# 6. COMMUNICATION CHANNELS

## 6.1. Introduction

The goal of any kind of communication is to send or to copy an information to another place (or often to more places) with the help of appropriate devices and equipment. The information sender is often called the *source* while the receiver is called the *sink*. The path between the source and the sink is the communication channel.

If there are only two partners taking part in the communication, we speak about point-to-point communication. There are also multipoint systems with one source sending information to more sinks. In *simplex* systems, the information can be passed in one direction only. The system is said to be *half-duplex* if the communication is possible in both directions, but not at the same time. Finally, in *duplex* systems the communication may happen in both directions without any time restrictions. In this chapter the characteristics of *simplex*, point-to-point communication will be discussed and considered as the communication channel.

The simplest communication channels consist of a transfer medium and of transducers interfacing the information to this medium. The characteristics of such channels are influenced mainly by the transfer medium. In this sense wirebound and wireless channels are distinguished. From the user's point of view, however, it is more important how he can be connected to the channel, what expected quality of the channel can be and what kind of information can be sent.

A channel is said to be analog if analog signals are transmitted and received at its input and output. On the contrary, a digital channel transmits and receives digital signals or series of symbols between the input and the output points. These input and output points are called the *interfaces*.

Usually, the transmission path from the source to the sink is built upon cascaded channel sections. It may also happen that by means of convenient transducers, a digital channel is built upon an analog channel or, on the contrary, an analog channel is formed from a digital channel (see Fig. 6.1.).

When a channel is to be characterized, first of all, the essential differences between analog and digital channels shall be taken into account. In the following, the channel properties are discussed from this point of view.

## 6.2. Analog Channels

As we have seen in Fig. 6.1., an analog channel is a section of the communication channel receiving analog signal at the input interface and reproducing analog signal at the output interface. Such a channel can be characterized by the specification of the signals, for which the channel can provide a satisfying operation. More detailed characterization can be given if the effects produced by the channel are defined with the help of simple models. Sometimes it is difficult to characterize a parameter of the channel in the desired depth. In this case the channel is said to be uncertain or unspecified from the point of view of the given parameter. However, this uncertainty does not affect definitely the quality of transmission, it only sets some restrictions on the transmitted signals. (For instance, telephone channels are not specified in the frequency range from 0 to 300 Hz but such a lack of specification is irrelevant since the transmitted signals do not contain such spectral components.) There is another restriction of the input signal if some values of a certain signal parameter disturb the operation of the channel or have influence on other channels.

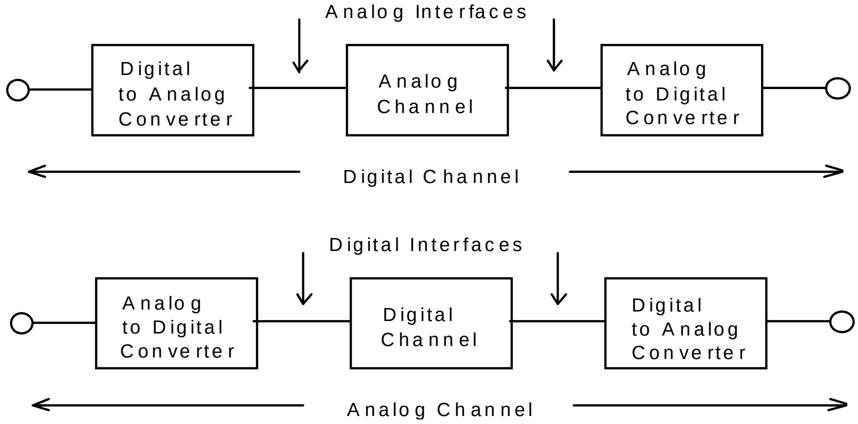


Fig. 6.1 Analog and Digital Channels Built Upon Each Other

Three effects are considered in commonly used channel models:

- linear distortion (which can be either time-invariant or time-variant),
- nonlinear distortion (which may be memoryless or looped),
- noises (noise is meant as an effect independent of the input signal).

### *6.2.1. Time-Invariant Linear Distortion*

This is a kind of distortion typically caused by the attenuation and time delay the signal suffers when passing through the transmission medium and the interfacing devices. This distortion is present almost always and generally it is not too harmful as the time delay (if it is short) usually does not cause any problem and the channel attenuation can be compensated by appropriate amplification. The distortion of the channel is generally frequency-dependent; this dependence can be described by the channel frequency response $H_C(f)$, $f \in (-\infty,\infty)$. *Attenuation* and *phase* are the quantities derived from the frequency response and used for the practical characterization of the channel:

$$a(f) = -20\lg\left|H_c(f)\right| \quad \text{and} \quad f(f) = -\mathrm{arc}\{H_c(f)\} \qquad (6.1)$$

It is easy to see that if the signal is attenuated by $a_o$ and suffers a delay $T$, the distortion can be modelled by the channel with

$$a(f) = a_0 \quad \text{and} \quad f(f) = 2pfT \qquad (6.2)$$

For those frequencies where the signal does not have any spectral components, the behaviour of the channel is indifferent.

It is an often case that different components of the signal have different attenuation and delay. This phenomenon is called *dispersion* and is described in detail by the frequency response $H_c(.)$. For the superficial characterization of the dispersion, test impulses shown in Fig. 6.2. are used.

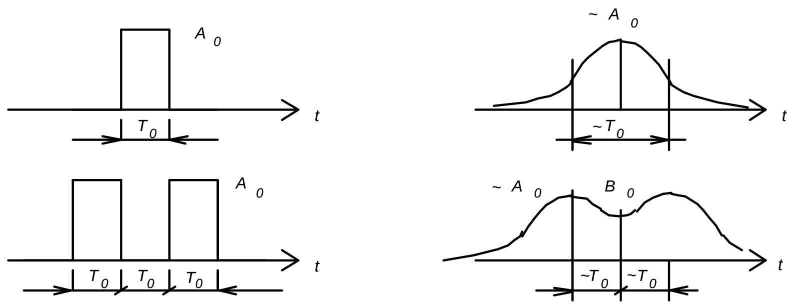


Fig. 2. Pulses Distorted by a Linear System

Sometimes the channel is used only in a narrow region near a frequency $f_o$ where

$$\left|H_c(f)\right| = A_0 \qquad \text{and} \quad \varphi(f) = \varphi(f_0) + 2\pi(f - f_0)\tau_0 \tag{6.3}$$

This is the case when the channel input signal is

$$x(t) = m(t)e^{j2\pi f_0 t} \tag{6.4}$$

where $m(.)$ is a slowly changing narrowband signal. Let $M$ be the Fourier transform of $m$ so that in the narrow band around $f_o$

$$X(f) = M(f - f_0)$$

and the signal at the channel output is then

$$Y(f) = X(f)H_c(f) = M(f - f_0)H_c(f)\,,$$

i.e.

$$Y(f) = A_0 e^{-j\varphi(f_0)} M(f - f_0)e^{-j2\pi(f - f_0)\tau_0}$$

The inverse transform is then

$$y(t) = A_0\, m(t - \tau_0)e^{j\left(2\pi f_0 t - \varphi(f_0)\right)} \tag{6.5}$$

Besides the amplification of the signal by $A_o$ and the shift of its harmonic factor by $\Phi(f_o)$, it is important to notice that the envelope of the harmonic signal $m(.)$ remained essentially undistorted but suffered a delay $\tau_o$. It follows from equation (6.3) that

$$\tau_0 = \frac{1}{2\pi}\frac{\partial\varphi(f)}{\partial f}\bigg|_{f - f_0} = \tau(f)\Big|_{f = f_0} \tag{6.6}$$

The function $\tau(f)$, $f \in (-\infty,\infty)$ which is the derivative of the phase characteristic is called *envelope delay* (or group delay). Generally, the envelope delay is a more illustrative term than the phase. If the envelope delay of the channel is frequency-dependent, the wideband signals may be significantly distorted. It is worth to remark that envelope delay of channels of a bandpass character has remarkable ripples at the edges of the passband.

*6.2.2. Echo and Reverberation*

Echo and reverberation are special kinds of linear time-invariant distortions which occur when the output signal is composed of several components of the input signal which have different delays and attenuations:

$$y(t) = \sum_i c_i x(t - T_i) \qquad (6.7)$$

This effect is often due to multipath propagation or caused by reflections from mismatched terminations. In the simplest case equation (6.7) has only two members and usually $a_o >> a_1$. Suppose that $a_o=1$ so that

$$y(t) = x(t) + c_1 x(t - T_1) \qquad (6.8)$$

In the frequency domain, this distortion corresponds to the following frequency response:

$$A(f) = 1 + c_1 e^{-j2pfT_1} \qquad (6.9)$$

which is periodical in $1/T_1$ so that the spectrum is periodically deformed by ripples having amplitude $a_1$. In the case of speech signals, if $T_1$>>50 ms, the delayed sound is perceived by the ear separately as an echo. If the delay is smaller, the sound is perceived as being one but of a particularly hollow sounding. In the case of video signals, the echo blurs the picture contours or produces a ghost picture.

The other usual form of reverberation is when the output signal contains components generated by multiple reflections:

$$y(t) = \sum_i c^i \cdot x(t\text{-}iT), \qquad |c| < 1 \qquad (6.10)$$

Speech or music is echoing in such a case.

*6.2.3. Time-Variant Linear Distortion*

It is also a usual condition that the transfer function of a channel cannot be supposed to be constant even for a short period of time. The simplest form of this case is when the gain (or attenuation) of the channel fluctuates:

$$y(t) = A(t)\ x(t) \qquad (6.11)$$

Even such a simple model enables us to set up and answer several interesting questions. Important parameters of such a type of interference having a multiplicative character are the rate and the extent of the changes of $A(.)$. If these are slow compared to the changing of $x$ and the amplitude varies just some few dB-s then this interference can be compensated relatively easily by automatic gain control (AGC). The real problem is caused by great (≥10 dB) and fast changes of $A$ which is typical for wireless communication.

A special type of the time-variant distortion is the so called *phase jitter* and *frequency shift*. This may happen typically when the output signal is a very particular transform of the input signal:

$$y(t) = x(t)\cos(m_t) - z(t)\sin(m_t) \qquad (6.12)$$

where $z(.)$ stands for the Hilbert transform of $x(.)$. (Hilbert transform is a linear distortion shifting the signal phase by $\pi/2$ rad at all frequencies.) Transformation given in (6.12) is time-variant because of the time dependence of $\mu_t$. The effect can be well illustrated when $x$ is a sinewave having frequency $f_o$. In this case

$$x(t) = \sin(2pf_0 t) \qquad \text{and} \qquad z(t) = -\cos(2pf_0 t)$$

so that

$$y(t) = \sin(2pf_0 t + m_t) \tag{6.13}$$

Several $\mu$ functions, having sometimes quite an unusual character might come about in real applications. If $\mu$ is a stationary process in the usual sense of the word then this effect is called *phase jitter*. Other important case is when $\mu$ varies linearly in time:

$$m_t = m_0 + \Delta t \tag{6.14}$$

In this case

$$y(t) = \sin\left(2p\,(f_0 + \Delta)t + m_0\right) \tag{6.15}$$

Note that all sinusoidal components are shifted by the same frequency $\Delta$. This shift essentially changes the signal shape, e.g. the transmitted data are so distorted that they even become unrecognizable. In musical signals, this leads to a particularly unpleasant sounding since the harmonic content characterizing musical sounds is significantly distorted even if $\Delta$ is small. In the speech signals the frequency shift is more acceptable since frequency offset of some few Hz does not degrade significantly the speech intelligibility.

### *6.2.4. Nonlinear Distortion*

Modelling of the systems by linear transformations is simple but sometimes imperfect. In more precise models nonlinear effects should be taken also into consideration. The simplest nonlinear models are memoryless, i.e. the output signal at an arbitrary time $t$ depends only on the input signal value in the same time:

$$y(t) = n(x(t)) \tag{6.16}$$

where $n(.)$ is a single variable function, usually continuous. *Saturation* and *dead zone* effect (see Fig. 6.3.) can be presented as typical examples of memoryless nonlinearities.

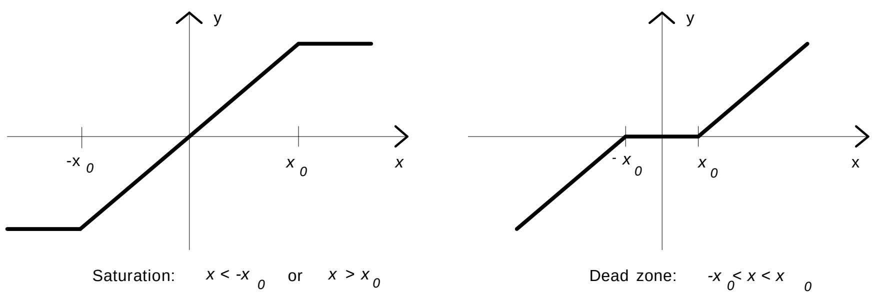


Fig. 6.3 Nonlinearity Caused by Saturation and Dead Zone

Function $n(.)$ describing the nonlinear behaviour can often be decomposed into Taylor series, more precisely it can be substituted by the first few members of the Taylor series:

$$n(x) = b_0 + b_1 x + b_2 x^{2+} + b_3 x^3 \tag{6.17}$$

To determine the effect of such nonlinearity, let us assume that the input signal is sinusoidal:

$$x(t) = U\cos(2pf_0 t)$$

The output signal will be then

$$y(t) = b_0 + b_1 U\cos(2pf_0 t) + b_2 U^2 \cos^2(2pf_0 t) + b_3 U^3 \cos^3(2pf_0 t)$$

or using trigonometric identities:

$$y(t) = b_0 + \frac{1}{2}b_2 U^2 + (b_1 U + \frac{3}{4}b_3 U^3)\cos(2pf_0 t) +$$
$$+ \frac{3}{4}b_2 U^2 \cos(2p\,2\,f_0 t) + \frac{1}{4}b_3 U^3 \cos(2p\,3\,f_0 t)$$

This simple example may serve for making some more general conclusions. Namely, it can be stated that because of nonlinear distortion, new sinusoidal components are generated in the output signal which were not present in the input signal (in the example above, $2f_o$ and $3f_o$ are such a components).

The amplitude of the fundamental harmonic is a nonlinear function of the input signal, and the amplitude of the harmonics are power functions of the input amplitude. These simple consequences enable us to characterize the nonlinearity by means of the 2nd, 3rd, etc. *harmonic distortion factor*, defined as the ratio of the corresponding harmonic amplitude to that of the fundamental component.

When examining the nonlinear behaviour of amplifiers, it is a common experience that increasing the input signal, the power of the components causing distortion starts to increase dramatically at a certain input level, thus indicating the start of the saturation. This effect can be used to define more precisely the overloading level of systems with higher complexity.

### *6.2.5. Additive Noise*

Generally, different interferences (crosstalk, thermal noise, man-made noise, etc.) influencing the output signal of the transmission systems have to be considered as being looped and non-linear. However, it is also usual that these effects can be collected as one common factor $\nu_t$ which is independent of the signal itself and can be simply added to it:

$$y(t) = x(t) + n_t \qquad (6.18)$$

Such a type of noise is called *additive* and can obviously be modelled by a stationary stochastic process. Generally, it is not possible to characterize the process by its distributions.

Sometimes $v$ is composed of several independent noise sources of approximately the same magnitude. In this case $v$ can be well approximated by a Gaussian process and the primary parameters of the process can be determined by secondary parameters (expected value, autocorrelation function). As a typical model, the process with zero-mean and constant spectral density over a wide range is used.

*Signal-to-noise ratio* which is the ratio of the powers of $x$ and $\nu$, is usually a good parameter to characterize the influence of the additive noise from the point of view of the sink:

$$\frac{S}{N} = \frac{P_x}{M(n_t^2)} \qquad (6.19)$$

The value of the signal-to-noise ratio is usually given in the logarithmic scale in decibels as

$$SNR = 10\lg(S/N) \quad \text{[dB]} \tag{6.20}$$

To characterize the signal and noise intensity, it is also convenient to introduce their power in logarithmic scale. The *absolute power level* of the signal $S$ is defined as

$$s_{\text{signal}} = 10\lg(S/S_0) \tag{6.21}$$

where $S_o$ is a reference power (usually 1 mW). Signal-to-noise ratio is then given as the difference of the signal and noise power level in dB:

$$SNR = s_{\text{signal}} - s_{\text{noise}} \quad \text{[dB]} \tag{6.22}$$

## 6.3. Digital Channels

Before discussing the properties of digital channels, we have to deal with characteristic features of digital sources and how the discrete sources are encoded to a digital signal acceptable by the channel.

### *6.3.1. Symbol Series as Information*

It is a frequent task to replace elements of a finite set of symbols by another set of symbols, e.g. to convert a text into a series of 0's and 1's. In the model of such a task, the elements of the set to be converted are called *source symbols* and the conversion procedure is called *coding*. The properties, possibilities and limits of coding are examined by the information theory, specifically by the source coding theory. This theory is motivated by the fact that the extent of the coded text, the *coding density*, is by far not indifferent for the user. To be able to define this question we have to characterize the source of the coded symbols.

The essential parameter of the source is the set of its symbols, the source alphabet. The source is well defined by listing all its symbols, e.g. $a_1$, $a_2$, ... $a_N$. The source message is understood as a series of symbols to be encoded. These may be so long that they even cannot be taken into consideration in source modelling; in this case we are speaking about infinite series of symbols. It is not a too good description of a real source but it is a simple and well handled model if the source symbols listed in message are supposed to be independent and to have equal random distribution. In this case the source is called *stationary* and *memoryless* and it is fully characterized by the source symbol distribution, i.e. the system of probabilities

$$p_k = P(y_i = a_k), \quad i = 0,\ 1, ... ;\ k = 1, 2, ... N$$

### *6.3.2. Coding Density*

Let us examine the binary coding of a source of $n$ elements with the distribution $P = (p_1, ... p_N)$. Obviously, the source symbols can be represented by series of 0's and 1's of the length $k$, if $2^k \geq N$ so that the code could be unambiguously decoded. Greater coding density can be achieved if codewords of different lengths are used. It is relatively easy to decode such codes for which one can decide after reading a certain number of code bits whether they form a valid code or not. This type of code is called the *prefix(free)* code.

Obviously, to be able to decode a code unambiguously, the code lengths $l_i$ ($i$ = 1, 2, ... $N$) must be longer than a certain minimum length. This relation is given by the Kraft's inequality stating that a code can be unambiguously decoded if it is true for the code lengths that

$$\sum_{i=1}^{N} 2^{-l_i} \leq 1. \qquad (6.23)$$

From the point of view of the total length of the message, the expected value of the codeword length is of importance. If the probability of an $l_i$ long code word is $p_i$ then the expected value of the code word length is

$$\lambda = \sum_{i=1}^{n} l_i\, p_i \qquad (6.24)$$

Obviously, it is reasonable to assign short code lengths to symbols with great probability and vice versa. It can be proved that for any unambiguous code

$$\lambda \geq \sum_{i=1}^{n} p_i \,\mathrm{ld}(1/p_i) \qquad (6.25)$$

It is also true that choosing $\mathrm{ld}(1/p_i) \leq l_i < \mathrm{ld}(1/p_i) + 1$, Kraft's inequality is satisfied so that a code exists for which

$$\sum_{i=1}^{n} p_i \,\mathrm{ld}(1/p_i) \leq \lambda < 1 + \sum_{i=1}^{n} p_i \,\mathrm{ld}(1/p_i) \qquad (6.26)$$

There is a particular function of $p_i$ probabilities, a feature of source distribution playing role in the limits obtained for average code length. This feature is called *source entropy*:

$$H(P) = \sum_{i=1}^{n} p_i \,\mathrm{ld}(1/p_i) \qquad (6.27)$$

### *6.3.3. Coding of Symbol Series*

If the source entropy is relatively small (≅ 1 bit/symbol) then the limit (6.26) given for the average code length is not too strict. For the practical point of view it is not indifferent whether the average length of a code is closer to the shortest or to the longest code length.

The limit given by equation (6.26) ensures the existence of highly efficient code in the case of high entropy sources. It is possible to create a source equivalent to the original one but having much higher entropy. As the source symbols, let us take the $K$ symbol messages of the original source. This extended source will have $n^K$ symbols and its entropy will be $KH(P)$ since the source is memoryless. So that for the average length of the best code representing the original $K$ set of symbols, the following inequality can be given:

$$KH(P) \leq \lambda^{(K)} < 1 + KH(P)$$

The average code length assigned to a single symbol of the original source can thus arbitrarily approximate the entropy of the original source.

The source entropy thus fully determines the possible coding density of the source and in this sense it is characteristic for the information content of the source messages. The length of the bit series of the code matching the source best may serve as a measure of the information content of the source messages.

### *6.3.4. Memoryless Channels*

Digital channels are systems capable to accept the previously defined $N$ different source symbols ($a_1$, $a_2$, ..., $a_N$) and -generally- producing $M$ different output symbols ($b_1$, $b_2$, ..., $b_M$). Such a channel can also be characterized by the rhythm the symbols are received and generated. This parameter is called the *symbol rate* ($v_s$). As $N$ symbols can generate bit series in length ld($N$), the so-called ***data transfer rate*** (or bitrate) is

$$n_{\mathrm{data}} = n_{\mathrm{s}}\,\mathrm{ld}(N) \qquad (6.28.)$$

Generally, it is not warranted that the channel output symbols characterize unambiguously the input symbols. However, it is often true that the output symbol $h$ depends solely on the actual input symbol $x$ and the instantaneous 'caprice' of the channel.

A channel is said to be ***memoryless***, if its output is independent of previous symbols and of the response to those symbols. Memoryless channel is characterized by conditional probabilities or the so-called ***system of transitional probabilities***:

$$p_{ij} = P(h = b_i | x = a_j), \quad i = 1,2,...,N_{\mathrm{out}} \qquad j = 1,2,...,N_{\mathrm{in}} \qquad (6.29)$$

A typical example of a memoryless channel is the binary symmetric channel (BSC), symbols of which can have two values, e.g. 0 and 1. Transitional probability is characterized by a single date, $p$:

$$p_{01} = p_{10} = p \qquad p_{11} = p_{00} = 1 - p$$

For channels with the same set of input and output symbols, the transmission errors can be evaluated by the so-called ***probability of error***. This term, denoted as $P_e$ means the probability of the event that the channel output signal is not equal to the input symbol:

$$P_e = P(h \neq x) \qquad (6.30)$$

Obviously, the probability of error depends on the probabilities with which the source generates the individual input symbols. If $p_i$ is the probability of sending the $i$th symbol then

$$P_e = \sum_{i=1}^{N} p_i P(h \neq a_i | x = a_i) = \sum_{i=1}^{N} p_i (1 - p_{ii}) \qquad (6.31)$$

It is interesting that for BSC the probability of error is independent of the source distribution: $P_e = p$.

There are several practical cases when the probability of error gives sufficient information about the usability of the channel. This is the case when the probability of error is small and the channel is used e.g. for transmission of coded speech. Occasional channel errors cause additional noise in the reconstructed speech signal but this may be tolerable and does not necessarily degrade the quality of the provided service.

The situation is quite different when the same channel is used e.g. for copying a computer program. If there is just one faulty bit in the copied program code, the program might become completely useless. In such a case it is necessary to recognize the errors caused by the channel and to correct the faulty symbols.

### *6.3.5. Principles of Error-Detection*

Let us divide the source symbols(bits) into consecutive blocks each of *k* symbols(bits) and assign a supplement of *n-k* symbols(bits) to each block according to some appropriate rule. The blocks of *n* symbols(bits) are transmitted through the channel and checked in the receiver whether the relations between the first *k* symbols(bits) and the remaining *n-k* symbols(bits) match the defined rule. If the rule used for the supplement generation is well chosen, not only can we detect the errors but also deduce which symbols(bits) are defective. Of course, it is not indifferent how much the original message has to be lengthened to detect and correct the faulty symbols(bits), since the transfer rate is decreased by the factor of *k*/*n*. The efficiency of error-correction is limited by the transition probabilities of the channel. This problem belongs to the coding theory and will be discussed in Chapter 7.

### *6.3.6. Capacity of Binary Symmetric Channels*

Let us determine the way and efficiency of error-free transmission via a binary symmetric channel having probability of error *p*! Suppose that the examined model is optimal; i.e. it consists of the above channel and an ideal backward channel informing about the received message (e.g. we are working on an unreliable keyboard but we can see the output on the display).

As a first step, let us send a message of $n_0$ bits, out of which $p \cdot n_0$ bits will be probably damaged. The simplest error-correcting message could be a series of $n_0$ bits, containing '0'-s on correctly received bit positions and '1'-s on faulty bit positions. In this case, the error-correcting message will have the distribution of  bits as follows:

$$P: \quad p_0 = 1 - p \qquad \text{and} \qquad p_1 = p\,.$$

Let us use a more efficient source coding for the error-correcting message!  Using the optimum source coding  (see Chapter 6.2.3), the length of the error-correcting message will be only $n_1 = n_0 H(P)$ bits. Of course, this message might also contain faulty bits in some positions but we can send again and again similar error correcting messages until all errors are corrected. Thus, for the error-free transmission of all *n* bits a total of

$$\sum_{i=0}^{\infty} n_i = n_0 + n_0 H^2(P) + \ldots = n_0 \frac{1}{1 - H(P)} \tag{6.32}$$

bits will be needed. Therefore, efficiency of the coding is

$$C = \frac{k}{n} = \frac{n_0}{\sum_{i=0}^{\infty} n_i} = 1 - H(P) \tag{6.33}$$

Obviously, efficiency will further be reduced if the error correcting messages are to be 'built into' the original message in advance. However it can be proved that if $k/n < C$ always such a code exists for which the probability of erroneous evaluation of blocks goes to zero if $n \to \infty$.

The above example can be generalized for even more complex channels. With certain codes the probability of error can be made arbitrarily small provided the coding efficiency *k*/*n* is smaller than the capacity, a limit determined by the channel.

## Control Questions

1. What is the difference between distortion and noise?
2. What are the conditions for a signal not to be distorted by a linear (time-invariant) distortion?
3 When is it reasonable to characterize the transmission quality by the signal-to-noise ratio?
4. What is coding density limited by?
5. Why is it more advantageous to encode series of symbols instead of encoding single symbols?
6. What are binary symmetric channels?
7. What is channel capacity and in what sense does it limit transmission efficiency?

## Exercises

1. A lowpass filter having bandwidth of $B$ [Hz] is tested by the double impulse given in Fig. 6.2. Estimate the value of $D$ if $B = 1/T$.
2. Determine the capacity of a binary erasure channel. The channel has three possible outputs: 0, 1 (representing the input symbols) and $x$ which stands for a non-readable symbol.
3. What is the entropy of the probability distribution $p_i = 2^{-i}$, $i$ =1,2, ...?

## References


[1] Gallager, R. G.: Information Theory and Reliable Communication. Wiley, New York, 1968.
[2] Csibi S. et al: Információ közlése és feldolgozása. Tankönyvkiadó, 1986.
[3] Gyõrfi L.-Vajda I.: A hibajavító kódolás és a nyilvános kulcsú titkosítás elemei. TUB lecture notes, 1991.