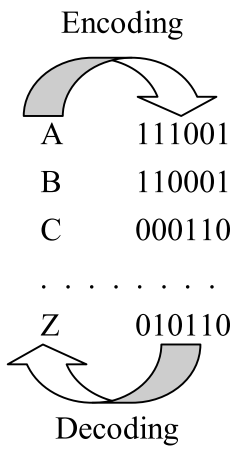
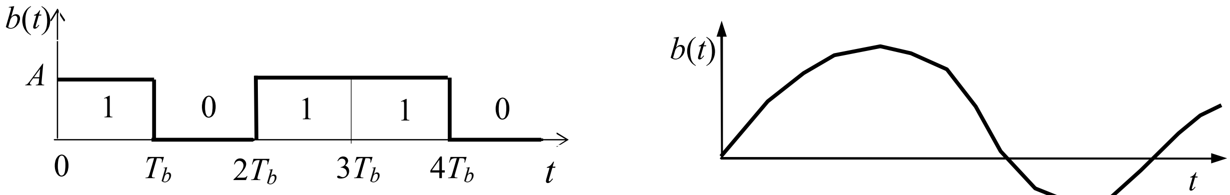


Figure 2 – Illustration of encoding and decoding

Continuous message represents a change of some time continuous magnitude (for example, sound pressure). Transformation of continuous messages to electric signals is carried out by different transducers (for example, a microphone). As a result of transformation we receive **a continuous baseband (analog) signal** (fig. 3, *b*). The basic characteristic of a continuous baseband signal is the maximum frequency of its spectrum $F_{\max}$ that characterizes its speed of change.



a) *b*)

Figure 3 – Baseband signals: *a* – digital signal; *b* – continuous signal

1.3 Communication channel

Communication channel is a set of devices for transfer of electric signals on distance.

If a baseband signal is a digital one then communication channel must be digital. If a baseband signal is continuous then communication channel must be continuous.

A continuous baseband signal can be transformed into a digital signal for transfer by a digital communication channel. In that case there will be analogue-digital conversion. So, in case of digital transfer of continuous messages a source encoder transforms messages into a baseband continuous signal, and then – in a digital signal. Digital transfer has a lot of advantages in comparison with analogue transfer. Digitalization of transmission systems takes place during several last decades.

Transformations discussed above are shown on fig. 1.

A communication channel can be simple or compound.

We shall consider construction of compound channel as part of a telecommunication network.

1.4 Communication network

Network is a set of nodes and links which provide information exchange between users (users are sources and recipients).

There are user nodes (UN) where a terminal equipment is used, and switch nodes (SN) where switching of channels or packages is carried out. On fig. 4 the fragment of logic topology of a network is shown. It takes part in transfer of information between the considered user nodes. In one UN there is a message source and a source encoder, and in other UN – a source decoder and a message recipient. User node is called also the terminal equipment.

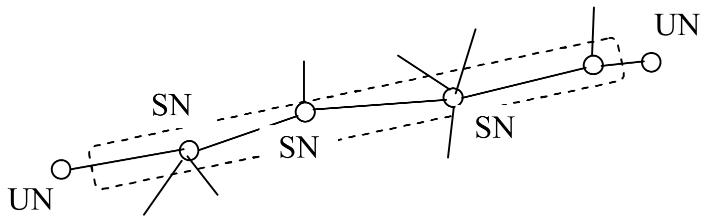


Figure 4 – Fragment of network topology: SN – switch node; UN – user node

Links that form a communication channel are traced by dotted line.

1.5 Transmission systems

Links of network are called **transmission systems**. Transmission system that connects user node with the nearest switch node is called access system.

On fig. 5 block diagram of typical digital transmission system is shown.

The transmission system is constructed on the basis of a transmission line. **Transmission line** is a physical circuit (cable) or a free space (radio communication), used for transfer of a signal on distance.

The baseband digital signal is coded by the error control code allowing at decoding to find out or correct errors, arising at transmission.

The **modulator** forms a signal, which is well suited for transmission over the line. In other words, the modulator coordinates the characteristics of the signal with channel characteristics, in particular bandwidth. The primary **demodulator** recovers a digital signal from the modulated signal.

In some cases, the channel **encoder** and **decoder** may be absent, if a demodulator recovered signal with high quality.

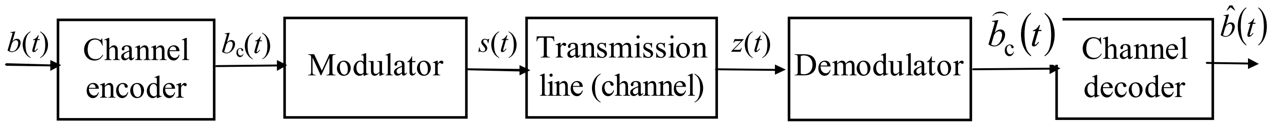


Figure 5 – Block diagram of a typical digital transmission system

1.6 Interference and distortion

Interference – any impact on the signal during transmission, which causes the random deviation of the signal values. Interference are classified according to the following criteria: origin, physical properties, nature of effects on the signal.

By origin interference can be divided on:

– thermal noise – noise from the equipment;

– interference from outside sources that enter to the input of demodulator.

By physical properties interference can be divided on:

– fluctuation of the obstacle (synonym) noise, a continuous oscillation that varies randomly;

– pulse–interference – isolated short pulses of random duration and intensity.

– concentrated (on the spectrum) interference – modulated or is a harmonic oscillation with a spectral width which is less than or comparable to the spectral width of the useful signal.

By nature of the impact on the signal interference can be divided on:

– additive interference – its instantaneous value is added to the instantaneous values of the signal.

– multiplicative interference – its instantaneous values are multiplied with the instantaneous values of the signal.

Signal **distortion** is not accidental change of the waveform due to non-ideal characteristics of electrical circuits and devices through which the signal passes. There are linear and nonlinear distortions that occur, respectively, in linear and nonlinear circuits. In the general case of distortion of the signal negatively affect the quality of messages recovering, and therefore they should not exceed the established norms.

1.7 Main characteristics of telecommunication system

Main characteristics of telecommunication system are:

− accuracy of message transfer or quality of message transfer;

− transmission rate or quantity of transferred information per 1 second.

In communication channel signal is distorted because of noise action. A noise is occasional influence on a signal that makes complicate signal transfer.

Distortion of digital signals will consist of occurrence of errors – instead of actually transferred symbols to the recipient other symbols come. Such kind of distortion is quantitatively characterized by probability of a symbol error. Distortion of continuous signals will consist of their form changing because of noise covering. Distortion of continuous signals can be characterized by an average square of difference between accepted and transmitted signals or the signal/noise ratio.

For messages transfer on a telecommunication channel it is necessary to spend a band of frequencies and power of a signal (the basic resources of a telecommunication channel). So, the main tasks of telecommunication theory following:

– how to provide necessary quality of message transfer on a telecommunication channel;

– how to provide necessary transmission rate on a telecommunication channel at limited resources of a telecommunication channel.

Questions for section 1

1. Define the concepts of information, message, signal.
2. What problem does the system of electric communication solve?
3. What is the difference between "communication" and "broadcasting"?
4. Define discrete and continuous messages.
5. How to get the baseband digital and analog signals?
6. Define the concept of communication channel.
7. Under what conditions should the communication channel be continuous and what be digital?
8. How is digital transmission of continuous signals carried out?
9. Define the concept of communication network.
10. Define the concept of terminal equipment and the switching node.
11. Define the concept of transmission system.
12. Define the concept of transmission line.
13. What is the purpose of digital signal encoding by error-control code?
14. Define the concept of modulator and demodulator.
15. Define the concept of noise and the classification of noise.
16. Define the concept of distortion.
17. What are the quantitative measures of noise impact estimation and the estimation of distortion on digital and continuous signals?

2 ELEMENTS OF GENERAL THEORY OF SIGNALS

2.1 Classification of signals

Signal is rather wide concept. Signal is a process of change in time of the physical phenomenon or a state of any technical object. Signals serve for mapping, registration and transmission of message. The common for signals is that in them the information is assumed in it. We shall consider, that a signal is an electric voltage or a current.

The most natural mathematical model of a signal is function of time $b(t)$, $s(t)$, $z(t)$ and etc. Such function of time establishes conformity between any moment of time t and size $s(t)$. Considering mathematical models of signals, we abstract from the concrete physical nature of a signal (a voltage, a current, an intensity of an electromagnetic field, etc.), considering, that function $s(t)$ completely reflects the important properties of a signal.

Depending on what values of a signal s and values of a variable t are possible, distinguished following:

1. A signal is **continuous** (fig. 6, *a*) if a set of values t is continuous, i.e. the argument accepts any values on an interval of existence of a signal.

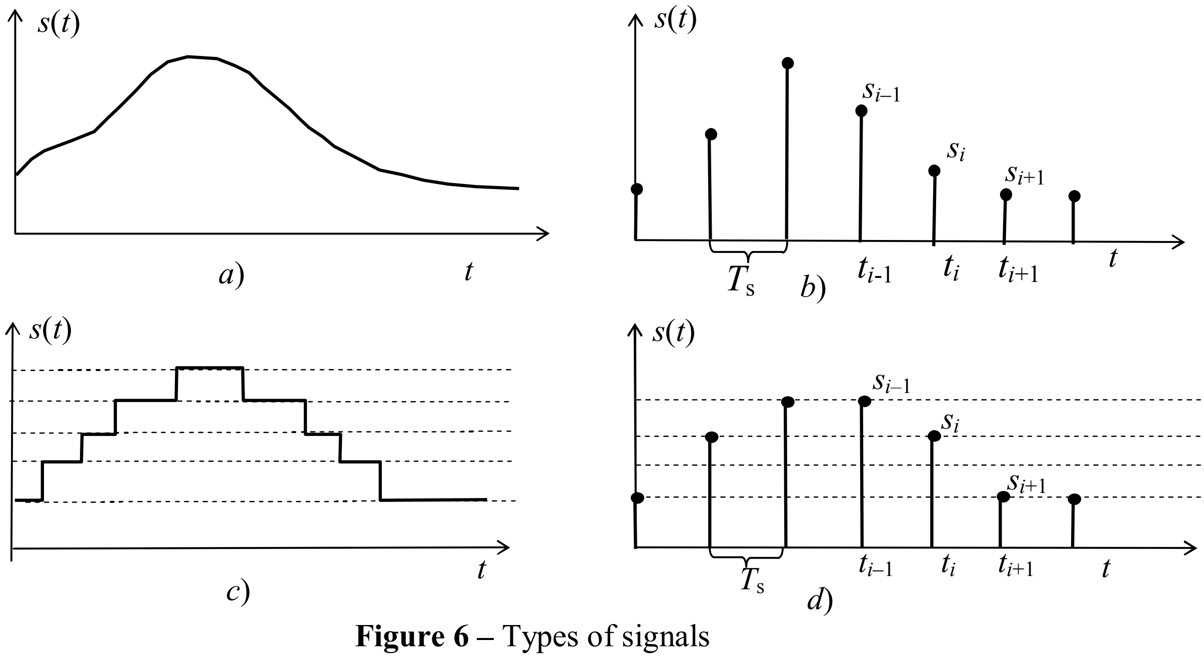


Figure 6 – Types of signals

2. A signal is **discrete** (fig. 6, *b*) if a set of values t is finite. Discrete signal is called also sequence, a time series. Such signal appears as a result of sampling a continuous signal. The step in time, through which discrete signal is given is called sampling interval T_s.

3. A signal is **quantized** (fig. 6, *c*) if value s accept final number of values. Such signal is a result of discretization on a level of a continuous signal.

4. A signal is **digital** (fig. 6, *d*) if it is also discrete, and quantized. Digital signals appear as a result of coding discrete messages, and also as a result of coding continuous signals for their representation by digital signals.

Signals are divided into:

1. **Baseband signal** is a representation of the message of not electric nature by an electric signal.

2. **Modulated signal** is a result of transformation of a baseband telecommunication signal in a signal for transfer by transmission line.

All signals are categorized as:

1. **Determined signals** – mathematical expression of determined (completely known) signal is completely certain function of time set by a formula, a plot or the table of values. For example, $s(t) = A_0\cos(2\pi f_0 t + \varphi_0)$, where A_0, f_0 and φ_0 are defined numbers. All values $s(t)$ are known at any moment.

2. **Stochastic signals** – mathematical representation of a stochastic signal is stochastic function of time, its values cannot be precisely predicted beforehand. Stochastic function of time (stochastic process) is described by statistical characteristics that characterize those or other properties of this function on the average. Stochastic processes are more full mathematical models of communication signals, than the determined functions of time. But many transformations of signals can be studied, using the determined functions of time. It can be simple function – harmonious waveform, a pulse, etc.

Any real signal $s(t)$ has **final duration** T_s. In many cases it is convenient to count, that the signal is **infinite on duration** and exists on an interval $(-\infty, \infty)$ or $(0, \infty)$.

Signals are **real** and **complex**. It is clear, that a signal is the real function of time which represents a state of some object. Sometimes for convenience of the mathematical analysis of signal transformations a complex signal is entered into consideration

$$s(t) = s_1(t) + j\, s_2(t),$$

where $s_1(t)$, $s_2(t)$ is the real functions of time; depending on a solved problem $s_2(t)$ is result of some transformations of function $s_1(t)$ or function, not dependent from $s_1(t)$.

Widely used complex exponent is an example

$$s(t) = e^{j2\pi f_0 t} = \cos 2\pi f_0 t + j \sin 2\pi f_0 t\,.$$

Frequently in devices of signal transformation the auxiliary signals, which are not containing the information, are used. Such signals are auxiliary and refer to as waveforms.

2.2 Energy characteristics of determined signals

The basic energy characteristics of a signal $s(t)$ are its **power** and **energy**. Instant power of the real signal is defined as a square of instant value $s(t)$:

$$p(t) = s^2(t),\ \mathrm{V}^2. \tag{2.1}$$

Power of a signal characterizes signal intensity, its ability to influence on devices, which register a signal.

Average power of a signal of final duration is defined by averaging (2.1) on an interval of existence of a signal $(0,\ T_s)$

$$P_s = \frac{1}{T_s}\int_0^{T_s} s^2(t)dt. \tag{2.2}$$

Energy of a signal of final duration is defined as

$$E_s = P_s T_s = \int_0^{T_s} s^2(t)dt. \tag{2.3}$$

From the last ratio it is clear, that energy of a signal takes into account both signal intensity, and time of its action.

Power and energy of a complex signal $s(t)$ are defined by ratios (2.1)...(2.3) in which instead of $s^2(t)$ it is necessary to substitute $s(t)\cdot s^*(t) = |s(t)|^2$, where $s^*(t)$ is the function in a complex conjugate with $s(t)$; $|s(t)|$ is the module of a signal $s(t)$.

The signal refers to **normalized**, if its energy

$$E_s = 1. \tag{2.4}$$

In addition to function of time $s(t)$, which completely defines a signal, other time characteristic – **correlation function** of a signal in some cases is used. For a real signal of final duration it is defined

$$K_s(\tau) = \int_0^{T_s} s(t)s(t+\tau)dt, \tag{2.5}$$

where τ – time shift which accepts both positive, and negative values. At $\tau = 0$

$$K_s(0) = E_s. \tag{2.6}$$

For a periodic signal with period T which energy indefinitely big, is used the following definition

$$K_{s\,\mathrm{per}}(\tau) = \frac{1}{T}\int_0^{T} s(t)s(t+\tau)dt. \tag{2.7}$$

Function $K_{s\,\mathrm{per}}(\tau)$ is periodic with period T, and

$$K_{s\,\mathrm{per}}(0) = P_s. \tag{2.8}$$

There are two signals $s_1(t)$ and $s_2(t)$. Concepts of **mutual correlation function** is introduced for them

$$K_{s1s2}(\tau) = \int_0^{T_s} s_1(t)s_2(t+\tau)dt \tag{2.9}$$

and **scalar product** are entered

$$(s_1, s_2) = \int_0^{T_s} s_1(t)s_2(t)dt. \tag{2.10}$$

From last ratio it is clear, that $(s_1, s_2) = K_{s1s2}(0)$.

Signals are **orthogonal** if scalar product of signals is $(s_1, s_2) = 0$.

Let consider definition of energy characteristics for the rectangular video pulse (fig. 7). Analytically record of the rectangular video pulse is:

$$s(t)=\begin{cases} A, & 0 \le t < T_s, \\ 0, & t<0,\ t \ge T_s. \end{cases}$$

Average power and energy are defined by ratio (2.1) and (2.3)

$$P_s = \frac{1}{T_s}\int_0^{T_s} A^2 dt; \quad E_s = A^2 T_s .$$

The correlation function of a signal is defined by a ratio (2.5). Let $0 < \tau < T_s$. Then

$$s(t)s(t+\tau)=\begin{cases} A^2, & 0 \le t < T_s - \tau, \\ 0, & t<0,\ t \ge T_s - \tau. \end{cases}$$

and

$$K_s(\tau)= \int_0^{T_s-\tau} A^2 dt = A^2(T_s - \tau).$$

When $\tau \ge T_s$ then $K_s(\tau) = 0$. In view of parity property of correlation function final expression is written as

$$K_s(\tau)=\begin{cases} A^2(T_s - |\tau|), & |\tau| < T_s, \\ 0, & |\tau| \ge T_s. \end{cases}$$

The graph of correlation function of the rectangular video pulse is shown on fig. 7.

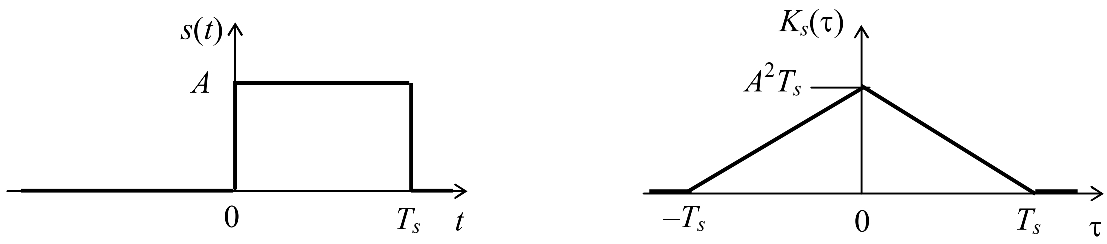


Figure 7 – Rectangular video pulse $s(t)$ and its correlation function $K_s(\tau)$

2.3 Representation of signals in orthogonal bases

Among various mathematical methods, which are used while describing electric circuits and signals, representation of any function as the sum of more simple ("elementary") functions is the most widely applied. Let $s(t)$ is the determined signal of duration T_s. Let present its weighed sum of some **basic functions** as

$$s(t) = \sum_{n=1}^{\infty} a_n \psi_n(t), \quad 0 \le t \le T_s, \tag{2.11}$$

where a_n – factors of decomposition; $\psi_n(t)$ – basic functions.

Basic functions are selected according to those or other reasons, and then factors of decomposition calculate. But factors of decomposition are calculated in more simple way if basic functions are orthogonal on an interval (0, T_s). Let multiply the left and right parts of equality (1) on ψ_k *(t)* also we make integration on an interval (0, T_s):

$$\int_0^{T_s} s(t)\psi_k(t)dt = \int_0^{T_s}\sum_{n=1}^{\infty} a_n\psi_n(t)\varphi_k(t)dt = \sum_{n=1}^{\infty} a_n \int_0^{T_s}\psi_n(t)\psi_k(t)dt\,.$$

As functions $\psi_n(t)$ and $\psi_k(t)$ are orthogonal integrals in the right part are equal to zero except for a case $n = k$ – in this case the integral is equal to energy of basic functions $\psi_k(t)$. Therefore last equality will be written down as

$$\int_0^{T_s} s(t)\psi_k(t)dt = a_k E_{\psi_k}\,.$$

Let return to an index n and we shall write down a rule of **calculation of decomposition factors**

$$a_n = \frac{1}{E_{\psi n}}\int_0^{T_s} s(t)\psi_n(t)dt\,; \quad n = 1, 2, 3, ... \tag{2.12}$$

If basic functions are orthonormal, so

$$a_n = \int_0^{T_s} s(t)\psi_n(t)dt = (s, \varphi_n), \quad n = 1, 2, 3, ... \tag{2.13}$$

A series (2.11), in which decomposition factors are defined according to the formula (2.12), is called as **generalized Fourier series**.

While practical using of decomposition (2.11) it is necessary to limit the number of terms

$$\hat{s}(t) = \sum_{n=1}^{N} a_n\psi_n(t)\,. \tag{2.14}$$

Thus the approximate representation of a signal $s(t)$ is got, which satisfies to the certain measure of accuracy

$$E_\varepsilon = \int_0^{T_s}[s(t) - \hat{s}(t)]^2 dt = E_s - \sum_{n=1}^{N} a_n^2 E_{\psi_n}\,. \tag{2.15}$$

It is usually considered, that the number N is selected in the way so that to satisfy to the given measure of accuracy, and in writing of decomposition of a signal $s(t)$, a sign of exact equality is used

$$s(t) = \sum_{n=1}^{N} a_n\psi_n(t)\,. \tag{2.16}$$

Depending on properties of a signal $s(t)$ different systems of orthogonal functions are used: trigonometrical functions, exponent functions, functions of samples and Walsh functions.

After series expansion of a signal $s(t)$, decomposition factors completely set a signal $s(t)$, i.e. according to factors it is possible to recover a signal. Factors of decomposition also allow to define energy characteristics of signal $s(t)$:

$$E_s = \int_0^{T_s} \left(\sum_{n=1}^{N} a_n \psi_n(t) \right)^2 dt = \sum_{n=1}^{N} a_n^2 E_{\psi_n} \tag{2.17}$$

and scalar product of signals $s_1(t)$ and $s_2(t)$

$$(\mathbf{s_1}, \mathbf{s_2}) = \int_0^{T_s} \left(\sum_{n=1}^{N} a_{n1} \psi_n(t) \right) \left(\sum_{n=1}^{N} a_{n2} \psi_n(t) \right) dt = \sum_{n=1}^{N} a_{n1} a_{n2} E_{\psi_n}\,, \tag{2.18}$$

where a_{n1} and a_{n2} – decomposition factors of signals $s_1(t)$ and $s_2(t)$ appropriately.

Definition of decomposition factors can be made with hardware as shown on fig. 8, *a*. This procedure is called the signal analysis. Predicted decomposition factors completely describe a signal. Knowing them, it is possible to synthesize a signal – to recover it according to decomposition factors (fig. 8, *b)*. The circuits shown on fig. 8 find application in communication techniques. In a transmitter the analysis of signals is performed, on a transmission channel the decomposition factors are transferred, in the receiver synthesis of a signal is performed.

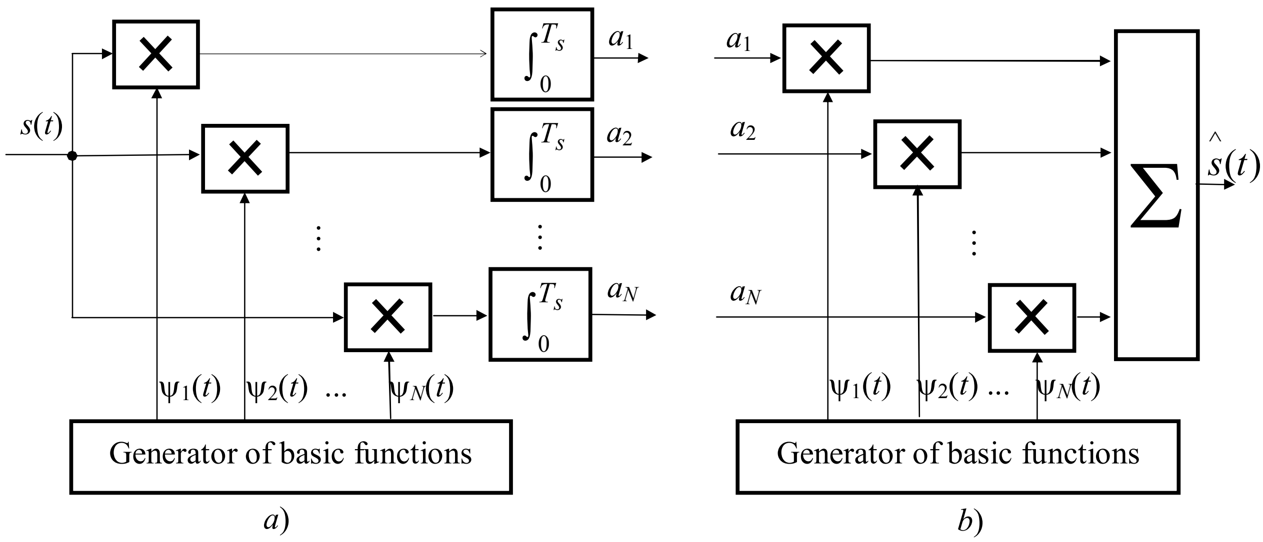


Figure 8 – Circuits of signal analysis (*a*) and synthesis (*b*)

2.4 Geometrical representation of signals

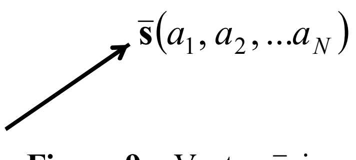


Figure 9 – Vector $\overline{\mathbf{s}}$ in N-dimensional space

While mathematical describing signals are convenient to consider it as **vectors or points in some space** (fig. 9). Let remember, that a vector is a segment of a given direction and lengths. Usually a vector is set by coordinates of its end. The system of coordinates should be set by unit vectors, angles between which are equal 90°. So, to a signal $s(t)$ the vector $\bar{\mathbf{s}}$ is put in conformity.

The basic ratios for N-dimensional linear metric space are:

– **length** (norm) of a vector $\bar{\mathbf{s}}$

$$\|\bar{\mathbf{s}}\| = \sqrt{\sum_{n=1}^{N} a_n^2}\ ; \tag{2.19}$$

– **distance** between vectors $\bar{\mathbf{s}}_1$ and $\bar{\mathbf{s}}_2$

$$d(\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2) = \sqrt{\sum_{n=1}^{N} (a_{1n} - a_{2n})^2}\ ; \tag{2.20}$$

– **scalar product** of vectors $\bar{\mathbf{s}}_1$ and $\bar{\mathbf{s}}_2$

$$(\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2) = \sum_{n=1}^{N} a_{1n} \cdot a_{2n}\ . \tag{2.21}$$

Total sum of all functions of time set on an interval (0, T_s) is called a functional space. These functions are considered as vectors in functional space. Coordinates of these vectors are values of functions of time. It is clear, that $N \to 0$, and ratios (2.19) – (2.21) are passing in the following

– **length** (norm) of a vector $\bar{\mathbf{s}}$

$$\|\bar{\mathbf{s}}\| = \sqrt{\int_0^{T_s} s^2(t)dt} = \sqrt{E_s}\ , \tag{2.22}$$

it is important to remember, that the length of a vector is equal to a root from energy of a signal;

– **distance** between vectors $\bar{\mathbf{s}}_1$ and $\bar{\mathbf{s}}_2$

$$d(\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2) = \sqrt{\int_0^{T_s} (s_1(t) - s_2(t))^2 dt}\ ; \tag{2.23}$$

– **scalar product** of vectors $\bar{\mathbf{s}}_1$ and $\bar{\mathbf{s}}_2$

$$(\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2) = \int_0^{T_s} s_1(t) s_2(t) dt\ . \tag{2.24}$$

Let address to decomposition of signals in generalized Fourier series

$$s(t) = \sum_{n=1}^{N} a_n \psi_n(t)\ . \tag{2.25}$$

We take into consideration, that basic functions are orthonormal and are normalized. Let rewrite a ratio in the vector form

$$\bar{\mathbf{s}} = \sum_{n=1}^{N} a_n \psi_n\ . \tag{2.26}$$

Hence, if a signal is decomposed into generalized Fourier series it can be presented in N-dimensional space.

2.5 Spectral analysis of periodic signals

In the theory and techniques of communication the **trigonometrical basis** is the most widely applied to the decomposition of signals in series. Wide application of these functions in the theory and techniques of communication is caused by that while passing through linear circuits the form of each of them does not change – only their levels and phases change (there is a shift in time).

For a periodic signal Fourier series:

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos 2\pi n f_1 t + b_n \sin 2\pi n f_1 t). \tag{2.27}$$

where $f_1 = 1/T$, T is period of a signal $s(t)$;

$$a_n = \frac{2}{T}\int_{-T/2}^{T/2} s(t)\cos 2\pi n f_1 t dt, \quad n = 0, 1, 2, ...; \quad b_n = \frac{2}{T}\int_{-T/2}^{T/2} s(t)\sin 2\pi n f_1 t dt, \quad n = 1, 2, ...$$

If to enter definitions

$$A_n = \sqrt{a_n^2 + b_n^2}\,, \; \varphi_n = -\mathrm{arctg}\frac{b_n}{a_n},$$

that a series (1) will be transformed in:

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} A_n \cos(2\pi n f_1 t + \varphi_n). \tag{2.27a}$$

Decomposition of a signal (2.27) and (2.27a) are called **Fourier series in a trigonometrical form**. It is more convenient to use a series (2.27a) as it directly establishes, what harmonious components a signal will consist of – what values of their frequencies nf_1, amplitudes A_n and initial phases φ_n. Series (2.27), defines a signal spectrum.

Obvious representation of a spectrum is given by fig. 10 which an **amplitude spectrum** is built on – dependence of amplitudes A_n on frequency and a **phase spectrum** – dependence of initial phases φ_n from frequency. Frequency f_1 is called a basic frequency of a signal, it is equal to number of the periods of a signal in a second. Harmonious fluctuations with frequencies nf_1 (n = 2, 3, ...) are called harmonics of a signal $s(t)$: $2f_1$ –second harmonic, $3f_1$ –third harmonic, etc.

The amplitude spectrum allows to define bandwidth, as dimension of frequency range where total energy of components equals the certain share of full energy. If a considered signal spectrum adjoins to zero frequency its bandwidth is expressed by number F_{max}. Using F_{max}, assume that the signal does not contain frequencies upper F_{max}.

Spectral representation of a periodic signal can be made, using exponential basic functions

$$\varphi_n(t) = e^{j2\pi n f_1 t}, \; n = ..., -1, 0, 1, 2, ...$$

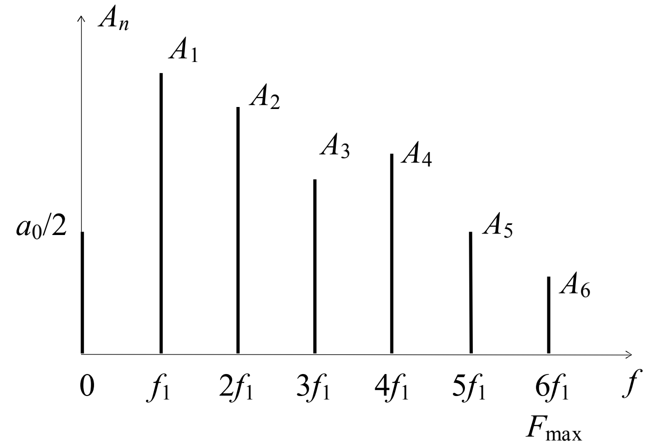


Figure 10 – Amplitude spectrum of a periodic signal

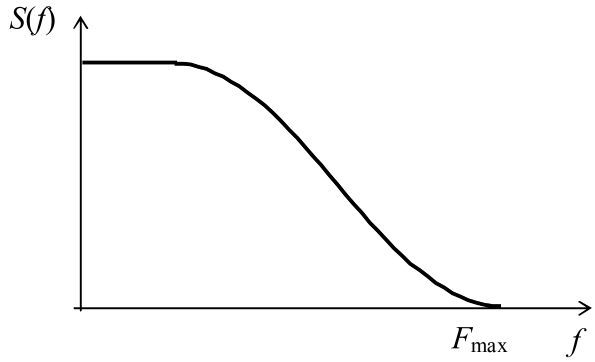


Figure 11 – Amplitude spectrum of a nonperiodic signal

Thus series is written as

$$s(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_1 t}. \tag{2.28}$$

Such decomposition of a signal is called **Fourier series in a complex form**. Coefficients of decomposition are defined as:

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} s(t) e^{-j2\pi n f_1 t} dt, \; n = ..., -1, 0, 1, 2, ...$$

Thus:

$$c_n = \frac{a_n}{2} - j\frac{b_n}{2} = \frac{A_n}{2} e^{j\varphi_n}.$$

Feature of Fourier series in a complex form is a compact record of series and coefficients of decomposition. Other feature is using of negative frequencies.

The spectrum, corresponding to series (2.28), refers to two-sided spectrum.

2.6 Spectral analysis of nonperiodic signals

Expressions:

$$S(j\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt \quad \text{and} \quad s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega) e^{j\omega t} d\omega \tag{2.29}$$

make a pair of **direct and inverse Fourier transforms**. Function $S(j\omega)$ refers to as spectral density of a signal. In general case spectral density $S(j\omega)$ is a complex function. It is determined on an interval $(-\infty, \infty)$. Let present it through a module and argument $S(j\omega) = S(\omega)e^{j\varphi(\omega)}$.

Function $S(\omega)$ refers to as an **amplitude spectrum** of a signal, and $\varphi(\omega)$ – a **phase spectrum** of a signal. Function $S(\omega)$ – even function of frequency.

Many signals possess even symmetry (it is achieved by a corresponding choice of a reference mark of time). Spectral density of such signals is a real function

$$S(\omega) = 2\int_0^{\infty} s(t)\cos\omega t dt\,.$$

Because of evenness of function $S(\omega)$ inverse Fourier transform is

$$s(t) = \frac{1}{\pi}\int_0^{\infty} S(\omega)\cos\omega t d\omega\,.$$

Last two integrals make a pair of **Fourier cosine-transforms**.

Basic difference of spectrum is (fig. 10 and 11), that a nonperiodic signal has a continuous spectrum, and periodic signal has a discrete spectrum, it contains harmonics of frequency $f_1 = 1/T$.

The apparatus of Fourier transforms is rather effective mathematical means to solve many problems of theory and techniques of communication. Note only some **properties of Fourier transforms**.

1. **Product of two signals** (a general case):

$$s(t) = s_1(t)s_2(t),$$

$$S(j\omega) = S_1(j\omega) * S_2(j\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_1(j\nu)S_2(j(\omega-\nu))d\nu = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_2(j\nu)S_1(j(\omega-\nu))d\nu \; -$$

multiplication of signals in time domain corresponds to convolution of their spectrum.

2. **Convolution of signals**

$$s(t) = s_1(t) * s_2(t) = \int_{-\infty}^{\infty} s_1(\tau)s_2(t-\tau)d\tau = \int_{-\infty}^{\infty} s_2(\tau)s_1(t-\tau)d\tau,$$

$$S(j\omega) = S_1(j\omega)\cdot S_2(j\omega).$$

3. Calculation of **signals energy**

$$E_s = \int_{-\infty}^{\infty} |s(t)|^2\, dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |S(j\omega)|^2\, d\omega \; -$$

this ratio is called Parseval relation.

4. **Scalar product** of signals:

$$(s_1, s_2) = \int_{-\infty}^{\infty} s_1(t)s_2(t)dt; \qquad (s_1, s_2) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_1(j\omega)S_2^*(j\omega)d\omega.$$

Having equated last ratio to zero, we shall receive a condition of orthogonality of signals given by spectral density.

2.7 Kotelnikov theorem and series

Kotelnikov theorem states: the signal $s(t)$, not containing frequencies higher than F_{max}, can be strictly recovered on the samples, taken through an interval $T_s \le 1/(2F_{max})$; T_s – sampling interval; $f_s = 1/T_s$ – sampling frequency.

As for any real signal it is possible to specify a highest frequency of a spectrum it is possible to suppose, that Kotelnikov theorem can be applied to all real signals.

It is possible to show, that a **spectral density of a discrete signal is periodic repetition with the period f_s of spectral density of a continuous signal** from which the discrete signal is received. It is illustrated by graph: on fig. 12, *a* an amplitude spectrum of any continuous signal with the maximal frequency F_{max} of a spectrum is shown; on fig. 12, *b* its periodic repetition is shown (figure is constructed for a case $T_s < 1/(2F_{max})$ or $f_s > 2F_{max}$. From fig. 12, *b* it is understand, that at $f_s \ge 2F_{max}$ on a discrete signal (samples) with LPF it is possible to recover an initial continuous signal (by dotted line it is shown amplitude response AR of the recovering filter). At $f_s < 2F_{max}$ there is an imposing periodic repetitions of a spectrum, and to recover without an error a continuous signal it is impossible. Thus Kotelnikov theorem is proved.

In time domain connection between a continuous and discrete signal is described by **Kotelnikov series**

$$s(t) = \sum_{n=-\infty}^{\infty} s(nT_s) \frac{\sin 2\pi F_{max}(t - nT_s)}{2\pi F_{max}(t - nT_s)}.$$

Values $s(nT_s)$ are coefficients of decomposition of a signal $s(t)$ on known system of orthogonal basic functions known from mathematics

$$\varphi_n(t) = \frac{\sin 2\pi F(t - nT_s)}{2\pi F(t - nT_s)}, \ n = ..., -1, 0, 1, 2, ...$$

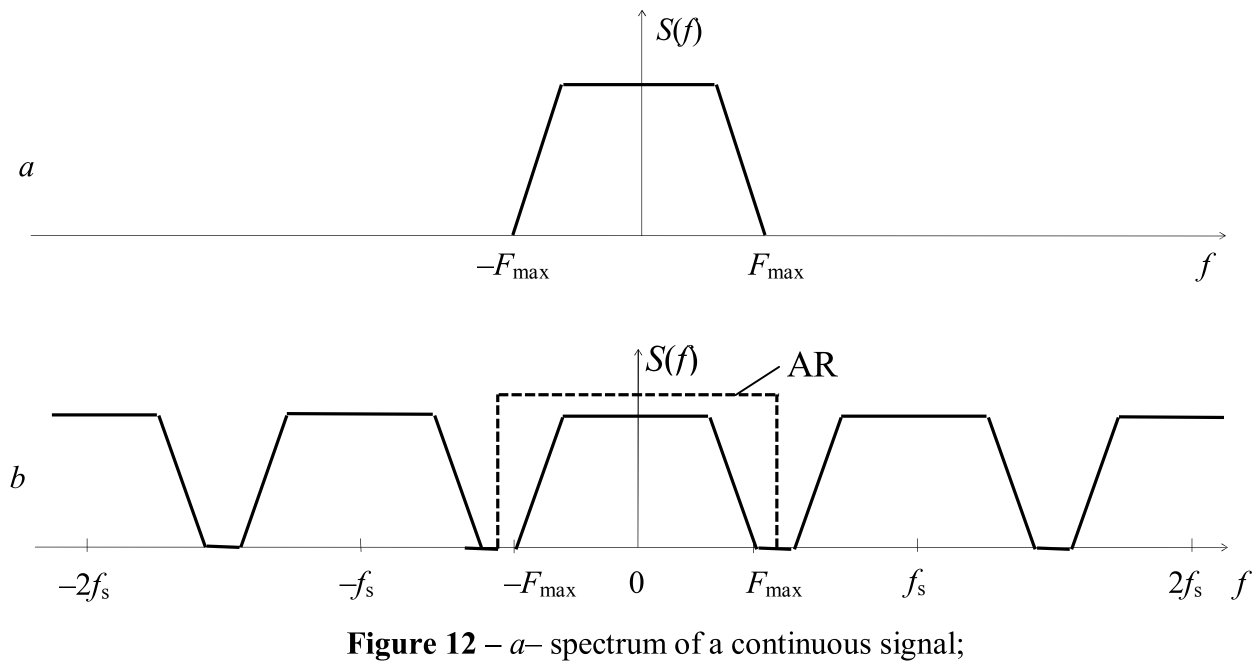


Figure 12 – *a*– spectrum of a continuous signal;
b – spectrum of a discrete signal and AR of a recovering filter

The graphic illustration of Kotelnikov series is given on fig. 13

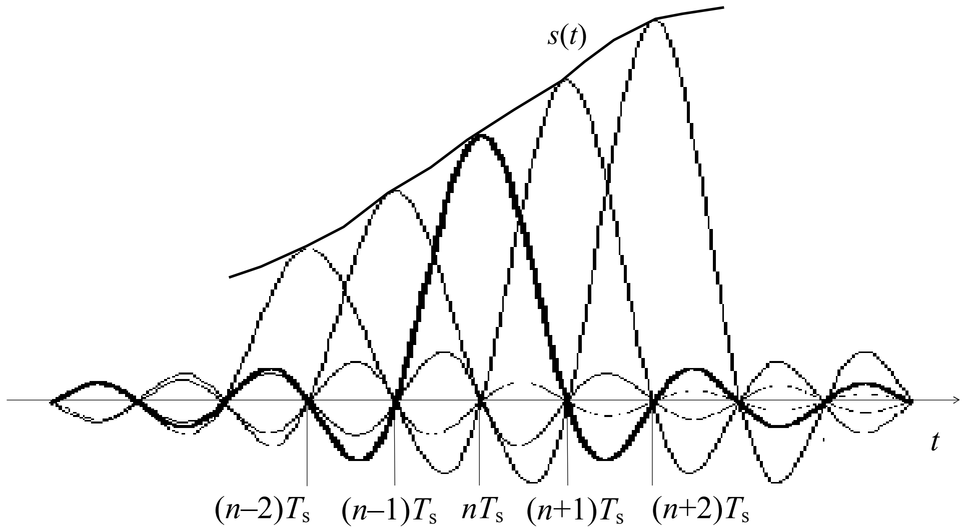


Figure 13 – Signal $s(t)$ and five components of Kotelnikov series

2.8 Representation of bandpass signals

A **signal refers to bandpass** if its spectrum does not adjoin to zero frequency. Their spectra are concentrated in a frequency band from f_{min} to f_{max}, and $f_{min} > 0$ (fig. 14). To describe bandpass signals such parameters are presented: a middle frequency of a spectrum $f_0 = 0{,}5(f_{min} + f_{max})$ and bandwidth of a spectrum $\Delta F = f_{max} - f_{min}$. As a rule, for bandpass signals the relation $\Delta F << f_0$ is carried out, and then they are refered to **narrow-band signals**. Narrow-band signals look like quasi-harmonic oscillations with middle frequency f_0 (fig. 15) in time domain.

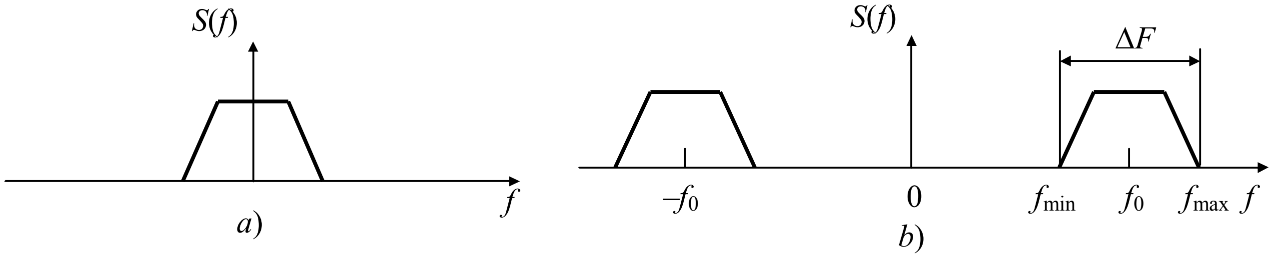


Figure 14 – Spectrum of low-frequency (a) and bandpass (b) signals

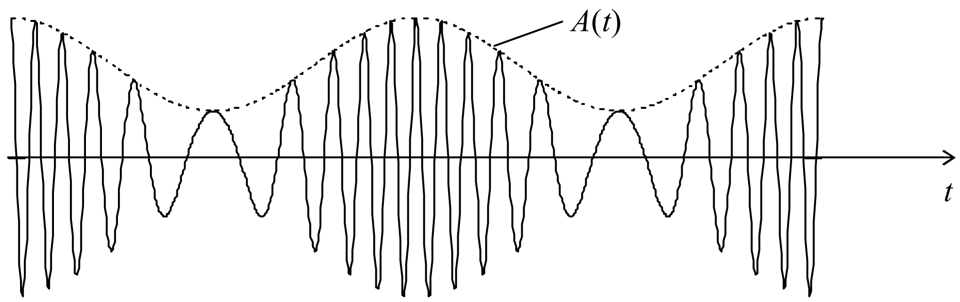


Figure 15 – Time diagram of a bandpass signal

Any bandpass signal can be presented by following mathematical expression:

$$s(t) = A(t)\cos[\psi(t)], \tag{2.30}$$

where $A(t)$ – an **envelope of a bandpass signal**;

$\psi(t)$ – a **full phase of a bandpass signal**.

An envelope of a bandpass signal is positively certain function, i.e. $A(t) \geq 0$, not being crossed with a signal, it has with it common points at moments, when instant value of a signal is maximal. A full phase of a bandpass signal will consist of three components:

$$\psi(t) = \omega_0 t + \varphi(t) + \varphi_0, \tag{2.31}$$

where $\varphi(t)$ – increment of a phase;

φ_0 – initial phase.

An increment of a phase causes changing of **instant frequency** of a signal. By definition frequency of a signal is a speed of its phase changing, i.e.:

$$\omega(t) = \frac{\partial\psi(t)}{\partial t} = \omega_0 + \frac{\partial\varphi(t)}{\partial t}. \tag{2.32}$$

The integral of instant frequency gives a full phase of a signal:

$$\psi(t) = \int_0^t \omega(t)dt + \varphi_0. \tag{2.33}$$

It is widely used **quadrature representation of bandpass signals**

$$\begin{aligned} s(t) &= A(t)\cos[\omega_0 t + \varphi(t) + \varphi_0] = \\ &= A(t)\cos[\varphi(t) + \varphi_0]\cos[\omega_0 t] - A(t)\sin[\varphi(t) + \varphi_0]\sin[\omega_0 t] = \\ &= I(t)\cos[\omega_0 t] - Q(t)\sin[\omega_0 t], \end{aligned} \tag{2.34}$$

where $I(t) = A(t)\cos[\varphi(t) + \varphi_0]$ – inphase or cosine component;

$Q(t) = A(t)\sin[\varphi(t) + \varphi_0]$ – quadrature or sinus component.

If quadrature components $I(t)$ and $Q(t)$ are known, then it is possible to find an envelope and full phase of a bandpass signal:

$$A(t) = \sqrt{I^2(t) + Q^2(t)};$$

$$\psi(t) = \omega_0 t + \mathrm{arctg}\left(\frac{Q(t)}{I(t)}\right).$$

One more form of representation of bandpass signals is the complex form $\dot{s}(t)$:

$$s(t) = \mathrm{Re}[\dot{s}(t)] = \mathrm{Re}\left[A(t)e^{j\psi(t)}\right] = A(t)\cos[\psi(t)].$$

While analysing of bandpass signals in the **complex form** a concept of complex envelope of a signal is entered:

$$\dot{s}(t) = A(t)e^{j\psi(t)} = A(t)e^{j(\omega_0 t + \varphi(t) + \varphi_0)} = A(t)e^{j\omega_0 t}e^{j(\varphi(t) + \varphi_0)} = \dot{A}(t)e^{j\omega_0 t}, \tag{2.35}$$

where $\dot{A}(t)$ – **complex envelope** of a bandpass signal.

Complex envelope has the following form:

$$\dot{A}(t) = A(t)e^{j(\varphi(t)+\varphi_0)} = A(t)\cos[\varphi(t)+\varphi_0] + jA(t)\sin[\varphi(t)+\varphi_0] = I(t) + jQ(t).$$

2.9 Analytical signal

Complex signal $\dot{x}(t) = x(t) + j\tilde{x}(t)$ refers to an **analytical** one, if $\tilde{x}(t)$ is **Hilbert transform** from $x(t)$. On figures a complex signal is represented as two circuits, as it is shown on fig. 16.

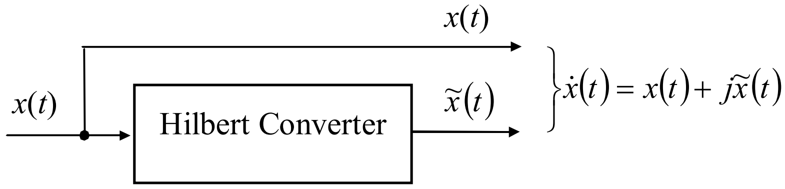


Figure 16 – Production of analytical signal

Hilbert converter is a linear circuit with the impulse response

$$g(t) = \frac{1}{\pi t}, \qquad -\infty < t < \infty.$$

Let take advantage of Duhamel integral

$$\tilde{x}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau.$$

This relation refers to **Hilbert transform** of a signal $x(t)$. Let we find transfer function of Hilbert converter like Fourier transformation from impulse response

$$H(j\omega) = \int_{-\infty}^{\infty} g(t)e^{-j\omega t}dt = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{-j\omega t}}{\pi t}dt = \begin{cases} -j & \text{at } \omega > 0, \\ j & \text{at } \omega < 0. \end{cases} \tag{2.36}$$

Or another form:

$$H(j\omega) = -j\ \text{sign}(\omega).$$

Let $S_x(j\omega)$ is a spectral density of a signal $x(t)$. Then spectral density of a signal $\tilde{x}(t)$ is defined as

$$S_{\tilde{x}}(j\omega) = \begin{cases} -j\ S_x(j\omega) & \text{at } \omega > 0, \\ j\ S_x(j\omega) & \text{at } \omega < 0. \end{cases} \tag{2.37}$$

Let we find spectral density of an analytical signal

$$S_{\dot{x}}(j\omega) = S_x(j\omega) + jS_{\tilde{x}}(j\omega) = \begin{cases} 2S_x(j\omega) & \text{at } \omega > 0, \\ 0 & \text{at } \omega < 0. \end{cases} \tag{2.38}$$

We have revealed the **important property of an analytical signal – its spectrum on negative frequencies is equal to zero** (fig. 17).

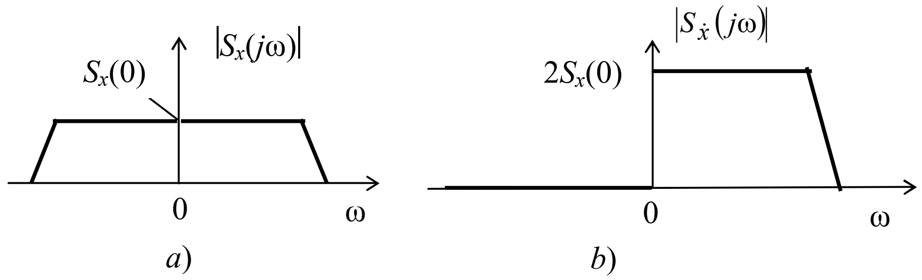


Figure 17 – Spectral density: *a* – a signal $x(t)$;
b – an analytical signal corresponding to it $|S_{\dot{x}}(j\omega)|$

Inverse Hilbert transform is

$$x(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\tilde{x}(\tau)}{t-\tau}d\tau.$$

The module of an analytical signal

$$A(t) = \sqrt{x^2(t) + \tilde{x}^2(t)}$$

is an envelope of a signal $x(t)$, and an argument

$$\varphi(t) = \operatorname{arctg}\frac{\tilde{x}(t)}{x(t)}$$

is a phase of a signal $x(t)$.

From last expressions follows that the analytical signal can be written down as:

$$\dot{x}(t) = A(t)e^{j\varphi(t)} \tag{2.39}$$

Thus, concepts of an envelope and a phase of a signal can be applied not only to bandpass signal, but also to baseband signals. An envelope satisfies two conditions: $A(t) \geq |x(t)|$ – function $A(t)$ does not cross function $x(t)$ anywhere and in points of contact of functions $A(t)$ and $x(t)$ their derivatives are equal: $A'(t) = x'(t)$, that is functions have the common tangents.

2.10 Sampling of bandpass signals

Representation of bandpass signals by discrete signals is necessary, when transformation of signals (a filtration, detecting, etc.) is carried out by digital signal processors. In case of bandpass signals, and especially narrow-band signals, a sampling frequency can be essentially less $2f_{\max}$.

The spectral density of a discrete signal writes down:

$$S_s(j2\pi f) = f_s\sum_{n=-\infty}^{\infty}S(j2\pi(f - nf_s)), \qquad -\infty < f < \infty, \tag{2.40}$$

where $S(j2\pi f)$ is spectral density of a continuous signal $s(t)$.

From this expression comes next: the **spectrum of a discrete signal is an infinite sum of periodic repetition of a spectrum $S(j2\pi f)$ of a continuous signal $s(t)$ with the period f_s,** and scale multiplier f_s.

On fig. 18,*a* the amplitude spectrum of any form $S(f)$ of bandpass signal is shown. It is concentrated on an interval (f_{min}, f_{max}). On fig. 18,*b* the amplitude spectrum of a discrete signal which can take place at sampling of a signal with a spectrum shown on fig. 18,*a* is represented.

Components of a discrete signal spectrum, which are caused by periodic repetition of frequency bands (f_{min}, f_{max}) and ($-f_{max}, -f_{min}$) are designated with "filling" different density for descriptive reasons. For impossibility of spectrum components with different *n* aliasing, a choice of value f_s is made on the basis of inequalities

$$\frac{2f_{max}}{k+1} \le f_s \le \frac{2f_{min}}{k}, \quad k = 0, 1, 2, ..., k_{max}. \tag{2.41}$$

In a case, when $k = 0$, a sampling frequency $f_s \ge 2f_{max}$, i.e. this condition of a sampling frequency choice for baseband signals, which satisfies Kotelnikov theorem. When $k > 0$, then representation of a bandpass signal by a discrete signal becomes more economical (smaller number of samples). The most economical representation of a signal will be, when $k = k_{max}$.

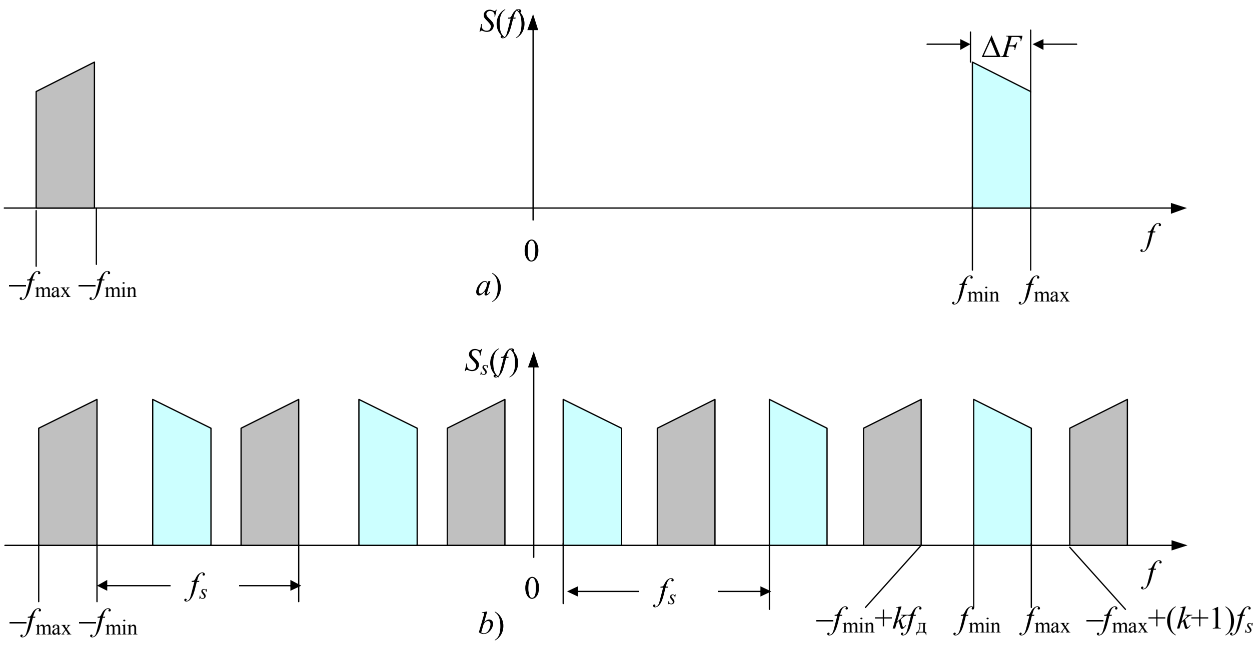


Figure 18 – Spectrum of continuous bandpass signal and discrete signal

A bandpass filter carries out recovering of a continuous bandpass signal on samples. Its lower cutoff frequency $f_{l\ cut}$ is no more than f_{min}, and upper cutoff frequency $f_{u\ cut}$ is not less than f_{max}.

Questions for section 2

1. Define continuous, discrete, quantized and digital signals.
2. Define baseband and modulated signals.
3. Define deterministic and random signals.

4. Define periodic and non-periodic signals.
5. What is the difference between signal and oscillation?
6. Define the instantaneous power of the signal. What is its dimension and what does it represent?
7. Define the average power and the signal energy.
8. Under what conditions the signal is called normalized?
9. Define correlation function of signal.
10. What is called the generalized Fourier series? Write down the expression and the expression of the calculation of the expansion coefficients.
11. Explain the concepts of analysis and synthesis of signal.
12. Write down the basic equations for an N-dimensional linear metric space.
13. Write down the form of the Fourier series at trigonometrically basis.
14. Define the amplitude and phase spectrum of the periodic signal.
15. Define the concepts of fundamental frequency and harmonics of the signal width of the signal spectrum.
16. How to conduct the spectral analysis of periodic and non-periodic signals? What is the fundamental difference in their spectra?
17. Give known properties of the Fourier transformation.
18. What is the feature of the spectrum of a discrete signal?
19. Formulate the sampling (Nyquist, Kotelnikov) theorem.
20. What is called to Nyquist series, and what does it illustrate?
21. Define bandpass and narrowband signal. As they are described, what are their basic features?
22. Write down the form of a bandpass signal.
23. What kind of signal is called analytical and what is necessary to obtain it?
24. Explain the features of the sampling bandpass signals.

3 DESCRIPTION OF RANDOM PROCESSES

3.1 Classification of random processes

Random (stochastic) process is random function of time. Main feature of random process is that its values cannot be precisely predicted beforehand. Random function of time is described by **statistical characteristics** that characterize those or other properties of this function on the **average**.

Random processes are more full mathematical models of communication signals, than the determined functions of time. Many tasks of the theory and techniques of communication can be solved only at the description of signals and noise by random functions. For example, the voltage of a noise on an output of a transmission line or on an output of a microphone is random function of time.

Let designate considered random process as $X(t)$. Separate supervisions of process give different functions $x(t)$ – different **realizations** of random process. Set all possible realizations of the given random process $\{x_k(t)\}$ is called **ensemble** (fig. 19).

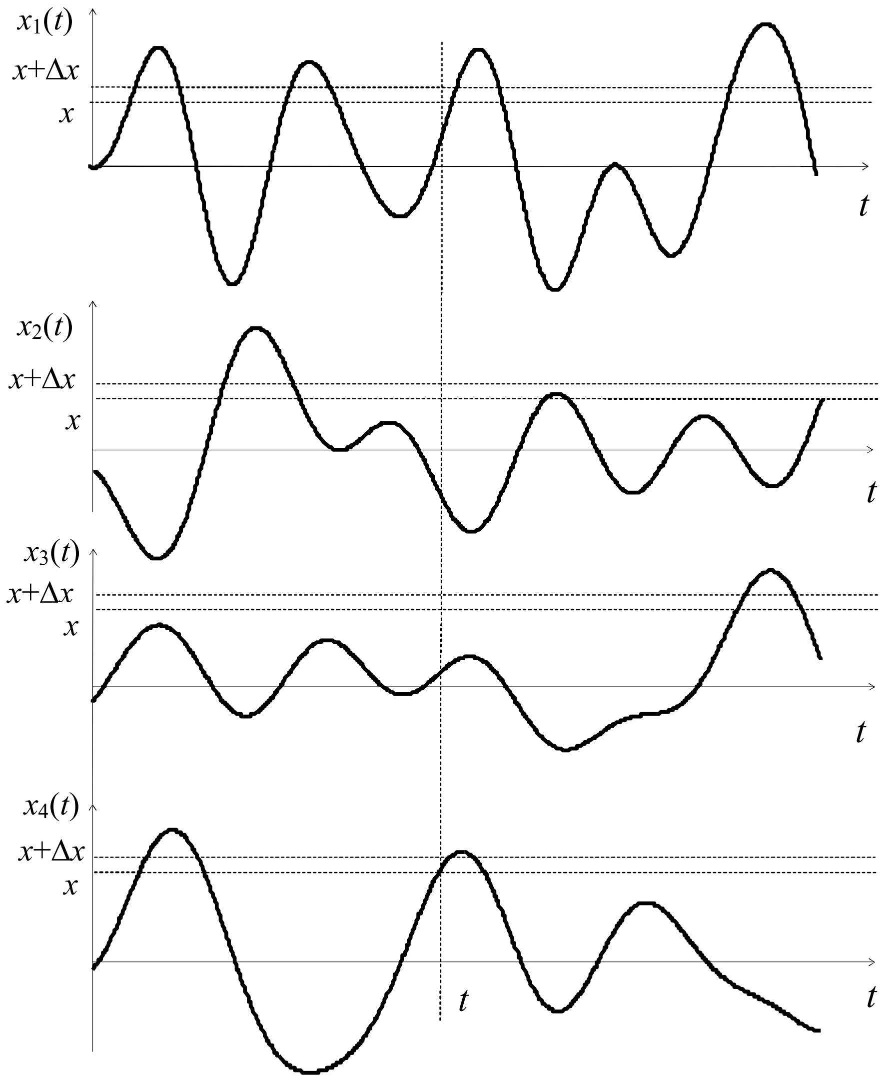


Figure 19 – Ensemble of realizations

The description of random process gives possibility to define some average characteristics of ensemble $\{x_k(t)\}$ as a whole. Such characteristics refer to statistical.

Random process refers to **stationary** if its statistical characteristics do not change with time.

Stationary random process refers to **ergodic** if its statistical characteristics are found by **averaging on ensemble**, coincide with the characteristics found by **averaging one realization in time.**

Let consider further, if it is not stipulated another, that considered random process is stationary and ergodic.

3.2 Probabilistic characteristics of random processes

Probabilistic characteristics are **probability distribution function** and **probability density function**. Probabilistic characteristics are the most used among statistical characteristics of random processes.

By definition, the value of probability distribution function $F(x)$ is equal to the probability of that in the arbitrary time moment process $X(t)$ will take on the value that does not exceed x:

$$F(x) = P\{X(t) \le x\}. \tag{3.1}$$

By definition, the value of probability density function $p(x)$ is equal to the limit of ratio of probability of that in the arbitrary time moment process $X(t)$ will take on the value on interval (x - $\Delta x/2$, $x + \Delta x/2$) to the interval length Δx when $\Delta x \to 0$:

$$p(x) = \lim_{\Delta x \to 0} \frac{P\{x - \Delta x/2 < X(t) \le x + \Delta x/2\}}{\Delta x}. \tag{3.2}$$

The properties of $F(x)$ and $p(x)$ functions shown on the table 1 are easy to prove using their definitional formulas (3.1) and (3.2).

Table 1 – The properties of the functions $F(x)$ and $p(x)$

	$p(x)$	$F(x)$
1	$P\{x < X(t) \le x + dx\} = p(x)dx$	$F(x) = P\{X(t) \le x\}$
2	$P\{x_1 < X(t) \le x_2\} = \int_{x_1}^{x_2} p(x)dx$	$P\{x_1 < X(t) \le x_2\} = F(x_2) - F(x_1)$
3	$\int_{-\infty}^{\infty} p(x)dx = 1$	$F(\infty) = 1; \quad F(-\infty) = 0$
4	$p(x) \ge 0$	$F(x_2) \ge F(x_1)$ when $x_2 > x_1$
5	$p(x) = \frac{dF(x)}{dx}$	$F(x) = \int_{-\infty}^{x} p(x)dx$

The considered functions (3.1) and (3.2) are **one-dimensional distributions** of probabilities. They characterize process only during one moment of time. **Two-dimensional distribution function** and **two-dimensional probability density function** characterize process during two moments of time t and $t + \tau$.

The two-dimensional probability distribution function of process $X(t)$ is defined as

$$F_2(x_1, x_2, \tau) = P\{X(t) \le x_1;\ X(t+\tau) \le x_2\}. \tag{3.3}$$

The two-dimensional probability density function of process $X(t)$ is defined as

$$p_2(x_1, x_2, \tau) = \frac{\partial^2 F_2(x_1, x_2, \tau)}{\partial x_1 \partial x_2}. \tag{3.4}$$

For n = 3, 4, ... moments of time by analogy with (3.3) and (3.4), ***n*-dimensional distributions of probability** can be found. The more value n, the more full random process is described. But consideration of n-dimensional distributions demands complex process of realizations $x_k(t)$. Knowledge of one-dimensional and two-dimensional distributions of probabilities is used for solving of many problems.

3.3 Numerical characteristics of processes

Many problems can be solved, using more "rough" characteristics of processes, than using of probability distributions.

Average value (or expectation) $\overline{X(t)}$ **of process** is equal to value around of which process accepts the values. Knowing probability density of process, we can define

$$\overline{X(t)} = \int_{-\infty}^{\infty} x p(x) dx. \tag{3.5}$$

Average value on time is designated by wavy line above dependence, which is averaged on time. So, average in time value of process $X(t)$ is defined by averaging on time of realization $x(t)$

$$\widetilde{X(t)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt. \tag{3.6}$$

Average power of process is average value of a square of process

$$P_X = \overline{X^2(t)} = \int_{-\infty}^{\infty} x^2 p(x) dx \tag{3.7}$$

or

$$P_X = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt. \tag{3.8}$$

Average value of a square of deviations from average value is a variance of process

$$D\{X(t)\} = \overline{[X(t) - \overline{X(t)}]^2} = \int_{-\infty}^{\infty} [x - \overline{X(t)}]^2 p(x) dx. \tag{3.9}$$

or

$$D\{X(t)\} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left[x(t) - \overline{X(t)}\right]^2 dt. \quad (3.10)$$

Positive root from variance is

$$\sigma_X = \sqrt{D\{X(t)\}}. \quad (3.11)$$

It is called a **root-mean-square deviation of a process**.

3.4 Correlation function of random process

Dependence between values $X(t)$ and $X(t + \tau)$ (τ – any shift in time) is statistically estimated by **correlation function** (CF) of process $X(t)$. CF is calculated as average value of product

$$K_X(\tau) = \overline{X(t)X(t+\tau)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_2(x_1, x_2, \tau) dx_1 dx_2 \quad (3.12)$$

or

$$K_X(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, x(t+\tau)\, dt. \quad (3.13)$$

CF properties of stationary process:

1. If in the expression (3.13) put $\tau = 0$ it passes to expression (3.8), therefore

$$K_X(0) = P_X. \quad (3.14)$$

2. As correlation function of stationary process does not depend on time t, average value of product is $\overline{X(t)X(t+\tau)} = \overline{X(t-\tau)X(t)}$, therefore

$$K_X(-\tau) = K_X(\tau) \quad (3.15)$$

correlation function of random process is even.

3. Let consider an average square of a difference of process values which will be distant in time τ

$$\begin{aligned} \overline{\varepsilon^2(\tau)} &= \overline{[X(t) - X(t+\tau)]^2} = \overline{X^2(t)} - 2\overline{X(t)X(t+\tau)} + \overline{X^2(t+\tau)} = \\ &= 2K_X(0) - 2K_X(\tau). \end{aligned} \quad (3.16)$$

Average square is always non-negative. Therefore

$$K_X(0) \geq K_X(\tau) \quad (3.17)$$

value of correlation function of any random process at argument $\tau = 0$ maximal.

4. Let answer a question: what are the difference in process values, which are distant on τ? The answer is in the ratio (3.21):

$$\overline{\varepsilon^2(\tau)} = 2[K_X(0) - K_X(\tau)] \quad (3.18)$$

the more difference between $K_X(\tau)$ and $K_X(0)$, the more average difference between values of process, which are distant on τ. Thus, correlation function of random process $K_X(\tau)$ characterizes a degree of statistical dependence between values of process, which are distant on τ.

It is obvious, that with growth $|\tau|$, statistical dependence between values $X(t)$ and $X(t+\tau)$ decreases and at rather big $|\tau|$ dependence disappears. At $|\tau| \to \infty$ function $K_X(\tau)$ tends to zero, decreasing monotonously or oscillating around of zero, as shown on fig. 20,*a*.

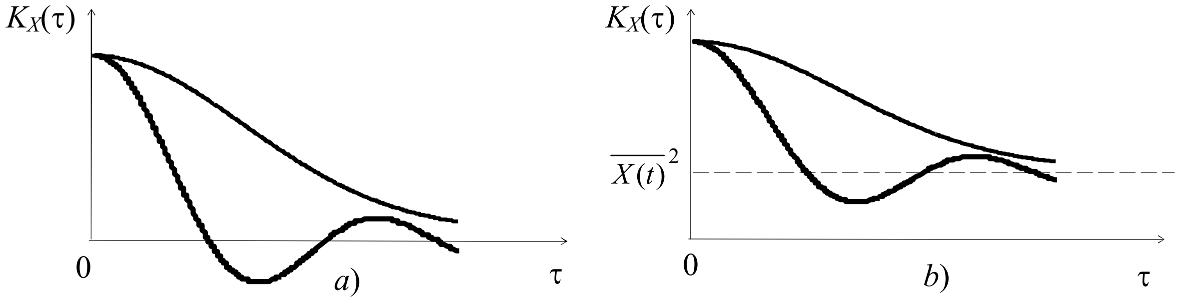


Figure 20 – Correlation functions of random processes:
a – at $\overline{X(t)} = 0$; *b* – at $\overline{X(t)} \neq 0$

5. Definition of statistical dependence is convenient for carrying out the **normalized correlation function**

$$R_X(\tau) = K_X(\tau)/K_X(0). \tag{3.19}$$

From ratio (3.19) it follows, that $-1 \le R_X(\tau) \le 1$. The closer value $R_X(\tau)$ to 1, the more strong correlated values of process, which are distant on τ.

For the rough description of correlation dependence it is introduced the concept of **correlation interval (time)** of process τ_c: values of process, which are distant on $\tau \le \tau_c$ are **essentially correlated** among themselves, and values of process, which are distant on $\tau > \tau_c$ are **uncorrelated**. Correlation interval is defined differently. One of the ways is the way, as duration of a pulse is estimated. So, it is possible to agree, that

$$\tau_к = \int_0^\infty |R_X(\tau)| d\tau. \tag{3.20}$$

Here τ_c is equaled to the basis of a rectangular with height $R_X(0) = 1$, having the same area, as the area under a curve $|R_X(\tau)|$ at $\tau > 0$ and an axis absciss. It is possible to define time of correlation τ_c as duration of function $|R_X(\tau)|$ at $\tau > 0$ at a level, for example, 0,1.

Function of mutual correlation for the characteristic of dependence between values of two random processes $X(t)$ and $Y(t)$, which are distant on τ, is defined in the similar way as correlation function of process

$$K_{XY}(\tau) = \overline{X(t)Y(t+\tau)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyp_2(x, y, \tau)dxdy, \tag{3.21}$$

where $p_2(x, y, \tau)$ – joint probability density of values of stationary processes $X(t)$ and $Y(t)$, which are distant on τ, or

$$K_{XY}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t+\tau)dt. \tag{3.22}$$

3.5 Power spectral density function of stationary random process

Let find Fourier transformation from realization of process $x_k(t)$, i.e. its spectral density

$$\lim_{T\to\infty} S_k(j\omega) = \lim_{T\to\infty} \int_{-T/2}^{T/2} x_k(t)e^{-j\omega t}dt. \tag{3.23}$$

But it will be the spectral characteristic only of realization $x_k(t)$, instead of process as a whole. It is possible to show, that

$$\lim_{T\to\infty} \frac{1}{T}\overline{|S_k(j\omega)|^2} = \int_{-\infty}^{\infty} K_X(\tau)e^{-j\omega\tau}d\tau, \tag{3.24}$$

where the direct line means averaging on ensemble of realizations. As correlation function characterizes process as a whole, the left part in (3.24) is also the spectral characteristic of all process. It is designated as

$$G_X(\omega) = \lim_{T\to\infty} \frac{1}{T}\overline{|S_k(j\omega)|^2}. \tag{3.25}$$

Function $|S_k(j\omega)|^2$ characterizes distribution of energy of process on frequency. As a result of division of this function on T we shall receive **distribution of power of process on frequency**. Function $G_X(\omega)$ is called spectral dencity of process power.

The expression (3.24) can be rewritten as direct and inverse Fourier transformations

$$\left.\begin{aligned} G_X(\omega) &= \int_{-\infty}^{\infty} K_X(\tau)e^{-j\omega\tau}d\tau, \\ K_X(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} G_X(\omega)e^{j\omega\tau}d\omega. \end{aligned}\right\} \tag{3.26}$$

On the basis (3.26) it is possible to write down

$$K_X(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G_X(\omega)d\omega = \int_{-\infty}^{\infty} G_X(f)df. \tag{3.27}$$

But $K_X(0) = P_X$. It follows from (3.27), that function $G_X(\omega)$ really characterizes distribution of power of process on frequency on an interval $(-\infty, \infty)$, and value of function $G_X(\omega)$ or $G_X(f)$ is equal power of process in bands in 1 Hz near of frequencies $+f$ and $-f$. Therefore function $G_X(\omega)$ refers to as power spectral density function of process. Thus, the **power spectral density function and correlation function of stationary random process are connected by Fourier transformations**. This statement is known as Khinchin-Wiener's theorem. Dimension of function $G_X(\omega)$ is V^2/Hz or Watt/Hz, coincides with dimension of energy and, probably, therefore sometimes function $G_X(\omega)$ is called energy spectrum of process.

As functions $K_X(\tau)$ and $G_X(\omega)$ are even instead of pair transformations (3.26) it is possible to write down a pair of Fourier cosine-transformations which are, as a rule, more simple in calculations, than ratio (3.26)

$$\left.\begin{aligned} G_X(\omega) &= 2\int_0^{\infty} K_X(\tau)\cos\omega\tau d\tau, \\ K_X(\tau) &= \frac{1}{\pi}\int_0^{\infty} G_X(\omega)\cos\omega\tau d\omega. \end{aligned}\right\} \tag{3.28}$$

Knowing function $G_X(\omega)$, it is possible to define **width of a spectrum of process** from some condition, for example, length of area of positive frequencies, outside of which value of function is not exceeded with values $0{,}1\max\{G_X(\omega)\}$. If the spectrum adjoins to zero bandwidth of a spectrum is defined as $F_{\max}$ (fig. 21, *a)*, and if a spectrum is bandpass, bandwidth of a spectrum is defined as ΔF (fig. 21, *b)*.

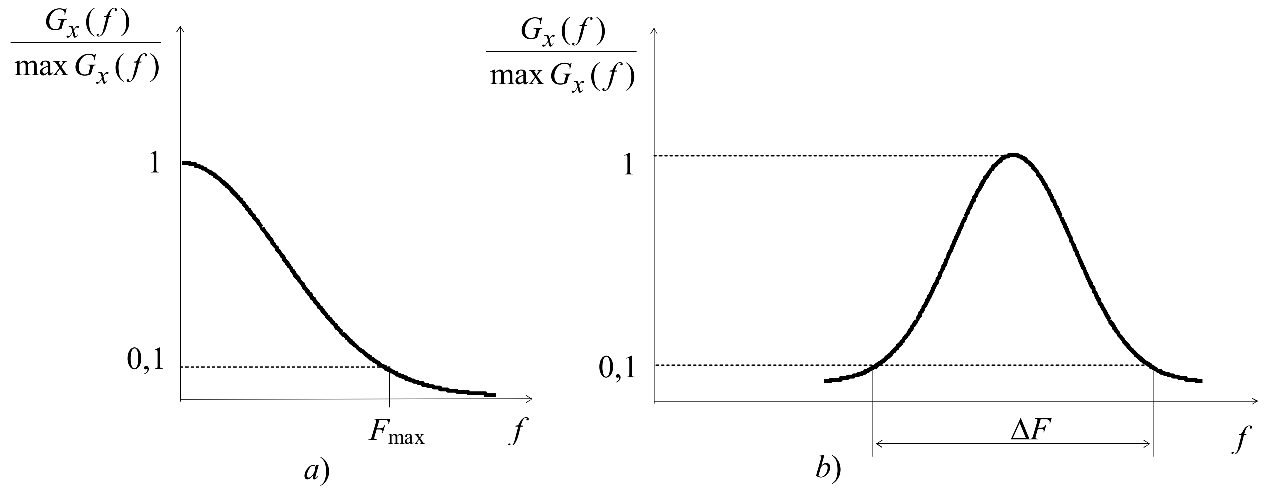


Figure 21 – Definition of bandwidth of process: *a* – the spectrum adjoins to zero frequency; *b* – bandpass spectrum

As functions $G_X(\omega)$ and $K_X(\tau)$ are connected by Fourier transformation according to property of change of scale, than the less correlation interval, the more wide a spectrum of process and on the contrary. In other words, the **correlation interval and bandwidth of process are inversely proportional values**.

3.6 Gaussian random process

Most frequently in the theory and techniques of communication meets so-called gaussian (or normal) random process. **Random stationary process *X*(*t*) re-**

fers to gaussian process if its one-dimensional and two-dimensional probability density functions are described by the following expressions

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2} \right), \tag{3.29}$$

$$p_2(x_1, x_2, \tau) = \frac{1}{2\pi\sigma^2\sqrt{1-R_X^2(\tau)}} \exp\left(-\frac{(x_1-a)^2 - 2R_X(\tau)(x_1-a)(x_2-a) + (x_2-a)^2}{2\sigma^2[1-R_X^2(\tau)]} \right), \tag{3.30}$$

where σ^2 – variance of process $X(t)$;

a – average value of process $X(t)$;

$R_X(\tau)$ – value of the normalized correlation function of process $X(t)$.

To define two-dimensional probability density of normal random stationary process, it is enough to know only its correlation function. Thus, normal stationary processes can differ one from another with kind of correlation function and power spectral density.

The one-dimensional probability distribution function of normal process is described

$$F(x) = 1 - Q\left(\frac{x-a}{\sigma} \right), \tag{3.31}$$

where
$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} \exp(-\frac{t^2}{2}) dt - \tag{3.32}$$

***Q*-function** or addition to probability distribution function. Graphs of functions (3.29) and (3.31) are shown on fig. 22.

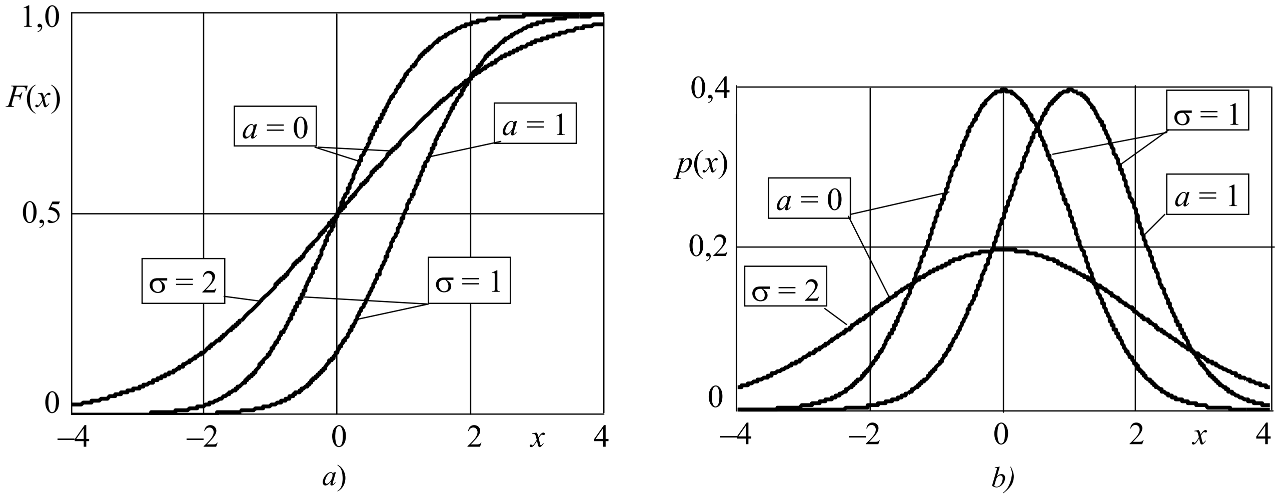


Figure 22 – Gaussian distribution:
a – probability distribution function; *b*– probability density function

Gaussian bandpass process is convenient to present through quadrature components

$$X(t) = A(t)\cos(\omega_0 t + \Phi(t)) = A_c(t)\cos\omega_0 t + A_s(t)\sin\omega_0 t = X_c(t) + X_s(t), \tag{3.33}$$

where $X_c(t)$ and $X_s(t)$ are quadrature components of process;

ω_0 – some frequency belonging to a band of process $X(t)$.

Quadrature components are uncorrelated processes having gaussian probability distribution. Their variances are identical and equal to half of variance of process $X(t)$.

Envelope $A(t)$ and phase $\Phi(t)$ are also uncorrelated processes. Envelope $A(t)$ has Reyleigh probability distribution (fig. 23,*a*)

In expressions (3.34) σ^2 is the variance of process $X(t)$. Average value of process $\overline{A(t)} = \sqrt{\frac{\pi}{2}}\sigma$, variance $D[A(t)] = (2 - \frac{\pi}{2})\sigma^2$, average power $P_A = 2\sigma^2$.

$$\left.\begin{aligned} p(a) &= \begin{cases} \dfrac{a}{\sigma^2}\exp\left(-\dfrac{a^2}{2\sigma^2}\right), & a > 0, \\ 0, & a \le 0; \end{cases} \\ F(a) &= \begin{cases} 1 - \exp\left(-\dfrac{a^2}{2\sigma^2}\right), & a > 0, \\ 0, & a \le 0. \end{cases} \end{aligned}\right\} \tag{3.34}$$

The phase $\Phi(t)$ has uniform probability distribution on interval $(0, 2\pi)$ (fig. 23,*b*)

$$p(\varphi) = \begin{cases} \dfrac{1}{2\pi}, & 0 \le \varphi < 2\pi, \\ 0, & \varphi < 0, \ \ \varphi \ge 2\pi. \end{cases} \tag{3.35}$$

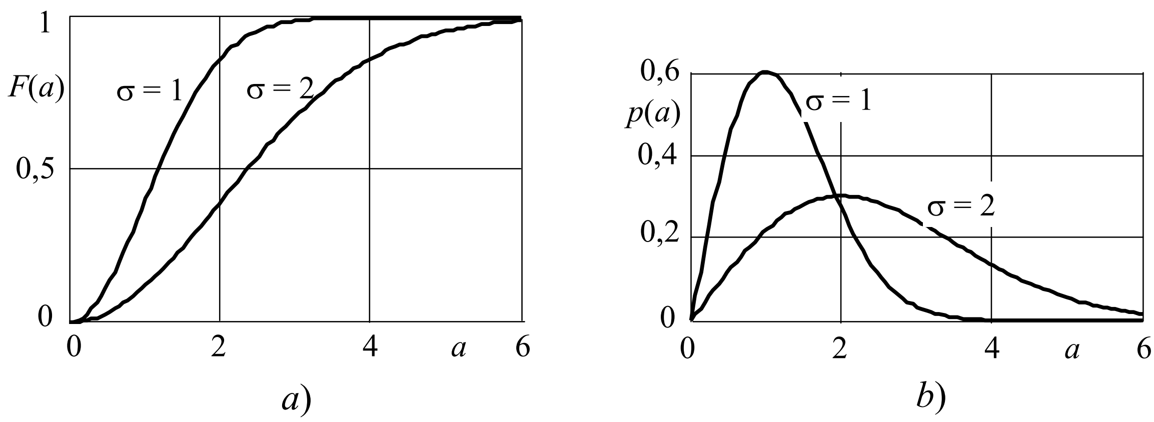


Figure 23 – Reyleigh distribution:
a – probability distribution function; *b*– probability density function

3.7 White noise

Random process refers to as **white noise**, if power spectral density function is a constant

$$G(\omega) = \frac{N_0}{2}, \qquad -\infty < \omega < \infty, \tag{3.36}$$

where N_0 is a power of process in a band equals 1 Hz.

Graphic dependences shown on fig. 24 correspond to expression (3.36).

The correlation function of a white noise is defined as inverse Fourier transform from (3.36)

$$K(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{N_0}{2}e^{j\omega\tau}d\omega=\frac{N_0}{2}\delta(\tau). \qquad (3.37)$$

On fig. 25 the graph of correlation function of white noise is represented.

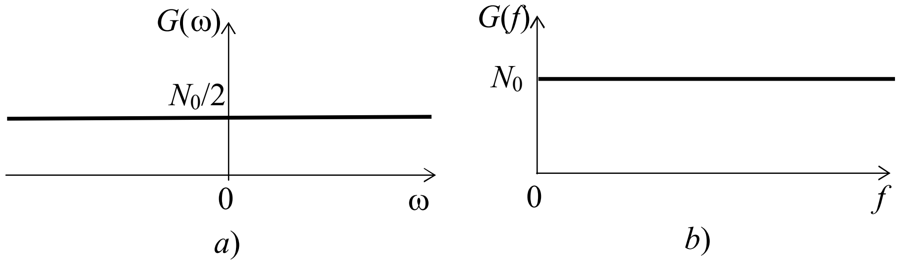


Figure 24 – Power spectral density function of white noise: *a* – two-sided spectrum; *b*– one-sided spectrum

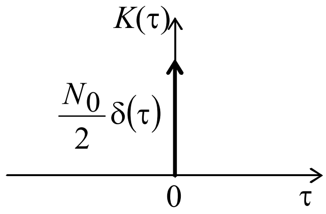
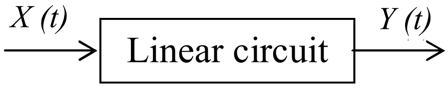


Figure 25 – Correlation function of white noise

3.8 Transformation of random processes by linear electric circuits

While studying of passage of random processes through linear circuits it is considered, that statistical characteristics of input random process $X(t)$ are known; transfer function of a linear circuit $H(j\omega)$ is known also. It is necessary to find characteristics of output process $Y(t)$.



Power spectral density function (PSDF) of process on output of a linear circuit is connected with PSDF of input process through square AR of a circuit

$$G_Y(\omega) = G_X(\omega)H^2(\omega). \qquad (3.38)$$

In particular, if input process is white noise, then PSDF of output process repeats square AR of a linear circuit.

Correlation function (CF) of process on output of a linear circuit is defined as Fourier transform from PSDF of process

$$K_X(\tau)=\frac{1}{\pi}\int_0^{\infty}G_X(\omega)\cdot\cos(\omega\tau)d\omega\ . \qquad (3.39)$$

Let $X(t)$ is white noise with one-sided PSDF $G_X(f) = N_0$, $0 \le f < \infty$, it acts on an input of ideal LPF with AR

$$H(f)=\begin{cases}H_0, & 0\le f<F_{\text{cut}},\\ 0, & f\ge F_{\text{cut}},\end{cases} \qquad (3.40)$$

where F_{cut} – cut off frequency of LPF. Then PSDF of process $Y(t)$:

$$G_Y(f) = G_X(f)\cdot H^2(f) = \begin{cases} N_0 H_0^2, & 0 \le f < F_{\text{cut}}, \\ 0, & f \ge F_{\text{cut}}. \end{cases} \tag{3.41}$$

PSDF of process $Y(t)$ is shown on fig. 26,*a*.
Average power of process $Y(t)$ is:

$$P_Y = \int_0^{\infty} G_Y(f)\,df = \int_0^{F_{\text{cut}}} N_0 H_0^2\,df = N_0 H_0^2\, F_{\text{cut}}. \tag{3.42}$$

Correlation function of process $Y(t)$ is:

$$K_Y(\tau) = N_0\, H_0^2\, F_{\text{cut}} \frac{\sin 2\pi F_{\text{cut}}}{2\pi F_{\text{cut}}}. \tag{3.43}$$

On fig. 26,*b* the normalized correlation function $R_Y(\tau) = K_Y(\tau)/K_Y(0)$ is shown. Correlation interval of process $Y(t)$ $\tau_c = 1/(2F_{\text{cut}})$.

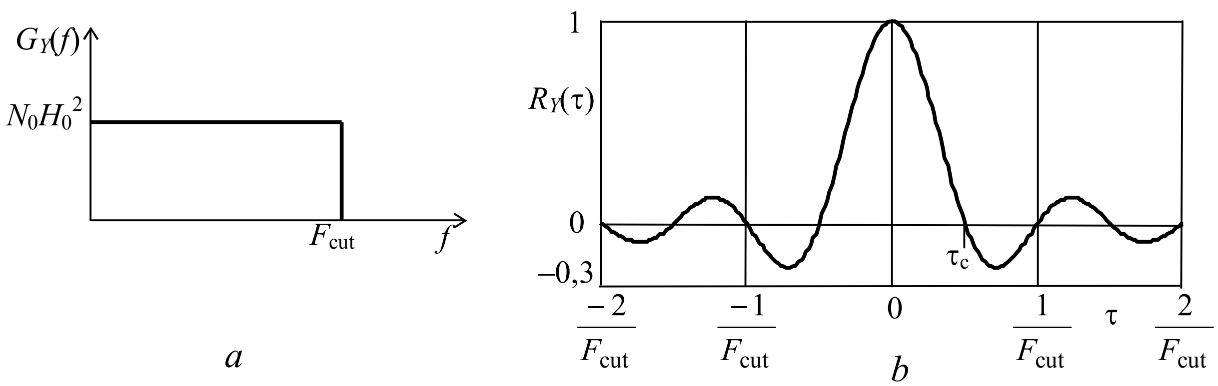


Figure 26 – Characteristics of process $Y(t)$ at a filtering with ideal LPF:
a – PSDF; *b* – CF

The concept **noise band of a linear circuit** is entered. Noise band of a circuit is equal to integral from a square normalized AR of circuit

$$F_n = \int_0^{\infty} \frac{H^2(f)}{H_{\max}^2}\,df, \tag{3.44}$$

where $H_{\max}$ is the maximal value of AR.

An ideal LPF has noise band $F_n = F_{\text{cut}}$. Noise band of a circuit allows easy to define power of process on an output of a circuit if on an input of a circuit white noise with one-sided PSDF N_0 acts:

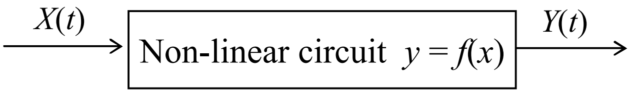
$$P_Y = N_0 \cdot F_n \cdot H_{\max}^2 \tag{3.45}$$

Consider probability distribution of process on output of a linear circuit. If on an input of a linear circuit Gaussian process acts, then output process will be also Gaussian – a type of distribution is not changed, only its parameters are changed. If on an input of a circuit process is not gaussian, then the distribution kind is changed, and output process has probability distribution closer to gaussian, then distribution of input process.

The filtering is narrow-band if bandwidth of circuit is much less than width of a spectrum of input process. At a narrow-band filtering the phenomenon of **normalization of process** takes place, which consists in the following – irrespective of a kind of distribution of input process, probability distribution of process on an output of a circuit is Gaussian.

3.9 Transformation of random processes by non-linear electric circuits

While researching of random processes passing through non-linear inertial circuit it is considered, that statistical characteristics of input process $X(t)$ and dependence $y = f(x)$ between instant values of input and output processes are known. It is necessary to find characteristics of output process $Y(t)$.



The most widespread function $f(x)$ for the description of non-linear transformations is the polynomial of degree n

$$f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n, \tag{3.46}$$

where a_0, a_1, a_2,..., a_n are coefficients of polynomial.

Factors and degree of a polynomial are defined as a result of approximation of the characteristic of a real electric circuit or proceeding from some assumptions. There are other dependences also used, except polynomial dependence (3.46).

Each of composed functions (3.46) brings the contribution to formation of values of reaction of a non-linear circuit on input action. So, a_0 describes occurrence of a constant component at $x = 0$; a_1x is linear composed element which provides proportional mapping of values x in y; a_2x^2 is square-law composed element, a_3x^3 is cubic composed, that are provided by the contributions proportional to x^2, x^3, etc.

The elementary action is a harmonious fluctuation $x(t) = A_1\cos 2\pi f_1 t$. In this case

$$y(t) = a_0 + a_1 A_1 \cos 2\pi f_1 t + a_2 A_1^2 \cos^2 2\pi f_1 t + ... + a_n A_1^n \cos^n 2\pi f_1 t. \tag{3.47}$$

If to take advantage of formulas of multiple arguments we shall receive

$$y(t) = Y_0 + Y_1 \cos 2\pi f_1 t + Y_2 \cos 2\pi 2 f_1 t + ... + Y_n \cos 2\pi n f_1 t, \tag{3.48}$$

where Y_0 is constant component of response;

Y_1, Y_2, ..., Y_n are amplitudes of the first, the second..., n-th harmonics of action.

Thus, response to harmonious action contains a constant component and harmonics of frequency of action – it essentially distinguishes non-linear circuits from linear in which new components do not arise.

In the case of biharmonic action

$$x(t) = A_1 \cos 2\pi f_1 t + A_2 \cos 2\pi f_2 t. \tag{3.49}$$

The approach to definition of response is the same, as well as used above, expression for $x(t)$ is substituted in a polynomial (3.46). While raising the sum (3.49) to such power as square, a cube, etc., degrees of cosine frequencies f_1 and f_2 will appear, that after transformations it gives expression of such kind (3.48) for fluctuations of frequencies f_1 and f_2. But products of cosines and their degrees are also appeared. Product of cosines gives components of summing and differential frequencies.

Components of **combinational frequencies** generally will take place

$$f_{\text{comb}} = |pf_1 \pm qf_2|, \tag{3.50}$$

where p, q are integers 0, 1, 2, ..., but such, that $p + q \le n$. Their sum $N = p + q$ is called the order of combinational frequency.

So, if $n = 3$, that in a spectrum of response there can be components of frequencies $f_1, f_2, 2f_1, 2f_2, |f_1 \pm f_2|, 3f_1, 3f_2, |2f_1 \pm f_2|, |f_1 \pm 2f_2|$. Amplitudes of components depend on amplitudes A_1 and A_2 and coefficients of a polynomial (3.46).

While passing of random process through a non-linear circuit the type of distribution of momentary values are significantly changed.

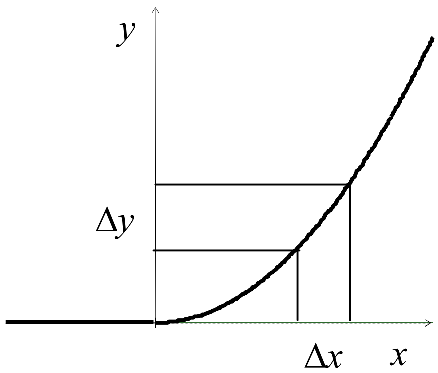


Figure 27 – The characteristic of nonlinearity

On fig. 27 non-linear dependence is shown as $y = f(x)$. All values of process $X(t)$, got on the interval Δx, are mapped in values of process $Y(t)$, got on the interval Δy. Therefore equality $p(x)\Delta x \approx p(y)\Delta y$ is correct. As for the infinitesimal increments dx and dy, we shall receive, that

$$p(y) = \frac{p(x)}{|dy/dx|} \tag{3.51}$$

It also is the general rule of calculation of probability density of output process.

To define power spectral density function of output process $G_Y(f)$ the next way is possible: to define correlation function of output process $K_Y(\tau)$, and then to perform with it Fourier transform. It follows from definition of correlation function

$$K_Y(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1) f(x_2) p_2(x_1, x_2, \tau) dx_1 dx_2, \tag{3.52}$$

where $f(x)$ is the function describing a non-linear circuit;

$p_2(x_1, x_2, \tau)$ is two-dimensional probability density of input process.

Methods of definition of characteristics of output process are stated. Certainly, in particular cases there can be mathematical difficulties.

Questions for section 3

1. Define random function and random process.
2. Define realization and ensemble of realizations of a random process.
3. What is the meaning of the statistical approach to the description of random processes?
4. Define probabilistic characteristics of random processes. What properties do they have?

5. What process is called stationary? Write down expressions that determine the numerical characteristics of such process.

6. Define correlation function of a random process. What properties does it have?

7. How is it roughly described the correlation between the values of random process in different points of time?

8. What process is called ergodic? Write down expressions that determine the numerical characteristics and CF of such process.

9. What is power spectral density of random process?

10. Formulate Khinchin-Wiener theorem, write down the corresponding mathematical expression.

11. Define Gaussian (normal) random process.

12. What is the probability distribution of envelope and phase of normal band-pass process?

13. Define white noise. What is the correlation function of white noise?

14. Define the linear electric circuit. How these circuits can be described?

15. How to determine PSDF, CF, and average power of the process at the output of the circuits?

16. Define the noise band of the linear circuit. What purpose is such feature introduced for?

17. How do linear transformations affect on the distribution of instantaneous values of a random process?

18. What is called the normalization of a random process, What are the conditions it occurs under?

19. Define nonlinear uninertial circuit. How can this circuits be described?

20. What fundamentally distinguishes nonlinear from linear circuit?

21. How do nonlinear transformations affect on the distribution of instantaneous values of a random process?

4 MODULATED SIGNALS

4.1 Classification of modulation types

In most transmission systems, telecommunication baseband signals cannot be passed directly by communication channels without transformation into other signals. Transformations have as an object to co-ordinate signals characteristics with communication channels characteristics. One of such transformations is **modulation**.

It is distinguished that if a modulating signal is continuous it belongs to **analog modulation**, and if modulating signal is digital it belongs to **digital modulation**.

4.2 General information on analog modulation

At analog modulation one of carrier parameters $u_{car}(t)$ gets increases, that are proportional to the values of modulating signal $b(t)$.

Carrier is auxiliary harmonic oscillation, necessary for implementation of modulation process

$$u_{car}(t) = A_0\cos(2\pi f_0 t + \varphi_0). \tag{4.1}$$

At such oscillation amplitude, frequency or initial phase can get increases.

Name of parameter which gets increases determines the name of type of modulation (amplitude, phase, and frequency).

While considering analog types of modulation we will consider that a modulating signal is a telecommunication baseband signal $b(t)$ with such characteristics:

– maximal frequency of signal spectrum F_{max} is given;

– a signal is normalized so, that maximal on the module value $|b(t)|_{max} = 1$;

– average value of signal $\overline{b(t)} = 0$;

– the coefficient of amplitude K_A, is given. It determines, in how many times the maximal on the module value of signal is exceeded its average quadratic value (root out of average power P_b):

$$K_{\mathrm{A}} = \frac{|b(t)|_{\max}}{\sqrt{P_b}}. \tag{4.2}$$

If a signal is normalized by the method indicated above

$$P_b = 1/K_{\mathrm{A}}^2. \tag{4.3}$$

4.3 The most common types of analog modulation

At amplitude modulation the amplitude of harmonious carrier are proportional to the instantaneous values of modulating signal, i.e. amplitude of the modulated signal is

$$A(t) = A_0 + \Delta A b(t), \tag{4.4}$$

where ΔA is a coefficient of proportionality, which is chosen so that amplitude $A(t)$ does not take on negative values. Frequency and initial phase of carrier remain constant.

In the case of arbitrary modulating signal, analytical expression of AM signal looks like

$$s_{\text{AM}}(t) = A_0[1 + m_{\text{AM}}b(t)]\cos(2\pi f_0 t + \varphi_0), \tag{4.5}$$

where $m_{\text{AM}} = \Delta A / A_0$ – amplitude modulation factor; $0 < m_{\text{AM}} \le 1$.

The time base diagram of AM signal is shown on fig. 28. Envelope of the modulated signal repeats the form of modulating signal.

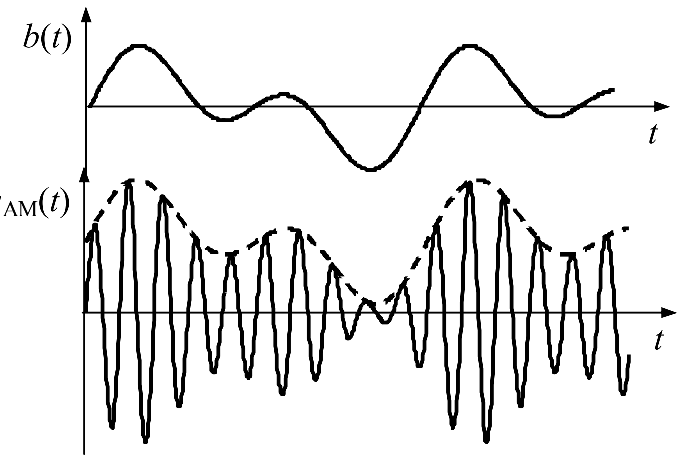


Figure 28 – Modulating $b(t)$ and modulated $s_{\text{AM}}(t)$ signals

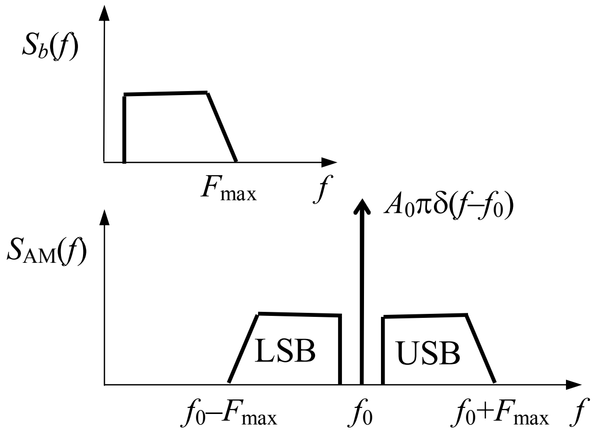


Figure 29 – Spectrums modulating and AM signals

We will pass to consideration of spectrum of AM signal in a case of complex modulating signal which complies with the real signals of telecommunication. The complex signal *b(t)* has finite or infinite sum of harmonious components. Every component causes appearance in the spectrum of the modulated signal two components – summing and differential frequencies. Their total sums create accordingly upper and lower sidebands of frequencies.

On fig. 29 arbitrary amplitude spectrum of modulating signal and its proper amplitude spectrum of AM signal are shown. It consists of harmonious oscillation of carrier frequency, upper sideband of frequencies (USB) and lower sideband of frequencies (LSB). Thus USB is the scale copy of spectrum of modulating signal, which is shifted on frequency on the value f_0. LSB is the mirror reflection of USB according to carrier frequency f_0. Fig. 29 gives an important result: the width of spectrum of AM signal ΔF_{AM} equals the doubled value of maximal frequency of spectrum of modulating signal $F_{\max}$, i.e.

$$\Delta F_{\text{AM}} = 2F_{\max}. \tag{4.6}$$

It is possible to show that average power of carrier $P_{\text{car}} = \dfrac{A_0^2}{2}$, and sideband $P_{\text{SB}} = \dfrac{A_0^2 m_{\text{AM}}^2}{2K_{\text{A}}^2}$. Then ratio of sideband power to total power of AM signal is $\dfrac{P_{\text{SB}}}{P_{\text{AM}}} = \dfrac{m_{\text{AM}}^2}{K_A^2 + m_{\text{AM}}^2}$. If to take on a maximally possible value $m_{\text{AM}} = 1$, and value of amplitude modulating signal coefficient $K_A = 5$ (voice signal), so part of sidebands power is $P_{SB}/P_{AM} = 0{,}04$ or 4%.

We see that while using of AM for transmission of telecommunication signals, prevailing part of power of AM signal is spent on oscillation of carrier frequency, although this oscillation does not carry information, as its level in the process of modulation remains constant – information is contained in the sidebands of frequencies. Therefore it is efficient to form a signal with a spectrum, consisting only of two sidebands of frequencies (without oscillation of carrier frequency), – such signal is called the signal of **double-sideband with suppressed carrier modulation**.

Analytical expression of signal double-sideband with suppressed carrier modulation (DSB-SC) looks like

$$s_{\text{DSB-SC}}(t) = A_0 b(t)\cos(2\pi f_0 t). \tag{4.7}$$

The time base diagrams of modulating and modulated signals are shown on fig. 30. From fig. 30 it is evident, that envelope of DSB-SC signal $A(t) = A_0|b(t)|$ (it is shown by dotted line) does not repeat modulating signal.

From comparison of mathematical expressions, describing AM and DSB-SC signals it is clear that the spectrum of DSB-SC signal differs from the spectrum of AM signal by absence of oscillation of carrier frequency.

Random amplitude spectrum of modulating signal and its proper amplitude spectrum of DSB-SC signal are shown on fig. 31. It consists of USB and LSB.

From fig. 31 it flows, that the width of spectrum of DSB-SC signal $\Delta F_{\text{DSB-SC}}$ equals the doubled value of maximal frequency of modulating signal spectrum F_{max}, i.e.

$$\Delta F_{\text{DSB-SC}} = 2F_{max}. \tag{4.8}$$

The width of spectrum of DSB-SC signal is the same as width of spectrum of AM signal.

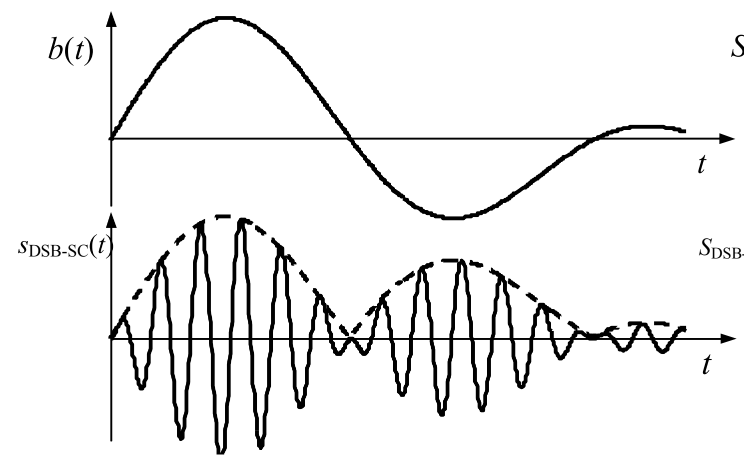


Figure 30 – Modulating $b(t)$ and modulated $s_{\text{DSB-SC}}(t)$ signals

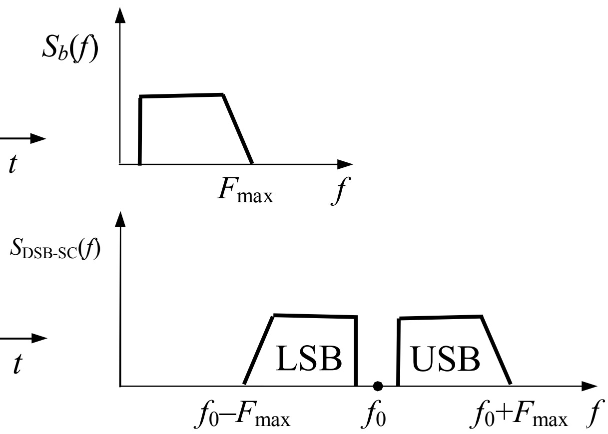


Figure 31 – Spectra of modulating and DSB-SC signals

Important advantage of DSB-SC signals in comparison with the AM signals is the advanced efficiency of the use of transmitter power, as considerable part of signal power is not spent on carrier oscillation, which is in the spectra of AM signals.

Without the losses of information about the signal $b(t)$ it is possible to take out one sideband (upper or lower) from the spectrum of DSB-SC signal. Thus we will get **single sideband modulation** (SSB).

In general case (for the arbitrary signal $b(t)$) SSB signal is written down as

$$s_{\text{SSB}}(t) = A_0 b(t)\cos(2\pi f_0 t + \varphi_0) \mp A_0 \tilde{b}(t)\sin(2\pi f_0 t + \varphi_0), \tag{4.9}$$

where sign "–" refers to description of signal with the upper sideband of frequencies, and sign "+" – with a lower sideband; $\tilde{b}(t)$ is a signal, conjugated on Gilbert with a signal $b(t)$.

On fig. 32 the time base diagrams of random modulating signal $b(t)$, the conjugated on Gilbert signal $\tilde{b}(t)$ and SSB signal calculated for it are shown. From fig. 32 it is evident, that SSB signal envelope $A(t) = A_0\sqrt{b^2(t)+\tilde{b}^2(t)}$ (shown by dotted line) does not repeat a modulating signal.

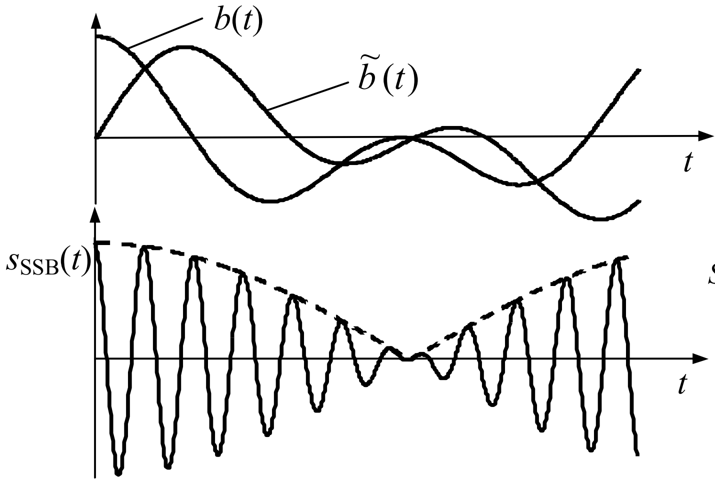


Figure 32 – Modulating $b(t)$ and modulated $s_{\text{SSB}}(t)$ signals

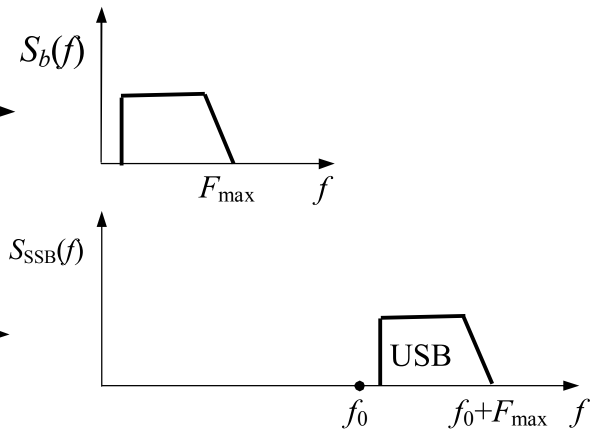


Figure 33 – Spectra of modulating and SSB signals

On fig. 33 amplitude spectrum of SSB signal, got from the spectrum of DSB-SC signal by the exception of lower sideband of frequencies (it is possible to eliminate the upper sideband of frequencies) is shown. So, such type of modulation, when the spectrum of the modulated signal coincides with the spectrum of modulating signal, shifted on carrier frequency, or is the inversion of the shifted spectrum respectively carrier frequency is called single sideband.

From fig. 33 it flows, that spectrum width of SSB signal ΔF_{SSB} equals maximal frequency of spectrum of modulating signal

$$\Delta F_{\text{SSB}} = F_{\max}. \tag{4.10}$$

Important advantage of SSB signal in comparison with DSB-SC and AM signals is twice decreased width of modulated signal spectrum, which allows the signals amount to increase twice in the set frequencies band. Therefore SSB is widely used in the systems of multichannel transmission with a frequency division. SSB is a single type of analog modulation, when the band of frequencies of signal is not broaden while modulation. Except of considered "clean" SSB in communication networks it was found the use of SSB signals with carrier (with pilot signal) and with partial suppression of one sideband of frequencies. It creates certain comforts at forming and detection of the modulated signals.

At **phase modulation** the increase of phase is proportional to the instantaneous values of modulating signal

$$\Delta\varphi(t) = \Delta\varphi_{\mathrm{d}}\, b(t), \tag{4.11}$$

where $\Delta\varphi_d$ is a coefficient of proportion, which is called **phase deviation**. As maximal on the module value $|b(t)|_{\max} = 1$, deviation of phase at PM is the most maximal deviation of phase from linear dependence in time.

Mathematical description of PM signal

$$s_{\mathrm{PM}}(t) = A_0 \cos(2\pi f_0\, t + \Delta\varphi_{\mathrm{d}}\, b(t) + \varphi_0). \tag{4.12}$$

At phase modulation instantaneous frequency depends on a modulating signal by next way

$$\Delta f(t) = \frac{1}{2\pi}\cdot\frac{d(\Delta\varphi_{\mathrm{d}} b(t))}{dt} = \frac{\Delta\varphi_{\mathrm{d}}}{2\pi}\cdot\frac{d(b(t))}{dt}. \tag{4.13}$$

At **frequency modulation** the increase of frequency is proportional to the instantaneous value of modulating signal

$$\Delta f(t) = \Delta f_{\mathrm{d}}\, b(t), \tag{4.14}$$

where Δf_{d} is coefficient of proportionality, which is called **frequency deviation** and determines the maximal deviation of instantaneous frequency of the modulated signal from carrier frequency f_0.

The increase of phase is at frequency modulation

$$\Delta\phi(t) = 2\pi\int_{-\infty}^{t} \Delta f(t)dt = 2\pi\Delta f_d \int_{-\infty}^{t} b(t)dt. \tag{4.15}$$

Will get mathematical description of FM signal by the substitution of expression (4.15) in formula (4.1):

$$s_{\mathrm{FM}}(t) = A_0 \cos(2\pi f_0 t + 2\pi\Delta f_{\mathrm{d}} \int_{-\infty}^{t} b(t)dt + \varphi_0). \tag{4.16}$$

From the given above descriptions of signals it follows, that FM and PM have a lot in common. Both FM and PM have the increases of frequency and phase. The name of modulation type is determined by which of parameters gets increases, proportional to modulating signal.

At angular modulations connection between the spectra of modulating and modulated signals is considerably more difficult, than at AM and its varieties. Therefore an analysis of spectra of angular modulations here is not given. We will discuss a final result. The analysis shows that theoretically the bandwidth of amplitude spectrum is infinite. However basic part of power of signal is concentrated in some limited frequency interval around f_0, which is considered as the width of signal spectrum. The spectrum width of FM and PM signals is calculated on formulas:

$$\Delta F_{\mathrm{FM}} = 2\,(m_{\mathrm{FM}} + 1)\,F_{\max}, \tag{4.17}$$

$$\Delta F_{\mathrm{PM}} = 2\,(m_{\mathrm{PM}} + 1)\,F_{\max}, \tag{4.18}$$

where

$$m_{\mathrm{FM}} = \Delta f_{\mathrm{d}} / F_{\max} - \tag{4.19}$$

is an index of frequency modulation, which is determined by ratio of frequency deviation of FM signal to frequency of modulating signal;

$$m_{\mathrm{PM}} = \Delta\varphi_{\mathrm{d}} - \tag{4.20}$$

is an index of phase modulation, which equals phase deviation of PM signal.

Only frequency modulation has got wide distribution. It is distinguished narrowband (in the case of $m_{\mathrm{FM}} < 1$) and broadband (in the case of $m_{\mathrm{FM}} >> 1$) modulations. Narrowband FM signal spectrum width is comparable with the AM signal spectrum width. If FM is broadband, then signal spectrum width approximately equals doubled deviation of frequency.

The **modulator diagram** realize a forming algorithm which follows from mathematical description of the modulated signal – expressions (4.5), (4.7), (4.9), (4.12) and (4.16).

Device which output voltage is proportional to some parameter of input bandpass signal is called **detector**. On the base of this determination, it is necessary to use definitions: amplitude, frequency and phase detectors.

4.4 General information on digital modulation

Modulation refers to digital if a modulating signal is the digital signal.

Digital signal (DS) is a sequence of digital symbols, which belong to the certain alphabet. As a rule, symbols are binary and designated 1 and 0, them name bits, and they act through interval T_{b}. DS appears as a result of coding signs on discrete messages or samples of continuous messages. DS is possible to present, writing down sequence, for example, 10110..., and specifying value T_{b}. Basic characteristic of DS is rate of a **signal or bit rate** $R = 1/T_{\mathrm{b}}$, bit/s.

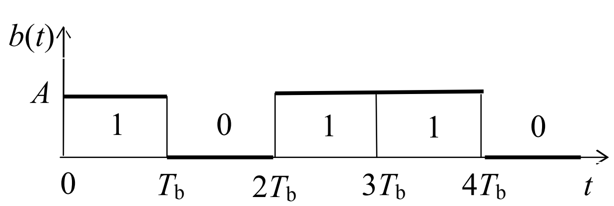


Figure 34 – Digital signal

In the terminal equipment of transmission systems, executed on logic microcircuits or processors, digital signals look like sequence of rectangular pulses. The example of such signal is resulted on fig. 34. Carrier of information is the rectangular pulse, and the information is displayed in amplitude of a pulse which accepts values A and 0.

For transfer of digital signals by communication channels methods of digital modulation, basically, with serial transfer are used.

The modulated signal for transfer by a communication channel is formed from signals $s_i(t)$ – elementary pulse signals belonging to ensemble $\{s_i(t)\}$, $i = 0, 1..., M - 1$, where M – quantity of elementary signals ($M \geq 2$). Elementary signals are called also **channel symbols**.

At formation of the modulated signal the sequence of binary symbols is broken into blocks from $n = \log_2 M$ bits. To each such block (the quantity of possible various blocks is $M = 2^n$) is put in conformity an elementary signal $s_i(t)$. Duration n bits makes **clock interval** $T = T_b \cdot \log_2 M$. The received sequence of the channel symbols acting through time T, forms the modulated signal

$$s(t) = \sum_{k=-\infty}^{k=\infty} s_i^{(k)}(t - kT), \tag{4.21}$$

where $s_i^{(k)}(t - kT)$ – i-th signal transmitting on a k-th clock interval.

On fig. 35 transition from DS to the modulated signal ($n = 4$, $M = 16$) is shown. On the plot $s(t)$ transmitting elementary signals $s_{13}(t)$, $s_1(t)$, $s_3(t)$, $s_6(t)$, ... are shown. Digital modulation is a display of blocks of bits in pulses-carriers.

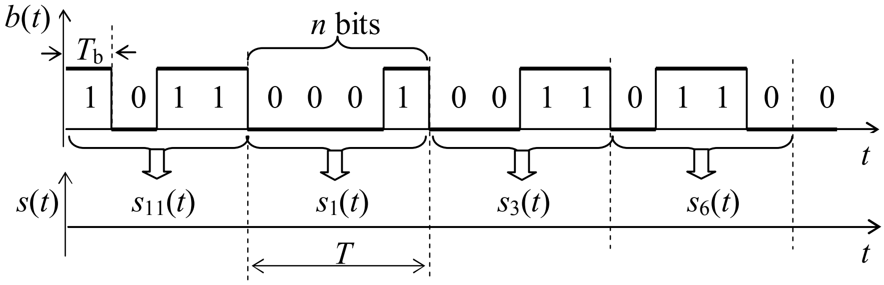


Figure 35 – Arrangement of transformation DS into the modulated signal

If $M = 2$, a signal $s(t)$ is binary; if $M > 2$, a signal $s(t)$ is multilevel or M-ary.

Serial transmission is considered. Also there is a **parallel-serial transmission** (OFDM). It will be considered later.

Key parameter of the modulated signal is the bandwidth. It depends on rate R and ensemble of channel symbols $\{s_i(t)\}$. The problem of a choice of digital modulation kind is reduced to a choice of ensemble of channel symbols.

4.5 Choice channel symbols forms

The information is displayed in amplitudes of pulses, instead of in their form. Therefore the form of a pulse-carrier is necessary for choosing under spectral and other characteristics.

Shown on fig. 34 DS does not approach for direct transmission on communication channels because of its spectral properties. On fig. 34 channel symbols are the rectangular pulses

$$A(t) = \begin{cases} 1, & 0 \leq |t| \leq T/2, \\ 0, & |t| > T/2; \end{cases} \tag{4.22}$$

the symbol 1 is represented by a pulse $A(t)$, and a symbol 0 – a pulse with zero amplitude.

Let's find spectral density of function $A(t)$:

$$S_A(jf) = \int_{-T/2}^{T/2} A(t) e^{-j2\pi ft} dt = T \frac{\sin \pi ft}{\pi ft}. \tag{4.23}$$

The amplitude spectrum of function $A(t)$ is

$$S_A(f) = T \left| \sin \pi ft / \pi ft \right|. \tag{4.24}$$

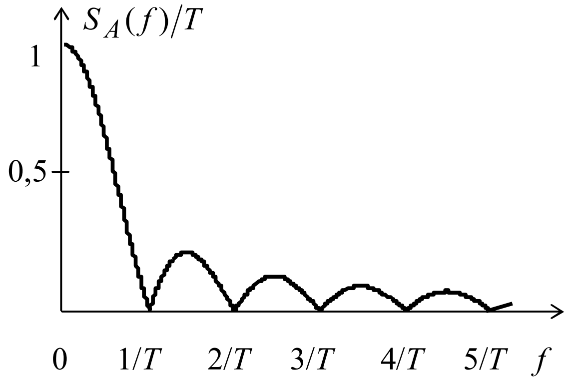


Figure 36 – Normalized spectrum of the rectangular pulse

On fig. 36 the plot of the normalized amplitude spectrum $S_A(f)/T$ is resulted. The spectrum of the rectangular pulse decreases extremely slowly – with a speed $1/f$. With the purpose of economy of a band of frequencies of a communication channel it is necessary to use pulses of the smoothed form.

The pulse $A(t)$ should satisfy to a condition

$$A(t) = \begin{cases} 1, & t = 0, \\ 0, & t = kT, \ k = \pm 1, 2, ... \end{cases} \tag{4.25}$$

The condition (4.25) is a sampling condition or a condition of absence of intersymbol interference (ISI). After sampling of a pulse $A(t)$, satisfying a condition (4.25), the discrete signal is formed $A(n)$ = ..., 0, 0, 1, 0, 0, ...

It is possible to show, that the spectrum of a pulse $A(t)$ should be skew symmetric

$$S_A(\omega_N - \omega) + S_A(\omega_N + \omega) = T, \quad 0 \le \omega \le 2\pi/T, \tag{4.26}$$

where $\omega_N = \pi/T$ – Nyquist frequency.

On fig. 37 examples of functions $S_A(\omega)$ with skew symmetry are resulted.

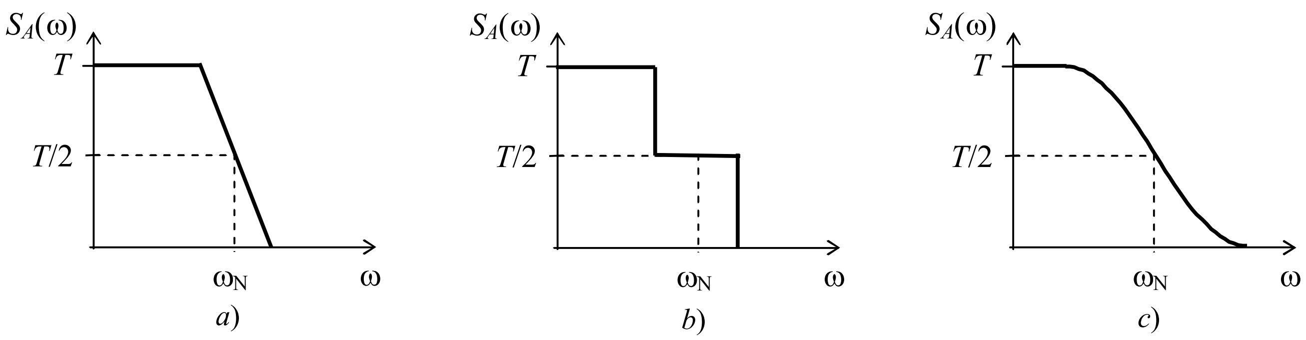


Figure 37 –Examples of functions $S_A(\omega)$ with skew symmetry

The spectrum of pulse signals satisfying a sampling condition (4.25), refer to as Nyquist spectrum. Them designate as $N(f)$. More often Nyquist spectrum describe by function

$$N(f)=\begin{cases} T, & 0\le|f|\le(1-\alpha)f_N, \\ 0{,}5T\left[1+\sin\left(\dfrac{\pi}{2\alpha}\left(1-\dfrac{|f|}{f_N}\right)\right)\right], & (1-\alpha)f_N<|f|<(1+\alpha)f_N, \\ 0, & |f|\ge(1+\alpha)f_N. \end{cases} \quad (4.27)$$

where $f_N = 1/(2T)$ is Nyquist frequency;

α – roll-off factor of a signal spectrum , $0 \le \alpha \le 1$.

Dependence (4.27) refers to „raised cosine“. On fig. 38 such dependences are resulted for $\alpha = 0$; 0,2; 0,5 and 1. From fig. 38 it is visible, that bandwidth of a pulse $F_{max} = (1 + \alpha)f_N$. Minimal possible bandwidth $\min F_{max} = f_N = 1/(2T)$.

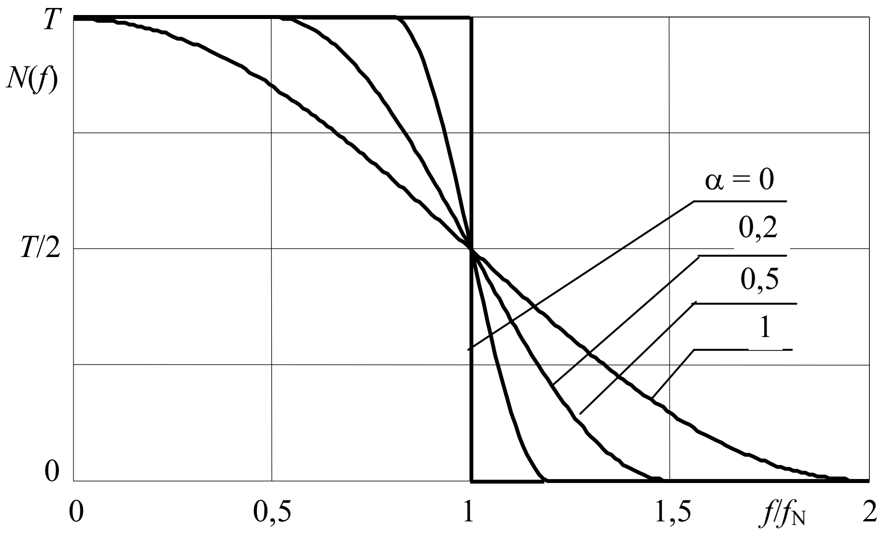


Figure 38 – Nyquist spectrum

Typical values of factor α lay in limits from 0,2 up to 0,4.

Function $A(t)$ can be received as inverse Fourier transform from $N(f)$

$$A(t) = \frac{\sin 2\pi f_N t}{2\pi f_N t}\cdot\frac{\cos 2\pi\alpha f_N t}{1-(4\alpha f_N t)^2}. \quad (4.28)$$

Pulses $A(t)$ name Nyquist pulses (fig. 39).

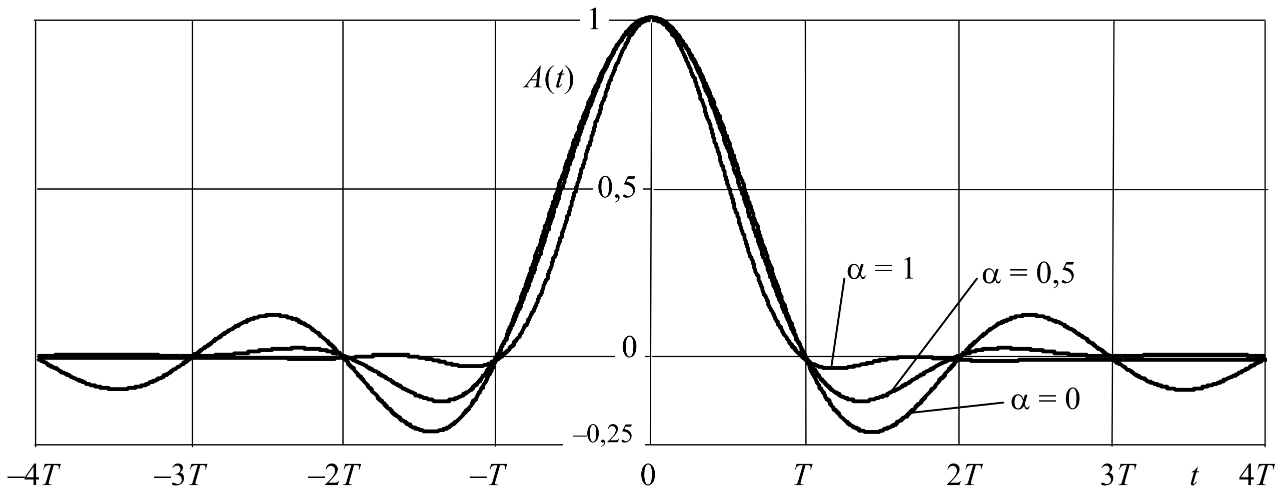


Figure 39 – Nyquist pulses

The type of pulse at the transmission of a baseband signal is considered. If the spectrum should be bandpass radio impulses $A(t)\cdot\cos(2\pi f_0 t)$ and $A(t)\cdot\sin(2\pi f_0 t)$ are used. Their amplitude spectrum have two side band which are copies of a spectrum of pulse $A(t)$ (fig. 40). If $A(t)$ has Nyquist spectrum the bandwidth of a radio pulse is defined $\Delta F = 2(1 + \alpha)f_N$.

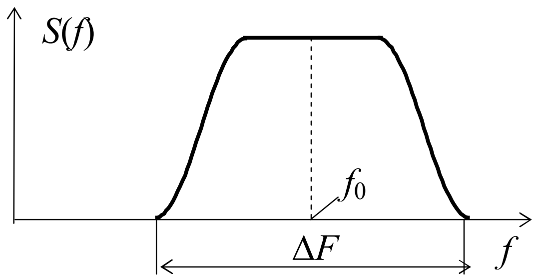


Figure 40 – Spectrum of radio pulses

4.6 Pulse-amplitude modulation

Modulation is named pulse-amplitude, if by channel symbols, used for forming of the modulated signal, are low-frequency impulses, i.e. their spectrum joins to a zero frequency. Channel symbols are described as

$$s_i(t) = a_i A(t), \quad i = 0, 1, ..., M - 1, \tag{4.29}$$

where $A(t)$ – impulse with certain time and spectral descriptions, the maximal value of which is equal 1; that intersymbol interference was not, impulse $A(t)$ must be the Nyquist impulse;

a_i is a coefficient, representing information.

The signals of pulse-amplitude modulation are designated as M-ary PAM, where M is a number of channel symbols.

Evident presentation of signals of digital modulation is signal constellation. On signal constellation each of channel symbols is represented a point, the co-ordinates of points are coefficients, which are, describe channel symbols. In the case of MPAM signals every channel symbol is described only one coefficient a_i, therefore for presentation of MPAM signals one-dimensional space is used. On a fig. 41 signal constellation of 2PAM or BPAM signal is shown. A mapping code, setting accordance between binary characters and coefficients, is specified also a_i.

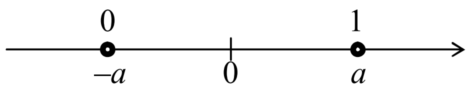


Figure 41 – BPAM signal constellation

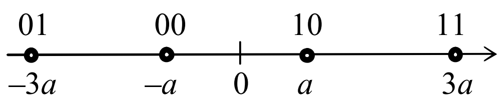


Figure 42 – QPAM signal constellation

On a fig. 42 signal constellation of 4PAM or QPAM signal is shown. A mapping code for the QPAM signal sets accordance between the pair of binary symbols (n = 2) and by the coefficients a_i. These pair also determines the signal number - binary symbols are the record of signal number in the binary notation scale. A mapping code must be a Gray code blocks from n bits, which correspond nearby signals, must

differ only in one bit. Gray code minimizes the amount of erroneous bits in the case of origin of error of decision about the passed channel symbol at demodulation.
Both at BPAM and at QPAM number a determines energies each of channel symbols and middle power of the modulated signal.

Like the considered examples it is possible to build signal constellations for M= 8, 16, ...

The chart of MPAM signal forming is resulted on a fig. 43. DS acts at the input. Mapper takes $n = \log_2 M$ is bits and gives out the coefficients of a_i rectangular pulses duration T. From these impulses a forming filter produces the impulses $a_i A(t)$. This procedure repeats oneself on every clock interval.

For different values M work of chart differs only a mapping code.

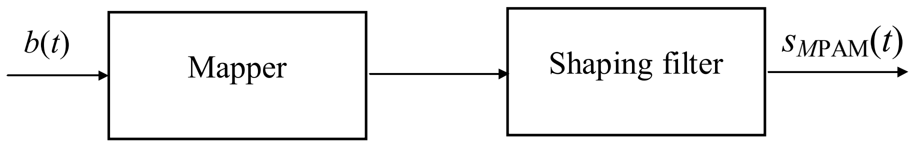


Figure 43 – Modulator of MPAM signal

MPAM signals bandwidth is determined by the width of impulses spectrum $A(t)$. As $T = T_\mathrm{b}\log_2 M$, then

$$\Delta F_{M\mathrm{PAM}} = \frac{1+\alpha}{2T_\mathrm{b}\log_2 M} = \frac{R(1+\alpha)}{2\log_2 M}. \tag{4.30}$$

4.7 One-dimensional bandpass signals of digital modulation

One-dimensional bandpass signals of digital modulation are MASK signals – M-ary amplitude shift keying ($M \geq 2$) and BPSK signals – binary phase shift keying.

At the MASK and BPSK signals channel symbols are radiopulses and they are written down:

$$s_i(t) = a_i\sqrt{2}A(t)\cos(2\pi f_0 t), \quad i = 0, 1, ..., M-1, \tag{4.31}$$

where a_i – number, representing $n = \log_2 M$ bits, transmitted the signal $s_i(t)$;
$A(t)$ – function, determining the form of radiopulses, its maximal value is equal 1;
f_0 – frequency of radiopulses.

As channel symbols differs only the coefficient a_i, signal constellations of these types of modulation appear in one-dimensional space, and the modulated signals are named one-dimensional. On a fig. 44 signal constellations of BPSK signals, BASK and QASK with pointing of mapping codes are resulted.

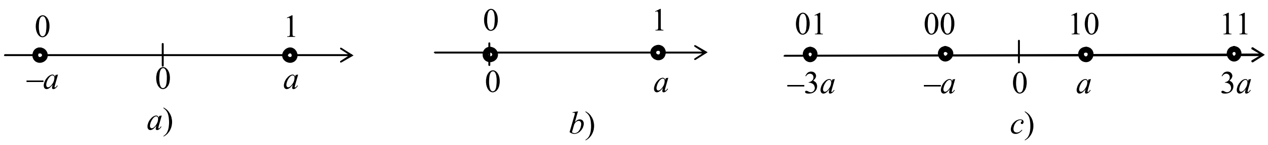


Figure 44 - Signal constellations of signals: a – BPSK; b – BASK; c – QASK

Energy of i-th channel symbol calculated

$$E_i = \int_{-\infty}^{\infty} s_i^2(t)dt = a_i^2. \quad (4.32)$$

Comfortably to compare the different methods of modulation on the size of minimum distance between the signals d, characterizing the difference of signals is quantitative. The size d is determined on signal constellation. The size d must be expressed through the physical parameters of signals. Comfortably to consider E_b such parameter is energy, expended on the transmission of one bit: $E_b = P_s T_b = P_s / R$. The last sizes are set on the system of transmission. Energy E_b is expressed through middle energy of signals E_{av} and amount bit, passed one signal n

$$E_b = E_{av}/n;\ E_{av} = \frac{1}{M} \cdot \sum_{i=0}^{M-1} E_i\ ; n = \log_2 M.$$

We will conduct calculations. For BASK:

$$d = a;\ E_0 = 0;\ E_1 = a^2;\ E_{av} = 0{,}5a^2;\ E_b = E_{av} = 0{,}5a^2;\ d = \sqrt{2E_b}.$$

For BPSK:

$$d = 2a;\ E_0 = E_1 = a^2;\ E_{av} = a^2;\ E_b = E_{av} = a^2;\ d = 2\sqrt{E_b}.$$

For QASK:

$$d = 2a;\ \ E_0 = E_2 = a^2;\ \ E_1 = E_3 = 9a^2;$$

$$E_{av} = 0{,}5(a^2 + 9a^2) = 5a^2;\ \ E_b = 0{,}5E_{av} = 2{,}5a^2;\ \ d = 2\sqrt{2/5 \cdot E_b} = 1{,}25\sqrt{E_b}.$$

Time diagrams of the examined channel symbols are resulted on a fig. 45. For obviousness it is accepted at a construction, that $A(t)$ – rectangular pulse of duration, equal to the clock interval. Like considered it is possible to build signal constellations and time diagrams for the 8ASK, 16ASK signals etc.

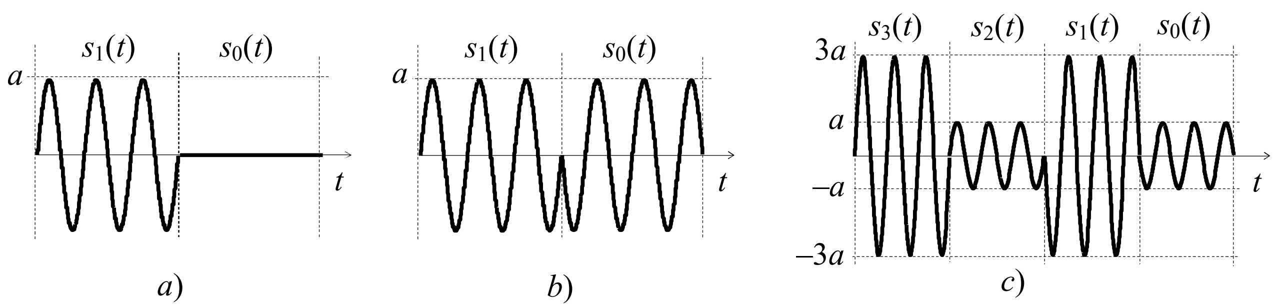


Figure 45 – Time diagrams of channel symbols: *a* – BASK; *b* – BPSK; *c* – QASK

We will consider the spectrum of carrier pulse $\sqrt{2}\,A(t)\cdot\cos(2\pi f_0 t)$, which channel symbols are built on the basis of (4.31). This carrier pulse is the signal of analogue DSB-SC, and that is why his spectrum consists of two sidebands, concentrated near frequency of radiopulse f_0, which can be considered frequency of carrier oscillation. Frequencies sidebands is the reflection of spectrum of impulse $A(t)$. So, spectral properties any of channel symbols $s_i(t)$ wholly determined the function $A(t)$.

Frequencies sidebands are the copies of Nyquist spectrum (fig. 46), and the spectrum width of *M*ASK and BPSK signals is determined:

$$\Delta F = 2f_{\mathrm{N}}(1+\alpha) = \frac{R(1+\alpha)}{\log_2 M}. \qquad (4.33)$$

An important conclusion follows from expression (4.33) – the increase of levels number of *M*ASK signal at the fixed speed *R* allows decreasing the spectrum width of channel symbols.

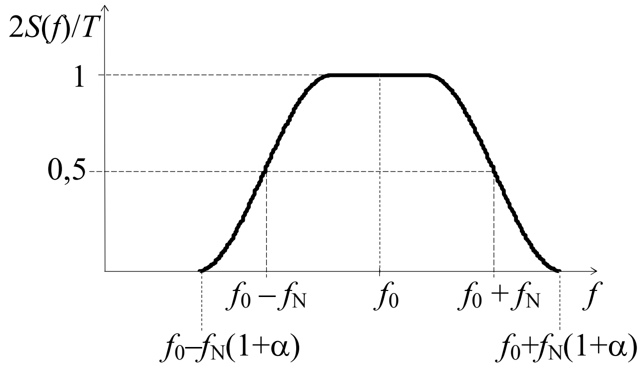


Figure 46 - Spectrums of *M*ASK and BPSK channel symbols

We will consider the chart of forming of MASK and BPSK signals. From comparison of expressions (4.29) and (4.31) follows, that at the MPAM signals carrier pulse $A(t)$, and at the bandpass signals carrier pulse $\sqrt{2}A(t)\cdot\cos(2\pi f_0 t)$. Thus, the chart of forming of one-dimensional bandpass signals (modulator) is built on the basis of chart of fig. 43 with addition the generator of carrying waveform *G* $\sqrt{2}\cos(2\pi f_0 t)$ and multiplier × (fig. 47).

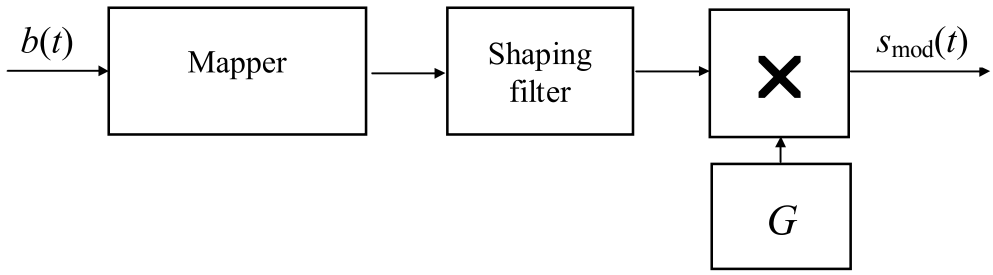


Figure 47 – Modulator of one-dimensional bandpass signals

So, on the basis of analysis of one-dimensional bandpass signals of digital modulation became obvious, that the name of modulation method specifies, what channel symbols parameter differs: *M*ASK – *M* signals differ in amplitudes, BPSK – two signals differ in initial phases (0 and 180°).

4.8 Two-dimensional bandpass digitally modulated signals

For two-dimensional bandpass signals of digital modulation the signals *M*-ary PSK ($M \geq 4$) and *M*-ary APSK (amplitude-phase shift keying) uses. At these types of modulation channel symbols are described the sum of cosine and sine radiopulses:

$$s_i(t) = a_i\sqrt{2}A(t)\cos 2\pi f_0 t + b_i\sqrt{2}A(t)\sin 2\pi f_0 t, \quad i = 0, 1, \ldots, M-1, \qquad (4.34)$$

where a_i, b_i – pair of coefficients, which jointly represents a sequence from $n = \log_2 M$ bits, transmitted the channel symbol $s_i(t)$;

$A(t)$ – function, determining the form of radiopulses, its maximal value is equal 1;

f_0 – radiopulses frequency.

As every channel symbol is described two coefficients a_i and b_i, signal constellations of these types of modulation appear in two-dimensional space, and the modulated signals are named two-dimensional.

Sum of cosine and sine radiopulses of monotonous forms in (4.34) can be transferable one radiopulse of the same form with the amplitude multiplier A_i and initial phase φ_i, determined:

$$A_i = \sqrt{2\left(a_i^2 + b_i^2\right)}, \qquad \varphi_i = -\mathrm{arctg}\frac{b_i}{a_i}, \qquad i = 0, 1, ..., M-1. \tag{4.35}$$

The identical amplitude multipliers $A_i = a$ have elementary MPSK signals for all i, and their initial phases φ_i differ with the step $2\pi/M$. On a fig. 48 signal constellations of MPSK signals are resulted with pointing of mapping codes. Evidently, that mapping codes are Gray codes.

The elementary MAPSK signals of differ or by the amplitude multipliers A_i, or initial phases φ_i, or amplitude multipliers and initial phases simultaneously. On a fig. 49 constellations of 16-ary of quadrature amplitude modulation (16QAM) are resulted. Signals of MQAM are the separate cases of MAPSK signals.

For signals MQAM attribute the signals of MAPSK, if points of signal constellation are in the knots of square grate. Such structure of constellation gives certain comforts at demodulation.

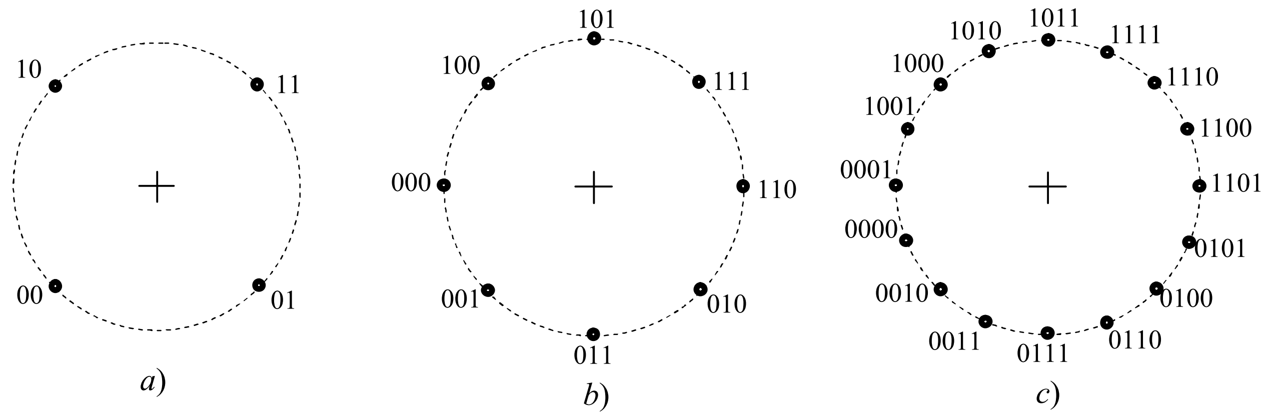


Figure 48 - Signal constellations of signals *a* – QPSK; *b* – 8PSK; *c* – 16PSK

For QPSK:

$$d = \sqrt{2}a\,;\ E_0 = E_1 = E_2 = E_3 = a^2;\ E_{av} = a^2;\ E_b = 0{,}5E_{av} = 0{,}5a^2;\ \ a = \sqrt{2E_b}\,;$$

$$d = 2\sqrt{E_b}\,;\ \ \max d = 2a = 2\sqrt{2E_b}\,.$$

For 8PSK:

$$d/2 = a{\cdot}\sin 22{,}5°;\ \ d = 0{,}765^{\circ};\ \ E_i = a^2;$$

$$E_{av} = a^2;\ \ E_b = E_{av}/3 = a^2/3;\ \ d = 0{,}765\sqrt{3E_b} = 1{,}36\sqrt{E_b}\,.$$

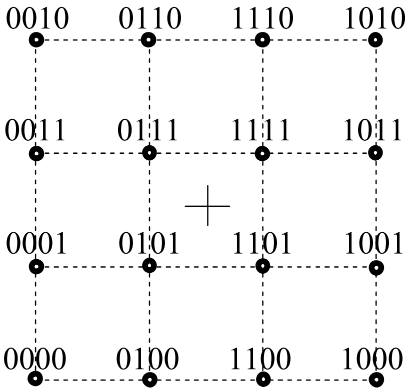


Figure 49 – 16QAM signal constellation

The followings MQAM signals of are used in practice: 4QAM (the same, that QPSK), 8QAM, 16QAM, 64QAM, 256QAM, 1024QAM.

Signals, described expression (4.34), are a sum two DSB-SC signals with identical amplitude spectra which are determined the signal $A(t)$ spectrum. In case if $A(t)$ – the Nyquist pulse, amplitude spectrum to each of constituents, and also their sum, looks like, resulted on a fig. 46. Therefore the spectrum width of channel symbols in the case of MPSK and MAPSK is described expression (4.33).

We will consider the chart of forming of MPSK and MAPSK signals. From comparison of expressions (4.34) and (4.31) flows out, that the chart of two-dimensional bandpass signals forming (modulator) is built on the basis of fig. 47 chart with addition of second subchannel and summarizing (fig. 50). Mapping code coder puts in accordance $n = \log_2 M$ to the entrance bits two rectangular impulse with amplitudes a_i and b_i; rectangular impulse is filtered forming filters, to get the Nyquist impulses; the impulses $a_iA(t)$ and $b_iA(t)$ act at the inputs of double sideband suppressed carrier modulator; the gotten modulated signals are added up.

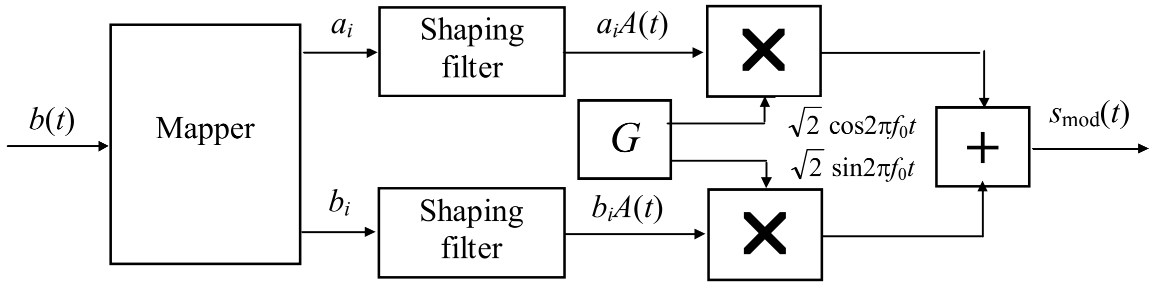


Figure 50 – Modulator of two-dimensional bandpass signals

Thus, made sure again, that the name of method of digital modulation specified what parameter (or by what parameters) channel symbols differ: MPSK – M signals differ in initial phases; MAPSK – M signals differ in amplitudes and/or initial phases. To the two-dimensional signals the signals of binary frequency modulation (BFSK) belong also. From the name of modulation follows that channel symbols are radiopulses, different frequencies:

$$\begin{aligned} s_0(t) &= aA(t)\cos\big(2\pi\big(f_0 - \Delta f/2\big)t + \varphi_0\big), \\ s_1(t) &= aA(t)\cos\big(2\pi\big(f_0 + \Delta f/2\big)t + \varphi_1\big), \end{aligned} \qquad (4.36)$$

where $s_0(t)$ – signal for the transmission of symbol 0;

$s_1(t)$ – signal for the transmission of symbol 1;

f_0 – middle frequency of radiopulses;

Δf – frequency spacing;

$A(t)$ – function, determining the form of radiopulses, its maximal value is equal 1;

a – coefficient which determines energy of signals;

φ_0, φ_1 – initial phases of impulses.

The modulated signal is written down

$$s_{\text{BFSK}}(t)=\sum_{k=-\infty}^{\infty} s_i^{(k)}(t-kT). \tag{4.37}$$

In order that at demodulation radiopulses can it was be divided on condition that their phases φ_0 and φ_1 are arbitrary, the spectra of radiopulses $s_0(t)$ and $s_1(t)$ must not be recovered. If the spectrums of signals are not recovered, such signals orthogonal. We will pass to vector representation of channel symbols

$$\bar{s}_i = a_i\bar{\psi}_0 + b_i\bar{\psi}_1, \tag{4.38}$$

where

$$\begin{aligned}\psi_0(t) &= A(t)\cos\bigl(2\pi\bigl(f_0-\Delta f/2\bigr)t+\varphi_0\bigr),\\ \psi_1(t) &= A(t)\cos\bigl(2\pi\bigl(f_0+\Delta f/2\bigr)t+\varphi_1\bigr),\end{aligned} \tag{4.39}$$

and a mapping code which does equivalent records (4.38) and (4.36) is resulted in table 2. On the basis of expression (4.38) and table 2 BFSK signal constellation looks like, shown on a fig. 51. Here the impulses $\psi_0(\tau)$ and $\psi_1(\tau)$ form the base of signals space.

Table 2 – BFSK mapping code

i	a_i	b_i
0	a	0
1	0	a

In order that a width of spectra of radiopulses was minimum and there was not intersymbol interference (ISI), an impulse $A(t)$ must be the Nyquist impulse. At that rate it is possible to consider that a spectrum of signal $s_{\text{BFSK}}(t)$ is the sum of spectra of two radiopulses of frequencies $f_0 - \Delta f/2$ and $f_0 + \Delta f/2$. On a fig. 52 the rationed spectrum of BFSK signal, from which follows, is presented, that frequency spacing of will be minimum, when the spectra of radiopulses join one to other, and evened:

$$\Delta f_{\min} = \frac{1+\alpha}{T}, \tag{4.40}$$

where T – a clock interval is equal T_b.

Then spectrum width of BFSK signal:

$$F_{\text{BFSK}} = \Delta f_{\min} + \frac{1+\alpha}{T} = \frac{2(1+\alpha)}{T}, \tag{4.41}$$

it is twice greater spectrum widths of BASK and BPSK signals.

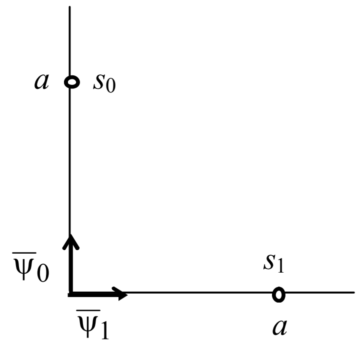


Figure 51 – BFSK signal constellation

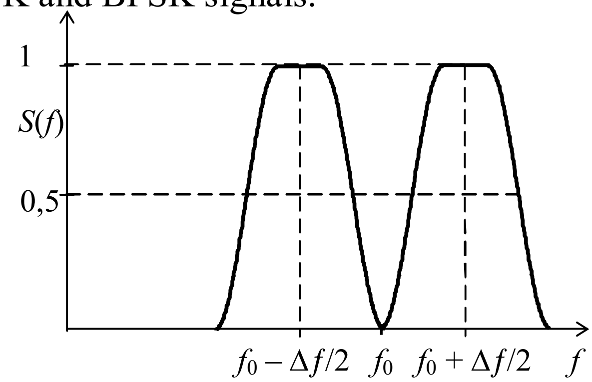


Figure 52 – Spectrum of BFSK signal, on the basis of Nyquist impulses

From expression (4.38) and table 2 the chart of BFSK modulator flows (fig. 53). Forming of BFSK signals differs from forming of *M*PSK signals work of mapping code coder and that generators of carrying waveform frequencies in double sideband suppressed carrier modulators differ on the size $\Delta f/2$ from frequency of bearing oscillation.

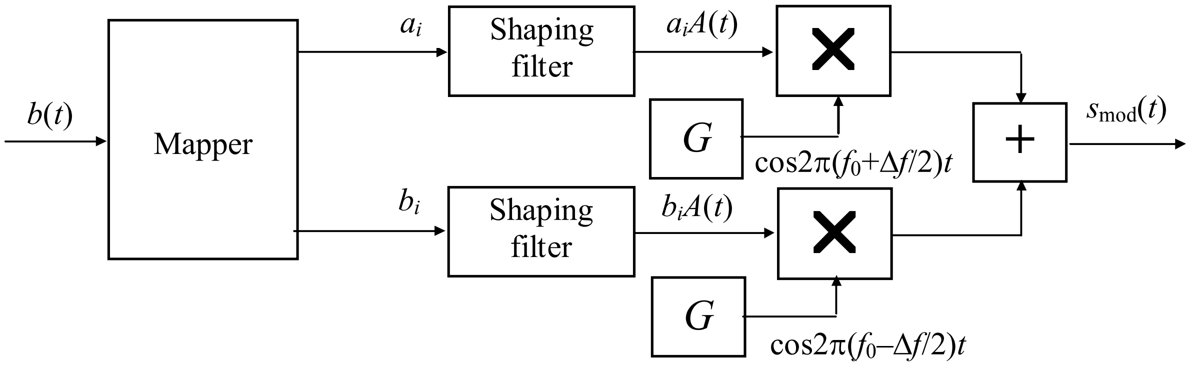


Figure 53 – BFSK signal modulator, if channel symbols are Nyquist impulses

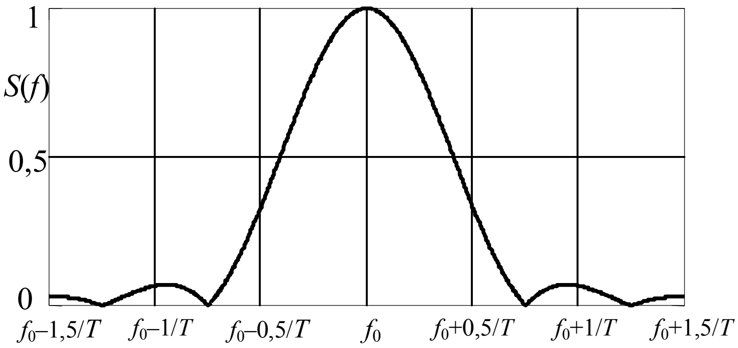


Figure 54 – MSK signal spectrum

If the function $A(t)$ is rectangular impulse, it is necessary to provide forming of BFSK signal the chart of modulator without the «open-phase fault». It is possible, when frequency spacing $\Delta f = k/(2T)$, where $k = 1, 2, 3, ...$; $T = T_b$. When $k = 1$, $\Delta f = 0{,}5/T$ and modulation is named modulation of minimal shift keying (MSK). In the case of MSK the rationed spectrum of signal is described expression

$$S(f) = \frac{\sqrt{1+\cos\left(4\pi(f-f_0)T\right)}}{\sqrt{2}\left(1-\left(4(f-f_0)T\right)^2\right)}. \tag{4.42}$$

Dependence (4.42) is resulted on a fig. 54. With the increase $|f - f_0|$ spectrum decreases at a speed $1/f^2$. If to define the ΔF_{MSK} spectrum width on the first zeros of dependence (4.42) then

$$\Delta F_{MSK} = 1{,}5/T. \tag{4.43}$$

4.9 Signals with spread spectrum

Broadband signals have been used in the systems of telecommunication, since 50-s. The properties of these signals for the first time were used in the systems of radio communication for removal effect of radiowaves multipath propagation.

The next stage of implementation of broadband signals in telecommunication systems was early 90-s, when these signals were used in mobile communication of second generation of standard of IS-95 networks for Code Division Multiple Access realization (CDMA). Code Division Multiple Access has appeared extraordinarily effective in mobile communication networks, therefore the systems of third generation of cdma2000 and UMTS also use broadband signals.

Since middle of 90-s implementation of wireless ports of access, which allow to connect various devices, for example, mobile telephone and computer was begun. The example of such port is Bluetooth, which provides connection of various devices in a radius of 100 m. The known property of broadband signals is their high stability to the narrow-band hindrances which are made by other devices of radio contact.

Broadband signals (BS) are called such signals, which spectrum width is more than minimum band of frequencies, necessary to pass a digital signal of the set rate:

$$\Delta f_{\text{BS}} >> \Delta f_{\min}, \tag{4.44}$$

where Δf_{BS} is a width of spectrum of broadband signal;

$\Delta f_{\min}$ is a minimum band of frequencies, necessary for a transmission; that is equal to the limit of Nyquist, which for binary bandpass signals is equal to transmission rate, i.e. $\Delta f_{\min} = R$.

One of widespread methods of broadband signals forming is Direct Sequence Spread Spectrum – DSSS. Signals, formed by such method, are used in mobile communication of standards of IS-95 and UMTS networks.

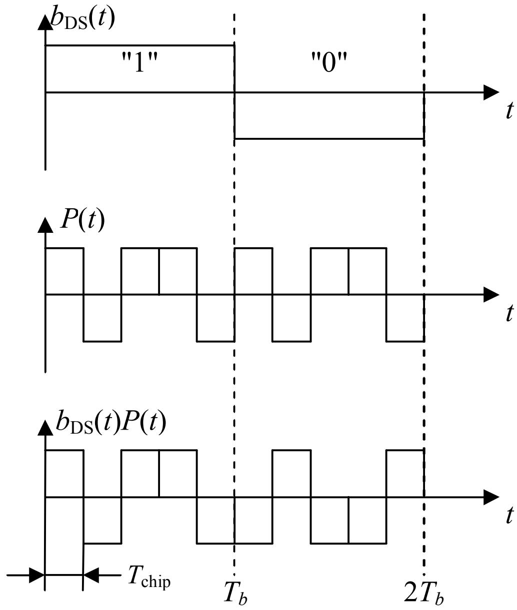


Figure 55 – Direct spread of spectrum: $b_{\text{DS}}(t)$ – digital signal; $P(t)$ – PNS

The principle of direct expansion is that information character is multiplied by a so called Pseudo-noise sequence (PNS) the period of which is equal to duration of character T_b (fig. 55). Expansion of spectrum takes a place due to that duration of element of PNS, which is called a chip, is less then duration of information character, i.e. $T_{\text{chip}} << T_b$.

PNS is a sequence of binary characters which values correspond a certain law. Such PNS are the function of Walsh, *m*-sequence etc.

After multiplying by PNS a signal is given on the input of standard modulator of BPSK, on the output of which a broadband signal appears with Direct Sequence Spread Spectrum:

$$s_{\text{DSSS}}(t) = b_{\text{DS}}(t)P(t)\sqrt{2}A(t)\cos(2\pi f_0 t), \tag{4.45}$$

where $b_{\text{DS}}(t)$ is digital signal;

$P(t)$ is PNS;

$A(t)$ is envelope signal with Direct Sequence Spread Spectrum;

f_0 is carrier frequency.

The chart of signal modulator with Direct Sequence Spread Spectrum is presented on fig. 56. Modulator contains forming LPF, which task is to form impulses for the transmission of chips with a compact spectrum. This filter forms envelope signal. It is chosen such AR LPF, that on his output a spectrum of chip (element of PNS) is the spectrum of Nyquist. In this case, the width of spectrum of signal with direct spread is calculated as:

$$\Delta f_{\mathrm{DSSS}} = \left(1+\alpha\right) R_{\mathrm{chip}} = \left(1+\alpha\right) N R_{\mathrm{b}}, \tag{4.46}$$

where $R_{\mathrm{chip}} = 1/T_{\mathrm{chip}}$ – chip rate, i.e. transmission rate of PNS elements;

$R_{\mathrm{b}} = 1/T_{\mathrm{b}}$ – rate of digital signal;

N is a number of chips on single informative symbol;

α is a Nyquist spectrum roll-off factor $0 \le \alpha \le 1$.

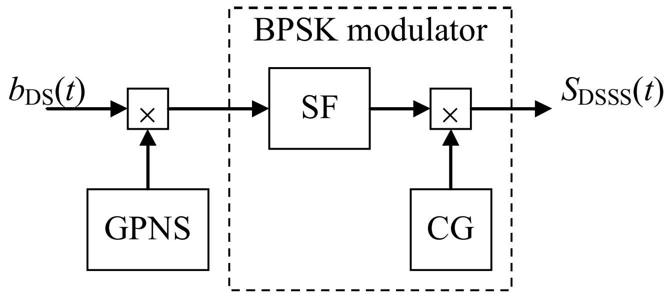


Figure 56 – Direct Sequence Spread Spectrum signal modulator: GPNS – PNS generator; CG – carrier generator

From the expression (4.46) follows, that signal spectrum with direct expansion is exactly in N times wider, than signal without spreading of spectrum, therefore a number is often called the coefficient of spectrum spreading.

Spreading of spectrum allows getting some useful properties.

Fig. 57 demonstrates spectral properties of broadband signals. From the figure follows, that under spreading of spectrum of a signal of value of his spectral density is reduced. Values of spectral density of a broadband signal under considerable spreading of a spectrum (hundreds and thousands times) become near to the value of spectral density of noise. In this case it is difficult to distinguish a signal from noise, if the parameters of signal are unknown, for example, bearing frequency is unknown. Broadband signals are in general called noise-like, as their properties are similar property of white noise.

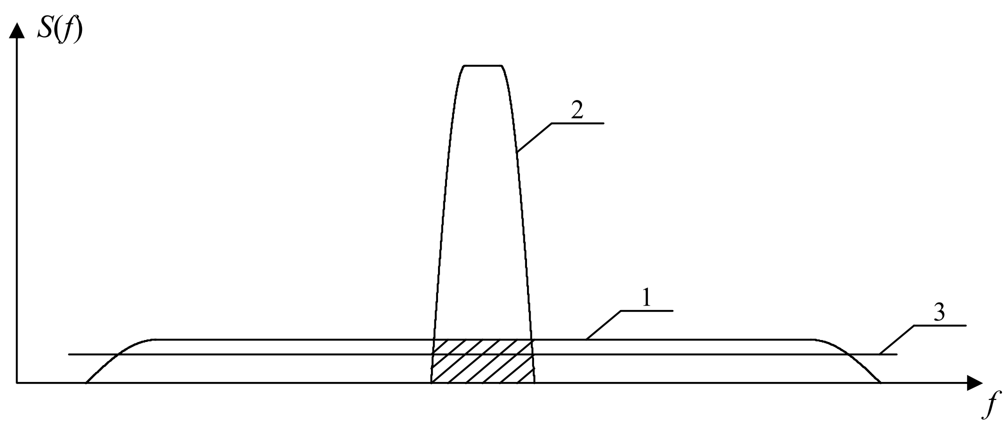


Figure 57 – Spectrums of: 1 – broadband signals; 2 – narrowband signal; 3 – white noise

Similarity of values of spectral density of broadband signals and value of spectral density of noise means, that these signals do not create considerable hindrances to signals without spectrum spreading. On the other side fig. 57 demonstrates another property of broadband signals, namely firmness to the narrow-band hindrances. On a figure it is possible to see, that the spectrum of narrow-band signal (2) destruct insignificant part of spectrum of broadband signal (shaded part), that allows effectively to remove influencing of such hindrance. In fact both systems in which broadband signals are used and the systems in which signals are used without spreading of spectrum can work in one band of frequencies.

At demodulation of signals with the spreading spectrum the so-called matched filters are used. Their properties are:

1) If on the input of a filter is given a signal which is matched with a filter, output response repeats the correlation function of this signal.

2) If on the input of a filter is given a signal which is not matched with a filter, output response repeats the mutual correlation function of input signal and signal which a filter is matched with.

On fig. 58 and fig. 59 correlation properties of broadband signals are demonstrated. The function of correlation of any signal $s(t)$ is determined by expression:

$$K(\tau) = \int_{-\infty}^{\infty} s(t)s(t-\tau)dt, \tag{4.47}$$

and mutual correlation function of two signals $s_1(t)$ and $s_2(t)$ is determined by expression:

$$K_{12}(\tau) = \int_{-\infty}^{\infty} s_1(t)s_2(t-\tau)dt. \tag{4.48}$$

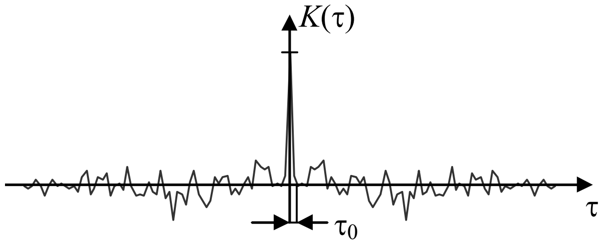


Figure 58 – Broadband signal correlation function

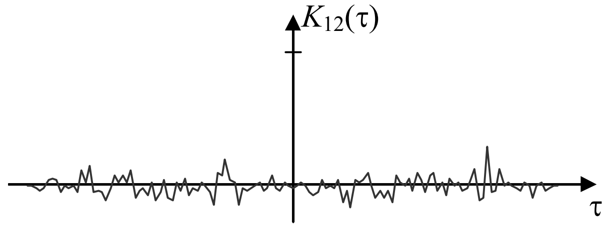


Figure 59 – Two PNS mutual correlation function

As it follows from fig. 58 a correlation function of broadband signal has main narrow signal spike and small on values sideband signal spikes. Duration of the main signal spike of correlation function τ_0 is inversely proportional the band of frequencies of broadband signal, i.e.:

$$\tau_0 \approx \frac{1}{\Delta f_{\text{BS}}}. \tag{4.49}$$

A function of mutual correlation of two broadband signals has small on values signal spikes (fig. 59). It allows realizing the code division of signals (wave-form separation).

Multipath propagation of radiowaves is characteristic for systems of radio contact and leads a few copies of a passed signal $s(t)$, which appear as a result of reflection of electromagnetic wave from various objects, enter into the input of receiver. These copies of a passed signal enter into input of receiver with different delays t_l:

$$s_{\text{rec}}(t) = \sum_{l=1}^{L} s(t - t_l), \tag{4.50}$$

where l is a ray number; a way of distribution of electromagnetic wave is called ray;
L is an amount of rays.

There will be observed sum of functions of correlation on the output of a matched filter (fig. 60). As a function of correlation of broadband signals has the narrow main bursts and small on values sideband bursts, signals of separate rays can be put together laid down – processed by Rake-receiver. Realization of Rake-receiver for processing of signals of two rays is represented on a fig. 61.

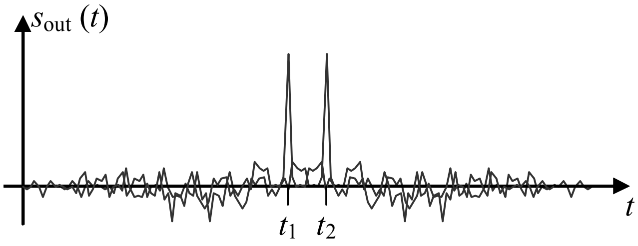


Figure 60 – MF output signal in a case of input sum of two PNS

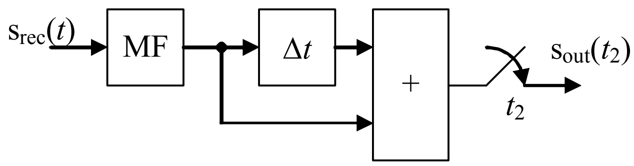


Figure 61 – Rake-receiver

In communication networks with multiple access with code division on the input of MF a few broadband signals of different subscribers which are transmitted simultaneously and in one band of frequencies, enter into input of matched filter:

$$s_{\text{rec}}(t) = \sum_{i=1}^{M} s_i(t), \tag{4.51}$$

where M is an amount of active subscribers.

For waveform separation of broadband signals property 2 of matched filters is used. As a value of function of mutual correlation of broadband signals tends to zero, signals, which are not matched with a filter, will not create considerable hindrances for a signal, which is matched with a filter.

4.10 OFDM

Orthogonal Frequency Division Multiplex (OFDM) realize a parallel-serial transmission (fig. 62). A communication channel is simultaneously transmit L of modulated signals with a serial transfer which has been examined before. For this purpose sequence of binary characters is demultiplexed in L parallel sequences $b_{(1)}(t)$, $b_{(2)}(t)$, …, $b_{(L)}(t)$... On the basis of each of such sequences the modulated signals *of s(1)(t)* are formed, $s_{(1)}(t)$, $s_{(2)}(t)$, ..., $s_{(L)}(t)$ as well as under a serial transfer. Sum of signals $s_{(l)}(t)$, $l = 1, 2, ..., L$ forms the modulated signal of parallel-serial transmission, which is written down as

$$s(t) = \sum_{l=1}^{L} s_{(l)}(t) = \sum_{l=1}^{L} \sum_{k=-\infty}^{k=\infty} s_{(l)i}^{(k)}(t - kT), \tag{4.52}$$

where $s_{(l)i}^{(k)}(t - kT)$ – i-th signal, passed by l-th subchannel on k-th clock interval.

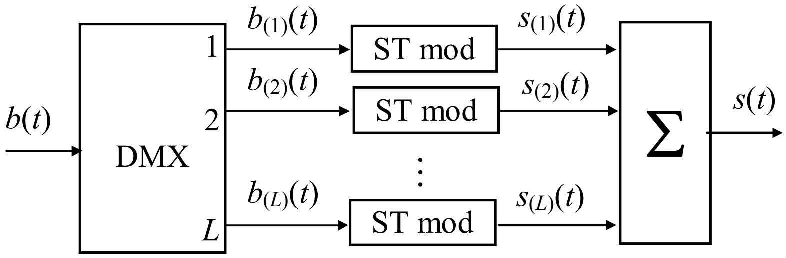


Figure 62 – Modulator of parallel-serial transfer:
DMX – demultiplexer;
ST mod – serial transfer modulator

The feature of this method of transmission are:

- rate of digital signals $b_{(l)}(t)$ lower, than signal $b(t)$ in L times;
- on the output of modulator L of the modulated signals $s_{(l)}(t)$ simultaneously present;
- the modulated signals of separate subchannels $s_{(l)}(t)$ must be such, that they can be divided out of a sum (4.52) for separate demodulation; these signals occupy different bands of frequencies.

In real systems the number L can make thousands. If AR and PR of communication channel are distorting, in the bands of frequencies for separate signals $s_{(l)}(t)$ a channel is practically nondistorted. At first glance complication of method of transmission actually requires simple transformations: all L of modulators are realized by one procedure of inverse fast Fourier transform. For demodulation direct fast Fourier transform is used. OFDM is used in systems of radio contact and digital sound and television broadcasting.

Questions for section 4

1. Define analog modulation. What is used as a carrier in the process of analog modulation?
2. Define amplitude modulation, select needed mathematical expressions. What are the spectral components of AM signal? How to determine the AM signal bandwidth?
3. Explain why is the use of AM signal inefficient?
4. What are the reasons of double sideband suppressed carrier modulation signals appearing? Write down mathematical description of DSB-SC signals. How to determine the DSC-SC signal bandwidth?
5. What are the reasons of single-sideband modulation signals appearing? Write down mathematical description of SSB signals. How to determine the SSB signal bandwidth?

6. Define frequency modulation of signal. Write down mathematical expression of FM signal.
7. Define phase modulation of signal. Write down mathematical expression of PM signal.
8. Define frequency deviation and phase deviation.
9. What is decisive in the construction of schemes of modern modulators?
10. Write down mathematical expression and draw modulator scheme of AM and DSB-SC signals.
11. Write down mathematical expression of SSB signal and explain the possible ways of SSB signal forming.
12. Write down mathematical expression and draw modulator scheme of FM and PM signal.
13. Define detection process.
14. Define digital signal. What digital signal is called binary? What is its main characteristics?
15. What is the reason for the rejection of the use of rectangular pulses in the transmission over communication channel?
16. What conditions should satisfy the impulses with compact spectrum, transmitted over communication channels? What are the names of corresponding impulses?
17. What is called "skew symmetry" of the spectrum? Why is the slope of the pulses spectrum smooth one in practical applications?
18. Give an example of the Nyquist pulse shape and its spectrum. What parameters describe the Nyquist pulse? What is the spectrum width of the Nyquist pulse?
19. Define pulse-amplitude modulation signal. Give the signal constellations of the MPAM ($M = 2, 4, 8$) signals, draw scheme for its formation. Write down formula for the spectrum width calculation.
20. Define modulation code and explain what is Gray code.
21. What kind of device is called a shaping filter? What is the purpose of its application?
22. Define one-dimensional bandpass signals of digital modulation. Draw signal constellations of MASK ($M = 2, 4, 8$) and BPSK signals, draw scheme for their formation. Write down formula for the spectrum width calculation.
23. Define two-dimensional bandpass signals of digital modulation. Draw signal constellations of MPSK ($M \geq 4$) and 16QAM signals, draw scheme for their formation. Write down formula for the spectrum width calculation.
24. Explain the relationship between the quantity of channel symbols and a spectral width of digital modulation signals.
25. Define broadband signal. What purpose is this signal in transmission systems used for?
26. Give an example of the formation of broadband signals.
27. Define parallel-to-serial transmission. What features are it characterized by?

5 NOISE IMMUNITY OF DIGITALLY MODULATED SIGNALS RECEPTION

5.1 General information about the demodulation of digital modulated signals

A transmission system is considered (Fig. 63). The input transmission system receives a baseband signal $b(t)$. The modulator converts the baseband signal $b(t)$ into the modulated signal $s(t)$ for transmission over a distance through the transmission line. Transmission line is undistorted. The demodulator input receives the sum of signal $s(t)$ and noise $n(t)$: $z(t) = s(t) + n(t)$. It is believed, unless expressly agreed that the noise $n(t)$ is AWGN (more precisely, the white noise in the passband of the transmission line). On signal $z(t)$ demodulator have to restore the baseband signal $\hat{b}(t)$, the least different from the transmitted signal $b(t)$. Demodulator must take into account all characteristics of signal $s(t)$ and noise $n(t)$ at demodulation.

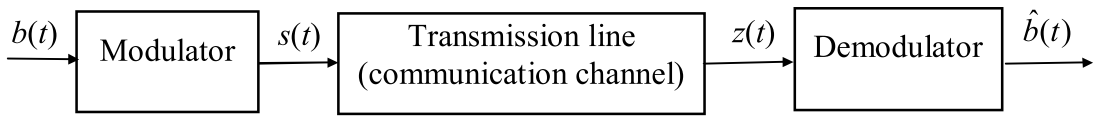


Figure 63 – Transmission system

Depending on the nature of the baseband signal (digital or analog) transmission system is divided into digital and analog. In these systems, different modulation and demodulation are performed; quantitatively accuracy of the baseband signal recovery is estimated in different ways. Properties and signals formation of analog and digital modulation types are considered in the section 4. This module examines the problems arising in the demodulation of signals:

– algorithms (schemes) of signals demodulation;
– noise immunity of transmission systems – the ability of transmission systems to resist the action of noise.

In digital transmission systems signal $s(t)$ is a sequence determined channel symbols, which reflect the baseband digital signal, and following through clock interval T:

$$s(t) = \sum_{k=-\infty}^{\infty} s_i^{(k)}(t - kT), \tag{5.1}$$

where $s_i(t)$, $i = 0, \ldots, M - 1$ – channel symbols;
$M = 2^n$ – number of channel symbols;
n – number of bits transmitting by channel symbol;
T – clock interval is time, which the channel symbols passed through;
$s_i^{(k)}(t - kT)$ – i-th channel symbol, transmitting on k-th clock interval.

At demodulation the numbers i of channel symbols are unknown. Channel symbols (a form of pulses, their frequency (frequencies), their phase, clock interval, moments of the reference clock intervals) are known. Each channel symbol $s_i(t)$, if it is convenient, designated as channel symbol s_i.

Demodulator must make a difference between channel symbols – to make decision $\hat{s}_j$ about numbers of transmitted channel symbols, and gives out a decision bits in accordance with the mapping code.

We assume that the demodulator makes a decision $\hat{s}_j$ on the transmitted channel symbol at every clock interval, independently of the decisions of other clock intervals. This method is called symbol-by-symbol demodulation. At symbol-by-symbol demodulation the work of demodulator can be considered at individual clock intervals, for example, when $k = 0$ demodulated signal has the form

$$z(t) = s_i^{(0)}(t) + n(t). \tag{5.2}$$

Here, number of channel symbol is unknown, and the demodulator is required to make a decision about number i. As on the demodulator input a distorted by noise signal is received, the decision about transmitted channel symbols will contain errors. Channel symbol error is the demodulator decision $\hat{s}_j$ when receiving on the demodulator input symbol is s_i and $j \neq i$. Errors appears randomly as the random nature of noise. Wrong decision by the demodulator about channel symbol reduces to one or more incorrect bits (depending on the mapping code) on the demodulator output. Bit error probability p is a quantitative measure of noise immunity of digital transmission systems. Bit error probability p, along with the signal rate R is a basic characteristic of digital transmission system.

5.2 Criterion of optimality of digital modulation signals demodulators decision rules

You can offer a lot of number of algorithms for demodulation of the signal (5.2). Any demodulation algorithms must take into account all or part of the initial data were: description of the channel symbols $s_i(t)$, $i = 0, ..., M - 1$; a prior probabilities of channel symbols $P(s_i)$, $i = 0, \ldots, M - 1$; statistical characteristics of noise $n(t)$, in particular the probability density $p(n)$.

Our task is to find the optimal demodulation algorithms, assuming that the criterion of optimality is the minimum total error probability of demodulator decision about channel symbol. To minimize the bit error probability at the demodulator output provided the minimum error probability decision about channel symbol one optimizes the mapping code.

Obviously, the minimum total error probability of demodulator decision about channel symbol is a decision on the maximum a posterior probability of channel symbols $P(s_i/z)$. Rule of maximum a posterior probability is formulated in the following way: demodulator makes a decision about symbol transfer $\hat{s}_i$, if a system of $M - 1$ inequalities is satisfied:

$$P(s_i/z) > P(s_j/z), \quad j = 0, 1, \ldots, M - 1; \; j \neq i\,. \tag{5.3}$$

Then use the Bayes formula

$$P(s_i/z) = \frac{P(s_i)p(z/s_i)}{p(z)}, \tag{5.4}$$

where $p(z)$ is unconditional probability density of the signal $z(t)$;

$p(z/s_i)$ is conditional probability density of the signal $z(t)$ provided that $z(t) = s_i(t) + n(t)$.

This assumes that a prior probabilities of channel symbol $P(s_i)$, $i = 0, ..., M - 1$ and the probability density of the noise $p(n)$ allow us to calculate the conditional probabilities density $p(z/s_i)$. Typically, channel symbols are equally probable in transmission systems, i.e.

$$P(s_i) = 1/M, \quad i = 0, ..., M - 1. \tag{5.5}$$

With this in mind we rewrite the system of inequalities (5.3) by deleting the entries unconditional probability density $p(z)$, that is included in the right and left sides, and not dependent on the index:

$$p(z/s_i) > p(z/s_j), \quad j = 0, 1, \ldots, M - 1;\ j \neq i. \tag{5.6}$$

The system of inequalities (5.6) is called usually the maximum likelihood rule, which is formulated in the following way: demodulator makes a decision about transfer symbol $\hat{s}_i$, if conditional probability density of the signal $z(t)$ provided that $z(t) = s_i(t) + n(t)$ is maximal. Conditional probability density is called also the likelihood function.

Note that the maximum likelihood rule is used for demodulators' construction, if the channel symbols are equiprobable (in this case is realized generally maximum a posterior probability) or a prior probabilities of channel symbols are unknown.

To move from decision rules (5.6) to the demodulation algorithm it is necessary to use representation of signals and noise in a multidimensional space. Full description of the set of channel symbols $s_i(t)$, $i = 0, ..., M - 1$ is their representation in the N-dimensional space (see section 4). They considered one-dimensional ($N = 1$) and two-dimensional ($N = 2$) signals as the most frequently used. For clarity of the description the signal constellations are used.

Realization of input demodulator sum of signal and noise $z(t)$ – the relation (5.2) – can be also represented in a multidimensional space formed by the orthonormal basis functions $\{\psi_k(t)\}$, which are used for the description of channel symbols $s_i(t)$:

$$s_i(t) = \sum_{k=0}^{N-1} a_{ik}\psi_k(t), \ i = 0, 1, \ldots, M - 1; \tag{5.7}$$

$$z(t) = \sum_{k=0}^{N-1} z_k \psi_k(t), \tag{5.8}$$

where

$$z_k = \int_0^{T_s} z(t)\psi_k(t)dt, \quad k = 0, 1, \ldots, N - 1 \ – \tag{5.9}$$

is decomposition coefficients of demodulated signal. This ratio suggests that the functions $\{\psi_k(t)\}$ form an orthonormal basis with orthogonality interval $(0, T_s)$, where T_s is duration of channel symbols.

The coefficients z_k are uncorrelated Gaussian random values. Therefore, the conditional probability densities $p(z/s_i)$, $i = 0, ..., M-1$ are N-dimensional distributions, determined by the product N one-dimensional distributions. The right and left sides of inequalities (6) can be rewritten as

$$\frac{1}{\left(\sqrt{2\pi}\sigma\right)^N}\exp\left(-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}\left(z_k - a_{ik}\right)^2\right) > \frac{1}{\left(\sqrt{2\pi}\sigma\right)^N}\exp\left(-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}\left(z_k - a_{jk}\right)^2\right), \quad (5.10)$$

$$j = 0,\ 1,\ \ldots, M-1;\ j \neq i,$$

where σ is a root mean square deviation (RMS) of the coefficients z_k.

Verification of inequalities (10) is equivalent to checking the inequalities

$$\sum_{k=0}^{N-1}\left(z_k - a_{ik}\right)^2 < \sum_{k=0}^{N-1}\left(z_k - a_{jk}\right)^2, \quad j = 0,\ 1, \ldots, M-1;\ j \neq i. \quad (5.11)$$

Sums in inequality (5.11) are nothing other than squares of the distances between the demodulated signal and the channel symbols in the N-dimensional space. Last relation expressed the maximum likelihood rule for the optimal demodulation: decision about the number of channel symbols shall be in favour of the signal, the distance between that and demodulated signal is minimal. Note that there is no need to take a square root of the left and right sides of inequality (5.11), the square of the distance can be also compared.

Rule of the demodulation can be interpreted in a following way:

– signal space is divided into M disjoint areas with the names s_i, $i = 0, ..., M-1$, each area s_i is a set of points that are closer to the signal $s_i(t)$, rather than to other signals;

– demodulator makes a decision about transfer signal $s_i(t)$, if the point $z(t)$ in N-dimensional space falls in the area s_i.

Notes:

1. The criterion of minimum total error probability of demodulator decision in the Russian-language literature is also called the criterion of the ideal observer (a term introduced by V.A. Kotelnikov).

2. If the error making by demodulator differently undesirable for various channel symbol, they used the criterion of minimum average risk (criterion minimal total error probability of demodulator decision assumes that all errors are equally undesirable).

Example 1. Let's find the decision rule for optimal demodulation of one-dimensional ($N = 1$) BPAM signal, whose channel symbols are described

$$s_i(t) = a_i A(t), \quad i = 0,\ 1, \quad (5.12)$$

where $A(t)$ is pulse, with set time and spectral characteristics, which maximum value and energy are equal to 1;

a_i is factors that show the transmitted bits: $a_1 = a$, $a_0 = -a$; value a determines the energy of channel symbols.

The function $A(t)$ plays the role of the basis functions $\psi_0(t)$ for the representation of the signal $s_i(t)$ in one-dimensional space.

On the demodulator input on a concrete clock interval (e.g., $k = 0$) comes

$$z(t) = a_i A(t) + n(t). \tag{5.13}$$

Find the coefficient submission signal $z(t)$ in the basis $A(t)$

$$z_0 = \int_{-\infty}^{\infty} z(t) A(t) dt = a_i + \zeta, \tag{5.14}$$

where ζ – random variable with Gaussian probability distribution, since it is the result of linear transformation noise with a Gaussian distribution.

In accordance with expression (5.11) for the restoration of binary symbol, transfer at this clock interval, you should compare the distances $z_0 - a_1$ and $z_0 - a_0$ and decide in favor of the minimal of these:

$$\text{if } z_0 - a_1 < z_0 - a_0, \text{ then } \hat{b}_i = 1;$$

$$\text{if } z_0 - a_1 > z_0 - a_0, \text{ then } \hat{b}_i = 0.$$

As follows from relation (5.14) z_0 is nothing short of the estimate $\hat{a}$ of the coefficient a_i. Let us discuss the decision on a comparison of conditional probabilities densities (6) (rule of maximum likelihood): demodulator makes a decision about the transmitted channel symbol $\hat{s}_1$, if $p(\hat{a}/s_1) > p(\hat{a}/s_0)$, and the decision $\hat{s}_0$, if $p(\hat{a}/s_1) < p(\hat{a}/s_0)$.

In fig. 64 the BPAM signal constellation and the conditional probability density estimates of the coefficient, which describes demodulated signal are shown. This figure shows that instead of comparing the conditional probability density demodulator can make a decision on the result of comparison of estimates $\hat{a}$ with threshold value λ according to the rule: if $\hat{a} > \lambda$, then a signal $s_1(t)$ was transmitted, and if $\hat{a} < \lambda$, then a signal $s_0(t)$ was transmitted.

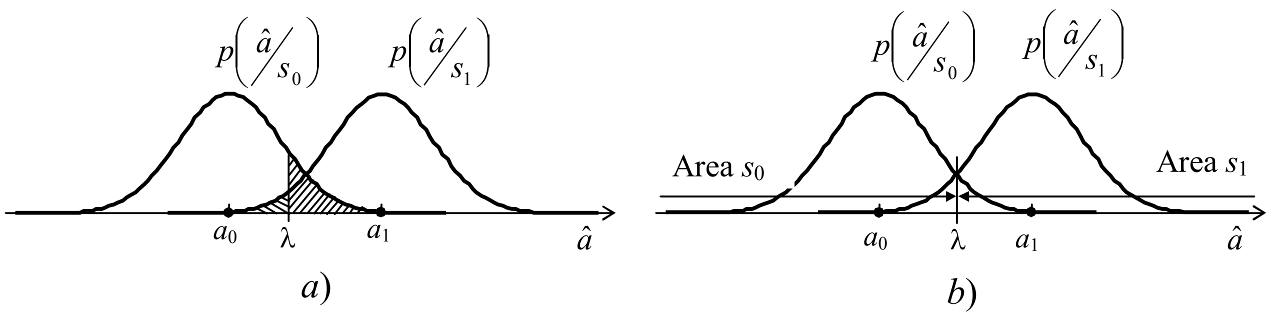


Figure 64 – BPAM signal constellation and the conditional probability density estimates: a – not optimal setup of threshold; b – optimal setup of threshold

The error probability in the transmission $s_1(t)$

$$P_{\text{er}}(s_1) = \int_{-\infty}^{\lambda} p(\hat{a}/s_1) d\hat{a}. \tag{5.15}$$

Similarly, we define the error probability in the transmission $s_0(t)$

$$P_{\text{er}}(s_0) = \int_{\lambda}^{\infty} p(\hat{a}/s_0) d\hat{a}. \tag{5.16}$$

The unconditional error probability of signal and bit

$$p = P(s_1)P_{\text{er}}(s_1) + P(s_0)P_{\text{er}}(s_0) = 0{,}5[P_{\text{er}}(s_1) + P_{\text{er}}(s_0)]. \tag{5.17}$$

From fig. 64, *a* can be seen that the probabilities in square brackets equal to the shaded areas. It is easy to see that the total area will be minimal when the threshold is midway between a_1 and a_0:

$$\lambda = 0{,}5(a_1 + a_0). \tag{5.18}$$

This value is shown in fig. 64, *b*. It is seen that $P_{\text{er}}(s_1) = P_{\text{er}}(s_0)$ in this case. It also shows the partition of the signals space (in this example, the number axis) in the signals areas: the range of values $\hat{a}$, where $p(\hat{a}/s_1) > p(\hat{a}/s_0)$, is an area of the symbol s_1, and the range of values $\hat{a}$, where $p(\hat{a}/s_1) < p(\hat{a}/s_0)$ is the area of symbol s_0.

Note. Fig. 64 shows that the probability of error depends on the values of RMS σ of estimate $\hat{a}$ – the less RMS, the less error probability. It will be shown that the calculation of estimates for the algorithm (5.14) provides a minimum value of RMS estimates.

5.3 Algorithm for optimal demodulation of digital modulation signals (general case)

The received above maximum likelihood rule is reflected in the form of M-ary signals optimal demodulator scheme (Fig. 65). Individual circuit blocks perform the following functions:

1. Determination of signal $z(t)$ coordinates in the space of channel symbols, on the basis of (5.9).
2. Determination of the squares of the distances between $z(t)$ and $s_i(t)$ in the space of channel symbols based on the ratio

$$d^2(z, s_i) = \sum_{k=0}^{N-1} (z_k - a_{ik})^2, \quad i = 0,\ 1, \ldots, M-1. \tag{5.19}$$

3. Comparison of the squares of distances (or distances), identification number j, which corresponds to the minimum value of $d^2(z, s_j)$, delivery of solutions $\hat{s}_j$.
4. Presentation channel symbol s_j by bits in accordance with the mapping code.

On the next clock interval mentioned actions are repeated.

The scheme of optimal demodulator (fig. 65) can be used to demodulate the signal of an arbitrary specified type of modulation – types of modulation have different values N and M, form of channel symbols. There is one limit. Channel symbols are equally probable, but this restriction is usually performed in practice.

Depending on the modulation type, by considering properties of channel symbols, coordinates of the signal $z(t)$ can calculate in different ways in the space of channel symbols. This diversity of ways of demodulators constructing, which will be devoted a significant place in the following sections. The remaining blocks of the demodulator: calculation of the square distances between $z(t)$ and $s_i(t)$, a decision on the basis of minimum distance value, decoding based on the mapping code are the standard for the demodulator, and thereafter will be merged into one unit which will be called "decision".

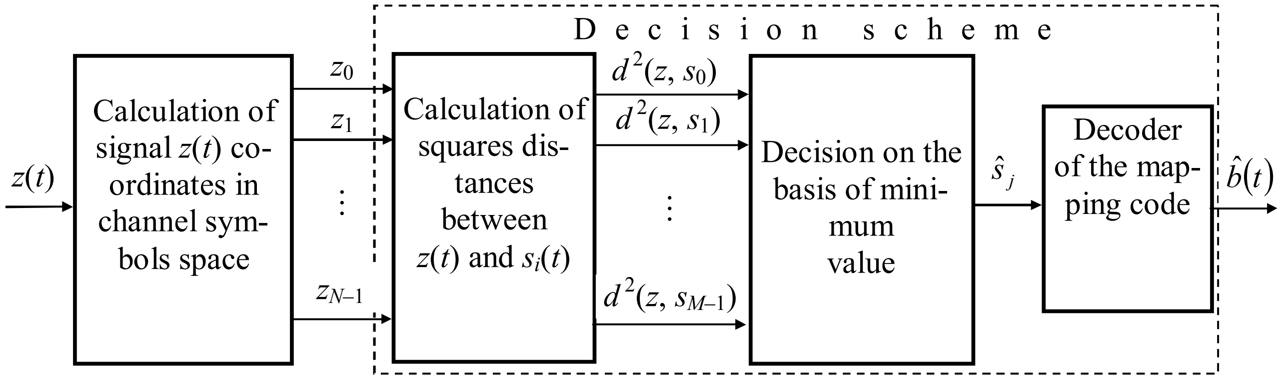


Figure 65 – Optimal demodulator of M-ary signal

Calculation of signal $z(t)$ coordinates in the space of channel symbols based on the ratio (5.9) can be satisfied by correlators schemes (fig. 66). The scheme of the correlator includes a signal generator $\psi_k(t)$ – a replica of the k-th basis function of channel symbol, multiplier and integrator with reset. Switch takes sample at the end of the signal $\psi_k(t)$, and the integrator is given in the zero state to be ready for the next signal processing.

The term "correlator" is due to the fact that the scheme calculates the value of cross-correlation function between signals $z(t)$ and $\psi_k(t)$.

The calculation corresponding to the relation (5.12), can be performed by linear electric circuit with a specially chosen impulse response $g_k(t)$. In the general case, the output signal $y(t)$ and the input signal $z(t)$ linear circuit are connected by the relation, called the Duhamel integral

$$y(t) = \int_{-\infty}^{\infty} z(\tau) g(t-\tau) d\tau, \tag{5.20}$$

where $g(t)$ is the impulse response of a circuit.

Let

$$g_k(t) = \psi_k(T_s - t). \tag{5.21}$$

Since the signal $\psi_k(t)$ exists on the interval $(0, T_s)$, then at the same interval there exists a function $\psi_k(T_s - t)$. Therefore, the limits of integration are 0 and T_s. Let's find the value $y_k(T_s)$

$$y_k(T_s) = \int_0^{T_s} z(\tau) \psi_k(T_s - T_s + \tau) d\tau = \int_0^{T_s} z(t) \psi_k(t) dt = z_k. \tag{5.22}$$

Linear electric circuit with the impulse response (5.21) is called a matched filter with the signal $\psi_k(t)$ (the impulse response is a mirror reflection of a signal).

Thus, the calculation of expansion coefficients of the signal $z(t)$ can be accomplished with the help of correlator or matched filter (fig. 67). Accordingly, the demodulator scheme will contain N correlators or N matched filter and decision scheme (fig. 68).

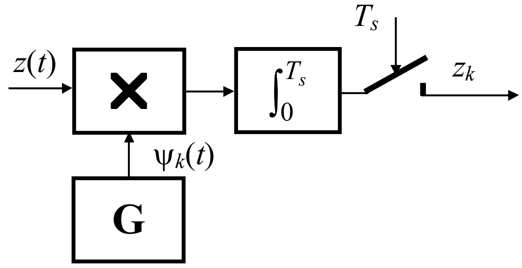


Figure 66 – Calculation of the coefficient z_k by correlator

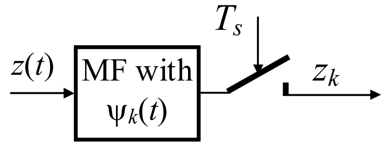


Figure 67 – Calculation of the coefficient z_k by matched filter

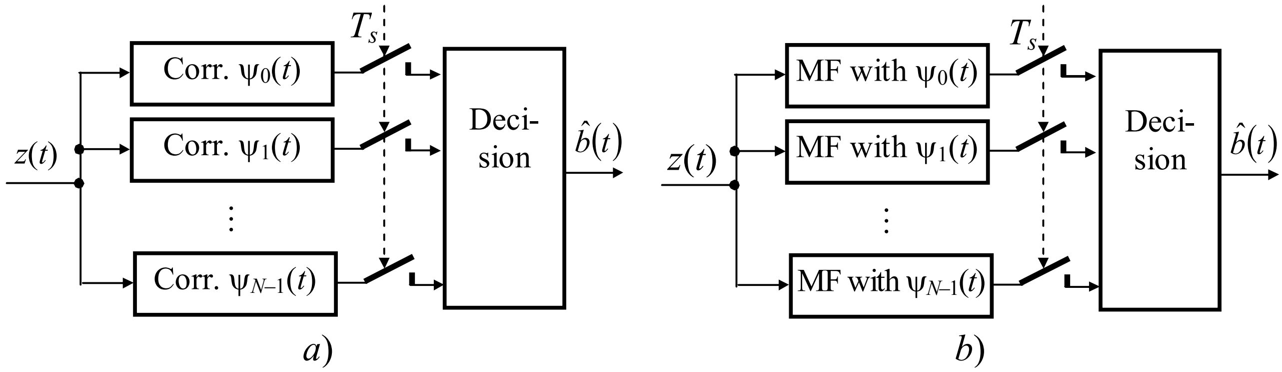


Figure 68 – Optimum demodulators:
a – on the basis of correlators; *b* – on the basis of matched filters

5.4 Matched filter

In section 5.3 the term "matched filter" (MF) was introduced as a device for coefficient calculation representation of demodulated signal in orthonormal basis. MF is more widely used in transmission systems equipment. Therefore we consider below MF with common positions.

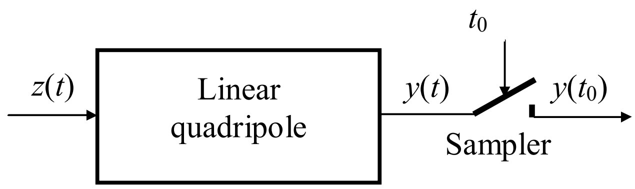


Figure 69 – Inclusion of quadripole and sampler

There is a linear quadripole (filter) with transfer function $H(j\omega)$. Its input is the sum of deterministic pulse signal $s(t)$ and noise $n(t)$: $z(t) = s(t) + n(t)$. There is a sum of responses to the signal and noise $y(t) = y_s(t) + y_n(t)$ on the quadripole output. Sampler is connected to the quadripole output to sample at the time t_0 (fig. 69). This device is used to reduce noise and to sample in order to determine the maximum value of the pulse.

The filter is called matched with a signal $s(t)$, if at submission on its input of the sum of the signal $s(t)$ and a noise $n(t)$ on its output at the certain moment (designated t_0) the maximum ratio of the instantaneous power of a signal $y_s^2(t_0)$ to the average power of noise $P_{n\,\text{out}}$ takes place: $\rho_{\text{peak}} = y_s^2(t_0)/P_{n\,\text{out}}$.

Matched filter is used not only for signal/noise ratio maximization, but for other important signal and noise transformations also. Therefore we will consider properties of MF:

1. Let's find the transfer function $H(j\omega)$. Signal $s(t)$ is specified and the noise $n(t)$ is white noise with power spectral density $N_0/2$.

Let

$$S(j\omega)=\int_{-\infty}^{\infty} s(t)e^{-j\omega t}dt \; - \qquad (5.23)$$

spectral density of a signal $s(t)$. Then spectral density of the output signal $y_s(t)$ is defined as

$$S_{\text{out}}(j\omega)=S(j\omega)H(j\omega). \qquad (5.24)$$

Sampling value of signal $y_s(t_0)$ is the inverse Fourier transforms from $S_{\text{out}}(j\omega)$

$$y_s(t_0)=\frac{1}{2\pi}\int_{-\infty}^{\infty} S(j\omega)H(j\omega)e^{j\omega t_0}d\omega. \qquad (5.25)$$

Noise power on the filter output and average square of noise sample $y_n(t_0)$ is defined as

$$P_{n\,\text{out}}=\overline{y_n^2(t_0)}=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{N_0}{2}\left|H(j\omega)\right|^2 d\omega. \qquad (5.26)$$

Signal to noise ratio is a signal power $y_s^2(t_0)$ to average square of noise sample ratio $P_{n\,\text{out}}$ in the sampling moment

$$\rho=\frac{\left(\dfrac{1}{2\pi}\right)^2\left[\int_{-\infty}^{\infty} S(j\omega)H(j\omega)e^{j\omega t_0}d\omega\right]^2}{\dfrac{1}{2\pi}\cdot\dfrac{N_0}{2}\int_{-\omega}^{\infty}\left|H(j\omega)\right|^2 d\omega}. \qquad (5.27)$$

We seek a transfer function $H(j\omega)$, in which there is a maximum value of the numerator in the ratio (5.27). For the next calculation it is necessary to take advantage of that the integral in the numerator – scalar product of two functions $S^*(j\omega)$ and $H(j\omega)e^{j\omega t_0}$ ($S^*(j\omega)$ is complex conjugate function with the function $S(j\omega)$). Scalar product is maximum if these functions coincide with accuracy to any positive coefficient c, i.e. $H(j\omega)e^{j\omega t_0}=c\cdot S^*(j\omega)$. Hence, the maximum of the numerator (5.27) takes place when the transfer function

$$H(j\omega)=c\cdot S^*(j\omega)e^{-j\omega t_0}. \qquad (5.28)$$

After substituting expression (5.28) in relation (5.27) we obtain

$$\rho=\frac{\left(\frac{1}{2\pi}\right)^2\left[c^2\int\limits_{-\infty}^{\infty}\left|S(j\omega)\right|^2d\omega\right]^2}{\frac{c^2}{2\pi}\cdot\frac{N_0}{2}\int\limits_{-\omega}^{\infty}\left|S(j\omega)\right|^2d\omega}=\frac{2E_s}{N_0}. \tag{5.29}$$

Here it is used, that energy of a signal is defined as

$$E_s=\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\left|S(j\omega)\right|^2d\omega. \tag{5.30}$$

We see that ratio (5.28) provides not only a maximum of the numerator of the SNR (5.27), but the maximum of this ratio (the value of ρ does not depend on the specific form of the transfer function $H(j\omega)$, included in the denominator). In such a way, the problem of determining the transfer function of MF $H(j\omega)$ is solved.

2. Value (5.29) determines the maximum possible signal to noise ratio at the output of the filter in the sampling moment. This ratio is called peak

$$\rho_{\text{peak}}=\frac{2E_s}{N_0}. \tag{5.31}$$

Let's define the gain in the signal/noise ratio, which shows how many times increases the signal/noise ratio provided with the MF,

$$g_{\text{MF}}=\frac{\rho_{\text{peak}}}{P_s/P_n}=\frac{2E_s\cdot P_n}{N_0\cdot P_s}=\frac{2P_sT_sN_0F_{\text{n}}}{N_0\cdot P_s}=2F_{\text{n}}T_s, \tag{5.32}$$

where F_n is band of noise on filter input;

T_s is signal $s(t)$ duration;

P_s and P_n are average powers of signal and noise on filter input.

From the expression (5.32) evidently, that at certain correlations between the noise band and duration of signal gain can take on large values.

3. Let's find amplitude response and phase response of MF. Transfer function of a linear circuit defines AR and PR circuit

$$H(j\omega) = H(\omega)\exp(j\varphi(\omega)), \tag{5.33}$$

where $H(\omega)$ is AR circuit, $\varphi(\omega)$ is PR circuit.

Let's present the spectral density of a signal through the module and the argument

$$S(j\omega) = S(\omega)\exp(j\psi(\omega)), \tag{5.34}$$

where $S(\omega)$ is an amplitude spectrum of a signal, $\psi(\omega)$ is a phase spectrum of a signal.

After substituting (5.33) and (5.34) in (5.28) we find that the frequency response of MF

$$H(\omega) = cS(\omega) \tag{5.35}$$

up to an arbitrary factor coincides with the amplitude of the signal the filter is matched with. MF transmission coefficient is bigger at those frequencies at which signal $s(t)$components is bigger.

Equality arguments of left and right sides of (5.28) gives

$$\varphi(\omega) = -\psi(\omega) - \omega t_0, \tag{5.36}$$

that is interpreted as follows: PR of MF up to a linear term is opposite in sign to the phase spectrum signal the filter is matched with.

For finding out of physical essence of MF PR will consider some signal harmonic of frequency f_i: $A_i\cos(2\pi f_i + \psi_i)$. This harmonic on the MF output is determined:

$$A_i H(f_i)\cos(2\pi f_i t + \psi_i + \varphi(f_i)) = A_i H(f_i)\cos(2\pi f_i t + \psi_i - \psi_i - 2\pi f_i t_0).$$

The full phase of oscillation is equal $2\pi f_i (t - t_0)$. At a moment $t = t_0$ the full phase of oscillation is equal to the zero independently of frequency. At this moment all harmonics are in a phase and at addition give the maximally possible value of response.

4. Let's find the impulse response of MF as the inverse Fourier transform of the transfer function

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(j\omega)e^{j\omega t}dt = \frac{c}{2\pi}\int_{-\infty}^{\infty} S(-j\omega)e^{-j\omega t_0}e^{j\omega t}dt = \frac{c}{2\pi}\int_{-\infty}^{\infty} S(j\omega)e^{j\omega(t_0 - t)}dt = cs(t_0 - t). \tag{5.37}$$

We see, that impulse response of matched filter is mirror reflection of a signal which the filter is matched with, about the point t_0 in scale c.

Example 2. We will build the graph of impulse response of filter, matched with a signal

$$s(t) = \begin{cases} 1 - t/T_s, & 0 \le t \le T_s, \\ 0, & t < 0,\ t > T_s. \end{cases}$$

Condition of a physical realization of a circuit is $g(t) \equiv 0$ at $t < 0$. The matched filter accumulates all of the components of a signal and in sampling moment forms of them the greatest possible value. It is obvious, that the filter accumulated all components of a signal, sampling moment should not be less than the duration of a signal: $t_0 \ge T_s$, where T_s is a signal $s(t)$ duration.

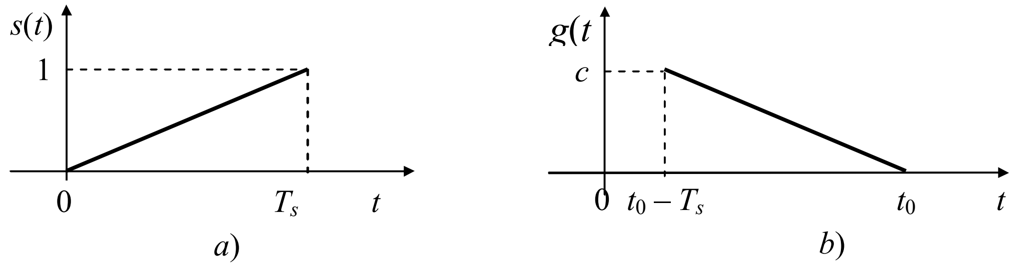


Figure 70 – Graphs: *a* – signal,
b – impulse response of filter matched with signal

5. Let on the MF input operates arbitrary signal $z(t)$. The response of the filter is determined by the Duhamel integral

$$y(t)=\int_{-\infty}^{\infty} z(\tau)g(t-\tau)d\tau = c\int_{-\infty}^{\infty} z(\tau)s(t_0-t+\tau)d\tau = cK_{zs}(t-t_0), \tag{5.38}$$

where $K_{zs}(\tau)$ – cross-correlation function of signals $z(t)$ and $s(t)$.

From the expression (5.38) follows that the shape of the signal on the MF output is determined by cross-correlation function of the input signal and the signal which the filter is matched with, namely, it repeats the cross-correlation function in the scale c and shifted to the right on t_0.

If in the ratio (5.38) put $c = 1$ and $t_0 = T_s$, it is easy to see that $y(T_s)$ gives the value of the scalar product of signals $z(t)$ and $s(t)$. This property of MF was used to determine the expansion coefficients in the ratio (5.22).

6. Let on an input of MF be the signal $s(t)$ with which the filter is matched. Then, on foundation (5.38) will write down

$$y(t)=cK_s(t-t_0), \tag{5.39}$$

where $K_s(\tau)$ – correlation function of signal $s(t)$.

Thus, if on the MF input we have signal, which filter is matched with, the response of the filter is determined by the correlation function of signal, namely, it repeats the correlation function in the scale c and shifted to the right on t_0.

Exercise 3. Let's illustrate the properties of MF considered by the example of a filter, matched with a rectangular pulse of amplitude A and duration T_s. Let $c = 1/A$ and $t_0 = T_s$. Impulse response of filter, matched with rectangular pulse have rectangular shape, amplitude 1 and duration T_s, i.e. impulse response coincides with the signal (fig. 71,*a*).

The spectral density of the rectangular impulse is defined with the aid of Fourier transform

$$S_r(j\omega) = \int_0^{T_s} A e^{-j\omega t}dt = \frac{A}{-j\omega}\left[e^{-j\omega T_s}-1\right] = \frac{2A}{\omega}\sin\frac{\omega T_s}{2}\cdot e^{-j\omega T_s/2}. \tag{5.40}$$

According to the ratio (5.28) we will get the expression for transfer function of filter, matched with rectangular pulse, if $c = 1/A$ and $t_0 = T_s$

$$H(j\omega) = \frac{1}{j\omega}\left[1-e^{-j\omega T_s}\right]. \tag{5.41}$$

Hence this relationship follows that the scheme of the filter, matched with the rectangular impulse consists of the integrator (with transfer function $1/j\omega$), delay device for the period of T_s (with transfer function $\exp(-j\omega T_s)$) and the subtracter (Fig. 71,*c*). In this figure the numbers denote the individual points of scheme to discuss its work.

It is easy to receive expression for AR of the filter, matched with the rectangular impulse. Final expression for AR after transition to a variable f looks like function $\sin(x)/x$

$$H_r(f) = T_s \frac{\sin \pi f T_s}{\pi f T_s}. \quad (5.42)$$

AR of MF and amplitude spectrum of the signal is shown in fig. 71,*b*.

Fig. 72,*a* shows the processes taking place in MF for its input δ-function. There is the impulse response on the output. Fig. 72,*b* shows the process taking place in MF for its input pulse which filter is matched with. On the scheme output is observed a response, which coincides with the correlation function of the rectangular pulse duration T_s (see subsection 2.2).

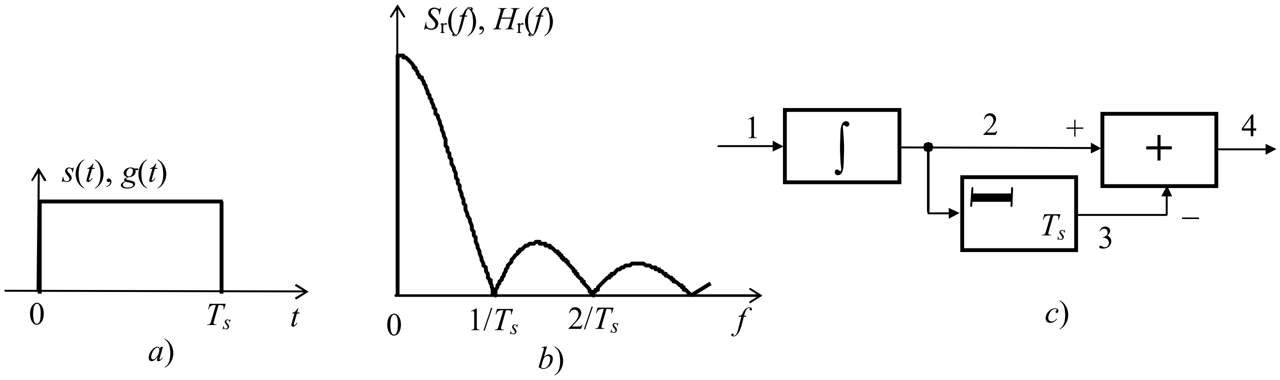


Figure 71 – Characteristics of the filter matched with rectangular pulse: *a* – impulse response; *b* – AFC; *c* – MF scheme

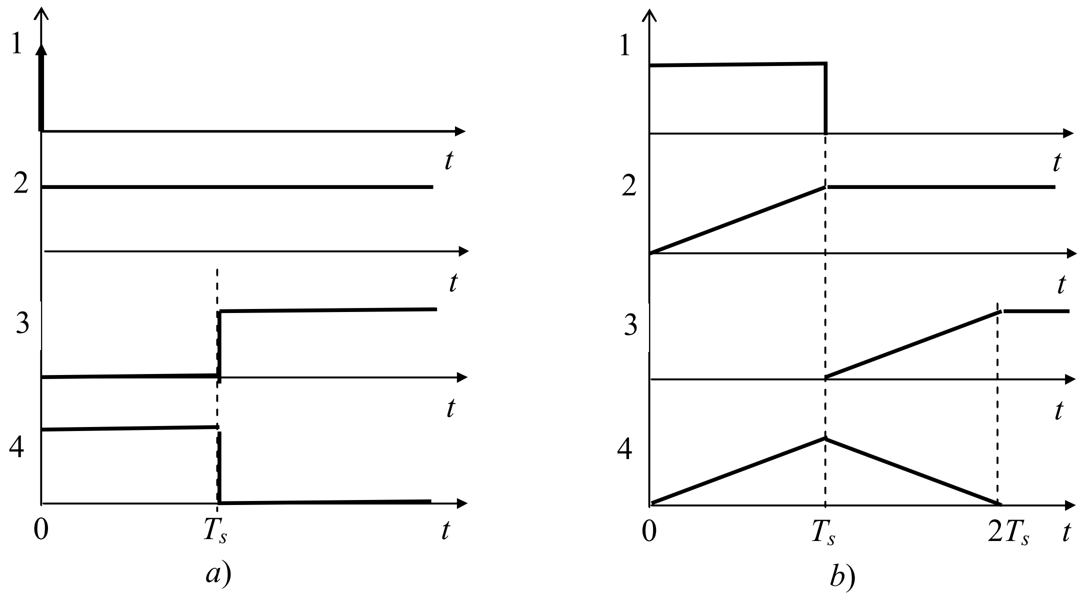


Figure 72 – Time diagrams, which illustrated work of the filter matched with rectangular pulse: *a* – δ-function, *b* – rectangular pulse on the input

5.5 Application of matched filter in the *M*PAM signal demodulator

Consider the joint scheme of the *M*PAM signal modulator and demodulator (fig. 73). Scheme of the modulator is based on the description of *M*PAM signals (discussed in section 4)

$$s_i(t) = a_i A(t), \quad i = 0, 1, ..., M-1, \quad (5.43)$$

where $A(t)$ – pulse with a certain frequency and temporal characteristics;

a_i – coefficient that map transferred bits.

In terms of signals space *M*PAM is a one-dimensional space, $A(t)$ is basic function, a_i are decomposition coefficients.

Mapper generates the coefficients a_i by the input digital signal. On each clock interval $n = \log_2 M$ bits block is assigned to coefficient a_i. This coefficient applied to the shaping filter (SF) input by a signal $a_i\delta(t)$. SF generates a pulse $a_iA(t)$.

Demodulator scheme is based on earlier sections material. The input filter receives a sum of signal and noise $a_iA(t) + n(t)$. Matched filter reduces noise, and on its output there is a useful signal $a_iP(t)$ and noise $\zeta(t)$. Sampler takes the sample and gives out an estimate $\hat{a}$ of the coefficient a_i. The maximum value of pulse $P(t)$ at the sample time equals to 1, $\hat{a} = a_i + \zeta$ (estimate $\hat{a}$ can be considered as the coefficient z_0 submission signal $z(t)$ in one-dimensional space on the basic function $A(t)$). The Sampler is controlled by a sequence of pulses from a clock recovery scheme (ClR), which ensures the samples in the time of maximum signal/noise ratio. Based on information received from the sampler estimate $\hat{a}$ decision scheme makes a decision about the number of transmitted channel symbol and gives the solution of binary symbols in accordance with the modulation code.

Since shaping filter is excited by δ-function, then the amplitude spectrum of pulse $A(t)$ is equal to the AR of SF

$$S_A(f) = H_{SF}(f). \tag{5.44}$$

Amplitude spectrum of pulse $P(t)$ s defined as

$$S_P(f) = S_A(f)\cdot H_{MF}(f), \tag{5.45}$$

where $H_{MF}(f)$ is an amplitude response of filter matched with pulse $A(t)$.

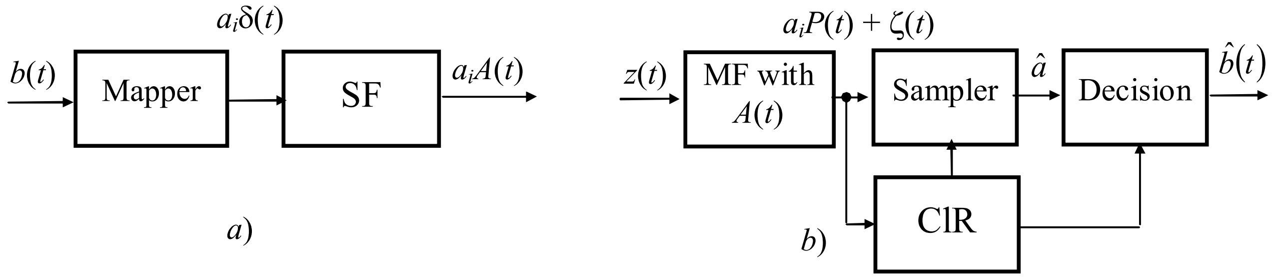


Figure 73 – Schemes: *a* – *M*PAM signals modulator; *b* – *M*PAM signals demodulator

Pulse on the MF output $P(t)$ must satisfy the absence of ISI, so we require that the spectrum $S_P(f)$ will be a Nyquist spectrum $N(f)$.

$$S_P(f) = N(f). \tag{5.46}$$

We use the property of MF: its amplitude response coincides with the amplitude spectrum of the signal, matched with (if $c = 1$)

$$H_{MF}(f) = S_A(f). \tag{5.47}$$

Given the equality (5.44) – (5.47) we conclude that

$$H_{SF}(f) = H_{MF}(f) = \sqrt{N(f)}. \tag{5.48}$$

As they say the AR of MF and SF is described by the "square root of the Nyquist spectrum" dependence.

Usually Nyquist spectrum is described by the "raised cosine" dependence.

$$N(f) = \begin{cases} T, & 0 \le |f| \le (1-\alpha) f_N, \\ 0{,}5T\left[1+\sin\left(\dfrac{\pi}{2\alpha}\left(1-\dfrac{|f|}{f_N}\right)\right)\right], & (1-\alpha) f_N < |f| < (1+\alpha) f_N, \\ 0, & |f| \ge (1+\alpha) f_N, \end{cases} \qquad (5.49)$$

where T – clock interval;

$f_N = 1/(2T)$ – Nyquist frequency;

α – roll-off factor.

Dependence "square root of the Nyquist spectrum" is described as

$$\sqrt{N(f)} = \begin{cases} \sqrt{T}, & |f| \le (1-\alpha) f_N, \\ \sqrt{T}\sin\left[\dfrac{\pi}{4\alpha}\left(1+\alpha-\dfrac{|f|}{f_N}\right)\right], & (1-\alpha) f_N < |f| < (1+\alpha) f_N, \\ 0, & |f| \ge (1+\alpha) f_{N.} \end{cases} \qquad (5.50)$$

Fig. 74 shows the dependences $N(f)$ and $\sqrt{N(f)}$ for α = 0,4. Fig. 74,*b* shows that the shaping and matched filters are a lowpass filters, but with a special response. If as MF and SF filters of Butterworth, Chebyshev, etc. are used, synthesized with a view to bringing them as close to the AR of rectangular, it will not be fulfilled the condition of absence of ISI.

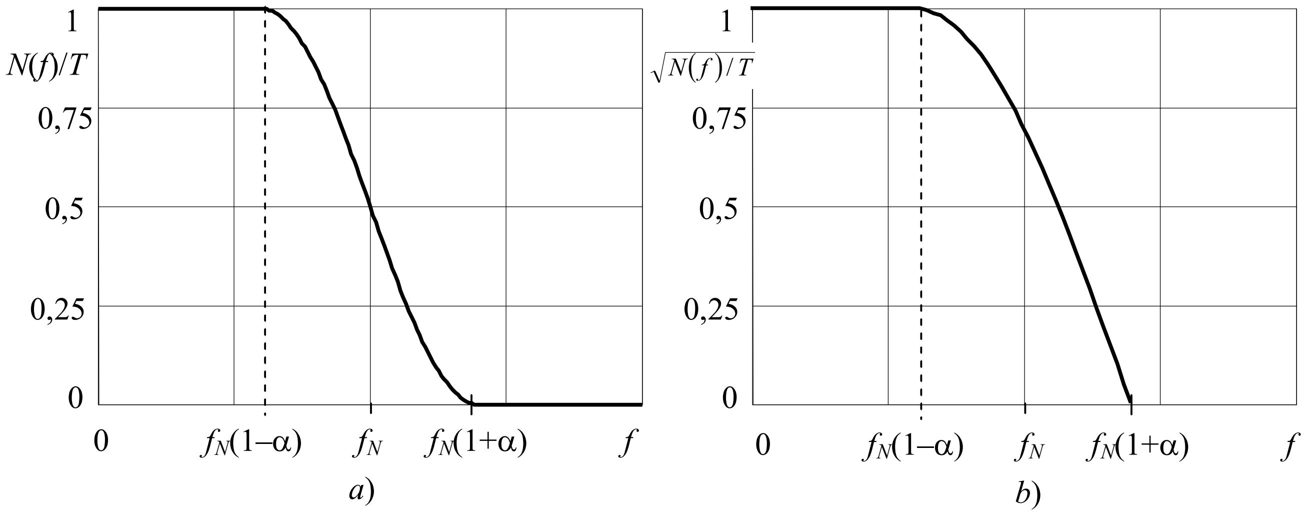


Figure 74 – Spectra: *a* – Nyquist spectrum, b – root of Nyquist spectrum

The expression for the pulse $A(t)$ can be obtained as the inverse Fourier transform of dependence $\sqrt{N(f)}$, assuming that the phase spectrum is identically equal to zero:

$$A(t)=\frac{\pi}{8\alpha+2\pi(1-\alpha)}\left\{\frac{8\alpha}{\pi(1-(8\alpha f_N t)^2)}\cdot\left[\cos(2\pi(1+\alpha)f_N t)+8\alpha f_N t\cdot\sin(2\pi(1-\alpha)f_N t)\right]+\right.$$
$$\left.+2(1-\alpha)\frac{\sin(2\pi(1-\alpha)f_N t)}{2\pi(1-\alpha)f_N t},\qquad -\infty<t<\infty.\right. \tag{5.51}$$

$P(t)$ function can be obtained as the inverse Fourier transform of $N(f)$, assuming that the phase spectrum is identically equal to zero:

$$P(t)=\frac{\sin\,2\pi f_N t}{2\pi f_N t}\cdot\frac{\cos\,2\pi\alpha\cdot f_N t}{1-(4\alpha\cdot f_N t)^2}. \tag{5.52}$$

Fig. 75 shows graphs of pulses $A(t)$ and $P(t)$. From graphs $P(t)$ shows that its peak value is 1. This means that when the pulse $a_i A(t)$ transfer sample on the output of the sampler is a_i. Fig. 75 shows that the pulse $P(t)$ takes zero value at $t=\pm kT$ (k = 1, 2, 3,…), i.e. satisfies the sampling condition. Impulse $A(t)$ doesn't take zero values at $t=\pm kT$ (k = 1, 2, 3,…).

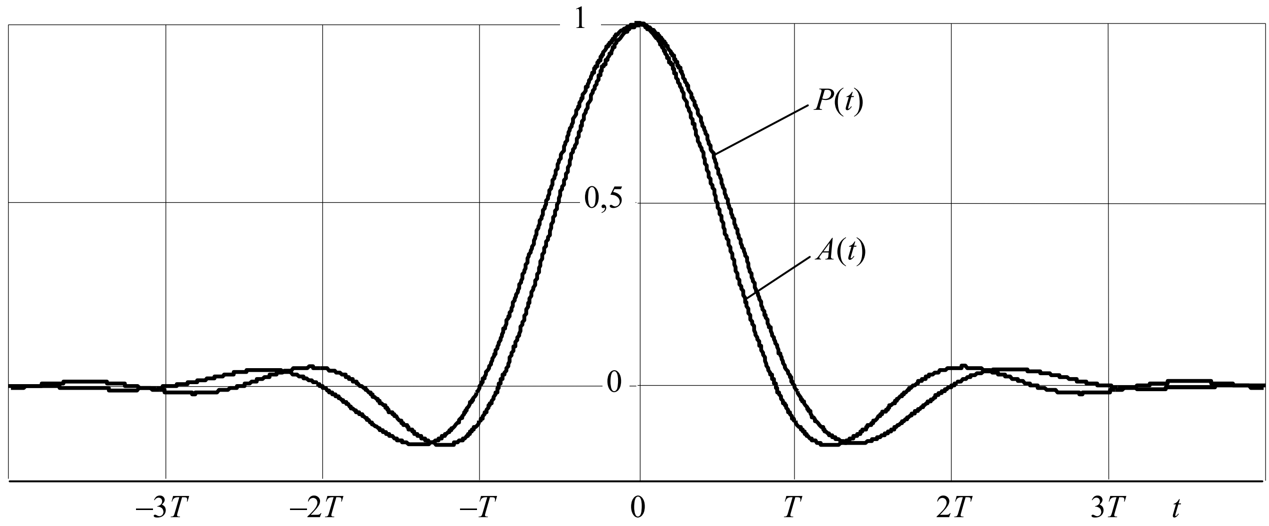


Figure 75 – $A(t)$ and $P(t)$ pulses

Define the value of the pulse $A(t)$ energy which require in further analysis,

$$E_A=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_A^2(\omega)d\omega=2\int_0^{\infty}S_A^2(f)df=2\int_0^{\infty}\left(\sqrt{N(f)}\right)^2df=2\int_0^{\infty}N(f)df=1. \tag{5.53}$$

The result was obtained, based on the fact that the integral equals the area under the curve described by the integrand (fig. 84, *a*). Since the function $N(f)$ has anti-symmetric slope, then this area is the area of a rectangle of height T and the base $f_N=1/(2T)$. This value allows you to easily determine the signal energy $s_i(t)=a_i A(t)$:

$$E_i=a_i^2. \tag{5.54}$$

For further analysis the value of noise average power is also required on the MF output, which response are described by the dependence $\sqrt{N(f)}$,

$$P_{n\,\text{out}} = 2\int_0^\infty \frac{N_0}{2} H_{MF}^2(f)df = N_0 \int_0^\infty N(f)df = \frac{N_0}{2}. \quad (5.55)$$

When integrating the same approach as in the calculation of the integral (5.53). The value of RMS noise on the MF output is equal to

$$\sigma = \sqrt{N_0/2}. \quad (5.56)$$

Given (5.54) and (5.56) is easily seen that the signal to noise ratio at the sampling moment

$$\frac{a_i}{\sigma} = \sqrt{\frac{2E_i}{N_0}} \quad (5.57)$$

corresponds to a matched filter property (5.31).

5.6 Correlator

In analyzing the properties of the matched filter has been related between the output signal $y(t)$ and input signal $z(t)$ – relation (5.38). Let $c = 1$, $t_0 = T_s$ and under these conditions, the relation (5.38) define $y(T_s)$. At the same time we take into account that the signal $s(t)$ exists on the interval $(0, T_s)$:

$$y(T_s) = \int_0^{T_s} z(t)s(t)dt. \quad (5.58)$$

Hence the last relation follows that you can make an optimal signal processing with the noise $z(t) = s(t) + n(t)$ by scheme, which work is described by the relation (5.58). The device, which work is described by the relation (5.58) is called correlator and it is considered in Section 5.3.

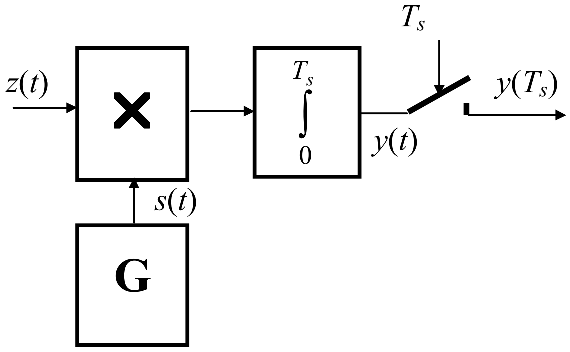


Figure 76 – Correlator

Fig. 76 shows the scheme of the correlator. It contains the signal $s(t)$ generator (a replica of the processed signal), multiplier and integrator with reset. At the end of the signal $s(t)$ sampler takes the sample, and the integrator is given in the zero state to be ready to process the next signal.

From the description of the scheme it is understandable that have not been shown circuit of synchronization signal generator $s(t)$, the control circuit of the sampler and the integrator reset.

Equivalence of signal processing with the noise by matched filter and correlator in both cases is

$$\rho_{\text{out}} = \frac{2E_s}{N_0}. \quad (5.59)$$

However, the processes taking place in the schemes of matched filter and correlator are different. Illustrate this for an example of the radio pulse $s(t)$ (fig. 77,a).

The output of matched filter response is observed in the form of the correlation function of the signal $s(t)$ (fig. 77,*b*). From the output of the filter sample $y_s(T_s)$ is taken. Fig. 77,*c* shows the signal at the output of the integrator with the reset of the correlator $y_s(t)$ and the sample value from the output of the correlator $y_s(T_s)$. Despite the fact that the processes are different, the outputs of both circuits of the signal/noise ratio equal

$$\frac{y(T_s)}{\sigma} = \sqrt{\frac{2E_s}{N_0}}, \tag{5.60}$$

where σ is RMS of sample $y(t)$.

Relationships (5.59) and (5.60) are valid for the correlator, if the signal processing time by correlator equals to the duration of the signal. Situation is different in the processing of pulses with a substantially limited range, for example, Nyquist pulses. In this case, the pulse duration can take values $T_s = (8\ldots20)T$, where T – clock interval duration. Processing time by correlator cannot be more than T at digital modulation signal demodulation.

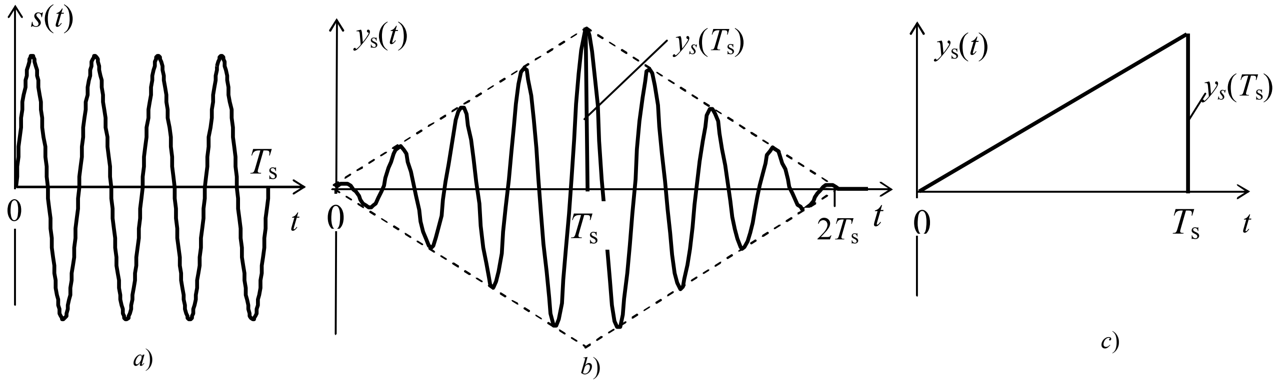


Figure 77 – Signals: *a* – on the processing device input; *b* – on the matched filter output; *c* – on the integrator output of the correlator

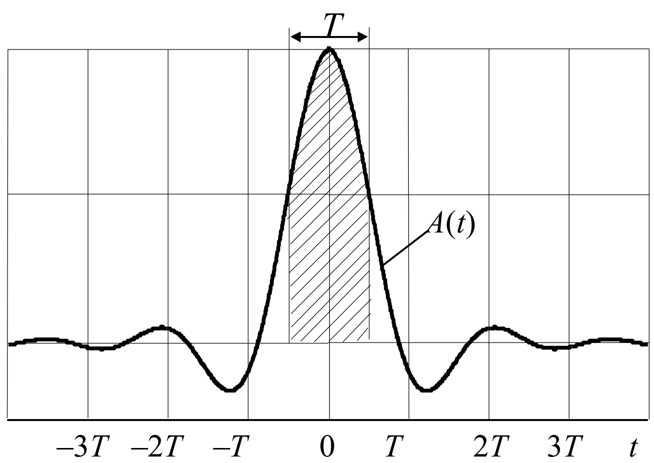


Figure 78 – Processing of the pulse by correlator

In fig. 78 pulse $A(t)$ is shown. The shaded area under the curve $A(t)$ on the interval $(-T/2,\ T/2)$ shows the result of integrating by correlator – in the processing of the pulse are not using the pulse value at intervals $(-\infty,\ -T/2)$ and $(T/2,\ \infty)$. Therefore, signal/noise ratio at the correlator output is less than the processing of a matched filter. Processing time of matched filter is equal to the duration of its impulse response, and it is equal to the duration of the signal T_s, and all the signal values used in processing. For this reason, the correlator was used in the digital modulation demodulator using weakly filtered rectangular pulses, when the duration of the signal is practically equal to the duration of the clock interval. Matched filters are applied when pulses with a substantially limited range are used.

5.7 Matched filtering of radio pulses

In the case of bandpass signals of digital modulation (*M*APSK, *M*PSK, *M*QAM etc.) channel symbols are based on pulse-carrier

$$s(t) = \sqrt{2}A(t)\cos 2\pi f_0 t\,. \tag{5.61}$$

Signal $s(t)$ is bandpass. Its spectrum is centered round the frequency f_0. We must perform an optimal filtering of the signal $s(t)$, coming together with the noise $n(t)$, and take sample. We suppose that the noise is white noise, its spectrum is concentrated in the bandwidth of the communication channel.

First method. Used bandpass filter which amplitude response is described by the relation (5.35). Engineering practice has shown that the bandpass filters have a low accuracy of implementation.

Second method. First is the coherent (synchronous) detection of the signal and noise sum $z(t) = s(t) + n(t)$, and then pulse filtering $A(t)$ with noise by low-pass matched filter (fig. 79). Carrier recovery (CR) scheme produces an oscillation $\sqrt{2}\cos 2\pi f_0 t$, which is necessary for the detector.

For the analysis of noise transmission through the synchronous detector we represent bandpass noise by quadrature components

$$n(t) = N_c(t)\sqrt{2}\cos 2\pi f_0 t + N_s(t)\sqrt{2}\sin 2\pi f_0 t, \tag{5.62}$$

where $N_c(t)$ is amplitude of the cosine component of the noise;

$N_s(t)$ is amplitude of the sine component of the noise.

The output multiplier we obtain

$$u_{\text{mult}}(t) = [A(t) + N_c(t)] + [A(t) + N_c(t)]\cos 2\pi 2 f_0 t. \tag{5.63}$$

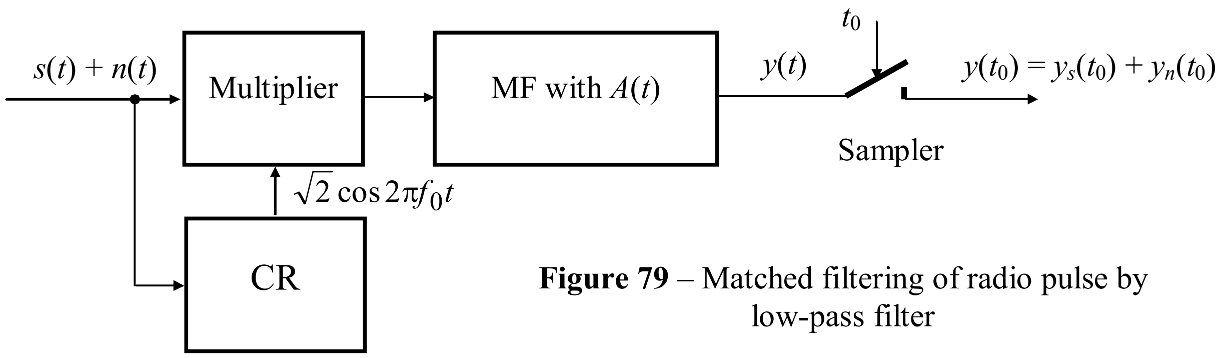


Figure 79 – Matched filtering of radio pulse by low-pass filter

The first term in (5.63) is low-frequency component, and the second term is DSB-SC signal with a carrier frequency $2f_0$. In transmission systems, carrier frequency is substantially greater than the maximum frequency of the signal spectrum $A(t)$, and the spectra of two terms in (5.63) do not overlap. For signal $A(t)$ after the multiplier one should include low-pass filter, transmitting $A(t)$ and weakening

$A(t)\cos 2\pi 2f_0t$. I.e. requires a low-pass filter with a cut off frequency is higher than the maximum frequency in the spectrum of the signal $A(t)$. The same filter after the multiplier (in fact LPF) performs optimal filtering of the signal $A(t)$. On the MF output due to the pulse $A(t)$ we obtain the pulse $P(t)$ (see section 5.5).

Let us discuss the passing of noise through the synchronous detector and matched filter. Noise $n(t)$ is white in the bandwidth of the communication channel F_{ch} with average frequency f_0. Its average power is determined $P_n = N_0F_{\text{ch}}$. This power is divided equally between the cosine and sine components. Thus, the power of cosine component

$$\overline{\left(N_c(t)\sqrt{2}\cos 2\pi f_0t\right)^2} = P_n/2. \tag{5.64}$$

Performing the averaging over the left-hand side of equation (5.64), we obtain

$$\overline{N_c^2(t)} = P_n/2. \tag{5.65}$$

Noise $N_c(t)$ is white in the frequency band $(0,\ F_{\text{ch}}/2)$. Its power density is

$$\frac{P_n/2}{F_{\text{ch}}/2} = \frac{P_n}{F_{\text{ch}}} = N_0. \tag{5.66}$$

An analysis of transformations of the signal and noise synchronous detector can be seen that the input filter, matched with $A(t)$, signal $A(t)$ and white noise with power density N_0 act. This filtering is considered in the previous section. MF provides the ratio of the instantaneous signal power to average noise power

$$\rho_{\text{peak}} = \frac{2E_A}{N_0}. \tag{5.67}$$

Define the signal energy $s(t)$

$$E_s = \int_{-\infty}^{\infty} s^2(t)dt = \int_{-\infty}^{\infty} \left(A(t)\sqrt{2}\cos 2\pi f_0t\right)^2 dt = \int_{-\infty}^{\infty} A^2(t)dt = E_A. \tag{5.68}$$

Thus, the circuit shown in fig. 79, provides the signal/noise ratio

$$\rho_{\text{peak}} = \frac{2E_s}{N_0}, \tag{5.69}$$

then there is a filter, matched with the bandpass signal $s(t)$.

This is precisely the scheme used in the demodulator of two-dimensional digital modulation signals for optimum signal filtering, as the inclusion of two such schemes with the reference oscillations $\cos 2\pi f_0t$ and $\sin 2\pi f_0t$ a division of cosine and sine of radio pulse and their subsequent separate processing performed.

5.8 Optimal demodulator for bandpass signals

Let's consider the one-dimensional *M*ASK and BPSK signals. Channel symbols are described as

$$s_i(t) = a_i\sqrt{2}A(t)\cos 2\pi f_0t, \quad i = 0, 1, ..., M-1, \tag{5.70}$$

where $\sqrt{2}A(t)\cos 2\pi f_0 t$ is radio pulse, with time and spectral characteristics, whose energy is equal to 1, the maximum value is equal to $\sqrt{2}$;

a_i – coefficients that reflect n = $\log_2 M$ transmitted bits in accordance with mapping code.

Above it was found that the main elements of the optimal demodulator are: a matched filter, sampler, and decision. We have said that the optimal filtering of radio pulse should perform circuit containing a synchronous detector and low-frequency matched filter. To synchronous detector it is necessary to use oscillation $\sqrt{2}\cos 2\pi f_0 t$ that generates by a carrier recovery scheme (CR). Coming from the clock recovery scheme (ClR) impulses controls the work of the sampler. Therefore, the demodulator circuit has the form shown in fig. 80.

Rule making decisions formulated on the basis of partitioning the space of signals into M disjoint areas s_i, $i = 0, \ldots, M - 1$; each area s_i is a set of points that are closer to the signal $s_i(t)$, rather than to other signals. Decision scheme gives solutions bits in accordance with the mapping code (fig. 81).

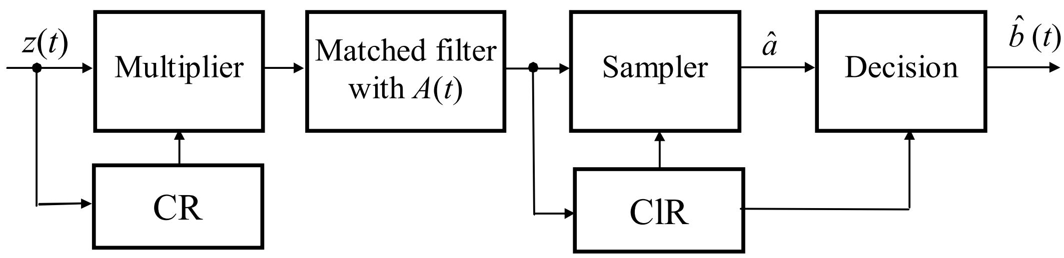


Figure 80 – Optimal demodulator of *M*ASK and BPSK signals

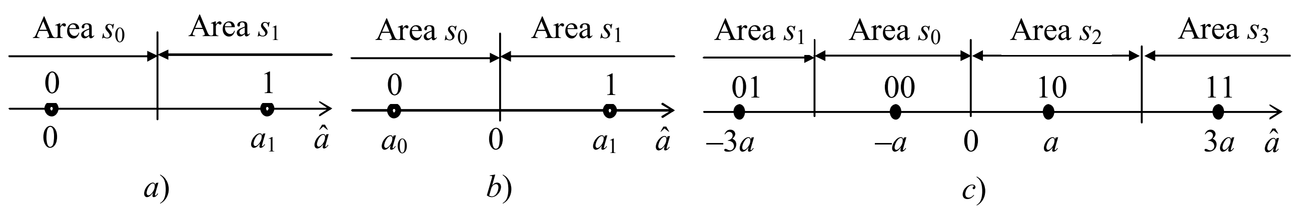


Figure 81 – Partitioning of signals space on the signals areas:
a – BASK; *b* – BPSK; *c*– QASK

In the case of two-dimensional signals *M*PSK ($M \geq 4$), *M*APSK, *M*QAM channel symbols are described as

$$s_i(t) = a_{ci}\sqrt{2}A(t)\cos 2\pi f_0 t + a_{si}\sqrt{2}A(t)\sin 2\pi f_0 t, \quad i = 0,\ 1, \ldots,\ M - 1, \qquad (5.71)$$

where a_{ci} and a_{si} – pair of coefficients that reflect n = $\log_2 M$ transmitted bits in accordance with mapping code.

Channel symbols consist of cosine and sinus impulses. They should be divided. This is done by two synchronous detectors, supporting different variations – if the reference oscillation $\sqrt{2}\cos 2\pi f_0 t$, that the detector does not respond to the sinus component of the input signal, if the reference oscillation $\sqrt{2}\sin 2\pi f_0 t$, that the detector does not respond to the cosine component of the input signal. Therefore, two detectors divide cosine and sine pulses and their subsequent separate processing in the two subchannel demodulator (fig. 82): matched filtering and sampling. Estimates coefficients $\hat{a}_c$ and $\hat{a}_s$ act at the decision scheme.

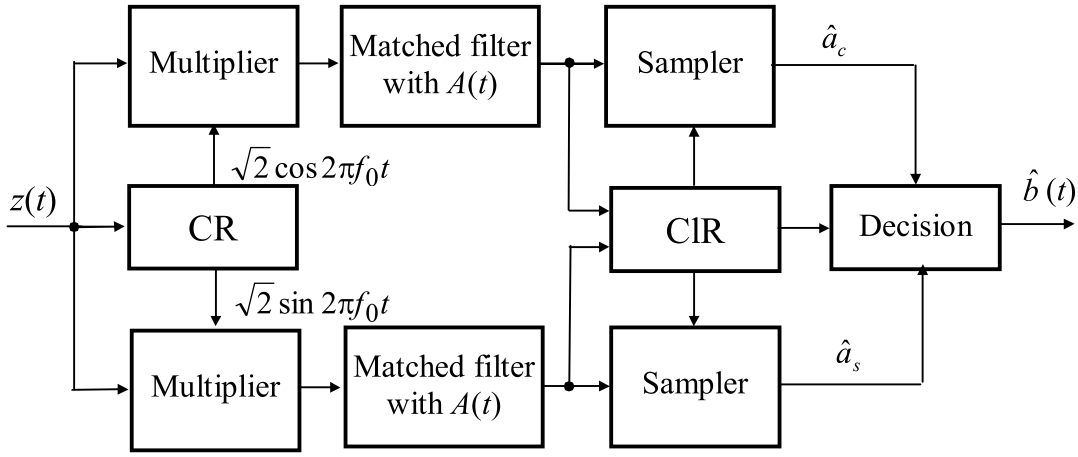


Figure 82 – Optimal demodulator of *M*PSK ($M \geq 4$), *M*APSK, *M*QAM signals

Just as in the case of one-dimensional signals, generally make decisions formulated on the basis of partitioning the space of signals into M disjoint areas s_i, $i = 0, \ldots, M - 1$; each area s_i is a set of points that are closer to the signal $s_i(t)$, rather than to other signals. The difference lies in the fact that the space is two-dimensional. If the point $(\hat{a}_c, \hat{a}_s)$ appears in the region signal s_i, then decides that the transmitted signal s_i. Decision scheme gives out solutions bits in accordance with the mapping code.

Fig. 83 shows a partition of signal space on the signals areas. Boundaries of the areas shown in heavy lines. In the case of QPSK signal the signals areas are 4 quadrant, in the case of 8PSK signal the signals areas are 8 sectors, in the case of 16QAM signal the signals area are formed by vertical and horizontal lines. Due to the simplicity of two-dimensional *M*QAM signal space partitioning on the signals areas they have received the widest distributed among the *M*APSK signals.

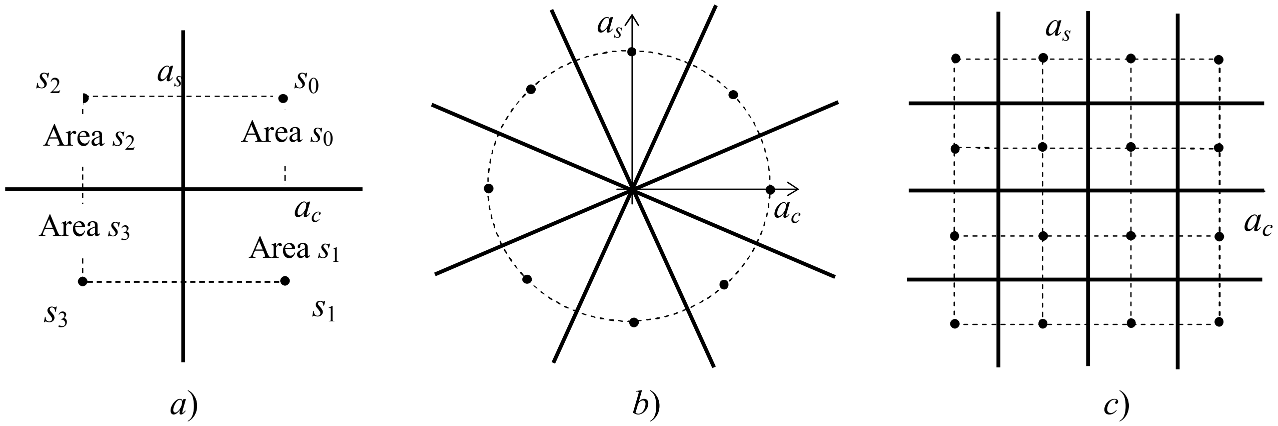


Figure 83 – Partitioning of signals space on the signals areas:
a – QPSK; *b* – 8PSK; *c* – 16QAM

BFSK signals referred to two-dimensional signals also. Channel symbols are described as

$$s_0(t) = a\sqrt{2}A(t)\cos\left(2\pi\left(f_0 - \Delta f/2\right)\right)t, \quad s_1(t) = a\sqrt{2}A(t)\cos\left(2\pi\left(f_0 + \Delta f/2\right)\right)t, \quad (5.72)$$

where Δf – frequency spacing.

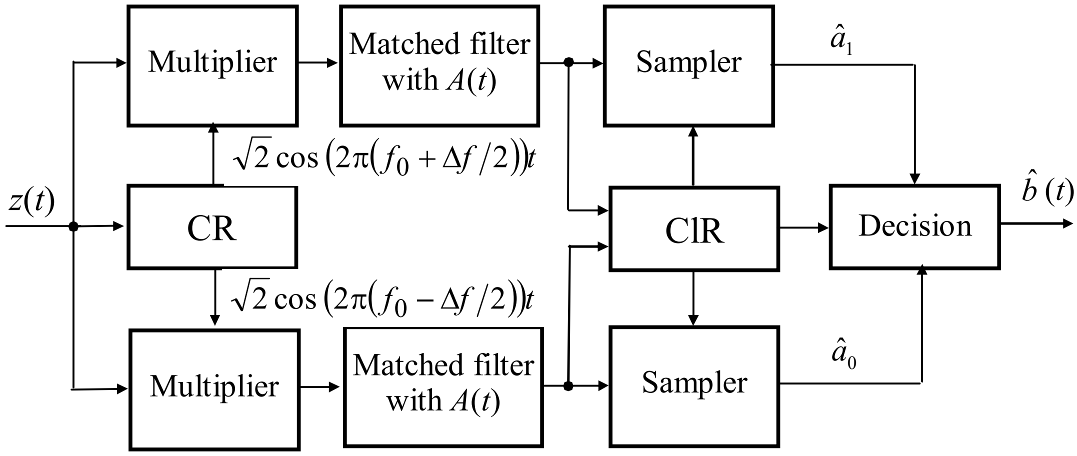


Figure 84 – Optimal demodulator of BFSK signal

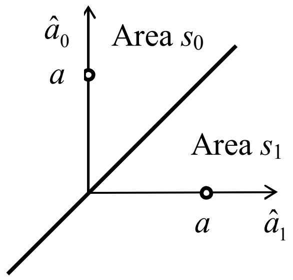


Figure 85 – Partitioning of BFSK signals space on the signals areas

If frequency spacing satisfies $\Delta f = k/(2T)$, where k = 1, 2, …, the signals are orthogonal, and they can be separated by a synchronous detector (similar to the separation of cosine and sine pulses). Scheme of optimal demodulator of BFSK signal is similar to the scheme shown in fig. 83, but differs in supporting synchronous oscillations in the detectors (fig. 84).

Fig. 85 shows a partition of the BFSK signal space on signals areas. A heavy line shows the border area. This figure shows that it is enough to compare estimates of the signal amplitudes $s_0(t)$ and $s_1(t)$ for a decision on the transmitted signal: if $\hat{a}_1 > \hat{a}_0$ demodulator decides $\hat{s}_1$ and on the contrary.

5.9 The error probability at the optimum demodulation

Bit error probability p is a quantitative measure of noise immunity of digital transmission systems. Analysis of bit error probability will start from the consideration of one-dimensional signals (MPAM, MASK, BPSK) – see fig. 64,*b* and relations (5.15)–(5.18).

The conditional probability density $p(\hat{a}/s_0)$ has a normal probability distribution with average equals to a_0. With this in mind (5.16) we can write

$$P_{er}(s_0)=\int_{\lambda}^{\infty} p(\hat{a}/s_0)d\hat{a}=Q\left(\frac{\lambda-a_0}{\sigma_\zeta}\right), \tag{5.73}$$

where σ_ζ – noise RMS on the matched filter output, defined earlier by the ratio (5.56).

We take into consideration the distance between signals

$$d=(a_1-a_0). \tag{5.74}$$

Express the threshold λ through distance d (fig. 64):

$$\lambda = a_0 + 0{,}5d. \tag{5.75}$$

In view of (5.56) and (5.75) ratio (5.73) can be rewritten

$$P_{\text{er}}(2) = Q\left(\frac{d}{\sqrt{2N_0}}\right), \tag{5.76}$$

where $P_{\text{er}}(2)$ – signal error probability in the binary transmission system.

The definition of Gaussian Q-function $Q(z) = \frac{1}{\sqrt{2\pi}}\int_z^{\infty} e^{-t^2/2}dt$ follows that for the larger value of the argument is the lower value of the function $Q(z)$. Signal error probability (5.76) will be decreased with distance d between signals increasing and the noise power density at the demodulator input N_0 decreasing.

It's easy (5.73) stated in terms of errors, when the line connecting the signal point is parallel to the axis a_s or the axis a_c. If the line connecting the signal point is not parallel to anyone of the axes, then the condition of the error is stated differently. But we can show that in such cases, the probability of error signal in the binary transmission system P_{er} (2) is defined by (5.76).

The task is to express the distance d through the physical parameters of the signal at the demodulator input – the average signal power P_s and the digital signal transmitted rate R. This problem is solved in section 4. To compare the noise immunity of different types of modulation signals is conveniently expressed in terms of distance E_b – average value of signal energy to transfer one bit.

For binary transmission systems $p = P_{\text{er}}$. Formulas are given in table 3.

Table 3 – Bit error probability in binary transmission systems

Modulation method	BPAM	BASK	BPSK	BFSK
Distance	$2\sqrt{E_b}$	$\sqrt{2E_b}$	$2\sqrt{E_b}$	$\sqrt{2E_b}$
Bit error probability	$p = Q(\sqrt{2}h_b)$	$p = Q(h_b)$	$p = Q(\sqrt{2}h_b)$	$p = Q(h_b)$

In a table $h_b^2 = E_b/N_0$ is signal/noise ratio;

$E_b = P_s \cdot T_b = P_s/R$.

We pass on now to multilevel transmission systems, i.e. $M > 2$. If the channel symbols are equiprobable, the signal error probability is determined

$$P_{\text{er}} = \frac{1}{M}\sum_{i=0}^{M-1}\sum_{\substack{j=0\\ j\neq i}}^{M-1} P_{\text{er}}(s_i, s_j), \tag{5.77}$$

where $P_{\text{er}}(s_i,\ s_j)$ is error probability in binary transmission systems, using signals s_i and s_j, error appears when the demodulator makes a decision on s_j, if s_i was transferred.

To simplify the calculations take into account only the transitions in the nearest signals. Table 4 shows the formulas of error probability for certain multilevel modulation types when mapping Gray code is used.

Table 4 – Bit error probability in multilevel transmission systems

Type of modulation	QPAM	QPSK	8PSK	16QAM
Bit error probability	$p = Q\left(\sqrt{2/5}\, h_b\right)$	$p = Q\left(\sqrt{2} h_b\right)$	$p = \frac{2}{3} Q(0{,}94 h_b)$	$p = Q(0{,}89 h_b)$

5.10 Carrier recovery system

Carrier recovery system (CR) of bandpass digital modulation signals demodulators is intended for the formation of the reference harmonic oscillation whose phase coincides with the phase of the carrier on which demodulated signal is formed.

Already in the 30-th years of the last century became clear that the BPSK signals have highest noise immunity. It was necessary to solve the problem of carrier recovery in demodulator for application of these signals in the transmission systems. The reference oscillation is necessary for the synchronous detector (fig. 86). Let on the detector input BPSK signal acts. Channel symbol is described as

$$s_i(t) = a_i \sqrt{2} A(t) \cos 2\pi f_0 t. \tag{5.78}$$

If the oscillations phase from the generator

$$u_{\text{ref}}(t) = \sqrt{2} \cos(2\pi f_0 t + \Delta\varphi). \tag{5.79}$$

differs from the carrier phase of the input signal by a value Δφ, the signal at the output of the synchronous detector receives the factor cosΔφ:

$$u_{\text{out}}(t) = a_i A(t) \cos \Delta\varphi. \tag{5.80}$$

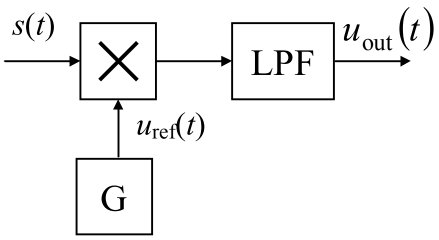


Figure 86 – Synchronous detector

As the maximum value of the cosine is one and is achieved only in the case Δφ = 0, presence of the phase difference leads to a decrease in the level of the signal at the detector output. If Δφ = π/2, the signal at the detector output is absent: $u_{\text{out}}(t) \equiv 0$. At Δφ=π $u_{\text{out}}(t) = -a_i A(t)$, i.e. changes its sign. The task is Δφ = 0.

First the carrier recovery scheme with frequency multiplication on 2 has been proposed (fig. 87).

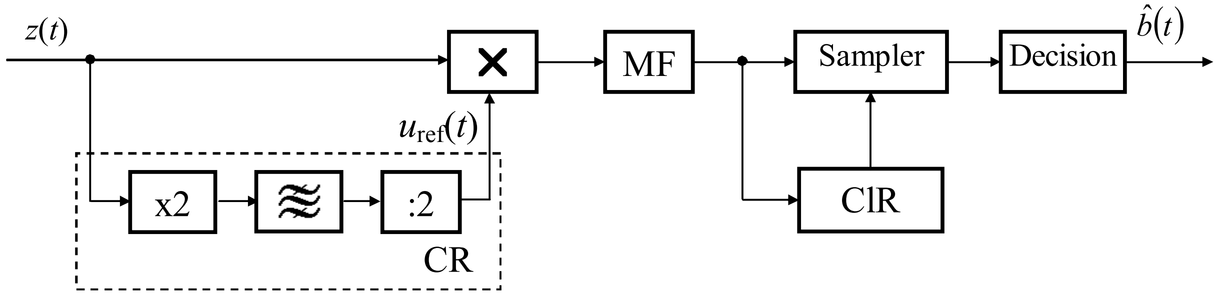


Figure 87 – Carrier recovery scheme with frequency multiplication on 2

BPSK signal constellation is given in fig. 81,*b*. Channel symbols are described by the formula (5.78):

$$s_1(t) = aA(t)\cos 2\pi f_0 t; \quad s_0(t) = aA(t)\cos(2\pi f_0 t + \pi). \tag{5.81}$$

For many decades "weak" filtered pulses $A(t)$ are used, they were similar in shape to the rectangular pulse at the interval duration T

$$A(t) = \begin{cases} 1, & 0 \le t \le T, \\ 0, & t < 0, t > T. \end{cases} \tag{5.82}$$

After multiplying the frequency by 2, the signal $s_1(t)$ and signal $s_0(t)$ give $a^2\cos 2\pi 2 f_0 t$. Narrow band filter has a centre frequency $2f_0$. It is designed to reduce noise. Frequency divider 2 can issue one of two possible reference oscillations:

– case 1: $u_{\text{ref1}}(t) = \cos 2\pi f_0 t$;
– case 2: $u_{\text{ref2}}(t) = \cos(2\pi f_0 t + \pi)$.

Both variants are possible, because the result depends on what the initial conditions arise in the scheme of the divisor. It is said that the reference oscillation has a phase uncertainty of the second order.

In case 1, the algorithm of optimal BPSK signal demodulation is implemented. In case 2, on the multiplier output, and then on the matched filter and the sampler will be tensions, opposite to those which occur in case 1. The scheme will make the inverse decision: instead of 1 will return 0 and vice versa. This phenomenon is called the inverse (opposite) work of demodulator. It turned out that in the demodulation process can occur random cuts from fluctuation $u_{\text{ref1}}(t)$ to oscillation $u_{\text{ref2}}(t)$ and vice versa.

In the QPSK signal demodulator must use a frequency multiplier on 4, a filter with a centre frequency $4f_0$ and the frequency divider on 4. After the frequency divider arises one of the reference oscillations, which differ in phase with the step 90°. There is phase uncertainty of the reference oscillation of fourth order.

Differential coding can be used for elimination of phase uncertainty of the reference fluctuations in the demodulator. Such transfer methods are called differential encoded phase modulation (see section 5.11).

Examined system with CR with exponentiation works well when the pulse amplitude $A(t)$ is close to a rectangular shape. Now, Nyquist pulses are used (pulses with substantially smoothed form $A(t)$). In this pulse form the CR system with exponentiation works bad.

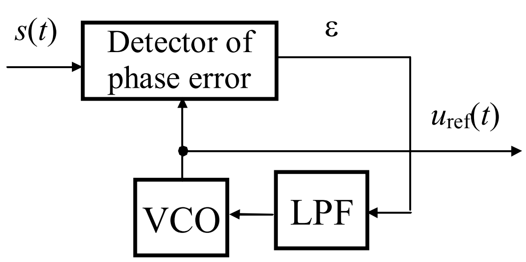


Figure 88 – PLL scheme

Now CR system is a system of phase locked-loop (PLL) (fig. 88) with a special detector of phase error, which is able to operate in the absence of a carrier in the signal spectrum. Here VCO is voltage-controlled oscillator. In the presence of phase error voltage ε adjusts the frequency and phase oscillations, which is produced by VCO, so as to reduce the amount of phase error.

Let's consider the construction of detector of phase error in the case of BPSK signal. Scheme of the detector contains one additional synchronous detector, the reference oscillation that is $\sqrt{2}\sin(2\pi f_0 t + \Delta\varphi)$. Recall that the work of the synchronous detector can be regarded as the calculation of the projection of $s(t)$ on $u_{\text{ref}}(t)$. Two synchronous detectors differ by supporting oscillations shifted in phase by 90°. Therefore, the voltages received from the outputs of synchronous detectors are the quadrature components of the detected signal.

Fig. 89 shows the signal constellation of demodulated BPSK signal and the calculated quadrature components in the sample time provided that the channel symbol with the amplitude a is demodulated: I – in-phase component, Q – quadrature component. In fig. 89,*a* phase error of the reference oscillation $\Delta\varphi = 0$; and synchronous detector calculates $I = a$, $Q = 0$. In fig. 89,*b* phase error of the reference oscillation $\Delta\varphi > 0$; and synchronous detector calculates $I = a\cdot\cos\Delta\varphi$, $Q < 0$. In fig. 89,*c* phase error of the reference oscillation $\Delta\varphi < 0$; and synchronous detector calculates $I = a\cdot\cos\Delta\varphi$, $Q > 0$.

We see that the sign for Q corresponds to the error phase: namely, if $Q < 0$, then $\Delta\varphi > 0$ and it is necessary to decrease the frequency and phase of the VCO, if the $Q > 0$, then $\Delta\varphi < 0$ and it is necessary to increase the frequency and phase of the VCO. Thus, the value of Q can be taken as the phase error ε. But the situation with the Q sign is opposite if channel symbol with the amplitude $-a$ is demodulated.

Costas suggested calculate a phase error of reference oscillation in BPSK signal demodulators

$$\varepsilon = Q \cdot \text{sign}(I). \tag{5.83}$$

Fig. 90 shows the scheme of BPSK signal demodulator with an open circuit of the carrier recovery.

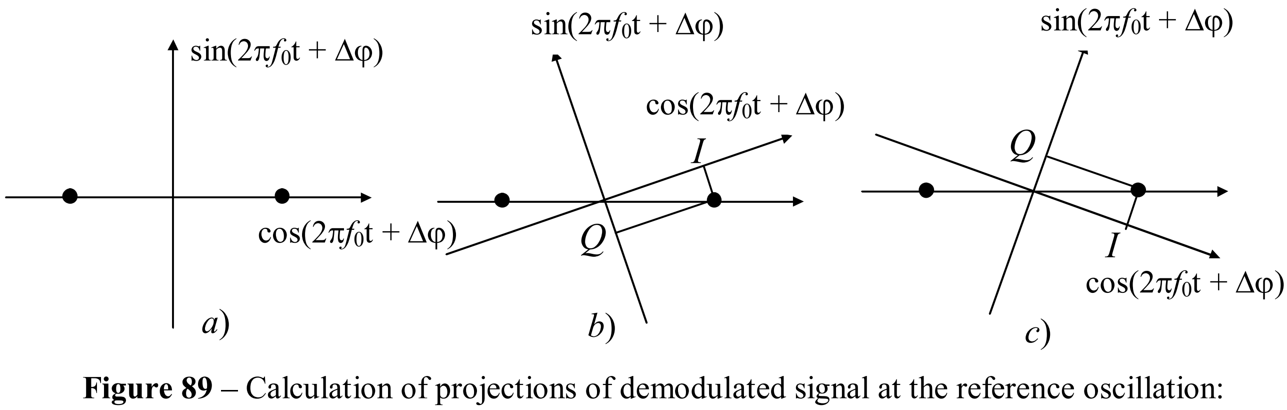


Figure 89 – Calculation of projections of demodulated signal at the reference oscillation: a – $\Delta\varphi = 0$; b – $\Delta\varphi > 0$; c – $\Delta\varphi < 0$

To construct the CR system of QPSK signal demodulator phase error detector is used, which is calculated by the Costas algorithm

$$\varepsilon = Q \cdot \text{sign}(I) - I \cdot \text{sign}(Q). \tag{5.84}$$

Here is a sign of the phase error of the reference oscillation is inequality of quadrature components I and Q. The same algorithm for calculating the phase error is used in the MQAM signals demodulators.

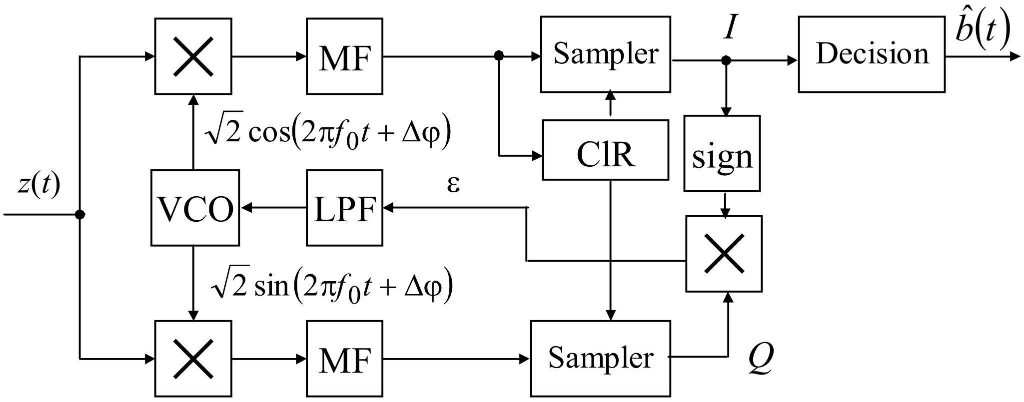


Figure 90 – BPSK signal demodulator with the CR system

5.11 Differentially encoded phase modulation

Solve the problem of inverse work of BPSK demodulator allows the transition to the difference method of transmission, in which the transmitted binary symbols (bits) appearing not in the initial phases of radio pulses (as in BPSK), but in the phase difference between adjacent radio pulses on time. The formation of modulated signal in this way is called the differentially encoded phase shift keying (DPSK).

The principle of modulation and demodulation of BDPSK signals is displayed in fig. 91. BDPSK modulator signal consists of the differential coder (DC) and the BPSK signal modulator, and the demodulator BDPSK signal consists of BPSK signal demodulator and differential decoder (DD).

Differential coder of BDPSK modem works on rule

$$b_k^{\mathrm{d}} = b_k \oplus b_{k-1}^{\mathrm{d}}, \tag{5.85}$$

where b_k – bit 1 or 0 on coder input on k-th clock interval;

b_k^{d} – symbol 1 or 0 on coder output on k-th clock interval;

$\oplus$ – mod 2 plus sign.

Differential decoder of BDPSK modem works on rule

$$\hat{b}_k = \hat{b}_k^{\mathrm{d}} \oplus \hat{b}_{k-1}^{\mathrm{d}}, \tag{5.86}$$

where $\hat{b}_k^{\mathrm{d}}$ – symbol 1 or 0 on differential decoder input on k-th clock interval;

$\hat{b}_k$ – bit 1 or 0 on differential decoder output on k-th clock interval.

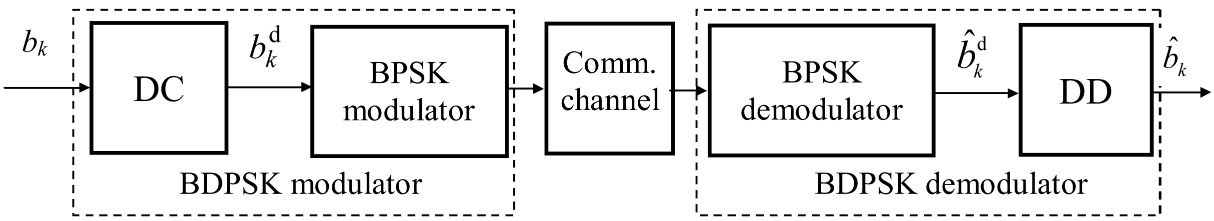


Figure 91 – The principle of BDPSK signal modulation and demodulation

The initial phase of the recovered carrying oscillation at the demodulator can coincide with the initial phase of the demodulated BPSK signal or differ from it in the angle π. In general terms, we can write that the phase of the reference signal is $p \cdot \pi$, where $p = 0$ or 1 is a value, which describe the phase shift. Assuming that there is no noise in the communication channel, the symbols in the output of BPSK demodulator will be determined by the ratio

$$\hat{b}_k^{\mathrm{d}} = b_k^{\mathrm{d}} \oplus p \tag{5.87}$$

for all k. Substituting (5.87) into formula (5.85), is easy to see that bit $\hat{b}_k$ does not depend on p.

Example of coding and decoding of an arbitrary sequence of bits is shown in table 5. Table illustrates the coding, starting with $k = 1$. As coding on the k-th interval of the pulse participates previous encoded symbol, in the second row of arbitrary decision $b_0^{\mathrm{d}} = 0$. Row 3 repeats the row 2 – demodulation without inverse work. The result of decoding is shown in row 4. Row 5 contains the inversion row 2 – demodulation inverse work. After decoding the reconstructed signal (row 6) coincides with the original signal (row 1). Thus, the transfer with a differential coding eliminates inverse work of BPSK demodulator.

Table 5 – Example of difference encoding and decoding

Row №		k						
		0	1	2	3	4	5	6
1	b_k		1	1	0	0	1	0
2	b_k^{d}	0	1	0	0	0	1	1
3	$\hat{b}_k^{\mathrm{d}}$	0	1	0	0	0	1	1
4	$\hat{b}_k^{\mathrm{d}}$		1	1	0	0	1	0
5	$\hat{b}_k^{\mathrm{d}}$	1	0	1	1	1	0	0
6	$\hat{b}_k$		1	1	0	0	1	0

In a case of M-ary phase modulation for the same reason as in the case of BPSK, there is phase uncertainty of M-th order. Similarly BDPSK information appears in the phase difference between adjacent pulses and it is called MDPSK. At $M > 2$ differential encoder and decoder work with M-ary symbols. The transition from bit of digital signal to the M-ary channel symbols takes place in the mapper.

Four elementary signals are used at QPSK signals transmission

$$s_i(t) = aA(t)\sqrt{2}\sin\left(2\pi f_0 t + q_i\frac{\pi}{2} + \frac{\pi}{4}\right), \tag{5.88}$$

where q_i are quaternary symbols that take values 0, 1, 2, 3;

$q_i\frac{\pi}{2} + \frac{\pi}{4}$ are initial phases of elementary signals that take values π/4, 3π/4, 5π/4, 7π/4.

The initial phase of recovery reference oscillation in the demodulator cannot be defined uniquely; it is determined to π/2. This is due to the symmetry of the QPSK signal constellation: demodulator does not know which of the four signals considered "zero".

To eliminate the influence of the phase uncertainty of the reference oscillations at QPSK signal demodulation, go to modulation QDPSK. The principle of modulation and demodulation of signals QDPSK is displayed in fig. 92. QDPSK signal modulator consists of a mapper, differential encoder and the QPSK signal modulator, and the QDPSK signal demodulator consists of a QPSK signal demodulator, differential decoder and the mapping decoder. In this figure, and hereafter the subscript k specifies the number of clock interval, and the shift q can take on a value 0, 1, 2 and 3.

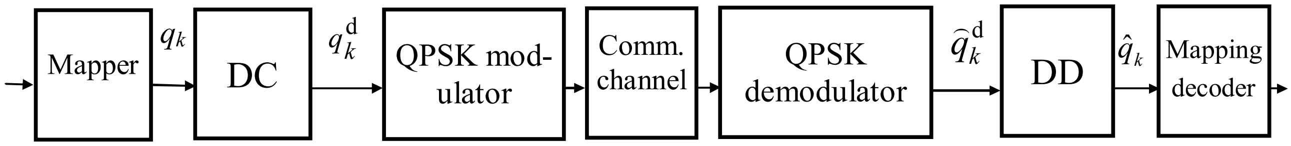


Figure 92 – Transmission system with QDPSK signal

Differential QDPSK coder implements the coding rule of quaternary symbols:

$$q_k^{\mathrm{d}} = q_k \oplus q_{k-1}^{\mathrm{d}}, \tag{5.89}$$

where $\oplus$ – mod 4 plus sign (remainder of division by 4 arithmetic sum of the parts).

Differential decoder implements the decoding rule of quaternary symbols:

$$\hat{q}_k = \hat{q}_k^{\mathrm{d}} \ominus \hat{q}_{k-1}^{\mathrm{d}}, \tag{5.90}$$

where $\ominus$ – mod 4 minus sign (remainder of division the difference by 4).

The initial phase of the recovered carrying oscillation at the demodulator can coincide with the initial phase of the demodulated QPSK signal or differ from it in the angle $p\cdot\pi/2$ (p = 0, 1, 2 or 3 is a value which describe phase shift). Assuming that there is no noise in the communication channel, the symbols in the output of QPSK demodulator will be determined by the ratio

$$\hat{q}_k^{\mathrm{d}} = q_k^{\mathrm{d}} \oplus p \tag{5.91}$$

for all k. Thus, because of the phase uncertainty of coherent oscillations in the QPSK signal demodulators all the symbols $\hat{q}_k^{\mathrm{d}}$ get increment p. Substituting (5.91) into formula (5.90), is easy to see that symbol $\hat{b}_k$ does not depend on p.

At QDPSK because of subtraction in the decoder, the value p (i.e. phase uncertainty) is excluded. As for the exclusion phase uncertainty value $\hat{q}_k$ defined as the difference between two next symbols, the encoding value q_k^{d} formed as the sum of the previous value q_{k-1}^{d} and transmitted symbol q_k.

Rules of addition on mod 4 are shown in table 6, and the rules of subtraction on mod 4 are shown in table 7.

Digital QPSK signal transmission the transition from pair of bits b_1b_2 to quaternary symbols q at each clock interval the modulation carried out according to Gray code, an example of which is shown in table 8.

Since at the QPSK signal demodulation the most likely error is its transitions to the nearest signal, then at Gray code using such transitions result in error in only one bit and it minimizes the bit error probability.

Table 6 – Addition on mod 4 ($a \oplus b$)

a	***b***			
	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

Table 7 – Subtraction on mod 4 ($a \ominus b$)

a	***b***			
	0	1	2	3
0	0	3	2	1
1	1	0	3	2
2	2	1	0	3
3	3	2	1	0

Table 8 – Mapping Gray code

Pair of bits b_1b_2	Quaternary symbol q	Initial signal phase
00	0	45°
10	1	135°
11	2	225°
01	3	315°

In table 9 an example of coding and decoding digital signal transmission by QDPSK method is shown. The transition from pair of bits to quaternary symbols is carried out according to table 8. It is assumed that the value p = 3, and $q_0^{\mathrm{d}} = 1$. From the data in table 9 that received bits coincide with the transmitted bits.

Table 9 – Example of difference encoding and decoding of quaternary symbols

Number of clock interval k	0	1	2	3	4	5	6	7	8
Sequence of transmitted bits		01	00	11	01	10	11	10	00
Sequence of symbols q_k		3	0	2	3	1	2	1	0
Sequence of symbols q_k^{d}	1	0	0	2	1	2	0	1	1
Sequence of symbols $\hat{q}_k^{\mathrm{d}}$	0	3	3	1	0	1	3	0	0
Sequence of symbols $\hat{q}_k$		3	0	2	3	1	2	1	0
Sequence of received bits		01	00	11	01	10	11	10	00

Assume that because of noise BPSK demodulator makes incorrect decision on the k-th clock interval. Each symbol that comes to the differential decoder input, at the decoding is used twice – on the k-th and (k + 1)-th clock interval. Therefore, if the demodulator decisions on (k – 1)-th and (k + 1)-th clock intervals are correct, then on the differential decoder output will be two erroneous bits. Thus, the difference decoder multiplies errors.

The bit error probability of the BDPSK and QDPSK transmission methods and for small values of error probability (p << 1) can be written

$$p_{\mathrm{BDPSK}} = p_{\mathrm{QDPSK}} = 2Q\left(\sqrt{2}h_b\right). \tag{5.92}$$

5.12 Incoherent demodulation of digital modulation signals

Until now the demodulators schemes that implements coherent demodulation are considered. Demodulation of signals is called incoherent if the demodulator does not use information about the initial phases of demodulated signals. Obviously, it is possible, at least, at the BASK and BFSK signal demodulation, in which information is built into the amplitude and frequency of pulses, respectively, but not in the initial phases of pulses.

The purpose of incoherent demodulation usage is simplification of the demodulator scheme due to absence of carrier recovery of reference oscillations. Thus, in a case of BASK signal demodulator scheme has the form in fig. 93.

The principal difference between incoherent demodulator scheme (fig. 93) and coherent is the absence of a synchronous detector, which contains a carrier recovery of reference oscillations. Envelope detector (ED) is used instead of the synchronous detector. At the analog circuit implementation the ED was carried out very simply (fig. 94).

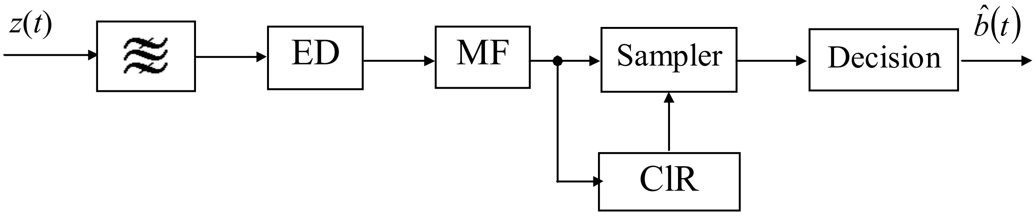


Figure 93 – Scheme of incoherent BASK signal demodulator

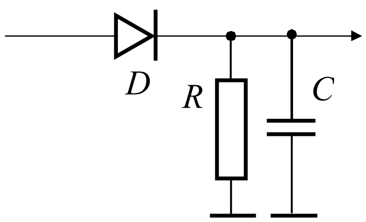


Figure 94 – Envelope detector

At the processor implementation the envelope is calculated as the square root of the sum of the quadrature components squares of the detected signal. Bandpass filter should pass to the detector signal and reduce noise at frequencies that are outside the frequency band of signal. Other elements of the scheme are the same as in the scheme of the coherent demodulator.

Scheme shown in fig. 93, can be used to the MASK signals demodulation, when $M > 2$, but with a boundary conditions on parameters: information coefficients a_i must be nonnegative, since the output voltage ED does not depend on the phase of the radio pulse. QASK signal constellation, for which the incoherent demodulation is possible, is shown in fig. 95.

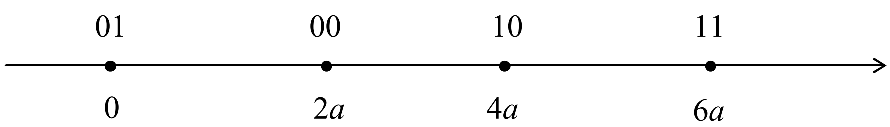


Figure 95 – QASK signal constellation

Analysis of noise immunity transmission systems with incoherent demodulation is difficult, because the noise distribution on the ED output is significantly different from normal. Definitely they can say that the signal/noise ratio at the output ED is in 2 times less than on the synchronous detector output. The final formula for the bit error probability on the demodulator output

$$p_{\text{BASK inc}} = \frac{1}{2}\exp\left(-h_b^2/2\right). \qquad (5.93)$$

Channel symbols (radio pulses) at BFSK are different in frequencies:

$$\begin{aligned} s_1(t) &= a\sqrt{2}A(t)\cos 2\pi f_1 t, \\ s_0(t) &= a\sqrt{2}A(t)\cos 2\pi f_0 t. \end{aligned} \qquad (5.94)$$

In the scheme of coherent BFSK signal demodulator a synchronous detector separates the radio pulses. In the scheme of incoherent demodulator bandpass filters are used for the separation of radio pulses (fig. 96).

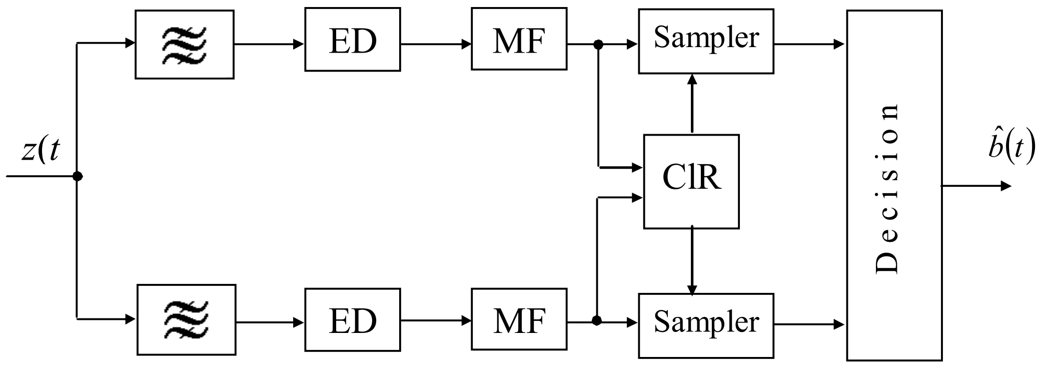


Figure 96 – Scheme of incoherent BFSK signal demodulator

Bandpass filters have average bandwidths f_1 and f_0. Scheme to the right of the envelope detector is the same as in the scheme of the coherent demodulator. The formula for the bit error probability on the BFSK demodulator output is the same as at BASK:

$$p_{\text{BFSK, inc}} = \frac{1}{2}\exp(-h_b^2/2). \qquad (5.95)$$

Incoherent demodulation of BDPSK signal is possible in the sense as defined above the concept of "incoherent demodulation" at the beginning of the subsection. At the BDPSK information lies not in the initial phases of radio pulse, but in the phase difference of adjacent in time radio pulses. So we can, without knowing the initial phases of radio pulses, perform demodulation by comparing the phases of adjacent radio pulses. BDPSK signal incoherent demodulator scheme is shown in fig. 97.

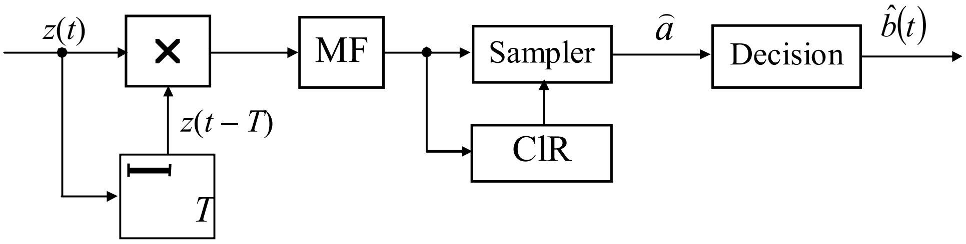


Figure 97 – Scheme of incoherent BDPSK signal demodulator

Delay circuit at the time of clock interval is used in the demodulator scheme. We can assume that the synchronous detector is made on multiplier, where the reference oscillation is a previous radio pulse. The polarity of the voltage on the multiplier output (and on the outputs of next blocks) depends on the phase difference oscillations, which are multiplied:

– case 1 – the same phases of next pulses:

$$a\sqrt{2}A(t)\cos 2\pi f_0 t \cdot a\sqrt{2}A(t-T)\cos 2\pi f_0 t = a^2A(t)A(t-T) + a^2A(t)A(t-T)\cos 2\pi 2f_0 t;$$

– case 2 – phases of next pulses differs on π:

$$-a\sqrt{2}A(t)\cos 2\pi f_0 t \cdot a\sqrt{2}A(t-T)\cos 2\pi f_0 t = -a^2A(t)A(t-T) - a^2A(t)A(t-T)\cos 2\pi 2f_0 t.$$

The first terms provide on the MF output voltage, the polarity of which depends on the same or different phase of the multiplied oscillations. From the above it is evident that:

1) rule of decision calculating:

$$\text{if } \hat{a} > 0, \text{ then } \hat{b}_i = 0,$$
$$\text{if } \hat{a} < 0, \text{ then } \hat{b}_i = 1;$$

2) differential decoder is not required.

The distribution of instantaneous values of noise on the multiplier output is no Gaussian because of presence of a component generated by the product of the noise $n(t)$ and $n(t-T)$. Analysis of error probability is difficult. The final formula for the bit error probability on the demodulator output has the form:

$$p_{\text{BDPSK, inc}} = \frac{1}{2}\exp(-h_b^2). \qquad (5.96)$$

Note that the following terminology is widely used: incoherent BDPSK signal demodulation is called the method of phase comparison reception, coherent BDPSK signal demodulation is called the method of polarities comparison reception.

Let's compare the noise immunity of the coherent and incoherent binary signals demodulators. We use dependence $p = f(h_b^2)$ for comparison (fig. 98).

The figure shows that passing from incoherent to coherent demodulation at keeping the error probability it is necessary to increase the signal/noise ratio by about 1 dB.

Incoherent demodulation is widely used for many decades, when the hardware was carried out on the vacuum tube and transistor. Now the incoherent demodulation is used, if the communication channel is rapidly changing the initial phase of the signal, for example, radio communication between objects that are moving relatively quickly.

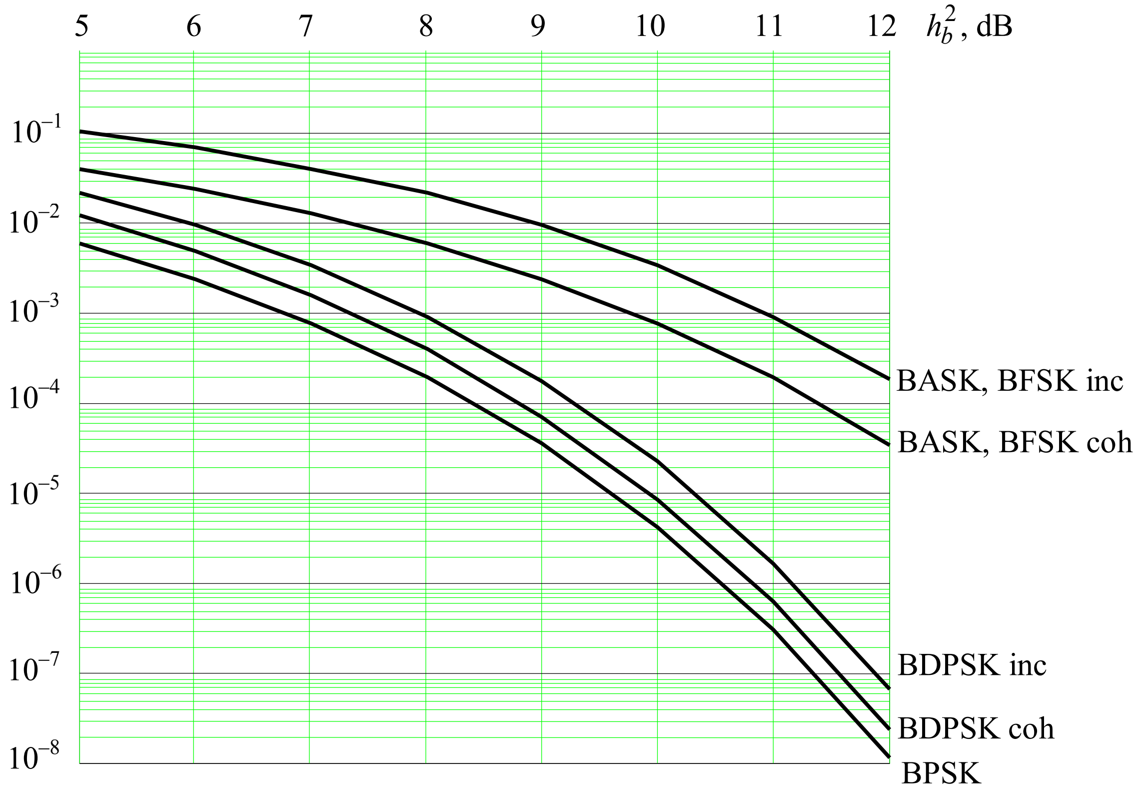


Figure 98 – Noise immunity of binary signals demodulators

5.13 Clock recovery systems

In the digital signal demodulator ClR system is designed to build sampling clock pulses, which provide the capture of signal samples at the output of matched filter at the time of maximum signal/noise ratio.

In demodulators of digital signals matched filtering of channel symbol $A(t-\tau)$ takes place (fig. 99). Delay of channel symbols τ depends on the length of the transmission line. The delay is a random (unknown) value. Under such conditions, there is a delayed on time τ pulse $P(t-\tau)$ on the MF output, maximum of which $P(0)$ is observed at unknown moment of time $t=\tau$.

ClR system contains a clock generator, which generates a short pulse $\delta(t-\tau_G)$ at every clock interval duration T, closing the key at the moment $t=\tau$. But the impulse is formed in an unknown arbitrary time $t=\tau_G$. As a result, the sample of pulse $P(t-\tau)$ not be maximal, since $\tau_G \neq \tau$, therefore $P(\tau-\tau_G) < P(0)$. In addition, there is intersymbol interference, as the samples are taken not at the times when $P(t)=0$.

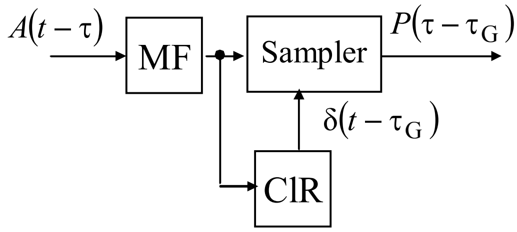


Figure 99 – Signal processing by matched filter

System ClR must ensure the correct choice of the sample moments. Modern demodulator implemented on the processors, this problem is solved by adjusting of the clock generator by error signal, formed by a special detector. That is, the basis for constructing a ClR system is a PLL system (fig. 88).

All principles of ClR systems construction are based on the properties of the digital signal to change sign. A fragment of the signal on the MF output in the symbols 1 and 0 transmission is shown in fig. 100, which implies that the change in the sign of the signal occurs approximately midway between the moments in which it is necessary to take samples.

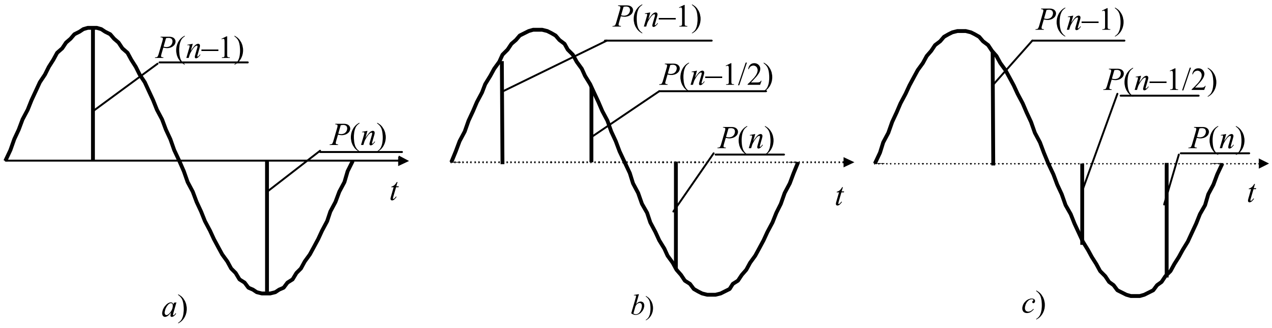


Figure 100 – Signal sampling after MF at ClR system working

Fig. 100,*a* shows the case where sample pulses are installed correctly. Samples are taken at the moments of maximum signal. This figure clearly shows that the digital signal is zero midway between the moments of taking samples. Detector of ClR system errors can be constructed based on the following principle. At each clock interval pulse is taken in the middle of the clock interval $P(n-1/2)$. If $P(n-1/2)=0$, then samples, which will be made to the decision taken in the best moments, otherwise an ClR system error signal is formed:

$$\varepsilon(n)=P(n-1/2)\{\mathrm{sign}[P(n)]-\mathrm{sign}[P(n-1)]\}, \qquad (5.97)$$

where $P(n)$ is sample, which decision is made on this the clock interval;

$P(n-1)$ is sample, which decision is made on previous the clock interval;

$\mathrm{sign}[x]=\begin{cases}+1, & x>0,\\ 0 & x=0,\\ -1 & x<0\end{cases}$ is sign determining function.

Need of multiplier $\{\mathrm{sign}[P(n)]-\mathrm{sign}[P(n-1)]\}$ in terms of formation of the error signal (5.97) due to the following: if a digital signal does not change sign, then $P(n-1/2)\neq 0$, even if the samples $P(n-1)$ and $P(n)$ taken in the best moments of samples. Therefore, in such situation it is necessary that the signal $P(n-1/2)$ is not used for error calculation. Indeed, if the signs of samples $P(n-1)$ and $P(n)$ identical, then their difference is equal to zero and sample $P(n-1/2)$ is not used.

Algorithm for error calculating (5.97) was obtained by Gardner. It corresponds to the ClR system, represented by fig. 101. On the diagram ED is error detector. The scheme takes into account that the frequency of taking samples twice the clock frequency.

An examination of the ClR scheme shows that the adjustment of the generator will be implemented only if the digital signal transitions from 1 to 0 and from 0 to 1. Therefore, the formation of the transmitted digital signal it is necessary to eliminate long sequences of identical symbols. To achieve this, the transmitted digital signal is converted into a special device – a signal scrambler with the characteristics of a random digital signal or use of a linear AMI code.

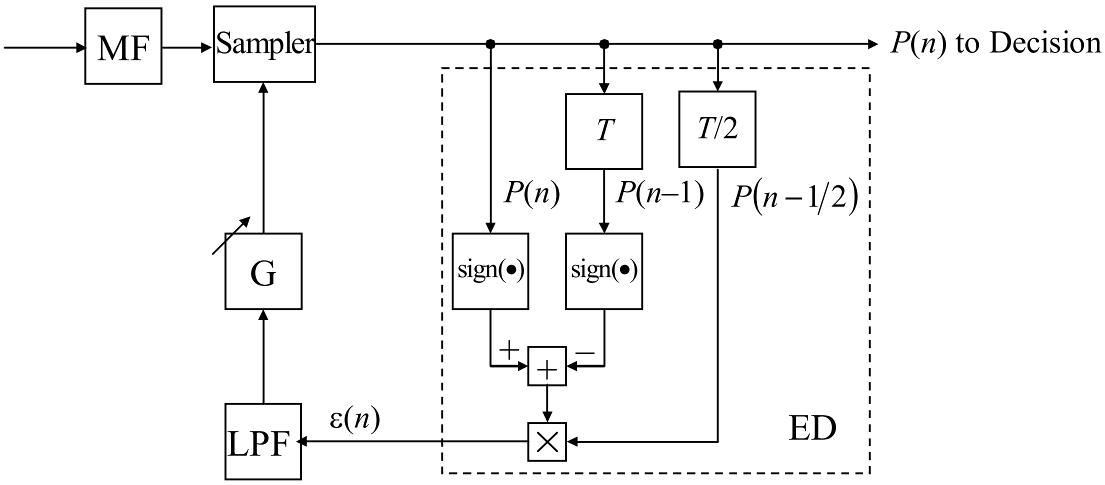


Figure 101 –CIR system with Gardner's error detector

5.14 Noise immunity in channels with variable parameters

A typical model of the radio channel in mobile communication systems and wireless access is a channel with variable parameters

$$z(t) = k(t)\cdot s(t) + n(t), \tag{5.98}$$

where $n(t)$ is additive noise, we believe, as before, that AWGN;

$k(t)$ is multiplicative noise, it plays the role of transfer coefficient of communication channel; mathematical function $k(t)$ is random function which varies slowly compared with changes in the signal $s(t)$ parameters.

They say that the relation (5.98) is a model of the fading channel (fading is the random variation of the signal level at the output of a communication channel). Multipath propagation of radio waves is a reason of fading. In the radio antenna receives two or more copies of the signal $s(t)$, past different ways; copies have different phases, as they were different distances. Adding, we can give a copy of an increase or attenuation of signal $s(t)$ depending on the difference of their phases.

Statistical characteristics is used to describe the function $k(t)$. It is important for us to know the probability distribution of values k. Most believe that $k(t)$ has Rayleigh distribution

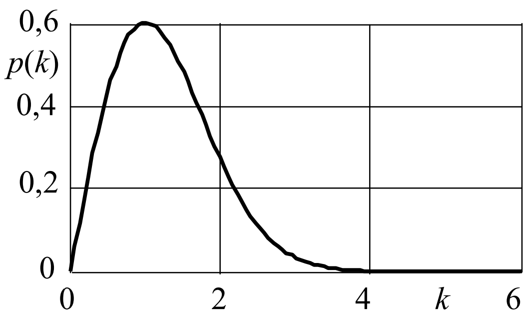


Figure 102 – Rayleigh distribution ($\overline{k^2} = 1$)

$$p(k) = \begin{cases} \dfrac{2k}{\overline{k^2}} \exp\left(-\dfrac{k^2}{\overline{k^2}}\right), & k \geq 0, \\ 0, & k < 0, \end{cases} \tag{5.99}$$

where $\overline{k^2}$ means a square of the transmission coefficient k.

Dependence (5.99) with $\overline{k^2} = 1$ shown in fig. 102. The graph shows that there may be situations when the signal is close to zero. Because of these situations, the problem is the recovery of the carrier signal, so demodulators are realized incoherent reception typically in such communication channels.

Signal/noise ratio depends on the transmission coefficient of communication channel, so the value of h_b is random with the same distribution as the k value:

$$p(h_b)=\begin{cases}\dfrac{2h_b}{\overline{h_b^2}}\exp\left(-\dfrac{h_b^2}{\overline{h_b^2}}\right), & h_b \geq 0,\\ 0, & h_b < 0.\end{cases} \tag{5.100}$$

Since h_b changes, then we can say about the conditional error probability. Thus, in the case of BFSK

$$P_{\text{er}}(h_b)=\frac{1}{2}\exp\left(-h_b^2/2\right). \tag{5.101}$$

Define the unconditional bit error probability

$$\begin{gathered}\overline{P_{\text{er}}}=\int_0^\infty p(h_b)P_{\text{er}}(h_b)dh_b=\int_0^\infty\frac{2h_b}{\overline{h_b^2}}\exp\left(-\frac{2h_b}{\overline{h_b^2}}\right)\frac{1}{2}\exp\left(-\frac{h_b^2}{2}\right)dh_b=\\ =\frac{1}{\overline{h_b^2}}\int_0^\infty h_b\exp\left(-h_b{}^2\cdot\frac{2+\overline{h_b^2}}{2\overline{h_b^2}}\right)dh_b=\frac{1}{\overline{h_b^2}}\left(-\frac{1}{2}\frac{2\overline{h_b^2}}{2+\overline{h_b^2}}\exp\left(-h_b^2\cdot\frac{2+\overline{h_b^2}}{\overline{h_b^2}}\right)\right)\Bigg|_0^\infty=\frac{1}{2+\overline{h_b^2}}.\end{gathered} \tag{5.102}$$

Compare the noise immunity of the reception in a channel with constant parameters – the formula (5.101) – and in a channel with variable parameters – the formula (5.102).

To achieve the error probability 10^{-6} in a channel with constant parameters required h_b^2= 26,2 or 14,2 dB, and in the channel with variable parameters $\overline{h_b^2}$= 10^6 or 60 dB, i.e. signal/noise ratio must be increased in 40000 times.

Similarly analyze the noise immunity of BDPSK signal reception in a channel with constant parameters

$$P_{\text{er}}(h_b)=\frac{1}{2}\exp\left(-h_b^2\right); \tag{5.103}$$

in a channel with variable parameters

$$\overline{P_{\text{er}}}=\frac{1}{2+\overline{h_b^2}}. \tag{5.104}$$

5.15 Non-optimal demodulators

We have considered the scheme of optimal demodulators and their noise immunity in all the previous sections. Recall that the criterion of demodulation optimality is the minimum error probability of signal and bit. Scheme of optimal MPSK ($M \geq 4$), MAPSK, MQAM demodulators are shown in fig. 82. To achieve minimum error probability in this scheme the following conditions are satisfied.

1. References oscillations, which are fed to the multiplier, is in phase with the cosine and sine components of the channel symbol, because this is a complete separation of cosine and sine components in order to follow their separate treatment, and the output voltage multipliers maximum.

2. Matched filter is matched with the pulse-carrier that provides maximum signal/noise ratio at the sample moment.

3. The ClR system provides taking samples at times when the instantaneous values of pulse-carrier are maximal.

4. Due to the fact that the shaping filters of the modulator and matched filter of the demodulator have the same AR, described by the dependence of the "root of the Nyquist spectrum", and transmission line does not distort the signal at the samplers' inputs we have impulses without intersymbol interference.

5. Decision rule at the decision circuit is based on the optimal partition of the space of signals on the signal areas.

Disturbance of any of these conditions leads to deterioration of noise immunity (increasing of the error probability). As a rule, all conditions in real demodulators in one way or another are not satisfied. So they say that the real demodulator is non-optimal.

1. References oscillations are generated from the noisy signal and do not correspond exactly in phase with the cosine and sine components of the channel symbol, so there is no complete separation of the cosine and sine components, there are transient noise, and the output voltage of multipliers is not maximal.

2. Because of the finite precision implementation of shaping and matched filters, signal distortion in a transmission line the matched filters are not exactly matched with the pulse-carrier and does not provide the maximal signal/noise ratio at the sample moment.

3. Sample pulses in the ClR system are formed from the noisy signal and don't provide taking samples at the instants when the instantaneous value of the pulse-carrier is maximum.

4. Because of the finite precision implementation of shaping and matched filter, signal distortion in a transmission line the intersymbol interference takes place at the samplers' inputs.

5. In the multilevel signal's demodulator there is not optimal partition of signals space to the signals areas as demodulated signal has levels that differ from the levels for which partitioning of the signals areas was calculated. To reduce this effect, the demodulator should have an automatic control of signal gain.

It is not possible to calculate the noise immunity of non-optimal demodulator. Noise immunity of non-optimal demodulator is determined experimentally or by simulation on a computer. In Fig. 103 shows the calculated dependence of the error probability on the signal/noise ratio for optimum demodulator and obtained experimentally.

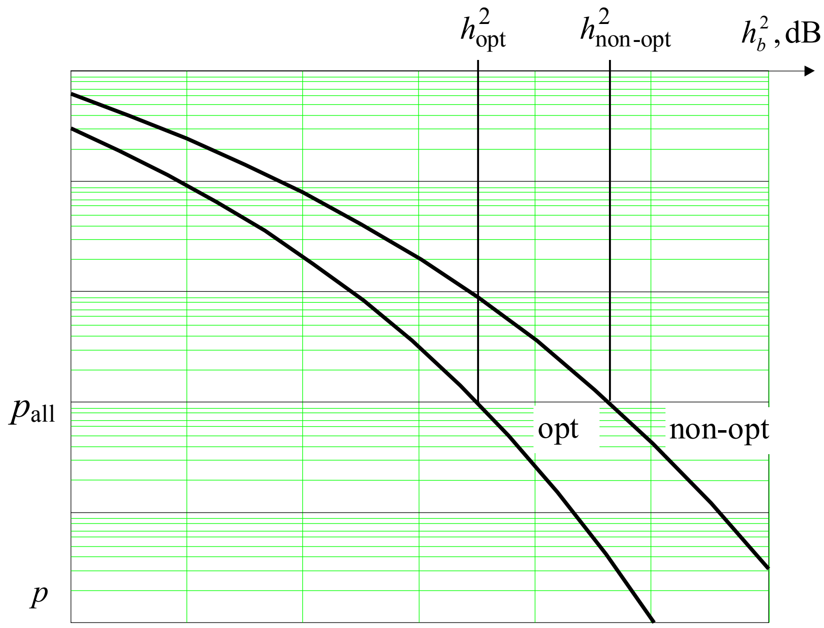


Figure 103 – Comparison of noise immunity of optimal and non-optimal demodulation

From the curves the value of the signal/noise ratio at the optimum reception h^2_{opt} and non-optimal reception $h^2_{\text{non-opt}}$ can be determined. The difference between these values is called the energy losses of demodulation (EL):

$$\text{EL} = h^2_{\text{non-opt}} - h^2_{\text{opt}}. \tag{5.105}$$

In modern well-implemented modems in the channel with AWGN for QPSK EL = 0,8...1,0 dB. With the increasing of signal levels number losses increases and for 64QAM, 256QAM they can reach 2 dB.

Questions for section 5

1. What is a quantitative measure of noise immunity of digital transmission systems?
2. What is a symbol-by-symbol reception of digital signals?
3. Formulate optimality criterion of the demodulators of digital modulated signals.
4. Formulate decision rule on the maximum a posteriori probability.
5. Formulate decision for maximum likelihood rule.
6. Why do we need to break up the space of the signals to signals areas?
7. List the functions that are performed by separate units of the scheme of optimal demodulator.
8. How to calculate the coordinates of the signal $z(t)$ in the space of channel symbols?
9. What rule does the decision scheme of the optimal demodulator work on?
10. What is the criterion of optimality of the matched filter?
11. List the properties of matched filter.

12. How is the peak signal-to-noise ratio in the output of the coherent filter determined?

13. Formulate the purpose of shaping and matched filters.

14. Formulate the purpose of clock recovery.

15. How is the RMS of noise at the output of MF determined?

16. Draw a scheme of the correlator and explain its work.

17. What is common and what is different in signal with interference processing by matched filter and correlator?

18. Describe two methods of matched filtering of radio pulses.

19. Explain the purpose of the scheme of carrier recovery.

20. Explain the purpose of separate blocks of MASK and BPSK signal optimum demodulator.

21. Formulate a decision rule on the basis of splitting the space of channel symbols into M areas.

22. Explain the purpose of the blocks of the optimal demodulator of two-dimensional MQAM and MPSK signals.

23. Write down and explain the formula for the error probability of channel symbol in a binary transmission system.

24. Explain the value $h_б^2$.

25. What simplifications are accepted in the analysis of a noise immunity of multi-position signals?

26. Write down and explain the formula of symbol error probability for binary signals 4PSK, 8PSK, 16QAM. Compare the noise immunity.

27. Explain why with the increasing in the number of signal positions the noise immunity is decreased.

28. Explain the operation of the carrier recovery scheme with frequency multiplication.

29. Explain what is the phase uncertainty of carrier.

30. Explain what is the Costas error detector.

31. Explain the principle of modulation and demodulation of BDPSK signal.

32. Explain the principle of modulation and demodulation of QDPSK signal.

33. How is the noise immunity of transmission systems with BDPSK and QDPSK methods determined?

34. Explain the principle of incoherent demodulation of BASK, BFSK and BDPSK signal.

35. Compare the noise immunity of coherent and incoherent methods of demodulation of BASK, BFSK and BDPSK signals.

36. Explain the purpose of the clock recovery system in the demodulator.

37. How is the error signal in the ClR system defined?

38. Why in the transmitted digital signal has to be transitions from 1 to 0 and from 0 to 1?

39. What is intersymbol interference in communication systems?

40. List the reasons of nonoptimality of real demodulators.

41. What is the energy loss of demodulation?

6. NOISE IMMUNITY OF ANALOG MODULATED SIGNALS RECEPTION

6.1 General information about the analog demodulation signal modulation types

An analog signal is a baseband signal $b(t)$, which can be continuously variable in time and accept any form. Such signals take place, for example, in telephony, broadcast, television and telemetry. The signal $b(t)$ in these cases is beforehand unknown: it is known only, that it belongs to some great number or is realization of some random process $B(t)$. Analog signals can be transmitted *directly* or with the use of *modulation.* In first case the transmitted signal is proportional the analog signal $s(t) = kb(t)$, where k is a constant factor. In second case a signal $s(t)$ is some function of analog signal $b(t)$ (see section 4.2).

Block-diagram of the analog transmission system is similarly as well as block-diagram of the digital transmission system (fig. 63). A signal on the demodulator input is total waveform of the transmitted signal $s(t)$ and noise $n(t)$

$$z(t) = s(t) + n(t). \tag{6.1}$$

Noise $n(t)$ is realization of stationary Gaussian process with the power spectral density N_0 in the signal frequencies band. Out of signal frequencies band noise PSD equal to the zero.

A task consists of that, on the input signal $z(t)$ to get (to recover) a baseband signal $\hat{b}(t)$, the least different in sense of some criterion, from the transmitted signal $b(t)$. The signal $\hat{b}(t)$ reproduced with some error is named the signal $b(t)$ estimation. Thus, the task of signal $s(t)$ demodulation can be examined as a task of receipt of signal estimation $\hat{b}(t)$. In special case, when a signal $s(t)$ is the function of some time-constant parameter λ, a task is taken to the signal parameter λ estimation. At the direct transmission of signal $s(t) = kb(t)$ calculation of estimation $\hat{b}(t)$ is taken to signal linear filtration. At a transmission the modulated signal $s(t)$ the signal estimation $\hat{b}(t)$ in demodulator is carried out by detection and signal processing. In many cases signals processing is taken to one or another methods of filtration and can be carried out both before the detector and after it.

As measure of noise immunity at analog signals transmission can be a degree of "deviation" of the got estimation $\hat{b}(t)$ from the transmitted signal $b(t)$. An average square deviation or an average square error is usually accepted

$$\overline{\varepsilon^2(t)} = \overline{\left[\hat{b}(t) - b(t)\right]^2}, \tag{6.2}$$

where averaging is executed on all possible realization of signals $\hat{b}(t)$ and $b(t)$.

Difference $\varepsilon(t)$ = $\hat{b}(t)$ – $b(t)$ is a noise on the demodulator output. Its average square $\overline{\varepsilon^2(t)} = P_\varepsilon$ is noise power on the demodulator output. Power of the transmitted signal $P_b = \overline{b^2(t)}$ is considered set. Then it is possible to define the powers of signal and noise ratio on the demodulator output ρ_{out} = P_b/P_ε. Powers of signal and noise ratio on the demodulator input ρ_{in} it is usually known. In many cases as a criterion of noise immunity accept not error average square $\overline{\varepsilon^2(t)}$, but signal/noise ratio ρ_{out}. Demodulator can improve signal/noise ratio ρ_{in}. This improvement depends not only on the method of the signal processing but also on the modulation type. Therefore comfortably to estimate noise immunity of the analog transmission systems by the signal/noise ratio gain

$$g = \frac{\rho_{\text{out}}}{\rho_{\text{in}}} = \frac{P_b / P_\varepsilon}{P_s / P_n}. \tag{6.3}$$

At $g > 1$ signal/noise ratio is improved in demodulator. On occasion $g < 1$, that means, that not gain, but losses take place at demodulation.

Other correctness criteria of continuous messages transmitting are used in practice. For example, criterion of speech intelligibility at transmitting of vocal messages, criterion of maximal error in a telemetry and other are used.

6.2 Optimum demodulator of analog modulated signals

The set forth below going near the construction of optimum demodulator is applicable to the case of weak noise (high signal/noise ratio on demodulator input). Such situation is incident to the analog transmission systems, which is used in communication and broadcasting networks.

The criterion of optimality is a minimum of error means a square of baseband signal recovery. Demodulator of analog modulation signals necessarily contains a detector, corresponding to the used modulation type – its output is proportional the values of informative parameter of the modulated signal. For error minimization of baseband signal recovery it is necessary to execute optimum processing:

– to recover the modulated signal by a pre-detector filter;

– to weaken unwanted components, taking place on the detector output by a post-detector filter.

The analog modulation signals optimum demodulator scheme follows from said (fig. 104)

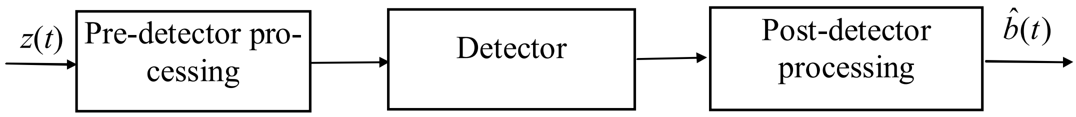


Figure 104 – Analog modulation signals optimum demodulator scheme

As devices of pre-detector and post-detector processing, as a rule, linear filters are used, although nonlinear processing is required on occasion. It is explained, foremost, by simplicity of realization of linear filters, which comparatively synthesized easily, and there is the developed theory of their construction, what cannot be said about nonlinear filters.

6.3 Optimum linear filtration of continuous signals

We will consider the theory of optimum linear filtration – Kolmogorov-Wiener filter or OLF.

Let on the linear filter input the sum of useful signal $s(t)$ and noise $n(t)$ operates with the transfer function $H(j\omega)$. It is required to find such function $H(j\omega)$, which provides a minimum of error mean square of recovered signal

$$\overline{\varepsilon^2(t)} = \overline{[\hat{s}(t) - s(t)]^2}\,, \tag{6.4}$$

where $\hat{s}(t)$ is an estimation of signal on the filter output.

We will suppose that $s(t)$ and $n(t)$ – stationary uncorrelated processes with the known power spectral density (PSD) $G_s(f)$ and $G_n(f)$. In such raising a task was decided A.N. Kolmogorov (1939) for discrete random sequences and N. Wiener (1941) for continuous processes. Therefore an optimum (in the indicated sense) linear filter is called the Kolmogorov-Wiener filter.

Operating on the filter input the signal $s(t)$ and noise $n(t)$ is passed through a filter independently and create on the filter output accordingly filtered signal $s^f(t)$ and noise $n^f(t)$. Taking into account it an error on the filter output will be written down

$$\varepsilon(t) = \hat{s}(t) - s(t) = \left(s^f(t) - s(t)\right) + n^f(t) = \Delta s(t) + n^f(t). \tag{6.5}$$

Element $\Delta s(t)$ reflects the error component, conditioned linear distortions of useful signal by a filter. Error mean square $\overline{\varepsilon^2(t)}$ equals

$$\overline{\varepsilon^2(t)} = \overline{\Delta s^2(t)} + \overline{\left(n^f(t)\right)^2}\,. \tag{6.6}$$

The size of linear distortions of useful signal by a filter depends on the degree of difference of filter AR from a constant value and degree of PR difference from linear dependence. The noise mean square on the filter output depends only of filter AR. In order that linear distortions of useful signal were minimum will accept filter PR linear

$$\varphi(\omega) = -\omega t_0, \tag{6.7}$$

where t_0 is a delay of signal in a filter.

We will pass to determination of filter AR. For this purpose we will define the power spectral density of the formula left and right parts (6.5)

$$G_\varepsilon(\omega) = G_{\Delta s}(\omega) + G_{nf}(\omega). \tag{6.8}$$

We will give components in right part correlations (6.8) through signal $s(t)$ and noise $n(t)$ PSD and necessary filter AR:

$$G_{nf}(\omega) = G_n(\omega) H^2(\omega); \tag{6.9}$$

$$G_{\Delta s}(\omega) = S^2_{\Delta s}(\omega), \tag{6.10}$$

where $S_{\Delta s}(\omega)$ is an error amplitude spectrum $\Delta s(t)$;

$$S_{\Delta s}(\omega) = S(\omega)[H(\omega) - 1], \tag{6.11}$$

where $S(\omega)$ is a signal amplitude spectrum $s(t)$;

$$G_{\Delta s}(\omega) = G_s(\omega)[H(\omega) - 1]^2 . \tag{6.12}$$

After the substitution of relations (6.9) and (6.12) in (6.8) will get

$$G_\varepsilon(\omega) = G_s(\omega)[H(\omega) - 1]^2 + G_n(\omega)H^2(\omega). \tag{6.13}$$

The recovery error mean square (average power) is calculated as

$$\overline{\varepsilon^2(t)} = P_\varepsilon = \frac{1}{\pi}\int_0^\infty G_\varepsilon(\omega)d\omega . \tag{6.14}$$

As function $G_\varepsilon(\omega) \geq 0$ on all frequencies, then, providing min $G_\varepsilon(\omega)$ on all frequencies, we will attain a minimum of size $\overline{\varepsilon^2(t)}$. Necessary AR $H(\omega)$ we will define from a condition

$$\frac{dG_\varepsilon(\omega)}{dH(\omega)} = 0 : \tag{6.15}$$

$$\frac{dG_\varepsilon(\omega)}{dH(\omega)} = 2G_s(\omega)[H(\omega) - 1] + 2G_n(\omega)H(\omega) = 0 . \tag{6.16}$$

After the decision of equalization (6.16) will get expression for filter AR

$$H(\omega) = \frac{G_s(\omega)}{G_s(\omega) + G_n(\omega)} . \tag{6.17}$$

On fig. 105 OLF AR, determined relation (6.17) is illustrated.

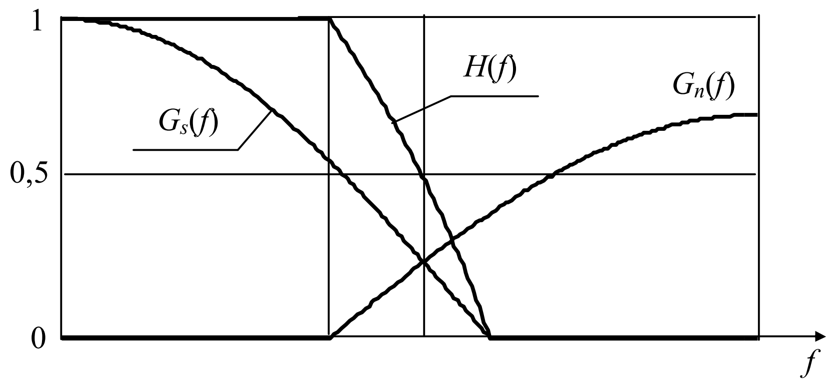


Figure 105 – Illustration of OLF AR

From fig. 105 the features of OLF AR are visible:

– in frequencies band, where $G_n(f) \equiv 0$, AR value $H(f) = 1$ – in this bandwidth a filter must not bring in distortions;

– in frequencies band, where $G_s(f) \equiv 0$, AR value $H(f) = 0$ – in this bandwidth a filter must fully weaken noise;

– on frequency on which $G_s(f) = G_n(f)$, AR $H(f) = 0{,}5$;

– on other frequencies AR value determined by calculations on a formula (6.7).

If to put expression (6.17) in relation (6.13), will get expression for error PSD

$$G_\varepsilon(\omega) = \frac{G_s(\omega)G_n(\omega)}{G_s(\omega)+G_n(\omega)}. \tag{6.18}$$

At the substitution of relation (6.18) in expression (6.14) it is possible to calculate the recovery error mean square of signal $\overline{\varepsilon^2(t)}$.

Easily to notice that error $\overline{\varepsilon^2(t)}$ = 0 only in that case, when $G_s(f) \cdot G_n(f) = 0$, i.e. when the signal and noise spectrums of are not overlapped. In all other cases an optimum filter transmits the oscillations of different frequencies with that less weight, than relation $G_n(f)/G_s(f)$ more on this frequency.

It is necessary to notice that optimum linear filters, determined equation (6.17), substantially differ from the matched filters, considered above. If the basic setting of the filters examined here consists of the best reproducing of signal form, the task of the matched filters consists in forming of maximal signal/noise ratio in the sampling moment.

At the use of OLF as a device of pre-detector processing such feature comes to demodulators of the analog transmission systems. The high ratio of signal and noise spectral densities takes place $G_s(f)/G_n(f) >> 1$. Expression for OLF AR (6.17) passes to the following

$$H(f) = \begin{cases} 1, & f_{\min} \le f \le f_{\max}, \\ 0, & f < f_{\min},\ f > f_{\max}, \end{cases} \tag{6.19}$$

where $f_{\min}$ and $f_{\max}$ are a scope frequencies of signal spectrum.

Thus, the device of pre-detector processing is a bandpass filter with rectangular AR.

6.4 Comparison of noise immunity of optimum demodulators analog signal modulation types

We found that the demodulator should contain:

– pre-detector processing filter;

– detector;

– post-detector processing filter.

Filters should be optimal for optimal demodulation. In the weak noise filters AR must be rectangular:

– pre-detector processing filter is a bandpass filter, passband edge frequency of which coincide with the boundary frequencies of the spectrum of the modulated signal;

– post-detector processing filter is a LPF, cut-off frequency of which coincides with the maximum frequency spectrum of the baseband signal $F_{\max}$.

Noise immunity defines in terms of the AWGN. Analysis of noise immunity is to determine the gain demodulator in the signal/noise

$$g = \frac{P_b / P_\varepsilon}{P_s / P_n}. \quad (6.20)$$

To determine the gain required to determine the 4 values: P_b, P_ε, P_s, P_n. In view of the cumbersome calculations (they can be found in textbooks on the theory of telecommunication) we will give the final expressions and compare the analog transmission system. Demodulator gain in signal/noise ratio g and the coefficient of expansion of the bandwidth at modulation $\alpha = \Delta F_s / F_{\max}$ are the main parameters on which transmitting systems are compared. For the considered modulation techniques, these parameters are summarized in table 10.

Let's make a comparison of numerical values of the parameters in some initial data: $K_A = 5$; $m_{FM} = m_{PM} = 6$; $m_{AM} = 1$.

Calculations give: $g_{AM} = 0{,}038$; $g_{DSB\text{-}SC} = 2$; $g_{SSB} = 1$; $g_{FM} = 60{,}5$; $g_{PM} = 20{,}2$.

Comparison of the numerical values of the gain shows that the lowest noise immunity has transmission system with AM: gain $g_{AM} \ll 1$, logically be called a losses. However, AM is used in the broadcasting system, where this deficiency is compensated by the simplicity of the demodulator on the basis of the envelope detector (a huge number of more simple radios and a radio transmitter with more power than at DSB-SC or SSB using).

Table 10 – The main parameters of analog transmission systems

Modulation method	g	α	Notes
AM	$\frac{2m_{AM}^2}{m_{AM}^2 + K_A^2}$	2	Synchronous detector
	$\frac{m_{AM}^2}{m_{AM}^2 + K_A^2}$	2	Envelope detector
DSB-SC	2	2	
SSB	1	1	
FM	$\frac{3m_{FM}^2}{K_A^2} \cdot \alpha$	$2(m_{FM} + 1)$	$\rho_{in} \geq \rho_{thr}$
PM	$\frac{m_{PM}^2}{K_A^2} \cdot \alpha$	$2(m_{PM} + 1)$	$\rho_{in} \geq \rho_{thr}$

The highest noise immunity has a transmission system with FM. "Payment" for high noise immunity is a wide bandwidth signal. Thus, when $F_{\max} = 3{,}4$ kHz and $\Delta F_{FM} = 47{,}6$ kHz, while the bandwidth of the signal SSB $\Delta F_{SSB} = 3{,}4$ kHz.

The threshold of noise immunity of the FM signals demodulator. From the relation of FM demodulator gain (table 10) follows that the larger index m_{FM}, the larger gain (though at the price of signal bandwidth increasing). It makes you think that this gives the opportunity to work demodulators with a weak signal (low signal/noise ratio). One should take into account the phenomenon of the threshold of noise immunity. It is as following: if the signal/noise ratio at the input of the demodulator ρ_{in} is higher than threshold signal/noise ratio ρ_{thr}, demodulator that provides high gain, defined by the relation given in table 10, but if $\rho_{in} < \rho_{thr}$, then gain decreases sharply, and this range of values ρ_{in} is nonworking.

In the demodulator, performed on the basis of the standard frequency detector ρ_{thr} =10 dB. So-called circuit threshold-extension FM demodulator has been proposed, which are called:

– demodulator with tracking filter;
– demodulator with frequency feedback;
– demodulator on the basis of synchronous phase detector.

In the demodulator threshold signal/noise ratio, depending on the initial data on the transmission system may be 5...7 dB. Reduction in the ρ_{thr} allows:

1) to operate demodulators with lower signal/noise ratio;
2) increase the gain, as

$$\rho_{\text{in}} = \frac{P_s}{N_0 \Delta F_{\text{FM}}} = \frac{P_s}{N_0 2F_{\max}\left(m_{\text{FM}} + 1\right)}, \tag{6.21}$$

and assuming a decrease signal/noise ratio ρ_{in}, is possible to increase the index m_{FM}, and increase m_{FM} leads to an increase in gain.

Questions for section 6

1. What is the quantitative measure of the noise immunity of the analog transmission systems?
2. What criteria of noise immunity are used in analog transmission systems?
3. What do blocks the scheme of optimal demodulator of analog modulated signals consist of?
4. Formulate optimality criterion of the OLF.
5. Under what conditions can the signal recovery error be reduced to zero?
6. Explain the difference between OLF and coherent filter from the point of view of noise reduction.
7. Compare the gain in the signal-to-noise ratio for AM, DSB-SC, SSB, PM and FM signal demodulators.
8. Explain what is the phenomenon of "the threshold of noise immunity of FM signal reception".
9. What is thresholds decreasing scheme?

7 DICTIONARIES

English-Russian dictionary

access system	система доступа
accidental phase	случайная фаза
amplifier	усилитель
amplitude factor	коэффициент амплитуды
amplitude modulation	амплитудная модуляция
amplitude modulation factor	коэффициент амплитудной модуляции
analog (амер.), analogue (англ.)	аналоговый, аналог
attenuation	ослабление
attenuator	аттенюатор
average power (of a process)	средняя мощность (процесса)
average value	среднее значение
band	полоса частот
bandpass noise	полосовой шум
bandpass signals	полосовые сигналы
bandwidth (of signal, process)	ширина спектра (сигнала, процесса)
baseband signal	первичный сигнал
binary modulation types	двоичные виды модуляции
bit rate of a signal	скорость цифрового сигнала
block diagram	структурная схема
BPSK (binary phase shift keying)	ФМ-2 (двоичная фазовая модуляция)
broadcasting	вещание
carrier	несущее колебание (несущая)
carrier frequency	частота несущего колебания (несущей)
carrier recovery (CR)	восстановление несущей (ВН)
channel bandwidth	полоса пропускания канала
channel symbols	канальные символы, элементарные сигналы (импульсы)
clock period, timing period, timing interval	тактовый интервал
clock recovery (ClR)	восстановление тактовой синхронизация (ТС)
coded block, codeword	кодовая комбинация (блок), кодовое слово
constant component	постоянная составляющая

continuous (time-continuous) signal	непрерывный (непрерывный по времени) сигнал
correlation characteristics	корреляционные характеристики
correlation meter	коррелометр
correlation time	интервал корреляции
decision	решение, решающая схема
decomposition factors	коэффициенты разложения
delay, delay time	задержка, время задержки
detecting signal	детектируемый сигнал
detector gain	выигрыш детектора
determined signals	детерминированные сигналы
deterministic signal	детерминированный сигнал
deviation	девиация
differential decoder	относительный декодер
differential encoder	относительный кодер
differentially encoded phase modulation (DPSK)	фазоразностная модуляция, относительная фазовая модуляция
digital signal	цифровой сигнал
discrete signal	дискретный сигнал
distortion (of signal)	искажение, изменение формы (сигнала)
double-sideband-suppressed-carrier modulation	балансная модуляция
duplex	двухсторонняя передача сообщений
duration (infinite/finite)	длительность (бесконечная/конечная)
duration of pulse	длительность импульса
energy losses (EL)	энергетические потери
energy per bit	энергия на бит
energy spectral density	спектральная плотность энергии
ensemble	ансамбль, совокупность
envelope	огибающая
envelope detector (ED)	детектор огибающей
equiprobable	равновероятный
ergodic	эргодический
error control code	корректирующий код
error detector	детектор ошибки
error ratio	коэффициент ошибок
even function	четная функция

expectation	математическое ожидание
fluctuation noise	флуктуационная помеха
Fourier series	ряд Фурье
Fourier transformation	преобразование Фурье
frequency detector	частотный детектор
frequency deviation	девиация частоты
frequency modulation index	индекс ЧМ
full duplex	полнодуплексный
gain in signal-to-noise ratio	выигрыш в отношении сигнал/шум
Gaussian Q-function	гауссовская Q-функция
half duplex	полудуплексный
harmonic oscillation, harmonious waveform	гармоническое колебание
Hilbert transform	преобразование Гильберта
impulse response	импульсный отклик
incoherent demodulation	некогерентная демодуляция
initial phase	начальная фаза
in-phase, cosine component	синфазная, косинусная составляющая
intersymbol interference	межсимвольная интерференция
joint probability density	совместная плотность вероятности
Kolmogorov-Wiener filter	фильтр Колмогорова-Винера
link	соединение
lower sideband of frequencies	нижняя полоса частот
low-frequency components	низкочастотные составляющие
mapper	кодер модуляционного кода
mapping code	модуляционный код
MAPSK – M-ary amplitude-phase modulation;	АФМ-M
M-ary amplitude modulation	M-ичная амплитудная модуляция
matched filter	согласованный фильтр
maximum a posteriori probability	максимум апостериорной вероятности
maximum likelihood rule	правило максимума правдоподобия
mean-square error	среднеквадратическая ошибка
MFSK – M-ary frequency modulation	ЧМ-M
middle frequency	средняя частота
modulating signal	модулирующий сигнал

monotone decreasing	монотонно убывающая
MPSK – *M*-ary phase modulation	ФМ-*М* – *М*-ичная фазовая модуляция
MQAM – *M*-ary quadrature-amplitude modulation	КАМ-*М* – *М*-ичная квадратурная амплитудная модуляция
MSK modulation – minimum shift keying modulation	ММС – модуляция минимального сдвига
mutual correlation	взаимная корреляция
narrow spectrum	узкий спектр
narrow-band signal	узкополосный сигнал
node	узел сети
noise immunity	помехоустойчивость
noise power spectral density	удельная мощность шума
Nyquist pulse	импульс Найквиста
one-dimensional	одномерный
one-way message transfer	односторонняя передача сообщения
optimal demodulator	оптимальный демодулятор
periodic signal	периодический сигнал
phase locked-loop (PLL)	фазовая автоподстройка частоты (ФАПЧ)
phase uncertainty	неопределенность фазы
power spectral density (PSD)	спектральная плотность мощности (СПМ)
power spectral density function	спектральная плотность мощности
probability density function	плотность вероятности
probability distribution function	функция распределения вероятности
pulse energy	энергия импульса
pulse-carrier	импульс-переносчик
QPSK (quaternary phase shift keying)	ФМ-4 (четверичная фазовая модуляция)
quadratic voltmeter	квадратичный вольтметр
quadrature splitter	квадратурный расщепитель
quadrature, sine component	квадратурная, синусная составляющая
raised cosine	поднятый косинус
random process	случайный процесс
Rayleigh probability distribution	Релеевское распределение вероятности
realizations of random process	реализации случайного процесса
recipient	получатель

rectangular radio pulse	радиоимпульс с П-образной огибающей
rectangular video pulse	П-импульс
reference waveform	опорное колебание
reliability	достоверность
roll-off factor	коэффициент ската
root-mean square deviation (RMS)	среднеквадратическое отклонение (СКО)
sampler	дискретизатор
sampling	дискретизация
sampling condition	условие отсчетности
sequence	последовательность
shaping filter	формирующий фильтр
shifted on frequency	сдвинут по частоте
signal constellation	сигнальное созвездие
signal points	точки сигнального созвездия
signal to noise ratio (SNR)	отношение сигнал/шум (ОСШ)
signalling alphabet	сигнальный алфавит
simplex transfer	только односторонняя передача сообщения
single-sideband modulation (SSB)	однополосная модуляция (ОМ)
spectral spreading	расширение спектра
spectrum	спектр
square root of the Nyquist spectrum	корень квадратный из спектра Найквиста
stationary	стационарный
statistical dependence	статистическая зависимость
stochastic signals	стохастические сигналы
symbol interval	тактовый интервал
symbol rate	символьная скорость, скорость модуляции
symbol-by-symbol demodulation	поэлементный прием
synchronous detector	синхронный детектор
terminal equipment	оконечное оборудование
threshold	порог
tracking filter	следящий фильтр
transducer	датчик

two-way message transfer	двухсторонняя передача сообщений
ultimate values	крайние значения аргумента
uncorrelated	некоррелированные
uniform distributing	равномерное распределение
upper sideband of frequencies	верхняя боковая полоса частот
Variance (dispersion)	дисперсия
voltage-controlled oscillator (VCO)	генератор, управляемый напряжением (ГУН)
waveform	колебание
Wiener-Khinchin theorem	теорема Хинчина-Винера

Russian-English dictionary

амплитудная модуляция	amplitude modulation
аналоговый (непрерывный первичный) сигнал	analogue signal
аналоговый, аналог	analog (амер.), analogue (англ.)
ансамбль, совокупность	ensemble
аттенюатор	attenuator
АФМ-*M* (*M*-ичная амплитудно-фазовая модуляция)	MAPSK – *M*-ary amplitude-phase modulation;
балансная модуляция	double-sideband-suppressed-carrier modulation
верхняя боковая полоса частот	upper sideband of frequencies
вещание	broadcasting
взаимная корреляция	mutual correlation
восстановление несущей (ВН)	carrier recovery (CR)
восстановление тактовой синхронизации (ТС)	clock recovery (ClR)
выигрыш в отношении сигнал/шум	gain in signal-to-noise ratio
выигрыш детектора	detector gain
гармоническое колебание	harmonic oscillation, harmonious waveform
гауссовская *Q*-функция	Gaussian *Q*-function
генератор, управляемый напряжением (ГУН)	voltage-controlled oscillator (VCO)
датчик	transducer
двоичные виды модуляции	binary modulation types
двухсторонняя передача сообщений	duplex, two-way message transfer
девиация	deviation
девиация частоты	frequency deviation
детектируемый сигнал	detecting signal
детектор огибающей	envelope detector (ED)
детектор ошибки	error detector
детерминированные сигналы, детерминированный сигнал	determined signals, deterministic signal
дискретизатор	sampler
дискретизация	sampling
дискретный сигнал	discrete signal
дисперсия	variance (dispersion)

длительность (бесконечная/конечная)	duration (infinite/finite)
длительность импульса	duration of pulse
достоверность	reliability
задержка, время задержки	delay, delay time
импульс Найквиста	Nyquist pulse
импульсный отклик	impulse response
импульс-переносчик	pulse-carrier
индекс ЧМ	frequency modulation index
интервал корреляции	correlation time
искажение, изменение формы (сигнала)	distortion (of signal)
КАМ-*M* (*M*-ичная квадратурная амплитудная модуляция)	MQAM – *M*-ary quadrature-amplitude modulation
канальные символы	channel symbols
квадратичный вольтметр	quadratic voltmeter
квадратурная или синусная составляющая	quadrature or sinus component
квадратурная, синусная составляющая	quadrature, sine component
квадратурный расщепитель	quadrature splitter
кодовая комбинация (блок), кодовое слово	code block, codeword
колебание	waveform
корень квадратный из спектра Найквиста	square root of the Nyquist spectrum
корректирующий код	error control code
коррелометр	correlation meter
корреляционные характеристики	correlation characteristics
коэффициент амплитудной модуляции	amplitude modulation factor
коэффициент амплитуды	amplitude factor
коэффициент ошибок	error ratio
коэффициент ската	roll-off factor
коэффициенты разложения	decomposition factors
крайние значения аргумента	ultimate values
максимум апостериорной вероятности	maximum a posteriori probability
математическое ожидание	expectation
межсимвольная интерференция	intersymbol interference
M-ичная амплитудная модуляция	*M*-ary amplitude modulation

модулирующий сигнал	modulating signal
модуляционный код	mapping code
модуляция минимального сдвига	MSK modulation – minimum shift keying modulation
монотонно убывающая	monotone decreasing
начальная фаза	initial phase
некогерентная демодуляция	incoherent demodulation
некоррелированные	uncorrelated
неопределенность фазы	phase uncertainty
непрерывный (непрерывный по времени) сигнал	continuous (time-continuous) signal
несущее колебание (несущая)	carrier
нижняя полоса частот	lower sideband of frequencies
низкочастотные составляющие	low-frequency components
огибающая	envelope
одномерный	one-dimensional
однополосная модуляция (ОМ)	single-sideband modulation (SSB)
односторонняя передача сообщения	one-way message transfer
оконечное оборудование	terminal equipment
опорное колебание	reference waveform
оптимальный демодулятор	optimal demodulator
ослабление	attenuation
относительная фазовая модуляция	differentially encoded phase-shift keying
относительный декодер	differential decoder
относительный кодер	differential encoder
отношение сигнал/шум (ОСШ)	signal to noise ratio (SNR)
первичный сигнал	baseband signal
периодический сигнал	periodic signal
П-импульс	rectangular video pulse
плотность вероятности	probability density function
поднятый косинус	raised cosine
полнодуплексный	full duplex
полоса пропускания канала	channel bandwidth
полоса частот	band
полосовой шум	bandpass noise
полосовые сигналы	bandpass signals
полудуплексный	half duplex

получатель	recipient
помехоустойчивость	noise immunity
порог	threshold
последовательность	sequence
постоянная составляющая	constant component
поэлементный прием	symbol-by-symbol demodulation
правило максимума правдоподобия	maximum likelihood rule
преобразование Гильберта	Hilbert transform
преобразование Фурье	Fourier transformation
равновероятный	equiprobable
равномерное распределение	uniform distributing
радиоимпульс с П-образной огибающей	rectangular radio pulse
расширение спектра	spectral spreading
реализация процесса	realization of a process
Релеевское распределение вероятности	Rayleigh probability distribution
решение, решающая схема	decision
ряд Фурье	Fourier series
сдвинут по частоте	shifted on frequency
сигнальное созвездие	signal constellation
сигнальный алфавит	signalling alphabet
символьная скорость, скорость модуляции	symbol rate
синфазная или косинусная составляющая	in-phase or cosine component
синфазная, косинусная составляющая	in-phase, cosine component
синхронный детектор	synchronous detector
система доступа	access system
скорость цифрового сигнала	bit rate of a signal
следящий фильтр	tracking filter
случайная фаза	accidental phase
случайный процесс	random process
совместная плотность вероятности	joint probability density
согласованный фильтр	matched filter
соединение	link
спектр, спектры	spectrum, spectra

спектральная плотность мощности	power spectral density function
спектральная плотность мощности (СПМ)	power spectral density
спектральная плотность энергии	energy spectral density
среднее значение	average value
среднеквадратическая ошибка	mean-square error
среднеквадратическое отклонение	root-mean-square deviation
среднеквадратическое отклонение (СКО)	root-mean square deviation (RMS)
средняя мощность (процесса)	average power (of a process)
средняя частота	middle frequency
статистическая зависимость	statistical dependence
стационарный	stationary
стохастические сигналы	stochastic signals
структурная схема	block diagram
тактовый интервал	symbol interval
тактовый интервал	clock period, timing period, timing interval
теорема Хинчина-Винера	Wiener-Khinchin theorem
только односторонняя передача сообщения	simplex transfer
точки сигнального созвездия	signal points
удельная мощность шума	noise power spectral density
узел сети	node
узкий спектр	narrow spectrum
узкополосный сигнал	narrow-band signal
усилитель	amplifier
условие отсчетности	sampling condition
фазовая автоподстройка частоты (ФАПЧ)	phase locked-loop (PLL)
фазоразностная модуляция	differentially encoded phase-shift keying
фильтр Колмогорова-Винера	Kolmogorov-Wiener filter
флуктуационная помеха	fluctuation noise
ФМ-M – M-ичная фазовая модуляция	MPSK – M-ary phase modulation
формирующий фильтр	shaping filter
функция распределения вероятности	probability distribution function
цифровой сигнал	digital signal

частота несущего колебания (несущей)	carrier frequency
частотный детектор	frequency detector
четная функция	even function
ЧМ-*M*	MFSK – *M*-ary frequency modulation
ширина спектра (сигнала, процесса)	bandwidth (of signal, process)
энергетические потери	energy losses (EL)
энергия импульса	pulse energy
энергия на бит	energy per bit
эргодический	ergodic

Education publication

Ivaschenko Peter Vasilyevich
Rozenvasser Denis Mikhailovich

TELECOMMUNICATION THEORY

Textbook
Part 1

Signed for printing 17.02.17
Format 60/88/16 Order №5982
Edition of 100 copies. Amount of quires 7,75.
Printed by the publishing RISO company equipment
Printed at the Editorial and Publishing Center of ONAT named after A. S. Popov