FUNDAMENTALS OF TELECOMMUNICATIONS

Prof. Michele Luglio
(realized on the basis of notes authored by Prof. Francesco Valdoni)

TABLE OF CONTENTS

1 INTRODUCTION

The course aims to provide the basic elements on the mathematical instruments which are useful to define the **signal** and for its elementary processing that is necessary in order that it could transfer information from one end to the other of a link or of a network.
Such instruments span from the basic functions and basic operations of the mathematical analysis, to the domain transformations, to the stocastic process theory. With them it is possible to implement signal shaping operations, to characterize the ideal, perfect and noisy transmisison and to implement the basic processing for the elementary analogue modulation.
Being the course Fundamentals of Telecommunicaitons is inserted in the path of **Internet Engineering**, in this short introduction it is useful and worth to provide also a vision which includes the role and underlines the importance of the signal.

1.1 TELECOMMUNICATION SYSTEMS AND SERVICES

Communication is the process by which **information** flows between two or more parties; they are people, physical equipment or logical entities capable of emitting or receiving: in the former case it is a **source** of information, while in the latter a **recipient** of information. Obviously, the same subject can simultaneously perform both functions, sending or receiving, appearing both as a source and a destination; however, the two functions can be, and will be considered separately.
If the nature of the end users is the same and the source is not of electric type, the remote communication generally begins with a **transducing** operation, that is, with the transformation of the output variable of an information source in an electromagnetic magnitude function of time, called **signal** and indicated with x(t); the communication ends then with a further transduction operation, complementary to the initial one, through which the received signal is converted into a proper variable accepted by the recipient. The transduction can only occur at a termination. It is out of the scope of this course: in fact, these operations reside in complex physical entities, called terminals, capable of emitting or receiving information in electromagnetic form, i.e. signals, which are the only focus of thus course. An emitting terminal, or a signal source that conveys the information, will be referred to simply as the source; a receiving terminal, i.e. a signal recipient, will be referred to simply as the recipient.
When the information flow among a plurality (at least two) of remote terminals is realized utilizing propagation phenomena, to cover the required distance, of a signal in the electromagnetic environment, the **telecommunication** occurs. The minimum distance necessary to set up a telecommunication process depends on the type of service. In fact, it is the needed distance to the utilized electronic devices for the information (data) exchange, spanning from a few millimeters in local environment (PC, printer, smart phone) or even several kilometres for long distance connections. The set of terminals, of the circuits and electromagnetic environments in which they propagate signals containing the information constitutes a **telecommunications infrastructure** (see Figure 1.1). In the most elementary case it is reduced to a one-way connection between source and recipient; when users become many and the possible interconnections necessary for full connectivity should grow numerically as the factorial, the **telecommunications network** arises, namely an infrastructure for the most part shared (branches and nodes) and in part for the exclusive use of the individual user (cable or radio link “last meter” and terminal equipment); therefore, more generally, there will be a large number of terminals connected to a very complex structure, divided into nodes, loops and branches and called **telecommunication network** (see Figure 1.2), which allows simultaneous execution of a plurality of communications with links variable as a function of time.

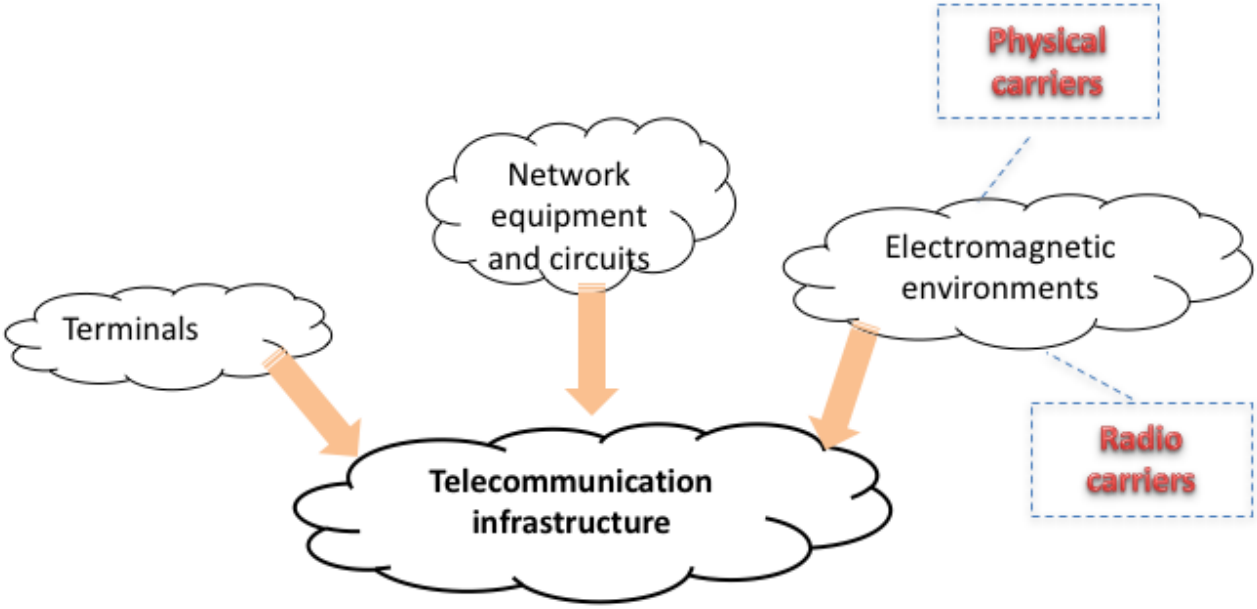


Figure 1.1: Telecommunication infrastructure

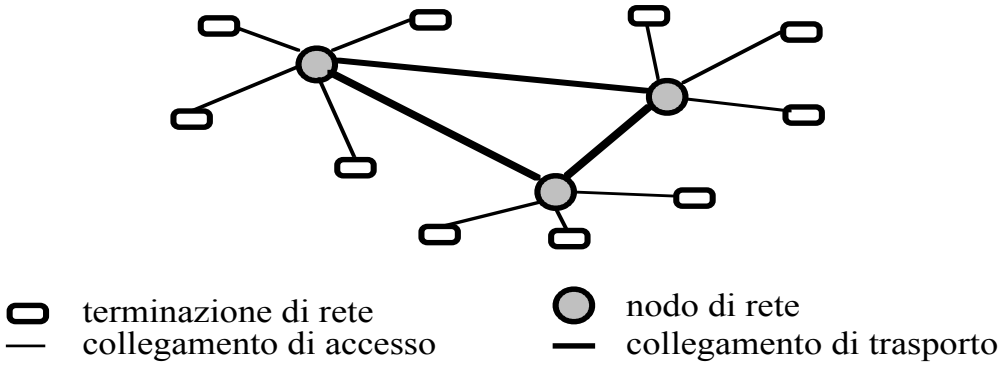


Figure 1.2: Telecommunication network general scheme.

The concepts of telecommunication network evolves to the telecommunication system if in addition to the infrastructure we consider the processing on the signal to make the transmission possible, the operations to allow information to travel and mainly services abd applications, which represent the interface between the system and the users and without which anything else wouldn't make sense. In fact, the user utilise the available tools of the system because they need to exchange the information created by the applications that in turn represent the operational modality to enjoy the services. Figure 1.3 shows a scheme of the just expressed concept.
Often the networks have the ability to operate with bi-directional mode, i.e. for each communication they make available a pair of links, in which the signals propagate in opposite directions; obviously, the connected terminals are then able to simultaneously operate both as sources and as recipients.
The manner a communication process is performed, with standardized procedures at international level, constitutes a **telecommunication service**; it is enjoyed by the terminals, so even named users of the service. The exploitation of a telecommunication service requires the implementation of a multiplicity of functions, some oriented to the processing of the information for its utilization, other rather oriented to its distance transport, namely the **transfer** of information. Limiting the attention to the latter, a **carrier service** occurs; it is however necessary, together with the functions of the other type, to provide full communication capabilities, namely a **teleservice**. The bearer services are thus offered by the telecommunication networks for the transfer of information between the **network terminations**, which are the points of the network interface towards the terminals.

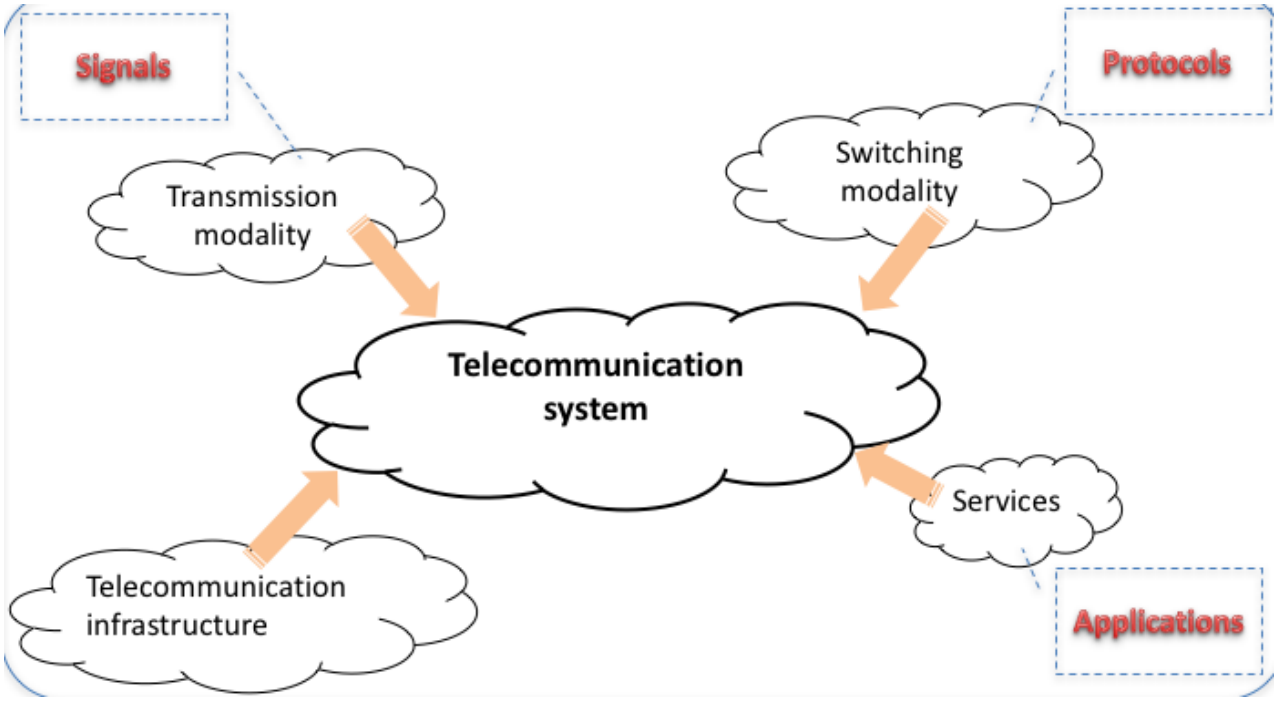


Figure 1.3: Telecommunication system

The network performs basically **switching** and **transmission** functions: the former, residents in the diagram nodes in Figure 1.1, allow the forwarding of information to implement the desired connection, while the latter allow the propagation of the signal that carries the information, among the network terminations and nodes or among nodes, respectively, with **access links** and **transport links**. Other relevant **signalling** and **management** functions complete the procedures for a telecommunication network: they allow the performance in real time and optimized the many services required.
Table 1.1 shows the main definitions concerning the concepts at the basement of a telecommunications system.

Table 1.1: Main definitions

Communication	→	Need (can develop particular requirements)
Distance	→	Condition
Information	→	Content
Transfer	→	Scope
Signal	→	Vehicle
Rules	→	Implementation and circulation modality (Internet)
Services	→	Aim and modality of process development
Applications	→	Service realization modality Tools that generate information
Applicative	→	Product which implements an application
Infrastructure	→	Physical locations where tlc processes develops

Hereinafter the attention will be focused solely on the transmission, leaving entirely out the other issues of networks and terminals for considering only the relevant aspects concerning transmission. It will also be considered one way transfer of information, considering that the operation in the opposite direction can take place in a similar manner.
Nevertheless, just to understand the role and the importanve that signals and transmission have in the telecommunication world, a very brief introduction on the concept of Internet will be provided.

1.2 Internet

Internet can be defined as an INFRASTRUCTURE over which APPLICATIONS runs to generate INFORMATION which are transferred by means of SIGNALS that travel over the INFRASTRUCTURE. The previous (circular) statement can be expressed also in the opposite direction i.e INFRASTRUCTURE over which SIGNALS travels to bring INFORMATION which are generated by APPLICATIONS that work over the INFRASTRUCTURE (see Figure 1.4).
Thus, Internet, to exist and to properly work, needs all the aspects component of the Information and Communication Technology (IC) ed in particular of the Telecommunications, for what concerns the infrastructure and the signals, and of the Computer Science, for what concerns the development of applications and the generated information.

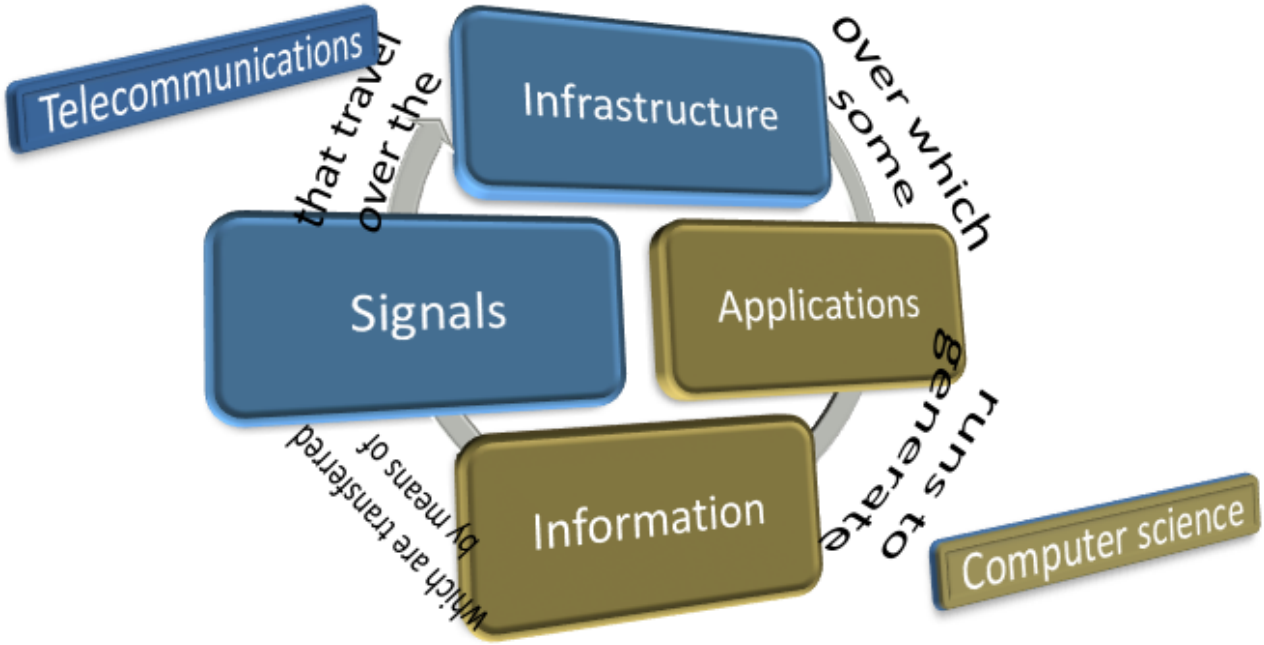


Figure 1.4: Circular vision of Internet

1.3 Ideal transmission of a signal

1.3.1 Definition of signal in strict sense

In strict sense, a **signal** is a physical quantity, real function of the independent real variable time, usable as a vehicle of the information that flows from a source to a recipient. In order to actually carry information it is necessary that the characteristic parameters of such physical entity are unknown to the recepient.

1.3.2 Ideal transmission

From external behaviour point of view, a generic infrastructure dedicated to the transmission of information from a source to a recepient, i.e. a generic **transmission system** utilized to exploit the link, can be always represented through a single block that, excited by the incoming signal x(t), which properly conveys the information, provides as output a response signal y(t). The transmission is assumed to be ideal if the following condition is verified:

$$[1.1]\ y(t) = g\,x(t - t_0),$$

with g and t_0 real contants, i.e. if the output signal is a replica of the input one, unless the multiplication factor (amplification/attenuation) g and the delay $t_0>0$. Such quantities have a

meaningless nature: in fact, they are determined, although not a priori known in reception, and therefore can not carry information, that must always be associated with random parameters. The relation [1.1] is a sufficient but not necessary condition in order to fully transfer correct information: in fact, the particular way to convey it through the signal x(t) can allow further alteration of the response y(t) with no consequences on the utilization of the received information.
The above mentioned relation between the outcoming and incoming signals globally summarises the processing for distance transmission that guarantees the perfect performance of the link.
Actually, to understand the most important elements it is better to represent the system with more details, considering several functional blocks, connected in cascade; part of them, sometimes reduced to just one block as shown in Figure 1.5, has one physical dimension as long as the distance to cover is converned, while the other can be assumed as zero dimension. The blocks of the first cathegory ideally represent the **transmission means**, main responsible of the distance propagation; the blocks of the second cathegory are conceptually equivalent to **transmission equipment**, usually including active electronic or optoelectronic circuits that can be divided in emitting equipment (incoming side) and receiving equipment (outcoming side).

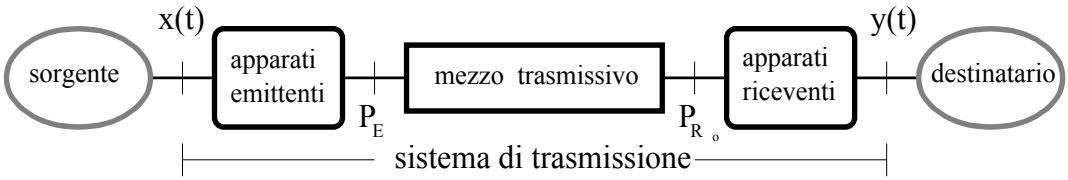


Figure 1.5: Essential scheme of a transmission system that realizes a link.

2 SIGNALS IN TIME DOMAIN

2.1 GENERALITIES ON SIGNALS

2.1.1 Complex notation

Referring to a signal located in a generic section of a transmission system, it assumes the following notation:

$$x(t) \quad , \quad \forall\, t \in T \quad , \quad x \in X,$$

where the real function x(t), with an example shown in Figura 2.1, is defined over the time domain T, often extended to the whole time axis, and the signal codomain (range) X, composed of all the values assumed by x(t), has electric nature.

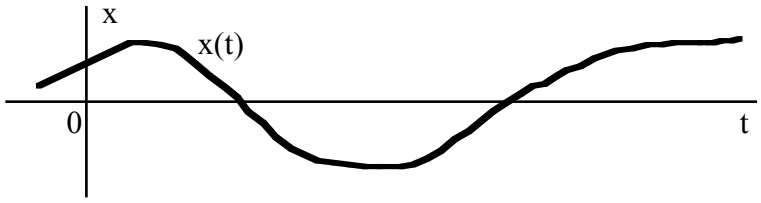


Figura 2.1: Example of real signal.

In wide sense a signal can be whichever pysical quantity representing a physical signal, but generally complex function of the time independent variable, expressed with the notation in real and imaginary part:

$$[2.1]\ x(t)=x_R(t)+jx_I(t) \quad , \ \forall\, t\in T \quad , \quad x_R \in X_R,\ x_I \in X_I,$$

where $j^2=-1$, X_R+jX_I is the complex codomain (range) of the signal and:

$$[2.2]\ x_R(t)\hat{=}\Re\{x(t)\}=\frac{1}{2}\left[x(t)+x^*(t)\right],$$

$$[2.3]\ x_I(t)\hat{=}\Im\{x(t)\}=\frac{1}{j2}\left[x(t)-x^*(t)\right],$$

having indicated with $x^*(t)$ the complex conjugate function of x(t). With the complex notation in modulus and argument it can be expressed as:

$$[2.4]\ x(t)=|x(t)|e^{j\arg[x(t)]},$$

where

$$[2.5]\ |x(t)|\hat{=}\sqrt{x_R^2(t)+x_I^2(t)}=\sqrt{x(t)x^*(t)}\ ,$$

$$[2.6]\ \arg[x(t)]\hat{=}\operatorname{artg}\left[\frac{x_I(t)}{x_R(t)}\right]+\frac{\pi}{2}\left[1-\frac{x_R(t)}{|x_R(t)|}\right].$$

Since the arctangent function provides a value included between $(-\pi/2, \pi/2)$, the second term in the second member of the previous expression is necessary to achieve the value of the argument in the whole round angle.
For a real signal the following well known relation is valid:

$$[2.7]\ x(t)=x^*(t) \quad , \ \forall\, t\in T \quad , \quad \text{for } x_I(t)\equiv 0,$$

while for a purely imaginery signal:

$$[2.8]\ x(t)=-x^*(t) \quad , \ \forall\, t\in T \quad , \quad \text{for } x_R(t)\equiv 0.$$

2.1.2 Basic operations and faithful signals

Being characterised by mathematical functions, it is possible to perform any mathematical operation, *on* the signals and *among* signals, allowed for the respective functions. In particular, on a generic signal x(t) all the elementary operations of multiplication by a constant and time shift, shown in Figura 2.2 are accomplishable. As above mentioned, once identified a physical real signal $x_0(t)$ in a transmission system, any other signal:

$$[2.9]\ x(t) = g x_0(t - t_0),$$

different from $x_0(t)$ only for a constant scale factor g, real and positive or negative, and/or for a time shift, in delay ($t_0>0$) or ahead of time ($t_0<0$), is assumed to be equivalent from transfer of information point of view. Thus, the elementary operations of multiplication by a real constant and time shift don't have substantial effect on the communication; the real quantities of the family defined through [2.9] are called ***faithful signals*** in strict sense.

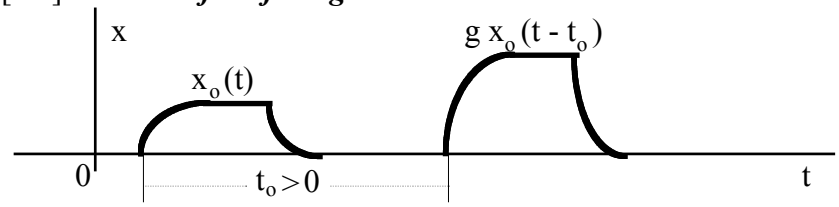


Figura 2.2: Real signal $x_0(t)$ and its faithful replica $g\, x_0(t - t_0)$.

If a complex signal is considered, the family of ***faithful signals*** in wide sense can be more broadly defined through the following expression:

$$[2.10]\ y(t) = g\, e^{-j\Gamma}\, x(t - t_0),$$

with Γ arbitrary real constant, considering elementary the broader operation of multiplication by a complex constant.

Considering other simple operations, it is easy to verify that the conjugation one has no effect if applied to a real signal; otherwise, it provides the signal $x^*(t)=x_R(t)-jx_I(t)$, different from x(t) and usually not faithful. Also the operation of time axis inversion, obtaining x(-t) from x(t), provides in general a different and not faithful signal (see Figura 2.3a); nevertheless, in case of real signal with symmetric pattern an exception can occur (as the example shown in Figura 2.3b).

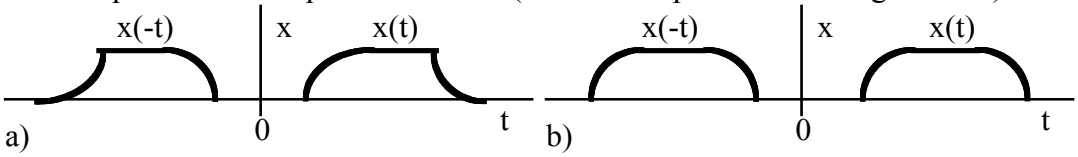


Figura 2.3 Real signal x(t) and signal x(-t) obtained by time axis inversion.

Among signals the elementary operations of sum, subtraction and multiplication are also applicable (executed point by point). Moreover, on the signals it is possible to apply exponentiation, square root, extraction of root, logarithm, derivative, integration, etc. Finally, the operation of convolution between two signals is defined as follows:

$$[2.11]\ x(t) * y(t) = \int x(v)\, y(t-v)\, dv\,.$$

where v is a time variable.

The meaning, the importance and the application of such an operation will be illustrated later on.

The ***convolution integral*** between two time continuous signals, in general complex, indicated with the notation $x(t) * y(t)$, enjoys the associative, distributive and commutative properties, which are typical of the usual product operation; in particular, the order of the two signals is not significant, so that:

$$[2.12]\ \int x(v)\, y(t - v)\, dv = \int y(v)\, x(t - v)\, dv\,.$$

2.2 Classification of signals

2.2.1 Time continuous and time discrete signals

A signal can be defined as continuous or discrete depending on the domain or codomain (range) characteristics of the function that represents it.
The signals considered so far were defined in every point of a time domain T, even over the whole time axis: they belong to the class named ***time continuous signals***.
Nevertheless, there are signals defined only in discrete time instants of the time variable, which belong to the class of ***time discrete signals***. Altough they have a just conceptual nature, they are very important in studying telecommunication systems, particularly considering the even greater importance of digital transmission techniques. The detailed characterization of each class will be dealt with separately. For sake of completeness and to keep generality the following definitions are used:

- If the domain is a continuous set (limited or unlimted), then the signal is time continuous
- If the domain is a discrete set (limited or unlimted), then the signal is time discrete
- If the codomain (range) is a continuous set (limited), then the signal is continuous
- If the codomain (range) is a discrete set (limited), then the signal is discrete or quantized.

The case of discrete domain corresponds to a signal existing only in some instants and between them it is not defined (it is NOT zero).

2.2.2 Deterministic signals and random signals

2.2.2.1 Deterministic signals

Often ***deterministic*** or ***definite signals*** are considered, belonging to the ***deterministic class***, characterized by evolutions that are ssumed fully known over the whole time axis; thus, in case they are real, they can be represented through a graph over the plane {x, t} with a single trace (sketch) or waveform (see Figura 2.1, Figura 2.2, Figura 2.3), or in case they are complex with a pair of waveforms over the planes $\{x_R(t), t\}$ and $\{x_I(t), t\}$ (or over the planes {|x(t)|, t} and {arg[x], t}).
Sometimes, deterministic signals can be represented through simple mathematical representations, as for example that valid over the entire time axis of the generic real harmonic signal, or simply ***real harmonic***:

$$[2.13]\quad x(t) = A\cos(\omega_0 t + \varphi_0),$$

where $\omega_0 = 2\pi f_0$ is the pulsation, f_0 is the frequency, $T_0 = 1/f_0$ is the period, A is the amplitude and φ_0 is the phase of the particular harmonic.
The deterministic signals are not considered in the operation of transmission systems because, being known by definition, their transfer would be useless but they are largely and effectively utilized in the analysis and design phases. Thanks to the full predictability of their expected evolution in ideal conditions, some kind of deterministic signals with easily measurable waveforms are utilized also in the performance validation and test phases of equipment or systems, evaluating the deviation with respect to the expected evolution in reception.

2.2.2.2 Random signals

In the operation of a transmission system ***random signals***, belonging to the ***stochastic class***, are utilized, with temporal evolutions never fully known in their individuality; in fact, the received signals are a combination of the useful signal with signals of the same nature (disturbs) produced either by the environment (transmission channel) or by the circuits through which the signal passes; in fact, as concerns the useful signals it is observed that the information resides in the occurrence of random events, namely uncertain a priori at the receiving side (otherwise the transmission would have no sense), while the disturbs are intrinsically unknown (considering how

they are generated) and it is obvious that they would no longer be really harmful if they were known and, consequently, removable. However, it is possible to specify some quantities characterizing the set of all the potentially and randomly generated signals from a same source and that, therefore, belong to the same generation of ***random process***; individually these signals are called ***realizations*** (***samples***) of the process.
It is possible an even more concise description in the condition of ***stationary process***, i.e. if the fundamental causes that generate the information or disturbs and the mechanisms which then form the corresponding signals remain unchanged in time. For example, if the particular source constituted by a speaker at the microphone are fixed, it is not possible to know the output waveform on the entire time axis, but it is possible to identify some common characteristics of all the audio signals that the considered source has the capability to emit, especially if the conditions of the sound information source or microphone or the environment don't change.
The random processes can be divided in isolated event processes, continuous processes and discrete processes.
In the first case the randomness is not associated to the values of the temporal evolution of the signal, but to one or more parameters that can randomly assume different values in the various realizations, of the continuous-time type, constituting the set of those that can be generated by the considered process. Once the source has determined with an isolated event some particular values of the random parameters, the waveform is completely known throughout its development, that implies that the emitted signal becomes certain; its transmission may, however, be useful, as it allows to know remotely the information of the isolated event, present in the particular values, or ***determinations***, set by the source for the random parameters. Analytically an isolated event process can be expressed with an explicit function of time, which contains the parameters in the form of ***random variables***, often indicated with the acronym r.v. A simple example is offered by a rectangular pulse, in which the parameters are the amplitude A, the duration T and the presentation time t_0; once the determinations of such r.v. are known the entire realization is known, as shown in Figura 2.4.

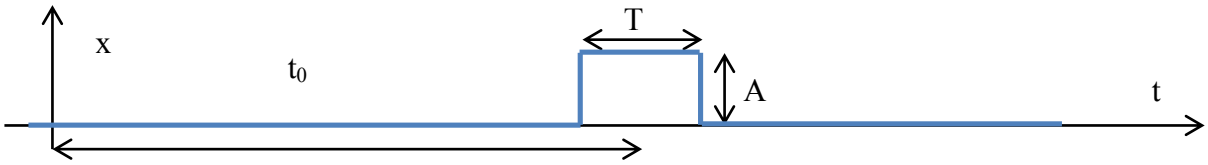


Figura 2.4: Example of realization of random process with isolated event.

In a **continuous process** the randomness is manifested in dynamic form for the effect of a series of random events distributed over the time, so that the knowledge of the waveform even starting from very distant times never allows the predictability of future evolutions. The most typical example is offered by the Gaussian noise, due to the emission in random moments of a very large number of waveforms of very short duration and small amplitude, that added together they give rise to realizations in which, up to the extreme limit, the knowledge of the entire past track has no influence on the present value. Generally, it is not possible to represent the process by means of explicit functions of time, while it is possible to use probabilistic descriptions, which allow to identify synthetic quantities characteristic of all the realizations, through a hierarchy of probability functions.
In a **discrete process** the randomness is still manifested in a dynamic form, but with essentially logical nature: the process is in fact constituted by a succession of random events distributed in discrete instants over the time axis. In the simplest case it can be represented mathematically by a series of unique r.v., so that the realizations (samples), typical discrete-time signals, are constituted by the corresponding series of determinations, i.e. the particular values that the source has determined for all individual random events. As an example you can consider the elementary case of a digital message source, which randomly emits at regular intervals one of the two possible binary digits 0 and 1; it has then a realization of the type:

....01101000101011101001011100....

The probabilistic description of the process is possible characterizing the succession of the mentioned r.v.

A continuous process can be conceived starting from a time succession of random events distributed in discrete instants, i.e. by a discrete process, but emitting for each event a time continuous signal typical of isolated events process; then, the process can be expressed analytically through a time series of explicit functions of time, containing one or more parameters in the form of a r.v. In a simple example the explicit functions are rectangular pulses, with fixed duration T, and regularly spaced presentation instants, with amplitudes corresponding to a succession of r.v.; Figura 2.5 shows a realization of the considered process. The probabilistic description of the process is based on that of the corresponding discrete process.

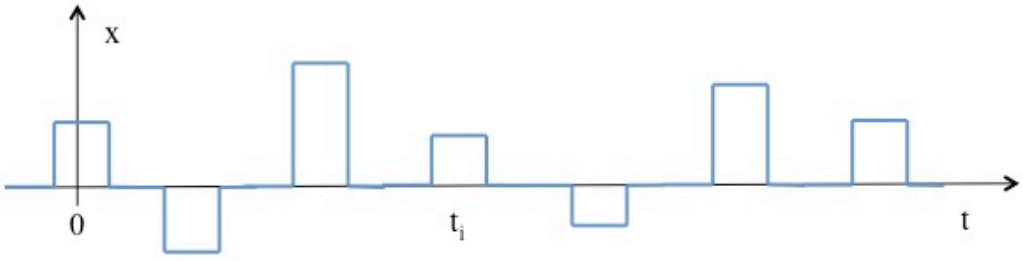


Figura 2.5: Example of time continuous process realization expressed through a time series of processes with isolated events.

2.3 TIME CONTINUOUS SIGNALS

2.3.1 Continuous signals

The time domain of a time continuous signal is a continuous set, which, unless differently specified, will be identified with the whole time axis. Such a signal x(t) is **regular** in the generic instant t_0 if the limit for t which tends to t_0 of x(t) exists, finite or infinite; the signal is then continuous in t_0 if it is verified that:

$$[2.14]\ \lim_{t\to t_0} x(t)=x(t_0).$$

If the previous expression is satisfied in every point of the time domain, it is a **continuous signal**. The real harmonic in [2.13] is an example of such kind of signals.

Due to the non ideal chatacteristics of devices and components utilized in real transmission systems, signals utilized in practical systems, deterministic test signals or realizations of random processes, are all continuous. Nevertheless, sometimes there are waveforms with extremely steep segments, which look, with good approximation, to have the evolution of non continuous signals, taken as reference in the theoretical analysis and system design phases.

2.3.2 Signals with discontinuities

2.3.2.1 Discontinuity in time continuous signals

A time continuous signal x(t) which doesn't converge in t_0 shows a ***first type discontinuity*** in that instant if its evolution is regular both on the left and on the right side of t_0, that means:

$$[2.15]\ \lim_{t\to t_0^-} x(t)=x(t_0^-),\ \lim_{t\to t_0^+} x(t)=x(t_0^+),$$

but the two limits are different with non zero value of the discontinuity

$$[2.16]\ d_0 \hat{=} x(t_0^+)-x(t_0^-)\neq 0\ .$$

For the non continuous signal, in t_0 the ***half value*** is defined:

$$[2.17]\ x_{em}(t_0) \hat{=} \frac{1}{2}\left[x\left(t_0^+\right)+x\left(t_0^-\right)\right] = x\left(t_0^-\right)+\frac{1}{2}d_0 = x\left(t_0^+\right)-\frac{1}{2}d_0 .$$

A discontinuous signal, as above mentioned, is very useful (but only for theoretical approach) and can present more than one first type discontinuity.
Broadening the previous considerations to the derivative of a time continuous signal, it may be possible to highlight one or more ***second type discontinuity***; if also the derivative is continuous, ***third type discontinuity*** may occur observing the second order derivative of the signal.

2.3.2.2 Examples of signals with discontinuity

The first example of discontinuous signal is the ***unit step***, with waveform shown in Figura 2.6a and indicated with the notation:

$$[2.18]\ u(t) \hat{=} \begin{cases} 0 \ , \ \forall \ t<0 \\ \frac{1}{2} \ , \quad t=0 \\ 1 \ , \ \forall \ t>0 \end{cases} .$$

The signal u(t) has the half value ½ in the discontinuity in the origin; it has the following property over the entire time axis:

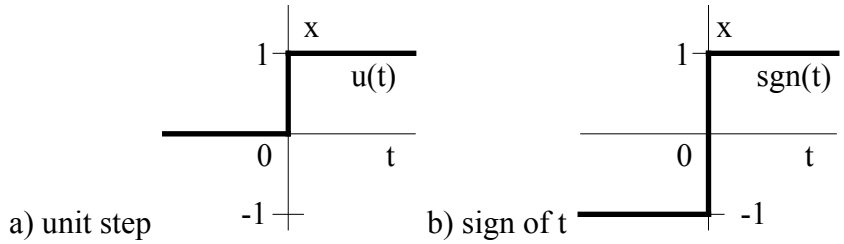
$$[2.19]\ u(t)+u(-t)=1.$$

A second example, with half value zero in the discontinuity point in the origin, is represented by the ***signal sign*** of the variable t, shown in Figura 2.6.b and indicated with the notation:

$$[2.20]\ \mathrm{sgn}(t) \hat{=} \begin{cases} -1 \ , \forall \ t<0 \\ 0 \ , \text{in } \ t=0 \\ 1 \ , \forall \ t>0 \end{cases} ;$$

Betweeen such signal and the unit step there is the following relation:

$$[2.21]\ \mathrm{sgn}(t) = u(t) - u(-t).$$



a) unit step b) sign of t

Figura 2.6: Examples of deterministic signals with first order discontinuity.

A further meaningful example of non continuous signal, with half values ½ in the discontinuity points |t|=T/2, is the signal named ***rectangular unit pulse*** of duration T (see Figura 2.7), expressed through the following notation:

$$[2.22]\ \mathrm{rect}\left(\frac{t}{T}\right) = \begin{cases} 1 \ , \forall \ |t| < T/2 \\ 1/2 \ , \ \text{in } |t| = T/2 \\ 0 \ , \forall \ |t| > T/2 \end{cases} ,$$

with the following relation with the unit step:

$$[2.23]\ \mathrm{rect}\left(\frac{t}{T}\right) = u\left(t+\frac{T}{2}\right) - u\left(t-\frac{T}{2}\right).$$

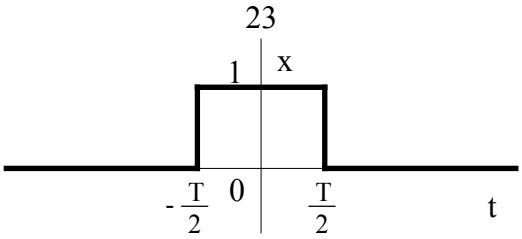


Figura 2.7: Unit rectangular pulse with two first order discontinuity.

Differently from the two signals previously defined the function rect(t/T) is characterised by the presence of a parameter: the duration T of the pulse. In case the value T is known, the signal is considered deterministic, as well as the unit step and the sign; instead, if T indicates a r.v., the function rect(t/T) bcomes the explicit expression of a random process with isolated event.
Every signal with discontinuity points of the first type in the instants t_k generally can be broken up into the sum:

$$[2.24]\quad x(t) = x_C(t) + x_G(t),$$

where $x_C(t)$ is a continuous signal and $x_G(t)$ is a ***step signal***, constant per segment (see example in Figura 2.8):

$$[2.25]\quad x_G(t)=\sum_{k=1}^{N} d_k u(t-t_k),\ d_k=x\left(t_k^+\right)-x\left(t_k^-\right).$$

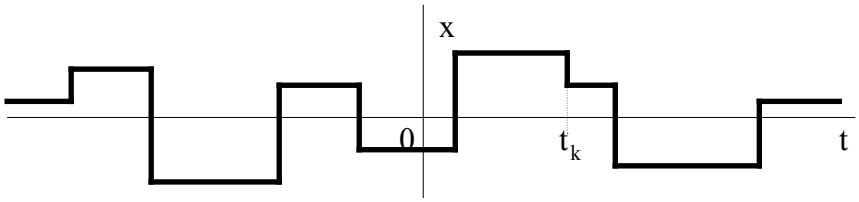


Figura 2.8: Waveform of a step signal.

Recalling the relations that allow to express the signal sign and the rectangular unit impulse as a function of the unit step u(t), it is possible to note that they are particular cases of the generic step signal expressed through the [2.25].

2.3.3 Signal duration

2.3.3.1 Time continuous signals with finite duration

The time continuous signal has finite duration if the evolution results strictly limted in time, i.e. if it is identically zero out of an interval (t_m, t_M), of limited duration $D = t_M - t_m$:

$$[2.26]\quad x(t)=\begin{cases} x \in X & ,\forall\ t \in (t_m, t_M) \\ x \equiv 0 & ,\forall\ t \notin (t_m, t_M) \end{cases}.$$

In case of short duration it is named impulsive signal or simply ***pulse***. If $t_m \geq 0$, the signal gains the attribute of ***causal signal*** with finite duration.
The first example is the ***rectangular pulse*** of duration T and finite amplitude A (see Figura 2.9a):

$$[2.27]\ x(t) = A\,\mathrm{rect}\left(\frac{t}{T}\right),$$

which generalises the unit rectangular pulse, adding a second parameter. Other examples are the ***triangular pulse*** of duration T and maximum value A (see Figura 2.9b):

$$[2.28]\ x(t) = A\,\mathrm{tri}\left(\frac{t}{T}\right),\ \text{con}\ \mathrm{tri}\left(\frac{t}{T}\right) = \left(1 - \frac{|t|}{T/2}\right)\mathrm{rect}\left(\frac{t}{T}\right),$$

and the ***raised cosine pulse*** of duration T and maximum value A (see Figura 2.9c):

$$[2.29]\ x(t) = A\,\frac{1}{2}\left[1+\cos\frac{2\pi t}{T}\right]\mathrm{rect}\left(\frac{t}{T}\right).$$

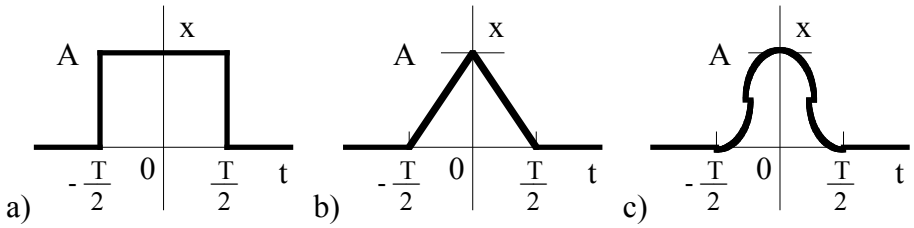


Figura 2.9: Waveforms of limited duration signals.

Observing the waveforms of the three examples of time continuous and finite duration signals shown in Figura 2.9, it is possible to note the presence of two discontinuities of first type in the first, three discontinuities of second type, i.e. of the derivative, in the second, while the third is continuous, as well as its derivative.

2.3.3.2 Time continuous and unlimited duration signals

Time continuous and unlimited duration signals, i.e. not strictly limited in time, are classified into three categories.

The ***bilateral signal*** is not zero over the whole time axis, except some points or limited number of intervals. A meaningful example of bilateral signal is the ***sinc signal***, defned through the expression:

$$[2.30]\ \mathrm{sinc}\left(\frac{t}{T}\right) = \frac{\sin(\pi t/T)}{\pi t/T}.$$

The considered signal, time continuous, is equal to 1 in the origin and is 0 each |t| multiple of the parameter T, as shown in Figura 2.10. The envelope of the considered function, i.e. the locus of the relative maxima of the absolute value |sinc(t/T)|, decreases following a law of inverse proportionality to the distance |t| from the origin. Thus, it is necessary to move away from the origin of a great number of intervalks T to achieve negligible values compared to unity for the sinc signal, for example the absolute value is always less than 0.01 if |t| is greater than 32T.

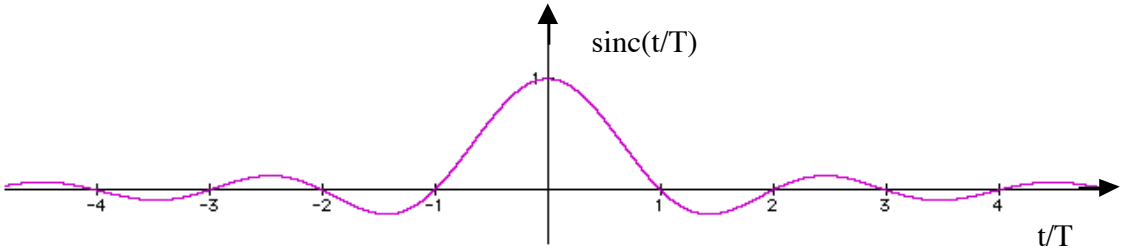


Figura 2.10: Waveform of signal $\mathrm{sinc}\left(\frac{t}{T}\right)$.

A second example of bilateral signal, stil of continuous type, is the ***unit gaussian signal***:

$$[2.31]\ \mathrm{gauss}(t/T) = e^{-\frac{1}{2}\left(\frac{t}{T}\right)^2},$$

with maximum value (1) in the origin and evolution decreasing even more rapidly as a function of increasing distance |t| from the origin, smaller is the value of parameter T. It is sufficient to move away from the origin for a small number of intervals T to achieve negligible values of the gaussian signal with respect to unit; for example, the absolute value is always less than 0.01 if |t| is greater than 3.1T.

The second category concerns the ***right monolateral signal***, which is always zero for each t less than an initial instant t_m, i.e.

$$[2.32]\ x_{MD}(t)=\begin{cases} x \equiv 0 & , \forall\ t<t_m \\ x \in X & , \forall\ t>t_m \end{cases} .$$

In case $t_m=0$ the considered signal is specified as right monolateral in the origin; if $t_m \geq 0$ often it is called ***causal signal*** with unlimited duration.
Figura 2.11 shows as an example two waveforms of right monolateral type; the former, with first type discontinuity, is the ***unit exponential decay signal***:

$$[2.33]\ u_e\,(t/T) = e^{-\frac{t}{T}}\,u(t),$$

where the parameter T>0 is called the time constant; the latter, time contonuous, but with second type discontinuity, is the ***ramp signal***:

$$[2.34]\ \text{ramp}(t/T) = \frac{1}{T}\int_{-\infty}^{t} u(\tau)d\tau = \frac{t}{T}\,u(t).$$

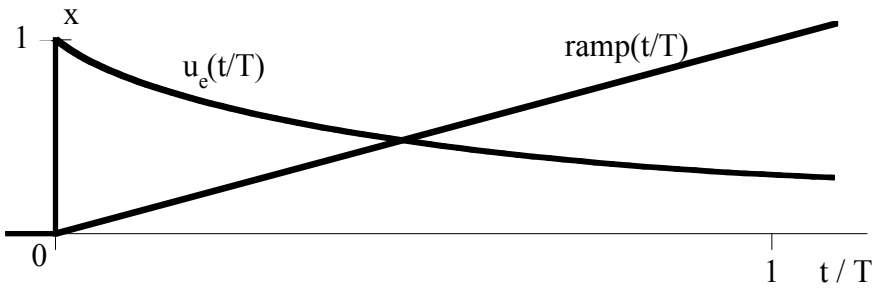


Figura 2.11: Waveforms of right monolateral signals from the origin.

Considering a final instant t_M, a left monolateral signal is always zero for each t greater than the mentioned t_M, i.e.:

$$[2.35]\ x_{MS}(t)=\begin{cases} x \in X & , \forall\ t< t_M \\ x \equiv 0 & , \forall\ t> t_M \end{cases} .$$

Recalling the relation [2.19], in case of bilateral signal it is possible to write:

$$[2.36]\ x(t) = x(t)[u(t) + u(-t)] = x_{MD}(t) + x_{MS}(t),$$

obtaining, as shown in Figura 2.12, the division of the signal into a pair of monolateral signals from the origin, one right and one left:

$$[2.37]\ x_{MD}(t) = x(t)\,u(t),\ x_{MS}(t) = x(t)\,u(-t).$$

Of course, the division can be realized starting from a generic instant t_0, utilizing the relation $u(t-t_0)+u(-t+t_0)=1$, obtained from [2.19] chaning the variable t in $t-t_0$.

2.3.3.3 Time continuous periodic signals

A meaningful subclass of the bilateral waveforms is represented by the time continuous ***periodic signals***, composed of the functions respecting the repetition property:

$$[2.38]\ x(t-kT_0) = x(t),\ \begin{cases} \forall\ t \in T \\ \forall\ k \in Z \end{cases},$$

where Z indicates the set of the integer numbers and the repetition interval of the waveform evolution T_0 is called ***period*** of the signal.

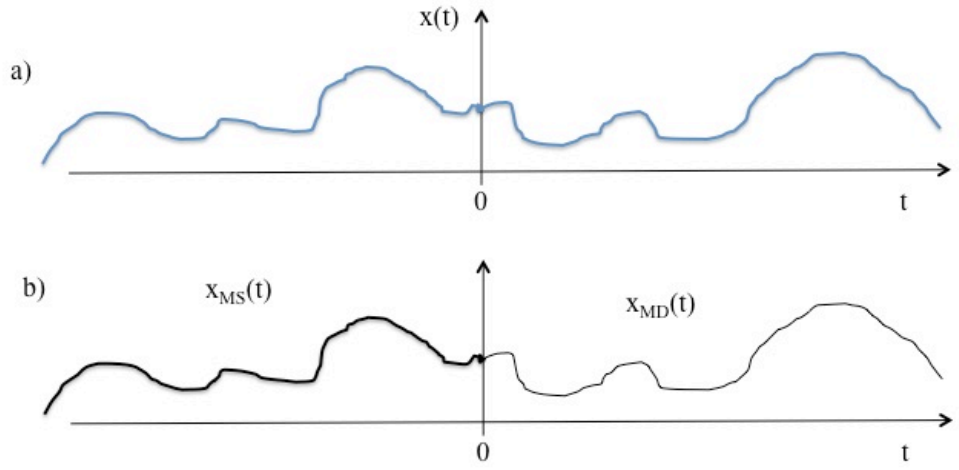


Figura 2.12: Bilateral signal (a) and division in monolateral signals (b).

The most typical examples of periodic signals are the harmonic signals: the already mentioned real signal (see [2.13]) and the ***complex harmonic***:

$$[2.39]\ A\,e^{j(\omega_0 t+\varphi_0)} = A\cos(\omega_0 t+\varphi_0) + jA\sin(\omega_0 t+\varphi_0).$$

A periodic signal can be obtained also from the infinite repetition of an elementary function. Given a signal g(t) and defined a repetition function at intervals T_0 (without the extremes in $\sum$ means that the the summation extends from $-\infty$ to $+\infty$):

$$[2.40]\ rep_{T0}\{g(t)\} = \sum_k g(t - kT_0),$$

it is possible to verify that a periodic signal $x(t) = rep_{T0}\{g(t)\}$ is generated, with period T_0 (see Figura 2.13). It is important to note that only in case g(t) is limited within an interval of duration $D_0 \leq T_0$, the evolution of the periodic signal is the same of the generating function, g(t), in the interval in which it is different from zero (see Figura 2.13b).

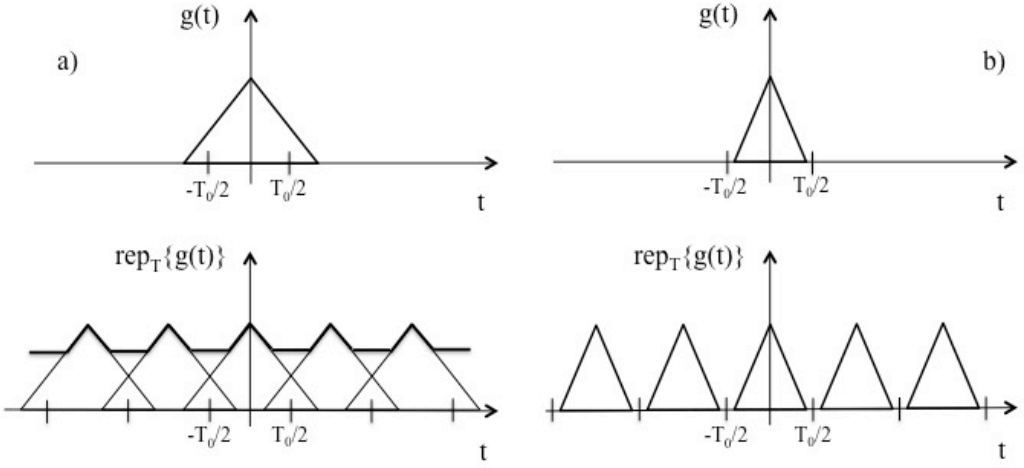


Figura 2.13: Periodic waveforms obtained by repetition of a generating function with a duration greater (a) and less (b) than the period.

2.3.4 Factorization of time continous signals

A generic time continuous real signal can be additively factorised in an ***even component*** and in a ***odd component***:

$$[2.41]\ x(t) = x_P(t) + x_D(t),$$

which are respectively characterized by the properties:

$$[2.42]\ x_P(t) = x_P(-t),\ x_D(t) = -\,x_D(-t),$$

highlithed in the vaweforms shown in Figura 2.14.

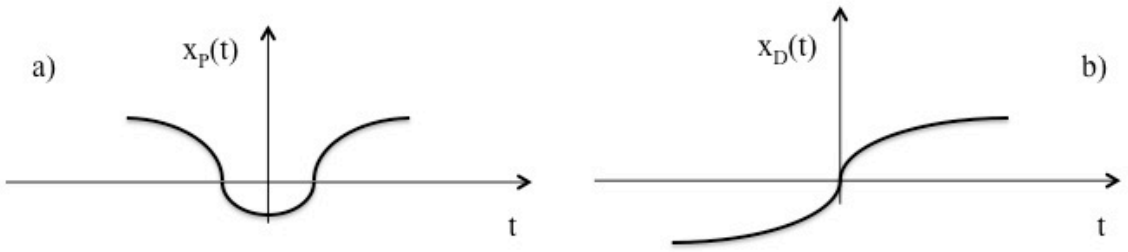


Figura 2.14: Even component (a) and odd component (b) of a real signal.

Since the just mentioned above properties allow to achieve:

$$[2.43]\ x(-t) = x_P(-t) + x_D(-t) = x_P(t) - x_D(t),$$

the following relations are valid:

$$[2.44]\ x_P(t)=\frac{1}{2}\left[x(t)+x(-t)\right],\ x_D(t)=\frac{1}{2}\left[x(t)-x(-t)\right].$$

In case either the odd component or the even component is identically equal to zero, respectively the signal is either ***even signal*** ($x_D(t)\equiv 0$) or ***odd signal*** ($x_P(t)\equiv 0$). The rectangular pulse Arect(t/T) and the sinc signal sinc(t/T) are examples of even signals while the signal sgn(t) is an example of odd signal.
Even a real causal signal can be factorised in even and odd components; as an example the unit step (see Figura 2.15a):

$$[2.45]\ u(t)=\frac{1}{2}\left[1+\mathrm{sgn}(t)\right].$$

In case of the rectangular pulse shifted in delay of a value $t_0>T/2$ it is achieved (see Figura 2.15b):

$$[2.46]\ \mathrm{rect}\left(\frac{t-t_0}{T}\right)=\frac{1}{2}\left[\mathrm{rect}\left(\frac{t-t_0}{T}\right)+\mathrm{rect}\left(\frac{t+t_0}{T}\right)\right]+\frac{1}{2}\left[\mathrm{rect}\left(\frac{t-t_0}{T}\right)-\mathrm{rect}\left(\frac{t+t_0}{T}\right)\right]$$

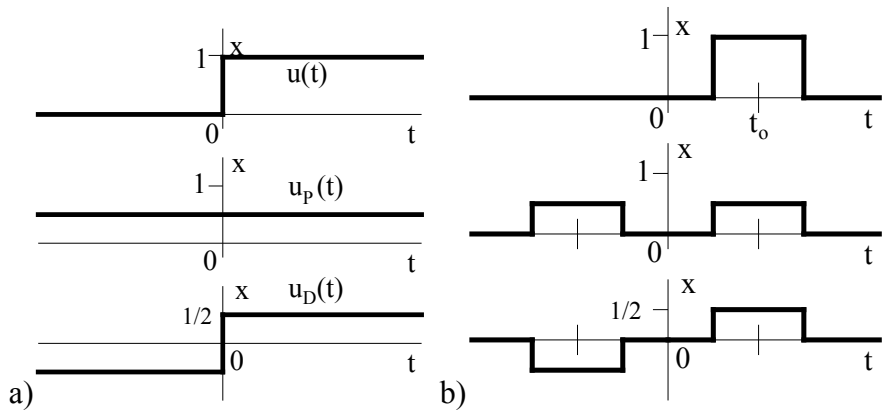


Figura 2.15: Factorization in even and odd parts of unit step a) and of rectangular unit pulse (b).

In case of a generic complex time continuous signal the factorization is generalized in a ***hermitian complex component*** and in an ***inverse hermitian complex component***:

$$[2.47]\ x(t) = x_H(t) + x_{AH}(t),$$

respectively characterized by the properties:

$$[2.48]\ x_H(t) = x_H^*(-t),\ x_{AH}(t) = -x_{AH}^*(-t).$$

It is possible to note that the hermitian component has the real part even and coefficient of the imaginery part odd, while the inverse hermitian component has real part odd and coefficient of the imaginery part even.

Since the above properties imply that:

$$[2.49]\ x^*(-t) = x_H^*(-t) + x_{AH}^*(-t) = x_H(t) - x_{AH}(t),$$

the following relations are valid:

$$[2.50]\ x_H(t)=\frac{1}{2}\left[x(t)+x^*(-t)\right],\ x_{AH}(t)=\frac{1}{2}\left[x(t)-x^*(-t)\right].$$

If either the inverse hermitian or the hermitian component is equal to zero the signal is respectively named ***hermitian signal***, characterized by:

$$[2.51]\ x(t) = x^*(-t),$$

or ***inverse ihermitian signal***, characterized by:

$$[2.52]\ x(t) = - x^*(-t).$$

The [2.51] implies that the simultaneous application of the simple operations of conjugation and time axis inversion produces no effect on a hermitian signal.

2.3.5 Ideal Dirac pulse

2.3.5.1 Properties of the ideal pulse

For the analysis of telecommunication systems, often a particular signal is used: it is the Dirac generalised function, or Dirac ***ideal pulse***, usually indicated as δ(t).
The ideal impulse is usually introduced through its sampling property, expressed by:

$$[2.53]\ \int_a^b f(t)\delta(t-t_0)dt = \begin{cases} f(t_0) , & a<t_0<b \\ \frac{1}{2}f(t_0) , & t_0=a, b \\ 0 , & t_0<a\ ,\ t_0>b \end{cases},$$

where the generic function f(t) is assumed to be continuous in $t=t_0$. Moreover, being the integral extended from -∞ to +∞, the following relation is valid:

$$[2.54]\ \int f(t)\delta(t-t_0)dt=f(t_0).$$

Since the last relation implies, in particular if f(t)≡1, that δ(t) is zero everywhere except in the origin and that its integral is equal to one, meaning that its area is:

$$[2.55]\ \int \delta(t)dt=1,$$

it is evident that the ideal pulse is not a function in ordinary sense. As shown in section 2.3.5.2, the ideal pulse δ(t) can be considered at the limit as the derivative of the unit step u(t).

2.3.5.2 The ideal pulse as limit of ordinary function

As above stated the ideal pulse is not an ordinary function: to approach it rigorously the concept of distribution, an entity able to match a value and a function, should be introduced. In order to use only ordinary functions, the entity δ(t) present as a factor in the integral [2.54] must be properly interpreted as limit of ordinary functions.
To this aim let's consider the unit step u(t), obtained as limit of a continuous function, assuming:

$$[2.56]\ u(t)=\lim_{T_0\to 0}\int_{-\infty}^{t/T_0} \mathrm{sinc}(\tau)d\tau,$$

where $T_0>0$ and the limit must be developed after the integration. Indicating with δ(t) the derivative with respect to t of the second member of [2.56] :

$$[2.57]\ \lim_{T_0 \to 0} \frac{1}{T_0} \operatorname{sinc}\left(\frac{t}{T_0}\right) = \delta(t),$$

one of the possible interpretations of the ideal pulse as limit of ordinary function is achieved. As a consequence δ(t) can be considered the derivative of the function u(t)-c, where c is constant. This relation, that once the integration constant is defined shows the definite integral

$$[2.58]\ \int_{-\infty}^{t} \delta(\tau)\, d\tau = u(t),$$

as a matter of fact is rigorous in the field of distributions.
The [2.57] implies that the ideal pulse is even, being such the signal sinc(t/T). Thus, considering a generic time shift t_0 the relation is achieved:

$$[2.59]\ \delta(t-t_0) = \delta(t_0-t).$$

2.3.6 Energy and power of time continuous signals

A ***functional space***, indicated with L_P, is the full set of time continuous signals x(t) for which the p^{th} norm exists, is finite and is defined by (being the integral extended from -∞ to +∞):

$$[2.60]\ \|x(t)\|_p \hat{=} \left[\int |x(t)|^p\, dt\right]^{1/p},$$

where p is a real positive number. Thus, L_1 is the functional space of the signals which can be integrated in their absolute value with finite result and L_2 is the functional space of the signals which can be integrated in their squared value with finite result. If the signals x(t) are defined only over a finite domain T, the functional space composed of these signals is indicated with the notation $L_P(T)$.
Having assumed that $|x(t)|^2$ provides the istantaneous power of x(t), in general complex, the ***energy of the signal*** is defined as:

$$[2.61]\ E_{xx} = \|x(t)\|_2^2 = \int |x(t)|^2\, dt.$$

The set of signals with finite E_{xx}, real by definition, is the first meaningful cathegory: such signals, which assemble the functional space L_2, are called ***finite energy*** or simply ***energy signals***. They can be characterised by finite duration, as in the examples of rectangular, triangular and raised cosine pulse, but also time unlimited, as in the case of sinc and gaussian signals.
In the case of time unlimited signal, it can be approximated with a ***practically finite duration signal***, identifying an interval, called ***practical duration*** of the signal, out of which x(t) is approximated with zero so that the integral of $|x(t)|^2$ calculated over just the practical duration provides a value corresponding to the actual energy E_{xx}.
If a signal x(t) is time unlimited and its energy is not finite, another quantity is considered, adopting the following definition of the time average power, or simply ***signal power***:

$$[2.62]\ W_{xx} = \lim_{T \to \infty} \frac{1}{T} \int |x_T(t)|^2\, dt,$$

where the truncated signal has been introduced:

$$[2.63]\ x_T(t) = x(t)\, \operatorname{rect}\left(\frac{t}{T}\right).$$

If the value of W_{xx}, real by definition, is finite, the considered type of signal is called with ***finite power*** or simply ***power signal***. For such signals also the following quantity is defined:

$$[2.64]\ x_{eff} = \sqrt{W_{xx}},$$

called ***effective value of the signal***.
In case of periodic signals, typical subclass of power signals, indicating with $E_{xx}(T_0)$ the energy of the signal x(t) calculated only within any time interval with duration equal to the period T_0 and independent on the particular selected interval, the power of the signal can be expressed through:

$$[2.65]\ W_{xx} = \frac{E_{xx}(T_0)}{T_0} = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} |x(t)|^2\, dt\ .$$

In the case of periodic signal generated by repetition of a generating function g(t) (see [2.40]), it is easy to note that $E_{xx}(T_0)$ is exactly the energy E_{gg} of the generating function, if its duration is not greater than T_0, otherwise in general it is different.
The integral of the signal over the entire time axis,

$$[2.66]\ \text{area}[x(t)] = \int x(t)\,dt\ ,$$

is called ***area of the signal***, which is in general complex and always finite for energy signals.
Only for power signals the time average value, or simply the ***signal average value***, can be different from zero and defined by:

$$[2.67]\ \bar{x} = \lim_{T\to\infty}\frac{1}{T}\int x_T(t)\,dt\ .$$

A signal with average value different from zero can be fractionated in two addends: one constant, equal to the time average value $\bar{x}$, and the other with zero average value:

$$[2.68]\ x_a(t) = x(t) - \bar{x}\ .$$

In the case of real signal, the two considered addends, $\bar{x}$ and $x_a(t)$, are called respectively ***continuous component*** and ***alternating component***.
Often, real power signals are considered, with finite codomain (range) X, called ***symmetric signals***: they have no continuous component and have opposite maximum and minimum values; defined the ***peak value*** as the common absolute value, i.e. $x_p=x_M=|x_m|$, the ***peak factor*** is defined as:

$$[2.69]\ F_p = \frac{x_p}{x_{eff}} = \frac{x_p}{\sqrt{W_{xx}}}\ .$$

As well known the peak factor of a real harmonic is equal to $\sqrt{2}$; for any other symmetric signal the peak factor is greater.
Note that the power signals are ideal, because actually there are no signals with no zero values in the most extreme parts of the time axis. In practice, however, it is often considered as power signals those energy waveforms which have much greater time duration than the interval (sometimes even reduced to the order of seconds) within which it is interesting to observe the signal and for which the power temporal average evaluated only in the range of observation practically does not vary if the extention of the signal varies; the power signal is then actually what is obtained by extrapolating the internal evolution outside of the observation time.

2.4 DISCRETE TIME SIGNALS

2.4.1 General case and sequences

From the typical evolution of a step signal it is easy to find out that it is completely characterized once known the discontinuity instants t_k and the set, correspondingly ordered, of the levels. On the basis of this concept it is worth to introduce the ***time discrete signal***, theoretical entity but very much valid on the logical-mathematical level, which is defined only on a countable discrete set

$\{t_n\}$ of time instants, that can belong either to a finite or infinite set of numbers, distributed over the time sparsely or equispaced between adjacent elements, as in Figura 2.16.

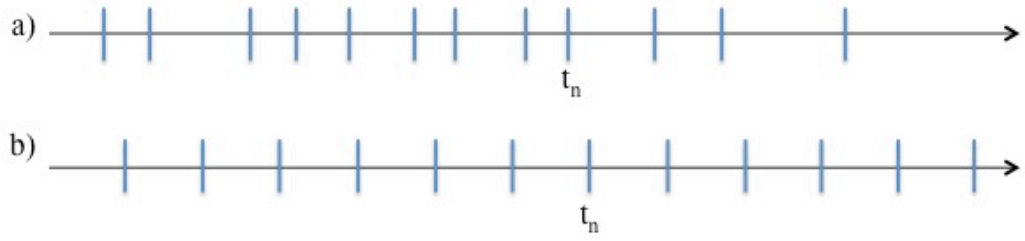


Figura 2.16: Discrete definition set of a time discrete signal, with sparse distribution (a) and equispaced over the time (b).

The most meaningful case (actually and as a matter of fact the only dealt with in literature) is that with equispaced discrete instants, with generic constant interval T, for which the following very simple biunivolcal relation with the set of integer numbers can be established:

$$[2.70]\ t_n = nT.$$

Then, the time discrete signal is called sequence and is indicated with the notation x(n), where the generic k^{th} element, or ***sample*** of the sequence, takes the value:

$$[2.71]\ x_k = x(kT),$$

which can be either real or complex.
All the elementary operations, such as multiplication by a constant and time shift, as well as conjugation and time axis inversion, are applicable to sequences, but executing them sample by sample. The concepts of real and complex sequences, of sequences reciprocally faithful, easily inferred from definitions provided for time continuous signals, are valid as well.
If the values of all the samples x_k are known, a (sure) ***deterministic sequence*** is obtained. Alternatively, the considered elements can be randon variables, indicated with the notation X_k: then, the succession of such r.v. becomes the representation of a discrete random process, as already mentioned in section 2.2.2.2.
The first elementary example of deterministic sequence is the ***unit sample***, shown in Figura 2.17a and expressed by the notation:

$$[2.72]\ \delta(n) = \begin{cases} 1\ , & n = 0 \\ 0\ , & \forall\ n \neq 0 \end{cases},$$

which corresponds in the discrete time domain to the Dirac impulse δ(t), introduced as elementary time continuous signal, but that absolutely doesn't have to be confused with it because the definition, the analitycal properies and the reference mathematical framework are extremely different.
As econd example is the ***unit step sequence***, shown in Figura 2.17b, analytically expressed as:

$$[2.73]\ u(n) = \begin{cases} 0\ , & \forall\ n < 0 \\ 1\ , & \forall\ n > 0 \end{cases}.$$

Note that in the origin, n=0, it is equal to 1, differently from the similar time continuous signal which is equal to the half value ½. As a consequence the following property stands:

$$[2.74]\ u(n) + u(-n-1) = 1.$$

Between the two particular sequences above introduced the following relation is valid:

$$[2.75]\ \delta(n) = u(n) - u(n-1).$$

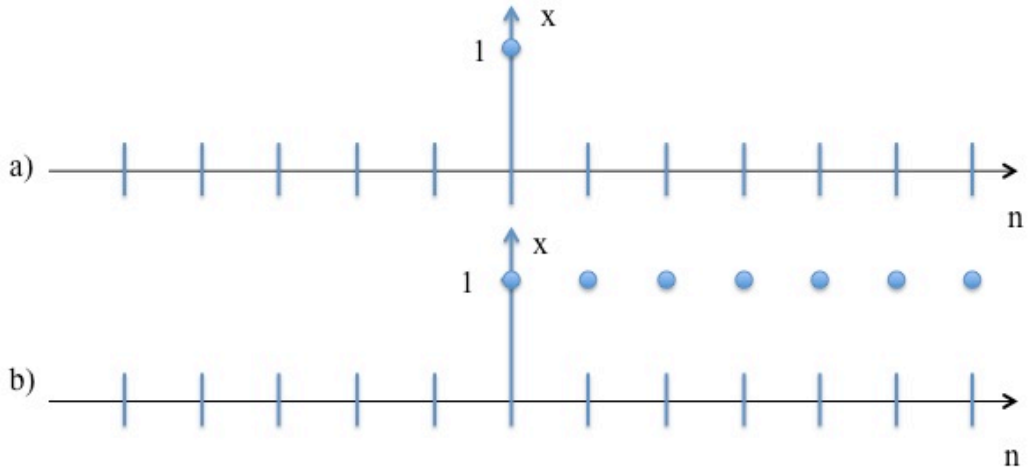


Figura 2.17: Unit sample (a) and unit step sequence (b).

2.4.2 Sequence lenght

Utilizing the series of shifted unit samples, a generic sequence with unlimited duration can be represented as:

$$[2.76]\ x(n) = \Sigma_k\, x_k \delta(n-k),$$

and called ***bilateral sequence***, with samples that may be not zero for any value of k and where the summation is intended to be extended from -∞ to +∞. In the case the samples are zero for each k lower than an integer number k_m and greater than another integer $k_M > k_m$, it is a finite lenght sequence, often called ***string*** of samples; in this case, in the expression like [2.76] the mentioned values k_m and k_M are the extremes of the summation.
In the frame of bilateral sequences the ***periodic sequences*** are characterised by the continuous repetition of the same string of samples and have the following property:

$$[2.77]\ x(n-hN_0) = x(n),$$

with h integer and repetition rate N_0, which must be a finite integer number greater than one. The time interval $T_0 = N_0 T$ is the period of the sequence.
Similarly to time continuous signals, starting from and integer number k_m a ***right monolateral sequence*** can be defined if its samples are all zero for $k < k_m$, so that:

$$[2.78]\ x_{MD}(n) = \sum_{k=k_m}^{\infty} x_k\, \delta(n-k);$$

with $k_m \geq 0$ a ***causal sequence*** is obtained. Considering an integer number k_M a ***left monolateral sequence*** can be defined if its samples are all zero for $k > k_M$, so that:

$$[2.79]\ x_{MS}(n) = \sum_{k=-\infty}^{k_M} x_k\, \delta(n-k).$$

Utilizing property [2.36] it is always possible to divide a bilateral sequence into a pair of monolateral sequences, one right and one left; this is like to break the summation in [2.76].
Finally, all the concepts expressed with respect to fractioning of time continuous signals can be applied too: thus, in case of real sequences ***even sequences*** ($x_k = x_{-k}$) and ***odd sequences*** ($x_k = -x_{-k}$) and in case of complex sequences ***hermitian sequences*** ($x_k = x_{-k}^*$) and ***inverse hermitian sequences*** ($x_k = -x_{-k}^*$) can be defined.

2.4.3 Energy and power of sequences

A sequence x(n), since it is different from zero only in a set of countable instants, has neither energy nor power, according to the definition adopted for time continuous signals. Nevertheless, it is possible to define in a conventional way the mentioned physical quantities also for the sequences, referring to the corresponding step signal, obtained from the time discrete one by means of the operation of ***hold of samples***, shown in Figura 2.18.

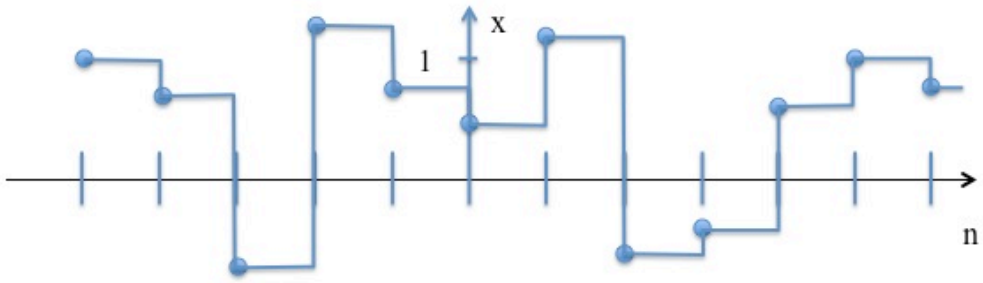


Figura 2.18: Step signal obtained from a sequence with the operation of hold of the samples.

Thus, with the mentioned procedure the conventional quantities are obtained:

$$[2.80]\ E_{xx} = T \sum_k |c_k|^2,$$

$$[2.81]\ W_{xx} = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} |c_k|^2 ,$$

called ***energy of the sequence*** and ***power of the sequence***, which are finite and different from zero respectively in the case of ***energy sequences*** or ***power sequences***. In the latter case the ***average value of the sequence*** is defined as:

$$[2.82]\ \bar{x} = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} c_k .$$

The power of a periodic sequence, as well as in case of time continuous signals, can be expressed on the basis of the $E_{xx}(T_0)$ of the sequence truncated in a generic interval corresponding to the period $T_0=N_0T$, obtaining:

$$[2.83]\ W_{xx} = \frac{E_{xx}(T_0)}{T_0} = \frac{1}{N_0} \sum_{k=1}^{N_0} |c_k|^2 .$$

2.5 AFFINITY BETWEEN SIGNALS

In the operations of the telecommunications it is important to quantitively evaluate the affinity between signals, when one is shifted with respect to the other.
Such an evaluation is exploited utilizing correlation functions properly defined, which depend on the time variable τ which represents the amount of the time shift of one signal with respect to the other.

2.5.1 Affinity between energy signals

A first important case concerns the energy signals which, as already mentioned, represent the very most of the the signals utilized in telecommunications. The measure of the affinity between energy signals represents an important factor in different applications spanning from radar to mobile radiocommunications.

2.5.1.1 Cross correlation function of energy signals

In the functional space L_2 two generic energy signals, x(t) e y(t), having respectively energies E_{xx} and E_{yy} are considered; the time ***cross correlation function***, or mutual time correlation, of the two signals is defined through the following expression:

$$[2.84]\ C_{xy}(\tau) = \int x(t+\tau)\, y^*(t)\, dt\,.$$

Also the equivalent notation can be used:

$$[2.85]\ C_{xy}(\tau) \equiv (x(t+\tau), y(t)).$$

The introduced function changes exchanging the order of the two signals, but the following property stands:

$$[2.86]\ C_{yx}(\tau) = C^*_{xy}(-\tau)\,.$$

The values assumed in the origin (for τ=0) by the two cross correlation funtions are called ***mutual energies***; thanks to the previous relation the following equations are obtained:

$$[2.87]\ E_{xy} \triangleq C_{xy}(0) = \int x(t)\, y^*(t)\, dt = \left[\int y(t)\, x^*(t)\, dt\right]^* = C^*_{yx}(0) \triangleq E^*_{yx}\,,$$

i.e. one mutual energy is the complex conjugate of the other.
The following property is valid for the cross correlation function:

$$[2.88]\ |C_{xy}(\tau)|^2 \le E_{xx}\, E_{yy};$$

The reader interested to the demonstration of [2.86] and of [2.88] can see section 2.5.1.2.
Introducing the corresponding normalized dimensionless function, called ***cross correlation index function***:

$$[2.89]\ \kappa_{xy}(\tau) \triangleq \frac{C_{xy}(\tau)}{\sqrt{E_{xx} E_{yy}}}\,,$$

as a consequence it can be maximized:

$$[2.90]\ \kappa_{xy}(\tau) \le 1.$$

In particular, for τ=0 and recalling [2.87], from [2.88] the following relations for the energies are obtained:

$$[2.91]\ |E_{xy}|^2 = |E_{yx}|^2 \le E_{xx}\, E_{yy},\ E_{xy}\, E_{yx} \le E_{xx}\, E_{yy}.$$

The cross correlation function can highlight the affinity in broad sense between two energy signals, meaning that such kind of affinity is measured by the maximum of the absolute value of the cross correlation index function. Two signals have no affinity in broad sense and are called not correlated if their cross correlation function is identically equal to zero:

$$[2.92]\ C_{xy}(\tau) \equiv 0, \text{ ossia } \kappa_{xy}(\tau) \equiv 0\ ;$$

instead they have maximum affinity in broad sense if:

$$[2.93]\ |\kappa_{xy}(\tau)|_{max} = 1.$$

It is very interesting to remark that such last event occurs if and only if the pair of signals is faithful and thus the relation $y(t)=g\, e^{-j\Gamma}\, x(t-t_0)$ is valid; the maximum unit value of $|\kappa_{xy}(\tau)|$ is then obtained just when τ assumes the particular value $\tau=-t_0$. The interested reader will find the demonstration in section 2.5.1.2. Definitively, the faithful signals are the only signals between which the maximum cros correlation can occur, i.e. the maximum affinity in broad sense.

2.5.1.2 Demonstration of the properties of the cross correlation function

Exchanging the order of x and y in the definition of the cross correlation function, conjugating and then applying the change of the variable t'=t+τ the following relation is obtained:

$$[2.94]C_{yx}(\tau)=\int y(t+\tau)x^{*}(t)dt=\left[\int x(t)y^{*}(t+\tau)dt\right]^{*}=\left[\int x(t'-\tau)y^{*}(t')dt'\right]^{*}=\left[C_{xy}(-\tau)\right]^{*};$$

which demonstrates the property [2.86].
Considering a pair of generic complex functions a(t) and b(t), belonging to the functional space L_2, the Schwartz disequality is valid:

$$[2.95]\ \left|\int a(t)b(t)dt\right|^{2}\le\int|a(t)|^{2}dt\int|b(t)|^{2}dt,$$

where the equality occurs if and only if the following relation is satisfied:

$$[2.96]\ b(t)=Ka^{*}(t),$$

with $K = g e^{j\Gamma}$ arbitrary complex constant. Applying the Schwartz disequality to the cross correlation function, i.e. with $a(t) \equiv x(t+\tau)$ and $b(t) \equiv y^{*}(t)$, the following relation is obtained:

$$[2.97]|C_{xy}(\tau)|^{2}=\left|\int x(t+\tau)y^{*}(t)dt\right|^{2}\le\int|x(t+\tau)|^{2}dt\int|y(t)|^{2}dt=\int|x(t)|^{2}dt\int|y(t)|^{2}dt=E_{xx}E_{yy},$$

which demonstrates the property [2.88].
On the basis of [2.96] the maximum value $\sqrt{E_{xx}E_{yy}}$ of $|C_{xy}(\tau)|$ is obtained if and only if the following relation occurs:

$$[2.98]\ y^{*}(t)=g\,e^{j\Gamma}x^{*}(t+\tau),$$

which corresponds to the well known definition of faithful signal for $\tau=-t_0$. Definitively:

$$[2.99]\ |C_{xy}(\tau)|_{max}=|C_{xy}(-t_0)|=\left|\int x(t-t_0)ge^{j\Gamma}x^{*}(t-t_0)dt\right|=|g|E_{xx}.$$

Taking into account [2.97], which together with the previous one provides:

$$[2.100]\ |g|=\sqrt{\frac{E_{yy}}{E_{xx}}},$$

it is thus demonstrated that if and only if the two signals are faithful with a delay t_0, the modulus of the cross correlation function can assume the value one for $\tau=-t_0$.

2.5.1.3 Autocorrelation function of an energy signal

Correlating an energy signal with itself the time autocorrelation function is obtained:

$$[2.101]\ C_{xx}(\tau)=\int x(t+\tau)x^{*}(t)dt.$$

As a particular case of [2.86] the following property is valid:

$$[2.102]\ C_{xx}(\tau)=C_{xx}^{*}(-\tau);$$

the autocorrelation function is thus hermitian, assuming a real value in the origin corresponding with the energy of the signal:

$$[2.103]\ C_{xx}(0)=E_{xx}.$$

Such a value is the maximum that the modulus of the autocorrelation function can assume, because as particular case of [2.88] the following disequality is verified:

$$[2.104]\ |C_{xx}(\tau)|\le E_{xx}.$$

If the signal is real the autocorrelation function is real and even, i.e. $C_{xx}(\tau)=C_{xx}^{*}(\tau)=C_{xx}(-\tau)$ occurs.
The autocorrelation function highlights for each τ the decreasing of affinity in strict sense that occurs between a signal and its waveform shifted in time, with respect to the full correlation occurring for $\tau=0$. It is also an indicator of how much memory the signal keeps of itself, i.e. how much globally the values at time $(t+\tau)$ depend on those at time t.
In the case of signals with limited duration D, the affinity in strict sense, i.e. the memory of the signal, is totally lost for any shift, ahead or in delay, greater than D; as a consequence, the autocorrelation function has limited duration but in a double time interval 2D, since for $|\tau|>D$ the two signals x(t) and $x(t+\tau)$ are non zero in intervals separated in time, so that the function under the integral is zero in the autocorrelation function. A typical example is the rectangular impulse A rect(t/T), with autocorrelation function being a triangular pulse depending on the independent variable τ, with maximum value A^2 and duration 2T (see Figura 2.19).

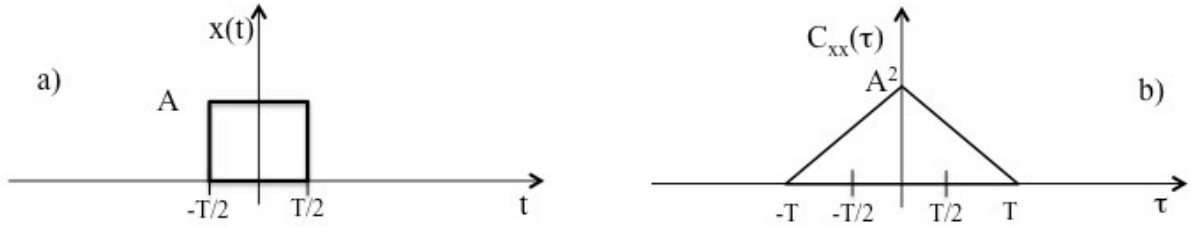


Figura 2.19: Rectangular pulse (a) and its autocorrelation function (b).

2.5.1.4 Scalar product of energy signals

The value in the origin (for $\tau=0$) of the cross correlation function between two signals, that is the constant quantity, in general complex, defined by:

$$[2.105]\ (x,y)=\int x(t)y^{*}(t)dt\,,$$

is called ***scalar product*** of the two signals.
Considering what shown above, the scalar products coincide with the mutual energies, i.e.: $(x,y)=C_{xy}(0)=E_{xy}$ and $(y,x)=C_{yx}(0)=E_{yx}=E_{xy}^{*}$. Thus, the two scalar products have the same modulus; in case of real signals they are the same. Using the property [2.88] the following relation is valid:

$$[2.106]\ |(x,y)|\le\sqrt{E_{xx}E_{yy}}\,,$$

where the equality occurs if and only if $y(t)=g\,e^{-j\Gamma}x(t)$ is verified, implying that the two signals are reciprocally faithful with no delay; considering the cross correlation index:

$$[2.107]\ \kappa_{xy}\triangleq\frac{(x,y)}{\sqrt{E_{xx}E_{yy}}}\,,$$

its modulus is lower than one, unless the two signals are related through the simple complex proportionality.
The cross correlation index provides a measure of the affinity in strict sense between two energy signals, i.e. not considering any time shift. Two vaweforms are called ***parallel signals*** when $\kappa_{xy}=1$, i.e. the relation is $y(t)=|g|\,x(t)$, and ***antipodal signals*** in the particular case $\kappa_{xy}=-1$, with g=-1 e $\Gamma=0$, that means y(t)=-x(t). Decreasing $|\kappa_{xy}|$ corresponds to a decreasing affinity in strict sense between the signals, up to the case of $\kappa_{xy}=0$, that means to have the scalar product euqal to zero (x,y)=0, in which the two waveforms are called ***orthogonal signals***.
The previous geometrical type denominations are carried from some considerations developed in section 2.5.1.5, which the reader willing to go more in deep is invited to read.

The orthogonality condition between energy signals, that is their scalar product equal to zero (x,y)=0, is often very meaningful; for this reason it is worth to develop the following considerations.

First of all two signals can be orthogonal but they can have some affinity in wide sense, i.e. they may not be uncorrelated (see [2.92]); in fact, for $\tau \neq 0$ the cross correlation function of a pair of orthogonal signals can be different from zero or the modulus can even assume the maximum value. As an example, this is the case of any signal with limited duration D and for a waveform faithful to it with a shift $|t_0|$ greater than 2D, which ensures the separation in the time domain of the two signals and as a consequence that the scalar product is zero, but full affinity in wide sense is implied.

Let's keep in mind that the separation in time of a pair of signals is a condition sufficient but not necessary for their orthogonality; in fact, two signals can be orthogonal even if they are not strictly time limited.

2.5.1.5 Geometrical formal similarities of the scalar product

Taking into account that the energy is the square of the second order norm (see [2.60]), for the signal sum x(t)+y(t) the following expression is achieved:

$$[2.108]\ \left\|x(t)+y(t)\right\|_2^2 = (x+y,x+y) = (x,x) + (y,y) + 2\Re\{(x,y)\} = E_{xx}+E_{yy}+2\sqrt{E_{xx}E_{yy}}\cos\theta = \left\|x(t)\right\|_2^2 + \left\|y(t)\right\|_2^2 + 2\left\|x(t)\right\|_2 \left\|y(t)\right\|_2 \cos\theta,$$

where the cosine is

$$[2.109]\ \cos\theta = \frac{\Re\{(x,y)\}}{\sqrt{E_{xx}E_{yy}}} = \Re\{\kappa_{xy}\} \leq 1.$$

In the case of parallel signals $\kappa_{xy}=1$ and then $\cos\theta = 1$; thus, the particular relation is achieved:

$$[2.110]\ \left\|x(t)+y(t)\right\|_2^2 = \left[\left\|x(t)\right\|_2 + \left\|y(t)\right\|_2\right]^2,$$

formally identical to that achieved considering a pair of parallel vectors in the space instead of parallel signals x(t) and y(t). In case of orthogonal signals, it is possible to verify that (x,y)=0 and, even more, then $\Re\{(x,y)\}=0$ and $\cos\theta = 0$; the following expression is thus achieved:

$$[2.111]\ \left\|x(t)+y(t)\right\|_2^2 = \left\|x(t)\right\|_2^2 + \left\|y(t)\right\|_2^2,$$

formally identical to that achieved considering a pair of orthogonal vectors in the space instead of two orthogonal signals x(t) and y(t).

2.5.1.6 Famiglie di segnali ortogonali

A set of energy signals $\{x_k(t)\}$ constitutes a ***family of orthogonal signals*** if the following relation is verified:

$$[2.112]\ (x_k, x_h) = E_{xkxh} = 0,\ \forall\, h \neq k\ ;$$

and, moreover, if

$$[2.113]\ (x_k, x_h) = E_{xkxh} = \begin{cases} 0\ , & \forall\, h \neq k \\ 1\ , & h = k \end{cases},$$

the sub case of a ***family of orthonormal signals*** is achieved.

An an example two important families of orthogonal signals will be dealt with, introducing some properties of theirs, postponing demonstrations.

The former family of orthogonal signals is the discrete and countable set $\{x_k(t)\}$ composed of the complex unit harmonics truncated in the generic interval T,

$$[2.114]\ x_k(t) = e^{jk\Omega t}\,\mathrm{rect}\left(\frac{t}{T}\right),$$

where k is integer and $\Omega=2\pi/T$, for which the following relation stands:

$$[2.115]\ (x_k, x_h) = \int e^{jk\Omega t}e^{-jh\Omega t}\mathrm{rect}\left(\frac{t}{T}\right)dt = \int_{-T/2}^{T/2} e^{j(k-h)\Omega t}\,dt = \begin{cases} 0\ , \forall\ h \neq k \\ T\ ,\ h = k \end{cases}.$$

The energies of all the considered signals have thus the common value:

$$[2.116]\ E_{xkxh} = T.$$

Moreover, it is verified that:

$$[2.117]\ \mathrm{area}[x_k(t)] = \int_{-T/2}^{T/2} e^{jk\Omega t}\,dt = \begin{cases} 0\ , \forall\ k \neq 0 \\ T\ ,\ k = 0 \end{cases},$$

i.e. the area is different from zero only for the function of the family with k=0, which is actually the unit rectangular pulse. Finally, it is possible to note that each complex member of the family, with $k\neq 0$,

$$e^{jk\Omega t}\,\mathrm{rect}\left(\frac{t}{T}\right) = [\cos k\Omega t + j\,\sin k\Omega t]\,\mathrm{rect}\left(\frac{t}{T}\right),$$

is actually composed of a pair of addends that are reciprocally orthogonal, as the reader can easily demonstrate.

The latter family of orthogonal signals is composed of the sinc(t/T) pulse and of its time shifted, of a quantity multiple of the finite parameter T, waveforms,

$$[2.118]\ x_k(t) = \mathrm{sinc}\left(\frac{t}{T}-k\right),$$

for which it is possible to verify that (the reader will be able to verify later on utilizing [3.102]):

$$[2.119]\ (x_k, x_h) = \int \mathrm{sinc}\left(\frac{t}{T}-k\right)\mathrm{sin}\,c\left(\frac{t}{T}-h\right)dt = \begin{cases} 0\ , \forall\ h \neq k \\ T\ ,\ h = k \end{cases}.$$

The energies of the considered real signals are all equal to the same value:

$$[2.120]\ E_{xkxh} = \int \mathrm{sinc}^2\left(\frac{t}{T}\right)dt = T.$$

Moreover, it is verified that:

$$[2.121]\ \mathrm{area}[x_k(t)] = \int \mathrm{sinc}\left(\frac{t}{T}\right)dt = T,$$

i.e. the area of all the functions assumes the same value. Finally, the particular property stands:

$$[2.122]\ \sum_k \mathrm{sinc}\left(\frac{t}{T}-k\right) = 1.$$

2.5.2 Affinity between power signals

2.5.2.1 Correlation for power signals

In the case of two generic power signals x(t) and y(t), having respectively power W_{xx} and W_{yy}, it is possible to define the time ***cross correlation function*** of the two signals:

$$[2.123]\ R_{xy}(\tau) \triangleq \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} x(t+\tau)\,y^{*}(t)\,dt\ ,$$

and, as particular case y(t)≡x(t), the time ***autocorrelation function*** of the power signal:

$$[2.124]\ R_{xx}(\tau) \triangleq \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} x(t+\tau)\,x^{*}(t)\,dt\ .$$

Paying the attention to replace the functions of type C(τ) with the similar of type $R(\tau)$ and the energies with powers, all the comments and properties developed on the affinity for energy signals exposed in sections 2.5.1.1 and 2.5.1.3 are applicable to power signals.
Since power signals may have time average values $\bar{x}$ and $\bar{y}$ different from zero, furthermore it is possible to define the time ***covariance funtion*** of the two signals:

$$[2.125]\ K_{xy}(\tau) \triangleq \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left[x(t+\tau)-\bar{x}\right]\left[y(t)-\bar{y}\right]^{*} dt\ ,$$

and, as particular case the time ***autocovariance funtion*** of the power signal:

$$[2.126]\ K_{xx}(\tau) \triangleq \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left[x(t+\tau)-\bar{x}\right]\left[x(t)-\bar{x}\right]^{*} dt\ .$$

The above introduced functions, applied only to the components $x_a(t)$ and $y_a(t)$ characterized by average values equal to zero, are the same of the previous ones in the event $\bar{x}=\bar{y}=0$; in general the simple relations are verified:

$$[2.127]\ R_{xy}(\tau) = K_{xy}(\tau) + \bar{x}\bar{y}^{*}\ ,$$

$$[2.128]\ R_{xx}(\tau) = K_{xx}(\tau) + W_c,$$

where $W_c=\bar{x}\bar{x}^{*}$ is the power associated just to the component $\bar{x}$ of the signal. Considering that $K_{xx}(0)$ is equal to the power of just the component $x_a(t)$, as a consequence the power of the generic power signal is figured out to be equal to the sum of the powers of the two components considered separately.
In the particular case that both signals are periodic with same period T_0, the cross correlation and autocorrelation functions are periodic as well in the independent variable τ, with period T_0, and assume the expressions:

$$[2.129]\ R_{xy}(\tau) = \frac{1}{T_o}\int_{-T_0/2}^{T_0/2} x(t+\tau)\,y^{*}(t)\,dt = R_{xy}(\tau - kT_0)\ ,$$

$$[2.130]\ R_{xx}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t+\tau)\,x^{*}(t)\,dt = R_{xx}(\tau - kT_0)\ .$$

The reader can demonstrate that for the complex unit harmonic $e^{jk\Omega_o t}$, with period $T_0=2\pi/\Omega_0$, and for its components, real and imaginery part ($\cos\Omega_0 t$ and $\sin\Omega_0 t$), the autocorrelation functions are those shown in Tabella 2.1.

Tabella 2.1: Autocorrelation function for harmonic signals

x(t)	$e^{j\Omega_0 t}$	$\cos\Omega_0 t$	$\sin\Omega_0 t$
$R_{xx}(\tau)$	$e^{j\Omega_0 \tau}$	$\frac{1}{2}\cos\Omega_0\tau$	$\frac{1}{2}\cos\Omega_0\tau$

2.5.2.2 Families of incoherent and uncorrelated signals

Two power signals are ***incoherent*** if their cross correlation function is identically equal to zero, i.e. this relation is valid:

$$[2.131]\ R_{xy}(\tau) \equiv 0\ ;$$

instead, they are called ***uncorrelated*** if the less severe condition is verified:

$$[2.132]\ K_{xy}(\tau) \equiv 0,$$

which disregargds the existence of time average values.
A set of power signals constitutes a ***family of incoherent signals*** if, indicating with x(t) and y(t) two different generic members, the [2.131] is verified; instead, the set composes a ***family of uncorrelated signals*** if the [2.132] is verified.
The continuous set of the complex unit harmonics $\{e^{j\omega t}\}$ is an example of family of incoherent signals. Considering two generic members $e^{j\omega_x t}$ and $e^{j\omega_y t}$, with $\omega_x=2\pi f_x \neq \omega_y=2\pi f_y$, applying [2.123] the following relation is achieved:

$$[2.133]\ R_{xy}(\tau) = e^{j\omega_x \tau} \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} e^{j(\omega_x-\omega_y)t}\, dt = e^{j\omega_x \tau} \lim_{T\to\infty} \mathrm{sinc}[(f_x-f_y)T],$$

which for $f_x \neq f_y$ is equal to zero for any τ because of the property of the sinc function to tend to zero when its argument tends to infinite, from the positive or negative side. The reader can demonstrate that both the continuous sets $\{\cos\omega t\}$ and $\{\sin\omega t\}$ form yet, but separately, two families of incoherent signals.

2.5.3 Affinity between energy and power signals

The mathematical definition of scalar product between two energy signals can be broadened to all the cases in which the operation of integration provides a finite value, even if one of the two is a power signal. The result of the operation:

$$[2.134]\ (x,y) = \int x(t)\, y^*(t)\, dt\,,$$

formally identical to the scalar product but differently applied, allows then to evaluate the affinity in strict sense, or ***correlation***, existing between the two signals.
If the result of the correlation is a finite value $(x,y)=E_{xy}$ and if y(t) is an energy signal, it is possible to define the signal, that is an energy one as well:

$$[2.135]\ x_y(t) = \frac{E_{xy}}{E_{yy}}\ y(t)\ ;$$

thus, the power signal x(t) can be fractionated into two addends:

$$[2.136]\ x(t) = x_y(t) + x_0(t),$$

the former proportional to y(t) by definition through the factor E_{xy} which measures the correlation, i.e. the affinity in strict sense between the two signals, and the latter, orthogonal to the former. In fact, the relation:

$$[2.137]\ E_{xy} = (x_y+x_0,y) = (x_y,y)\ (x_y,y)\ + (x_0,y) = \frac{E_{xy}}{E_{yy}}\ E_{yy} + (x_0,y),$$

is achieved, that implies that $(x_0,y)=0$.
The orthogonal component $x_0(t)$ is a power signal considering that x(t) is a power signal as well.

2.5.4 Affinity between sequences

For a given pair of sequences, x(n) e y(n), the sequences function of the independent integer variable ν can be defined:

$$[2.138]\ C_{xy}(\nu) \equiv x \otimes y = T \sum_k x(k+\nu)\, y^*(k),$$

$$[2.139]\ R_{xy}(\nu) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} x(k+\nu)\, y^*(k),$$

called time ***cross correlation sequences***, respectively of energy or power, between the two sequences. In the case of power sequences, in addition, the time ***covariance sequence*** can be defined:

$$[2.140]\ K_{xy}(\nu) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} \left[x(k+\nu)-\bar{x}\right]\left[y(k)-\bar{y}\right]^* = R_{xy}(\nu) - \bar{x}\bar{y}^*$$

In the case of x(n)≡y(n), the three above expressions provide respectively the ***autocorrelation sequence***, $C_{xx}(\nu)$ or $R_{xx}(\nu)$, and the time ***autocovariance sequence***, $K_{xx}(\nu)$, of the given sequence.
The above defined cross correlation, autocorrelation, covariance and autocovariance sequences have the same meanings of the corresponding functions concerning time continuous signals: they can evaluate the affinity in wide sense between sequences.
In the case of energy sequences, similar relations to those obtained in sections 2.5.1.1 and 2.5.1.3 for time continuous signals are applicable, but replacing functions of type C(τ) with sequences of type C(ν).
The above consideration are applicable to power sequences as well, considering sequences of type $R(\nu)$ and $K(\nu)$ instead of functions of type $R(\tau)$ and $K(\tau)$.
In particular, the following properties for power real sequences are valid:
⇒ the cross correlation and covariance sequences are real and even and thus:

$$[2.141]\ R_{yx}(\nu) = R_{yx}(-\nu) = R_{xy}(\nu) = R_{xy}(-\nu),$$

$$[2.142]\ K_{yx}(\nu) = K_{yx}(-\nu) = K_{xy}(\nu) = K_{xy}(-\nu);$$

⇒ the mutual powers are real and equal:

$$[2.143]\ W_{yx} = W_{xy} = R_{xy}(0) = K_{xy}(0) + \bar{x}\bar{y};$$

⇒ the limitation is valid:

$$[2.144]\ R_{xy}^2(\nu) \le W_{xx}\, W_{yy};$$

⇒ the autocorrelation and autocovariance sequences are real and even:

$$[2.145]\ R_{xx}(\nu) = R_{xx}(-\nu),$$

$$[2.146]\ K_{xx}(\nu) = K_{xx}(-\nu);$$

⇒ the power is given by:

$$[2.147]\ W_{xx} = R_{xx}(0) = K_{xx}(0) + \bar{x}^2;$$

⇒ the limitation is valid, with maximum in ν=0:

$$[2.148]\ R_{xx}^2(\nu) \le W_{xx}^2.$$

3 REPRESENTATION OF SIGNALS

3.1 REPRESENTATION OF SIGNALS IN TIME SERIES

3.1.1 Fourier series of periodic signals

A generic periodic power signal x(t), with period T_0, can be represented utilizing a discrete countable set of infinite complex unit harmonics:

$$[3.1]\ e^{jk\Omega_0 t} = \cos k\Omega_0 t + j \sin k\Omega_0 t,$$

with k integer and $\Omega_0=2\pi/T_0$. Through a linear combination with constant complex coefficients of such functions, the representation:

$$[3.2]\ x(t) = \sum_k C_k e^{jk\Omega_0 t},$$

is achieved (as usual without extremes the summation is intended to be extended from $-\infty$ to $+\infty$) and named development in ***Fourier series***, with the ***Fourier coefficients*** C_k. It is possible to note that such development allows to break up the periodic signal into infinite addends which constitute its ***harmonic components***, at the frequencies $f_k=k/T_0$ and with amplitudes $|C_k|$ and phases $\arg\{C_k\}$.
To figure out the value of the Fourier coefficients C_k let's consider the particular energy functions with limited duration:

$$[3.3]\ \psi_k(t) = e^{jk\Omega_0 t}\ \mathrm{rect}(t/T_0),$$

obtained from the unit complex harmonics truncated in the interval $(-T_0/2, T_0/2)$, which compose an orthogonal set, with the common value of the energy $E_{\psi k \psi h}=T_0$ (see section 2.5.1.6); then, let's exploit for each k the correlation between the signal x(t) and each $\psi_k(t)$, operation which highlights the affinity between the periodic signal and the k^{th} function of the orthogonal set. In fact, exchanging the order of intergal and summation, it is achieved:

$$[3.4]\ (x,\psi_k) = \int \sum_h C_h e^{jh\Omega_0 t} e^{-jk\Omega_0 t} \mathrm{rect}\left(\frac{t}{T_0}\right) dt = \sum_h C_h (\psi_h,\psi_k) = T_0 C_k,$$

from which it is possible to obtain $C_k=(x,\psi_k)/T_0$ and thus the aimed expressions of the Fourier coefficients:

$$[3.5]\ C_k = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t) e^{-jk\Omega_0 t}\, dt,$$

where the correlation can be extended just to the interval $(-T_0/2, T_0/2)$, removing from the integrand the function $\mathrm{rect}(t/T_0)$, considering that all the functions $\psi_k(t)$ are equal to zero outside that interval. Thus, the coefficients C_k represent the measure of the affinity between the periodic signal x(t) and each $\psi_k(t)$.
Inserting the representation in series [3.2] into the general expression of the power of the periodic signals (see [2.65]), the following expression is obtained:

$$[3.6]\ W_{xx} = \frac{1}{T_o} \sum_k \sum_h C_k C_h^* \int_{-T_0/2}^{T_0/2} e^{j(k-h)\Omega_0 t}\, dt.$$

The integrand harmonic functions for $h\neq k$ have periods which are submultiples of T_0 and thus the integrals are all zero, excluding that for h=k, which is equal to T_0; as a consequence the following property of the Fourier coefficients is achieved:

$$[3.7]\ \sum_k |C_k|^2 = W_{xx}.$$

3.1.2 Representation in series of orthogonal functions

3.1.2.1 Energy signals represented by means of a base

As well as the case just considered for periodic signals, once fixed a certain class of signals on the basis of the definition domain and other properties, also energy signals can be represented in time series through the use of proper auxiliary functions, belonging, as well as those considered in the previous case, to a set with determined characteristics.
When the utilized functions belong to a discrete set, as in the case of the development in Fourier series for periodic signals, the case is of particular interest and will be considered hereinafter.
The functional space of energy signals, indicated with $L_2(T)$ where the definition domain T can even be extended to the entire time axis, other than linear and normal is a ***metric space*** once defined the ***distance between two*** generic ***signals*** x(t) and y(t) belonging to it:

$$[3.8]\ d(x,y) \triangleq \left\| x(t)-y(t) \right\|_2 = \left[\int_T \left| x(t)-y(t) \right|^2 dt \right]^{1/2}.$$

Let's consider a discrete set of functions $\{\psi_k(t)\}$, with subscript k=1, 2,.., N, belonging to $L_2(T)$ which respects the orthonormality relation:

$$[3.9]\ (\psi_k,\psi_h) = \int \psi_k(t)\psi_h^*(t)dt = \begin{cases} 0\ , & \forall\ h \neq k \\ 1\ , & h = k \end{cases}.$$

Considering a signal x(t), belonging to $L_2(T)$ and defined for any value t in the domain T, on the basis of the set $\{\psi_k(t)\}$ it is possible to realize its representation in time series, in general not perfect, by means of the linear combination of the orthonormal functions $\psi_k(t)$:

$$[3.10]\ \tilde{x}(t) = \sum_{k=1}^{N} \alpha_k \psi_k(t),$$

where the coefficients α_k, that can be calculated on the basis of the minimization of the mean squared error of approximation criterion, i.e. making minimum the square of the distance $d(x, \tilde{x})$, are the results of the scalar products (the reader interested to the demonstration can go to section 3.1.2.2):

$$[3.11]\ \alpha_k = (x,\psi_k) = \int_T x(t)\psi_k^*(t)dt.$$

It is possible to note that to calculate the coefficients α_k the signal x(t) must be known over the whole domain T, i.e. the representation concerns the deterministic signals.
The set $\{\psi_k(t)\}$ is considered satisfactory to represent a certain class of signals x(t) if the accuracy of the representation improves indefinitely increasing the number N of the utilized finctions; the orthonormal set $\{\psi_k(t)\}$ constitutes then a ***base*** for the considered class of signals. Thus, it is usual to write that for N which tends to infinite:

$$[3.12]\ x(t) = \sum_k X_k \psi_k(t), \text{ con } X_k = \alpha_k = (x,\psi_k),$$

where the change in the notation α_k to X_k indicated that the satisfactory condition occurrs and the equality in the representation is intended in the sense of the linear mean approximation(1); the following property is then valid:

$$[3.13]\ \sum_k |X_k|^2 = E_{xx}.$$

(1) It is possible to state that the points where the representation doesn't converge to the value of the signal are included in an empty set with respect to T.

It is worth to highlight that for particular classes of signals the satisfactory representation may occur even with a finite number N of orthogonal functions; in that case the set $\{\psi_k(t)\}$ constitutes a complete system, the finite number N is named ***dimension*** of the base and the representation assumes the following expression:

$$[3.14]\ x(t) = \sum_{k=1}^{N} X_k\, \psi_k(t), \text{ con } X_k = \alpha_k = (x, \psi_k),$$

Then, instead of [3.13], the following expression is obtained:

$$[3.15]\ \sum_{k=1}^{N} |X_k|^2 = E_{xx}.$$

The considered representations in time series can be applied also utilizing a set of functions just orthogonal, i.e. with energies $E_{\psi k \psi h} \neq 1$. As a consequence, in such a case the expressions of the coefficients are modified and become:

$$[3.16]\ X_k = \alpha_k = \frac{1}{E_{\psi_k \psi_k}} (x, \psi_k) = \frac{1}{E_{\psi_k \psi_k}} \int_T x(t) \psi_k^*(t) dt,$$

with the following properties:

$$[3.17]\ \sum_k |X_k|^2\ E_{\psi k \psi k} = E_{xx},\ \sum_{k=1}^{N} |X_k|^2\ E_{\psi k \psi k} = E_{xx}.$$

As above mentioned, the coefficients X_k can be achieved only if the evolution of the signal x(t) is known over the entire domain T, i.e. the representations in time series through a set of orthogonal functions concern the deterministic signals. Nevertheless, it is possible to consider the [3.12] and [3.14] as explicit representations of a time continuous random process, assuming that X_k are random variables.

3.1.2.2 Optimization of the representation with orthogonal functions

The criterion to choose the coefficients α_k to optimize the representations $\tilde{x}(t)$ is the minimization of the square of the its distance from the signal x(t), expressed by:

$$[3.18]\ d^2(x, \tilde{x}) = \int_T |x(t) - \tilde{x}(t)|^2\, dt = (x - \tilde{x}, x - \tilde{x}) = (x,x) - (x, \tilde{x}) - (x, \tilde{x})^* + (\tilde{x}, \tilde{x}).$$

Utilizing [3.10] and taking into account the orthogonality of the set $\{\psi_k(t)\}$ it is possible to achieve:

$$[3.19]\ d^2(x, \tilde{x}) = (x,x) - \sum_{k=1}^{N} \alpha_k^* (x, \psi_k) - \sum_{k=1}^{N} \alpha_k (x, \psi_k)^* + \sum_{k=1}^{N} \sum_{h=1}^{N} \alpha_k \alpha_h^* (\psi_k, \psi_h) =$$

$$= E_{xx} - 2 \sum_{k=1}^{N} \Re\left\{ \alpha_k (x, \psi_k)^* \right\} + \sum_{k=1}^{N} |\alpha_k|^2\, E_{\psi_k \psi_k} =$$

$$= E_{xx} - 2 \sum_{k=1}^{N} \Re\{\alpha_k\} \Re\{(x, \psi_k)\} - 2 \sum_{k=1}^{N} \Im\{\alpha_k\} \Im\{(x, \psi_k)\} + \sum_{k=1}^{N} [\Re\{\alpha_k\}]^2\, E_{\psi k \psi k} + \tilde{g}(t)\, E_{\psi k \psi k}.$$

The value d^2_{min} of the square of the distance is obtained when the partial derivatives of [3.19] with respect to $\Re\{\alpha_k\}$ and $\Im\{\alpha_k\}$ are equal to zero; the following system is obtained:

$$[3.20]\ \sum_{k=1}^{N} \Re\{(x, \psi_k)\} - \sum_{k=1}^{N} \Re\{\alpha_k\}\, E_{\psi k \psi k} = 0,$$

$$[3.21]\ \sum_{k=1}^{N} \Im\{(x,\psi_k)\} - \sum_{k=1}^{N} \Im\{\alpha_k\} E_{\psi k \psi k} = 0,$$

which provides the optimum solution expressed in [3.16].
With the identified solution and still due to the orthogonality of the set $\{\psi_k(t)\}$, for any k it is possible to get:

$$[3.22]\ (x-\tilde{x}, \psi_k) = (x,\psi_k) - (\tilde{x},\psi_k) = \alpha_k E_{\psi k \psi k} - \sum_{h=1}^{N} \alpha_h (\psi_h,\psi_k) = 0,$$

that means that the error signal $x(t)-\tilde{x}(t)$ of the imperfect representation is orthogonal to any function $\psi_k(t)$ and, as a consequence, it is orthogonal also to $\tilde{x}(t)$. This implies the relation:

$$[3.23]\ d^2_{min} = \|x\|^2 - \|\tilde{x}\|^2 = E_{xx} - \sum_{k=1}^{N} |\alpha_k|^2 E_{\psi k \psi k}.$$

The representation in time series of orthogonal functions is thus satisfactory for a given class of signals if

$$\lim_{N\to\infty} d_{min} = 0\ ,$$

i.e. if the Parceval equality is verified:

$$[3.24]\ \sum_{k=1}^{\infty} |\alpha_k|^2 E_{\psi k \psi k} = E_{xx}.$$

3.1.3 Fourier series of signals with limited duration

In the functional space L_2 the orthogonal set of the complex harmonics truncated in the interval (-T/2, T/2) is considered:

$$[3.25]\ \psi_k(t) = e^{jk\Omega t}\ \mathrm{rect}\left(\frac{t}{T}\right),$$

where k is integer, $\Omega=2\pi/T$ and all the $\psi_k(t)$ have the same energy $E_{\psi k \psi k}=T$; it is worth to note that, being T generic, such functions broaden those utilized in the case of periodic signals. Considering any energy signal g(t) with finite duration D, the representation in time series like [3.10], can be obtained but it will be in general imperfect:

$$[3.26]\ \tilde{g}(t) = \sum_{k=1}^{N} \alpha_k\ e^{jk\Omega t}\ \mathrm{rect}\left(\frac{t}{T}\right),$$

where the coefficients α_k are given by:

$$[3.27]\ \alpha_k = \frac{1}{E_{\psi_k \psi_k}} (x,\psi_k) = \frac{1}{T} \int_{-T/2}^{T/2} g(t) e^{-jk\Omega t}\, dt\ ,$$

having removed from the integrand the rectangular function after having properly replaced the integration extremes.
If the series is extended up to infinite, i.e. for $N\to\infty$, the considered representation becomes satisfactory for the class of energy signals which are zero out of an interval (t_m, t_M) all contained within (-T/2, T/2), condition which necessarily implies that the duration $D= t_M - t_m$ is not greater than T; in Figura 3.1 a signal of the mentioned class is shown. In fact, considering such a waveform g(t) as that one from which the periodic function $x(t)=rep_T[g(t)]$ is generated, then, it is allowed to express g(t) as:

$$[3.28]\ g(t) = x(t)\,\mathrm{rect}\left(\frac{t}{T}\right);$$

then, utilizing the representation of x(t) in Fourier series (see [3.2]), it is possible to express g(t) as:

$$[3.29]\ g(t) = \sum_k C_k e^{jk\Omega t}\ \mathrm{rect}\left(\frac{t}{T}\right),$$

which, compared with the [3.26], shows that the representation, named ***Fourier series with limited duration***, is satisfactory for the considered class of signals, with duration $D \leq T$, as long as the Fourier coefficients $C_k=\alpha_k$ are used, i.e the following expression is valid:

$$[3.30]\ C_k = \frac{1}{T}\int_{-T/2}^{T/2} g(t)e^{-jk\Omega t}\,dt\ .$$

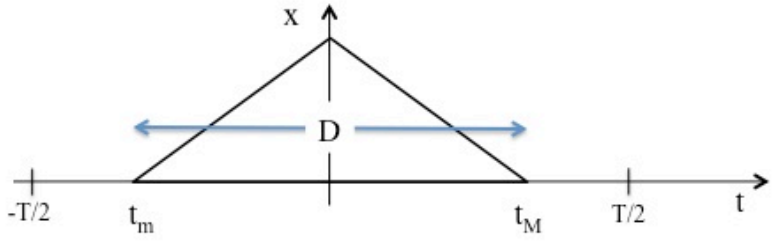


Figura 3.1: Energy signal with limited duration within the interval (-T/2, T/2).

Moreover, recalling the [3.17], the following property is valid:

$$[3.31]\ T\sum_k \left|C_k\right|^2 = E_{gg},$$

where as usual E_{gg} is the energy of the signal g(t).
It is important to highlight that for a given energy signal of the considered class the development in Fourier series is not unique: in fact, varying the duration T of the truncated unit complex harmonics, as long as the choice respects the condition that the interval (-T/2, T/2) includes (t_m, t_M), the coefficients C_k are different as well as, as a consequence, the truncated harmonic components which, summing up, rebuild the waveform of the signal g(t).
Moreover, it is possible to note that wherever the finite interval (t_m, t_M), out of which the signal g(t) is zero, is positioned, and independently on its finite duration D, it is always possible to get a satisfactory representation in Fourier series but ensuring that the orthogonal functions are different from zero in an interval of duration T which includes entirely the interval (t_m, t_M); as an example it is possible to consider in the [3.25] the truncation function rect[(t-t_0)/D] with opportune time shift t_0, at the center of the interval (t_m, t_M), as shown in Figura 3.2.

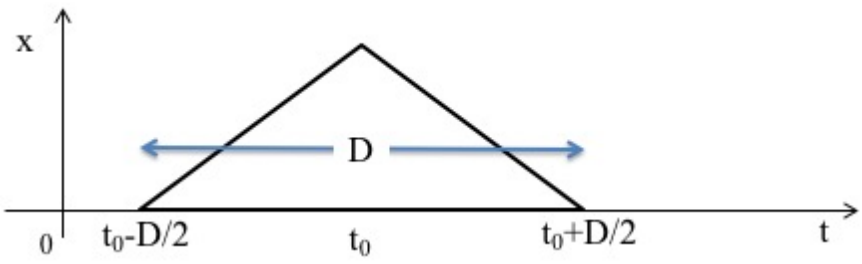


Figura 3.2: Energy signal with limited duration exactly contained in the interval (t_0-T/2, t_0+T/2).

Finally, it is worth to highlight that the imperfect representation with a finite number N of addends, applicable also in the case of signals only practically time limited, can be useful in case of numerical calculus, but only if a sufficient approximation is achievable.

3.1.4 Representation by means of samples interpolation

In the functional space L_2 let's consider the set of the real orthogonal functions, with limited duration and generic parameter T_c:

$$[3.32]\ \psi_k(t) = \mathrm{sinc}\left(\frac{t}{T_c}-k\right),$$

having all the same value of the energy $E_{\psi k \psi k}=T_c$. Referring to any real energy signal and with finite duration, i.e. belonging to the same space L_2, the representation in time series like [3.10], in general imperfect is acheved:

$$[3.33]\ \tilde{x}(t)=\sum_{k=1}^{N}\alpha_k\ \mathrm{sinc}\left(\frac{t}{T_c}-k\right),$$

where the corfficients α_k are real and given by:

$$[3.34]\ \alpha_k=\frac{1}{E_{\psi_k\psi_k}}\ (x,\psi_k)=\frac{1}{T_c}\int x(t)\,\mathrm{sinc}\left(\frac{t}{T_c}-k\right)dt\ .$$

The particular property of the considered orthogonal functions, on the basis of which in the instant kT_c the k^{th} function is equal to one and all the others with $h\neq k$ are equal to zero, implies that the [3.33] for $t=t_k=kT_c$ provides $\alpha_k=\tilde{x}(kT_c)$: thus, the coefficients are the instantaneous values of the representation $\tilde{x}(t)$ in the equispaced instants t_k.

Aiming to identity the conditions to make the considered representation satisfactory, prelimirarly it is highlighted that usually an infinite number N of orthogonal functions must be utilised because the signals to be represented are unlimited.

With an infinite N, then, pushing T_c to zero the obtained representation is satisfactory.

In fact, recalling the [2.57], from [3.34] it is possible to obtain:

$$[3.35]\ \alpha_k=\lim_{T_c\to 0}\frac{1}{T_c}\int x(t)\,\mathrm{sinc}\left(\frac{t}{T_c}-k\right)dt\ =\int x(t)\,\delta(t-kT_c)\,dt\ =x(kT_c),$$

and since for the properties of the considered functions also $\alpha_k=\tilde{x}(kT_c)$ is achieved, the representation $\tilde{x}(t)$ in instants extremely close between each other results to have the same values of the signal x(t), so that, at least for $T_c\to 0$, it is possible to achieve

$$[3.36]\ x(t)=\sum_k c_k\,\mathrm{sinc}\left(\frac{t}{T_c}-k\right),$$

where the real oefficients are the values of the signal x(t), or ***samples*** of the signal

$$[3.37]\ c_k=x(kT_c),$$

in the equispaced ***sampling instants*** $t_k=kT_c$.

As demonstrated later on, the above considered representation, released by Nyquist, is satisfactory with T_c not infinitesimal for a very large class of real energy signals with limited duration, as long as T_c is chosen lower than a definable upper limit value T_N, named ***Nyquist interval***. The representation in time series [3.16] rebuilds in this way the signal through ***sample interpolation*** c_k with ***sampling interval*** $T_c<T_N$, by means of infinite functions obtained replicating the sinc pulse with opportune time shift kT_c (see Figura 3.3).

Recalling the [3.17], in addition the following propoerty is valid:

$$[3.38]\ T_c\sum_k c_k^2\ =E_{xx},$$

where, as usual, E_{xx} indicates the energy of the signal x(t).

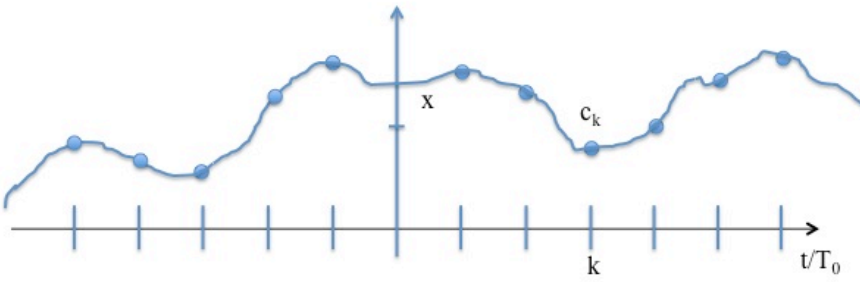


Figura 3.3: Representation of a signal by interpolation of its samples.

As in the case of the Fourier series of limited duration signals, the representation by sample interpolation is not unique: varying the sampling interval T_c, which determines the selection of the orthogonal functions, respecting the condition $T_c<T_N$ the samples c_k are in fact different as well as the interpolating functions $\psi_k(t)$.
Moreover, the representation can be applied also to an opportune class of power signals, for which a non zero value of the Nyquist interval can be determined; in such a case the satisfactory accuracy can be demonstrated implementing the interpolation in series for a finite value of N and, then, passing to the limit for $N\to\infty$. As an immediate example, the [3.33] is satisfactory in the elementary case x(t)=c real constant; in fact, for $N\to\infty$, taking into account the properties of the considered orthogonal functions (see section 2.5.1.6), all having area equal to T_c and the summation is equal to 1 for any t, from [3.34] the coefficients α_k are all equal to the constant c and thus from [3.33] $\tilde{x}(t)=c=x(t)$ is obtained.
Finally, it is worth to highlight that the imperfect representation of the truncated evolution of the signal x(t), obtained with the actual finite N samples c_k, can be usefully utilized in case of numerical calculus; in such a case the approximation of the signal $\tilde{x}(t)$ achieved through interpolation is not adequate only in the final segments.

3.2 Representation in the space of signals

Let's consider again the representation $\tilde{x}(t)$, with coefficients $\alpha_k=(x,\psi_k)$. This latter expression, thanks to the orthonormality of the familiy $\{\psi_k(t)\}$ implies that the N functions $\psi_k(t)$ corrispond to the following n-tuple of values

$$(1,0,..,0),\ (0,1,..,0),..,\ (0,0,..,1),$$

to which it is possible to associate N versors, $\boldsymbol{\psi}_k$, all orthogonal among one another, that define an Euclidean space with N dimensions, named ***space of the signals***. If the N functions $\psi_k(t)$ are real, the relation $\alpha_k=(x,\psi_k)$ establishes a correspondence between a real signal x(t) and a n-tuple of real values, $(\alpha_1, \alpha_2, ..., \alpha_N)$, to which it is possible to associate a vector $\tilde{\mathbf{x}}$, that is the projection of the signal in the N dimensional space just defined above. Every coefficient α_k is thus the k^{th} component of the vector $\tilde{\mathbf{x}}$ along $\boldsymbol{\psi}_k$, i.e the following relation is achieved:

$$[3.39]\ \alpha_k = \tilde{\mathbf{x}}\cdot\boldsymbol{\psi}_k,$$

which even more justifies the denomination of scalar product of the operation $\alpha_k=(x,\psi_k)$ in the time domain.
As already previously highlighted, in general an infinite number of orthonormal functions is needed to satisfactorily represent any signal in its definition interval; nevertheless, the desired accuracy can be reached for particular cathegories of signals, even for finite N. In such a case, the correspondence between a signal, which can be represented in the time domain with the following representation:

$$[3.40]\ x(t) = \sum_{k=1}^{N} x_k \psi_k(t),\ x_k = (x,\psi_k),$$

and its vectorial representation on the euclidean base $\{\boldsymbol{\psi}_k\}$, becomes biunique, as illustrated in Figura 3.4, indicated with the notation:

$$[3.41]\ \mathbf{x} = \sum_{k=1}^{N} x_k \boldsymbol{\psi}_k\ ,\ x_k = \mathbf{x}\cdot\boldsymbol{\psi}_k.$$

It is also possible to obtain the following relation between the vector modulus $|\mathbf{x}|$ and the energy of the signal:

$$[3.42]\ |x|^2 = \mathbf{x}\cdot\mathbf{x} = \sum_{k=1}^{N} x_k^2 = \mathcal{E}_{xx}.$$

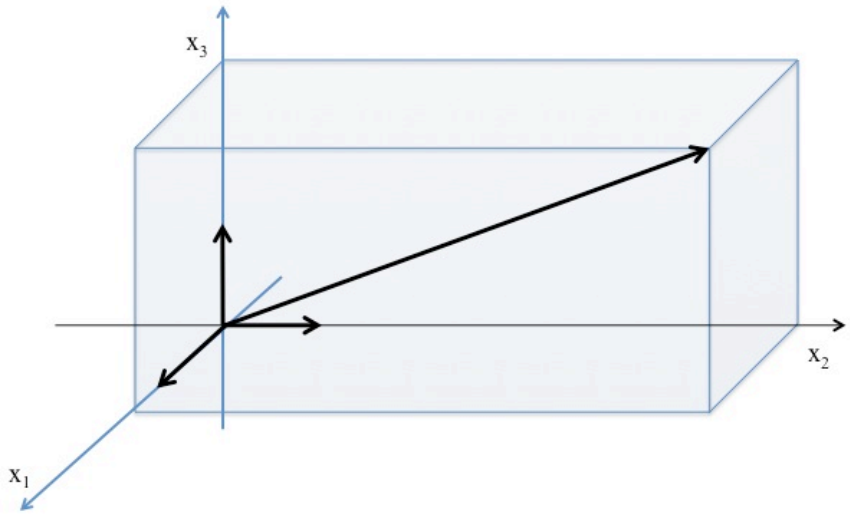


Figura 3.4: Representation of a signal in the signal space

The introduced representation is very useful because there is no time variable.
Considering a pair of signals, x(t) and y(t), which both can be represented through a N dimensional base $\{\psi_k(t)\}$ with the n-tuple x_k and y_k (k = 1, 2,..,N), the scalar product of the two corresponding vectors, **x** and **y**, in the signal space, is obtained as follows:

$$[3.43]\ \mathbf{x}\cdot\mathbf{y} = \sum_{k=1}^{N} x_k y_k\ .$$

Introducing into the definition of the scalar product in the time domain the signals represented as linear combination of the base functions $\psi_k(t)$, the same result is obtained, i.e. achieving the following relation:

$$[3.44]\ \mathbf{x}\cdot\mathbf{y} = (x,y).$$

In fact, due to the orthonormality of the real base functions it is possible to obtain:

$$[3.45]\ (x,y) = \int x(t)\,y^*(t)\,dt = \int \sum_{k=1}^{N} x_k \psi_k(t) \sum_{h=1}^{N} y_h \psi_h(t)\,dt = \sum_{k=1}^{N}\sum_{h=1}^{N} x_k y_h \int \psi_k(t)\,\psi_h(t)\,dt = \sum_{k=1}^{N} x_k y_k\ .$$

Note that in case (x,y)=0 the two representative vectors **x** and **y** are actually orthogonal in the signal space.

3.2.1 Gram-Schmidt Orthogonalization procedure

It is worth to recall an important case of signal cathegory which can be satisfactorily represented through an orthonormal base, i.e. in a finite dimension space: it is the generic discrete set, $\{x_i(t)\}$, composed of a finite number M of signals, for which the Gram-Schmidt orthogonalization method guarantees the full validity of the representation utilizing a finite number N≤M of particular

orthonormal functions. The method consists in applying some operations on every $x_i(t)$ to carry out a base function $\psi_i(t)$ which is orthonormal to all the functions previously determined. In the first step the following relation is imposed:

$$[3.46]\ \psi_1(t) = \frac{x_1(t)}{\|x_1(t)\|}.$$

In the second step the signal orthogonal to $x_1(t)$ is determined:

$$[3.47]\ w_2(t) = x_2(t) - (x_2, \psi_1)\,\psi_1(t),$$

achieving in this way the second base function:

$$[3.48]\ \psi_2(t) = \frac{w_2(t)}{\|w_2(t)\|}.$$

At the k^{th} step the signal orthogonal to all the previous ones is achieved:

$$[3.49]\ w_k(t) = x_k(t) - \sum_{i=1}^{k-1}(x_k, \psi_i)\,\psi_i(t),$$

as well as the corresponding base function:

$$[3.50]\ \psi_k(t) = \frac{w_k(t)}{\|w_k(t)\|}.$$

The $\psi(t)$ are linear combinations of the signals $x_i(t)$; if these are linearly independent, then all the $\psi(t)$ exist; otherwise just $N \leq M$ of them are not identically equal to zero.

3.3 REPRESENTATION OF TIME CONTINUOUS SIGNALS IN THE FREQUENCY DOMAIN

3.3.1 Linear transformation of time continuous signals

Let's consider a complex function x(v), with codomain (range) X, of an independent real variable v and a complex function w(z), with codomain (range) W, of an independent variable z, in general complex and different from v. In general, a ***transformation***, indicated with the notation

$$[3.51]\ w(z) = T\{x(v), z\},$$

is defined as a correspondence law established by a mathematical entity, named ***operator***, between the input functions x(v) and the output ones w(z). The x(v) is named ***operand***, the w(z) is named ***transform***.
Hereinafter the attention is limited on the ***linear transformations***, for which the operator satisfies the conditions of additivity, so that the well known principle of superimposition of the effects is valid:

$$[3.52]\ T\{[c_1x_1(v) + c_2x_2(v)], z\} = c_1T\{x_1(v), z\} + c_2T\{x_2(v), z\}.$$

In presence of linearity and supposing that the domains of the independent variables are continuous, the transformation can assume the expression:

$$[3.53]\ T\{x(v), z\} = \int x(v)\,h_n(z,v)\,dv = \left(x(v), h_n^*(z,v)\right),$$

where the function $h_n(z, v)$, usually complex, is named ***nucleus*** of the linear tranformation.
Assuming that the operand x(t) is a time continuous signal (v=t) and that the variable z is a time variable, specifically a time continuous ***linear transformation of signal*** is achieved and the transform, usually indicated with capital letter:

$$[3.54]\ X(z) \equiv T\{x(t), z\} = \int x(t) h_n(z,t)\, dt,$$

assumes the meaning of representation of the signal x(t) in the continous domain z, different from the time domain.
A typical example of linear transformation of time continuous signal is that offered by its representation in the frequency domain with the Fourier transform, in which the variable $z=f=\omega/2\pi$ is real and the nucleous of the transformation becomes the function:

$$[3.55]\ h_n(f, t) = e^{-j\omega t}.$$

A second example is the Laplace transform, which allows the representation of signals in the domain of the complex variable $s=\alpha+j\omega$.

3.3.2 Fourier transform and its properties

3.3.2.1 Fourier transform

Both in the case of periodic signals and in the case of finite duration signals a harmonic type representation was obtained, in which the signal is the sum of a discrete series of complex harmonic components, infinitely extended over the time or truncated.
Also in the case of bilateral energy signals, with unlimited duration, a harmonic analysis can be applied, named ***Fourier transform*** of the signal x(t), as long as the continuous set of functions:

$$[3.56]\ e^{j\omega t} = \cos\omega t + j \sin\omega t,$$

incoherent among one another, composed of the complex harmonics with unit modulus at the generic frequencies $f=\omega/2\pi$ and with phases equal to zero is considered and as long as instead of the discrete complex coefficients the complex function:

$$[3.57]\ X(f) = \int x(t) e^{-j\omega t}\, dt = \left(x(t), e^{j\omega t}\right),$$

belonging to the functional space L_2 in the frequency domain is considered.
The transformation considered in [3.57] is reversible, i.e. there is the ***inverse Fourier transform***:

$$[3.58]\ x(t) = \int X(f) e^{j\omega t}\, df = \left(X(f), e^{-j\omega t}\right),$$

which is a satisfactory representation with reference to the continuous set of the functions [3.56]. In fact, it is possible to note that in [3.58] the signal is composed of infinite, and infinitesimally close, complex harmonic components $X(f)\ e^{j\omega t}\, df$.
Often the following notations are used:

$$X(f) \equiv F\{x(t)\},\ x(t) \equiv F^{-1}\{X(f)\}.$$

The biunivocal correspondence between the signal in the functional space L_2 in the time domain and its Fourier transform in the functional space L_2 in the frequency domain is then indicated hereinafter with the notation:

$$x(t) \Leftrightarrow X(f).$$

It is possible to note that to get the Fourier transform it is mandatory to know the signal x(t) over the entire time domain, implying that the considered representation is applicable to the deterministic signals; nevertheless, it may be formally applicable also to random energy signals as long as the randomness concerns only some parameters.
The complex function X(f), usually expressed in modulus, |X(f)|, and in argument, arg{X(f)}, can be graphically represented with a pair of traces |X(f)| and arg{X(f)}, respectively on the planes |*F*|, f and arg{*F*}, f, named ***amplitude spectrum*** and ***phase spectrum*** of the signal (see example in Figura 3.5). The Fourier transform is thus indicated also as ***spectrum*** of the signal.

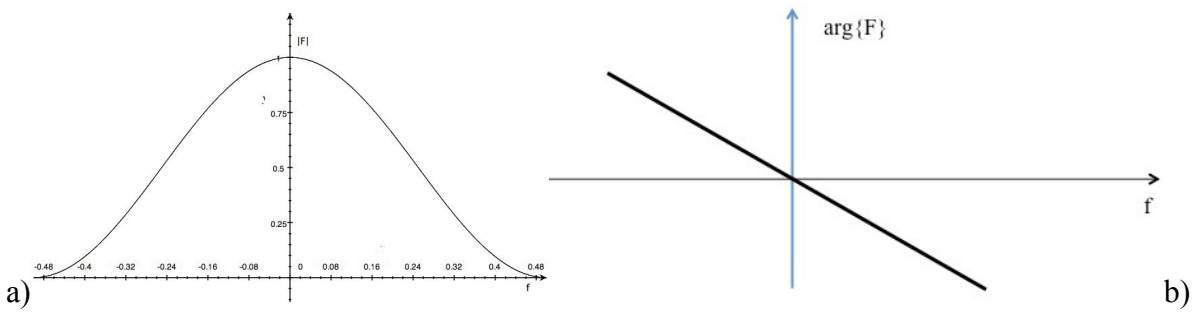


Figura 3.5: Amplitude spectrum (a) and phase spectrum (b) od the signal.

The integral that provides the function X(f) converges for the energy signals belonging to the functional space L_2 in the time domain, i.e. which can be quadratically integrated, but it also converges in the case of signals belonging to L_1, which includes signals that can be integrated in modulus. Moreover, the Fourier transform can exist also for other kind of signals, as it will be shown later on.

3.3.2.2 General properties of the Fourier transform

The operators that lead to the Fourier transform and inverse transform first of all enjoy the ***linearity*** property, i.e. to a linear combination with constant coefficients of many signals corresponds the linear combination, with the same coefficients, of the respective spectra of the signals, and vice versa; then it is allowed to write:

$$[3.59]\ c_1x_1(t)+c_2x_2(t)+... \Leftrightarrow c_1X_1(f)+c_2X_2(f)+...$$

In particular, for a complex signal, the following biunivocal correspondence is valid:

$$[3.60]\ x_R(t)+jx_I(t) \Leftrightarrow F\{x_R(t)\}+jF\{x_I(t)\}.$$

Moreover, for the pair of operators the ***duality*** property is valid; in fact, assuming that $F\{x(t)\} = X(f)$, exchanging the time variable and the frequency variable in the functions the following relation is achieved:

$$[3.61]\ F\{X(t)\}= x(-f).$$

In Tabella 3.1 other meaningful general properties of the Fourier transform are listed. The reader is invited to follow as an exercise the demonstrations reported in section 3.3.3.

Tabella 3.1: Some meaningful properties of the Fourier transform

PROPERTY	SIGNAL	TRANSFORM
Scaling	$x(at)$	$\frac{1}{\lvert a\rvert}X\left(\frac{f}{a}\right)$
Sign inversion	$x(-t)$	$X(-f)$
Time shift	$x(t+t_0)$	$X(f)\,e^{j2\pi ft_0}$
Frequency shift	$x(t)\,e^{j2\pi f_0t}$	$X(f-f_0)$
Conjugation in time	$x^*(t)$	$X^*(-f)$
Conjugation in frequency	$x^*(-t)$	$X^*(f)$
Time derivative	$\frac{d^n}{dt^n}x(t)$	$(j2\pi f)^n X(f)$
Frequency derivative	$(-j2\pi t)^n x(t)$	$\frac{d^n}{df^n}X(f)$
Integration in time	$\int x(t)dt$	$\frac{X(f)}{j2\pi f}$

3.3.2.3 Fourier transform of a real signal

In the frequent case of real signal, with $x(t)=x_R(t)$ and $x_I(t)=0$, which can be Fourier transformed, it is possible to obtain:

$$[3.62]\ X(f) = \int x_R(t)\cos\omega t\,dt - j\int x_R(t)\sin\omega t\,dt = X_R(f) + jX_I(f),$$

and exchanging f with -f:

$$[3.63]\ X(-f) = \int x_R(t)\cos\omega t\,dt + j\int x_R(t)\sin\omega t\,dt = X_R(f) - jX_I(f).$$

Thus, the transform X(f) of a real signal is a hermitian function, i.e. the particular property is valid:

$$[3.64]\ x(t) = x^*(t) \Leftrightarrow X(f) = X^*(-f);$$

the amplitude spectrum is thus an even function, $|X(f)|=|X(-f)|$, while the phase spectrum is an odd function, $\arg\{X(f)\}=-\arg\{X(-f)\}$.
From the observation of the evolutions of the spectra shown as examples in Figura 3.5, it is possible to find out that the the signal which the spectra correspond is real.
In addition, if the signal is real and even function, also its transform has the same peculiarity. i.e. the following property, even more particular, is valid:

$$[3.65]\ x(t) = x^*(t) = x(-t) \Leftrightarrow X(f) = X^*(f) = X(-f).$$

3.3.3 Demonstrations of some properties of thr Fourier transform

Scaling property

$$[3.66]\ F\{x(at)\} = \int x(at)e^{-j\omega t}\,dt\,;$$

changing the variable t'=at, it follows that:

$$[3.67]\ F\{x(at)\} = \frac{1}{a}\mathrm{sgn}(a)\int x(t')e^{-j\frac{\omega}{a}t'}\,dt' = \frac{1}{|a|}X\left(\frac{f}{a}\right);$$

Sign inversion property
Applying the scaling property with a=-1 it is easy to obtain:

$$[3.68]\ F\{x(-t)\} = X(-f)$$

Time shift property
Applying the definition:

$$[3.69]\ F\{x(t+t_0)\} = \int x(t+t_0)e^{-j\omega t}\,dt$$

changing the variable $t'=t+t_0$, it follows that:

$$[3.70]\ F\{x(t+t_0)\} = \int x(t')e^{-j\omega(t'-t_0)}\,dt' = e^{j\omega t_0}X(f)$$

Frequency shift property
It is possible immediately to achieve:

$$[3.71]\ F\left\{x(t)e^{j\omega_0 t}\right\} = \int x(t)e^{j\omega_0 t}e^{-j\omega t}\,dt = \int x(t)e^{-j(\omega-\omega_0)t}\,dt = X(f-f_0)$$

Conjugation in time property
It is possible immediately to achieve:

$$[3.72]\ F\left\{x^*(t)\right\} = \int x^*(t)e^{-j\omega t}\,dt = \left[\int x(t)e^{j\omega t}\,dt\right]^* = X^*(-f)$$

Conjugation in frequency property
It is possible to achieve:

$$[3.73]\ X^*(f)= \left[\int x(t)e^{-j\omega t}\,dt\right]^* = \int x^*(t)e^{j\omega t}\,dt$$

changing the variable t'=-t it is possible to achieve:

$$[3.74]\ X^*(f)= \int x^*(-t')e^{-j\omega t'}\,dt' = F\{x^*(-t)\}$$

Time derivation property
It is possible to achieve:

$$[3.75]\,F\left\{\frac{d^n x(t)}{dt^n}\right\} = \int \frac{d^n x(t)}{dt^n} e^{-j\omega t}\,dt\,;$$

integrating per parts, it is possible to achieve:

$$[3.76]\ F\left\{\frac{d^n x(t)}{dt^n}\right\} = \left[\frac{d^{n-1} x(t)}{dt^{n-1}} e^{-j\omega t}\right]_{-\infty}^{\infty} + j\omega \int \frac{d^{n-1} x(t)}{dt^{n-1}} e^{-j\omega t}\,dt$$

where the first term is zero because x(t) is an energy signal; iterating it is possible to achieve:

$$[3.77]\ F\left\{\frac{d^n x(t)}{dt^n}\right\} = (j\omega)^n \int x(t)e^{-j\omega t}\,dt = (j\omega)^n X(f)$$

Frequency derivation property
It is possible immediately to achieve:

$$[3.78]\ \frac{d^n X(f)}{df^n} = \frac{d^n}{df^n}\left[\int x(t)e^{-j\omega t}\,dt\right] = \int x(t)\frac{d^n e^{-j\omega t}}{df^n}\,dt = (-j\omega)^n X(f)$$

Integration in time property
It is possible to achieve:

$$[3.79]\ F\left\{\int_{-\infty}^{t} x(u)\,du\right\} = \int \int_{-\infty}^{t} x(u)\,du\; e^{-j\omega t}\,dt = \int \int_{-\infty}^{t} x(u)\,du\;\frac{1}{-j\omega} d\left(e^{-j\omega t}\right) =$$

integrating per parts, it is possible to achieve:

$$[3.80]\ F\left\{\int_{-\infty}^{t} x(u)\,du\right\} = \left[-\frac{e^{-j\omega t}}{j\omega}\int_{-\infty}^{t} x(u)\,du\right]_{-\infty}^{\infty} + \frac{1}{j\omega}\int x(t)e^{-j\omega t}\,dt = \frac{1}{j\omega}X(f)$$

if area [x(t)]=0, otherwise a further addend {area[x(t)δ(t)}/2 must be considered.

3.3.3.1 Fourier transform of the convolution

The introduced operator, that is very significant as will be evident later on, is applicable to a pair of energy signals as well as to an energy signal and a power signal; in the former case an energy signal is achieved while in the latter case a power signal is obtained.
Applying the Fourier transform operator to the convolution integral between a pair of energy signals and utilizing the time shift property, it is possible to achieve:

$$[3.81]\ (x(t) * y(t), e^{j\omega t}) = \iint x(v)\,y(t-v)\,e^{-j\omega t}\,dv\,dt = \int x(v)\left[\int y(t-v)e^{-j\omega t}\,dt\right]dv =$$
$$\int x(v)\,Y(f)e^{-j\omega v}\,dv = X(f)\,Y(f),$$

demonstrating the important property of ***time convolution***:

$$[3.82]\ F\{x(t)*y(t)\} = X(f)\,Y(f).$$

With similar procedure and applying the inverse Fourier transform operator to the convolution integral between the transforms X(f) and Y(f) of two energy signals, the ***frequency convolution*** property is demonstrated:

$$[3.83]\ F\{x(t)y(t)\} = X(f)*Y(f).$$

3.3.4 Relevant examples of Fourier transforms

3.3.4.1 Transform of the unit rectangular pulse and of the sinc signal

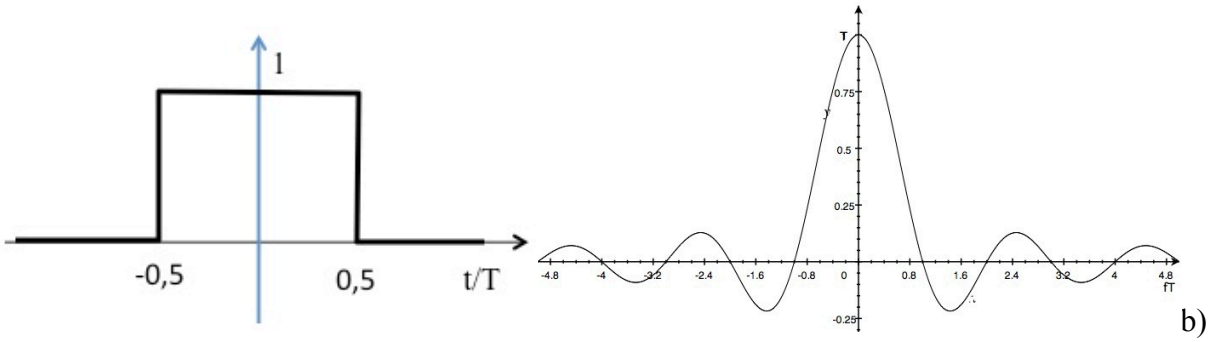
The unit rectangular impulse rect(t/T), with the time evolution shown in Figura 3.6a, is a typical example of signal belonging to the space L_2 and with limited duration. Applying the operator of the Fourier transform to the unit rectangular pulse the result is:

$$[3.84]\ (\text{rect}(t/T),\ e^{j\omega t}) = \int \text{rect}\left(\frac{t}{T}\right) e^{-j\omega t}\,dt = \int_{-T/2}^{T/2} e^{-j\omega t}\,dt = \frac{e^{-j\omega T/2} - e^{j\omega T/2}}{-j\omega},$$

determining the following expression of the Fourier transform:

$$[3.85]\ F\{\text{rect}(t/T)\} = T\,\text{sinc}(fT).$$

Being the signal real and even, the Fourier transform is real and even as well, with just the amplitude spetrum shown in Figura 3.6b.



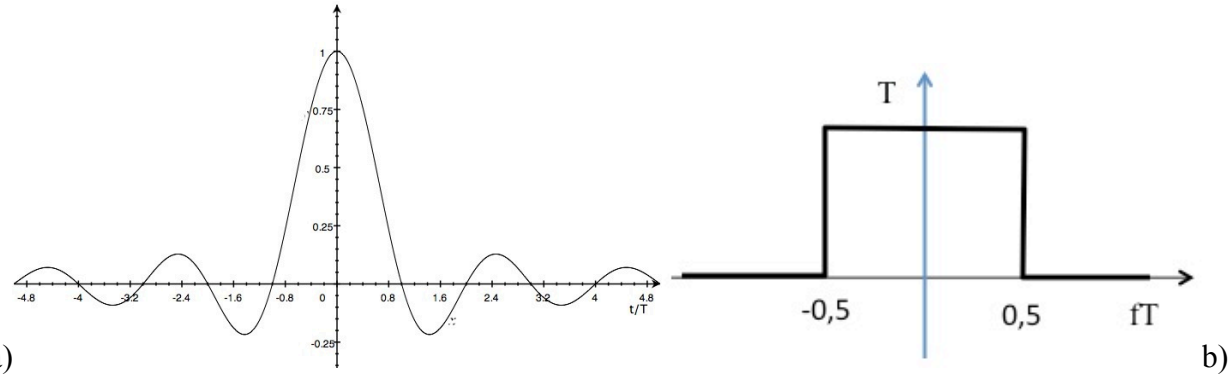
a) b)

Figura 3.6: Unit rectangular impulse in time (a) and its spectrum (b).

The signal sinc(t/T), with the waveform shown in Figura 3.7a, is a further typical example of signal belonging to the space L_2, but with finite duration. Starting from [3.67], applying the duality and scaling properties of the Fourier transform the following expression is directly obtained:

$$[3.86]\ F\{\text{sinc}(t/T)\} = T\,\text{rect}(fT),$$

still real and even, with just the amplitude spectrum shown in Figura 3.7b.



a) b)

Figura 3.7: Signal sinc(t/T) in time (a) and its spectrum (b).

Definitively, the waveforms of two considered signals are mutally exchanged in representing the time evolution and the frequency evolution of both. In particular, the former is limited in time and unlimited in frequency, while the latter is unlimited in time and limited in frequency.

3.3.4.2 Transform of particular signals

As already mentioned, the Fourier transform can converge also in case of particular signals, which can be integrated neither quadratically nor in modulus, thus belonging neither to the functional space L_2, nor to L_1.
In the case of the ideal pulse, recalling its sampling property, for example the result is:

$$[3.87]\ (\delta(t),\ e^{j\omega t}) = \int \delta(t) e^{-j\omega t}\, dt\ = e^{-j\omega t}\Big|_{t=0} = 1,$$

i.e. the Fourier transform of the Dirac pulse is constant in frequency and equal to 1:

$$[3.88]\ F\{\delta(t)\} = 1.$$

Considering the inverse Fourier transform applied to the ideal pulse in the frequency domain, δ(f), similarly the result is:

$$[3.89]\ F^{-1}\{\delta(f)\} = \int \delta(f) e^{j\omega t}\, df\ = e^{j\omega t}\Big|_{f=0} = 1,$$

which actually highlights the Fourier transform of a unit time constant signal:

$$[3.90]\ F\{1\} = \delta(f).$$

Aiming to determine the Fourier transform of other particular signals it is worth to consider the unit signal with exponential decay, $e^{-t/T}u(t)$ with T>0, which belongs to the space L_2; as the reader can achieve as an exercise, the transform of the considered signal is:

$$[3.91]\ F\{e^{-t/T}u(t)\} = \frac{T}{1+j\omega T}\ = \frac{T}{1+j2\pi fT}.$$

On the basis of such last result and considering the function sign(t) expressed as:

$$[3.92]\ \mathrm{sgn}(t) = \lim_{T\to\infty}\{e^{-t/T}u(t) - e^{t/T}u(-t)\},$$

its Fourier transform is:

$$[3.93]\ F\{\mathrm{sgn}(t)\} = \lim_{T\to\infty}\left\{\frac{T}{1+j\omega T} - \frac{T}{1-j\omega T}\right\} = \lim_{T\to\infty}\left\{\frac{-j2\omega T^2}{1+\omega^2T^2}\right\} = \frac{1}{j\pi f}.$$

Recalling the expression u(t)=[1+sgn(t)]/2 and the [3.72], the Fourier transform of the unit step is immediately determined:

$$[3.94]\ F\{u(t)\} = \frac{1}{2}\delta(f) + \frac{1}{j2\pi f}.$$

Also a generic complex harmonic $A e^{j(\Omega_o t+\varphi)}$, at frequency $f_0=\Omega_0/2\pi=1/T_0$, with amplitude A and phase φ, although it is a power signal, can be transformed with Fourier. Since it is possible to achieve:

$$[3.95]\ (A e^{j(\Omega_0 t+\varphi)}, e^{j\omega t}) = Ae^{j\varphi}\int e^{j(\Omega_0-\omega)t}\, dt,$$

calculating the integral with extremes -T/2 and T/2 and then passing to the limit for T→∞ (the interested reader can refer to [2.56]) in fact the result is:

$$[3.96]\ F\{Ae^{j(\Omega_o t+\varphi)}\} = A e^{j\varphi}\delta(f-f_0).$$

In the case of a generic real harmonic, taking into account that it is possible to express:

$$[3.97]\ A\cos(\Omega_0 t+\varphi)=\frac{1}{2}A\left[e^{j(\Omega_0 t+\varphi)}+e^{-j(\Omega_0 t+\varphi)}\right],$$

the Fourier transform is obtained:

$$[3.98]\ F\{A\cos(\Omega_0 t+\varphi)\}=\frac{1}{2}A\,e^{j\varphi}\,\delta(f-f_0)+\frac{1}{2}A\,e^{-j\varphi}\,\delta(f+f_0),$$

which is a hermitian function in the frequency domain, as supposed since the signal is real. The spectrum, equal to zero for any f except that for the discrete values $\pm f_0$, is named ***striped spectrum***, considering that it is conventionally drawn, as shown in Figura 3.8, with strips as high as the modulus or equal to the argument of the area of each discrete component.

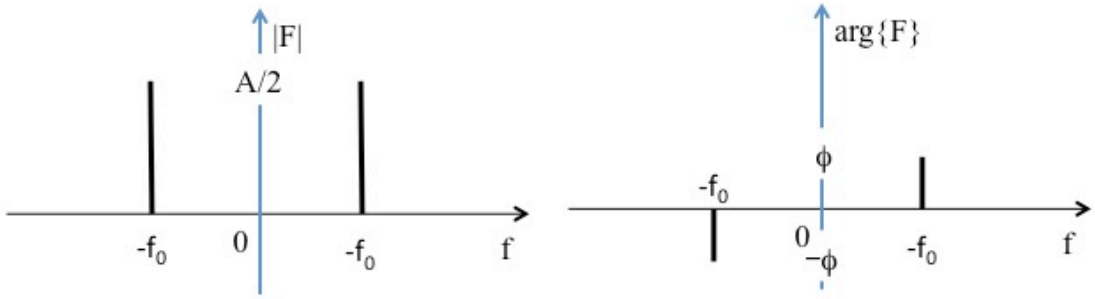


Figura 3.8: Conventional representation of the discrete amplitude (a) and phase (b) spectrum of a real harmonic signal.

3.3.5 Affinity between energy signals represented in the frequency domain

In section 2.5.1 interesting considerations concerning the affinity of two generic energy signals x(t) and y(t) have been developed, starting from the cross correlation function $C_{xy}(\tau)$ that can be calculated from their time evolution. Now, supposing to know the frequency representations of the signals, the aim is to find out if the same information can be carried out using the Fourier transforms X(f) and Y(f).

If the variable is changed from t to -v in the integral that defines the cross correlation function between a pair of energy signals, the function assumes the form of the particular convolution product (see **Errore. L'origine riferimento non è stata trovata.**]):

$$[3.99]\ C_{xy}(\tau)=(x(t+\tau),y(t))=x(\tau)*y^*(-\tau),$$

establishing in this way an interesting relation between the two functions. On the basis of the property of the convolution in time (see [3.82]) of the Fourier transform, then the result is:

$$[3.100]\ F\{C_{xy}(\tau)\}=X(f)\,Y^*(f).$$

As a consequence the following expression of the cross correlation function as a function of the spectra of the signals is obtained:

$$[3.101]\ C_{xy}(\tau)=F^{-1}\left\{X(f)Y^*(f)\right\}=\int X(f)Y^*(f)e^{j\omega\tau}\,df\ ;$$

considering in particular $\tau=0$, the scalar product assumes the expressions:

$$[3.102]\ (x,y)=\int x(t)y^*(t)dt=\int X(f)Y^*(f)df\,,$$

which shows that such a quantity in the frequency domain has a dual form with respect to its definition in the time domain.

The last member of the [3.101] allows to identify an interesting sufficient condition so that the two signals have no affinity, i.e. they are ***uncorrelated***: in fact, if the spectra of the two signals are different from zero in separated frequency intervals $X(f)Y^*(f)\equiv 0$ is obtained, becasue the integrand

is zero for any frequency, so that the cross correlation function is identically equal to zero for any τ, i.e. $C_{xy}(\tau)\equiv 0$ occurs.
With similar approach on the [3.102], the mentioned separation in frequency of a pair of signals is sufficient condition, but not necessary, for their orthogonality, as well as already stated for the separation in time, which impies $x(t)y^*(t)\equiv 0$.
It is worth to recall that the maximum affinity in wide sense occurs if and only if the signals are faithful between each other, i.e. the relation $y(t)=g\,e^{-j\Gamma}\,x(t-t_0)$ is valid; switching to the representations in frequency the correspondent relation is achieved:

$$[3.103]\ Y(f) = gX(f)\ e^{-j(\omega t_o+\Gamma)},$$

for which the generic family of faithful energy signals is characterized in the frequency domain by proportional amplitute spectra and by differences of the phase spectra with linear evolutions (proportional in the case of real signals with $g>0$).

3.3.6 Energy spectra

3.3.6.1 Energy spectral density of a signal

Considering $y(t)\equiv x(t)$ in the expression [3.101] the autocorrelation function is achieved as a function of the spectrum X(f) of the energy signal:

$$[3.104]\ C_{xx}(\tau) = F^{-1}\left\{X(f)X^*(f)\right\} = \int \left|X(f)\right|^2 e^{j\omega\tau}\,df\,.$$

Recalling that the value in the origin $C_{xx}(0)$, per $\tau=0$, provides the energy E_{xx} of the signal, the following expressions are achieved:

$$[3.105]\ E_{xx} = \int \left|x(t)\right|^2 dt\ = \int \left|X(f)\right|^2 df\,;$$

they show that in the frequency domain the energy has a dual form with respect to its definition in the time domain (Parseval theorem) and justify that the square of the amplitude spectrum is named ***energy spectral density***, or ***energy spectrum***:

$$[3.106]\ E_{xx}(f) = X(f)\,X^*(f) = |X(f)|^2,$$

function, always real, that shows with its evolution how the energy is distributed over the frequency axis.
Transforming both members of [3.104] the following relation in the frequency domain, named Wiener-Khintchine, is achieved:

$$[3.107]\ E_{xx}(f) = F\{C_{xx}(\tau)\} = \int C_{xx}(\tau)e^{-j\omega\tau}\,d\tau\,,$$

which directly shows that the energy spectrum is the Fourier transform of the autocorrelation function.
In the particular but usual case that the signal is real the autocorrelation function is real and even; As a consequence also the energy spectrum has the same characteristics, i.e $E_{xx}(f)=E_{xx}(-f)$ occurs.
In the graphical representations the function $E_{xx}(f)$ is preferred with respect to the amplitude spectrum $|X(f)|$, and often it is not traced directly but through its expession in decibel:

$$E[dB] = 10\log\frac{E_{xx}(f)}{E_{xxM}},$$

taking as reference it maximum value, indicated with E_{xxM}. In this way, as a matter of fact, it is possible to highlight very small values with respect to the maximum, as shown in the example in Figura 3.9 which is referred to the energy spectrum of the rectangular pulse, $E_{xx}(f)=T^2\,\mathrm{sinc}^2(fT)$.

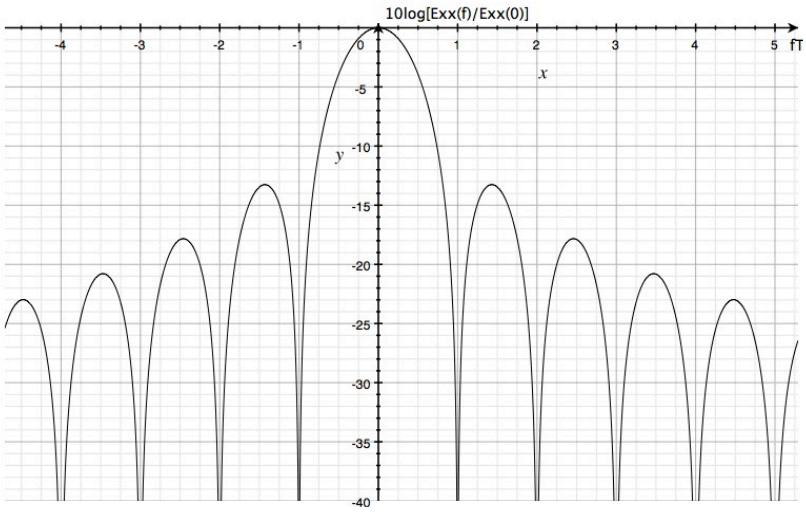


Figura 3.9: Representation in decibel of the energy spectrum of the rectangular impulse.

3.3.6.2 Energy spectral density of the sum of two signals

Let's consider the signal z(t)=x(t)+y(t), sum of two energy signals. Thanks to the linearity property of the Fourier transform the spectrum of z(t) is the sum of the spectra of the addends, i.e. Z(f)=X(f)+Y(f), in general this is not true for the energy spectrum (or amplitude) because:

$$[3.108]\ E_{zz}(f) = [X(f) + Y(f)]\,[X^*(f) + Y^*(f)] = E_{xx}(f) + E_{yy}(f) + 2\Re\{E_{xy}(f)\},$$

where for the ***energy mutual spectral densities***, or ***energy mutual spectra***, between the two signals the following relation is valid:

$$[3.109]\ E_{xy}(f) = X(f)\,Y^*(f),\ E_{yx}(f) = Y(f)\,X^*(f) = E^*_{xy}(f),$$

both complex functions. Transforming both members of the [3.101] it is possible to achieve the following relation in the frequency domain:

$$[3.110]\ E_{xy}(f) = F\{C_{xy}(\tau)\} = \int C_{xy}(\tau)e^{-j\omega\tau}\,d\tau.$$

In general, neither the energies of the two signals can be summed up, considering that performing the integration over the entire frequency axis the result is:

$$[3.111]\ E_{zz} = \int E_{zz}(f)df = E_{xx} + E_{yy} + 2\int \Re\left[E_{xy}(f)\right]df.$$

If the two signals have no affinity, i.e. they are uncorrelated, as in the case they are separated in frequency, the mutual spectral densities are identically equal to zero: as a consequence the signals can be summed both in energy spectrum and in energy. The uncorrelation is condition necessary and sufficient in order that the two signals can be summed up in energy spectrum, while it is sufficient condition, but not necessary, in order that they can be summed up in energy. In fact, such a situation occurs also if the following relation is valid:

$$[3.112]\ \int \Re\left[E_{xy}(f)\right]df = 0,$$

with energy mutual spectral density not identically equal to zero. Taking into account the [3.110], the integral in the first member of [3.112] corresponds to the value in the origin (for τ=0) of the cross correlation function, i.e. to the scalar product; thus, it is possible to state that two orthogonal signals, for which $C_{xy}(0)$=0, can be always summed up in energy.

3.3.7 Power spectra

3.3.7.1 Power spectral density of a signal

Referring to a power signal, considering the correspondent truncated energy signal $x_T(t)$ and its Fourier transform, the ***power spectral density***, or in alternative ***power spectrum***, of the signal is defined:

$$[3.113]\ W_{xx}(f) = \lim_{T\to\infty}\frac{1}{T}X_T(f)X_T^*(f) = \lim_{T\to\infty}\frac{1}{T}|X_T(f)|^2,$$

which is a function always real which indicates how the power is distributed over the frequency axis.

Considering the time autocorrelation function $R_{xx}(\tau)$, assuming that it is possible to exchange the order of the integral and the limit, the Wiener-Khintchine relation is obtained:

$$[3.114]\ W_{xx}(f) = F\{R_{xx}(\tau)\} = \int R_{xx}(\tau)e^{-j\omega\tau}\,d\tau,$$

which directly shows that the power spectrum is the Fourier transform of the autocorrelation function. Swhitching to the time domain, for $\tau=0$ in addition the following relation is achieved:

$$[3.115]\ W_{xx} = R_{xx}(0) = \int W_{xx}(f)df.$$

In the particular, but usual, case that the signal is real the time autocorrelation function is real and even; as a consequence, also the power spectrum of the signal has the same characteristics, i.e. $W_{xx}(f)=W_{xx}(-f)$ is obtained.

3.3.7.2 Power spectral density of the sum of two signals

With similar considerations carried out in the case of energy signals, for the signal $z(t)=x(t)+y(t)$, sum of two power signals, the following expression is obtained:

$$[3.116]\ W_{zz}(f) = W_{xx}(f) + W_{yy}(f) + 2\Re\{W_{xy}(f)\},$$

where the ***power mutual spectral densities***, or ***power mutual spectra***, between the two signals are expressed as:

$$[3.117]\ W_{xy}(f) = \lim_{T\to\infty}\frac{1}{T}X_T(f)Y_T^*(f),\ W_{yx}(f) = \lim_{T\to\infty}\frac{1}{T}Y_T(f)X_T^*(f) = W_{xy}^*(f),$$

complex functions for which the following relations in the frequency domain are valid:

$$[3.118]\ W_{xy}(f) = F\{R_{xy}(\tau)\},\ W_{yx}(f) = F\{R_{yx}(\tau)\}.$$

In general, neither the powers of the two signals can be summed up, considering that performing the integration over the entire frequency axis the result is:

$$[3.119]\ W_{zz} = \int W_{zz}(f)df = W_{xx} + W_{yy} + 2\int\Re\left[W_{xy}(f)\right]df.$$

If the signals are incoherent, as in the case the two signals are separated in the frequency domain which, applied to the [3.117], implies $W_{xy}(f) \equiv W_{yx}(f)\equiv 0$, the signals can be summed up both in power spectrum and in power. The incoherence is necessary and sufficient condition in order that the two signals can be summed up in power spectrum, while it is sufficient, but not necessary, condition in order that they can be summed up in power: in fact, such a situation occurs also if it is verified that:

$$[3.120]\ \int\Re\left[W_{xy}(f)\right]df = 0,$$

with power mutual spectral density not identically equal to zero. Taking into account the [3.118], the integral in the first member of [3.120] corresponds to the value in the origin (for $\tau=0$) of the autocorrelation function; it is possible to state that two signals for which $R_{xy}(0)=0$ occurs can be always summed up in power.

3.3.8 Spectral extension of real signals

3.3.8.1 Remarks on the spectral extension of energy signals

Considering an energy signal, the properties of the Fourier transform imply that if the duration is limited, which means that it is strictly time limited, its energy spectrum E(f) is unlimited in the frequency domain (see as an example Figura 3.6). Due to the duality property, if the spectrum is different from zero only within a finite frequency interval, i.e. the signal is strictly limited in the frequency domain, the signal is time unlimited (see as an example Figura 3.7). In addition, the case of unlimited representation in both domains, time and frequency, may occur (see as an example the unit signal with exponential decay).
Limiting hereinafter the attention just to real signals, it is possible to note that a generic real energy signal is completely characterized even over only one half of the frequency axis, considering that the energy spectrum, other than real, is an even function.
Then, referring just to positive frequencies, in case E(f) is strictly limited in frequency, i.e. the signal is ***strictly limited in frequency***, it is possible to easily determine the maximum value, f_M, beyond which the energy spectrum is zero; moreover, a minimum value, f_m, may exist under which the energy spectrum is zero, as shown in the example in Figura 3.10a; otherwise, $f_m=0$ is supposed, as in the case in Figura 3.10b.

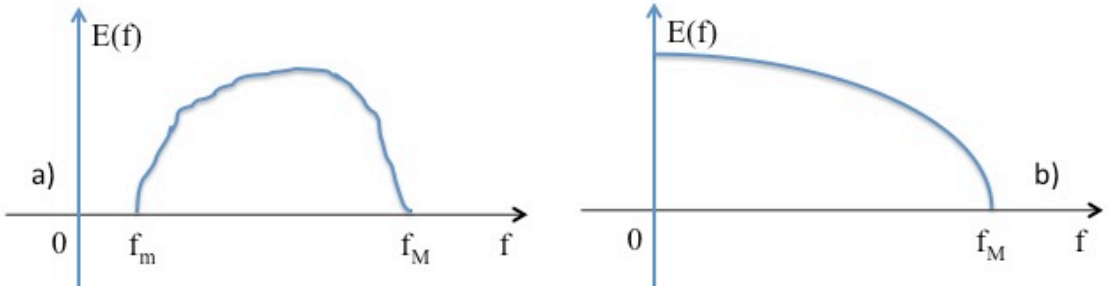


Figura 3.10: Energy spectra of real signals strictly limited in frequency.

Still considering just the right frequency half axis, in the case of energy spectrum theoretically unlimited it is anyway possible (see as an example Figura 3.11a and Figura 3.11b) to disregard the spectral components beyond an opportune positive value of the frequency, which is assumed to be the maximum, f_M, and seldom (see as an example Figura 3.11b) also those below another positive value, assumed as the minimum, f_m; in such cases the denomination ***practically limited in frequency signal*** is adopted.
For a signal just practically limited in frequency the maximum value f_M and the minimum value f_m, if different from zero, are not exactly defined; usually, it is opportune that the mentioned values are chosen so that the integral of 2E(f) calculated between such extremes provides a value practically equal to the value of the energy E of the signal, i.e. the result is:

$$[3.121]\quad 2\int_{f_m}^{f_M} E(f)\,df = E_{m,M} \cong E,$$

with relative difference much less than one (for example less than 0.01).

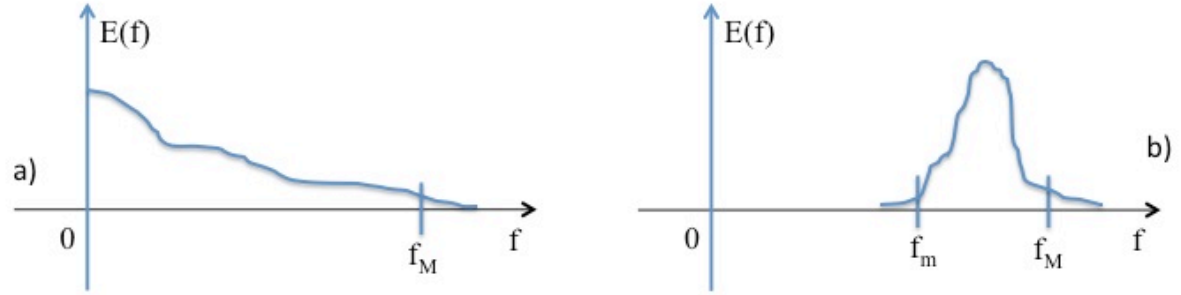


Figura 3.11: Energy spectra of real signals practically limited in frequency.

Rigorously speaking, the hypotesis that an energy signal is simultaneously strictly limited both in time and in frequency must be excluded. Nevertheless, in the case of physical signals it is

mandatory to suppose a limited duration, with corresponding theoretical extension of the spectra, while, since it is impossible to realize physical circuits with a response significantly different from zero for frequencies high enough, it is always possible to identify a limited practical maximum frequency. With satisfactorily approximation it is always possible to consider a ***signal practically limited in time and in frequency***, which has both the time evolution and the spectrum included into finite intervals.

3.3.8.2 Bandwidth of real signals

As previously discussed it is possible, rigorously or practically, to identify an interval (f_m, f_M) out of which the the energy spectral density of a generic energy signal is at least negligible. In the case of real power signals, theoretically always time unlimited, it is still possible to identify an interval out of which the power spectral density W(f) is zero or can be disregarded; in such an event the bounds of the interval are selected so that the result is:

$$[3.122]\quad 2\int_{f_m}^{f_M} W(f)\,df = W_{m,M} \cong W,$$

still with relative difference with respect to the power W much lower than one (for example less than 0.01).
For any real signal, it is then possible to define the monolateral bandwidth, or simply ***band***, through the:

$$[3.123]\quad B = f_M - f_m,$$

naming it ***band limited signal***. Introducing the central frequency of the band:

$$[3.124]\quad f_a = \frac{1}{2}(f_M + f_m),$$

often also the relative band of the real signal can be considered:

$$[3.125]\quad \frac{B}{f_a} = 2\,\frac{f_M - f_m}{f_M + f_m}.$$

Still considering a generic real signal, hereinafter the following definitions will be adopted. A ***baseband signal*** has the bounds of the monolateral band, either theoretical or practical, which respect the condition:

$$[3.126]\quad 0 \le f_m << f_M\ ;$$

in this case the maximum frequency can be identified as the signal bandwidth, with $B \cong f_M$, and the central band frequency can be identified with one half of its bandwidth. Usually in such cases (see examples in Figura 3.10 and Figura 3.11a) it is not interesting to consider the relative bandwidth, which anyhow is about 2. A ***signal with limited relative bandwidth*** is characterized by a relative badwidth, either theoretical or practical, which respect the condition:

$$[3.127]\quad \frac{B}{f_a} < \frac{2}{3},\ \text{i.e. } f_M < 2f_m;$$

it is immediate to find out that the bandwidth is smaller than the minimum frequency, i.e. $B < f_m$ (see for example Figura 3.11b). Within such latter cathegory, signals with ***narrow relative bandwidth***, for which $B << f_a$, and signals with ***very narrow relative bandwidth***, with B smaller than f_a for more than one order of magnitude, can be distinguished.
Usually, signals directly provided by information sources or signals obtained as a result of processing over multiple sources with the scope to get advanced waveforms, capable to drive all the information contained in the whole set of the processed signals over the same physical means, belong the the former cathegory, with $f_M >> f_m$.

The signals that after particular processing implemented on their baseband waveform assume a new representation, very often with $B<<f_a$, more suitable for the characteristics of the long distance transmission means, belong to the latter cathegory, with $f_M<2f_m$.

3.3.9 Discrete spectra of periodic signals

3.3.9.1 Spectrum of power signal represented in Fourier series

In section 3.3.4.2 it was demonstrated that a generic harmonic, altough it is a power signal, can be transformed with the Fourier transform. Thanks to the linearity property, this result is applicable to any generic power signal which can be represented through the development in Fourier series.
Let's consider a periodic signal x(t) with period T_0, represented through the development in Fourier series (see [3.2]) where as usual $\Omega_0=2\pi f_0=2\pi/T_0$; recalling that $F\{e^{j\Omega_o t}\}=\delta(f-f_0)$, the Fourier transform is:

$$[3.128]\ X(f)=\int \Sigma_k C_k e^{jk\Omega_o t} e^{-j\omega t}\,dt=\Sigma_k C_k\, F\left\{e^{jk\Omega_o t}\right\}=\Sigma_k C_k \delta(f-kf_0),$$

which shows that the representation in the frequency domain of a generic periodic signal is of discrete type, for the values of the harmonic components kf_0, with spectrum composed of stripes all equally spaced between each other of an interval $f_0=1/T_0$ and with values, in general complex, provided by the Fourier coefficients.
As shown in section 3.3.9.2, the power spectrum can be obtained as the Fourier transform of the autocorrelation function. In this way the discrete power spectrum is obtained:

$$[3.129]\ W_{xx}(f)=\sum_k |C_k|^2\,\delta(f-kf_0),$$

with an example shown in Figura 3.12.

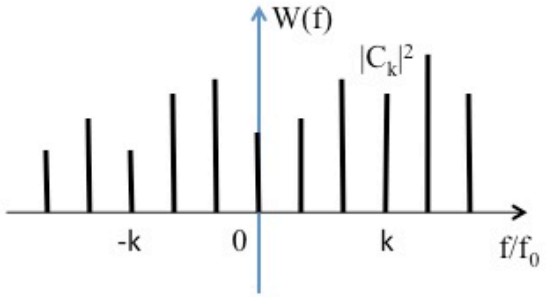


Figura 3.12: Conventional depiction of the discrete power spectrum of a periodic signal.

Integrating the power spectral density $W_{xx}(f)$ expressed by [3.114], it is immediate to achieve that the power of the periodic signal can be calculated summing up all the squares of the modula of the Fourier coefficients (see [3.7]).

3.3.9.2 Evaluation of the discrete power spectrum

Unitilizing the expression [2.130] of the autocorrelation function, in which the representation of a periodic signal in Fourier series is introduced, the result is:

$$[3.130]\ R_{xx}(\tau)=\frac{1}{T_0}\sum_k\sum_h C_k C_h^*\, e^{jk\Omega_o\tau}\int_{-T_0/2}^{T_0/2} e^{j(k-h)\Omega_o t}\,dt\,,$$

and again, since the integrals are all zero for h=k:

$$[3.131]\ R_{xx}(\tau)=\sum_k |C_k|^2\, e^{jk\Omega_0\tau}\,.$$

Through the Wiener-Khintchine relation $W_{xx}(f)=F\{R_{xx}(\tau)\}$, applying again the $F\{e^{j\Omega_0 t}\}=\delta(f-f_0)$, then the discrete power spectrum of the periodic signal is demonstrated to be like in [3.114].

3.3.9.3 Particular expressions of the Fourier coefficients

Often the periodic signal is expressed as a repetition like $x(t)=rep_{T0}\{g(t)\}$, with no limitations on the duration of the generating energy function g(t). Then, the Fourier coefficients are achieved (see [3.5]):

$$[3.132]\ C_k=\frac{1}{T_0}\int_{-T_0/2}^{T_0/2}\sum_n g(t-nT_0)e^{-jk\Omega_0 t}dt\ ;$$

exchanging the order of the integral and summation, changing the variable $v=t-nT_0$ and taking into account that $e^{-jk\Omega_0 nT_0}=1$, then the result is:

$$[3.133]\ C_k=\frac{1}{T_0}\sum_n\int_{-nT_0-T_0/2}^{-nT_0+T_0/2} g(v)e^{-jk\Omega_0 v}\,dv=\frac{1}{T_0}\int g(v)e^{-jk\Omega_0 v}dv=\frac{1}{T_0}G(kf_0),$$

which allows to calculate the Fourier coefficients on the basis of the Fourier transform G(f) of the generating function g(t).
Definitively, utilizing the previous expressions of the Fourier coefficients, the representation in time series, named Poisson sum, is achieved:

$$[3.134]\ x(t)=rep_{T0}[g(t)]=\frac{1}{T_0}\sum_k G(kf_0)e^{jk\Omega_0 t},$$

and as a consequence the discrete spectral representation of the periodic signal:

$$[3.135]\ X(f)=F\{rep_{T0}[g(t)]\}=\frac{1}{T_0}\sum_k G(kf_0)\delta(f-kf_0).$$

Moreover, considering that no hypotesis have been expressed on possible limitation the duration of the generating function, it can be particularized in the truncated signal in one period $x_{T0}(t)=x(t)rect(t/T_0)$; indicating with $X_{T0}(f)$ the corresponding Fourier transform, then the coefficients are:

$$[3.136]\ C_k=\frac{1}{T_0}X_{T0}(kf_0).$$

3.3.9.4 Spectrum of the signal repetition of the ideal pulse

Let's consider the particular periodic signal defined as repetition with interval T_0 of the ideal pulse in time:

$$[3.137]\ rep_{T0}[\delta(t)]=\sum_h\delta(t-hT_0).$$

On the basis of the time representation with the Poisson sum (see [3.134]) and recalling that the the Fourier transform of the particular generating function, i.e. the ideal pulse, is the unit constant, the following expression is achieved:

$$[3.138]\ rep_{T0}[\delta(t)]=\frac{1}{T_0}\sum_k e^{jk\Omega_0 t},$$

from which it is possible to find out that all the Fourier coefficients of the particular considered periodic signal are equal to the inverse of the period. Comparing the two previous expressions the following relation is achieved:

$$[3.139]\ T_0\sum_h\delta(t-hT_0)=\sum_k e^{jk\Omega_0 t}=1+2\sum_{k=1}^{\infty}\cos k\Omega_0 t\,.$$

Considering $C_k=1/T_0$ in the expression [3.128] of the spectrum of the generic periodic signal, the result is:

$$[3.140]\ F\{\mathrm{rep}_{T0}[\delta(t)]\} = \frac{1}{T_0}\Sigma_k\delta(f-kf_0) = \frac{1}{T_0}\,\mathrm{rep}_{f0}[\delta(f)],$$

which demonstrates that the discrete spectrum of the repetition with interval T_0 of the ideal pulse is the repetition at interval f_0 of the ideal pulse in frequency, but multiplied by the factor $1/T_0$. Sometimes, the considered spectrum is named spectral strips ***comb*** and the corresponding repetition in the time domain time strips ***comb***.
Since $F\{\delta(t)\}=1$, thanks to the linearity and time shift properties of the Fourier transform, the result is:

$$[3.141]\ F\{\mathrm{rep}_{T0}[\delta(t)]\} = \sum_h e^{-jh2\pi fT_0};$$

comparing such an expression with the previous one the following relation is achieved:

$$[3.142]\ \frac{1}{T_0}\sum_k\delta(f-kf_0) = \sum_h e^{-jh2\pi fT_0} = 1+2\sum_{h=1}^{\infty}\cosh\omega T_0,$$

which in the frequency domain has a similar form to that of the relation [3.124] in the time domain.

3.4 SAMPLING THEOREM

3.4.1 Sampling in the ferquency domain

As mentioned above (see [3.29]), an energy signal g(t) with finite duration in the interval (-D/2, D/2) can be represented in the time domain through the development in Fourier series with limited duration within an interval (-T/2, T/2), but respecting the condition T>D, with the Fourier coefficients C_k given by [3.30]. Indicating with G(f) the Fourier transform of the considered signal g(t), as demonstrated in section 3.4.2 it is possible to achieve:

$$[3.143]\ G(f) = \sum_k G(k/T)\mathrm{sinc}[(fT-k)],$$

where the ***samples*** G(k/T) at the equally spaced ***sampling frequencies*** $f_k=k/T$ have a simple relation with the coefficients of the Fourier transform through the:

$$[3.144]\ G(k/T) = TC_k.$$

The representation [3.143], known as the sampling theorem in the frequency domain, perfectly rebuilds the spectrum by interpolation of its samples at the frequencies multiple of 1/T, but only if the condition concerning the duration D of the signal is satisfied:

$$[3.145]\ T>D.$$

The representations obtained with samples of the spectrum must be considered also for signals just practically time limited, for which (for example in case of numerical calculus) the existence of a practical finite duration is assumed.

3.4.2 Demonstration of the sampling theorem in the frequency domain

Taking into account the Fourier transform of the considered signal g(t):

$$[3.146]\ G(f) = \int g(t)e^{-j\omega t}\,dt = \int_{-T/2}^{T/2} g(t)e^{-j\omega t}\,dt,$$

and the expression of the Fourier coefficients C_k, repeated herein:

$$[3.147]\ C_k = \frac{1}{T}\int_{-T/2}^{T/2} g(t)e^{-jk\Omega t}\,dt,$$

it is possibile to observe that putting $\omega_k=k\Omega=2\pi k/T$ in the last member of the [3.146] the elementary relation [3.144] between the coefficients C_k and the samples G(k/T) of the spectrum at the discrete frequencies multiple of 1/T is obtained. Thus, the representation in the time domain in Fourier series becomes:

$$[3.148]\ g(t) = \frac{1}{T}\sum_k G(k/T)e^{jk\Omega t}\,\mathrm{rect}\left(\frac{t}{T}\right).$$

Applying the Fourier transform to both members of the previous expresion and recalling the property of the convolution in the frequency domain, the result is:

$$[3.149]\ G(f) = \frac{1}{T}\sum_k G(k/T)F\{e^{jk\Omega t}\} * F\{\,\mathrm{rect}(t/T)\,\} = \frac{1}{T}\sum_k G(k/T)\delta(f-k/T) * T\,\mathrm{sinc}(fT) =$$

$$= \sum_k G\left(\frac{k}{T}\right)\int \delta\left(v-k/T\right)\mathrm{sinc}\left[\left(f-v\right)T\right]dv\,;$$

because of the fundamental property of the ideal pulse, finally the representation by sample interpolation [3.143] is achieved.

3.4.3 Sampling in the time domain

The representation by sample interpolation in the time domain, dual with respect to [3.143] in the frequency domain, was already introduced in section 3.1.4 as particular case of time seroes with orthogonal functions. Nevertheless, the representation can be rigorously satisfactory only for an opportune class of real energy signals with unlimited duration, as was introduced but not demonstrated. Now, such statement, which represents the sampling theorem in the time domain, will be demonstrated.

Applying the Fourier transform to both members of the representation by sample interpolation (see [3.36]), the result is:

$$[3.150]\ X(f) = \sum_k c_k\, F\left\{\mathrm{sinc}\left(\frac{t}{T_c}-k\right)\right\} = T_c\sum_k x(kT_c)\,e^{-jk2\pi fT_c}\,\mathrm{rect}(fT_c),$$

in which, considering the variable t instead of the frequency f and the parameter T instead of $f_c=1/T_c$ a representation in Fourier series like [3.29] can be recognized. Thus, on the basis of what was previously demonstrated, it is valid only if f_c is greater or equal to the extension of the interval out of which X(f) must be zero; supposing that the spectrum of the considered signal is strictly limited with maximum frequency f_M, i.e. within the interval extended $2f_M$, then the following condition in order that the [3.33] is satisfactory, and as a consequence the representation by sample interpolation in time, is achieved:

$$[3.151]\ f_c > 2f_M,$$

i.e.

$$[3.152]\ T_c < T_N = \frac{1}{2f_M},$$

where T_N is the already mentioned ***Nyquist interval***.

Summarizing, the representation by sample interpolation in the time domain is rigorously satisfactory for the class of real energy signals with limited duration, but with spectrum strictly limited in frequency, only if the sampling interval respects the condition [3.152].

The representation by samples is valid also for real power signals, but only if they are strictly limited in the frequency domain and the condition [3.152] is satisfied over the entire sampling interval.

Finally, in the case of just practically limited bandwidth signals more negligible are the spectral components out of the practical band better is the approximation of the representation.

3.5 COMPLEX REPRESENTATIONS OF TIME CONTINOUS SIGNALS

3.5.1 Representative complex signals

3.5.1.1 Analytical signal

Let's consider a real energy signal x(t), which respects the condition to have the Fourier transform equal to zero in the origin, i.e. the X(0)=0 is valid. Considering that X(f) is a hermitian function, the representation in the frequency domain is redundant, because it is sufficient to know even just one of the two portions of the spectrum present respectively only on either the frequency negative half axis or the positive one, respectively indicated with $X_-(f)$ and $X_+(f)$ (see Figura 3.13a).

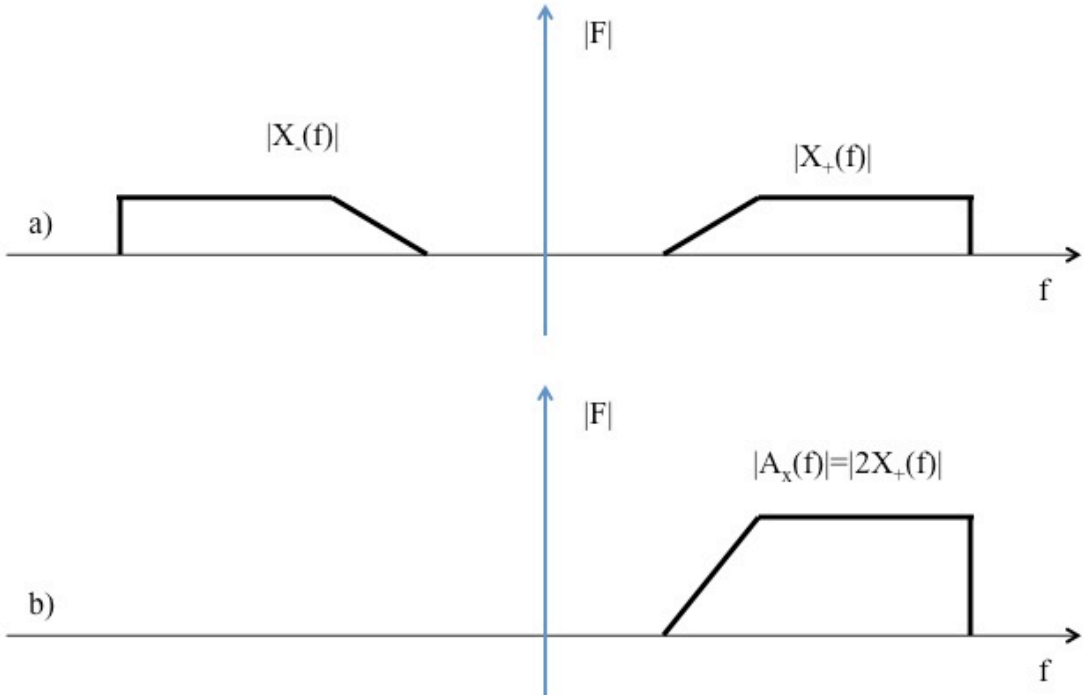


Figura 3.13: Example of amplitude spectrum of a real signal (a) and corresponding spectrum different from zero, and doubled, just on the frequency positive half axis (b).

The mentioned redundancy in the frequency domain can be removed considering the spectrum, that is completely representative too, defined by:

$$[3.153]\ A_x(f) = X(f)\,[1+\mathrm{sgn}(f)] = 2X_+(f),$$

different from zero just on the positive axis where it is equal to the double of the spectrum of x(t) (see Figura 3.13b); since $A_x(f)$ is not hermitian, the signal $a_x(t)=F^{-1}\{A_x(f)\}$, achievable as inverse Fourier transform of $A_x(f)$ and named ***analytical signal***, results to be complex.
Starting from [3.153], going on applying the property of the convolution in the time domain, the following expression of the analitic signal is achieved:

$$[3.154]\ a_x(t) = F^{-1}\{X(f)\,[1+\mathrm{sgn}(f)]\} = x(t) + x(t) * F^{-1}\{\mathrm{sgn}(f)\},$$

which can be written as (the reader interested to go more in deep can see section 3.5.1.5):

$$[3.155]\ a_x(t) = x(t) + j\,\hat{x}(t),$$

where $\hat{x}(t)$ is a real signal in biunivocal correspondence with x(t), named its ***Hilbert transform***. The analytical signal is thus certainly representative of the real one x(t), which in fact can be derived from $a_x(t)$ with the simple relation:

$$[3.156]\ x(t) = \Re\{a_x(t)\}.$$

It is worth to note that while the real signal x(t) is redundant in its representation in the frequency domain, by definition the corresponding complex signal $a_x(t)$ has no longer this characteristic, but instead it is redundant in the time domain, being representative even just its imaginery part.
Although introduced on the basis of its Fourier transfrom, the analytical signal has a biunivocal correspondence with x(t) through relations in the time domain that are valid even if the Fourier operators are not applicable. In this way, it is possible to use the complex representation with $a_x(t)$ even in the case of a power signal, but only if it is real and with mean value equal to zero, i.e. $\overline{x}=0$ is true.

3.5.1.2 Complex envelope

Let's consider a real energy signal x(t) with limited monolateral bandwidth, with extremes f_m and f_M, and its representation by means of the spectrum $A_x(f)=2X_+(f)$ of the analitical signal (see [3.153]), which is by definition different from zero just in the interval (f_m, f_M) and must be zero in the origin in the case $f_m=0$. Assuming a known arbitrary frequency f_c as reference, but only if it belongs to the mentioned interval ($f_m \le f_c \le f_M$), then the shifted spectrum $A_x(f+f_c)$, which in this way includes the origin of the frequency axis, can be considered, as shown in the example in Figura 3.14.

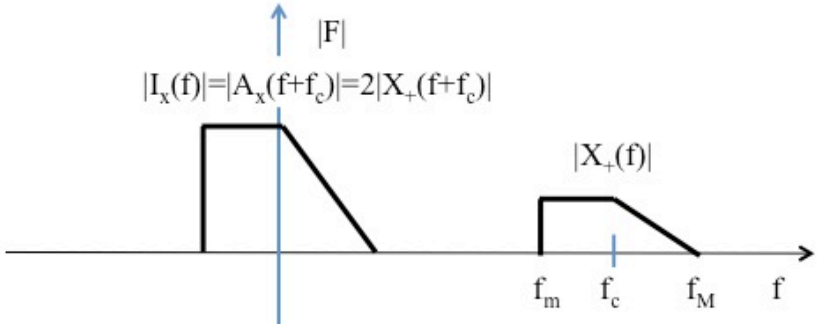


Figura 3.14: Example of amplitude spectrum of the complex envelope.

The complex function in the time domain defined applying the inverse Fourier transform:

$$[3.157]\ \iota_x(t) = a_x(t)\ e^{-j\omega_c t} = F^{-1}\{A_x(f+f_c)\},$$

results to be still fully representative of x(t) and assumes the name of representative complex envelope or simply ***complex envelope*** of the real signal x(t).
Recalling that $x(t)=\Re\{a_x(t)\}$, from the definition of the complex envelope the following relation can be achieved:

$$[3.158]\ x(t)=\Re\left\{\iota_x(t)e^{j\omega_c t}\right\},$$

which allows to rebuild in a simple way the real signal x(t) starting from its correspondent complex envelope $\iota_x(t)$, once the reference arbitrary frequency f_c is known.
The evolution of the spectrum X(f) of the signal can be achieved from that of the complex envelope, $I_x(f)=A_x(f+f_c)$, applying the operations of frequency shift of the amount f_c, axis inversion and conjugation, highlighted in the:

$$[3.159]\ X(f) = (1/2)\ [\ I_x(f-f_c) + I_x^*(-f-f_c)].$$

The above spectrum reconstruction is pictorially shown in Figura 3.15, where it is possible to note how the two addends are obtained, referring for sake of simplicity just to the modulus. As a matter of fact $I_x(f-f_c)$ can be obtained with a simple frequency shift, while $I_x^*(-f-f_c)$ requires in addition an axis inversion with respect to the origin and a conjugation.

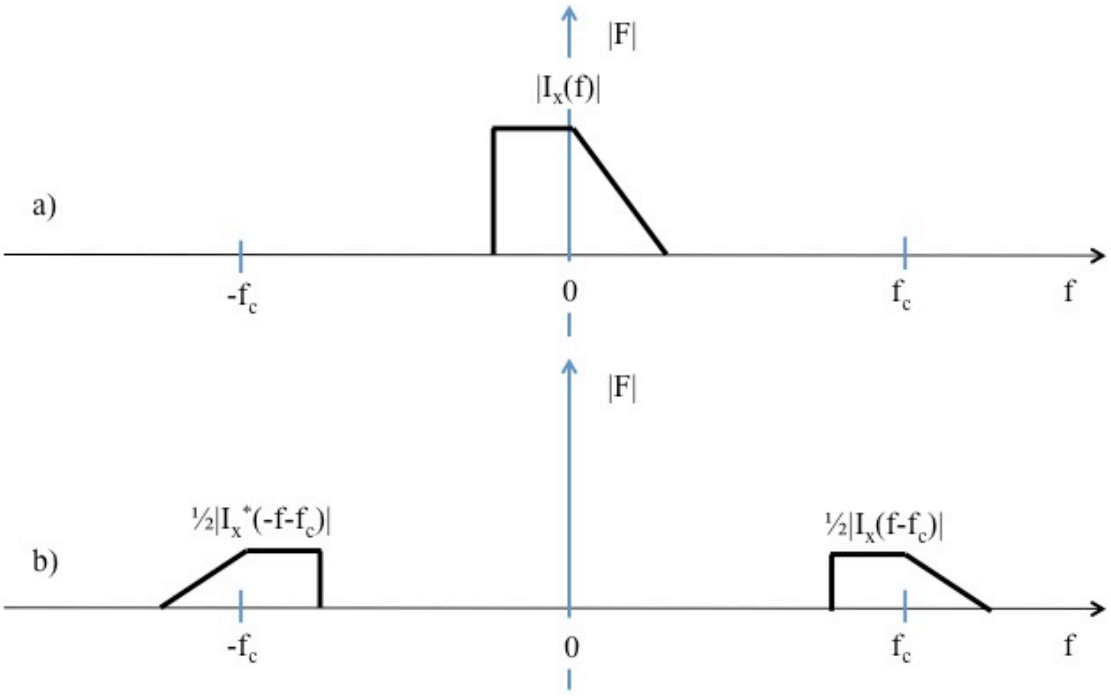


Figura 3.15: Example of amplitude spectrum of the complex envelope (a) and of reconstruction of the amplitude spectrum of the real signal (b).

It is possible to note that while the spectrum X(f) is hermitian, instead the same property is in general not satisfied for $I_x(f)$, unless it is possible to choose the value of the frequency f_c so that the particular condition is satisfied:

$$[3.160]\ X_+(f_c+f) = X_+^*(f_c-f),$$

in which the half positive spectrum $X_+(f)$ results to be symmetric in amplitude and inversely simmetric in phase with respect to the frequency f_c. Then, considering that the spectrum $I_x(f)$ is hermitian, the complex envelope becomes real.

Although introduced on the basis of the Fourier transform of the analitical signal, the complex envelope results to be in biunivocal correspondence with the analitical signal $a_x(t)$, and thus with the signal x(t), through relations in the time domain which are still valid even if the Fourier operators are not applicable. In this way it is possible to use the complex representation by means of $\mathbf{\iota}_x(t)$ even in the case of a power signal, but only if it is real and with mean value equal to zero, i.e. $\bar{x}=0$ is true.

Definitively the relation among time domain signal, analytic signal and complex envelope can be represented as shown in Figura 3.16.

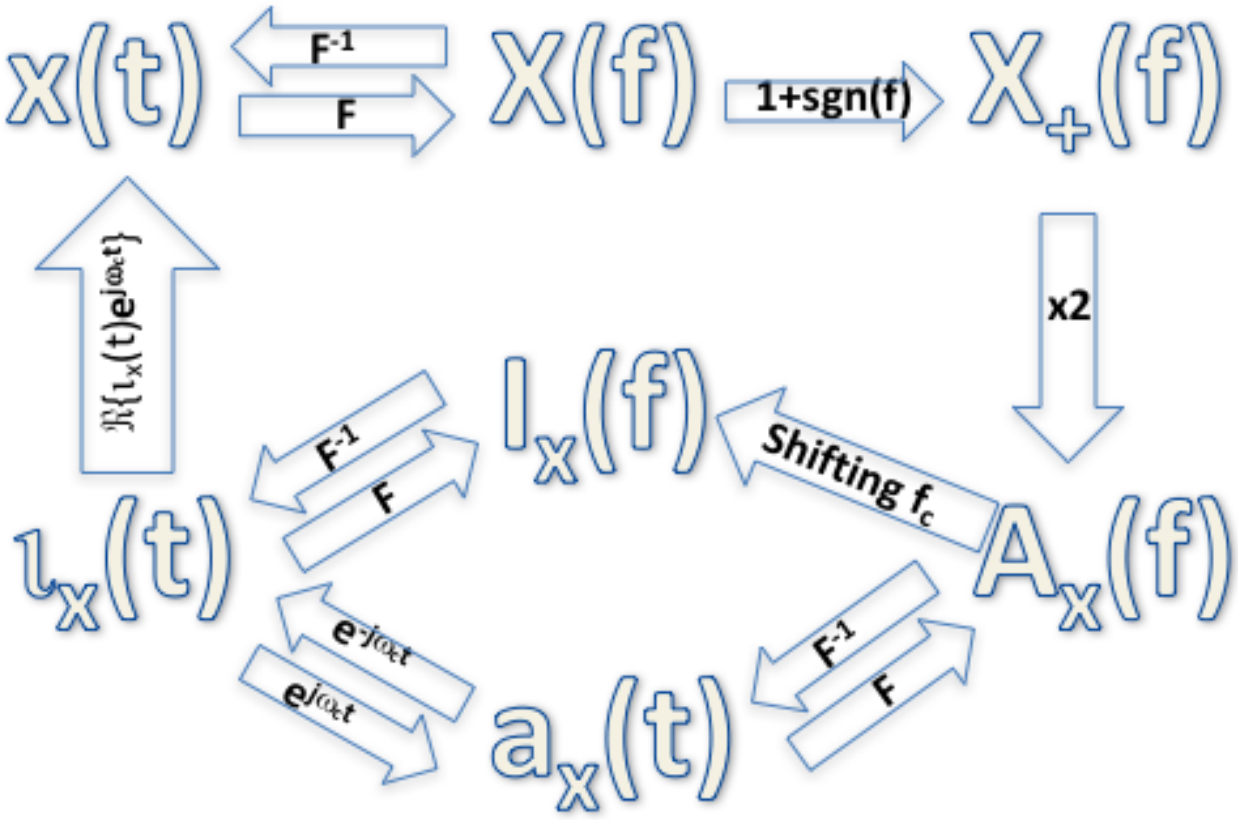


Figura 3.16: Relation among time domain signal, analytic signal and complex envelope

3.5.1.3 Characteristics of the complex envelope

First of all, it is possible to note that the representation by means of the complex envelope allows to extend to signals the complex Steinmetz notation, typical of the circuit theory, for which the relation is:

$$[3.161]\quad a(t)=\Re\left\{\mathbf{A}e^{j\omega_c t}\right\},$$

where the complex phasor $\mathbf{A}=Ae^{j\alpha}$, with modulus A and argument α real and constant, is representative only of the function of type $a(t)=A\cos(\omega_c t+\alpha)$, pure harmonic at the frequency $f_c=\omega_c/2\pi$, with amplitude A and phase α. Expressing a generic complex envelope $\iota(t)$ in modulus, $|\iota(t)| = \iota(t)$, and argument, $\arg\{\iota(t)\}=\alpha(t)$:

$$[3.162]\quad \iota(t) = \iota(t)\, e^{j\alpha(t)},$$

in the complex plane $\Re,\Im$ the considered signal can be in fact represented through a vector (see Figura 3.17): nevertheless it is no longer constant, as the phasor **A** would appear, but variable in time.

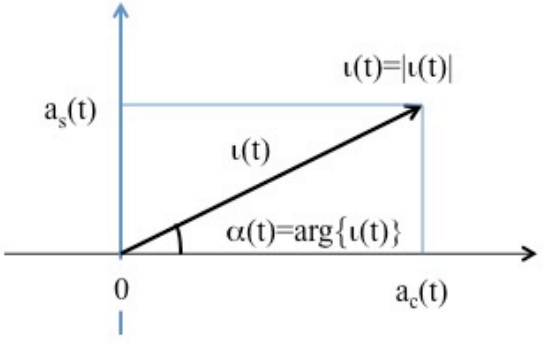


Figura 3.17: Representation of the complex envelope in the complex plane.

Of course, also the coordinates of the vector representing the complex envelope on the $\Re,\Im$ plane are dependent on time, i.e. its real part, $\Re\{\iota(t)\}=a_c(t)$, and the coefficient $\Im\{\iota(t)\} = a_s(t)$ of its imaginery part, and thus the following notation:

$$[3.163]\ \iota(t) = a_c(t) + j a_s(t),$$

The energy of the complex envelope is the same of the analytical signal, because their spectra are different only for the shift of the amount f_c which doesn't impact the expression of the energy in the frequency domain (see [3.90]). Taking this into account and utilizing the [3.153] the result is:

$$[3.164]\ E_\iota = \int \left[1+\mathrm{sgn}(f)\right]^2 X(f) X^*(f)\,df = 4\int_0^\infty E_{xx}(f)\,df,$$

i.e. the energy of the complex envelope is the double of that of the signal of which it is representative:

$$[3.165]\ E_\iota = \int \iota^2(t)\,dt = 2E_{xx};$$

such a result is consistent with the redundancy in the time domain, and no longer in the frequency domain, of the complex envelope, like that of the analitical signal to which it is strictly linked by definition. Moreover, the energy is the sum of the energies of its real $a_c(t)$ and imaginery $a_c(t)$ components; in fact, starting from the [3.163] the result is:

$$[3.166]\ E_\iota = \int a_c^2(t)\,dt + \int a_s^2(t)\,dt = E_{acac} + E_{asas}.$$

In the case of real power signals with zero mean, expressions similar to the two previous ones can be obtained just replacing the energies with the powers.

3.5.1.4 Right and left analytical signals

Once indicated with $x_-(t)=F^{-1}\{X_-(f)\}$ and $x_+(t)=F^{-1}\{X_+(f)\}$ respectively the inverse transforms of the two portions of the spectrum of x(t) respectively extended only on either the frequency negative or positive half axis, being $X(f)=X_-(f)+X_+(f)$ thanks to the linearity of the Fourier operator the result is:

$$[3.167]\ x(t) = x_-(t) + x_+(t),$$

i.e. also in the time domain the signal can be factorized in two complex signals, named ***left analitical signal*** and ***right analytical signal***.

Considering the relation $X_-(f)= X_+^*(-f)$, which is valid because the transform X(f) is hermitian, applying to it the property of conjugation in the time domain the result is:

$$[3.168]\ x_-(t) = x_+^*(t),$$

i.e. one signal is the conjugate of the other. Recalling the relation $A_x(f)=2X_+(f)$, in the time domain $a_x(t)=2x_+(t)$ is achieved and then on the basis of [3.155] the following relations can be achieved:

$$[3.169]\ x_+(t) = \frac{1}{2}\left[x(t) + j\,\hat{x}(t)\right],\ x_-(t) = \frac{1}{2}\left[x(t) - j\,\hat{x}(t)\right];$$

the left and the right analytical signals are thus both representative of the real signal x(t). Moreover, the inverse relations are:

$$[3.170]\ x(t) = 2\Re\{x_-(t)\} = 2\Re\{x_+(t)\},$$

which allow to determine very simply the real signal x(t) once any of the two considered complex signal is known.

Considering that by definition the spectra of $x_-(t)$ and $x_+(t)$ are separated on the frequency axis, the energy mutual spectral densities are identically equal to zero; thus, the two signals are reciprocally uncorrelated, and even more, orthogonal.

3.5.1.5 Hilbert transform

In section 3.5.1.1 the real signal $\hat{x}(t)$ has been introduced and the relation (see [3.154] and [3.155]):

$$[3.171]\ x(t) * F^{-1}\{\mathrm{sgn}(f)\} = j\,\hat{x}(t)$$

has been established. Applying the duality property to the expression [3.75] the result is:

$$[3.172]\ F^{-1}\{\mathrm{sgn}(f)\} = \frac{j}{\pi t}\ ;$$

achieving as a consequence:

$$[3.173]\ x(t) * F^{-1}\{\mathrm{sgn}(f)\} = jx(t) * \frac{1}{\pi t}.$$

Comparing with the relation [3.171] the definition of the ***Hilbert transform*** of x(t) is derived:

$$[3.174]\ \hat{x}(t) = x(t) * \frac{1}{\pi t} = \frac{1}{\pi}\int \frac{x(\tau)}{t-\tau} d\tau,$$

which is a real signal fully representative of x(t), either energy or power signal if x(t) is respectively of the former or the latter type.
Directly from [3.172] the following relation can be achieved

$$[3.175]\ F\left\{\frac{j}{\pi t}\right\} = \mathrm{sgn}(f);$$

assumed that x(t) is an energy signal, applying to [3.174] the Fourier operator and utilizing the property of the convolution in the time domain, the spectrum of the Hilbert transform $\hat{x}(t)$ can be achieved:

$$[3.176]\ \hat{X}(f) = -j\,\mathrm{sgn}(f)\,X(f),$$

as well as the inverse relation,

$$[3.177]\ X(f) = j\,\mathrm{sgn}(f)\,\hat{X}(f).$$

From this relation, passing into the time domain, the result is:

$$[3.178]\ x(t) = \hat{x}(t) * \frac{-1}{\pi t}\ ;$$

the ***inverse Hilbert transform*** can be then expressed by mrans of the following relation:

$$[3.179]\ x(t) = \hat{x}(t) * \frac{-1}{\pi t} = -\frac{1}{\pi}\int \frac{\hat{x}(\tau)}{t-\tau} d\tau,$$

which is valid independently on the signal type, either energy or power.
From [3.176] the equality $|\hat{X}(f)| = |X(f)|$ is immediately achieved, realizing that the signal and its Hilbert transform have the same energy spectrum and, thus, the same autocorrelation function:

$$[3.180]\ E_{\hat{x}\hat{x}}(f) = E_{xx}(f),\ C_{xx}(\tau) = C_{\hat{x}\hat{x}}(\tau).$$

Taking into account also the [3.177], the following relations concerning the energy mutual spectra and the cross correlation functions can be achieved:

$$[3.181]\ E_{x\hat{x}}(f) = -E_{\hat{x}x}(f) = j\,\mathrm{sgn}(f)\,E_{xx}(f),\ C_{x\hat{x}}(\tau) = -C_{\hat{x}x}(\tau).$$

Placing $\tau=0$ in the last relation and taking into account that the cross correlation functions are real, the signals x(t) and $\hat{x}(t)$ turn out to be orthogonal; considering that both the energy mutual spectra are purely imaginery, i.e. $\Re\{E_{x\hat{x}}(f)\}=0$ is obtained, the two considered signals can be then also added up in energy spectrum (see [3.93]), other than in energy.

Let's recall that the Hilbert transform, as well as the analytical signal, is applicable also to real power signals, but only if their mean value is equal to zero. In such a case the relations similar to the [3.180] and [3.181] are valid, obtained replacing the functions like E(f) and C(τ) with the correspondent of type W(f) and R(τ); in particular, the power mutual spectral densities, $W_{\hat{x}x}(f) = W^{*}_{x\hat{x}}(f)$, are purely imaginery, i.e. $\Re\{W_{x\hat{x}}(f)\}=0$ is achieved; as a aconsequence, the signal x(t) and its Hilbert transform $\hat{x}(t)$ can be added up in power spectrum (see [3.101]), other than in power.

3.5.1.6 Details on the complex envelope

On the basis of the definition of the complex envelope and recalling the [3.155], the following relation is achieved:

$$[3.182]\ \iota(t) = [x(t) + j\,\hat{x}(t)]\,e^{-j\omega_c t} = [x(t) + j\,\hat{x}(t)][\cos(\omega_c t) - j\sin(\omega_c t)],$$

and splitting the real and the imaginery parts:

$$[3.183]\ a_c(t) = x(t)\cos(\omega_c t) + \hat{x}(t)\sin(\omega_c t),$$

$$[3.184]\ a_s(t) = \hat{x}(t)\cos(\omega_c t) - x(t)\sin(\omega_c t).$$

In the case of energy signal, applying the Fourier operator to [3.183] and utilizing the [3.176] the spectrum of $a_c(t)$ is achieved:

$$[3.185]\ A_c(f) = X(f) * F\{\cos(\omega_c t)\} - j\,[\mathrm{sgn}(f)X(f)] * F\{\sin(\omega_c t)\};$$

being $F\{\cos(\omega_c t)\} = [\delta(f - f_c) + \delta(f + f_c)]/2$ and $F\{\sin(\omega_c t)\} = -j\,[\delta(f - f_c) - \delta(f + f_c)]/2$ then it is possible to obtain:

$$[3.186]\ A_c(f) = \frac{1}{2}\left\{X(f)\left[1-\mathrm{sgn}(f)\right]\right\} * \delta(f - f_c) + \frac{1}{2}\left\{X(f)\left[1+\mathrm{sgn}(f)\right]\right\} * \delta(f + f_c) =$$

$$= X_{-}(f) * \delta(f - f_c) + X_{+}(f) * \delta(f + f_c),$$

and due to the fundamental property of the ideal pulse the result is:

$$[3.187]\ A_c(f) = X_{-}(f-f_c) + X_{+}(f+f_c).$$

Going forward similarly, starting from [3.184] the result is:

$$[3.188]\ A_s(f) = j\,[X_{-}(f - f_c) - X_{+}(f + f_c)].$$

It is immediately possible to verify that, as obvious, the following relation is valid:

$$[3.189]\ I(f) = A_c(f) + jA_s(f) = 2X_{+}(f + f_c) = A(f + f_c).$$

If the particular condition [3.160] is updated taking into account that X(f) is hermitian, it becomes:

$$[3.190]\ X_{+}(f_c + f) = X_{-}(f - f_c),$$

from which it is possible to find out that from [3.188] $A_s(f)\equiv 0$ turns out and then from [3.187]

$$[3.191]\ A_c(f) = 2X_+(f+f_c) = I(f),$$

turns out, demonstrating the mentioned condition on the basis of which I(f) is hermitian and the complex envelope $\iota(t)$ is real.

3.5.2 Real representations through the complex envelope

3.5.2.1 Signals in shifted band

Although the complex envelope is a representation with general validity, it is typically utilized in the case of band limited signals, both energy and power signals, achieved from real baseband signals through proper transformations which, based on the actual use of a harmonic oscillation at the frequency f_c, provide processed signals with spectral components different from zero only in a band around such frequency. Hereinafter, a real processed signal of the mentioned type, often indicated with the notation s(t) and usually represented by means of the complex envelope (see [3.158]):

$$[3.192]\ s(t)=\Re\left\{\iota(t)e^{j\omega_c t}\right\},$$

is named ***signal in shifted band***.
Often, since f_c is chosen quite high, in the mentioned transformations the relative bandwidth is very small after the operation so that the result is a signal in narrow shifted band; nevertheless, also signals in shifted band achieved with values of f_c such that even after the transformation the lower bound is much smaller than the upper one can exist, as well as for the signal which has been processed: definitively it may happen that a signal in shifted band is actually in baseband.

3.5.2.2 General characteristics of signals in shifted band

Let's consider a real signal s(t), of shifted band type, represented through its complex envelope $\iota(t)$, which can be considered a signal of baseband type on the basis of the position of its amplitude spectrum. If the complex envelope is expressed in terms of real part, $a_c(t)$, and imaginery part, $a_s(t)$, from the [3.192] the following particular representation of the signal is obtained:

$$[3.193]\ s(t) = a_c(t)\cos(\omega_c t) - a_s(t)\sin(\omega_c t).$$

It is possible to note that the two real baseband functions $a_c(t)$ and $a_s(t)$ provide the assessment of the time variable entity of the two harmonics in quadrature at the frequency $f_c=\omega_c/2\pi$; Thus, they are respectively named ***instant amplitude in phase*** and ***instant amplitude in quadrature***.
Utilizing the expression of the complex envelope in modulus, $\iota(t)$, and argument, $\alpha(t)$, still from the [3.192] a new form of the signal in shifted band is achieved, with the expression:

$$[3.194]\ s(t) = \iota(t)\cos[\omega_c t + \varphi(t)],$$

where the real function $\varphi(t)$, defined also out of the interval $(0, 2\pi)$ and introduced to generalize the representation, corresponds to the argument $\alpha(t)$ but considering an integer multiple of the circle angle and is expressed as:

$$[3.195]\ \varphi(t) = \alpha(t) + 2\pi n.$$

From [3.194] it is possible to note that the modulus $\iota(t)$ of the complex envelope is a function which passes through all the relative maxima of s(t), in the instants t_n in which $\omega_c t_n+\varphi(t_n)=2\pi n$ occurs; thus, $\iota(t)$ is named instantaneous envelope or, briefly, ***envelope*** of the shifted band signal; in Figura 3.18 as an example a study case is shown. Still observing the [3.194], the denomination ***instantaneous phase*** of the shifted band signal appears to be well justified, assigned to the function $\varphi(t)$ strictly connected to the argument of the complex envelope. Also the two real quantity $\iota(t)$ and $\varphi(t)$ can be considered baseband type signals.

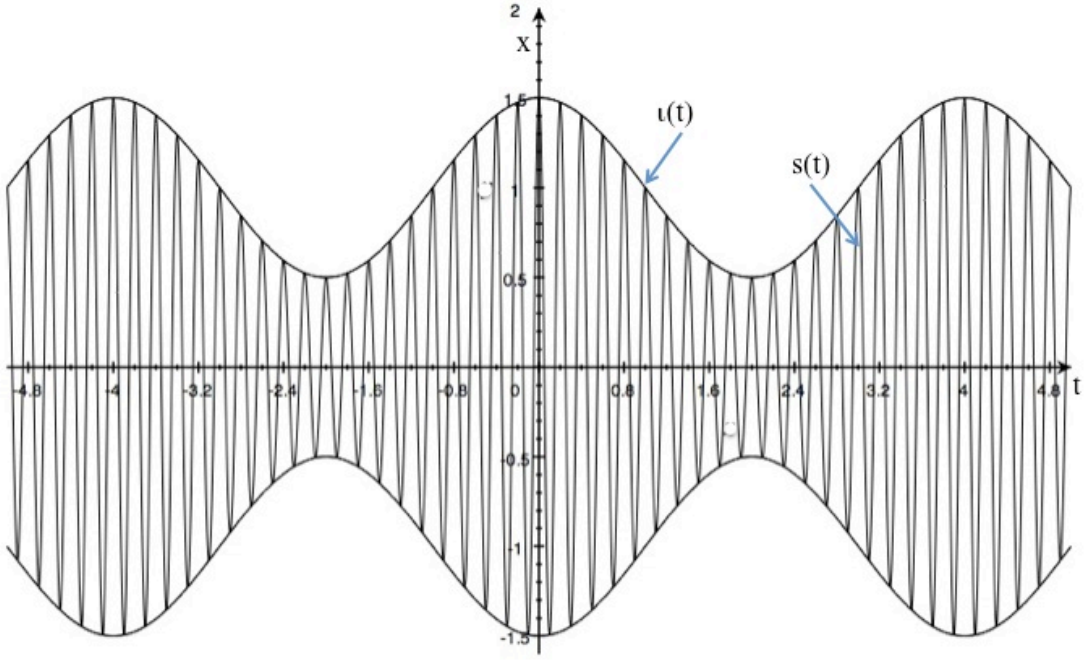


Figura 3.18: Example of envelope of shfted band signal.

To characterize the shifted band signal, just one of the pairs of quantities, either $a_c(t)$ and $a_s(t)$ which appear in the form [3.193] or $\iota(t)$ and $\varphi(t)$ which appear in the form [3.194] is sufficient. The following relations between the two pairs are valid:

$$[3.196]\ a_c(t) = \iota(t)\cos[\varphi(t)],$$

$$[3.197]\ a_s(t) = \iota(t)\sin[\varphi(t)],$$

as well as the inverse ones:

$$[3.198]\ \iota(t) = \sqrt{\left[a_c^2(t)+a_s^2(t)\right]},$$

$$[3.199]\ \varphi(t) = \text{artg}\left[\frac{a_s(t)}{a_c(t)}\right] + \frac{\pi}{2}\left\{1\text{-sgn}\left[a_c(t)\right]\right\} + 2\pi n\ ;$$

the last expression, in which the polidrom function arctg(x) is assumed to take the first determination in the interval $(-\pi/2;\ \pi/2)$, provides the value of the argumemt $\alpha(t)$ of the complex envelope, but considering an integer multiple of the circle angle.
In the following sections, especially in those focused of the fundamentals of modulation, it will be frequently possible to realize that once f_c is known, the information content can be fully attributed to the pair of signals $a_c(t)$ and $a_s(t)$, as well as to the correspondent pair composed of $\iota(t)$ and $\varphi(t)$. Anyway, it is possible to verify that, in particular, just one of the two quantities can be responsible of the information transfer, as for example in the case of unique instant amplitude, for which, disregarding a meaningless constant phase, it is possible to set:

$$[3.200]\ s_a(t) = a_c(t)\cos(\omega_c t),$$

or in the case just the envelope varies, typically with phase equal to zero,

$$[3.201]\ s_\iota(t) = \iota(t)\cos(\omega_c t),$$

or once again if just the argument varies, typically with constant envelope,

$$[3.202]\ s_\alpha(t) = A_c\cos[\omega_c t + \varphi(t)].$$

3.5.2.3 Particular characteristics of phase continuous signals

Looking at the expression [3.199] of the instantaneous phase φ(t), it is possible to note that in the instant in which $a_c(t)$ becomes zero with inversion of the sign, the discontinuity equal to π produced by the latter addend in the second member compensates that introduced by the former, but only if in the same instant $a_s(t)$ is different from zero; the function φ(t) instead turns out to be discontinuous in every particular instant in which the simultaneous inversion of the sign of both instant amplitudes occurs, or just one of them, in the case the other is identically zero, when necessarily also the envelope ι(t) becomes zero (see [3.198]). Of course further occurrence of discontinuity in the instant phase can occur in the instants in which $a_c(t)$ and $a_s(t)$ are discontinuous. In conclusion, it is possible to have a ***phase continuous signal***, characterized by no discontinuity of the function φ(t), if the instant amplitudes are continuous and the envelope is never equal to zero (or if it becomes zero, but without inversion of the sign of the amplitudes).
It is possible to observe that the frequency f_c of a harmonic oscillation can be defined as the derivative with respect to time of the argument of the circular function, related to the circle angle. In the case of a shifted band signal with continuous phase, in addition to the instant phase function φ(t), it is possible to refer to the ***instant frequency*** of the signal, which is defined as an extension of the harmonic case through the:

$$[3.203]\ f(t) = \frac{1}{2\pi}\,\frac{d\left[\omega_c t+\varphi(t)\right]}{dt} = f_c + \frac{1}{2\pi}\,\frac{d\varphi(t)}{dt}.$$

Once introduced also the ***frequency instant deviation*** of the signal:

$$[3.204]\ \Delta f(t) = f(t) - f_c,$$

the following relation is achieved:

$$[3.205]\ \Delta f(t) = \frac{1}{2\pi}\,\frac{d\varphi(t)}{dt}.$$

In general, it shows that an exact relation of linear type between the frequency instant deviation and the instant phase of the continuous phase signal exists, so that any processing concerning the latter impacts also the former quantity, and viceversa. Referring to $\Delta\omega(t)=2\pi\Delta f(t)$ instead of φ(t),it is also possible to achieve the following expression of the phase continuous signal:

$$[3.206]\ s(t) = \iota(t)\cos\left[\omega_c t+\int_{-\infty}^{t}\Delta\omega(t)\,dt\right].$$

It is possible to note that through the inverse of the intervals between two consecutive increasing crossing of the time axis of a phase continuous signal it is possible to know the values, averaged in the intervals, of the instant frequency, as shown in the example with constant envelope in Figura 3.19, as well as it is possible to observe the value of the frequency of a harmonic oscillation through the inverse of the period, identified by the consecutive instants in which the growing time evolution crosses the axis of the abscissa.

3.5.3 Representation of the complex envelope through samples

In the case of generic real signal strictly limited in band, with maximum frequency f_M and minimum $f_m \neq 0$, both in the event it has been transformed into a shifted bandwidth signal or not having been processed in such a way, it is anyway valid the representation of type [3.193]:

$$[3.207]\ x(t) = a_c(t)\cos(\omega_a t) - a_s(t)\sin(\omega_a t),$$

in which for the complex envelope $\iota(t) = a_c(t)+ja_s(t)$ the central frequency $f_a=(f_M+f_m)/2$ of the monolateral bandwith is chosen as reference.

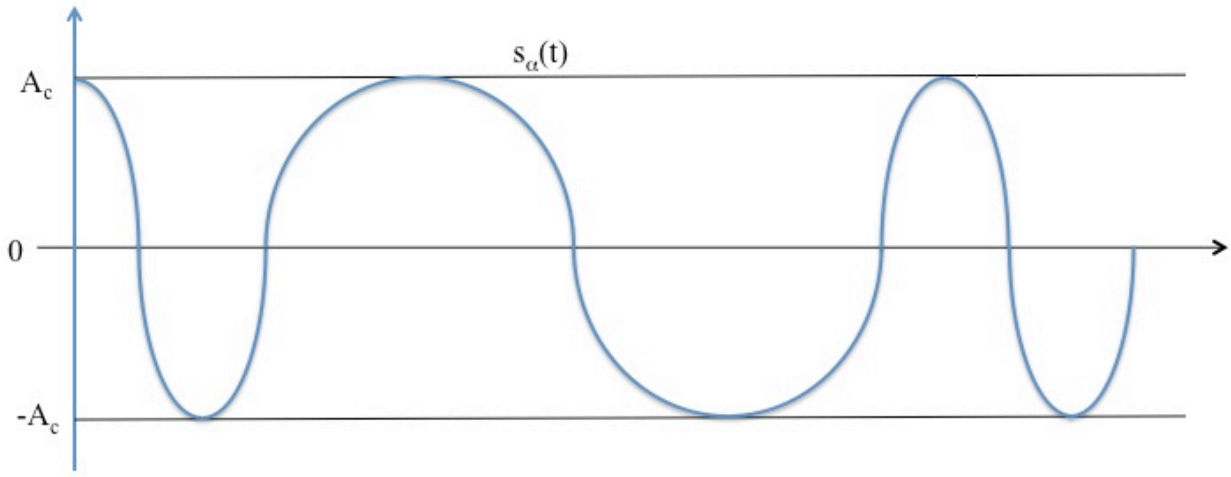


Figura 3.19: Example of frequency shifted signal with constant envelope.

Thus the instant amplitudes in phase and in quadrature, $a_c(t)$ and $a_s(t)$, turn out to be real signals of baseband type with maximum frequency $f_{aM} = f_M - f_a = B/2$ and minimum frequency $f_{am}=0$. As a consequence they can be in turn represented by their sample interpolation, through the following expressions:

$$[3.208]\ a_i(t) = \sum_k c_{ik} \operatorname{sinc}\left(\frac{t}{T_{ac}} - k\right),\ i = c, s,$$

where k is integer, the values of the samples are given by $c_{ik}=a_i(kT_{ac})$ and the sampling interval respects the condition:

$$[3.209]\ T_{ac} < \frac{1}{2f_{aM}} = \frac{1}{B}.$$

Anyway, assuming the maximum allowed sampling interval, the representation applied directly to the signal x(t) permits to rebuild it starting from the sequence of the samples at interval $T_N=1/2f_M<1/2B$, while the representation applied separately on the real part and on the coefficient of the imaginery part of the complex envelope, needs two sequences of samples instead of just one, but with interval $T_{aN}=1/2f_{aM}=1/B$ at least the double, saving on the total number of necessary samples for unit of time. As a matter of fact, there is no saving if the signal x(t) is of baseband type, for which $B \cong f_M$ can be assumed, while it is very significant in the case of narrow relative band, with $B << f_a$.

As already mentioned, it is possible to assume that every physical signal in practice is both limited in duration T, and in band B; as a consequence it can be always represented with good approximation through a finite set of real samples. The minimum number N of such elements represents the number of the ***freedom degrees*** of the considered signal; since in general two sequences of samples relative to the complex envelope are needed, it corresponds to the double of the integer number which approximate the ratio T/T_{aN}. Definitively, the result is:

$$[3.210]\ N \cong 2BT,$$

independently on the position of the signal bandwidth over the frequency axis. What just pointed out is very meaningful to evaluate the quantity of information that can be transferred with a signal.

4 ELEMENTS ON SOURCE SIGNALS

4.1 ANALOG OR DIGITAL SIGNALS

Aiming to examine from the information content point of view the characteristics of the signals transmitted by terminals which operate as sources, it is opportune to briefly address the information flows which being non electromagnetic quantities can access the trasducers present inside such sources.
Depending on the nature of the information, first of all it is possible to classify the signals in two main cathegories:

- ***continuous information flows***, in which there is a non electromagnetic quantity which varies in the time domain assuming values in a continuous codomain (range);
- ***discrete information flows***, or ***data flows***, in which there is a time sequence of countable symbols in discrete instants, as for example in the case of printed characters.

Thus, depending on the characteristic of the generated information flow, the corresponding waveform transmitted by the source can be:

- ***analog signal***, with time evolution similar to that of the information flow, i.e. with instant values proportional to the quantity and variable with continuity;
- ***digital signal***, with time evolution of several types, but created to be anyway perfect vehicle of the data flow.

In general, in an analogue source signal every instantaneous value is associated to the information, so that the ideal transfer of the information can occur only delivering to the destination terminal the unaltered waveform or a faithful version (fully correlated), with the exemption of the theoretical possibility of a different evolution, but with the same significant samples, rarely pursued in practice. Instead, in the digital case it is possible to achieve the ideal transfer of the data flow even delivering to the destination terminal a signal significanltly different from that created by the source, as it can be verified approaching digital communications.
This section is focused just on the collection of some basic elements on analog signals of audio and image sources, on the development of basic concepts on data flows and on the correspondent digital signals and on some considerations concerning the analog to digital signal conversion.

4.2 ELEMENTS ON ANALOG SOURCE SIGNALS

4.2.1 Audio signals

A signal is named ***audio signal*** if it is generated by an acoustic-electrical transducer, typically a microphone, capable to release the electrical quantity which is time dependent similarly to a continuously variable information flow and arrives in form of acoustic wave to the considered source; the final user of such a signal is an electrical-acoustic transducer, either the speaker or the earphones in case a very low acoustic level is sufficient. The audio signals, of analog type, can be modeled as power waveforms, with zero mean and with continous and symmetrical evolution, i.e. with equal modulus of maximum and minimum values.
In telecommunications the audio signals are characterized more for the range assigned to their spectrum than for the particular waveforms or powers, very variable for different voices (men and women), different languages or different possible music sources; in fact, it is known that the hearable spectral components range from a tenth of Hz up to about 15000 Hz, but often it is not necessary to transmit the audio information on the entire hearable band. As it will be discussed in a naxt chapter, the reduction within some standardized transmission bands, typical of a limited number of different applications, is obtained through a linear processing with cut of the band, both upper and lower, i.e. with appropriate band-pass filter.
The ***telephone signal*** is the audio signal with the narrowest assigned bandwidth. The voice power spectral density W(f), relative to the total power W, is distributed like that shown in Figura 4.1, but aiming to recognize the speaking person, i.e. to reach the scope to guarantee at least the commercial quality in a conversation by telephone, it is sufficient to transmit just the spectral

components present within the ***net telephone band***, with monolateral range from 300 Hz to 3400 Hz, while to achieve just the intelligibility the maximum frequency can be decreased down to 2000 Hz. Often the reference is the ***gross telephone band***, extended from zero up to 4 kHz, aiming to define the interval whitin which the spectral components of other signals potentially disturbing don't have to be present.

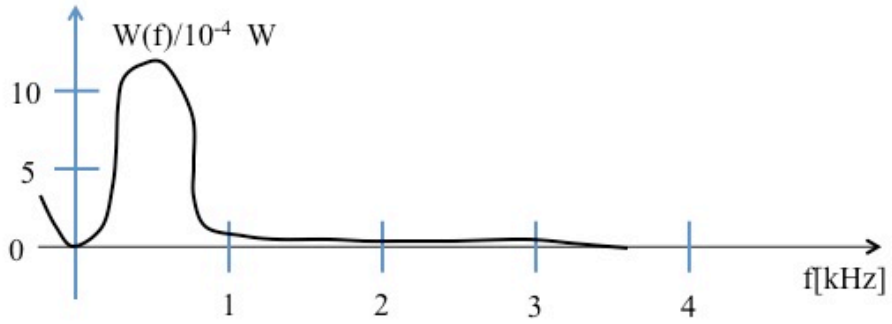


Figura 4.1: Power spectrum of a voice signal.

The spectral extension of the ***audio signal***, characterized by a pleasant listening quality, is greater and the bandwidth can be increasingly assigned as a function of the target desired quality. In the ***normal audio*** version, typical of the classical sound radiobroadcast, the band ranges from 50 Hz up to 4500 Hz; in the ***music audio*** version, typical of the quality sound radiobroadcast, the spectrum is wider with lower bound around 30 Hz and upper bound around 12 kHz; finally, in the case of ***high fidelity*** signal the bandwidth is the widest, with extremes out of the interval from 20 Hz to 15 kHz.

4.2.2 Image signals

4.2.2.1 Image signal by scanning

Many ***image signals***, that can be modeled as power signals, derive from the analogical interpretation of a time sequence of fixed optical images, realized inside the source terminal by transducers with the method of ***scan by lines***. As schematically shown in Figura 4.2, an optical-electrical transducer explores one by one each plane image one line after the other, sufficiently distant between each other, moving along each of them with constant speed v.

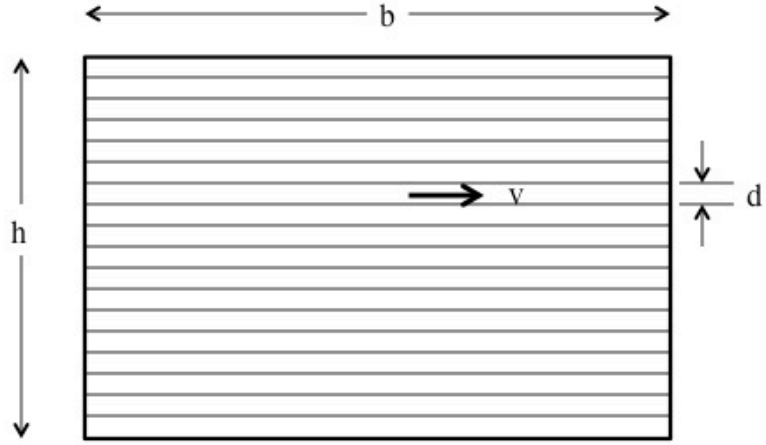


Figura 4.2: Fixed image under scan by raws.

A first characteristic quantity of the operation is the ***image time***, indicated with T_{im}, sum of the exploration time and of the dead time necessary to arrange the sensor to scan a new image; the inverse quantity $f_{im}=1/T_{im}$ is the ***image frequency***. A second quantity is the ***line time***, indicated with H, during which a line is examined (actual line time) and the sensor is arranged to scan the next line (dead line time); the inverse quantity $f_{ri}=1/H$ is the ***line frequency***. The ratio between the two intervals, or the inverse one between the two considered frequencies:

$$[4.1]\ N \triangleq \frac{T_{im}}{H} = \frac{f_{ri}}{f_{im}},$$

named ***number of lines***, is thus greater that the actual number of lines N_e in which the image is fractionated, due to the dead time before the beginning of the new one, so that it is possible to write:

$$[4.2]\ N_e \triangleq \eta_{im} N,$$

with η_{im} a bit less than one. Similarly, concerning the actual line time it is possible to set:

$$[4.3]\ H_e \triangleq \eta_{ri} H = \frac{\eta_{ri}}{f_{ri}},$$

still with η_{ri} a bit less than one.
Supposing for sake of simplicity that the sensor produces with continuity an analog signal with instant value x(t) proportional to just one optical quantity, for example the brightness of the little area observed at the moment, as a result of the scanning by lines an evolution like that shown in Figura 4.3 is for example obtained, in which the minimum value of the signal (x_m=x_N=0) corresponds to the black (no brightness) and the maxmum value of the signal (x_M=x_B=A_M) corresponds to white (maximum brightness), while through the ***cancellation pulses*** nulls are imposed corresponding to the line dead time.

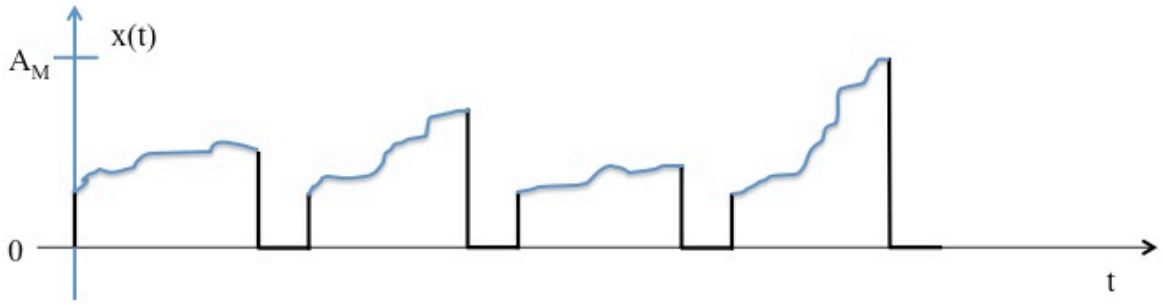


Figura 4.3: Example of piece of signal obtained by scanning by lines.

It is possible to note that in the considered image signal, in addition to the component of information due to the brightness variation of of the little area instantly explored, there are also a DC component, due to the mean brightness of the image sequence, and two series of harmonic components, with main periods T_{im} and H, i.e. with frequencies multiple of image and line ones. The spectral extension is only theoretically unlimited: in fact it is possible to determine an upper finite limit of the band f_M, both considering a finite resolution of the scanning and assimilating the discontinuities due to the cancellation with segments of signal with high steepness, but finite; the interested reader can go more in deep in section 4.2.2.2.

4.2.2.2 Bandwidth of a scanned image signal

Let's note that the distance d between two adjacent scanned lines would provide the vertical sampling distance of the image if the line had no thickness; actually, as the sensor scans a very small, but finite, area it is realistic that the image vertical resolution is worse, assuming as the ***sampling distance*** the quantity:

$$[4.4]\ d_c \triangleq \sqrt{2}d = \sqrt{2}\frac{h}{N_e},$$

where it is known that h is the image height. Imposing the same resolution also for the horizontal direction, to cover the sampling distance d_c the sensor needs a time, named ***sampling interval*** of the signal, which is:

$$[4.5]\ T_c = \frac{d_c}{v} = \frac{d_c}{b} H_e,$$

where v is the exploration speed of a line, with a lenght equal to b which is the base of the image.
The maximum frequency of the band needed by just the information component is achieved by the following expression:

$$[4.6]\quad f_M = \frac{1}{2T_c} = \frac{1}{2\sqrt{2}}\frac{b}{h}\frac{N_e}{H_e} = \frac{1}{2\sqrt{2}}\frac{b}{h}\frac{\eta_{im}}{\eta_{ri}}Nf_{ri} = \frac{1}{2\sqrt{2}}\frac{b}{h}\frac{\eta_{im}}{\eta_{ri}}N^2 f_{im},$$

in which N^2 is justified because the image is bidimensional.
Taking into account the [4.1], from the previous expression approximately $f_M \approx Nf_{ri}$ is carried out; then, if the number of lines N is very high, in the band limited by f_M a very large number of harmonics of the line frequency can be accommodated, so that the cancellation pulses can be surely considered highly steep. Definitively, the [4.6] provides thus the maximum frequency of the band of the considered signal.

4.2.2.3 Black and white video signal

Thanks to the persistence of the images on the human eye's retina it can perceive the continuity of the vision evolution even if a sequence of different fixed images is presented: to reach this goal it is necessary that $f_{im}>15$ Hz in order to achieve the sensation of continuity of the motion and to have at least 45 enlightenments per second to avoid the flickering inconvenience (periodic variation of brightness). In the case of television the interleaved lines image scanning method is applied, dividing the image in two frames, one for the odd lines, explored first, and the other for the even lines, explored after that. In the European analog standard $f_{im}=25$ Hz is adopted as the image frequency, which allows a good sensation of continuous motion, but to remove flickering a double ***frame rate*** is necessary.
In the old European analog standard with b/h=4/3, the number of lines N is equal to 625, so that the following figures are obtained:

$$f_{ri} = 15.625 \text{ Hz}, H = 64\ \mu s;$$

the actual number of raws N_c is fixed to be equal to 585, while the actual line time H_e is eqaul to 82% of H, so that the result is:

$$\eta_{im} = 0{,}935,\ \eta_{ri} = 0{,}82.$$

With such results a huge spectral extension is achieved, with maximum frequency of about 5 MHz (the interested reader can use the [4.6]).
In the case of black and white signal, the instant value of the signal x(t) coming out from the transducer, named ***vision signal***, is proportional to the brightness of the little area simultaneously explored, excluding during the dead times, during which the cancellation pulses are produced, necessary for the sensor to come back horizontally and vertically. In order to guarantee the synchronization to the receiving transducer which reproduces the image during the horizontal dead times, with duration 0.18H=11.52 μs, ***line synchronization pulses*** are inserted (see Figura 4.4), almost rectangular with duration 0.09H with one level corresponding to black ($x_N=0$) and the other one with a lower level ($x_m=A_m<0$, with $A_m/A_M=-3/7$), while in the vertical dead times between one frame and the following, lasting $0.064/2f_{im}=1.28$ ms, other ***frame synchronization signals*** are added, composed of several quasi rectangular pulses with the two mentioned levels, 0 and $A_m<0$. The so completed signal consitutes the ***black and white video signal***.
In the case of color, the scanning is of the same type, but with a multiple sensor based on the triple-color. As a matter of fact three different simultaneous signals are achieved, that together with the synchronization signal are manipulated in a complex manner to obtain a ***color video signal*** of compatible type, i.e. also usable by an equipment arranged only for black and white signal. In fact, the colored signal is composed of a luminance signal, which carries the information of brightness and has the same temporal and spectral characteristics of the black and white signal, and a pair of chrominance signals that are processed and added to the former with an appropriate technique, which allows to separate and to give back each of them, also sharing the bandwidth, which therefore remains unchanged as extension.

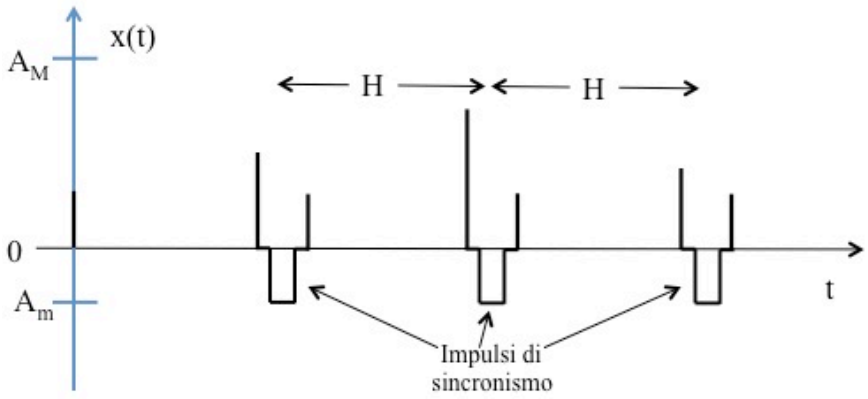


Figura 4.4: Synchronization pulses inserted in the cancellation intervals.

The video signals in addition to have a DC component have spectral components which extend down to zero frequency, without which the information slowly varying relative to the average brightness of the scene would be lost; in practice, however, it is possible to transmit video signals by removing the lower part of their spectrum, but only if below f_{IM} = 25 Hz, so that f_m>0 is obtained: in fact, it is possible to restore the absolute levels of the original signal on the basis of the height measurement of the synchronization pulses of the signal transmitted with the lower bandwidth limit, given that the considered height is a deterministic quantity.

4.3 BASIC CONCEPTS ON DIGITAL SOURCE SIGNALS

4.3.1 Synchronous data flows and digital sequences

Let's consider an information source which generates a synchronous data flow D(n) composed of a regular series, at constant time interval T, of ***simbols*** D_k which belong to a discrete countable and finite set $\{D_q\}$, with q=1, 2.., M. Such set of symbols, named ***M-nary alphabet*** because its cardinality is M≥2, can be of any type, as for example the latin letters alphabet, the set of the characters of a keyboard or, in the usual case of M=2, the binary alphabet composed of the binary digits, or ***bits***, corresponding to the numbers 0 and 1.
Once established a biunivocal correspondence between the symbols of the source alphabet and the same number of different signal values, indicated with the notation

$$[4.7]\ \{D_q\} \Leftrightarrow \{x_q\},\ q=1,2,..,M,$$

a real sequence x(n) can be associated to the information flow D(n) with the generic sample x_k assuming one of the values of the set $\{x_q\}$, of course discrete, countable and finite as well; in this way a sequence with discrete values, or ***digital sequence***, can be achieved and that can be supposed to be the vehicle of the symbols of the corresponding synchronous data flow, with the same timing.
As a new symbol is generated for each constant time interval T, such a quantity, common characteristic of the data flow of the digital sequence, is named ***symbol time***; the inverse quantity,

$$[4.8]\ R = \frac{1}{T},$$

which provides the number of symbols of the information flow per time unit, as well as the number of real discrete samples of the corresponding digital sequence per time unit, is then named ***symbol rate***. In the case of binary alphabet, the interval specifically named T_b, is named in particular ***bit time*** and the inverse quantity,

$$[4.9]\ R_b = \frac{1}{T_b},$$

which provides the number of bits of the information flow per time unit, as well as the number of real binary samples of the corresponding digital sequence per time unit, is then the ***bit rate***. The unit of measure of the latter is the bit per second so that the used notation is bit/s; often the multiples are utilized, i.e. kbit/s (10^3), Mbit/s (10^6) e Gbit/s (10^9).

4.3.2 Source multilevel signals

The digital sequences x(n), being discrete time signals, are very useful from logical point of view, but they are abstract. However, it is immediate to derive a type of waveforms to which the time continuous digital signals, generated by terminals in which the information source generates a synchronous data flow, can actually tend towards: supposing to use simple electronics circuits with instant snap and hold for an interval T, typically as output of the terminals ***synchronous step signals*** like the following are obtained:

$$[4.10]\quad x(t) = \sum_k x_k \, \mathrm{rect}\left(\frac{t}{T} - k\right),$$

in which each discrete sample x_k of the sequence really occurs through a ***level*** of signal having the same value, which is kept constant over all the symbol time T. The synchronous form like [4.10] often is named in general ***multilevel signal***; considering the cardinality M, the particular cases of ***binary signal*** (M=2), ***ternary signal*** (M=3), ***quaternary signal*** (M=4) and so on can be obtained.
Considering the close relation between the multilevel signal and the corresponding digital sequence, the symbol rate R, or the bit rate R_b, and the symbol time T, or the bit time T_b, are characteristic quantities of the multilevel signal, or of the binary signal, too.
As an example, Figura 4.5a shows a part of a quanternary signal, while Figura 4.5b shows the corresponding string of the symbol flow; it is easy to verify that the biunivocal correspondence [4.7] corresponds to the associations: $D_1 \Leftrightarrow x_1=-3$, $D_2 \Leftrightarrow x_2=-1$, $D_3 \Leftrightarrow x_3=1$, $D_4 \Leftrightarrow x_4=3$.

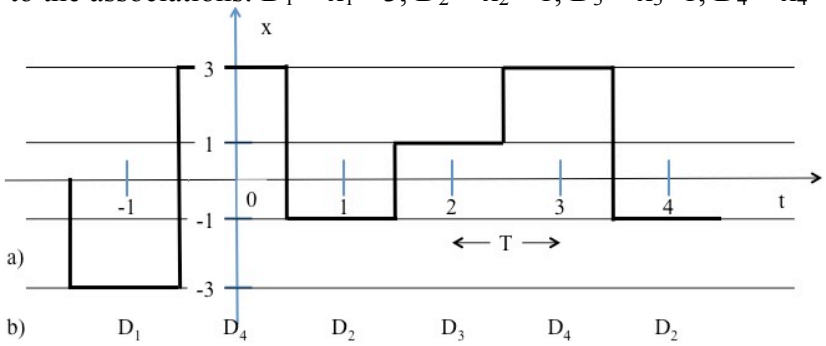


Figura 4.5: Example of a part of multilevel signal.

The multilevel signals, being achieved through linear combination of unit rectangular pulses shifted in time (see [4.9]), have a spectral extension theoretically unlimited, i.e. they are ideal. As a consequence of the practical limitation of the bandwidth, the ***digital signals*** actually generated by sources that work on the basis of synchronous data flows have evolutions that are slightly different from the ideal ones, with steep transitions instead of discontinuities and values nearly constant within each symbol time (see Figura 4.6); the actual bandwidth results to be limited but nevetheless very large, with upper bound f_M greatly higher that the inverse of T.
The multilevel signals, or better the digital signals that well represent them, have waveforms that are well suitable to be processed within a single equipment, becasue the level detection through sampling performed in presence of errors in the reading instants is slightly critical; instead, it is difficult to handle them for the long path transmission, because very often they implies bandwidth requirements too severe especially in case low cost transmission systems are utilised. As it will be presented in a folllowing chapter, it is possible to get different forms of digital signals, with reduced spectral extension, without degradation of the associated data flows.

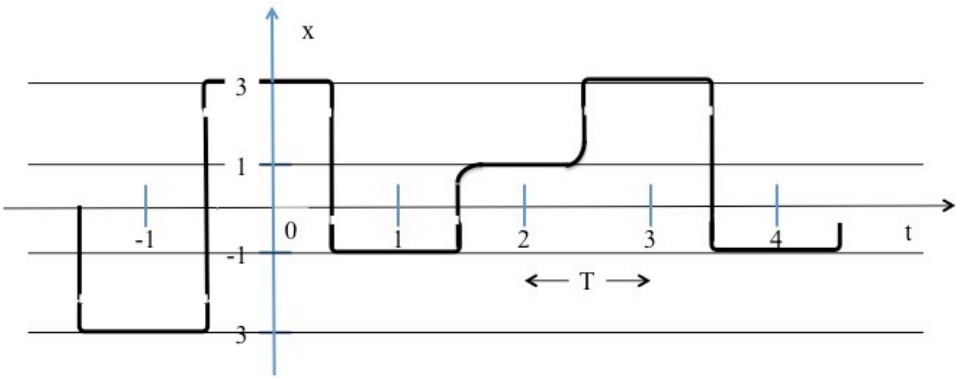


Figura 4.6: Example of a part of source digital signal with reduced bandwidth.

In practical transmission systems the signal to transfer is very often of binary type, considering that in the source terminals the use of particular processing is usual, which are named source coding and out of the scope of this text, which also convert the M-nary alphabets into the binary one.

4.3.3 Binary source signals

Often the biunivocal correspondence between the symbols s_q of the alphabet, with cardinality that can be expressed by

$$M_S=2^S$$

and the as many **binary words** of S binary digits or **bits** (0 and 1) are used; for example

$$s_1 \Leftrightarrow 00,\ s_2 \Leftrightarrow 01,\ s_3 \Leftrightarrow 10,\ s_4 \Leftrightarrow 11.$$

Being S an integer, the number M_s is equal to 2 or 4, 8, 16, 32, 64, etc.

Thus, a binary flow b(n) with bit time $T_b \triangleq T_S/S$, with $T_b \leq T_S$ results to be associated to the data flow s(n), with symbol time T_s.

The inverse quantity $R_b \triangleq RS$ is named **bit rate** and is measured in bit/s.

For example, associating $x_1 \Leftrightarrow 0$, $x_2 \Leftrightarrow 1$, a **binary sequence** x(n) with $T_S=kT_b$ results to be associated to the flow s(n); then, considering the time continuous form the source **binary signal** is obtained.

The bit time T_b and the bit rate R_b are characteristic quantities of the binary signal too. Of course a binary signal is the particular case of a multilevel signal for the cardinality of the alphabet M=2. In the example in Figura 4.7 the assumtions are:

$s_1 \Leftrightarrow 000$	$s_2 \Leftrightarrow 001$	$s_3 \Leftrightarrow 010$	$s_4 \Leftrightarrow 011$
$s_5 \Leftrightarrow 100$	$s_6 \Leftrightarrow 101$	$s_7 \Leftrightarrow 110$	$s_8 \Leftrightarrow 111$
bit $0 \Leftrightarrow x_1=0$		bit $1 \Leftrightarrow x_2=A$	

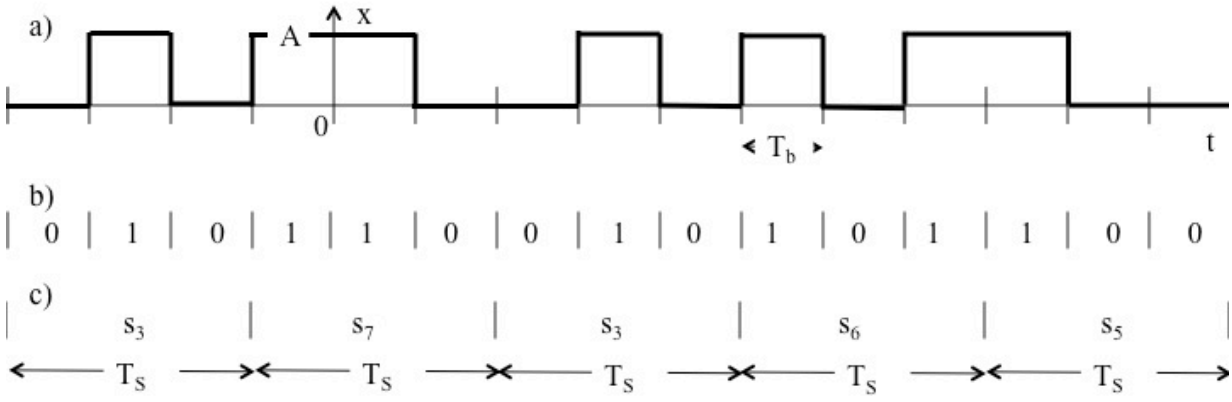


Figura 4.7: Source binary signal

4.3.4 Synchronous and asynchronous signals

All the multilevel signals expressed through the time series [4.10] have the same property to keep always unaltered the time interval between adjacent addends: thus, the attribute ***synchronous***, adjective usually implied considering the extremely low number of cases in which the mentioned property doesn't apply, is confered to each of them. Nevertheless, when the property doesn't apply the digital signals are named ***asynchronous***. In such latter type of signals it is not possible to refer to just one time interval, considering that the more general represention is valid:

$$[4.11]\ x_a(t) = \sum_k x_k \, \mathrm{rect}\left(\frac{t-t_k}{T_k}\right),$$

where the instants t_k are no longer constrained by regularity. Instead of the symbol rate, the ***signalling speed*** is then considered:

$$[4.12]\ R_V = \frac{1}{T_V},$$

being T_V the minimum value of the intervals t_k-t_{k-1}. The measure unit of R_V is the ***baud***, which expresses the maximum number of symbols that can be carried on by the asynchronous signal in the unit of time.
An example of binary asynchromous signal is shown in Figura 4.8a, corresponding to the hystorical message S O S (three dots, three lines, three dots) in the Morse alphabet. One present case, still binary, is instead that one exemplified in Figura 4.8b, relative to the emission of keyboard characters that occurs as asynchronous succession, with random starting times, of short energy signals associated to strings of a few bits (the initial bit, or start, and the final one, or stop, are always the same, while the word composed of the intermediate 8 bits of each string is typical of the character).

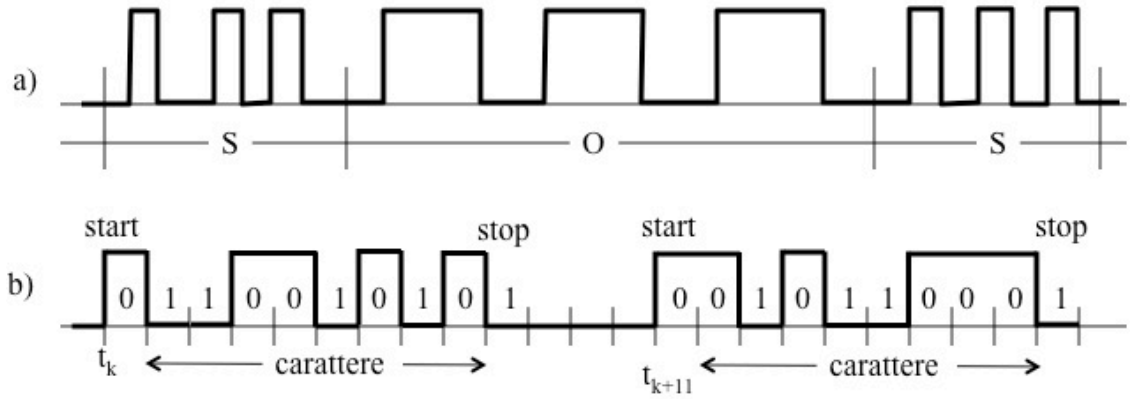


Figura 4.8: Examples of part of binary asynchronous signals.

5 SIGNALS IN LINEAR BIPOLES AND QUADRUPOLES

5.1 LINEAR TRANSFORMATION BETWEEN TIME CONTINUOUS SIGNALS

Referring to what already introduced in section 3.3.1, the following expression of a generic linear transformation is recalled:

$$[5.1]\ T\{x(v), z\} = \int x(v)h_n(z, v)dv = \left(x(v), h_n^*(z, v)\right),$$

which acting on an operand x(v), complex function of an independent real and continuous variable v, leads to the transform, complex function of an independent continuous variable z. Assuming that the operand is a signal x(v) time continuous (v=t) and that the transform is still a function of the independent real variable t, specifically a ***linear transform between time continuous signals*** is achieved:

$$[5.2]\ y(t) \equiv T\{x(v), t\} = \int x(v)h_n(t,v)dv = \left(x(v), h_n^*(t, v)\right),$$

where to not generating confusion the time variable in input keeps to be indicated with the notation v. Such kind of transformation is typically experienced by a signal in input, or ***excitation***, when it interacts with a linear system, which provides at the output a transformed version, named ***response***.
A linear system receiving at the input the ideal pulse at the instant v_0 provides the response:

$$[5.3]\ \int \delta(v-v_0)h_n(t,v)dv = h_n(t,v_0);$$

the nucleous of the transformation calculated in the value v_0 can be then interpreted as the response to the ideal pulse applied in such an instant. In particular, assuming at the input the excitation δ(v) with v_0=0, the corresponding signal in output is named ***pulse response*** (see Figura 5.1), indicated with:

$$[5.4]\ h(t) = h_n(t,0).$$

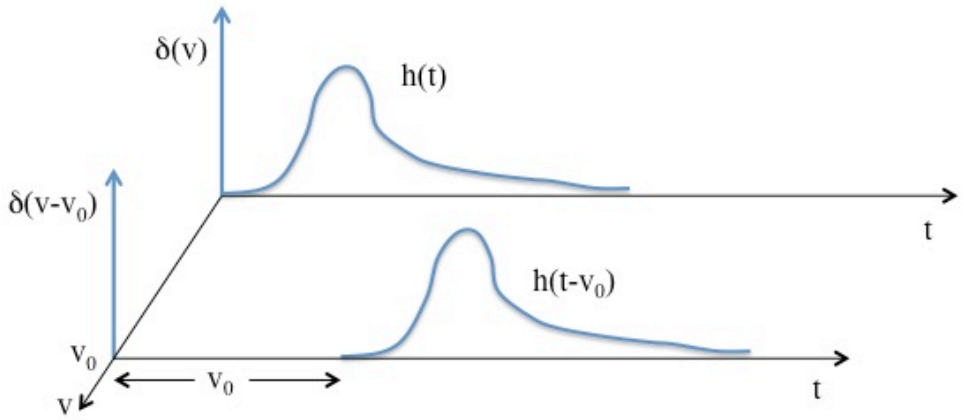


Figura 5.1: Impulse response h(t) and responce to the ideal impulse applied at time v_0.

A transformation between signals can also enjoy the property of ***time invariance*** or ***shift invariance***, which occurs whenever applying a signal at the input of a system the response is always the same regardless the instant of application, but implying a time shift of the same amount of that occurred to the input signal.
In the case of ***linear and time invariant*** (LTI) transform or ***linear shift invariant*** (LSI) transform the parameters characterizing the operator are fixed in time and the property of the substantial time unchangability of the effect, i.e. the response to an excitation shifted in time of a generic t_0 is always the same but with a time shift of the same entity, is valid. Then, the response $h_n(t, v)$ to the ideal pulse at the generic time v must be the same of the impulsive response but shifted of the quantity v, so that the nucleous assumes the particular expression (see Figura 5.1, with $v=v_0$):

$$[5.5]\ h_n(t,v) = h(t-v).$$

A LTI transformation between signals can be expressed as:

$$[5.6]\ y(t) = \int x(v)h(t-v)dv = \left(x(v), h^*(t-v)\right),$$

i.e. the response is the result of the ***convolution*** operation (see [2.11]) between the excitation signal x(t) and the impulsive response h(t):

$$[5.7]\ y(t) = h(t) * x(t).$$

Usually, considering an impulsive response with duration D not zero and at least practically finite, the transformation [5.6] shows that the instant value y(t) of the response at time t is the result of the interpolation of all the instant values of the input signal x(v) corresponding to the time interval, of D extension, in which the function h(t-v) is different from zero (see Figura 5.2); thus, a LTI transformation is in general assumed to be of the type ***with memory***.

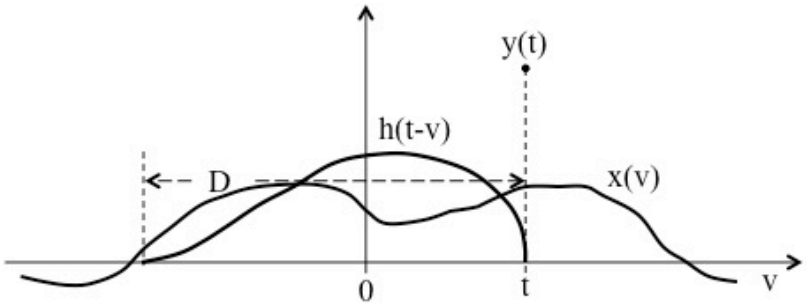


Figura 5.2: Instant value y(t) of the response calculated through the continuous interpolation of the values of the excitation x(v) with the values of the function h(t-v).

A particular very meaningful case is the ***memoryless*** LTI transformation, characterized by a pulse response of infinitesimal duration $g\delta(t-t_0)$, with g and t_0 constant, i.e. with $h(t-v)=g\delta(t-v-t_0)$ as nucleous; it is possible to verify immediately that the responses are faithful signals:

$$[5.8]\ y(t) = \int x(v)g\delta(t-v-t_0)dv = gx(t-t_0),$$

in which to figure out each instant value at the output just one corresponding instant value of the input is sufficient.
Finally, let's remind that for the generic ***linear time variant transformation*** the response $h_n(t,v)$ to the ideal pulse at the generic time v may no longer ensure to remain unchanged as a consequence of a time shift of amount v.

5.1.1 Remarks on the electric nature of the signals

5.1.1.1 Direct voltage and reflected voltage

Let's consider the elementary transfer of information in a generic section of a transmission system in which the connection between the two ends is realized with a pair of electrical cables. As in the scheme of Figura 5.3, the ends before and after the connection can be represented respectively by a generator bipole G and by an utilizer bipole U, while in the section the pair of real quantities, composed of the voltage v(t) between the two connection wires and the electrical current i(t) going in one of them and going out from the other, can be pointed out. It is important to underline that both the voltage and the current contribute to the information transfer, so that in general in each section of a transmission system a pair of real quantities function of time must be considered.

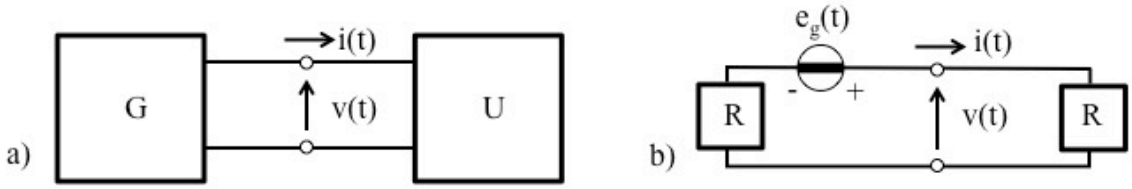


Figura 5.3: Representation of a two wires connection in general (a) and in the case of ideal matching (b).

The information results to be correctly transferred from one end to other end of the system if the connection respects the ***ideal matching*** condition, according to which the generator is a source of electromotive force $e_g(t)$ with internal resistance constant, real and positive R and the user can be represented by a resistance of the same value (see Figura 5.3b). In fact, in such a situation the different quantities are related between each other through $v(t)=Ri(t)=e_g(t)/2$, which indicates that v(t) and i(t) are not only proportional, and thus faithful between each other, but also faithful to $e_g(t)$; then, the maximum transfer of power is achieved, i.e. the instant power $w(t)=v(t)i(t)$ is completely absorbed by the user. Therefore, in the condition of ideal matching, regarding information just one of the two quantities can be considered in the identified section.

Let's consider the case, represented in the scheme in Figura 5.4, in which the generator is ideal but instead not the user as well, for which anyway the latter can be represented through a LTI passive bipole (Linear Time Invariant), characterized with the impedence $Z(f)\neq R$. The connection is no longer ideal: being v(t) and i(t) no longer proportional the signals, they cannot be assumed to be faithful between each other and as well not faithful to the electromotive force $e_g(t)$.

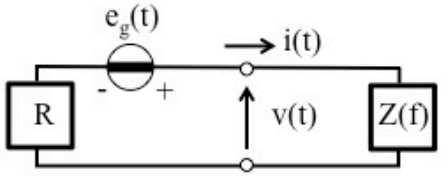


Figura 5.4: Two wires connection with ideal generator and LTI user.

Looking at Figura 5.4, it is possible to note that the electromotive force $e_g(t)$ is divided in the voltage drop Ri(t) on the internal resistance of the ideal generator and in the voltage v(t) at the user bipole clips, i.e. the relation is $e_g(t)=Ri(t)+v(t)$; then, any linear combination of v(t) and i(t) in the form k[v(t)+Ri(t)], with k real constant, results to be faithful with respect to the electromotive force $e_g(t)$, redardless the impedence Z(f) of the user. On the basis of such consideration the pair of real quantities in the connection section can be defined:

$$[5.9]\ v_d(t) = \frac{1}{2}[v(t) + R\,i(t)],\ v_r(t) = \frac{1}{2}[v(t) - R\,i(t)],$$

in which the former results to be anyway faithful to the electromotive force $e_g(t)$, and is the meaningful quantity for the information transfer, while the latter becomes even zero if the connection is ideal (Z(f)=R), showing to be associated to the deviation with respect to such a situation. Inverting the [5.9] the following relations are obtained:

$$[5.10]\ v(t) = v_d(t) + v_r(t),\ i(t) = i_d(t) + i_r(t),$$

where $i_d(t)$ and $i_r(t)$ are respectively:

$$[5.11]\ i_d(t) = \frac{1}{R}\,v_d(t),\ i_r(t) = -\frac{1}{R}\,v_r(t).$$

Utilizing the previous relations, the instant power transferred to the user LTI passive bipole assumes the expression:

$$[5.12]\ w(t) = [v_d(t) + v_r(t)][i_d(t) + i_r(t)] = v_d(t)\,i_d(t) + v_d(t)\,i_r(t) + v_r(t)\,i_d(t) + v_r(t)\,i_r(t) =$$

$$= \frac{1}{R}\,[v_d^2(t) - v_d(t)\,v_r(t) + v_r(t)\,v_d(t) - v_r^2(t)] = w_d(t) + w_r(t),$$

having assumed:

$$[5.13]\ w_d(t) = \frac{1}{R} v_d^2(t),\ w_r(t) = -\frac{1}{R} v_r^2(t).$$

It is evident that the instant power $w_d(t)$ is always positive, i.e. the significant quantity $v_d(t)$ provides immediately through the former of the [5.13] the power that in each instant the source sends to the user, while the instant power $w_r(t)$ is always negative, so that $v_r(t)$ is responsible to carry out from the user the power that it is supposed to receive on the basis of the latter of the [5.13]. For this reason the two quantities $v_d(t)$ and $v_r(t)$, which are the components of the voltage v(t) in the section, are respectively named ***direct voltage*** and ***reflected voltage***.

5.1.1.2 Direct signal and reflected signal

The quantities v(t) and i(t), and the pair $v_d(t)$ and $v_r(t)$ as well, are no longer significant in the case the connection is realized over a structure in which the propagation of the electromagnetic field with TEM (TEM=Transverse Electro Magnetic) mode, i.e. with electrical and magnetic vectors orthogonal between each other and both located over the plane of the section, is not allowed, as instead usually in the case of a structure constituted of two conductors isolated between each other. The mentioned situation occurs, for example, if in the connection section there is the typical configuration of an optical fiber or of a metallic waveguide.
Considering in general the pair of actual electromagnetic modes, one of which in the section is in charge of the direct propagation, with progressive wave from the generator to the user, and the other is in charge of the reflected propagation, with regressive wave in the opposite direction, anyway the respective instant powers can be defined with full physical sense: the direct one $w_d(t)$, always positive, the reflected one $w_r(t)$, always negative. Aiming to the widest generality, it is opportune to utilize the pair of real quantities defined through the:

$$[5.14]\ x_d(t) = \sqrt{w_d(t)},\ x_r(t) = \sqrt{-w_r(t)},$$

having as unit of measure the square root of the Watt, named respectively ***direct signal*** and ***reflected signal*** in the considered section.
In the case of ideal connection that occurs with no instant power reflected by the user, i.e. with $w_r(t) \equiv 0$ e $w(t) \equiv w_d(t)$, the following relation is obtained:

$$[5.15]\ x_d(t) = \sqrt{w(t)},\ x_r(t) = 0,$$

where $x_d(t)$ is the signal significant to transfer information and $x_r(t)$, which indicates the deviation from the ideal connection, is zero.
If the direct and reflected voltage are physically significant, through the [5.13] the following simple relations are obtained:

$$[5.16]\ x_d(t) = \frac{\pm 1}{\sqrt{R}} v_d(t),\ x_r(t) = \frac{\pm 1}{\sqrt{R}} v_r(t),$$

in which the ambiguity of the sign is not important, taking into account the considerations about the faithful signals.
Then, through the definitions [5.9] the following expressions of the signals as a function of v(t), i(t) are obtained:

$$[5.17]\ x_d(t) = \frac{1}{2}\left[\frac{1}{\sqrt{R}} v(t) + \sqrt{R} i(t)\right],\ x_r(t) = \frac{1}{2}\left[\frac{1}{\sqrt{R}} v(t) - \sqrt{R} i(t)\right],$$

which show how the direct and reflected signals are yet a linear combination with constant coefficients of the voltage and of the current in the section of the connection.

The use of the pair of direct and reflected signals, with the specified nature in the definitions immediately above, has the following advantages:

- the direct and reflected signals are homogeneous quantities, of general validity;
- the direct signal $x_d(t)$, the only one present in the ideal connection with $w_r(t) \equiv 0$, assumes the role of useful signal in the section;
- the reflected signal $x_r(t)$, identically zero in the ideal connection, highlights the deviation from the ideal case and is considered an undesired signal in the section;
- the electrical instant power transfered in the section is in general expressed by means of the simple relation $w(t) = x_d^2(t) - x_r^2(t)$, which in ideal condition is reduced to:

$$[5.18]\ w(t) = x_d^2(t), \text{ per } x_r(t) \equiv 0.$$

Moreover, the more general and simpler graphical single wire representation shown in Figura 5.5a can be used, indicating the way on which the connection between the generator and the user is realized, with the possibility to always identify an input or an output ***gate***, while not always the identification of a pair of poles to attach conductors is physically meaningful. In such a representation the ideality of the connection, which very often is considered to be satisfied, can be synthetically indicated omitting the reflected signal and removing the underscript in the notation of the direct signal (see Figura 5.5b).

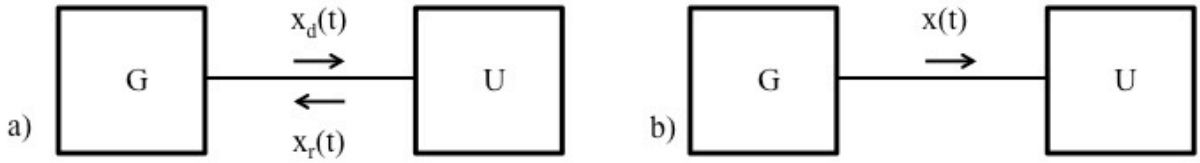


Figura 5.5: Representation of a connection with single wire scheme, in the generic case (a) and in condition of ideal connection (b).

Hereinafter, mainly signals with electrical nature corresponding to the sqaure root of the instant power carried by them will be addressed and, unless explicitly stated the contrary, the condition of ideal connection will be considered to be satisfied. Moreover, what above stated will be extended to the complex signals.

5.2 LTI TRANSFORMATIONS IN THE BIPOLES

5.2.1 Reflected response in the time and frequency domains

In the development of some basic considerations on the LTI transformations between signals it is worth to begin from the most elementary one, that occurs in an electrical subsystem with a single access section, or ***gate***, represented as shown in Figura 5.6, where the scheme with the access way over which the pair of direct and reflected signals are attached was used. Hereinafter, a single gate subsystem, on the basis of the possibility to represent it with an equivalent bipole, will be often named ***bipole*** even if the identification of a pair of access poles is not meaningful. The signals can be either both energy or power ones; the notation x(t) for the direct one, with the role of input exicitation signal, and y(t) for the reflected one, with the role of output signal, or response, are utilized.

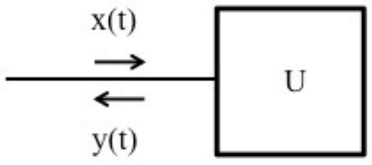


Figura 5.6: Scheme of a bipole with direct excitation signal x(t) and reflected response signal y(t).

Assuming that the considered subsystem concerns the user side after a generic section of a transmission system, the hypotesis that the representative bipole is passive, linear and characterized with fixed parameters in time is formulated; then, the reflected response y(t) results to be associated to the excitation x(t) through a transformation between signals of linear and time invariant (LTI) type.
It is known (see section 5.1.1) that in the case the LTI bipole is connected to a generator with equivalent internal resistance real and constant R, in the ideal condition of impedence of the user bipole equal to R, $y(t)\equiv 0$ is achieved for any x(t), implying that there is no reflected response; otherwise a transformed signal with the following expression is obtained:

$$[5.19]\ y(t) = \int x(v)\,h_\rho(t-v)\,dv = h_\rho(t) * x(t),$$

where in the integral the time variable in input has been indicated with v. Thus, the reflected signal can be obtained through the ***convolution*** operation between the direct signal x(t) and the ***reflected pulse response*** $h_\rho(t)$; the latter function, which characterizes the LTI bipole in the time domain with reference to the resistance R, is defined as the reflected signal corresponding to the excitation with the ideal pulse at the instant v=0 [$y(t)\equiv h_\rho(t)$ for $x(v)\equiv\delta(v)$], generated by a source with internal resistance R.
The transformation [5.19], usually of type with memory, is an operation applicable both in case of excitation energy signal and power signal, with response respectively of the former or of the latter type.
A passive LTI bipole is named ***physically feasible*** if its pulse response is real, belongs to the functional space L_1 and in addition is ***causal***, corresponding to the condition $h_\rho(t)=0\ \forall\ t<0$; instead, it is named ***ideally feasible*** if only the condition concerning the causality of the response is not valid. In the case of feasible LTI bipole, both ideally and physically, the Fourier transform of the reflected pulse response exists:

$$[5.20]\ \rho(f) = F\{h_\rho(t)\},$$

that also is characteristic function of the considered bipole with reference to the resistance R, but in the frequency domain. Supposing that it is possible to carry out the Fourier transform X(f) of the excitation signal of a passive LTI bipole, applying the property of the convolution in the time domain to the expression [5.19] of the transformation the corresponding relation in the frequency domain is achieved:

$$[5.21]\ Y(f) = F\{h_\rho(t)\}X(f) = \rho(f)X(f).$$

Taking into account that the excitation x(t) is the direct signal and that the response y(t) is the reflected signal, the function ρ(f) is then named ***reflection coefficient*** of the bipole.
Assuming that $h_\rho(t)$ is real, the reflection coefficient respects the property:

$$[5.22]\ \rho(f) = \rho^*(-f),$$

which implies that ρ(f) is a hermitian function, with even modulus and odd argument; thus, to know it only on the positive half axis of the frequency is exhaustive.
In ***ideal matching*** condition in the connection section the reflected response y(t) is identically zero for any excitation x(t): as a consequence the bipole must have a reflected pulse response $h_\rho(t)\equiv 0$, i.e. the reflection coefficient $\rho(f)\equiv 0$ over the whole frequency domain. The $h_\rho(t)\equiv 0$ is a sufficient but not necessary condition, in order to obtain $y(t)\equiv 0$; in fact, in the case of input signals of frequency limited type, with monolateral bandwidth extended from f_m to f_M, since the spectral components out of the band are zero, the reflected response is identically zequal to ero, as long as the much less severe condition:

$$[5.23]\ \rho(f) = 0,\ \forall |f| \in (f_m; f_M).$$

is satisfied. It is possible to note that in such condition of ***perfect matching*** in the connection section the reflected pulse response of the bipole is usually not identically zero.

5.2.2 Relation between reflection coefficient and impedence

If the quantities voltage and electrical current at the ends of the passive bipole representative of the user system have a physical meaning, the knowledge of the relation between the reflection coefficient ρ(f) and the impedence Z(f) can be useful, being also the latter a characteristic function of the bipole in the frequency domain.
Recalling the existing relations between the direct signal $x_d(t)$ and the reflected one $x_r(t)$ and the pair voltage v(t) and electrical current i(t) at the end of the bipole (see [5.16]), with the notations $x(t) \equiv x_d(t)$ and $y(t) \equiv x_r(t)$ the result is:

$$[5.24]\ x(t) = \frac{1}{2}\left[\frac{1}{\sqrt{R}}v(t)+\sqrt{R}i(t)\right],\ y(t) = \frac{1}{2}\left[\frac{1}{\sqrt{R}}v(t)-\sqrt{R}i(t)\right],$$

In the hypotesis of harmonic excitation at frequency f and applying the symbolic method, from the [5.21] and [5.24] the corresponding relations between phasors are carried out:

$$[5.25]\ \mathbf{Y} = \rho\mathbf{X},$$

$$[5.26]\ \mathbf{X} = \frac{1}{2}\left[\frac{1}{\sqrt{R}}\mathbf{V}+\sqrt{R}\mathbf{I}\right],\ \mathbf{Y} = \frac{1}{2}\left[\frac{1}{\sqrt{R}}\mathbf{V}-\sqrt{R}\mathbf{I}\right].$$

Since for the bipole the well known constraint $\mathbf{V} = Z\,\mathbf{I}$ is valid, from the previous expressions the following relations are obtained:

$$[5.27]\ \mathbf{Y} = \frac{1}{2}\left[\frac{1}{\sqrt{R}}Z\mathbf{I}-\sqrt{R}\mathbf{I}\right] = \rho\mathbf{X} = \rho\,\frac{1}{2}\left[\frac{1}{\sqrt{R}}Z\mathbf{I}+\sqrt{R}\mathbf{I}\right];$$

thanks to the linearity, the following relation between the reflection coefficient and the impedence of the user bipole over the whole frequency domain results to be valid:

$$[5.28]\ \rho(f) = \frac{Z(f)-R}{Z(f)+R}.$$

Considering that the bipole is assumed to be passive, the real part of Z(f) is positive; as a consequence the modulus of the reflection coefficient results to be limited:

$$[5.29]\ |\rho(f)| \le 1,$$

reaching the value 1 only for $\Re\{Z(f)\}=0$.

5.2.3 Further details on the reflection coefficient

Considering the case of energy signals, recalling the relation valid between the cross correlation function and the convolution (see [3.99]) the following relation is achieved:

$$[5.30]\ C_{yx}(\tau) = y(\tau)*x^*(-\tau) = x(\tau)*h_\rho(\tau)*x^*(-\tau) = h_\rho(\tau)*C_{xx}(\tau);$$

in the case of power signals a similar expression in then obtained. Unless the ideal or perfcet matching is ensured, the reflected response y(t) results thus to be affine to the direct input signal, depending on this latter and on the reflection coefficient of the bipole.
Passing to the frequency domain, from the [5.30] and from the similar expression the following relations are derived:

$$[5.31]\ E_{yx}(f) = \rho(f)\,E_{xx}(f),\ W_{yx}(f) = \rho(f)\,W_{xx}(f),$$

where with E(f) or W(f) the energy or power spectral density of the considered signals are respectively indicated. From the relation [5.21] the following relations are then carried out:

$$[5.32]\ E_{yy}(f) = |\rho(f)|^2 E_{xx}(f),\ W_{yy}(f) = |\rho(f)|^2 W_{xx}(f)\ ;$$

obtaining in this way the reflected energy or power:

$$[5.33]\ E_{yy} = \int \left|\rho(f)\right|^2 E_{xx}(f)\,df\ ,\ W_{yy} = \int \left|\rho(f)\right|^2 W_{xx}(f)\,df\ ,$$

reminding that it is the one from the user to the generator.
In the case of physical feasibility, the causality implies that $h_0(t)$ is right monolateral, allowing to use the expression $h_0(t)=h_0(t)sgn(t)$; applying the Fourier transform, being $F\{sgn(t)\} = -j/\pi f$, the following relation is obtained:

$$[5.34]\ \rho(f) = \rho(f) * (-j/\pi f).$$

Thus, making equal respectively the real and the imaginery parts the result is:

$$[5.35]\ \rho_R(f) = H\{\rho_I(f)\} = \hat{\rho}_I(f),\ \rho_I(f) = -H\{\rho_R(f)\} = -\hat{\rho}_R(f)\ ,$$

i.e. the real part $\rho_R(f)$ of $\rho(f)$ and the coefficient of its imaginery part $\rho_I(f)$ are related between each other through the Hilbert transform in the frequency domain.

5.3 LTI TRANSFORMATIONS IN THE QUADRUPOLES

5.3.1 Responses in the time and frequency domains

The transformations between signals occurring in an electrical subsystem with two access sections, or ***ports***, are very important becasue they are representative of transit equipment, that in general in a network are more than the bipole elements. The scheme is shown in Figura 5.7, where the direct and reflected signals on the access ways at the ports have been utilized; the notations $x_i(t)$ for the excitation input signal at the port i=1,2 and $y_i(t)$ for the corresponding output response have been used. It is well known that a subsystem with two ports, on the basis of the possibility to represent it with an equivalent quadrupole, often is named ***quadrupole*** even if the identification of a real pair of access poles has not physical meaning.

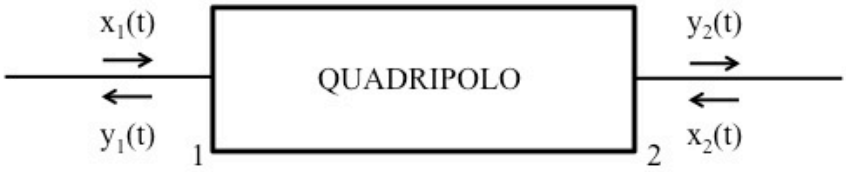


Figura 5.7: Scheme of a quadrupole, with a pair of input, or excitation, signals $x_i(t)$, and a pair of output, or response, signals $y_i(t)$.

In the most general case the two transformations which lead to the responses $y_1(t)$ and $y_2(t)$ have quite complex expressions and depend both on the excitations $x_1(t)$ and $x_2(t)$, as well as on the independent time variable.
Assuming the restrictive hypotesis, but very frequently occurring in practical systems, that the quadrupole is linear and characterized with fixed parameters in the time domain (time invariance), the contraints imposed by the particular implementation of the systems between the signals at the two ports lead to the pair of LTI transformations, usually of type with memory:

$$[5.36]\ y_i(t) = \int x_i(v)\,h_{ii}(t-v)\,dv + \int x_j(v)\,h_{ij}(t-v)\,dv\ ,\ i=1,\ 2\ e\ j=2,\ 1,$$

which express the output signals, or responses, as a function of the two input excitation signals. Note that in the convolution integrals the independent input time variable was indicated with v in order to distinguish it with respect to the output one t.

The four functions $h_{ij}(t)$, which characterize in the time domain a generic ***LTI quadrupole*** with reference to a resistance R, are named ***reflected pulse responses*** (for i=j) and ***trasmitted pulse responses*** (for i≠j), considering that they are the same output signals achievable when a single excitation corresponding to the ideal pulse at the instant v=0 emitted by a generator with internal impedence R reale and constant; in fact, $y_1(t)\equiv h_{11}(t)$ and $y_2(t)\equiv h_{21}(t)$ is obtained assuming the excitations $x_1(v)\equiv\delta(v)$ and $x_2(v)\equiv 0$, and $y_1(t)\equiv h_{12}(t)$ and $y_2(t)\equiv h_{22}(t)$ considering $x_1(v)\equiv 0$ and $x_2(v)\equiv\delta(v)$.
A LTI quadrupole, which is characterized by the two transformations [5.36] that are hereafter expressed with the two equivalent notations:

$$[5.37]\ y_1(t) = h_{11}(t)*x_1(t) + h_{12}(t)*x_2(t),\ y_2(t) = h_{21}(t)*x_1(t) + h_{22}(t)*x_2(t),$$

is ***physically feasible*** if all the four pulse responses which characterize it in the time domain belong to the functional space L_1, are real and moreover ***causal***, i.e. $h_{ij}(t)=0\ \forall t<0$; instead it is just ***ideally feasible*** if the causality of the responses is not verified. Thus, for a LTI quadrupole, both ideallly and physically feasible, the Fourier transforms exist:

$$[5.38]\ H_{ij}(f) = F\{h_{ij}(t)\},$$

that also are characteristic functions of the considered quadrupole with reference to the resistance R, but in the frequency domain. Assuming that $h_{ij}(t)$ are real the property:

$$[5.39]\ H_{ij}(f) = H_{ij}^{*}(-f),$$

is valid, i.e. the functions $H_{ij}(f)$ are hermitian, with even modulus and odd argument.
Reminding the definition, for every harmonic frequency, of the ***diffusion parameters*** S_{ij} of a LTI quadrupole (the reader can anyway refer to section 5.3.2), the result is that the ***reflectances***, that are the parameters $S_{11}(f)$ and $S_{22}(f)$, are identified with the Fourier transforms of $H_{11}(f)$ and $H_{22}(f)$ of the reflected pulse responses $h_{11}(t)$ and $h_{22}(t)$, while the ***transmittances***, i.e. the parameters $S_{12}(f)$ and $S_{21}(f)$, are identified with the Fourier transforms $H_{12}(f)$ and $H_{21}(f)$ of the transmitted pulse responses $h_{12}(t)$ and $h_{21}(t)$.
Once the transmittance of a LTI quadrupole is expressed through the respective modulus $g_{ij}(f) = |H_{ij}(f)|$ and the opposite of the argument $\Gamma_{ij}(f) = -\arg\{H_{ij}(f)\}$, so that for i≠j the result is:

$$[5.40]\ H_{ij}(f) = g_{ij}(f)\, e^{-j\Gamma_{ij}(f)},$$

the following real quantities are defined:

$$[5.41]\ G_{ij}(f) = g_{ij}^{2}(f) = |H_{ij}(f)|^2,$$

$$[5.42]\ t_{0ij}(f) = \frac{1}{2\pi}\frac{d\Gamma_{ij}(f)}{df} = -\frac{1}{2\pi}\frac{d}{df}\left[\arg\left\{H_{ij}(f)\right\}\right],$$

respectively named ***insertion gain*** and ***insertion group delay*** from the port j to the port i. In a feasible quadrupole the function $G_{ij}(f)$ is limited everywhere; if it is also physically feasible the function $t_{0ij}(f)$ is non negative everywhere; considering that $\Gamma_{ij}(0)=0$, the function $\Gamma_{ij}(f)$ is then positive (or negative) for any positive (or negative) value of the frequency.
Assuming that the Fourier transforms $X_i(f)$ of the excitation signals $x_i(t)$ of a LTI quadrupole exist, applying the time convolution property from [5.37] it is possible to obtain the corresponding relations in the frequency domain:

$$[5.43]\ Y_1(f) = H_{11}(f)\,X_1(f) + H_{12}(f)\,X_2(f),\ Y_2(f) = H_{21}(f)\,X_1(f) + H_{22}(f)\,X_2(f).$$

5.3.2 Diffusion parameters in LTI quadrupoles

The behaviour of a LTI quadrupole can be described, for any frequency $f=\omega/2\pi$, considering the simple harmonic excitations at its ports 1 and 2:

$$[5.44]\ x_1(t) = A\cos(\omega t+\alpha_1),\ x_2(t) = A_2\cos(\omega t+\alpha_2),$$

which produce the output signals, also them harmonics:

$$[5.45]\ y_1(t) = B_1\cos(\omega t+\beta_1),\ y_2(t) = B_2\cos(\omega t+\beta_2).$$

Through the Steinmetz notation usually the previous harmonic signals are indicated with phasors, named wave intensity when referred to the signals, at the input at frequency f_a:

$$[5.46]\ \mathbf{a}_1 = A_1 e^{j\alpha_1},\ \mathbf{a}_2 = A_2 e^{j\alpha_2},$$

and at the output at frequency f_b:

$$[5.47]\ \mathbf{b}_1 = B_1 e^{j\beta_1},\ \mathbf{b}_2 = B_2 e^{j\beta_2}.$$

Having assumned that the quadrupoles are LTI and that the same reference resistance R is present at the two ports, the constraints imposed on the wave intensities take the very simple forms:

$$[5.48]\ \mathbf{b}_1 = S_{11}\mathbf{a}_1+S_{12}\mathbf{a}_2,\ \mathbf{b}_2 = S_{21}\mathbf{a}_1+S_{22}\mathbf{a}_2,$$

where the adimensional quantities S_{ij}, characteristic of the quadrupole, are named ***diffusion parameters***. The parameters with identical pair of subscripts and those with difefrent subscripts are named respectively the ***reflectances*** and the ***transmittances*** of the quadrupole. Switching to consider the functions $S_{ij}(f)$ defined for any frequency and taking into account the linearity property of the Fourier tranform, comparing the previous relations with the [5.43] the result is that the diffusion functions correpond to the respective Fourier transforms of the pulse response of the LTI quadrupole.
The feasibility condition, either physical or ideal, of the quadrupole is carried out in the form:

$$[5.49]\ S_{ij}(f) = S_{ij}^{*}(-f);$$

i.e. the diffusion parameters must be hermitian functions. Then, all of them have limited modulus over the whole frequency axis; the upper bound limit is 1 in the case of passive quadrupoles. Going forward as for the bipole, in the case of physical feasibility it is possible to demonstrate that the real part of every function $S_{ij}(f)$ and the coefficient of its imaginery part are related between each other by the Hilbert transform in the frequency domain:

$$[5.50]\ S_{ijR}(f) = H\{S_{ijI}(f)\} = \hat{S}_{ijI}(f),\ S_{ijI}(f) = -H\{S_{ijR}(f)\} = -\hat{S}_{ijR}(f).$$

The reciprocity of the quadrupole implies that the transmittances are equal:

$$[5.51]\ S_{12}(f) = S_{21}(f),$$

while the electrical symmetry implies that the following relations are valid:

$$[5.52]\ S_{12}(f) = S_{21}(f),\ S_{11}(f) = S_{22}(f),$$

i.e. the equality of the reflectances are added to that of the transmittances.
A quadrupole for which the reflectances are zero:

$$[5.53]\ S_{11}(f) = S_{22}(f) = 0,$$

is named ***matched quadrupole***; if it is also reciprocal, then it is also symmetrical.
For a passive quadrupole with no losses, the following relations are valid:

$$[5.54]\ |S_{11}(f)|^2 + |S_{21}(f)|^2 = |S_{22}(f)|^2 + |S_{12}(f)|^2 = 1,$$

$$[5.55]\ S_{11}(f)\ S_{12}^{*}(f) + S_{21}(f)\ S_{22}^{*}(f) = 0.$$

Still for any frequency the determinant is also defined:

$$[5.56]\ \Delta(f) = S_{11}(f)\,S_{22}(f) - S_{12}(f)\,S_{21}(f).$$

In the case it is reciprocal, passive and lossless the modulus $|\Delta(f)| = 1$; then, it is $\Delta(f) = -1$ if the quadrupole is reduced to the simple direct connection between the two ports, in which the reflectances are zero and the transmittamces are equal to 1.

5.3.3 Transfer in matching conditions

5.3.3.1 Transfer in a quadrupole

Skipping the analysis of the general case, let's assume that the port 2 of a LTI quadrupole is connected to a passive linear bipole perfectly matched, i.e. having a reflection coefficient $\rho(f)$ equal to zero in the bandwidth of the signal; then, since the reflected signal from the bipole results to be identically equal to zero, and it is also the excitation signal of the quadrupole at the port 2, the pair of transformations which describes the behaviour of the latter is simplified in the:

$$[5.57]\ y_1(t) = h_{11}(t) * x_1(t), \text{ per } \rho(f)=0\ \ \forall |f| \in (f_m; f_M),$$

$$[5.58]\ y_2(t) = h_{21}(t) * x_1(t), \text{ per } \rho(f)=0\ \ \forall |f| \in (f_m; f_M).$$

Disregarding the reflected response at the port 1 (such a signal is very often identically equal to zero because the reflectance $H_{11}(f)$ is zero in the band of the signal), the attention is concentrated just on the transformation (see scheme in Figura 5.8):

$$[5.59]\ y(t) = h(t) * x(t),$$

where $y(t)=y_2(t)$ has been expressed as the only meaningful response, i.e. the one at the output from the port 2, $x(t)=x_1(t)$ has been expressed as the only excitation signal, at the input to the port 1, and finally the following notation:

$$[5.60]\ h(t)=h_{21}(t), \text{ per } \rho(f)=0\ \ \forall |f| \in (f_m; f_M),$$

has been used for the only characteristic function of the considered system in the time domain, i.e. the ***transfer pulse response.*** Note that the [5.59] is still the only interesting transformation even in the case there is no user bipole with perfect matching, but at the condition that $H_{22}(f)=0$ within the bandwidth of the signal.

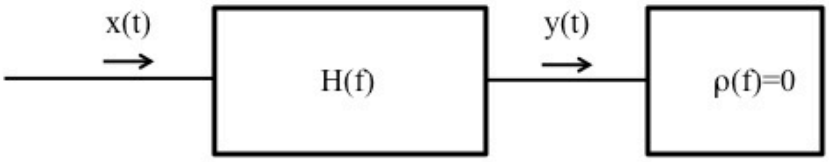


Figura 5.8: Scheme of a LTI quadrupole with perfect matching in output.

It is usual to name ***transfer function*** or ***frequency response*** of the quadrupole with perfect matching in output, either from the user side or from the quadrupole side, the Fourier transform of the transfer pulse response:

$$[5.61]\ H(f)=F\{h(t)\} = H_{21}(f),$$

i.e. the transmittance from the input port to the output port, that is also the only meaningful characteristic function of the system in the considered conditions, but in the frequency domain. Considering the modulus and the opposite of the argument of the transfer function, for which $H(f)=g(f)\, e^{-j_r(f)}$ is valid, the real quantities:

$$[5.62]\ G(f)=g^2(f) = |H(f)|^2,$$

$$[5.63]\ t_0(f)= \frac{1}{2\pi}\frac{d\Gamma(f)}{df} = -\frac{1}{2\pi}\frac{d}{df}\left[\arg\{H(f)\}\right],$$

hereinafter are simply named ***gain*** and ***group delay*** of the quadrupole; also the inverse of the gain, i.e. the transfer ***attenuation*** of the quadrupole:

$$[5.64]\ A(f)= g^{-2}(f) = |H(f)|^{-2}.$$

If the Fourier transform X(f) of the signal x(t) exists, for the property of convolution in the time domain from the [5.59] the corresponding relation in the frequency domain is obtained:

$$[5.65]\ Y(f) = H(f)\, X(f).$$

5.3.3.2 Transfer in quadrupoles in cascade

Let's consider the cascade of two LTI quadrupoles Q_1 and Q_2, with the notation shown in Figura 5.9.

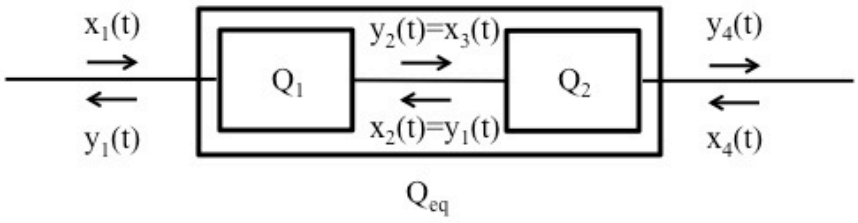


Figura 5.9: Equivalent quadrupole to the cascade of two quadrupoles.

The identities $y_2(t)\equiv x_3(t)$ and $x_2(t)\equiv y_3(t)$ on the four signals at the internal ports shown in Figura 5.9, which result to be imposed when the connection of the two quadrupoles is implemented, reduce in general the global behaviour of the cascaded system to just two transfromations, characteristic of the only quadrupole Q_{eq} equivalent to the cascade of Q_1 and Q_2 which has the external ports of the single quadrupoles as accessible ports.
In the hypotesis that Q_2 is terminated on a bipole with perfect matching and that both quadrupoles are matched, implying that the reflectances are zero within the band of the signal, all the reflected signals are identically equal to zero, implying that the following identities are valid:

$$[5.66]\ x_4(t) \equiv 0,\ x_2(t) \equiv y_3(t) \equiv 0,\ y_1(t) \equiv 0.$$

Indicating with $h_i(t)$ the transfer pulse response of the quadrupole Q and having posed $x(t)=x_1(t)$ and $y(t)=y_4(t)$, it is possible to obtain:

$$[5.67]\ y_2(t) = h_1(t)*x(t),\ y(t) = h_2(t)*x_3(t)\ ;$$

on the basis of the identity $y_2(t)\equiv x_3(t)$ then the result is:

$$[5.68]\ y(t) = h_2(t)*x_3(t) = h_2(t)*[h_1(t)*x(t)] = h_{eq}(t)*x(t),$$

where the transfer pulse response of the equivalent quadrupole has been expressed as:

$$[5.69]\ h_{eq}(t)=h_2(t)*h_1(t).$$

Applying the Fourier transform to both members of the previous expression, for the property of the convolution in the time domain the following simple relation beteween the transfer function of the quadrupole equivalent to the cascade and those of the component quadrupoles is achieved:

$$[5.70]\ H_{eq}(f) = H_1(f)\, H_2(f).$$

It is important to note that exchanging the positions of the two LTI quadrupoles in cascade the whole behaviour doesn't change.
Considering the casccade of m LTI quadrupoles, always in the condition of perfect matching in all the sections of the connection, the [5.70] can be generalised in the following expression:

$$[5.71]\ H_{eq}(f) = \prod_{i=1}^{m} H_i(f).$$

from which it is evident that the whole cascade is equivalent to a quadrupole with perfect matching and that transfer function is obtained as the simple product of those ones of the component quadrupoles.

5.3.3.3 Transfer in shifted band in a quadrupole

In the case a LTI quadrupole, either ideally or physically feasible, is utilised with high frequency signals, usually represented through the complex envelopes referred to a frequency $f_c=\omega_c/2\pi$ within the band, assuming that the evolution in the output connection section is perfect, often the characterization by means of the function named ***base band equivalent transfer function*** is preferred:

$$[5.72]\ \underline{H}(f)=H_+(f+f_c).$$

As shown in the example in Figura 5.10, the function $\underline{H}(f)$ is obtained simply shifting $H_+(f)$, i.e. just the part on the positive half frequency axis of the transfer function $H(f)=H_{21}(f)$, of f_c towards the origin.

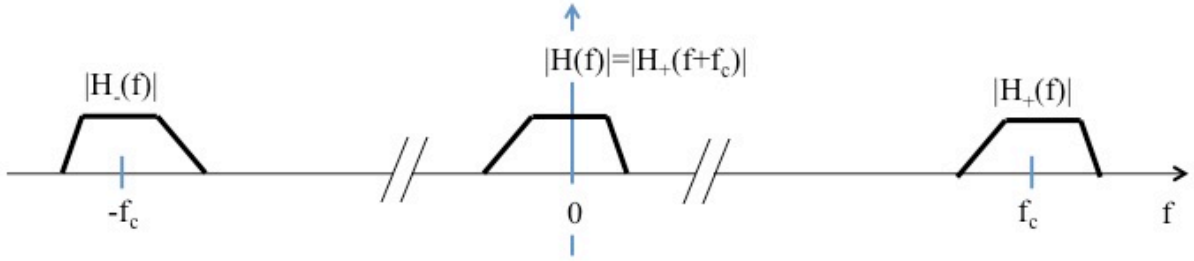


Figura 5.10: Equivalent base band transfer function of a LTI quadrupole with perfect matching in output used with shifted band signal.

Note that in general the function $\underline{H}(f)$ is not hermitian, that means:

$$[5.73]\ \underline{H}(f)\neq\underline{H}^*(-f);$$

as a consequence usually the equivalent base band pulse response is complex:

$$[5.74]\ \underline{h}(t)=F^{-1}\{\underline{H}(f)\} = h_c(t) + jh_s(t).$$

Nevertheless, in practice the case of quadrupoles for which, at least approximately, the property $\underline{H}(f)=\underline{H}^*(-f)$ is valid is frequent; then, $\underline{h}(t)$ results to be real, implying that $\underline{h}(t)\equiv h_c(t)$ and $h_s(t)\equiv 0$. Inviting the reader interested to the demonstration to see the section 5.3.3.4, the equivalent base band pulse response allows to carry out the complex envelope $\iota_y(t)$ of the response from that $\iota_x(t)$ of excitation by means of simple transformation:

$$[5.75]\ \iota_y(t) = \underline{h}(t)*\iota_x(t).$$

Assuming that the Fourier transform $I_x(f)$ of $\iota_x(t)$ exists, as usual it is possible to obtain the corresponding simple relation in the frequency domain:

$$[5.76]\ I_y(f)=\underline{H}(f)\, I_x(f)$$

5.3.3.4 Details on the transfer in shifted band

The equivalent base band transfer function can be broken up in its hermitian and non hermitian components, $\underline{H}(f) = \underline{H}_H(f)+\underline{H}_{AH}(f)$, for which:

$$[5.77]\ \underline{H}_H(f) = \frac{1}{2}\left[\underline{H}(f)+\underline{H}^*(-f)\right],\ \underline{H}_{AH}(f) = \frac{1}{2}\left[\underline{H}(f)-\underline{H}^*(-f)\right].$$

Then, performing the inverse transfrom it is possible to achieve:

$$[5.78]\ F^{-1}\{\underline{H}_H(f)\} = \frac{1}{2}\left[\underline{h}(t)+\underline{h}^*(t)\right] = h_c(t),$$

$$[5.79]\ F^{-1}\{\underline{H}_{AH}(f)\} = \frac{1}{2}\left[\underline{h}(t)-\underline{h}^*(t)\right] = jh_s(t),$$

from which also the following expressions of the real part and of the coefficient of the imaginery part of the base band equivalent pulse response can be achieved:

$$[5.80]\ h_c(t) = \frac{1}{2}F^{-1}\left\{\underline{H}(f)+\underline{H}^*(-f)\right\},\ h_s(t) = \frac{1}{j2}F^{-1}\left\{\underline{H}(f)-\underline{H}^*(-f)\right\}.$$

Considering the analitycal signals at the two ports (see [3.157]):

$$[5.81]\ a_x(t) = \iota_x(t)\ e^{j\omega_c t},\ a_y(t) = \iota_y(t)\ e^{j\omega_c t},$$

instead of the transformation between the real signals $y(t)=h(t)*x(t)$, the transformation addressing only the positive frequency half axis is carried out:

$$[5.82]\ a_y(t) = h_+(t)*a_x(t),$$

where

$$[5.83]\ h_+(t) = F^{-1}\{H_+(f)\} = \underline{h}(t)\ e^{j\omega_c t}\ ;$$

then, it is possible to achieve:

$$[5.84]\ \iota_y(t)\ e^{j\omega_c t} = \int \iota_x(v)h_+(t-v)e^{j\omega_c v}\,dv,$$

i.e.

$$[5.85]\ \iota_y(t) = \int \iota_x(v)h_+(t-v)e^{-j\omega_c(t-v)}\,dv,$$

which taking into account the [5.83] leads to the transformation [5.75] between the complex envelopes.

5.4 IDEAL QUADRUPOLE AND PERFECT QUADRUPOLES

5.4.1 Ideal quadrupole

Reminding that the ideal transfer is achieved when the condition of faithful signal [1.1] is verified, on the basis of what presented in the previous sections, considering a signal passing through a telecommunication system assimilated to a LTI system, it is possible to verify that such a condition is carried out only if:

$$[5.86]\ h_0(t) = g\,\delta(t-t_0) = F^{-1}\{g\,e^{-j2\pi f t_0}\},$$

In fact:

$$[5.87]\ y(t) = g\,x(t-t_0) = h_0(t)*x(t) = g\delta(t-t_0)*x(t) = F^{-1}\{g\,e^{-j2\pi f t_0}\}*x(t).$$

This expression characteristic of $h_0(t)$ implyes that the quadrupole has the reflectances at the two ports and the transmittance from the port 2 to the port 1 all identically zero (i.e. with $H_{110}(f)\equiv 0$, $H_{120}(f)\equiv 0$ and $H_{220}(f)\equiv 0$) and that moreover has from the port 1 to the port 2 the transmittance, and thus the transfer function, in the form:

$$[5.88]\ H_{210}(f) = H_0(f) = g\,e^{-j2\pi f t_0},$$

as usual with g and t_0 real constants.
In such a way, for any excitation signal x(t) and generic behaviour of the user bipole the quadrupole would provide on the basis of the convolution [5.59] the faithful response.
The considered block, satisfactory if aiming to the transmission for any condition at the ports and also for signals with unlimited bands, cannot represent any system really built and as such is named ***ideal quadrupole***. The inconvenience of the impossibility to being able to approximate its ideal behaviour with actual physical systems is nevertheless drastically scaled down if, as usual in practical transmission systems, signals with limited, or at least practically limited, monolateral spectrum within the two finite extremes, f_m e f_M of the monolateral band are considered. In the

following sections the quadrupoles capable to provide responses at least practically faithful to excitation signals first of base band type and then of shifted band type are just considered.

5.4.2 Perfect quadrupoles

5.4.2.1 Base band perfect quadrupole

Assuming that both the actual two port system and the passive user bipole are linear and time invariant, if the excitation signal x(t) is in base band, with extremes of the monolateral bandwidth satisfying the condition $f_M >> f_m$, the spectral components of the signals at the two ports are zero for any |f| out of the interval (f_m, f_M); thus, the evolutions of the reflectances and of the transmittances of the representative quadrupole at the frequencies not interested by the excitation cannot have any influence on the behaviour. The relations:

$$[5.89]\ H_{11}(f) = H_{22}(f) = H_{12}(f) = 0,\ H_{21}(f) = g\, e^{-j2\pi f t_0},\ \forall\ |f| \in (f_m; f_M),$$

corresponding to those that define the ideal quadrupole, but only in the intested band, are then sufficient to obtain the ideal transmission from the port 1 to the port 2, without reflected signals and the desired faithful response $y(t)=g\,x(t-t_0)$.
Summarizing, a two port system utilized for the transmission of base band signals is capable to allow the ideal transmission, from its input port 1 to the output one 2, when it can be represented by a LTI quadrupole which has in the useful band (f_m, f_M):

1) reflectances $H_{11}(f)$ and $H_{22}(f)$ and transmittance $H_{12}(f)$ equal to zero,
2) transfer function $H(f)=H_{21}(f)$ with
   - constant modulus,
   - argument proportional to the frequency (with possible addend $-\pi$ in case $g<0$).

The condition on the reflectances is named ***perfect matching condition*** of the quadrupole. The conditions on the transfer function, i.e.:

$$[5.90]\ |H(f)| = |g|,\ \arg\{H(f)\} = -2\pi f t_0 + \frac{\pi}{2}\left[\mathrm{sgn}(g) - 1\right],\ \forall\ |f| \in (f_m; f_M),$$

are named ***non distorsion in base band conditions*** of the quadrupole, considering that if they are not respected the transferred base band signal is not faithful, i.e. the time invariant ***linear distorsion*** of the response occurs.
If all the above listed conditions are satisfied the quadrupole is ***perfect in base band***. Note that the perfect evolutions of the functions $H_{ij}(f)$, required only within the useful limited band, can be well approximated in practice by those ones of the characteristic parameters of the representative quadrupole of a real system operating in base band, even if with increasing difficulties with greater f_M in the event $f_m=0$ or, differently, with a greater ratio f_M/f_m.
Seldom, in condition of full matching in the connection sections, the two way transfer is interesting, i.e. that the pair of direct and reflected signals result separately faithful, but fully deccoupled because vehicle of two distinct information that travel in opposite directions; then, it is immediate to add the further electrical symmetry condition, obtaining in this way a symmetrical perfect base band quadrupole.
The properties [5.90] imply that both the gain function and the group delay function are constant within the band (see [5.62] and [5.63]):

$$[5.91]\ G(f) = g^2,\ t_0(f) = t_0,\ \forall\ |f| \in (f_m; f_M),$$

Definitively, from the transmission point of view a perfect quadrupole in base band is characterised in the useful band by just two quantities, both real and positive: the gain $G=g^2$ and the group delay t_0.

5.4.2.2 Perfect quadrupole in shifted band

In the case of a LTI quadrupole which represent a two port system used to transmit a shifted band signal, with extremes f_m and f_M of the monolateral band, the considerations concerning the reflections in the connection sections are still valid: thus, aiming to the ideal transmission the same conditions relative to the perfect matching of the quadrupole within the useful band are applicable. Instead, with regard to the transfer from a port to the other of the quadrupole, it is opportune to refer to the less severe faithfulness condition:

$$[5.92]\ \iota_y(t) = g\ e^{-j\Gamma}\ \iota_x(t-t_0),$$

where Γ is an arbitrary real constant, expressed on the complex envelopes $\iota_x(t)$ and $\iota_y(t)$ instead of on the corresponding real signals x(t) and y(t); if in the definition of the complex envelopes a common reference frequency f_c in input and in output is selected, the mentioned condition is in general satisfied if the behaviour of the quadrupole can be expressed through the memoryless LTI transformation, with equivalent base band pulse response $\underline{h}(t) = g e^{-j\Gamma}\delta(t-t_0)$, corresponding in the frequency domain to the ideal equivalent baseband transfer function:

$$[5.93]\ \underline{H}_0(f) = g\, e^{-j\Gamma}\ e^{-j2\pi f t_0}.$$

Considering the limitation of the excitation bandwidth and the linearity, the trasfer function of the equivalent base band quadrupole usually must satisfy just the less severe condition:

$$[5.94]\ \underline{H}(f) = g\, e^{-j(2\pi f t_0+\Gamma)},\ \forall\ f \in (f_m - f_c; f_M - f_c);$$

shifting such a function at the frequency f_c, the following relation is achieved:

$$[5.95]\ H_+(f) = \underline{H}(f-f_c) = g\, e^{-j(2\pi f t_0+\Gamma-2\pi f_c t_0)} = g\, e^{-j(2\pi f t_o+\gamma_c)},\ \forall\ f \in (f_m; f_M),$$

where $H_+(f)$ for $f \geq 0$ is exactly the trasmittance $H_{21}(f)$ of the quadrupole in shifted band and $\gamma_c = \Gamma - 2\pi f_c t_0$ assumes an arbitrary value, being Γ arbitrary as well.
Summarizing, a system with two ports used for the transmission in shifted bandwidth can allow the ideal transmission from the port 1 to the port 2, with no change of the reference frequency in the representations with the complex envelopes, when it can be represented by a LTI quadrupole which in the useful bandwidth (f_m ; f_M) has:

1) reflectances $H_{11}(f)$ and $H_{22}(f)$ and transmittance $H_{12}(f)$ equal to zero,
2) transfer function $H(f)=H_{21}(f)$ with
   - constant modulus,
   - argument proportional to the frequency but with an arbitrary addend (thus linear).

The condition on the reflectances is stil the ***perfect matching condition*** of the quadrupole. The conditions on the transfer function, i.e.:

$$[5.96]\ |H(f)| = |g|,\ \arg\{H(f)\} = -2\pi(f-f_c)t_0 - \Gamma,\ \forall\ |f| \in (f_m; f_M),$$

are named ***non distorsion condition in shifted band*** of the quadrupole, considering that should them not being respected the signal in shifted band is not faithful, implying that the time invariant ***linear distorsion*** of the response occurs.
If all the above listed conditions are satisfied the quadrupole is ***perfect in shifted band***. It is important to note that the perfect evolution of the functions $H_{ij}(f)$ required just within the useful limited band can be well approximated in practice by those ones of the characteristic parameters of the representative quadrupole of a real system; with respect to the base band case, the presence in the second of the [5.96] of an arbitrary Γ makes the approximation of arg{H(f) easier and, moreover, all the difficulties are greatly reduced in the case, very frequent, of signals with very narrow relative bandwidth for which $B \ll f_a$.
Seldom, if full matching in the connection sections is realized, the two way transfer is interesting, i.e. the pair of direct and reflected complex envelopes are required to be separately faithful, but

fully decoupled because they are vehicles of two different information travelling towards opposite directions; then, it is immediate to add the further condition of electrical symmetry, obtaining a perfect symmetrical LTI quadrupole in shifted band.
The non distorsion conditions for a LTI quadrupole in shifted band are still the constant gain and group delay within the bandwidth.

5.4.2.3 Perfect non linear quadrupoles in narrow relative band

Still regarding the transmission of signals in shifted band, but only if with narrow relative bandwidth, let's consider the particular case of a subsystem, named ***frequency converter***, with the behaviour described hereinafter. The reflectances at both ports of the representative quadrupole are equal to zero in the band of the signal; the transfer from one port to the other can be expressed through a linear time invariant relation between the complex envelopes of the actual real signals y(t) and x(t) at the output and at the input, of the type:

$$[5.97]\ \iota_y(t) = \iota_x(t) * \underline{h}_C(t);$$

finally let's define these signals with reference to two different values of the frequency, indicated with f_{ce} and f_{cu} (see Figura 5.11). A converter, thus, shows a non linear behaviour with respect to the real signals at the ports, but it is instead linear, other than time invariant, if the corresponding complex envelopes are considered.

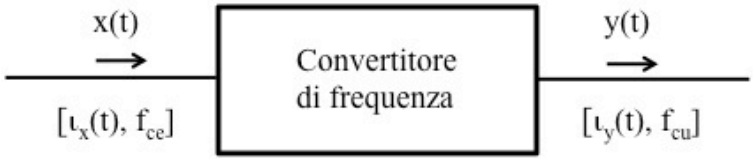


Figura 5.11: Frequency converter used with narrow relative bandwidth signals, with different reference frequencies at the input and at the output.

It is possible to observe that, because of the linearity of the [5.97], the two real signals at the input and at the output of a frequency converter have the same bandwidth B, as shown in the example in Figura 5.12 which is referred to energy signals.

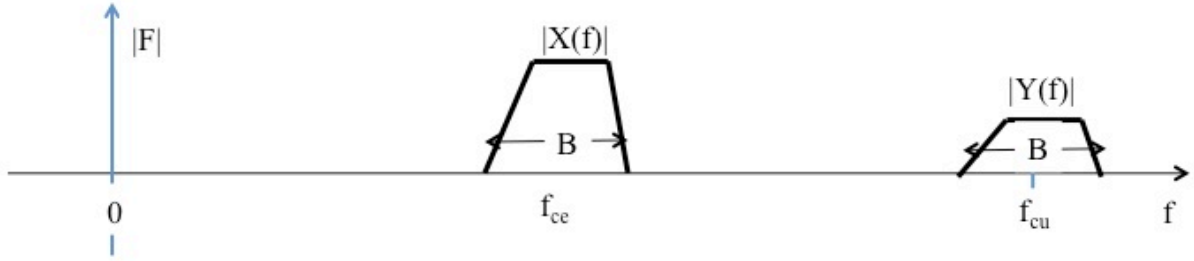


Figura 5.12: Example of amplitude spectra with narrow relative bandwidth at the input and at the output of a frequency converter.

Assuming the unessential hypotesis that the reference frequencies are chosen at the center of the respective spectra, a frequency converter, although it is surely non linear because at the output there are spectral components not present at the input, can allow the ideal transmission in narrow relative band only if the spectra of the input and of the output signals result to be separated, that means that the condition $|f_{ce}-f_{cu}|>B$ is satisfied and that the equivalent transfer function in base band of the representative quadrupole, obtained applying the Fourier transform to $\underline{h}_C(t)$, satisfies the condition similar to the [5.94]:

$$[5.98]\ \underline{H}_C(f) = g_C\, e^{-j(2\pi f t_o + \Gamma)},\ \forall\ f \in (-B/2;B/2).$$

In the case in which $f_{cu}>f_{ce}$ the quadrupole is a ***perfect up converter*** (UC=Up Converter), if instead $f_{cu}<f_{ce}$ the quadrupole is a ***perfect down converter*** (DC=Down Converter). Such a category of perfect quadrupoles cannot experience the symmetrical electrical behaviour.

As usual, summarizing, a two port system, operating on signals characterized by narrow relative bandwidth with meaningful variation of the reference frequency of the complex envelope at the output with respect to the input one can allow the ideal one way transmission, if a non linear time invariant quadrupole representative of its characteristics exists with a transfer function equivalent in base band $\underline{H}_C(f)$ and if, in the useful bands separated on the frequency axis, the following conditions are respected:

1) reflectances $H_{11}(f)$ and $H_{22}(f)$ equal to zero,
2) transfer function equivalent in base band $\underline{H}_C(f)$ with
    - constant modulus,
    - argument proportional to the frequency but with an arbitrary addend.

Applying the well known expressions [5.62] and [5.63] to the function $\underline{H}_C(f)$ the converter gain, $G_C(f)$, and its group delay, $t_{0C}(f)$ can be defined; then, the non distorsion conditions for such a quadrupole imply that the gain, $G_C = g_C^2$, is constant within the band as well as its group delay, t_{0C}.

6 FUNDAMENTALS OF TRANSMISSION

6.1 IDEAL TRANSMISSION

6.1.1 Conditions for the ideal transport of the information

The electrical transfer of the information in a link from a source to a destination separated by a meaningful distance can be sketched in elementary way as shown in Figura 6.1, where between the generator G and the user U a generic quadrupole has been inserted to represent, as synthetically as possible, a ***transmission system*** of the electrical signals. At the ends of the transmission system the pairs of excitation signals $x_1(t)$ and $x_2(t)$, and of **response,** $y_1(t)$ and $y_2(t)$ are assumed as electrical quantities. The signals can be either of power or energy type, but real in any case.

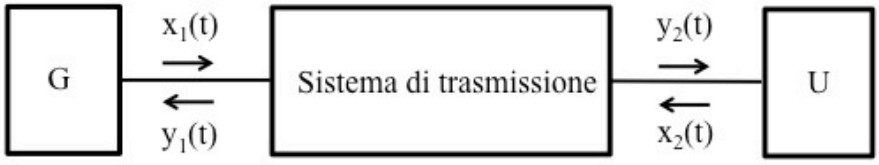


Figura 6.1: Synthetic scheme of a link.

Recalling the above carried out remarks and supposing that the transmission occurs only in one way, the situation is optimal if:

- in the connection sections the ideal matching is verified, i.e. the reflected signals are identically equal to zero:

$$[6.1]\ y_1(t) \equiv 0,$$

$$[6.2]\ x_2(t) \equiv 0.$$

- the two direct signals result to be faithful between each other (with maximum affinity in broad sense, see [2.10] and [2.93]), implying that the following relation is valid:

$$[6.3]\ y(t) = gx(t - t_0),$$

where for sake of simplicity $x(t) \hat{=} x_1(t)$ and $y(t) \hat{=} y_2(t)$ have been set.
It is important to remind that in the [6.3] g is an arbitrary, real and constant, factor and t_0 is an arbitrary real, constant and positive time shift; the former alteration can be considered with no impact, also because it can be easily compensated if the disturbs are not taken into account; the latter, as long as its value is limited within opportune limits depending on the type of information, must be accepted taking into account the unavoidable delay that the signals exprerience in the transmission, due to the covered distance not negligible in relation to the finite, although very large, propagation speed.
Assuming that x(t) is a signal which correctly transmits the information, it is clear that the three relations [6.1], [6.2] are [6.3] constitue a set of ***sufficient conditions for the optimal transport*** of the information from the generator to the user. In fact, the first two guarantee the maximum utilization of the instant power that the generator and the transmission system make available in the respective connection sections, while the third implies that the direct signal going out arrives at the user in the desired form. Disregarding the possible existence of the reflected signal $y_1(t)$, just the [6.2] and [6.3] represent the ***sufficient conditions for the ideal transfer***; instead, focusing just on the performance that are exclusive responsibility of just the transmission system, even in presence of reflection by the user bipole, just the [6.1] and [6.3] represent the ***sufficient conditions for the ideal transmission***.
Note that the relation [6.3], which concerns faithfulness, remains valid with the excitation signal multiplied by any constant and generically time shifted, i.e. it assumes that the transmission system is linear and time invariant (LTI). Moreover, if the generator is switched off. i.e. $x(t) \equiv 0$, on the basis of the presented criteria all the signals at the ports of the quadrupole must result identically equal to zero; then, it is deduced that the conditions for the ideal transfer or

transmission imply that the transmission system is also with no disturbing waveform having origin independent on the signal.
In this section the attention is focused on the signals at the ends of a link, up to determining the properties that they must have to guarantee the optimum performance to transport information. Hereinafter the implications that the identified conditions entail on the characteristics of the main subsystems which constitute with their cascade the transmission system are carried out.

6.1.2 Perfect transmission systems

6.1.2.1 Base band perfect transmission systems

In a transmission system in which the signal on the line propagating along the transmission means is of base band type, the transmitting and receiving equipment can be roughly distinguished in linear functional blocks, either of active or passive type, which together with the transmission means constitute the ***base band transmission channel*** linear as a whole, and in a pair of functional blocks which implement non linear processing; thus, the reference scheme is presented in Figura 6.2, where the ideal matching in all the connection sections is straightforward.

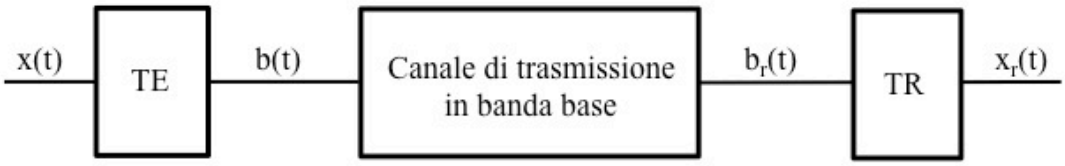


Figura 6.2: System with base band transmission channel.

The processing operations executed by the outer functional blocks, called transmitting terminal (TT) and receiving terminal (RT) are respectively named ***base band modulation*** and ***base band demodulation***.
The base band terminals, which transform the signal x(t) transferring the information on the different signal b(t) and viceversa, are often very complex in the digital case and they are worth to be inserted just for the achievable improvements on the global performance of the system, taking into account the impairments of the base band channel actually utilized.
In fact, the modulation is implemented to obtain the on line signal b(t) with time and spectral characteristics that better fit the real behaviour offered by the selected channel: for example to assure that b(t) is opportunely band limited, if x(t) is unlimited, or to get waveforms with redundancy of the carried information, more resilient to the channel disturbs.
The set composed of transmitting and receiving terminals is indicated as ***base band modem***; such a name means both the two subsystems which operate in a one way transmission and the set on the same side of the transmitting terminal and of the receiving one of a two way communication system.
Considering the [6.3], the behaviour of the base band channel is optimal if the following relation is valid:

$$[6.4]\ b_r(t) = g_C\, b(t - t_C),$$

where the factor g_C and the delay t_C are characteristics of an ideal base band channel, which must be linear and time invariant. As demonstrated immediately hereinafter, the whole transport of the information occcurs ideally if the terminals which compose the modem have the characteristic to be ***complementary***, which means that the latter block of the pair in theory performs, on a signal faithful to the one processed by the former, a processing which restores the initial form, but with a constant multiplicative factor and a delay. Such a property is shown in Figura 6.3, for a generic signal processor E_* and its complementary one $\bar{E}_*$. A processing, although the variation of the signal may be very deep, can be considered reversible if the complementary processing exists.

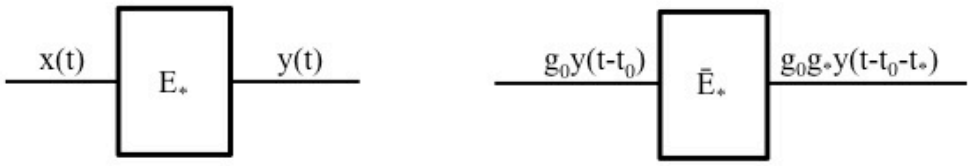


Figura 6.3: Properties of a pair of complementary signal processors.

The signal $x_r(t)$ output of the whole system in Figura 6.2 results to be faithful to the excitation x(t) if the base band channel is characterized by the relation $b_r(t) = g_C b(t-t_C)$ and the terminals are complementary; in fact, thanks to the latter property $x_r(t) = g_C g_M x(t-t_C-t_M)$ is obtained, where g_M and t_M are the characteristic factor and delay of the modem. Definitively, the optimal transport of the information is allowed in an ideal base band transmission system composed of the cascade connection, with ideal matching in every section, of:

- a base band transmission channel, with LTI behaviour and response faithful to the excitation;
- a pair of generic base band terminals, with non linear behaviour, as long as they are complementary.

6.1.2.2 Perfect transmission system in shifted band

In the case the in line signal propagating through the transmission means is of shifted band type, the transmitting and receiving equipment are usually more complex. As shown in Figura 6.4, in which the ideal matching in all the connection sections is straighforward, other than the ***transmission channel in shifted band***, including the transmission means and some functional blocks of the equipment, in the system at least two pairs of functional blocks with non linear processing are considered.

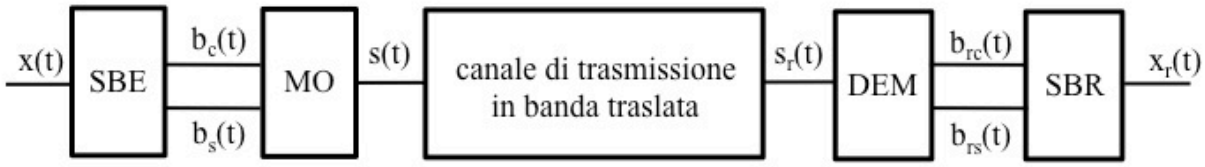


Figura 6.4: System with transmission channel in shifted band.

Representing the in line signal s(t) through the complex envelope with a harmonic reference f_c, expressed in its real and imaginery parts $\iota(t) = a_c(t)+ja_s(t)$, it is possible to note that in general it includes two real base band signals able to transport information, i.e. the shifted band channel offers two conceptual transmission paths over the same physical way. The processing operations executed by the inner functional blocks, i.e. by the harmonic modulator (MO) and by the harmonic demodulator (DEM), which are respectively named ***harmonic modulation*** and ***harmonic demodulation***, are just due not only to the need of the transformation between signals with different allocation of the respective spectra, around the harmonic reference in the channel and around the origin of the frequency axis in the external side, but due also to the opportunity to make sure that in general two base band signals are present, $b_c(t)$ and $b_s(t)$, on the transmitting side and their corresponding ones, $b_{rc}(t)$ e $b_{rs}(t)$, on the receiving side. The processing operations executed by the outer blocks, i.e. by the transmitting base band section (SBE) and by the receiving base band section (SBR) of the end equipment, add the functionality of distributing over two physical paths, or to bring back together on one path, the information to be transported to similar functionalities of a system with base band channel.

The set of the four blocks which realizes the transmitting and the receiving terminals is indicated as ***shifted band modem***; as usual such a name indicates both the two subsystems which operate in a one way transmission and the set on the same side of the transmitting terminal and of the receiving one of a two way communication system.

In the case of signals in shifted band, utilizing the representation of direct signals s(t) and $s_r(t)$ by means of the respective complex envelopes and imposing only to them to be faithful, instead of the [6.3] the less restrictive condition (see [2.10]):

$$[6.5]\ \iota_r(t) = g_C\, e^{-j\Gamma}\, \iota(t-t_C),$$

is obtained, where the factor g_C and the delay t_C are characteristic of an ideal channel in shifted band and Γ is an arbitrary real constant. The behaviour of the considered channel is then optimal if such a relation is valid. Then, as in the previous case, the whole transport of the information occurs ideally if both the pair in the base band sections and the pair harmonic modulator/demodulator are complementary, where for the two latter the complex envelopes instead of the real in line signals are considered.
Definitively, the optimal transport of the information is allowed in an ideal shifted band system composed of the cascade connection, with ideal matching in every section, of:

- a transmission channel in shifted band in which the complex envelope of the response is faithful to the one of the excitation;
- a harmonic modulator and a harmonic demodulator adjacent to the channel, with non linear behaviour, as long as they are complementary and consider the correct reference in frequency;
- a pair of generic base band terminals at the ends of the system, with non linear behaviour, as long as they are complementary.

In the above remarks developed so far the possibility that some subsystem of the cascade composing the channel with in line signal in shifted band has non linear nature was not considered.
In the case of signals in narrow relative bandwidth, with bandwidth B much lower than the reference frequency, nevertheless the configuration of the transmission channel may foresee the insertion in cascade also of one or more frequency converter. As already mentioned, the ideal transmission with respect to the complex envelopes is then allowed if thenon linear subsystems which implement the frequency conversion are perfect within the interval with width B, which results nevertheless centered on a frequency f_{ce} in input and on a different frequency f_{cu} in output.
In the case just one conversion is inserted, the global channel, although perfect with respect to the [6.5], has a non linear behaviour with respect to the corresponding real signals s(t) and $s_r(t)$, which have in fact spectra allocated around the two different frequencies f_{ce} and f_{cu}. Moreover, if in the channel two perfect frequency converters in the same frequency B, but complementary (i.e. able to provide with the pair in cascade the ideal transmission of the physical real signal considering that in the latter the reference frequency in output is just that f_{ce} in input to the channel) are inserted, the channel shows at the end points a global behaviour LTI like.

6.1.3 Perfect transmission means

In the simplest case the transmisison system is reduced just to the transmission means, that can usually be represented with a LTI passive and symmetrical quadrupole, thus characterized by identical reflectances and by the transmittances $H_{12}(f,r)=H_{21}(f,r)$, the latter ones dependent on the covered distance r. In such an event the ideal transmission occurs if the reflectances are zero in the useful bandwidth of the signal and the transfer function, which is then identified with the transmittances, assumes in band the particular expression:

$$[6.6]\ H_0(f,r) = g(r)\, e^{-j2\pi f t_0(r)},\ \forall\ |f| \in (f_m; f_M),$$

where:

- the factor g(r) depends only on r, is positive and satisfies the $g(r)<<1$, so that in practice the reflected signal $y_1(t)$ at the input port with respect to the one of excitation x(t) can be neglected;
- also the delay $t_0(r)$ depends only on r.

In fact, the considered conditions define a ***perfect transmission means***, taking into account that other than $y_1(t)\cong 0$ the signal at the output $y(t) = g(r)\, x[t - t_0(r)]$ occurs, faithful to the one at the input x(t), although significantly weakened.
Considering that in the generic perfect means, both physical and radio carrier, the delay results to be proportional to r, it is possible to state that:

$$[6.7]\ t_0(r) = \frac{r}{v},$$

where the proportionality constant is the inverse of the propagation speed v along the means itself. Such a propagation speed, constant not only with respect to the variation of r, but also with respect to the variation of f in the useful band, assumes a value that depends on the particular structure: the radio segments are all characterized by v=c, while in the physical carriers its value is lower than c, depending on the dielectric material.
Instead, the modulus of the transfer function, or better the attenuation $A_0(r)=1/g^2(r)$ of the perfect transmission means, can be expressed by two different laws:

$$[6.8]\ A_0(r) = e^{2r}, \qquad \text{for physical carrier,}$$

$$[6.9]\ A_0(r) = A(r_0)\left(\frac{r}{r_0}\right)^2, \qquad \text{for radio carriers,}$$

where the quantity $\alpha>0$, independent on r and on f in the useful band, is named the ***attenuation constant*** of the particular physical structure, while $A(r_0)$ is the attenuation that would occur for the same oriented pair of antennae with a prefixed reference value r_0 of the distance (for example r_0=1 km).
For the physical carrier means the existence of the exponential law of the attenuation can be deduced noting that whatever segmentation in two segments can be adopted, long r_1 and r_2, for a uniform structure long $r=r_1+r_2$, the total attenuation keeps to be unchanged; since the latter is given by the product of the attentuations in the two segments in cascade, the result is that $A_0(r_1+r_2)$ must be equal to the product $A_0(r_1)A_0(r_2)$, condition that can be satisfied only with the exponential law.
For a radio path the typical dependance of the attenuation on the square of the distance derives from the assumption that the propagation of the signal is exploited in the free space with no losses and, thus, the integral of the radiated power flux over any spherical surface of radius r, with the center in the origin of the spherical radiated electromagnetic wave, is kept constant.
The attenuation of the transmission means usually assumes values much greater than the unit. In practice, very often the evaluation in logarithmic units is adopted, using neper [Np] or, more frequently, decibel [dB]:

$$A[\mathrm{Np}] \mathrel{\hat{=}} \frac{1}{2}\ln(A), \qquad A[\mathrm{dB}] \mathrel{\hat{=}} 10\log(A),$$

where ln and log indicate respectively the natural logarithm and the one in base 10. In particular, in the case of perfect physical carrier on the basis of the [6.8] the result is:

$$[6.10]\ A_0[\mathrm{Np}] = \alpha\, r,\ A_0[\mathrm{dB}] = \alpha[\mathrm{dB/m}]\, r,$$

where it is evident that the attenuation constant α is measured in Np/m and the assumption:

$$\alpha[\mathrm{dB/m}] = 20\log(e)\,\alpha \cong 8{,}686\,\alpha$$

is adopted. Finally, it is possible to note that with the actual transmisison means it is possible to get close to the target of the perfect matching at the two ports, while it is often difficult to get it in the useful band of the transfer functions H(f,r) that well approximate the perfect behaviour highlighted in the [6.6], especially if the relative bandwidth is very large.

6.1.4 Perfect linear channels

In its simplest version, an ideal transmission system is constituted just by the transmission means (central block in **Errore. L'origine riferimento non è stata trovata.**), while a non elementary transmission system includes, in addition to the irreplaceable transmission means, also other subsystems.
In the first chapter the typical structure of a transmission system (see **Errore. L'origine riferimento non è stata trovata.**) was introduced, highlighting that it is composed of a transmission means, with the task to trasnport the signal at long distance, and of transmisison equipment inserted before and after, which implement other fundamental functions usually distributed in multiple blocks in cascade. Usually, the transmission means is of linear type, to get close to the target of the perfect behaviour, but often also the functional blocks of the transmitting and receiving equipment directly connected to the carrier means have the same property, so that the whole set of blocks constitutes a ***transmission channel*** linear as a whole.
In the case all the blocks composing the channel can be singularly represented by LTI like quadrupoles, perfect in the same band, the global behaviour can be assumed as the one of a ***perfect transmission channel***, or simply ***perfect channel***, still perfect LTI in the same band; its transfer function is equal to the product of the transfer functions of all the the m representative quadrupoles which compose the cascade (see [5.71]): indicating with G_i the gain and with t_{0i} the group delay of the i^{th} quadrupole, the total gain and group delay of the channel, both constant in band, are given by the following expressions:

$$[6.11]\ G = \prod_{i=1}^{m} G_i\ ,\ t_0 = \sum_{i=1}^{m} t_{0i}\ .$$

Observing that in a perfect channel the extremes of the band f_m and f_M within which it operates are always specified, it is possible to distinguish perfect channels in base band, as usual with $f_M >> f_m$, and perfect channels in narrow, or very narrow, relative band.
From a purely conceptual point of view such a situation is acceptable for any finite distance r to cover, because in theory, even in case of very great value of the attenuation of the ideal means, the received signal keeps the characteristic to be faithful as required by the [1.1], for any small value of the factor g; nevetheless, it is reasonable, even neglecting the presence of undesired signals in reception, to have the objective to recover the power level of the received signal y(t).
In its simplest configuration, a perfect channel includes at least an active block in addition to the transmission means: it is a ***perfect amplifier*** with gain $G_A = g_A^2$, that can be represented with a LTI quadrupole with the elementary transfer function $H_A(f)=g_A$ constant in the useful band. The amplifier is inserted to recover the significant attenuation of the signal going out from the transmission means with attenuation $A(r)>>1$ (actually the total attenuation is due to multiple elements among which the transmission means provides the largest contribution; the other contributions regard the waveguides, passive elements, signal splitters, other equipment); putting the two subsystems in cascade, in fact, in the useful band the total gain of the channel equivalent quadrupole is given by the following expression:

$$[6.12]\ G_{eq} = \frac{G_A}{A(r)}\ ,\ \forall\ |f| \in (f_m; f_M)\ ,$$

which with $G_A=A(r)$ allows to get a transferred faithful signal after the cascade with the same power of the excitation one, leaving as only effect of the ideal transmission in the considered channel, with total gain equal to one, the presence of the delay $t_0(r)$ introduced by the transmission means. What just stated is valid disregarding the location of the amplifier, at the beginning or at the end of the channel.
The presence of the attenuation $A_0(r)>>1$ can then be easily compensated with an ***amplifier***, capable to provide at the output a signal z(t) copy of the one at the input y(t) but different for a factor $g_A>>1$, that means having a gain:

$$[6.13]\ G_A = \frac{W_{zz}}{W_{yy}} = g_A^2\ ;$$

taking into account also that $A_0(r)=W_{xx}/W_{yy}$, in fact, as a whole the output power is:

$$[6.14]\ W_{zz} = G_A\, W_{yy} = \frac{G_A}{A_0(r)} W_{xx},$$

which in the case of the choice $G_A=A_0(r)$ would provide $W_{zz}=W_{xx}$, as shown in the scheme in Figura 6.5a. The result of the compensation of the attenuation is the same if the amplifier is put before the transmisison means, as shown in Figura 6.5b.

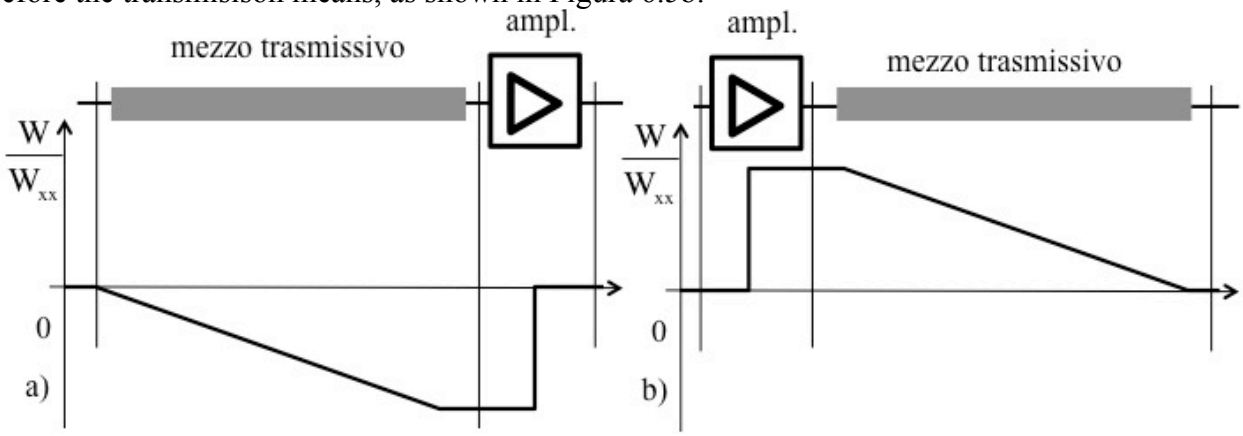


Figura 6.5: Transmission channels composed of an amplifier and a transmission means, with evolution of the signal power along the channel.

With no impairments in the amplifier and in the transmission means, the order of the two blocks, both linear, which constitute the ***transmission channel***, is not meaningful; with the same signal power W_{xx}, assumed equal at the beginning and at the end of the considered channel, nevertheless it is possible to note that putting the amplifier before (see Figura 6.5b) at the output a stronger power G_AW_{xx} is achieved, difficult to obtain keeping the linearity of the active block, so that the other solution would seem to be better (see Figura 6.5a), which, on the other hand, appears intuitively more risky for the very low value of the output power from the transmission means W_{xx}/G_A, considering the presence of a disturb at the output of the transmission means having the same nature and power independent on the signal one.

Actually, taking into account both the mentioned impairments, i.e. the potential lack of linearity of the amplifier and the presence of a disturb in the transmisison means, it is anyway reasonable that the preferred channel configuration considers the compensation of the attenuation of the transmission means in part at the beginning, with an active block named ***power amplifier***, and for the remaining part at the end, with another block named ***reception amplifier***.

Going forward with increasing complexity, it is possible to add other non linear functional blocks in the transmitting and receiving equipment, often creating pairs of deep transformations of the signal in transit, but reversible considering that a block in the receiving equipment works in a complementary way with respect to the corresponding block in the transmitting equipment, so that the information globally transferred is not altered. From logical point of view, the configuration of the equipment depends on the selection of the transmission means: in fact, it results to be defined on the basis of the whole technical-economical optimization of the transmission system, once known the type of carrier (either physical or radio) and its real performance.

In the case very long distances must be covered, the gradual recover of the attenuation produced by the transmission means may be convenient, dividing the whole path in many segments and inserting a number of active ***repeater equipment*** along its path. In this way, the subdivision of the link in several transmission systems in cascade is achieved.

As already mentioned in the introductory chapter, it is very common to recover the attenuation of the transmission means in part at the beginning, with a ***perfect power amplifier*** with gain G_E, and in part at the end, with a ***perfect reception amplifier*** with gain G_R, obtaining the total gain of the channel:

$$[6.15]\ G_{eq} = \frac{G_E G_R}{A(r)},\ \forall\ |f| \in (f_m; f_M).$$

In such a case, as in the scheme in Figura 6.6, a perfect channel with unit gain assuming $G_E G_R = A(r)$ is obtained.

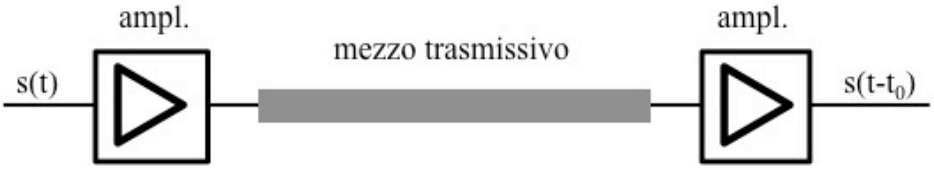


Figura 6.6: Configuration of a perfect channel with total unit gain, composed of all perfect subsystems.

6.2 LINEAR PROCESSING OF TIME CONTINUOUS SIGNALS

6.2.1 Linear processing with no cut of the band

6.2.1.1 Effects of LTI transformation with memory

The simplest processing applicable on a generic real signal x(t) is its LTI transformation with memory through the transit in a quadrupole Q physically feasible, with transfer function H(f); the operation is implemented with direct and time continuous action on the instantaneous physical quantity of the excitation: thus, also the name ***continuous time linear signal processing*** is adopted. Indicating as usual with $h(t)=F^{-1}\{H(f)\}$ the relative pulse response, real and causal, as a consequence of the transformation, as well known, at the output the real signal provided by the convolution is obtained:

$$[6.16]\ y(t) = h(t) * x(t).$$

The [6.16], to which a similar expression applicable in the case of signals s(t) in shifted band corresponds, is independent on the digital or analogical nature of the information carried by the signal.

With reference to the case of energy signals, reminding the relation between the cross correlation function and the convolution (see [3.99]) the following relation is valid:

$$[6.17]\ C_{yx}(\tau) = y(\tau) * x^*(-\tau) = h(\tau) * x(\tau) * x^*(-\tau) = h(\tau) * C_{xx}(\tau);$$

then, a similar expression with the functions $R_{yx}(\tau)$ and $R_{xx}(\tau)$ is obtained, valid in the case of power signals. Thus, the response y(t) results to be correlated to the excitation signal x(t), but since the transformation is of type with memory the output function is a ***distorted signal***, with waveform not faithful with respect to the input one, while, instead, it would be if the transformation were memoryless (see [5.8]). Therefore, the considered processing produces a time invariant ***linear distorsion*** and the system, that introduces it, is a ***LTI distorting quadrupole***.

On the other hand the aim is that the transmission system as a whole behaves so that the output signal is faithful to the input one, but each block is designed and realized to implement specific functions. Moreover, let's take into account that the memoryless transfer funtion is just ideal and as a consequence not feasible.

With similar procedure, in the case of energy signal, the autocorrelation of the output signal is:

$$[6.18]\ C_{yy}(\tau) = h(\tau) * h^*(-\tau) * C_{xx}(\tau),$$

and a similar expression with the functions $R_{yy}(\tau)$ and $R_{xx}(\tau)$ can be carried out in the case of power signals. Applying the Fourier transform and reminding the Wiener-Khintchine relations (see [3.107] and [3.114]) the effects of the linear processing in the frequency domain are obtained:

$$[6.19]\ E_{yy}(f) = |H(f)|^2 E_{xx}(f),\ W_{yy}(f) = |H(f)|^2 W_{xx}(f),$$

where with E(f) or W(f) the energy or the power spectral densities of the considered signals are respectively indicated. The expressions of the energy and of the power are carried out by means of the integration of the respective spectral densities:

$$[6.20]\ E_{yy}= \int \left|H(f)\right|^2 E_{xx}(f)\,df\ ,\ W_{yy}= \int \left|H(f)\right|^2 W_{xx}(f)\,df\ .$$

In some cases the processing is applyied, taking hint from the [6.19], just to opportunely shape the spectrum of the output signal, that means to distribute in a different way its energy or power within the same band; often this is implemented without modifying the total energy or power. In the case the processing is without cutting the bandwidth, the transfer function H(f) of the utilized LTI quadrupole must be different from zero over all the signal bandwidth.
Instead, in some other cases the scope can be to cut part of the spectrum of a signal, utilizing a LTI quadrupole which at the considered frequencies has a transfer function equal to zero, or almost zero in modulus: the linear processing is then of type with cut of the bandwidth.
One way to observe the effect produced by a generic linear processing in the time domain consists in expressing both the excitation signal x(t) and the response y(t) with the sample forms:

$$[6.21]\ x(t) = \sum_k x_k \mathrm{sinc}\left(\frac{t}{T_N}-k\right),\ y(t) = \sum_k y_k \mathrm{sinc}\left(\frac{t}{T_N}-k\right),$$

where x_k and y_k are the samples fully representative of the signals:

$$[6.22]\ x_k=x(kT_N),\ y_k=y(kT_N),$$

being $T_N=1/2f_M$ the Nyquist interval. From the information transfer point of view, a linear processing operating in a continuous way on the instant physical quantity of the signal, can be put in correspondence with a transformation which operates at discrete time between an input sequence x(n) and an output y(n), having as elements the respective samples of the signals x(t) and y(t), with time interval T_N unchanged both in input and in output.
As demonstrated in section 6.2.1.2, where the reader interested to get more details can go, performing the transformation with a physically feasible distorting LTI quadrupole the corresponding discrete time processing on the signal samples assumes the particular linear form:

$$[6.23]\ y_k= \sum_{n=0}^{\infty} h_n x_{k-n} = h_0x_k+h_1x_{k-1}+...+h_mx_{k-m}+...,$$

in which h_n are the elements of a right monolateral sequence characteristic of the quadrupole, named ***equivalent discrete pulse response***. From the [6.23] it is possible to find out that a LTI transformation with memory performs the convolution between its characteristic monolateral sequence h(n) and the one x(n) of the input signal. Each sample y_k of the sequence y(n) of the distorted output signal is given by a linear combination with real constant coefficients of a number, in theory infinite, of the samples x_k of the input signal, starting from the one corresponding to the same and proceeding backward for all the past ones. In practice, the the equivalent pulse response can be assumed to be a monolateral sequence h(n) of finite lenght.

6.2.1.2 Equivalent discrete pulse response

Let's consider a LTI quadrupole with transfer function H(f), with a signal x(t) upper bounded in band with finite extreme f_M at the input.
Thus, it allows the maximum sampling interval:

$$[6.24]\ T_N = \frac{1}{2f_M}.$$

As concerns the outcome of the transformation, the quadrupole can be replaced with an equivalent one, with transfer function made periodical in frequency with period $2f_M=1/T_N$:

$$[6.25]\ H_{eq}(f) \hat{=} \mathrm{rep}_{2fM}\{H(f)e^{j2\pi ft_0}\,\mathrm{rect}(fT_N)\}\,e^{-j2\pi ft_0},$$

considering that this is different from the former only at the frequencies in which the excitation is equal to zero. The modulus of $H_{eq}(f)$ is continuous, considering that $|H(f)|$ is a even function; to achieve also the continuity of the argument of $H_{eq}(f)$ it is opportune that the delay t_0 is chosen so that the $\arg\{H(f_M)\}=-2\pi ft_0$ is satisfyied.
The [6.25] can be expressed in Fourier series, achieving:

$$[6.26]\ H_{eq}(f) = \sum_{n=0}^{\infty} h_n\ e^{-j2\pi nfT_N},$$

where the Fourier coefficients are indicated with h_n:

$$[6.27]\ h_n = \frac{1}{2f_M}\int_{-f_M}^{f_M} H_{eq}(f)e^{j2\pi nfT_N}\,df = T_N\int_{-f_M}^{f_M} H(f)e^{j2\pi nfT_N}\,df,$$

zero for $n<0$ considering that $H_{eq}(f)$ keeps the physical feasibility.
Performing the inverse transform on the [6.26] the time discrete function is obtained:

$$[6.28]\ h_{eq}(t) = \sum_{n=0}^{\infty} h_n\,\delta(t-nT_N),$$

named ***equivalent discrete pulse response***, to which a right monolateral sequence h(n) with time interval T_N corresponds. Utilizing this function in the convolution $y(t)=h(t)*x(t)$, where the excitation is expressed in the sampled form [6.21], the response still in sampled form is obtained:

$$[6.29]\ y(t) = h_{eq}(t)*\sum_k x_k \mathrm{sinc}\left(\frac{t}{T_N}-k\right) = \sum_k \sum_{n=0}^{\infty} h_n x_k\,\delta(t-nT_N)*\mathrm{sinc}\left(\frac{t}{T_N}-k\right) = \sum_k \sum_{n=0}^{\infty} h_n x_k\,\mathrm{sinc}\left(\frac{t}{T_N}-k+n\right)$$

$$= \sum_k y_k \mathrm{sinc}\left(\frac{t}{T_N}-k\right),$$

where in the last equality the change of the discrete variable from k-n to k was performed and the elements y_k relative to the transformed signal are given by:

$$[6.30]\ y_k = \sum_{n=0}^{\infty} x_{k-n}h_n.$$

6.2.1.3 Channel equalization

It is important to note that a generic linear processing with no cut of the band, for which the distorting quadrupole Q is assumed to not cutting any spectral component of the signal, has the property to be reversible, implying that to obtain from the processed signal y(t) a faithful signal to the one of the excitation x(t) is possible.
In fact, being $|H(f)|>0$ for every $|f|$ within the extrems f_m and f_M of the band of the signal, then, the matched LTI quadrupole which has the transfer function:

$$[6.31]\ \bar{H}(f) \hat{=} \frac{g}{H(f)}e^{-j2\pi f\bar{t}},\ \forall |f| \in (f_m;f_M),$$

inverse of the H(f) but with a complex factor having constant modulus g and argument proportional to the frequency through the delay $\bar{t}$, is physically feasible, so that from the cascade of the two complementary quadrupoles Q and $\bar{Q}$ the following relation is obtained:

$$[6.32]\ H_{eq}(f) = H(f)\,\bar{H}(f) = g\,e^{-j2\pi f\bar{t}},\ \forall |f| \in (f_m; f_M),$$

which guarantees the ideal transmission as a whole.
The distorting quadrupole $\bar{Q}$, which is able to compensate the linear distortion with no cut of the band introduced by the quadrupole Q, is named time invariant ***equalizer***.
The equalization technique is largely utilized in actual transmisison channels, which otherwise, due to the above mentioned non perfect behavioiur of the transmission means, with variable attenuation and group delay as a function of the frequency in the useful band, would cause distortion of the signal at their output. The most common configuration of the LTI channel is thus shown in Figura 6.7.

Figura 6.7: Configuration of a linear channel with unit total gain, in which the perfect behaviour is acquired through equalization

6.2.2 Examples of linear processing with no cut of the band

The linear processing without cut of the band can be of several types, having the possibility to choose among a cospicuous set of transfer functions H(f), but respecting the reversibility of the operation. Hereinafter some examples will be examined, significant for the transformation that characterizes them. The LTI transformations respectively performed through the operations of derivation and integration in the time domain are two very common examples:

$$[6.33]\ y(t) = \frac{dx(t)}{dt},\ y(t) = \int_{-\infty}^{t} x(u)\,du\,.$$

It is possible to note that such processing are complementary between each other, but differing for a constant, which means that each one allows the reversibility of the other.
Applying the Fourier transform to the [6.33], the following transfer functions of the ***ideal derivator*** and of the ***ideal integrator*** are determined:

$$[6.34]\ H_{Di}(f) \triangleq j2\pi f,\ H_{Ii}(f) \triangleq \frac{1}{j2\pi f}\,.$$

Nevertheless, the respective quadrupoles are neither ideally feasible, as it appears from the previous characteristic functions each one having a pole, respectively for $f \to \infty$ and for $f=0$.
Starting from band limited signals with non zero f_m, the transformations [6.33] can be achieved with reversibility with LTI quadrupoles with memory physically feasible, which have transfer functions of type:

$$[6.35]\ H_D(f) \triangleq H_{Di}(f)\,g\,e^{-j2\pi f t_0},\ \forall |f| \in (f_m; f_M),$$

$$[6.36]\ H_I(f) \triangleq H_{Ii}(f)\,g\,e^{-j2\pi f t_0},\ \forall |f| \in (f_m; f_M),$$

where g and t_0 are real and positive constants. It is possible to verify that, in the band of the signal, the modulus of the transfer functions is limited and different from zero and the group delay results to be equal to t_0. These complementary quadrupoles, named respectively ***perfect derivator*** and ***perfect integrator***, deeply modify the energy or the power distribution of the signal within the

same bandwidth; in fact, on the basis of the [6.19] for a derivator the following relation is obtained:

$$[6.37]\ E_{yy}(f) = (2\pi f)^2 g^2 E_{xx}(f),\ W_{yy}(f) = (2\pi f)^2 g^2 W_{xx}(f),$$

and for an integrator:

$$[6.38]\ E_{yy}(f) = (2\pi f)^{-2} g^2 E_{xx}(f),\ W_{yy}(f) = (2\pi f)^{-2} g^2 W_{xx}(f),$$

with the consequent magnification of the spectral components respectively at the high or at the low frequencies. In Figura 6.8 the effects on the output spectral densities in the case of excitation signal with uniform spectrum in band are shown.

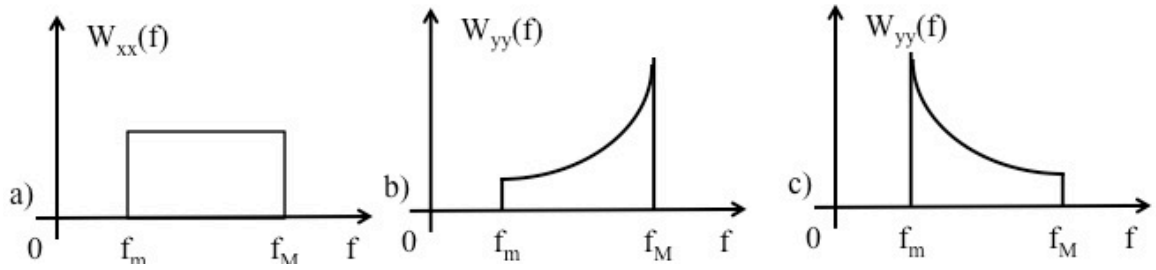


Figura 6.8: Input spectral density (a) and output spectral density respectively from a derivator (b) and from an integrator (c).

Other examples of analogue processing without cutting the band and reversible are those offered by the pair of LTI quadrupoles with memory physically feasible, still complementary between each other, with the respective transfer functions, having constant and unit modulus in a band from f_m up to f_M:

$$[6.39]\ H_H(f) \hat{=} -j\,\mathrm{sgn}(f)\, e^{-j2\pi f t_0} = e^{-j\left[2\pi f t_o + \frac{\pi}{2}\mathrm{sgn}(f)\right]},\ \forall |f| \in (f_m; f_M),$$

$$[6.40]\ \bar{H}_H(f) \hat{=} j\,\mathrm{sgn}(f)\, e^{-j2\pi f t_0} = e^{-j\left[2\pi f t_o - \frac{\pi}{2}\mathrm{sgn}(f)\right]},\ \forall |f| \in (f_m; f_M);$$

nevertheless, the transformations are applicable only if the excitation signals have no spectral components near the origin, which means that still $f_m > 0$ occurs.
It is possible to verify that in the considered band the group delay is equal to the positive constant t_0. The two considered quadrupoles, which don't modify the spectral distribution of the energy or of the power, produce the fundamental effect to introduce a step addend in the argument of the transfer functions, with consequent additional phase shift of $\pm\pi/2$ of each output spectral component; thus, they are named ***phase shifters of $\pm\pi/2$***.
The reader interested to further details can verify that the considered phase shifters quadrupoles perform the Hilbert and the inverse Hilbert transform (see [3.174] and [3.179]), but with a delay t_0.
Note that in the interval $(f_m; f_M)$ the functions $H_D(f)$ and $\bar{H}_H(f)$, as well as the pair $H_I(f)$ and $H_H(f)$, have arguments with the same type of evolution, of which in Figura 6.9 some examples are shown.
The quadrupoles that is possible to build to obtain in the band of the signal with good approximation, but not exactly, the evolutions of the transfer functions of the four types of the considered linear processors have characteristics with no discontinuities, both in modulus and in argument, over the whole frequency axis. For an assigned tolerance on the small deviations in band between the actual and the theoretical transfer functions, the circuit complexity grows as the value of the parameter:

$$[6.41]\ Q_H \hat{=} \frac{B}{f_m} = \frac{f_M - f_m}{f_m},$$

which greater it is more difficult and expensive the realization is; the approximation is thus easy in the case of transformations to perform on signals in narrow relative band ($B<f_m$), and much more if very narrow.

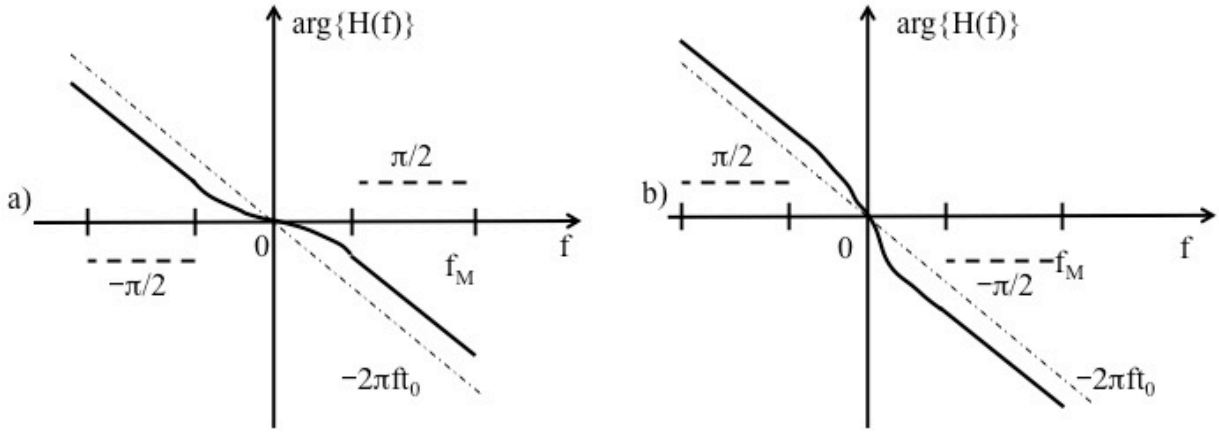


Figura 6.9: Evolutions of the argument of the transfer functions $H_D(f)$ and $\bar{H}_H(f)$ (a) and of the transfer functions $H_I(f)$ and $H_H(f)$ (b).

6.2.3 Elements on filters

Aiming to separate a useful signal x(t) with limited band within the upper extreme f_M from potential other additional signals, that instead have non zero spectral components only out of the mentioned interval, the transfer function is immediately identified (see Figura 6.10):

$$[6.42]\ H_{ti}(f) \triangleq \text{rect}(f/2f_t)\, e^{-j2\pi ft_0},$$

with cut off frequency $f_t=f_M$, characteristic of the LTI quadrupole named ***rectangular low-pass filter***. In fact, the system allows the transparent transfer, but with a constant delay t_0, of all the spectral components withing the ***pass band***, between 0 and f_t within which $|H_{ti}(f)|=1$, and doesn't allow any transfer in the ***stop band***, between f_t and infinite within which instead $H_{ti}(f)=0$.

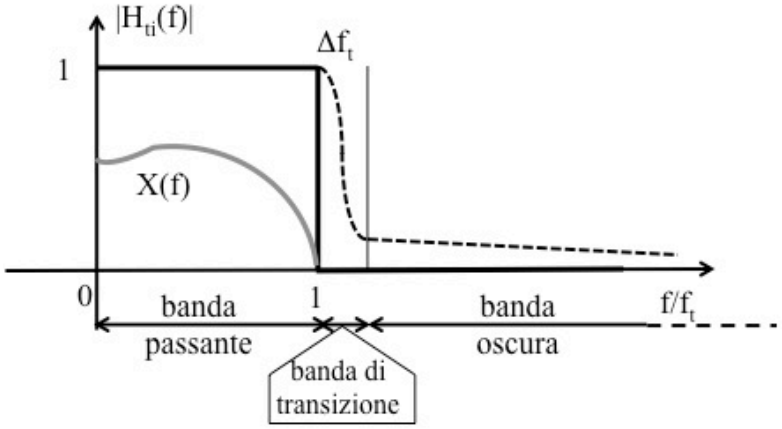


Figura 6.10: Modulus of the transfer function of the rectangular low pass filter, with continuous line, and of the perfect low pass filter, with dashed line.

Considering that the transfer function [6.42] is strictly limited in band, the pulse response is time unlimited: with finite t_0 the causality is not verified, implying that the rectangular low pass filter is just ideally feasible. Nevertheless, at least in the useful band of the signal the considered performance is guaranteed by a LTI quadrupole physically feasible, named ***perfect low pass filter***, for which the modulus of the transfer function is still equal to one up to the cutoff frequency, meaning that:

$$[6.43]\ H_t(f) \triangleq H_{ti}(f),\ \forall |f| < f_t,$$

and is near to zero for $|f|>f_t$, except that in the small transition interval Δf_t, adjacent to the pass band, within which it passes from one to near to zero values, as shown in the part in dashed line in Figura 6.10.
Passing to the quadrupoles that is actually possible to build with which the target performance can be obtained with good approximation, it is evident why the filters are considered among the systems that perform linear processing with memory, with no cut of the band on the useful signal. In fact, in practice it is not possible to achieve a steep transition between pass band and stop band without producing small, but unavoidable, effects of distortion in the output signal. A cathegory of practical filters, named maximum flatness or Butterworth, has for example the following evolution of the attenuation:

$$[6.44]\ A_B(f) = 1 + k^2(f/f_t)^{2n},$$

with $k^2<1$ and n positive integer, which for any value of n provides at the cutoff frequency $A_B(f_t)=1+k^2$; even if a small value of k^2 and a high value of n are fixed, the attenuation is not rigorously constant in the band of the useful signal, which as a consequence results distorted in output, even if not significantly. Adopting, as determined for the cathegory of the Tchebysheff filters, an evolution $A_T(f)$ of the attenuation with small oscillations in the pass band, all with amplitude $1+k^2$ which is still the value at the cutoff frequency, it is possible to achieve with the same order of the polynomial in frequency expressing the function $A_T(f)$ a steeper growth in the transition interval, but with a distorting effect more emphasised, with the same value of k^2. Moreover, there are other types of filters with even more complex developments.
Once established the tolerance of the deviation with respect to the behaviour of the ideal filter, often as a first step a mask is sketched (see Figura 6.11) which at least fixes the maximum value $A = 1+ k^2$ allowed for the variable attenuation in the useful band (slightly greater than one), the extension of the transition interval Δf_t and the minimum allowed value for the attenuation in the stop band (much greater than one). Then, it is possible to identify the minimum value to assign to the order 2n of the polynomial $A_B(f)$ to satisfy the requirements, i.e. so that the development of the filter attenuation is within the mask. The same methodology can be applied also to other cathegories of filters.

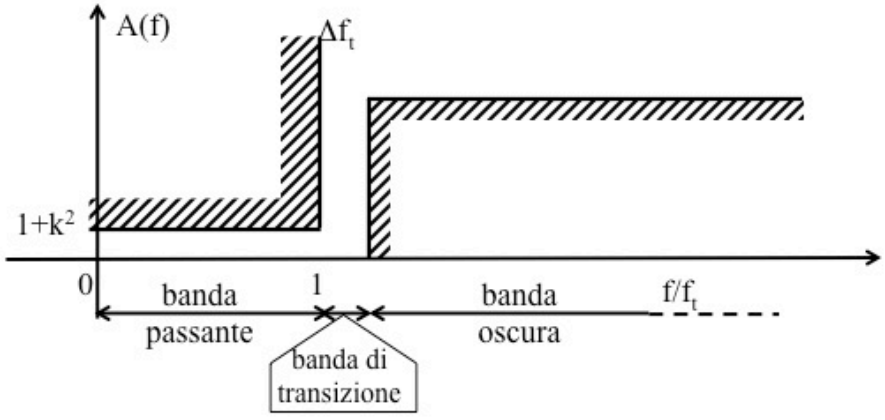


Figura 6.11: Attenuation of a low pass filter with maximum flatness, with development within an assigned mask.

With the same tolerances, the circuit complexity of the filter, which increases with n, is in general greater as a function of the value of the parameter:

$$[6.45]\ Q_t \triangleq \frac{f_t}{\Delta f_t}.$$

In addition to the low pass filter, other quadrupoles which perform linear processing without cut of the bandwidth are considered, still with characteristics strongly selective in frequency. With respect to their behaviours, shown in Figura 6.12 through the evolutions of the attenuation

A[dB]=10logA expressed in decibel, the four main types are named ***low pass filter*** (a), ***high pass filter*** (b), ***band pass filter*** (c) e ***stop pass filter*** (d).

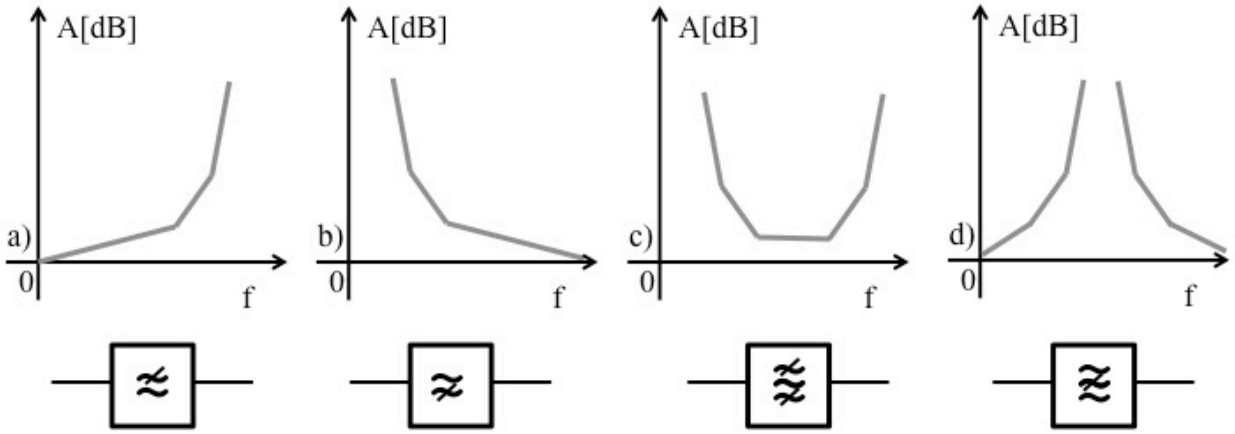


Figura 6.12: Development of the attenuations in decibel in the main types of filters – low pass (a), high pass (b), band pass (c) and stop pass (d).

At the bottom of Figura 6.12 the usual symbols used in the representations with block schemes are also shown.

The attenuation functions of the type of filters just introduced can be carried out from the expression of the attenuation A(f) of the low pass type through simple replacement of the independent variable:

- with (f_t/f) instead of (f/f_t) the high pass type attenuation is obtained, with cutoff frequency f_t;
- with $(f-f_a)/f_t$ instead of (f/f_t) the band pass type attenuation is obtained, with pass bandwidth $B_t=2f_t$ centered on the frequency f_a;
- with $f_t/(f-f_a)$ instead of (f/f_t) the stop pass type attenuation is obtained, with stop bandwidth $B_t=2f_t$ centered on the frequency f_a.

6.2.4 Linear processing with cut of the band

6.2.4.1 Methodology in the processing with cut of the band

Some signals, and in particular those digital with a step shape, often can be transmitted at a long distanza more conveniently after having experienced a processing which limits the signal band occupation at the output up to an opportune finite value f_{yM}, but ensuring that the suppression of a part, even significant, of the spectrum produces a distorsion effect not meaningful at all, or at least acceptable, with respect to the transmitted information. Then, the linear processing with cut of the band is adopted, which can be implemented by means of a LTI quadrupole physically feasible, with transfer function practically zero starting from the considered value f_{yM}.

In some cases, typically in that one of the step signal, the spectrum of the input signal is only practically limited in frequency with upper bound f_M, while the available bandwidth for the transmission has an extension f_{yM} much less than f_M; considering the significant distortion effect produced by the cut of the band, then, it is necessary to operate caring particularly to keep the transmitted information. As presented in the next section, in the digital case fortunately it is possible to implement some linear processing with significant cut of the band, but of reversible type, which allow the possibility to rebuild a signal faithful to the original one.

The simplest methodology to reduce the bandwidth concerns the use of a low pass filter, but with cutoff frequency $f_t=f_{yM}$ which is lower, although not significantly, than the maximum frequency f_M of the excitation signal x(t) on which it is intended to work. The produced distorsion on the waveform of the signal, kept within tolerable limits for the transmission of the information, can't be in general compensated with any post processing, that means that the reversibility is not guaranteed in the considered type of processing.

6.2.4.2 Effects of the cut of the band

Let's consider the transfer function $H_{ti}(f)$ (see [6.42]) of a rectangular low pass filter, with cutoff frequency f_t and delay t_0. Supposing to have at the input the signal $x(t)\equiv\delta(t)$, at the output the pulse response is obtained:

$$[6.46]\ h_{ti}(t) = F^{-1}\{H_{ti}(f)\} = 2f_t \mathrm{sinc}[2f_t(t-t_0)],$$

with time development very different from the excitation one as a consequence of the linear distorsion due to the cut of the band. Since $H_{ti}(f)$ is hermitian, i.e. the rectangular low pass filter is at least ideally feasible, the response is real; as shown in Figura 6.13a, in output a pulse with a maximum occuring at the delay time t_0 is obtained and with elongations of meaningful duration, so that the output signal appears even before the instant t=0, in which the ideal pulse was applied at the input. Thus, the quadrupole is not causal, i.e. it is not physically feasible. The effect of the suppression of the spectral components beyond the cutoff frequency f_t appears also in the finite steepness of the output pulse: referring to the time between the instant in which the response reaches the maximum and that one immediately before in which the signal is equal to zero, the value $\Delta t=1/2f_t$ is obtained, inversely proportional to the cutoff frequency.

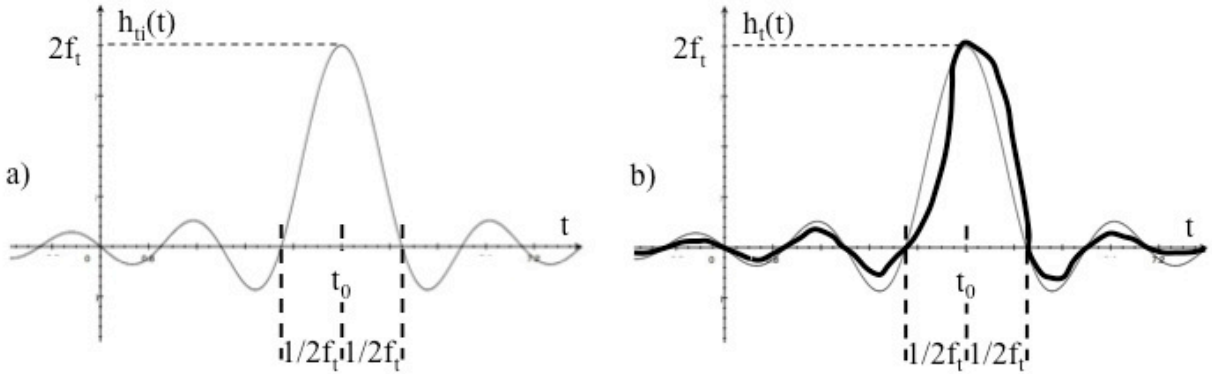


Figura 6.13: Pulse response of the rectangular low pass filter with delay t_0 (a) and of the filter physically feasible which approximates it (b).

A low pass filter physically feasible with transfer function that well approximates the rectangular one has a pulse response similar to that of Figura 6.13a. As shown in Figura 6.13b, the output is still a pulse with a rise time equal to $\Delta t=1/2f_t$ and a maximum corresponding to an instant t_0 (with $t_0>>\Delta t$), which can be assumed as the delay time of the filter; nevertheless, some significant diffferences occur, concerning a slight asymmetry of the pulse with respect to its maximum value point and, mainly, the elongations, which are shorter: the one ahead is zero for $t\leq 0$, while that one in delay is meaningless for $t>2t_0$. Definitively, the pulse response of an actual low pass filter can be considered practically limited in duration in the interval $(0; 2t_0)$.

Note that better the transfer function of the actual filter approximates the ideal one (with even higher value of the parameter [6.45]), implying that more its pulse response is similar to $h_{ti}(t)$, greater must be the delay time t_0 and, thus, the practical duration of the pulse response itself. In fact, the ideal low pass filter with rectangular transfer function implies a pulse response of infinite duration. Therefore, to be able to fully exploit the convolution an infinite time needs to pass.

The effect in the time domain caused by the transit of a generic base band signal x(t) through the rectangular low pass filter can be calculated, from time to time, through the convolution $h_{ti}(t)*x(t)$. Passing to an actual low pass filter, the distorted output signal is obtained with the same operation, but utilizing the actual pulse response $h_t(t)$, of the type shown in Figura 6.13b. In general, the following remarks can be stated.

In the case of energy signal strictly limited in an interval of D duration, and therefore theoretically unlimited in band, the distortion due to the cut of the band is surely present and is qualitatively highlighted under two aspects:

- the smoothing of the output time waveform, as greater as lower the cutoff frequency f_t is, mainly evident in the segments of greater steepness of the input signal;
- the time elongation with a practical duration that can be assumed equal to $D+2t_0$, as it is possible to find out from $y(t)=h_t(t)*x(t)$ taking into account that the pulse response has a practical duration $2t_0$ (remind that the duration of the convolution between two signals with finite duration is equal to the sum of the two durations).

In the case of signals strictly limited in band, with maximum frequency f_M, the former of the two above aspects appears of course only if $f_t<f_M$; the latter becomes meaningless, because even assuming that the duration D is practically finite its value is usually so big that the increase $2t_0$ can be neglected. Then, if the suppressed portion of the spectrum in not so large, as for example shown in Figura 6.14, the effect of the distorsion with smoothing of the waveform is weak, so that it can be accepted in the transmission, as happens in the case of the telephone analogue signal which is upper limited in spectrum at 3.4 kHz.

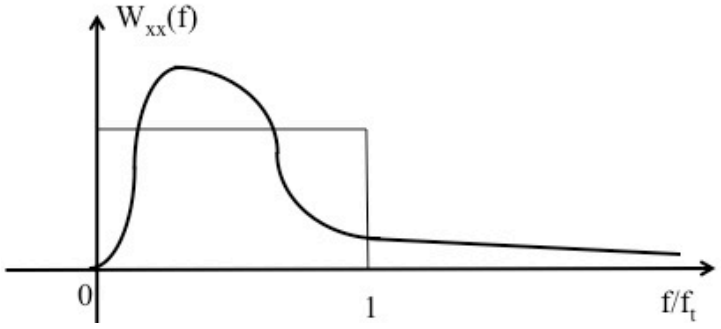


Figura 6.14: Power spectral density of a signal subject to the cut of the band.

6.3 PROCESSING OF STEP SIGNALS

6.3.1 Reversibility of the processing on step signals

6.3.1.1 Reversibility in the time domain

A power signal in the step shape of known duration T:

$$[6.47]\ a(t) = \sum_k a_k \mathrm{rect}\left(\frac{t}{T}-k\right),$$

since it is composed of a series of rectangular functions in the time domain has a spectrum infinitively large, which can be just practically limited in frequency, with finite upper bound f_M much greater than 1/T, i.e. as greater as shorter the duration T of each addend of the series is. The signals of the considered type often can be transmitted more conveniently after having experienced a processing which limits, in strict or practical sense, band occupation of the signal b(t) at the output up to an opportune finite value f_{bM} lower than f_M, but ensuring that the suppression of a part, even significant, of the signal spectrum produces an acceptable distorsion effect, or even not meaningful at all, with respect to the transmitted information. The issue concerns mainly the digital case, considering that the analogue step signals are rarely used in practice and only in the frame of processing inside the equipment.

From the [6.47] it is possible to observe that in the step signals the knowledge of just all the sequence of the samples $a_k=a(kT)$ ideally allows to faithfully rebuild the signal. A processing that starting from a(t) produces a signal in the similar, but different expression in time series:

$$[6.48]\ b(t) = \sum_k a_k \psi(t-kT),$$

with the same coefficients a_k and replacing the functions rect(t/T-k) with other real energy functions $\psi_k(t) \hat{=} \psi(t-kT)$, can still allow to faithfully rebuild the excitation signal but only if from the shape of b(t) it is possible in some way to recognize all the values a_k: then, the reversibility of the processing would be ideally guaranteed. It is important to underline that reversibility means the recovery of the information and not the recovery of the cut frequency components.
Following the same method of the representations of the energy signals through a base (see [3.14]), with the extension of the scalar product to the cross correlation between the power signal a(t) and the h^{th} energy function $\psi_h(t)$, exchanging the order of the integral and of the summation, it is possible to obtain:

$$[6.49]\ (b, \psi_h) = \int b(t)\psi(t-hT)dt = \sum_k a_k \int \psi(t-kT)\psi(t-hT)dt\,.$$

Then, if the set $\{\psi_k(t)\}$ is orthogonal, that means:

$$[6.50]\ \int \psi(t-kT)\psi(t-hT)dt = \int \psi(t+\nu T)\psi(t)dt = 0 \quad , \quad \text{per} \quad \nu \hat{=} h-k \neq 0\,,$$

each correlation (b,ψ_h) provides just the value a_h, although with the difference of a known coefficient $E_{\psi\psi}$ which corresponds to the energy of the function $\psi(t)$, from which all the other functions of the set $\{\psi_k(t)\}$ are derived by shift of kT.
The reversibility of the processing that leads to the shape [6.48] is thus ideally guaranteed by respecting the condition [6.50]; reminding the definition of the autocorrelation function of $\psi(t)$, the following property:

$$[6.51]\ C_{\psi\psi}(\nu T) = 0, \text{ per } \nu \neq 0,$$

corresponds to the orthogonality condition. It is important to note that in the mentioned conditions the [6.48] broadens the representation through a base to a power signal, but only if it is meaningful only with respect to the sequence of the values a_k carried by itself. Thus, usually the set $\{\psi_k(t)\}$ is considered a base and the function $\psi(t)$ is named ***base waveform***.
The reader interested to the corresponding reversibility condition in the frequency domain can see the section 6.3.1.2, to find out that there are no base vaweforms with upper bound of the bandwidth $f_{\psi M}$ which is lower than 1/2T. Since the processed signals b(t) are composed of a linear combination with constant coefficients of functions achieved from the base one $\psi(t)$ by time shifting, i.e. their bandwidth cannot be greater than $f_{\psi M}$, the following important constraint follows:

$$[6.52]\ f_{bM} = f_{\psi M} \geq f_0 \hat{=} \frac{1}{2T}\,.$$

The base waveform with minimum bandwidth has the expression:

$$[6.53]\ \psi_0(t) \hat{=} \text{sinc}(t/T);$$

in fact, the Fourier transform of such a function is $\Psi_0(f)=\text{Trect}(fT)$, that implies maximum frequency f_0, and the orthogonality of the shifted waveforms starting from sinc(t/T) is already well known (see section 2.5.1.6). Inserting the $\psi_0(t)$ in the [6.48], the well known representation by sample a_k interpolation, surely satisfactory, is obtained. Nevertheless, as discussed hereinafter, due to feasibility difficulties, the waveform $\psi_0(t)$ must be discarded in practical applications.
The above mentioned type of reversible processing is very meaningful in current practical transmission systems, in which the signals to transfer are even more of step type, with digital information: excluding the channels which utilize monomodal optical fibers, which offer behaviours practically perfect in very large bandwidths, those which utilize metallic pairs result upper bounded in bandwidth by the effect of the growing attenuation with frequency, while almost all the actual channels based on radio link can allow the transmisison only in assigned and well delimited bandwidths, often even with small extension.

6.3.1.2 Reversibility in the frequency domain

Let's consider a generic base waveform $\psi(t)$ and its autocorrelation function $C_{\psi\psi}(\tau)$, which thanks to the Wiener-Khinchine relation (see [3.107]) can be expressed by:

$$[6.54]\ C_{\psi\psi}(\tau) = \int E_{\psi\psi}(f)e^{j2\pi f\tau}\,df$$

where $E_{\psi\psi}(f)$ is the energy spectral density of $\psi(t)$.
Setting $f_T \hat{=} 1/T$ and indicating with:

$$[6.55]\ W(f) \hat{=} \mathrm{rep}_{fT}[E_{\psi\psi}(f)] = \sum_n E_{\psi\psi}(f-nf_T) = \sum_n E_{\psi\psi}(f-n/T),$$

the power spectral density obtained for replica at interval f_T of the energy spectrum $E_{\psi\psi}(f)$, it is possible to observe that W(f) is periodical in frequency, with period f_T, and thus can be represented in Fourier series:

$$[6.56]\ W(f) = \sum_\nu C_\nu e^{j2\pi\nu f/f_T},$$

with the coefficients:

$$[6.57]\ C_\nu = \frac{1}{f_T}\int E_{\psi\psi}(f)e^{-j2\pi\nu f/f_T}\,df.$$

Reminding that $f_T \hat{=} 1/T$ and the [6.54] the result is:

$$[6.58]\ C_\nu = T\int E_{\psi\psi}(f)e^{j2\pi f(-\nu T)}\,df = TC_{\psi\psi}(-\nu T);$$

then, from the [6.56] the following relation is achieved:

$$[6.59]\ W(f) = T\sum_\nu C_{\psi\psi}(-\nu T)\,e^{j2\pi f\nu T}$$

In conclusion, if the function $C_{\psi\psi}(\tau)$ in the time domain has the property [6.51], from [6.59] immediately results $W(f)=TC_{\psi\psi}(0)=TE_{\psi\psi}$, i.e. on the basis of the [6.55] the corresponding property in the frequency domain is valid:

$$[6.60]\ \sum_n E_{\psi\psi}(f-n/T) = TE_{\psi\psi},$$

named property of the vestigial symmetry.
Viceversa, if the latter is valid, that means if $W(f)=TE_{\psi\psi}$, being $C_{\psi\psi}(\tau)$ a real and even function from the [6.59] the next relation follows:

$$[6.61]\ TE_{\psi\psi} = TE_{\psi\psi} + 2T\sum_{\nu=1}^{\infty} C_{\psi\psi}(\nu T)\cos(2\pi f\nu T)$$

which implies the [6.51] because to be true the second term of the second member must be zero and therefore all the terms of $C_{\psi\psi}$ for $\nu\neq 0$ must be zero.
First of all, it is possible to note that the property of vestigial symmetry, being valid for every f, must be satisfied also at the particular frequency $f_0 \hat{=} 1/2T$ implying that the energy spectrum $E_{\psi\psi}(f)$ of a base waveform is necessarily extended at least up to f_0. Then, it is possible to demonstrate that there are no base waveforms with upper limit of the bandwidth lower than 1/2T; instead, there is no general condition for upper limitation the value assumed by $f_{\psi M}$, so that a base waveform can be unlimited in frequency. In the case in which $E_{\psi\psi}(f)$ is extended just up to f_0, the property of vestigial symmetry concerns just one addend for each frequency, in other words just one addend is present in the sum for each f; considering n=0 the energy spectrum results then to be constant in the interval $(-f_0;f_0)$ and is equal to T^2: thus, the base waveform with minimum bandwidth is the function $\psi_0(t) \hat{=} \mathrm{sinc}(t/T)$.
In the hypothesis that the energy spectrum $E_{\psi\psi}(f)$ is extended within $2f_0=1/T$, the property [6.60] concerns for every frequency only pairs of adjacent shifted spectra; for the value $f=f_0+\Delta f$ with Δf limited between 0 and f_0, the [6.60] assumes in fact the expression:

$$[6.62]\ E_{\psi\psi}(f_0+\Delta f) + E_{\psi\psi}(f_0+\Delta f-2f_0) = E_{\psi\psi}(f_0+\Delta f)+E_{\psi\psi}(\Delta f-f_0)=TE_{\psi\psi};$$

considering that for $\Delta f=f_0$ the result is $E_{\psi\psi}(0)=TE_{\psi\psi}$ and taking into account that $E_{\psi\psi}(f)$ is even the following equality is valid:

$$[6.63]\ E_{\psi\psi}(f_0+\Delta f) + E_{\psi\psi}(f_0-\Delta f) = E_{\psi\psi}(0),\ 0\leq\Delta f\leq f_0,$$

which expresses the property of the vestigial symmetry in the particular considered case.

6.3.1.3 Signals satisfying the Nyquist criterion

From the processed signal $b_0(t)$ that can be achieved, only theoretically, starting from the base waveform with minimum bandwidth $\psi_0(t)=\text{sinc}(t/T)$ it is possible to figure out the values a_k even by the simple observation of the instant values in hT; in fact the following relation is valid:

$$[6.64]\ b_0(hT) = \sum_k a_k \text{sinc}[(hT-kT)/T] = \sum_k a_k \text{sinc}(h-k) = a_h,$$

thanks to the particular property:

$$[6.65]\ \text{sinc}(\nu) = 0,\ \text{per}\ \nu \hat{=} h - k \neq 0.$$

The property [6.65] is applicable not only to the waveform sinc(t/T): for example the function x(t)sinc(t/T) can be considered, with x(t) energy signal, to verify that it has at least the same zeros in the instants νT for $\nu\neq0$. Identified a generic energy function q(t), which is different from zero in the origin and has the property:

$$[6.66]\ q(\nu T) = 0,\ \text{per}\ \nu \neq 0,$$

a signal represented by the following expresion:

$$[6.67]\ p_N(t) = \sum_k a_k q(t-kT),$$

still allows to figure out the values a_k by means of the simple observation of the instant values, because, at least in the instants hT, $p_N(hT)=a_h q(0)$ is obtained, as it is immediate to verify. The property [6.66] is usually indicated as ***Nyquist criterion*** in memory of the person who was the first to highlight it.

A signal of the type $p_N(t)$, constituted by a linear combination with constant coefficient of functions shifted of kT achieved starting from a function that satisfies the Nyquist criterion, is named ***signal satisfying the Nyquist criterion***. As the coefficients a_k are located in time in correspondence to the symbols of a digital sequence and can be individually distinguished in the evolution of the signal, usually it is stated that it is without ***Inter Symbol Interference*** (ISI).

Anyway, it is possible to note that in the type of signals introduced through the [6.67] no constraint has been expressed on the orthogonality of the shifted functions, that means that in general it is not satisfied. Nevertheless, signals of the type b(t) may exist, expressed in series of orthogonal shifted functions, in which simultaneously the base waveform respects the Nyquist criterion. The ideal reversibility of the processing that produces a signal of such a cathegory can thus be based also on the simplest direct observability of its instanteneous values.

6.3.2 Linear processing on step signals

6.3.2.1 Feasibility of linear shapers

Aiming to reduce the requirements in terms of bandwidth of the channel to use for the transmission of a step signal a(t), it is possible to resort to the linear processing with cut of the bandwidth, that can be performed by a LTI quadrupole with memory physically feasible. Considering the meaningful distorsion effect produced by the cut of the bandwidth, in particular with cutoff frequency significantly lower than the upper bound of the bandwidth, just practically limited, of the spectrum of the step signal, it is not worth to use just a low pass filter, but instead it is convenient to find out which types of developments of the quadrupole transfer functions must be

realised to effectively reduce the output bandwidth, getting close to the theoretical lower limit $f_0=1/2T$, but still keeping the reversibility of the processing.
Taking into account what above explained, if a function $\psi(t)$ could be identified satisfying the [6.50] and with Fourier transform $\Psi(f) \hat{=} F\{\psi(t)\}$ strictly limited with maximum frequency $f_{\psi M}$, a processing with cut of the bandwidth would be determined which ideally would allow the complementary operation. In this way, without the risk to miss information, from a step signal a(t) just practically limited in frequency, a processed signal b(t) with spectral components different from zero just in the finite bandwidth $(0; f_{\psi M})$ is obtained, being $f_{bM}=f_{\psi M}$.
Assumed, with immediate possibility to verify as soon as possible, that a linear processing with cut of the bandwidth of reversible type is possible, it is exploited putting in input a(t) to a LTI quadrupole with memory (see Figura 6.15), having a transfer function $H_F(f)$ equal to zero beyond the cutoff frequency $f_{\psi M}$. Thanks to the linearity, which defines always the same convolution relation between each pair of corresponding addends of the input and output series, the response results in the expression [6.48] but with a delay t_0, if the following relation is satisfied:

$$[6.68]\ \psi(t-t_0) = F^{-1}\{H_F(f)\} * \mathrm{rect}(t/T);$$

applying the Fourier transform then the result is:

$$[6.69]\ H_F(f) = \frac{1}{T}\frac{\Psi(f)}{\mathrm{sinc}(fT)}\, e^{-j2\pi f t_0}.$$

It is easy to verify that due to the presence of $\Psi(f)$ at the numerator the transfer function $H_F(f)$ is equal to zero for $|f|>f_{\psi M}$; moreover, since the denominator is equal to zero at the frequencies n/T, with n integer different from zero, the modulus $|H_F(f)|$ is verified to keep to be limited but respecting the condition that the maximum frequency $f_{\psi M}$ of the base waveform is lower that the frequency in which the first zero occurs, i.e. lower than 1/T.

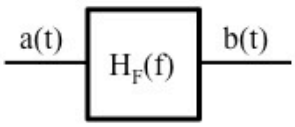


Figura 6.15: Functional scheme of a linear shaper.

Definitively, having verified the above assumption, reversible linear processing with cut of the bandwidth of step signals are possible, and a LTI quadrupole that implements it, named ***linear shaper***, is ideally feasible, as long as the upper bound of the base waveform is constrained by the limits (see also [6.52]):

$$[6.70]\ \frac{1}{2T} < f_{bM} = f_{\psi M} < \frac{1}{T}.$$

Thus, the linear shapers are significantly efficient for the task to obtain, respecting the reversibility, signals with reduced bandwidth.
A function $\psi(t)$ which other than respecting the condition of orthogonality of the shifted waveforms (see [6.50]) satisfies the [6.70] is a ***normalized bandwidth base waveform***.
Once demonstrated the feasibility of the linear shapers, it is worth to highlight an important characteristic of the signals provided by them. Reminding the effect of a linear transformation with memory (see [6.23]), in general the instant values of a generic signal b(t) linearly processed no longer depend on just one of the values a_k, as instead happens in every instant in a(t) and at least in one instant for each interval of duration T in the signals satisfying the Nyquist criterion, but they result for every t a linear combination of many of them. Therefore, it is usual to state that a signal b(t) going out from a linear shaper is usually affected by intersymbol interference, or that it is a signal with ISI. Such an impairment is the counterpart of the possibility to achieve a very effective reduction of the bandwidth of the processed signal.

6.3.2.2 Linear shapers of the raised cosine family

Looking for suitable normalized bandwidth base waveforms the orthogonality condition [6.50] is useful, once it is expressed in the frequency domain, in the considered case with $f_{\psi M}$ limited within 1/T. Inviting the interested reader to see the demonstration in section 6.3.1.2, the energy spectral density $E_{\psi\psi}(f)$ of a normalized bandwidth base waveform can be verified to be subject to the condition of vestigial symmetry, expressed by the:

$$[6.71]\ E_{\psi\psi}(f_0+\Delta f) + E_{\psi\psi}(f_0-\Delta f) = E_{\psi\psi}(0),\ 0\le\Delta f\le f_0.$$

The vestigial symmetry implies that the residual part of the spectrum, for every deviation Δf beyond the value f_0, compensates the missing part of the spectrum, for every deviation Δf within the value f_0, to reach the value in the origin. In the practice, one of the functions of the set $\{\psi_\alpha(t)\}$ is adopted as normalized bandwidth base waveforms, named ***raised cosine family***, which indeed enjoys the property of the vestigial symmetry; in fact, fixed a value between 0 and 1 of the parameter α, named roll-off factor, the normalized energy spectra $E_{\alpha\alpha}(f)/T^2$ take unit value for $|f|<1-\alpha$, are equal to zero for $|f|>1+\alpha$ and between such two extremes they follow the typical evolution of the cosine function in the first half period, raised by one, as shown in the example in Figura 6.16a. Remind that the evolution as a raised cosine concerns the energy spectral density and not the function in the time domain $\psi_\alpha(t)$.

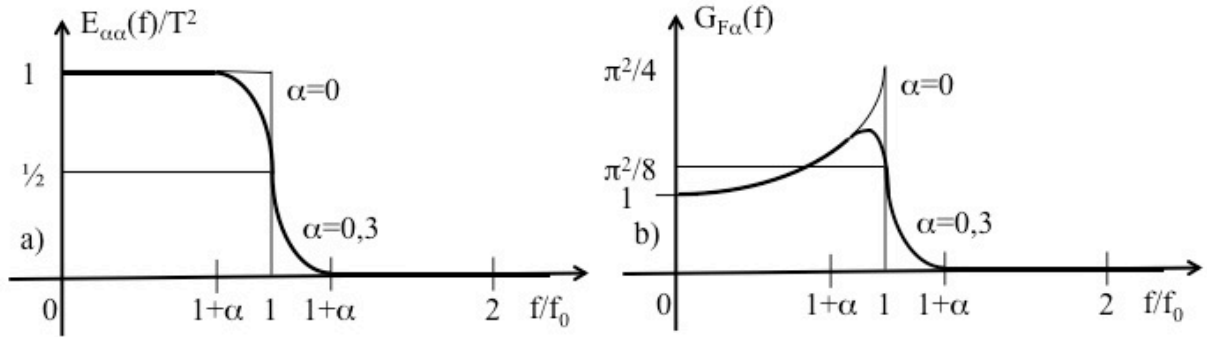


Figura 6.16: Energy spectra of base waveforms of the raised cosine family (a) and gain of the relative shapers (b).

The upper bound of the spectrum of a base waveform of the raised cosine family takes the following expression as a function of the roll-off factor α:

$$[6.72]\ f_\alpha \triangleq f_{\psi\alpha M} = \frac{1+\alpha}{2T} = (1+\alpha)f_0.$$

Note that also the function sinc(t/T), achieved for the value $\alpha=0$ and for this reason it is just indicated as $\psi_0(t)$, belongs to the considered family); in fact, it has energy spectral density $E_{00}(f)=T^2\mathrm{rect}(fT)$ (see thin line in Figura 6.16a) and assumes the minimum allowable value of the bandwidth $f_\alpha=f_0$.

The gain of the linear shaper relative to a waveform $\psi_\alpha(t)$ of the raised cosine family takes the expression:

$$[6.73]\ G_{F\alpha}(f) = \frac{1}{T^2}\frac{E_{\alpha\alpha}(f)}{\mathrm{sinc}^2(fT)};$$

its evolution is sketched in Figura 6.16b for the two values $\alpha=0$, case of the waveform $\psi_0(t)$, and $\alpha=0.3$ of the roll-off factor.

Inserting the rectangular energy spectrum of the $\psi_0(t)$ in the denominator of the [6.73], the gain $G_{F0}(f)=|H_{F0}(f)|^2$ of the corresponding linear shaper is determined (see thin line in Figura 6.16b), which shows a strong discontinuity in correspondence to the cutoff frequency f_0. The huge complexity that as a consequence should be faced to build the actual subsystem having behaviour

that adequately approximates the one of the shaper quadrupole corresponding to the waveform with minimum bandwidth, implies to not opting for $\psi_0(t)$ in practical applications.
The performance of linear shapers of the raised cosine family, excluding the case with $\alpha=0$ already examined, can be approximated with actual subsystems; for values of α just slightly different from zero, the transitions of $G_{F\alpha}(f)$ around the cutoff frequency f_α are very light and don't represent a difficulty in construction. Finally, it is possible to find out that anyway it is not physically possible to get a gain exactly equal to zero in the stop band; otherwise, the pulse response of the shaper would be time unlimited and the system would be not causal. In fact, the maximum allowed value of the delay, although it may be very large, it is always finite.

6.3.2.3 Matched filters

From a signal b(t) obtained starting from a generic base waveform $\psi(t)$ which respects the condition of orthogonality of its shifted replicas (see [6.50]), it is possible to carry out with a clock (timer) kT the samples a_k of the originating step signal a(t) with a physical system which, at least with good approximation, is capable to perform the due correlations (b, ψ_k) (see [6.49]). Nevertheless, the system results to be very complex to be realized.
Let's remind that the autocorrelation function $C_{\psi\psi}(\tau)$ of a base waveform in general satisfies the Nyquist criterion (see [6.51]); then, it would be possible to identify the values a_k of a linearly processed signal b(t), with the observation of the instant values of the signal with no ISI:

$$[6.74]\ b_N(t) = \sum_k a_k C_{\psi\psi}(t-kT),$$

that can be obtained after having processed b(t) to replace each of the shifted functions $\psi(t-kT)$ with the corresponding function $C_{\psi\psi}(t-kT)$. In fact, thanks to the mentioned property the following relations are valid:

$$[6.75]\ b_N(hT) = \sum_k a_k C_{\psi\psi}(hT-kT) = \sum_k a_k C_{\psi\psi}(\nu T) = a_h E_{\psi\psi},$$

where $E_{\psi\psi}=C_{\psi\psi}(0)$ is the known energy of the base waveform $\psi(t)$.
Going ahead as in the case of the linear shapers, the desired processing, which leads to the signal satisfying the Nyquist criterion, is demonstrated to be linear, which implies that can be implemented passing through a LTI quadrupole (see Figura 6.17); the transfer function $H_\psi(f)$ of this quadrupole, that can be achieved executing the convolution:

$$[6.76]\ C_\psi(t-t_0) = F^{-1}\{H_\psi(f)\} * \psi(t),$$

has the expression:

$$[6.77]\ H_\psi(f) = \frac{E_{\psi\psi}(f)}{\Psi(f)}\, e^{-j2\pi f t_0} = \Psi^*(f)\, e^{-j2\pi f t_0},$$

being $F\{C_\psi(t)\}=E_{\psi\psi}(f)=\Psi(f)\Psi^*(f)$. It is possible to verify immediately that the transfer function $H_\psi(f)$ is equal to zero for $|f|>f_M$ and that the identified LTI quadrupole, for its chacteristic function named filter matched to the base waveform, or simpy ***matched filter***, is ideally feasible.

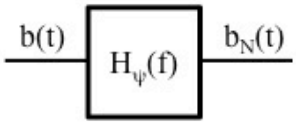


Figura 6.17: Functional scheme of the matched filter.

The gain of the matched filter in the case of a waveform $\psi_\alpha(t)$ of the family with raised cosine has the simple expression:

$$[6.78]\ G_{\psi\alpha}(f) = E_{\alpha\alpha}(f);$$

its evolution is therefore of the type that is exemplified in Figura 6.16a for the two values $\alpha=0$, case of the waveform $\psi_0(t)$, and $\alpha=0.3$ of the roll-off factor. For values of α that are slightly

different from zero, around the cutoff frequency f_α transitions even more delicate than the ones of the linear shaper occur, causing just a few difficulties in the construction of the matched filter. Finally, it is possible to still find out that anyway it is not physically possible to get a gain $G_\psi(f)$ exactly equal to zero in the stop band; otherwise, the pulse response of the matcged filter would be time unlimited and the system would be not causal. In fact, the maximum allowed value of the delay, although it may be very large, it is always finite.

6.3.3 Non linear processing of step signals

The processing starting from a step signal a(t) to provide a different representation (see [6.48]) in series of energy functions obtained shifting by kT a base waveform, becomes conceptually very simple if the waveform is at least practically limited in time with duration lower than T. In fact, in such a case, that is highlighetd using for the base waveform the different notation $\phi(t)$ and expressing the the processed signal as follows:

$$[6.79]\ b_\phi(t) \hat{=} \sum_k a_k \phi(t-kT),$$

the condition of orthogonality which guarantees the reversibility (see [6.50]) is satisfied immediately, thanks to the time separation of all the $\phi(t-kT)$, without considering any contratint on the evolution of $\phi(t)$.

Thanks to the limited duration within T, the waveform $\phi(t)$ surely ejoys, still with no constraints on the evolution, the property indicated as Nyquist criterion (see [6.66]), and as a consequence all the waveforms like $b_\phi(t)$ are signals satisfying the Nyquist criterion, that means with no ISI. Thus, the considered processing produces signals expressed in series of orthogonal shifted functions, in which simultaneously the base waveform respects the Nyquist criterion: the ideal reversibility can thus be based on the simplest direct observability of the instant values.

As counterpart of the above highlighted favourable characteristic, in every processed signal $b_\phi(t)$ the inconvenient of the theoretically infinite extension of the bandwidth is still present, but from a practical point of view an upper finite limit $f_{\phi M}$ can be assumed, which fixed T results to be as lower as less the discontinuities of the waveform $\phi(t)$ and of its derivatives of any order are. With reference to the already shown examples, the practical bandwidth is reduced passing from the rectangular waveform (see Figura 2.9a) to the triangular pulse with no discontinuities (see Figura 2.9b) and even to the raised cosine pulse (see Figura 2.9c), which is continuous together with its derivative. In a different way it is possible to state that the distorsion effect produced on the processed signal [6.79] due to the suppression of the spectral components beyond the frequency:

$$[6.80]\ f_{\phi TM} = \frac{2}{T} = 4f_0,$$

which would be meaningful in the case of rectangular step signal, becomes tolerable for signals processed with triangular pulse and negligible with the raised cosine one. Comparing the practical upper bound $f_{\phi TM}$ with the theoretical lower limit f_0, the achievable bandwidth reduction, although with little distorsion, is anyway significant.

The processed signals, obtained with the base waveforms with limited duration without discontinuities at the ends of the interval T, have time developments which are equal to zero at least in every instant of separation between the addends of the series [6.79], which expresses the signals themselves. Thus, they are named signals with return to zero, or simply ***signals RZ*** (RZ=Return to Zero); in Figura 6.18 an example of digital signal of such a type is reported. Instead, a step signal a(t) is equal to zero just in correspondence of transitions between levels of opposite sign: therefore, it is also indicated as ***signal NRZ*** (NRZ=No Return to Zero).

A signal processed with base waveforms $\phi(t)$, with duration at least practically limited within the interval T, can be obtained multiplying the step signal a(t) by the periodical signal $rep_T[\phi(t)]$, achieved by a pulse train generator named pulse carrier; in fact, the following relation is valid:

$$[6.81]\ a(t)rep_T[\phi(t)] = \sum_k \sum_h a_k rect[(t-kT)/T]\phi(t-hT)] = \sum_k a_k \phi(t-kT)],$$

considering that the the limitation of the duration within T ensures that the products of the rectangular functions by the base ones are zero for different time shifts and carry out the base waveform itself, when both have been shifted of the same time interval (h=k). Thus, a non linear subsystem which carries out RZ signals from a functional point of view can be represented with the scheme of the ***RZ shaper*** shown in Figura 6.19.

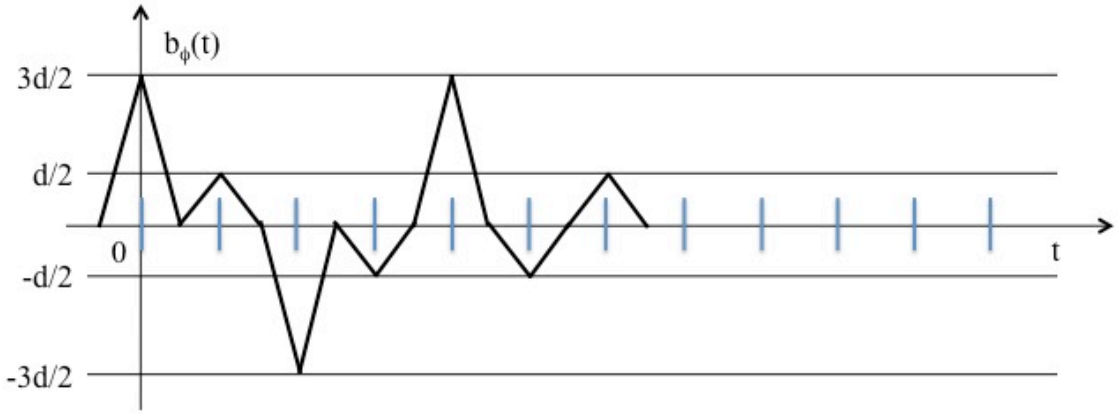


Figura 6.18: Example of digital signal of type RZ.

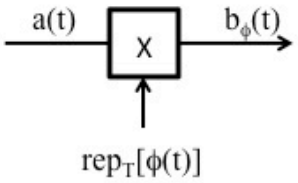


Figura 6.19: Functional scheme of the RZ shaper.

The bandwidth requirement of the RZ signals expressed by the [6.80], about three times larger than the one necessary utilizing a linear shaper with the same T (see [6.72]), in many cases is acceptable, particularly when the transmission means is a metallic pair; in fact, the product shaper's cost is similar to the linear shaper cost and, moreover, the absence of ISI allows to carry out the significant values a_k from the simple observation of the instant values of the processed RZ signals, with no need to put matched filters before.

6.3.4 Return of the step waveform

6.3.4.1 Step shaper

Aiming to facilitate the transmission of the generic step signal a(t), some processing has been introduced, with linear shapers and matched filters as well as with non linear RZ shapers, being confident on their ideal reversibility guaranteed by the possibility to recognize anyway all the values a_k, significant of a(t), in the processed signals. Moreover, it was highlighted that they can be achieved by means of the simple observation of the instant values of the signals at interval T, as long as these signals have been shaped or processed again respecting the Nyquist criterion (see [6.66]), i.e. they are with no ISI.
Thus, the complementary processing with respect to the above mentioned ones is the operation that acting on the signals satisfying the Nyquist criterion provides at the output the originating step waveform, but differing for a real factor g and a positive delay t_0.
The return of the step waveform a(t) starting from a signal with no ISI can be performed utilizing a quadrupole named ***step shaper***, hereinafter indicated in the schemes with a block having as symbol Π. Such a quadrupole cyclically samples the input signal, actually it observes the instant values, both if they are $b_N(hT)$ or $b_s(hT)$, in the reading instant $t_h=hT$, caring then in the output signal y(t) to hold such samples for a duration as large as T itself, as shown in Figura 6.20 and expressed by:

$$[6.82]\ y(t) = \sum_h b_x(hT)\,\mathrm{rect}\left(\frac{t-t_0}{T}-h\right),\ \text{per } x = N, \phi.$$

The reader interested to further technical elements on the sampling can see section 6.3.4.2.

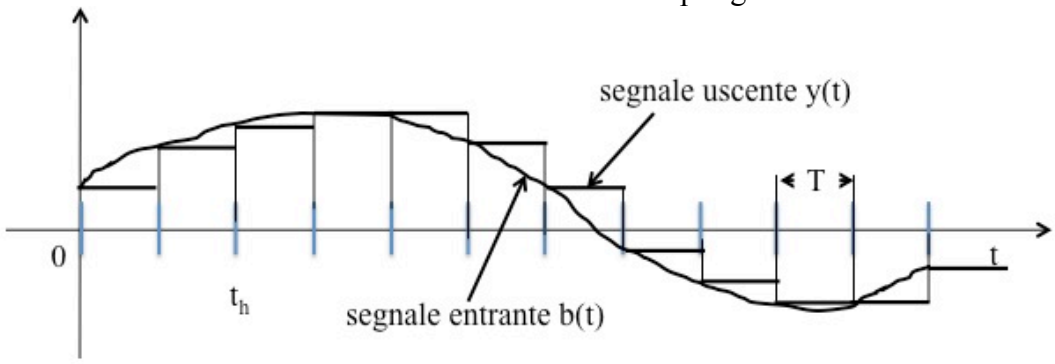


Figura 6.20: Step shaper, input signal and output signal.

It is meaningful to note that aiming to the correct return the step shaper must be synchronized with the input signal, in the sense that not only the sampling must occur with interval T_c equal to the symbol interval T of the signal, or the condition T_c=T must be respected, but also the reading instants must be positioned in each interval in the most opportune moment, in which the signal doesn't suffer from ISI. Therefore, in addition to the circuits that implement the sampling and hold, the considered processor must include a harmonic oscillator at the frequency $1/T_c$, nominally equal to the symbol rate 1/T, which generates a ***synchrosignal*** (clock signal) with control functions with respect to the clock devices of the sampler, and an auxiliary circuit that synchronizes the oscillator after having extracted the necessary elements from the timing of the input signal.

6.3.4.2 Elements on the samplers

The ideal sampling operation isolated at the reading instant t_0 of a time continuous signal x(t) consists in multiplying this signal by the function $\delta(t-t_0)$, generating the pulse $x(t)\delta(t-t_0)$. By means of a further integration processing, the isolated sampling with unlimited hold is achieved, which produces at the output the step signal:

$$[6.83]\ \int_{-\infty}^{t} x(v)\delta(v-t_0)\,dv = x_0 u(t-t_0),$$

in which the value of the isolated sample of the signal, $x_0=x(t_0)$, is evident. With the scheme in Figura 6.21, in which the first of the two operations is made similar to a switch that is supposed to close the circuit only at the instant t_0 for a very short duration Δt, the isolated sampling with hold is approximated.

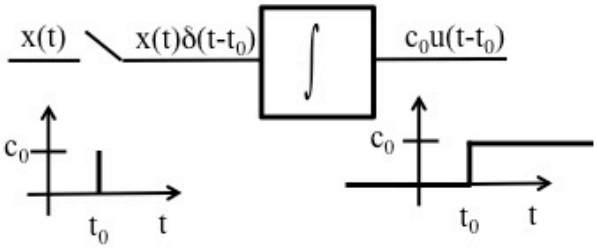


Figura 6.21: Scheme of isolated sampler with unlimited hold.

Inserting on the output of the integrator a second switch which closes to ground only at the instant t_0+T, as shown in Figura 6.22, an isolated sampler with hold of duration T is obtained, approximating at the output the rectangular shape:

$$[6.84]\ x_0[u(t-t_0)-u(t-T-t_0)] = c_0\,\mathrm{rect}\left(\frac{t-t_o}{T}-\frac{1}{2}\right).$$

Finally, activating the switches in Figura 6.22 repetitively at interval T, the cyclic sampling with hold for a duration T is obtained.

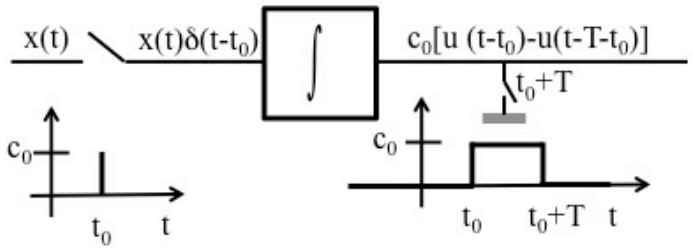


Figura 6.22: Scheme of isolated sampler with hold for a duration T.

6.3.5 Total processing with cut of the band

Concerning the ideal transmission of the signal a(t), but with cut of the band which allows to minimize the bandwidth requirements of the channel, the return to the step shape is carried out after a matched filter, that provides a signal $b_N(t)$ with no ISI. Indicating with t_0 the delay of the step shaper, thanks to the [6.75] the desired signal at the output of the step shaper is obtained:

$$[6.85]\ y(t) = \sum_h b_N(hT)\,\mathrm{rect}\left(\frac{t-t_0}{T}-h\right) = \sum_h a_h E_{\alpha\alpha}\,\mathrm{rect}\left(\frac{t-t_0}{T}-h\right) = E_{\alpha\alpha}a(t-t_0),$$

where the factor E_α corresponds to the known energy of the base waveform $\psi_\alpha(t)$, of the raised cosine family, utilized in the previous processing.
Differing for meaningless delays, the configuration of the total processing with cut of the bandwidth of step signals shown in Figura 6.23 is achieved; it is possible to note from it that the set composed of the matched filter and of the step shaper constitutes a subsystem that, inserted immediately after the linear shaper, implements the inverse processing, providing a signal at the output in faithful shape with respect to the one at the input of the chain. Broadening a concept already introduced (see [6.32]), the two functional blocks in Figura 6.23a and in Figura 6.23b are thus a pair of complementary quadrupoles.

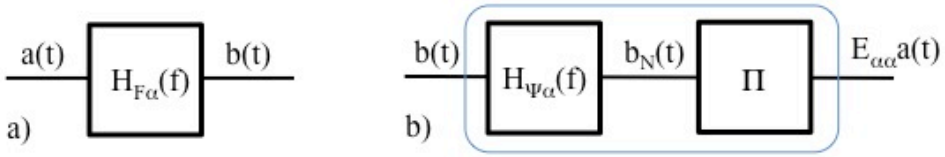


Figura 6.23: Linear shaper (a) and complementary quadrupole (b), composed of the corresponding matched filter and of the step shaper.

In a signal $b_N(t)$ with limited bandwidth, within the maximum frequency $f=(1+\alpha)/2T$ of its base waveform of the raised cosine family, the absence of intersymbol interference is in theory ensured only in one instant in each interval of duration T and in pratice strictly nearby. If the signal is observed on the screen of an oscilloscope with permanence time of the trace much higher than T and in which the time horizontal axis is synchronized with the reading instants (at the center of the screen), in the case of just two possible values of a_k an image of the type shown in Figura 6.24a is obtained; the shape is such to justify the usual name of ***eye diagram***. Note that the values of the levels a_1 and a_2 are neither maximum nor minimum points of the signal.

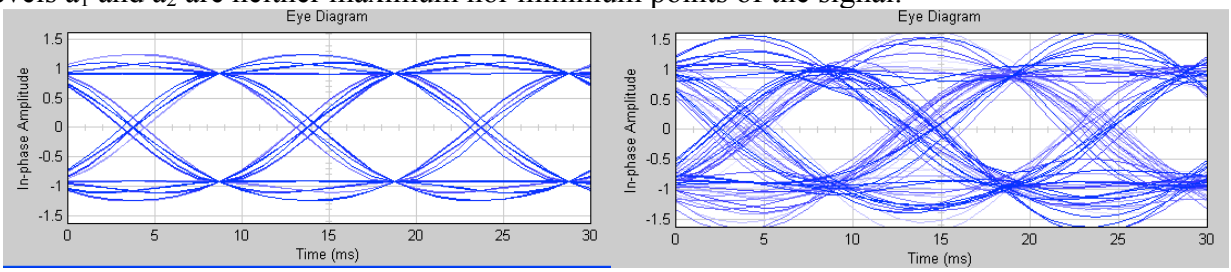


Figura 6.24: Eye diagrams for binary modulation: ideal case (a) and with intersymbol interference (b).

The Figura 6.24a points out that the synchronization of a step shaper is critical, also in the theoretical case in which the maximum "aperture of the eye" occurs, equal to the distance between the levels. Any deviation of the linear shaper and of the matched filter with respect to the theoretical performance, or impairments in the transmission among such subsystems, causes distorsion with consequent occurrence of intersymbol interference, which as shown in Figura 6.24b is pointed out by a reduction of the eye aperture, as more meaningful as stronger the mentioned phenomena are.

6.3.6 Total processing with reduction of the practical band

In the ideal transmission of the signal a(t) without cut of the band, but with a RZ signal with such a shape to well tolerate the suppression of the spectral components at the higher frequencies that can be caused by a channel, the return to the steps is performed just after it, which provides a signal $b_\phi(t)$ with negligible distorsion, practically still with no ISI. Indicating with t_0 the delay of the step shaper, as from the [6.79] the instant values $b_\phi(hT)=a_h\phi(0)$ are carried out, the desired signal is obtained at the output of the step shaper:

$$[6.86]\ y(t) = \sum_h b_\phi(hT)\,\mathrm{rect}\left(\frac{t-t_0}{T}-h\right) = \sum_h a_h\phi(0)\,\mathrm{rect}\left(\frac{t-t_0}{T}-h\right) = \phi(0)a(t-t_0),$$

where the factor $\phi(0)$ corresponds to the value for t=0 of the base waveform $\phi(t)$, with duration at least practically equal to T.
Differing for meaningless delays, the configuration of the total processing with reduction of the practical bandwidth of step signals is achieved as shown in Figura 6.25; from the figure it is possible to note that the step shaper constitutes a subsystem which, inserted immediately after the RZ shaper, implements the inverse processing, providing at the output a signal in faithful shape with respect to the input one to the chain. The two functional blocks in Figura 6.25a and in Figura 6.25b constitute thus a pair of complementary quadrupoles.

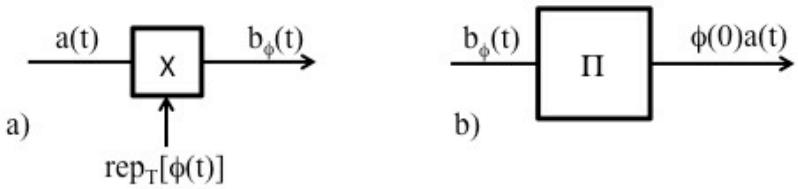


Figura 6.25: RZ shaper (a) and complementary quadrupole (b), consituted by the step shaper.

In a signal $b_\phi(t)$ obtained with a base waveform of duration T the absense of the intersymbol interference is in theory always ensured, but it is obvious that it is more convenient to establish the reading instants corresponding to the maximum value assumed by $\phi(t)$, which is supposed to be $\phi_M=\phi(0)$. Considering that the value of the shape varies around its maximum, the synchronization of the step shaper is still critical, differently from the case in which the step shaper were applied on a signal also step shaped, as often happens to perform read/write operations in the frame of the equipment of the information technology.

6.4 Multiplexing

Applying more widely the concept of the complementary transformations, which means that a reversible operation is followed by another operation capable to give back the starting waveform (in case differing just for a multiplying factor and for a delay), a further set of signal processing can be introduced.
In fact, in addition to the processing already mentioned in section 6.1.4 and anyway executed on a single signal, usually of the same type, in the transmission the signals frequently are subject to

operations which involve a set of several signals. These operations are, from the transmitting side the ***multiplexing*** operation, which transfers the carried information by N independent signals $x_i(t)$, at the input of the same number of physical lines and named ***contributor signals***, in a reversible way on a single developed x(t) on a single physical way, named ***multiplexed signal***; on the receiving side the complementary operation is exploited, ***demultiplexing***, which restores the individuality of the signals. The Figura 6.26, neglecting the meanigless constants g and t_0, shows a transmission system to which a pair of complementary equipment, ***multiplexer*** (Mux) and ***demultiplexer*** (Demux), is added.

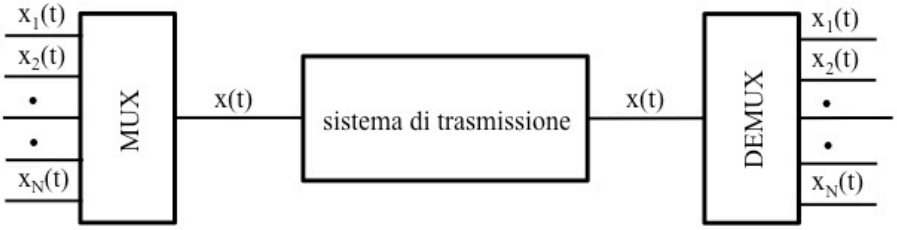


Figura 6.26: Scheme of a transmission system of multiplexed signals.

The aim of the further processing techniques introduced is to transform the signals in more advanced forms, particularly concerning the quantity of the transported information, which allow the best technical-economical use of the transmission means; thus, the operations are dependent on the actual performance offered by the transmission means. The introduction of the multiplexing provides a greater cost of the equipment per transmitted signal (the additional direct contribution of the mux-demux is usually greater than the decrease obtained replacing N equipment dimensioned for the contributors with the only one dimensioned for the multiplexed signal); on the other hand it leads to a significant save of the cost of the transmission means per transmitted signal and per lenght unit. As a consequence, greater the lenghts of the links are, more sofisticated the multiplexing processing that can be conveniently introduced are; moreover, the low cost use of the multiplexing is supposed to be dependent on having excessed opportune minimum distances. In the practice the multiplexing is always used in the transport links.
Typically the multiplexing consists in the sharing of a single transmission system (transmission means and equipment) identifying as resource to share the frequency, the time or the code. Thus, respectively the techniques are named ***frequency division multiplexing*** (FDM), ***time division multiplexing*** (TDM) or ***code division multiplexing*** (CDM). Instead, when the requirement is to set up N links on the same track, in alternative to the multiplexing which would allow to share a single transmission system, N distinct carrier means must be available. In such a case the simple ***space division multiplexing*** (SDM) is realized, used in practice only in very short access links.

6.5 SHORT INTRODUCTION TO THE ANALOGUE TO DIGITAL CONVERSION

Considering that almost all the operational transmission systems adopt the digital technique the analogue source signals must be necessarily transformed in digital signals. The transformation consists in three main steps. The sampling in the time domain of an analogue signal x(t) strictly limited in bandwidth can be considered as a first step towards its ***analogue to digital conversion***. Starting from the sequence x(n) of the samples $x_k=x(kT_{ca})$, obtained with sampling interval T_{ca} not greater than the Nyquist interval, in fact the step signal can be achieved:

$$[6.87]\ x_{ca}(t) = \sum_k x_k \,\mathrm{rect}\left(\frac{t}{T_{ca}} - k\right),$$

formally similar to a multilevel signal, but with the presence of the analogue samples x_k as meaningful difference, which assume values in a finite and continuous codomain (x_m; x_M), instead of the discrete levels. Thus, it is necessary to proceed to the ***quantization*** processing of the samples that, providimg them the discrete nature represents actually the core of the digital conversion.

The quantization is a processing on the sequence x(n) of the samples which consists in the partitioning of the codomain (x_m; x_M) in M parts, named ***quantization intervals***, and in the identification of all the values that belong to the q^{th} interval with just one of such values x_q, named q^{th} ***quantization level***. At the end, instead of the sequence of the analogue samples, a sequence of discrete values is achieved, and the following multilevel signal is put in correspondence to it:

$$[6.88]\ x_q(t) = \sum_k x_{qk}\, \mathrm{rect}\left(\frac{t}{T_{ca}} - k\right),$$

where the quantized levels $x_{qk} \cong x_k$ belong to a discrete, countable and finite set $\{x_q\}$ with M cardinality, that can be interpreted as M-nary alphabet of symbols.
The quantization can be uniform (x_{qk}-$x_{q(k-1)}$ = constant) or uneven in the case the analogue signal shows greater variation dynamics around some values of the codomain (range).
Note that while the sampling operation is reversible, the quantization is not reversible as well: in fact, it is no longer possible obtain back exactly the values of the samples x_k, once the quantized levels x_{qk} are known. The differences x_{qk}-x_k introduce an additive undesired signal, named ***quantization noise***, conceptually unavoidable: as a consequence, the analogue-digital (A/D) conversion cannot be considered an ideal processing. Nevertheless, it is intuitive to find out that the degradation implied by the quantization can be restrained within limits as small as desired, selecting a value of M sufficiently high.
As concerns the cardinality, once adopted the opportune value that can be expressed as:

$$[6.89]\ M = 2^S,$$

with exponentiation S integer, the last step to get the conversion of the analogue signal in a digital binary signal can be implemented with the same method of the complementary processing of the simple symbol coding (see section 6.3.4.1). From the symbol interval T_x=T_{ca} of the multilevel signal $x_q(t)$ then the bit interval T_a=T_{ca}/S of the corresponding binary signal a(t) at the output of the whole analogue-binary conversion operation is carried out.
For example, fixed the correspondence shown in Tabella 6.1 between the quantized levels x_q and the binary strings $y_1y_2y_3$, for q=1,2,..,M=8, and making the bit 0 correspond to the level a_1=0 and the bit 1 the level a_2=A, the part of the binary signal shown Figura 6.27c is obtained, while in Figura 6.27a and in Figura 6.27b the corresponding elements of the sequence of the quantized levels and of the binary flow are shown.

Tabella 6.1: Fixed correspondence between quantized levels and binary strings, for M = 8.

q	1	2	3	4	5	6	7	8
$y_1y_2y_3$	000	001	010	011	100	101	110	111

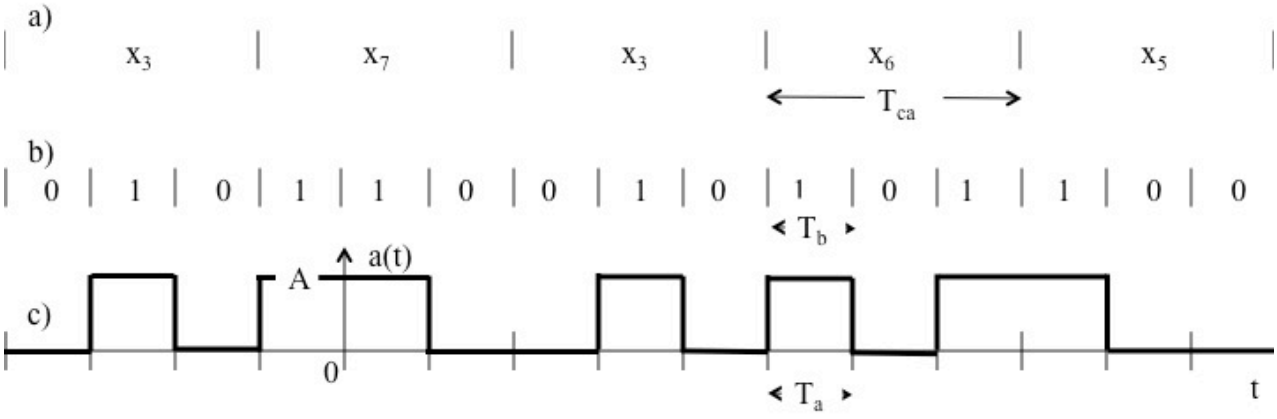


Figura 6.27: Example of a part of a binary signal (c) and the corresponding elements of the sequence of the quantized levels (a) and of the binay flow (b).

The analogue-digital conversion is largely applied, especially on sound source signals and on image source signals. In particular, in the case of telephone signal the sampling is performed with interval T_{ca}=125μs (f_{ca}=8 kHz), the quantization of the samples is performed with non uniform intervals in the codomain (x_m; x_M) and adopting the cardinality M=256 (S=8), and in conclusion the transformation in binary signal, indicated as ***telephone signal PCM*** (PCM = Pulse Code Modulation) is performed with the bit time that results T_b=15.625 μs, which corresponds to the bit rate R_b=64 kbit/s.

6.6 SHORT INTRODUCTION TO THE CHANNEL CODING

6.6.1 Coding with redundancy on the binary flows

6.6.1.1 Introduction to the methods of coding with redundancy

At the ends of a digital transmission system often some operations concerning the input and the output sequences, supposed to be binary, are performed leaving unaltered the physical form, of the step type with two levels, of the signals that carry them. Such a type of processing, as a matter of fact of logical nature, named ***coding***, consists in the preliminary processing, with microelectronics circuits, of the sequence x(n) of binary digits associated to the input physical signal x(t) with the aim to generate a different sequence y(n), still binary, that then is transmitted. In order to be able to perform the complementary processing of ***decoding***, which at the other end of the transmisison system gives back the originating sequence, the number of binary digits per unit of time carried by the signal y(t) associated to y(n) must be necessarily equal or greater to the one at the input. In the case, hereinafter considered, of the ***coding with redundancy***, the processed sequence is modified mainly because it carries more bits than those strictly necessary, with consequent redundancy of the information.
The coding with redundancy can be inserted in a digital transmission system with the general aim to allow the improvement of the achievable performance in the case of imperfect transmission channel. In fact, in such real conditions it is possible that some elements of the sequence are affected by errors; then, the decoder placed at the end of the system can be able to recognize such events and also to remediate, so that the actual ***bit error rate*** (BER=Bit Error Rate) is lowered, that is the value of the ratio between the number of error bits and the one contained in a very long, but finite, string of the decoded sequence. Such a measurable quantity corresponds “a posteriori” to the ***bit error probability***, which allows to evaluate “a priori” the behaviour of an imperfect digital transmission system.
Thus, the coding with redundancy has the general objective of the ***error detection*** and also of the ***error correction***. The functionality is based on the fact that the introduction of redundant elements confers to the processed sequence y(n) particular properties, that can be recognized after the transmission: in the event the expected characteristics are not observed the error may be detected. If an erroneous bit is detected in a finite length string, but not knowing exactly the position, two possible actions can follow: either the mere indication that the string contains errors or the request of retransmisison of the string itself, which once satisfied successfully implies the cancellation of the inconvenience. Considering more the transport to long distance of huge information flows than the requirements of the transmission concerning single applications, the coding techniques capable to correct errors are more interesting: to remove the errors is immediate, simply changing the erroneous binary digits with the complementary ones, if it is possible to exactly know the position of the detected errors. A solution that works in the considered way is named ***Forward Error Correction*** (FEC).
The introduction of redundant binary digits in the processed flow implies the variation of the timing at the output with respect to the input, i.e. $T_v<T_x$ is achieved, being T_v the bit time in output and T_x that at the input to the coder. The amount of the introduced redundancy is usually measured through the ***coding ratio***, defined by:

$$[6.90]\ R_c \triangleq \frac{k}{n} < 1,$$

where n is the number of binary digits of the output flow corresponding to a number k of those before the processing. Being for time continuity $nT_v = kT_x$, the redundancy causes an increase of the bit rate of the signal y(t) which carry the coded sequence with respect to the input value $R_x = 1/T_x$; in fact, the result is:

$$[6.91]\ R_y = \frac{1}{T_y} = \frac{n}{kT_x} = \frac{n}{k} R_x = \frac{R_x}{R_c} > R_x.$$

Such an increase, or better the decrease of the bit rate in input for the same value R_v accepted by the transmission system, is the counterpart to the performance improvement in terms of achieved BER.

Hereinafter two main methods are briefly introduced: the ***block coding*** and the ***convolutional coding***. The two methods are nor alternative; other than as stand alone, they can be jointly used concatenated, that means one outer and one inner.

6.6.1.2 Introduction on linear block coding

In the case of the ***linear block coding***, on the basis of the k elements of the binary sequence x(n), exctracted from the input flow at every interval of constant duration kT_x, the coder arranges in a register with k stages a string $x_1x_2...x_k$ of the same number of bits, named ***block***. The most significant part of the operation consists in the processing, through linear combinations, of every input block in an output string $y_1y_2...y_n$, named ***code word***, composed of a number n of bits greater than k (see Figura 6.28); the considered block coding is thus identified through the pair of values (n,k).

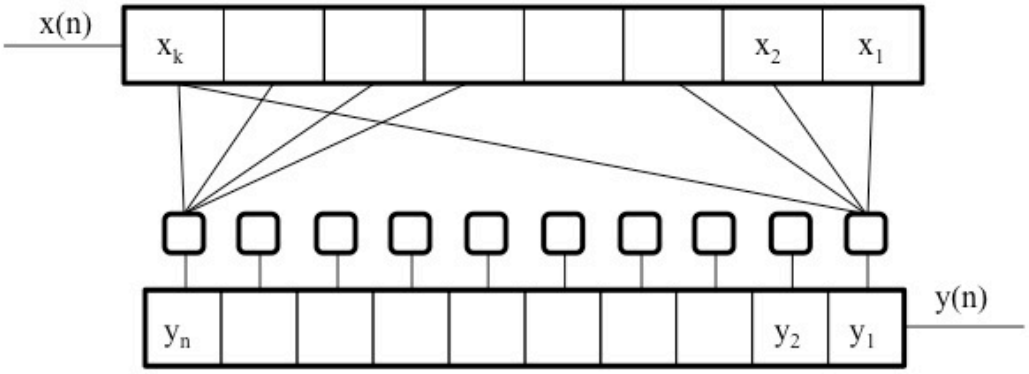


Figura 6.28: Input block and corresponding output code word.

A block coding is performed on the basis of a predefined rule of biunivolcal association between every different possible block in input and each output code word, that as a matter of fact corresponds to a well defined structure of the logical network, placed in between the two registers in Figura 6.28, which implements the desired linear combinations. The operational mode, invariant for the different blocks on which it is applied, produces a result always independent on the values of the binary digits not included in the current block; therefore, this processing is defined as memoryless coding. Note that the processed sequence results to be structured in strings of lenght n.

The number of different possible blocks is equal to 2^k and the different code words, biunivocally associated to the blocks by the particular selected coding rule, are of the same quantity; Considering that the different 2^n configurations of a string of n bits are many more than the possible blocks, there is margin to optimize the rule. Taking conveniently advantage of the presence of the $n-k$ redundant bits it is possible to maximise the cases in which the introduction after the imperfect transmisison of a single error (or more than one) provides a received string which is one of the $2^n–2^k$ configurations that don't belong to the code: the string affected by errors, which is not found in the set of all possible strings resulting having applied the association rule, is

detected in the decoder and, if the redundancy is sufficient, also the position of the erroneous bit (or more than one) within the string can be identified, with consequent forward correction.
As a counterpart of the performance improvement achieved in terms of BER, the linear block coding causes two impairments. The former, and more meaningful, is due to the quicker timing of the coded output sequence ($T_v<T_x$), already pointed out in the [6.91]. The latter and less meaningful impairment consists in an additional delay in the information transfer, surely not lower than $2kT_x$: even supposing to neglect the processing time to generate each code word (in transmission) or each block (in reception) and to realize the output signal which carries the elements in time series, in fact to extract from the input signal and to put in a register all the bits composing the block (in transmission) or the binary code word (in reception) neccessarily requires time that is mandatory to spend.
In the frame of the linear block coding, some details are provided on the type named ***systematic coding***, which has the minimum complexity with the same performance. The first k binary digits of the code words correspond in the same order to those ones of the corresponding blocks ($y_i=x_i$, for i=1, 2, .,,, k) and the remaining $n-k$, named digits for the parity check, are obtained from the previous ones by means of linear combinations, in arithmetic modulus-2, opportunely defined; for example, in the case of only one redundant bit the criterion can be to have the total number of digits, in the code words, equal to 1 either even or odd. Even assuming $n-k=1$, it is evident that every error introduced by the imperfect transmission which is unique within a code word can be detected, but not corrected, with a little complexity but at the cost of increasing the bit rate of a factor $(k+1)/k$ (see [6.91]). To get the capability to corerct errors a significant number of parity check digits is necessary, which implies high values of k, and thus with meaningful delays, if the coding rate is not worth to be decreased or if the aim is to not increase too much the bit rate R_v, after the coding. Anyway, note that the complexity of the logical network depends more on the difference $n-k$ than on k.
The systematic block coding of type (2,1), with the rule that the number of digits equal to 1 is odd, constitutes the elementary case, named ***Manchester code***; it has the minimum complexity, but as counterpart it causes even to double the bit rate of the coded signal. The code words corresponding to the two possible blocks, that are reduced to the single digits $x_1=0$ and $x_2=1$, are constituted by the pairs of alternate bits 01 and 10, both with probability ½ if each one is associated to one of the two digits in input supposed equiprobable; the pairs of identical bits 00 and 11 are the remaining two configurations, having anyway probability equal to zero. Every error, not followed by any other adjacent error, causes the reception of one of the configurations with zero probability and thus it is detectable; nevertheless, it cannot be corrected, being impossible to find out which bit of the pair is wrong. The Manchester code has actually very scarce capabilities to satisfy the objective to improve performance, so that in practice it is applied more for the particular properties that it provides to the signal after the following modulation operation.
A more meaningful example of systematic block coding is offered by the ***Hamming coding*** (7,4), still of systematic type ($y_i=x_i$, per i=1, 2, 3, 4), which is based on the rule established by the $n-k$ linear relations of parity check:

$$[6.92]\quad y_5=x_1\oplus x_2\oplus x_3,\ y_6=x_2\oplus x_3\oplus x_4,\ y_7=x_1\oplus x_2\oplus x_4,$$

where the symbol $\oplus$ indicates the sum modulus-2. The redundancy is significant: only $2^4=16$ code words among the $2^7=128$ possible configurations are utilized, with a coding rate (see [6.90]) slightly greater than the one of the Manchester coding. The code words are chosen so that the difference among one another is at least in three different positions; such a property allows the detection both of just one and also of two errors introduced in the received string, but also the correction of the single error, because just one code word which differs in a single position from the received string exists.

6.6.1.3 Introduction on the convolutional encoding

Aiming to the forward error correction capability, better results are usually obtained with the same circuit complexity with a second method of binary encoding with redundancy, which is named ***convolutional encoding*** considering that the processing concerns the repetitive operation of convolution of the input sequence with opportune known sequences of finite lenght.

As shown in the scheme in Figura 6.29a, the elements x_k, extracted from a binary signal at the input with rate $R_x=1/T_x$, fill a shift register of L stages, where L is the ***constraint lenght*** of the code. At every interval T_x, starting from the L bits simultaneously registered, through opportune linear combinations n binary digits are composed in parallel, of which n-1 redundant; such elements, by means of a logic circuit of ***parallel-series transformation*** (P/S), finally form one sequence y(n). The transformed output signal which carries the encoded sequence results to have the bit rate increased by a factor n:

$$[6.93]\ R_v = nR_x > R_x.$$

More in general, as shown in Figura 6.29b, it is possible to have k<n shift registers, still with lenght equal to L, which by means of a circuit of ***series-parallel transformation*** (S/P) are filled at every interval kT_x and with the digits present in the registers the n>k>1 digits, then serialized in output, are always formed. Thus, the bit rate of the encoded binary signal is the following:

$$[6.94]\ R_y = \frac{n}{k} R_x > R_x,$$

which, although much less increased, shows the presence of redundancy, because of the coding rate $R_c=k/n$.

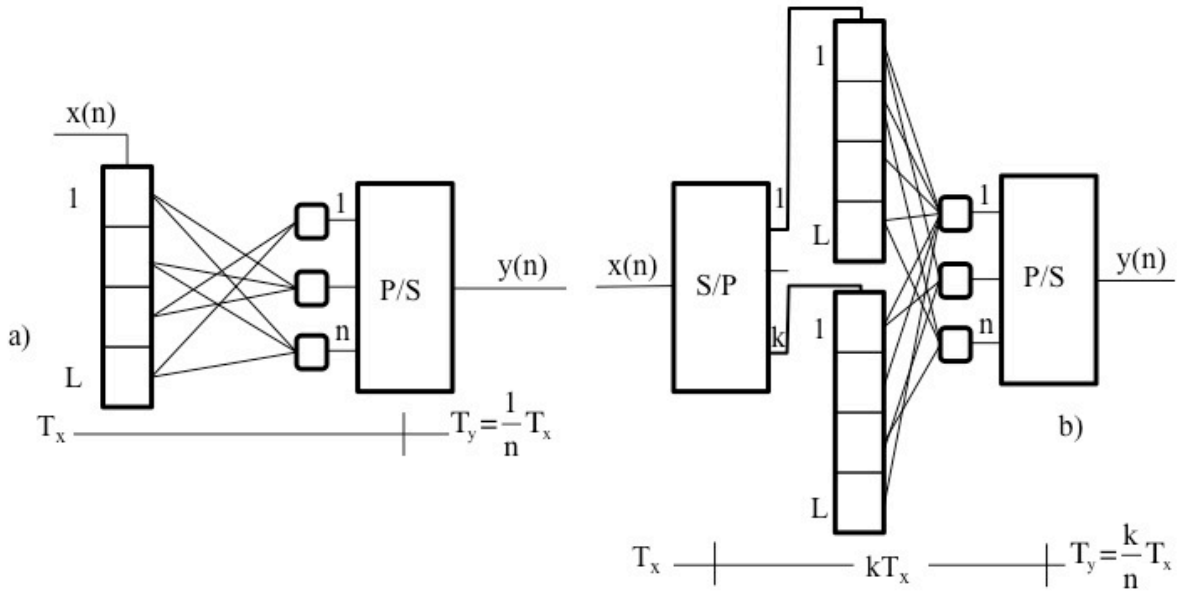


Figura 6.29: Block schemes of convolutional encoders.

The convolutional encoding is of type with memory, considering that operates on the basis not only of the current bit, but also of the previous kL-1. It is possible to note that the linear block coding is a particular case of the convolutional encoding for L=1; in this case there is no longer memory and in fact every block is processed separately.

The convolutional encoding, characterized by the pair of values (L, k/n), allows to acquire with reduced complexity (usually n is equal to a few units) the same performance improvements of the ones above considered in the block coding, by means of error detection and correction, still with the increase of the bit rate as counterpart. Also the introduced delays, still a little greater than $2kT_x$, are lower than those ones in the block coding with the same achievable performance.

6.6.2 Encoding with modification of the cardinality

6.6.2.1 Introduction on the symbol coding

The processing of digital sequences previously considered concerned only operations in which the alphabet of the symbols in input and in output, in particular of binary type, remains unchanged. On the other hand, encoding processing with modification of the symbol cardinality are possible: without loss of generality such kind of processing, referred as a generic ***symbol encoding***, can be fractionated in an opportune outer binary encoding and in a symbol coding without redundancy, or ***simple symbol encoding***, which is the only responsible of the alphabet change. Of course, the complementary processing of ***symbol decoding*** must exist, and must be fractionable in the two corresponding operations as well.

A ***simple symbol encoder*** change an input binary sequence y(n), with bit time T_y, in an output sequence z(n), with symbol time T_z, in which the generic element z_k assumes one of the values of the set of symbols $\{z_q\}$, with q=1, 2, ..., M_L, which constitutes the desired alphabet. Addressing only the logic function, the encoder has the scheme shown Figura 6.30.

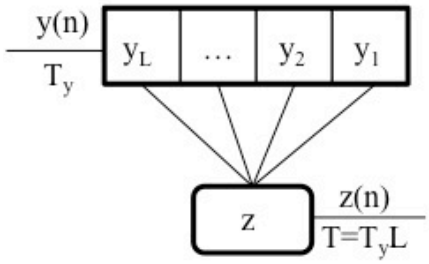


Figura 6.30: Functional logical scheme of a simple symbol encoder.

As illustrated in Figura 6.30, selected an integer number L, at every interval of duration:

$$[6.95]\ T = T_yL,$$

a string $y_1y_2...y_L$ of L binary digits is created in a register with the same number of stages, fed by the input sequence; by means of a linear logical circuit, which operates at interval T on the basis of all the L elements of the register, then the value of the corresponding symbol z of the output sequence is determined, which has therefore the symbol time T. The symbol belongs to an alphabet of cardinality:

$$[6.96]\ M_L \triangleq 2^L,$$

because the number of distinct elements is the same of the different configurations of the string in the register, in general equal to 2^L. Definitively, a reversible rule of symbol encoding results established, with biunivocal association between the M_L symbols and the as many different strings of binary digits,

$$[6.97]\ z_q \Leftrightarrow y_1y_2...y_L,\ q=1,2..,M_L.$$

Considering that the simple encoder operates always with the same method at the rate of the symbol time T, with influence of just the values of the binary digits of the input sequence that determine the string contained in the interval of duration T, the considered processing is memoryless.

The simple symbol encoding is introduced for the advantages that can derive from the slowing of the timing after it: with the same bit time T_v at the input, in fact at the output a step signal having transitions at interval $T=T_vL$ is obtained, with reduction of the practical bandwidth of a factor L. As a counterpart a little total delay is introduced, considering also the decoder, slightly greater than 2T.

Note that since L is integer the cardinality M_L of the achieved alphabet with the considered simple encoding can assume only the values 4, 8, 16, 32 and so on; for this reason the particular notation M_L with subscript L was used. Such a limitation can be removed in a generic ***symbol encoder***, putting before of the simple encoder an opportune binary encoder which introduces redundancy in every string of lenght L in which the sequence at its output is structured. If for example the binary sequence x(n) is changed in a binary sequence y(n) structured in strings of two bits having the

property that the pair 11 is always excluded, after the simple encoder with L=2, that implies M_L=4, the result is that the quaternary symbol that the rule [6.97] associates to the word 11 has probability equal to zero, so that the actual cardinality M≥2 of the alphabet of the sequence z(n) is reduced to the value M=3.

6.6.2.2 Examples of rules of symbol encoding

As well as it is usual to use the symbols 0 and 1 for the binary alphabet, in the case of alphabets with cardinality $M_L=2^L$ it is frequently assumed that the set $\{z_q\}$ of the symbols is composed of the series of the M_L odd integer numbers, distant between each other of two units, symmetrically placed with respect to the zero (see Figura 6.31); then, the algebraic expression of the symbols is the following:

$$[6.98]\ z_q = 2q - M_L - 1,\ \ q = 1, 2.., M_L,$$

for which the set $\{z_q\}$ results ordered with growing values of z_q with growing q.

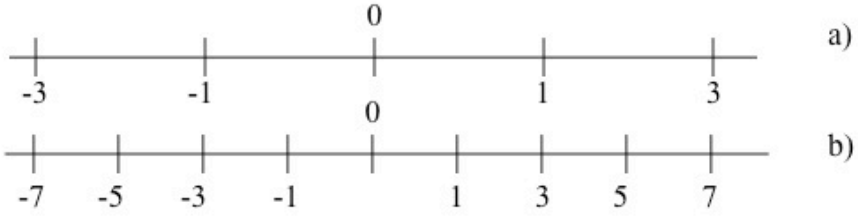


Figura 6.31: Examples of digital symbols of a quaternary (a) and eigthnary (b) alphabet.

Very often the rule [6.97] of symbol encoding is chosen so that to strings $y_1y_2...y_L$ different among one another just for one bit correspond always adjacent elements of the set of symbols [6.98]; then, the rule is named ***Gray encoding***.
A first example of simple symbol encoding is that offered by the Gray encoding with M_L=4, which associates the symbols z_q to the strings y_1y_2 with the rule:

$$[6.99]\ z = 4y_1 + 2(y_1 \oplus \overline{y}_2) - 3,$$

where the sign ⊕ indicates the operation of sum modulus-2, the notation $\overline{y}$ corresponds to the operation of inversion of the binary digit and the two binary digits can assume the values 0 or 1. It is immediate to verify that it is a Gray rule, considering that the associations shown in Tabella 6.2 are obtained.

Tabella 6.2: Associations established by the Gray rule for M_L = 4.

q	1	2	3	4
z_q	-3	-1	1	3
$y_1 y_2$	01	00	10	11

The reader can make the exercise to verify that the Gray rules to get the symbols z_q given by the [6.98] in the cases M_L=8 and M_L=16 have the expressions:

$$[6.100]\ z = 8y_1 + 4y_1 \oplus \overline{y}_2 + 2y_1 \oplus \overline{y}_2 \oplus y_3 - 7,\ M_L=8,$$

$$[6.101]\ z = 16y_1 + 8y_1 \oplus \overline{y}_2 + 4y_1 \oplus \overline{y}_2 \oplus y_3 + 2y_1 \oplus \overline{y}_2 \oplus y_3 \oplus \overline{y}_4 - 15,\ M_L=16.$$

In the case M_L=8 the associations shown in Tabella 6.3 are obtained.

Tabella 6.3: Associations established by the Gray rule for M_L=8.

q	1	2	3	4	5	6	7	8
z_q	-7	-5	-3	-1	1	3	5	7
$y_1 y_2 y_3$	010	011	001	000	100	101	111	110

6.7 INTRODUCTION TO HARMONIC MODULATION

6.7.1 Introduction to the methods of harmonic modulation

Many transmission channels, due to the nature of the transmission means itself, require that the signal in line s(t) is of the type in shifted band, with narrow relative bandwidth around a frequency f_c very high, as happens for the optical fibers and the radio links; moreover, for particular utilization circumstances it is necessary to not having spectral components near the origin, as in the case of the transmission on a telephone channel. On the other hand, at the beginning of the system, or better after a processing (see block SBE in Figura 6.4), in general a pair of signals, $b_c(t)$ and $b_s(t)$, is obtained with the commmon characteristic to be of base band type, often with identical minimum frequency f_{bm} equal or near to zero and the same maximum frequency f_{bM} lower than f_c: thus, in the considered cases it is necessary a processing on the mentioned pair (see block MO in Figura 6.4) which produces a different signal s(t) having all the spectrum contained within the bandwidth offered by the channel; obviously aiming to the correct transfer of the information the operation must be reversible, i.e. the complementary processing that acting on s(t) gives back two faithful forms to $b_c(t)$ and $b_s(t)$ must exists.
Considering that after the considered processing spectral components not present before appear in s(t), the considered processing must necessarily be implemented by means of a subsystem with non linear behaviour; also the complementary subsystem will be non linear as well. As a matter of fact in this considered case of the ***harmonic modulation*** non linear operations which involve $b_c(t)$ and $b_s(t)$ in transmission are implemented, or s(t) in reception, with a known harmonic signal:

$$[6.102]\ c(t) \mathrel{\hat{=}} \cos(\omega_c t+\varphi_c),$$

named ***carrier signal*** and characterized by the frequency $f_c=\omega_c/2\pi$ and by an unessential phase φ_c; the amplitude, unessential as well, can be assumed to be equal to one. The modulated signal s(t) in the channel can be conveniently represented referred to the frequency f_c, in its known shifted band signal forms (see [3.193] and [3.194]), here repeated for convenience:

$$[6.103]\ s(t) = \iota(t)\cos[\omega_c t+\varphi(t)] = a_c(t)\cos(\omega_c t) - a_s(t)\sin(\omega_c t),$$

or by means of its complex envelope:

$$[6.104]\ \iota(t)= \iota(t)\, e^{j\alpha(t)} = a_c(t) + j a_s(t),$$

in which the argument $\alpha(t)$ is equal to $\varphi(t)$ plus a multiple of the circle angle.
The harmonic modulation methods can be classified in two cathegories: ***modulation by product***, in which the product operation with harmonic signals at frequency f_c is implemented, and ***angle modulation***, in which the operation is implemented on the angle which is the argument of the oscillation. Hereinafter just the main aspects concerning the modification of the spectral allocation are dealt with that disregard the analogue or digital nature of the signals; the schemes of the modulators and demodulators (MO and DEM) and the effect produced on the harmonic modulation by the processing in the base band sections (SBE and SBR) are the object of a next chapter.

6.7.2 Introduction on the modulation by product

By means of one product between the carrier signal [6.102] and only one real signal in base band $b(t)\equiv b_c(t)$, then assuming that $b_s(t)\equiv 0$, the simple product modulation is obtained, or ***product modulation***, with modulated signal:

$$[6.105]\ s_x(t) \mathrel{\hat{=}} b(t)c(t) = b(t)\cos(\omega_c t+\varphi_c),$$

and complex envelope:

$$[6.106]\ \iota_{sx}(t) = b(t)\ e^{j\varphi_c}\ .$$

Note that usually b(t) assumes both positive and negative sign, so that the real envelope $\iota_{sx}(t)=|\iota_{sx}(t)|=|b(t)|$ has an evolution different from b(t).
Supposing that the Fourier transform B(f) of b(t) exists, from the [6.106] it is possible to immediately obtain the spectrum of the complex envelope $I_{sx}(f)=B(f)\,e^{j\varphi_c}$, so that reminding the [3.159] it is possible to determine the spectrum of the product modulated signal:

$$[6.107]\ S_x(f) = \frac{1}{2}\ [B(f-f_c)\,e^{j\varphi_c} + B^*(-f-f_c)\,e^{-j\varphi_c}\ ].$$

The evolution of the modulus $S_x(f)$ is like the one shown in Figura 6.32. The spectrum of the modulated signal is thus composed, but with a factor 1/2, of two replicas of the modulating one B(f) shifted of $\pm f_c$; from the same figure it is possible to observe immediately that such replicas result to be separated between each other but if the following condition is respected:

$$[6.108]\ f_c > f_{bM},$$

that therefore represents the reversibility condition of the modulation by product. Thus, with the modulation by product, sometimes also named Double Side Band (DSB), a signal which results redundant in the frequency domain is obtained, and therefore more robust with respect to channel impairments; such redundancy is the advantageous counterpart of the requirement of a double channel bandwidth with respect to the base band one (which is of course a disadvantage).

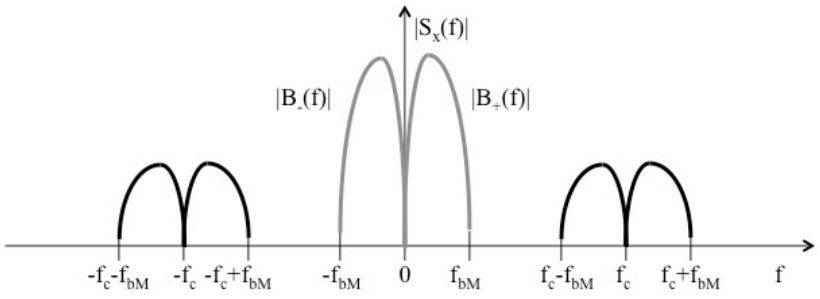


Figura 6.32: Amplitude spectra of the modulated signal by product (dark line) and of the modulating signal in base band (grey line).

In the case of power signal similar considerations are valid on the power spectra. Definitively, the bandwidth B_x of $s_x(t)$ in the channel results to be centered on f_c and has a double extension of the maximum frequency in base band:

$$[6.109]\ B_x = 2\,f_{bM}.$$

An interesting particular case occurs when b(t) is obtained with a simple linear processing like:

$$[6.110]\ b(t) = A_c[1+kx(t)] \geq 0,$$

i.e. with A_c and k constant and the latter chosen so that the following disequality $|kx(t)|\leq 1$ results always verified and therefore:

$$[6.111]\ \iota_{sx}(t)=|b(t)|=1+kx(t).$$

Then, with respect to x(t) the subcase of ***Amplitude Modulation*** (AM) is obtained, characterized by the simple proportionality, but differing for a multiplicative contant, between the envelope of the modulated signal and the originating signal. Nevertheless, note that in the AM on the channel the carrier signal $A_c\cos(\omega_c t+\varphi_c)$ is transmitted, that, being known, is not useful to transfer information while it requires more power than the one actually useful, regarding only the side bands. A typical time evolution of an AM signal is shown in Figura 3.18.
The ***modulation by product in quadrature*** is more general, seldom also named instant quadrature amplitude modulation (QAM = Quadrature Amplitude Modulation), which needs a processing capable to provide the two real signals in base band with mean value equal to zero and two

products with harmonic signal in phase quadrature; in this way the signal in shifted band in the general form is obtained (see [6.103]):

$$[6.112]\ s(t) = b_c(t)\cos(\omega_c t) - b_s(t)\sin(\omega_c t),$$

where $b_c(t)$ and $b_s(t)$ result to be the instant amplitudes of the two carriers in quadrature. Each of the two addends in s(t) are signals modulated by product with the same carrier frequency and as a consequence their spectra share the same channel bandwidth, for an extension equal to the double of the maximum frequency f_{bM}. With the modulation by product in quadrature the redundancy in the frequency domain still occurs, but thanks to the overlapping of the spectra there is no cost in terms of channel bandwidth requirement. Considering that the two addend signals in s(t) although overlapped in frequency can be splinted, as demonstrated in one of the following chapters, the processing is still reversible in the respect of the condition $f_c > f_{bM}$.

6.7.3 Introduction on the harmonic angle modulation

The second cathegory of processing with harmonic signal, i.e. the ***angle modulation***, is characterized by the variability in the time domain of just the argument of the oscillation, which means that the envelope is contant. Once it is imposed that $\iota_{s\alpha}(t) \equiv A_c$ where A_c is a constant, the modulated signal and its complex respectively assume the expressions:

$$[6.113]\ s_\alpha(t) \hat{=} A_c\cos[\omega_c t + \varphi(t)],$$

$$[6.114]\ \iota_{sx}(t) = A_c\, e^{j\varphi(t)}.$$

The information is carried through the intant phase φ(t), which is in fact made dependent on only one signal in base band b(t) by means of a generic linear transformation:

$$[6.115]\ \varphi(t) = K_\varphi h_\varphi(t) * b(t),$$

where K_φ is a constant and $h_\varphi(t)$ is an arbitrary function which can be Fourier transformed $H_\varphi(f) = F[h_\varphi(t)]$. A typical time evolution of an angle modulated signal is that one already shown in Figura 3.19.

The non linear dependence, through the cosine function, of the signal on the channel $s_\alpha(t)$ from the modulating one b(t) makes the determination of the bandwidth of the angle modulated spectrum very difficult. Such a spectrum, which, referring only to the positive half frequency axis shows anyway in general a symmetrical evolution with respect to f_c (see for example Figura 6.33) and cannot be extracted from that one of b(t), is in theory infinitively extended, but adopting a practical rule as a matter of fact can be assumed to be limited within the interval, named Carson bandwidth:

$$[6.116]\ B_\alpha = 2(\Delta f_p + f_{bM}),$$

where Δf_p is the frequency deviation, i.e. the peak value of the instant frequency deviation Δf(t) linked to the instant phase by the already known relation [3.205], reported below:

$$[6.117]\ \Delta f(t) = \frac{1}{2\pi}\,\frac{d\varphi(t)}{dt}.$$

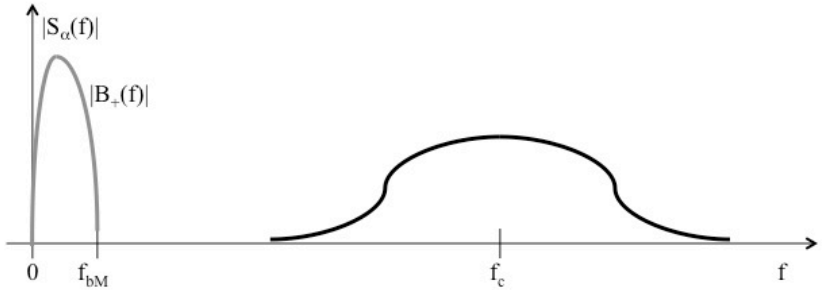


Figura 6.33: Amplitude spectrum of the angle modulated signal (dark line) and of the base band modulating signal (grey line).

To guarantee the reversibility of the angle modulation is thus opportune to satisfy the:

$$[6.118]\ f_c >> \Delta f_p + f_{bM}.$$

Note from [6.115] that growing K_φ the instant phase $\varphi(t)$ varies more and thus both the peak value Δf_p and the practical bandwidth B_α grow, with growing introduction of redundancy in the frequency domain of the modulated signal and consequent greater protection with respect to channel impairments and with the counterpart of the increase of the bandwidth requirement of the latter.

The simplest particular case, from analythic point of view, is the ***Phase Modulation*** (PM), for which in the [6.115] $h_\varphi(t) \equiv \delta(t)$ is assumed, i.e. $H_\varphi(t) \equiv 1$, so that the instant phase results proportional to the modulated signal:

$$[6.119]\ \varphi(t) = K_\varphi b(t), \text{ for PM.}$$

One more particular case is the ***Frequency Modulation*** (FM), in which the linear transformation which leads to $\varphi(t)$ is the integration in the time domain, so that reminding the [6.117] results fixed that the instant frequency deviation is proportional to the modulated signal:

$$[6.120]\ \Delta f(t) = \frac{K_\varphi}{2\pi} b(t), \text{ for FM.}$$

7 ELEMENTS OF PROBABILITY, RANDOM VARIABLES AND STOCHASTIC PROCESSES

7.1 ELEMENTS OF PROBABILITY

7.1.1 Basic definitions

The probability of an event A is a number P(A) assigned to the event itself which indicates a priori the percentage of times the event can occur with respect to a set of events which it belongs to. In the case of execution of an experiment, for large number of repetition of the experiment, it can be interpreted as the frequency the event occurs

$$[7.1]\ P(A) \cong \frac{n_A}{n}$$

Probablility can be dealt with according to three approaches, as summarized in Table 7.1.

Table 7.1: Approach to probability

Classical	The probability of an event is the ratio between the positive occurrences and the number of possible cases (as long as the latter are equally possible).
Frequency	The probability is the limit of the relative frequency of the event, when the number of experiments tends to infinite.
Subjective	The probability P(E) of an event E is the measure of the degree of confidence that an individual assigns, on the basis of information and opinions he/she has, that the event E will occur.

The introduced figures can be interpreted in a different way according to the kind of analysis we are carrying on, either statistical or probability, as shown in Table 7.2.

Table 7.2: Interpetations of the introduced figures

	Statistic	**Probability**
n_A	number of times the event occurs	number of favorable events
n	total number of repetitions of the event	total number of events
	A posteriori	**A priori**
	From the results of an experiment some characteristics of the population are determined	From the known characteristics of the population the result of another experiment is foreseen

7.1.2 Axiomatic theory of the probability

The axiomatic theory of the probability utilizes the concepts of the set theory to chracterize random phenomena. Given a set S, the elements contained in S are considered the possible outcomes of the experiments. Also the subsets of S are events. The certain event S always occurs. The event resulting by the union of two events A and B, namely the union event A+B, occurs when one of the two or both occur. The intersection event resulting from two events A and B, namely AB, occurs when both occur simultaneously. The events A and B are mutually exclusive if they cannot occur simultaneously. The definitions are summarized in Table 7.3.

Table 7.3: Axiomatic probability definitions

Definitions	
S space that contains all the events ➔ sure event Elements of S ➔ possible outcomes of the experiment	Subsets of S ➔ events The empty set {∅} ➔ impossible event The event $\{z_i\}$ ➔elementary event

The following axioms must be respected:

$P(A) \geq 0$; $P(S)=1$; A and B two events | $AB=\{\varnothing\}$ ➔ $P(A+B) = P(A)+P(B)$

The following properties are verified:

$P(\{\varnothing\})=0$; $P(A)=1-P(\bar{A})\leq 1$; $P(A+B) = P(A)+P(B) - P(AB) \leq P(A)+P(B)$

They can be demonstrated utilising the above axioms.

7.1.3 Conditional probability

The concept of conditional probabilities applies in all the situations in which the event A we are interested to observe, of which we want to carry out the relevant statistics, occurs assuming that another event M occurred. Thus, there is a preliminary condition to make the occurrence of A significant for our analysis, as often it is verified with signals.

The *conditional probability* of the occurrence of the event A assuming that the event M occurred, indicated as P(A|M) is defined as:

$$[7.2]\ P(A|M) = \frac{P(AM)}{P(M)}$$

where P(M) cannot be equal to zero.

Immediately from the definition the following properties are verified:

- If M⊂A then P(A|M)=1 because AM=A;
- If A⊂M then $P(A|M) = \frac{P(A)}{P(M)} \geq P(A)$

The frequency interpretation can help in understanding the concept. Let's denote by n_A, n_M and n_{AM} the number of occurrences of the events A, M and AM respectively and by n the total number the experiments that is run. Then:

$$[7.3]\ P(A) = \frac{n_A}{n}, P(M) = \frac{n_M}{n}, P(AM) = \frac{n_{AM}}{n}$$

and as a aconsequence

$$[7.4]\ P(A|M) = \frac{P(AM)}{P(M)} = \frac{n_{AM}/n}{n_M/n} = \frac{n_{AM}}{n_M}$$

In practice, discarding the trials in which M doesn't occur, P(A|M) equals the relative frequency of occurrence n_{AM} / n_M of the event A in the subsequence of trials in which M occurs.

Of course the conditional probability respects the axioms:

$P(A|M)\geq 0$; $P(S|M)=1$; $P(A+B|M) = P(A|M) + P(B|M)$.

In fact, to prove the third one, let's consider the events A and B mutually exclusive so that also the events AM and BM are mutually exclusive as well. Then,

$$[7.5]\ P(A+B|M) = \frac{P[(A+B)M]}{P(M)} = \frac{P(AM)+P(BM)}{P(M)}$$

7.1.4 Total probability and Bayes' Theorem

Denoting a partition of S by $\Omega=[A_1, A_2, \ldots, A_n]$ and identifying B as an arbitrary event, then:

$$[7.6]\ P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

In fact, $B=BS = B[A_1, A_2, \ldots, A_n]= BA_1 + BA_2 + \ldots + BA_n$ and the events BA_k and BA_j are mutually exclusive because A_k and A_j are mutually exclusive as well.

Thus,

$$[7.7]\ P(B)=P(BA_1)+P(BA_2)+\ldots+P(BA_n)$$

and since for the conditional probability

$$[7.8]\ P(BA_k)=P(B|A_k)P(A_k)$$

the [7.6] follows and is called the total probability theorem.

Since

$$[7.9]\ P(BA_k)=P(B|A_k)P(A_k)= P(A_k|B)P(B)$$

$$[7.10]\ P(A_k|B) = P(B|A_k)\frac{P(A_k)}{P(B)}$$

Finally, the Bayes' theorem is obtained

$$[7.11]\ P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B|A_1)P(A_1)+\cdots+P(B|A_n)P(A_n)} = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Often $P(A_k)$ is named a priori probability while $P(A_k|B)$ is named a posteriori probability.

7.1.5 Independence between events

Two events are independent if

$$[7.12]\ P(AB)=P(A)P(B)$$

and as a consequence

$$[7.13]\ P(A|B)=P(A)$$

In the case of three events, they are mutually independent if

$$[7.14]\ P(A_i\ A_j)=P(A_i)P(A_j) \text{ for any } i\neq j \text{ and } P(A_1\ A_2\ A_3)=P(A_1)P(A_2)P(A_3)$$

7.2 RANDOM VARIABLES

A random variable (r.v.) is a function $x(\zeta)$, whose domain is the set S of outcomes, which assigns a number to each possible result ζ of the experiment. The subsets of S are events and the codomain (range) is a set of numbers (integer, real, complex, etc.).
As an example, in the die experiment it is possible to assign to the six outcomes the numbers $\mathbf{x}(f_i)=2i$. Therefore, $\mathbf{x}(f_1)=2$, $\mathbf{x}(f_2)=4$, $\mathbf{x}(f_3)=6$, $\mathbf{x}(f_4)=8$, $\mathbf{x}(f_5)=10$, $\mathbf{x}(f_6)=12$.
In approaching the analysis of r.v. often the meaningful case concerns to know the probability of events like $\{\mathbf{x}\leq x\}$ or $\{x_1\leq\mathbf{x}\leq x_2\}$, where $\{\mathbf{x}\leq x\}$ is the subset of S that contains all the possible outcomes of the experiment ζ so that $\mathbf{x}(\zeta)\leq x$ and $\{x_1\leq \mathbf{x}(\zeta)\leq x_2\}$ is the subset of S that contains all the possible outcomes of the experiment ζ so that $\{x_1\leq\mathbf{x}\leq x_2\}$. Finally, the notation $\{\mathbf{x}=x\}$ indicates the subset of S for which $\mathbf{x}(\zeta) = x$.
Indicating with Ξ a set of numbers, then the event $\{\mathbf{x}\in\Xi\}$ represents a subset of S that contains all the possible outcomes for which $\mathbf{x}(\zeta)\in\Xi$.
The function which defines a r.v. (assigns a number $\mathbf{x}(\zeta)$ to the outcome ζ of an experiment) must satisfy the following two properties:

- The set $\{\mathbf{x}\leq x\}$ is representative of an event in S for every value x
- The probability of the events $\{x=\infty\}$ and $\{x=-\infty\}$ are equal to zero.

7.2.1 Distributions and density functions

The value of the probability of the event $\{\mathbf{x}\leq x\}$, indicated with $P\{\mathbf{x}\leq x\}$, depends on x. The **cumulative distribution function** of the random variable x is defined as $F_x(x)=P\{\mathbf{x}\leq x\}$ for every x between $-\infty$ and ∞.
It must satisfy the following properties:

1. $F(\infty)=1$ and $F(-\infty)=0$
2. if $x_1<x_2$ ➔ $F(x_1)\leq F(x_2)$ (non decreasing function)
3. if $F(x_0)=0$ ➔ $F(x) = 0\ \forall\ x\leq x_0$
4. $P\{\mathbf{x}>x\}=1-F(x)$
5. $F(x^+)=F(x)$ (continuity from right)
6. $P\{x_1\leq\mathbf{x}\leq x_2\}=F(x_2)-F(x_1)$

7. $P\{\mathbf{x}=x\}=F(x)-F(x^-)$
8. $P\{x_1<\mathbf{x}\leq x_2\}=F(x_2)-F(x_1^-)$

A r.v. is of continuous type if F(x) is continuous. In this case $F(x^-)=F(x)$ and $P\{\mathbf{x}=x\}=0$ for every x. A r.v. is of discrete type if F(x) is staircase shaped. Identifying x_i as the discontinuity points, it is possible to obtain

$$F(x_i)-F(x_i^-)=P(\mathbf{x}=x_i)=p_i$$

A r.v. is of mixed type if F(x) is discontinuous but not as a staircase. A r.v. is surely of discrete type if the set S contains many but finite elements, while it may be still discrete even in case S contains infinite elements.

The **density function** of a r.v. is defined as the derivative of the distribution function F(x):

$$[7.15]\ f(x)=\frac{dF(x)}{dx}$$

In case of discrete type r.v. assuming x_i as values with corresponding probability p_i, the density function is so obtained:

$$[7.16]\ f(x)=\sum_i p_i\delta(x-x_i) \qquad p_i=P(\boldsymbol{x}=x_i)$$

The density function must satisfy the following properties:

1. $f(x)\geq 0$
2. $\int f(x)dx=1$
3. $F(x_2)-F(x_1)=P(x_1\leq\mathbf{x}\leq x_2)=\int_{x1}^{x2}f(x)dx$
4. $P\{x\leq\mathbf{x}\leq x+\Delta x\}\cong\Delta x f(x)$

7.2.2 Conditional distributions

As well as the conditional probability, the conditional distribution of a r.v. can be introduced and defined as:

$$[7.17]\ F(x|M)=P(\boldsymbol{x}\leq x|M)=\frac{P(\boldsymbol{x}\leq x,M)}{P(M)}$$

where $P(\mathbf{x}\leq x, M)$ is the intersection event between $(\mathbf{x}\leq x)$ and M that means the event including all the outcomes $\zeta\in S$ such that $\mathbf{x}(\zeta)\leq x$ and $\zeta\in M$.

The definition of the conditional distribution is the same of the distribution function already introduced, but replacing all the probabilities with the conditional probabilities. Also the properties are the same:

$F(\infty|M)=1$ and $F(-\infty|M)=0$

$$[7.18]\ P(x_1<\boldsymbol{x}\leq x_2|M)=F(x_2|M)-F(x_1|M)=\frac{P(x_1<\boldsymbol{x}\leq x_2,M)}{P(M)}$$

By definition the conditional density f(x|M) is the derivative of the the conditional distribution F(x|M)

$$[7.19]\ f(x|M)=\frac{dF(x|M)}{dx}=\lim_{\Delta x\to 0}\frac{P(x<\boldsymbol{x}\leq x+\Delta x|M)}{\Delta x}$$

The function has the same properties of the density function and thus it is nonnegative and its integral is equal to 1.

If $M=(\mathbf{x}\leq a)$ and $F(a)\neq 0$

$$[7.20]\ F(x|\boldsymbol{x}\leq a)=P(\boldsymbol{x}\leq x|\boldsymbol{x}\leq a)=\frac{P(\boldsymbol{x}\leq x,\boldsymbol{x}\leq a)}{P(\boldsymbol{x}\leq a)}$$

If $x\geq a$, then $(\mathbf{x}\leq x, \mathbf{x}\leq a)=x(\mathbf{x}\leq a)$ and as a consequence

$$[7.21]\ F(x|\boldsymbol{x}\leq a)=\frac{P(\boldsymbol{x}\leq a)}{P(\boldsymbol{x}\leq a)}=1$$

Instead, if $x<a$, then $(\mathbf{x}\leq x, \mathbf{x}\leq a)=x(\mathbf{x}\leq x)$ and as a consequence

$$[7.22]\ F(x|\boldsymbol{x} \le a) = \frac{P(\boldsymbol{x} \le x)}{P(\boldsymbol{x} \le a)} = \frac{F(x)}{F(a)} = 1$$

Still for x<a, being f(x) the derivative of F(x) and applying the operation to the previous

$$[7.23]\ f(x|\boldsymbol{x} \le a) = \frac{f(x)}{F(a)} = \frac{f(x)}{\int_{-\infty}^{\infty} f(x)dx}$$

If M=(b<**x**≤a)

$$[7.24]\ F(x|b < \boldsymbol{x} \le a) = \frac{P(\boldsymbol{x} \le x, b < \boldsymbol{x} \le a)}{P(\ b < \boldsymbol{x} \le a)}$$

if x≥a, then (**x**≤x, b<**x**≤a)=(b<**x**≤a) and as a consequence:

$$[7.25]\ F(x|b < \boldsymbol{x} \le a) = \frac{F(a) - F(b)}{F(a) - F(b)} = 1$$

if (b≤**x**<a), then (**x**≤x, b<**x**≤a)=(b<**x**≤x) and as a consequence:

$$[7.26]\ F(x|b < \boldsymbol{x} \le a) = \frac{F(x) - F(b)}{F(a) - F(b)}$$

if x<b, then (**x**≤x, b<**x**≤a)=∅ and as a consequence

$$[7.27]\ F(x|b < \boldsymbol{x} \le a) = 0$$

The density function in the case (b≤**x**<a) can be calculated as

$$[7.28]\ f(x|b < \boldsymbol{x} \le a) = \frac{f(x)}{F(a) - F(b)}$$

for (b≤**x**<a) and 0 otherwise.
The **total probability theorem** can be expressed in its continuous version as:

$$[7.29]\ \int_{-\infty}^{\infty} P(A|\boldsymbol{x} = x) f(x) dx = P(A)$$

while the continuous version of the **Bayes' theorem** is the following:

$$[7.30]\ f(x|A) = \frac{P(A|\boldsymbol{x}=x)}{P(A)} f(x) = \frac{P(A|\boldsymbol{x}=x) f(x)}{\int_{-\infty}^{\infty} P(A|\boldsymbol{x}=x) f(x) dx}$$

7.2.3 Functions of random variable

Considering **x** a r.v. and g(x) a function of the real variable x it is possible to define a new r.v. as

$$[7.31]\ \mathbf{y}=g(\mathbf{x})$$

The meaning is the following: considering ζ∈S, **x**(ζ) is a number and g[**x**(ζ)] is another number, determined by **x**(ζ) and g(x), which is the value **y**(ζ)=g[**x**(ζ)] assigned to the new variable **y**. For a certain value of y, the values of x such that g(x)≤y belong to a set on the x axis denoted as I_y.
From the above considerations in order that g(x) is a r.v. it must have the following properties:

- its domain must include the range (codomain) of the r.v. x
- the events **y**=g(**x**)=±∞ must have probability equal to zero.

To express the distribution function $F_y(y)$ of the r.v. **y**=g(**x**) in terms of the distribution function $F_x(x)$ of the r.v. **x** and of the function g(x) it is necessary to determine the set R_y of the x axis such that g(x)≤y and then the probability that the r.v. **x** belongs to this set.
To determine the probability density function of **y**=g(**x**), $f_y(y)$, in terms of the probability density function of **x**, $f_x(x)$, it is necessary to consider only the values of **y** such that there are some values of **x** for which g(x)=y (for values of y out of that interval $f_y(y)=0$).
To find the expression of $f_y(y)$ for a specific y it is necessary to solve the equation y=g(x) and to indicate the real roots as x_n ($y=g(x_1)=g(x_2)=\ldots=g(x_n)$). It is possible to demonstrate that:

$$[7.32]\ f_y(y) = \sum_i \frac{f_x(x_i)}{|g'(x_i)|}$$

where $g'(x_i)$ indicates the derivative of g(x).

7.2.4 Moments

Moments are synthetic parameters that are very important to study and characterize a r.v. because with a single figure a lot of information concerning its behaviour are given.
The expected value or mean of a r.v. for a continuous type variable is defined as:

$$[7.33]\ E(x) = \int_{-\infty}^{\infty} xf(x)dx = \eta = \eta_x$$

In case of a discrete type variable the expected value assumes the following expression:

$$[7.34]\ E(x) = \sum_i p_i x_i$$

The conditional mean of a r.v. **x**, with M the condition event, is obtained by the [7.33] but replacing the density function with the conditional density function:

$$[7.35]\ E(x|M) = \int_{-\infty}^{\infty} xf(x|M)dx$$

while in case of discrete type r.v. the expression is the following:

$$[7.36]\ E(x|M) = \sum_i P(\boldsymbol{x} = x_i|M)x_i$$

If a r.v. **y** is a function of another r.v. **x** according to a function g(x) such that **y**=g(**x**) it is possible to demonstrate that

$$[7.37]\ E(y) = \int_{-\infty}^{\infty} yf_y(y)dy = E\{g(x)\} = \int_{-\infty}^{\infty} g(x)f_x(x)dx$$

The second most used moment is the variance which provides a measure of the variability of the r.v. around the expected value. The expression is the following:

$$[7.38]\ \sigma^2 = \int_{-\infty}^{\infty}(x-\eta)^2 f(x)dx$$

The parameter σ (or σ_x) is the standard deviation of the variable **x**.
In the case the r.v. is of discrete type the variance assumes the following expression:

$$[7.39]\ \sigma^2 = \sum_i p_i(x_i-\eta)^2$$

In addition to the two very important synthetic parameters above introduced, also other parameters are of interest in the characterization of random variables. Generalising the above expressions it is possible to obtain:

Moments (general)	[7.40]	$m_n = E(\boldsymbol{x}^n) = \int_{-\infty}^{\infty} x^n f(x)dx$
Central moments	[7.41]	$\mu_n = E\{(\boldsymbol{x}-\eta)^n\} = \int_{-\infty}^{\infty}(\boldsymbol{x}-\eta)^n f(x)dx$
Absolute moments	[7.42]	$E\{\lvert\boldsymbol{x}^n\rvert\}\quad E\{\lvert\boldsymbol{x}-\eta\rvert^n\}$
Generalized moments	[7.43]	$E\{(\boldsymbol{x}-a)^n\}\quad E\{\lvert\boldsymbol{x}-a\rvert^n\}$

It is possible to note that: $\mu_0=m_0=1$; $m_1=\eta$; $\mu_1=0$; $\mu_2=\sigma^2$.

7.2.5 Characteristic function and moment generating function of a random variable

The characteristic function of a continuous r.v. is defined by the following expression:

$$[7.44]\ \Phi(\omega) = \int_{-\infty}^{\infty} f(x)e^{j\omega x}dx$$

The above function assumes its maximum value at the origin because f(x)≥0 and therefore $|\Phi(\omega)| \leq \Phi(0)=1$.
The moment generating function of the random variable x is defined replacing jω with s:

$$[7.45]\ \boldsymbol{\Phi}(s) = \int_{-\infty}^{\infty} f(x)e^{sx}dx$$

The second characteristic function of the r.v. **x** is defined as

$$[7.46]\ \Psi(\omega)=\ln\Phi(\omega)=\boldsymbol{\Psi}(j\omega)$$

It is easy to note that $\Phi(\omega)=E(e^{j\omega x})$ and $\boldsymbol{\Phi}(s)= E(e^{sx})$.
Utilising the moment theorem, expanding $\boldsymbol{\Phi}(s)$ into a series near the origin, if all moments are finite and the series converges absolutely near s=0, it is possible to obtain

$$[7.47]\ \boldsymbol{\Phi}(s) = \sum_{n=0}^{\infty} \frac{m_n}{n!} s^n$$

from which it is possible to calculate $\boldsymbol{\Phi}'(0)=m_1=\eta$ and $\boldsymbol{\Phi}''(s)=m_2=\sigma^2+\eta^2$.
In the case the r.v. is of discrete type the characteristic function assumes the expression

$$[7.48]\ \Phi(\omega) = \sum_i p_i e^{-j\omega x_i}$$

If the discrete random variable **x** assumes only integer values the moment generating function is defined according to the following expression:

$$[7.49]\ \Gamma(z) = \sum_i p_i z^i$$

Also in this case, applying the moment theorem it is possible to achieve $\Gamma'(1)=E(n)$ and $\Gamma''(1)=E(n^2)-E(n)$

7.2.6 Examples of distribution and density functions

Some random variables are particularly important to characterize physical phenomena occurring in the telecommunication systems (signal, propagation channel, network behaviour, disturbs, etc.). In this section for eight random variables the expressions of the main parameters are reported: the density function, the distribution function, the expected value, the variance and the characteristic function.

Table 7.4: Uniform distribution

Density function	[7.50]	$p(x)=\frac{1}{b-a}, \quad a \le x \le b$
Distribution function	[7.51]	$P(x)=\frac{x-a}{b-a}, \quad a \le x \le b$
Expected value	[7.52]	$\eta=\frac{1}{2}(a+b)$
Variance	[7.53]	$\sigma_x^2=\frac{1}{12}(b-a)^2$
Characteristic function	[7.54]	$\Xi(x)=\frac{e^{j\omega b}-e^{j\omega a}}{j\omega(b-a)}$

To show the parameters of the Gaussian distribution it is necessary to introduce the error function

$$[7.55]\ erf(x) \triangleq \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$$

the complementary error function

$$[7.56]\ erfc(x) \triangleq 1 - erf(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} dt$$

and the main properties
erf(-x)=-erf(x), erf(-∞)=-1, erf(0)=0, erf(∞)=1,
erfc(-x)=1-erf(-x)=1+erf(x)=2-erfc(x), erfc(-∞)=2, erfc(0)=1, erfc(∞)=0

Table 7.5: Gaussian (normal) distribution

Density function		$p(x)=\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-\eta_x)^2}{2\sigma_x^2}}$
Distribution function		$P(x)=1-\frac{1}{2}\operatorname{erfc}\left(\frac{x-\eta_x}{\sqrt{2}\sigma_x}\right)$
Expected value		η_x
Variance		σ_x^2
Characteristic function		$\Xi(x)=\exp\left(j\omega\eta_x-\frac{1}{2}\omega^2\sigma_x^2\right)$

Table 7.6: Rayleigh distribution

Density function		$p(x)=\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}},\quad x\geq 0$
Distribution function		$P(x)=1-e^{-\frac{x^2}{2\sigma^2}},\quad x\geq 0$
Expected value		$\eta_x=\sqrt{\frac{\pi}{2}}\sigma$
Variance		$\sigma_x^2=\left(2-\frac{\pi}{2}\right)\sigma^2$
Characteristic function		$\Xi(x)=2\int_0^\infty ue^{-\left(u-2j\omega\sigma^2\right)u}du$

Table 7.7: Exponential distribution

Density function		$p(x)=\frac{1}{2\sigma^2}e^{-\frac{x}{2\sigma^2}},\quad x\geq 0$
Distribution function		$P(x)=1-e^{-\frac{x}{2\sigma^2}},\quad x\geq 0$
Expected value		$\eta_x=2\sigma^2$
Variance		$\sigma_x^2=4\sigma^4$
Characteristic function		$\Xi(x)=\frac{1}{1-2j\omega\sigma^2}$

Indicating with $I_0(x)$ the modified Bessel function of order 0 and with Q the Marcum function

Table 7.8: Rice distribution

Density function		$p(x)=\frac{x}{\sigma^2}e^{-\frac{x^2+s^2}{2\sigma^2}}I_0\left(\frac{sx}{\sigma^2}\right),\quad x\geq 0$
Distribution function		$P(x)=1-Q\left(\frac{s}{\sigma},\frac{x}{\sigma}\right),\quad x\geq 0$
Expected value		$\eta_x=\frac{\pi}{2}\sigma e^{-\frac{s^2}{4\sigma^2}}\left[\left(1+\frac{s^2}{2\sigma^2}\right)I_0\left(\frac{s^2}{4\sigma^2}\right)+\frac{s^2}{2\sigma^2}I_1\left(\frac{s^2}{4\sigma^2}\right)\right]$
Variance		$\sigma_x^2=2\sigma^2\left(1+\frac{s^2}{2\sigma^2}\right)-\eta_x^2$
Characteristic function		$\Xi(x)=\sigma^2\int_0^\infty uI_0\left(\sigma^2vu\right)e^{-\frac{\sigma^2}{2}\left(u^2+v^2-2j\omega u\right)}du$

Table 7.9: Lognormal distribution

Density function		$p(x)=\frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{1}{2}\left(\frac{\ln x-\eta}{\sigma}\right)^2}, \quad x\geq 0$
Distribution function		$P(x)=1-\frac{1}{2}\operatorname{erfc}\left(\frac{\ln x-\eta}{\sqrt{2}\sigma}\right), \quad x\geq 0$
Expected value		$\eta_x=e^{\eta}e^{\frac{1}{2}\sigma^2}$
Variance		$\sigma_x^2=e^{2\eta}\left(e^{2\sigma^2}-e^{\sigma^2}\right)$
Characteristic function		$\Xi(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}e^{-\left(\frac{u-\eta}{\sqrt{2}\sigma}\right)^2+j\omega\exp(u)}du$

Table 7.10: Binomial distribution

Density function		
Distribution function		$P_k=\operatorname{Prob}\{x=k\}=\binom{K-1}{k}p^k q^{K-1-k}$
Expected value		$\eta_x=Kp$
Variance		$\sigma_x^2=Kpq$
Characteristic function		$\Xi(x)=\left(q+pe^{j\omega}\right)^K$

Table 7.11: Binary distribution (binomial for K=2)

Density function		$p_x(x)=(1-p)\delta(x)+p\delta(x-1)$
Distribution function		$P_x(x)=\begin{cases}0 & x<0\\ 1-p & 0<x<1\\ 1 & x>1\end{cases}$
Expected value		$\eta_x=2p$
Variance		$\sigma_x^2=2p(1-p)$
Characteristic function		$\Xi(x)=\left(q+pe^{j\omega}\right)^K$

Table 7.12: Moments of order k of some probability distributions

Distribution	**Moment of order k, $E\{x^k\}$, $k\geq 0$**
Uniform	$\frac{1}{k+1}\frac{b^{k+1}-a^{k+1}}{b-a}$
Gaussian	$\frac{1}{\sqrt{\pi}}\sum_{n=0\ (neven)}^{k}\binom{k}{k-n}2^{n/2}\eta_x^{k-n}\sigma_x^n\Gamma\left(\frac{n+1}{2}\right)$
Rayleigh	$(2\sigma^2)^{k/2}\Gamma\left(1+\frac{k}{2}\right)$
Exponential	$2^k\sigma^{2k}k!$
Rice	$(2\sigma^2)^{k/2}e^{-s^2/2\sigma^2}\Gamma\left(1+\frac{k}{2}\right)\ {}_1F_1\left(1+\frac{k}{2},1;\frac{s^2}{2\sigma^2}\right)$
Lognormal	$e^{k\eta}e^{\frac{1}{2}k^2\sigma^2}$

Binomial	$\sum_{n=0}^{K-1} n^k \binom{K-1}{n} p^n q^{K-1-n}$
Note:	$\Gamma(x)$ indicates the gamma function and ${}_1F_1(\alpha,\beta;x)$ the ipergeometric function

7.2.7 Sequences

Generalizing the case of a single r.v. to the case of n random variables the sequences of random variables are introduced. A random vector is defined as a vector $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ with r.vs. as components.
The probability shall be referred to the case of a n-dimensional space in the sense that **X** must belong to a n-dimensional region D:

$$[7.57]\ P\{\boldsymbol{X} \in D\} = \int_D f(X)dX \qquad \mathbf{X} = [\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}]$$

where

$$[7.58]\ f(X) = f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1,\ldots,x_n)}{\partial x_1,\ldots,\partial x_n}$$

is the joint (multivariate) probability density function of the r.vs. $\mathbf{x}_i$ and

$$[7.59]\ F(X) = F(x_1, \ldots, x_n) = P\{\boldsymbol{x}_1 \le x_1, \ldots, \boldsymbol{x}_n \le x_n\}$$

is the joint probability distribution function.
In F(X) replacing some r.v. with $+\infty$ the joint probability distribution function of the remaining random variables is obtained. Instead, integrating $f(x_1, \ldots, x_n)$ with respect to some variables, the joint probability density function of the remaining random variables is obtained.

7.2.8 Transformations

Considering k functions $g_1(X), \ldots, g_k(X)$ with $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ the random variables $\mathbf{y}_1 = g_1(\mathbf{X}), \ldots, \mathbf{y}_k = g_k(\mathbf{X})$ can be generated.
If k<n as first step the joint probability density function of the n r.v. $\mathbf{y}_1, \ldots, \mathbf{y}_k, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_n$ is determined and then the integral is calculated on the variables x_i to eliminate them.
If k>n the r.v. $\mathbf{y}_{n+1}, \ldots, \mathbf{y}_k$ can be expressed in terms of $\mathbf{y}_1, \ldots, \mathbf{y}_n$ (thus k=n can be assumed).
To carry out the joint density function $f_y(y_1, \ldots, y_n)$ of the random vector $\mathbf{Y}=[\mathbf{y}_1, \ldots, \mathbf{y}_n]$ the equation system $g_1(X)=y_1, \ldots, g_n(X)=y_n$ is solved. If there are no solutions, $f_y(y_1, \ldots, y_n)=0$. If there is just one solution:

$$[7.60]\ f(y_1, \ldots, y_n) = \frac{f(x_1,\ldots,x_n)}{|J(x_1,\ldots,x_n)|}$$

with

$$[7.61]\ J(x_1, \ldots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

being the Jacobian of the transformation.

7.2.9 Alternative transformations

Alternatively, it is possible to carry out the probability density of a random variable **y** function of another variable **x** knowing the pdf of **x** and the functional relation between the two, in the following way.
Consider that **x** is a r.v. with known pdf $p_x(x)$, **y** is the r.v. with pdf $p_y(y)$ to figure out and y=f(x) is the relation between the two variables.

If f(x) is growing monotone $p_y(y)dy=p_x(x)dx$ and as a consequence

$$[7.62]\ p_y(y) = p_x(x)\left(\frac{dy}{dx}\right)^{-1} = \frac{p_x(x)}{f\prime(x)}$$

Broadening the model including decreasing functions

$$[7.63]\ p_y(y) = p_x(x)\left|\left(\frac{dy}{dx}\right)^{-1}\right| = \frac{p_x(x)}{|f\prime(x)|}$$

Finally, including both growing and decreasing monotone segmented functions, the generalized model can be expressed as:

$$[7.64]\ p_y(y) = \sum_{i=1}^{m} p_x(x_i)\left|\left(\frac{dy}{dx}\right)_{x_i}^{-1}\right| = \sum_{i=1}^{m}\frac{p_x(x_i)}{|f\prime(x_i)|}$$

with $x_i=f^{-1}(y)$, i=1, 2, …, m.

A further interesting case concerns the vectorial variables. The transformation formula can be extended as follows. Consider two random vectors $\underline{x}=[x_1, x_2, \ldots, x_n]^T$ and $\underline{y}=[y_1, y_2, \ldots, y_n]^T$ with probability density functions respectively $p_{\underline{x}}(\underline{x})$ and $p_{\underline{y}}(\underline{y})$. Moreover, the functions $y_i=f_i(\underline{x})$, with i=1, 2, …, n, are assigned. Assuming that all these functions are monotone, every f_i can be inverted, that means:

$$[7.65]\ x_i = f_i^{-1}\left(\underline{y}\right) \quad i = 1, 2, \ldots, n$$

Then, the transformation is carried out:

$$[7.66]\ p_{\underline{y}}(\underline{y})=|\mathbf{J}_y|p_{\underline{x}}(\underline{x}) \text{ with } \underline{x}=f^{-1}(\underline{y})$$

and $|\mathbf{J}_y|$ is the Jacobian determinant of the transformation, corresponding to the matrix

$$\boldsymbol{J}_y = \begin{vmatrix} \frac{\partial f_1^{-1}}{\partial y_1} & \cdots & \frac{\partial f_n^{-1}}{\partial y_1} \\ \cdots & \cdots & \cdots \\ \frac{\partial f_1^{-1}}{\partial y_n} & \cdots & \frac{\partial f_n^{-1}}{\partial y_n} \end{vmatrix}$$

To find $p_y(y)$ also the characteristic function can be utilized. It can be written starting from [7.44]

$$[7.67]\ \Phi_y(\omega) = E\{e^{j\omega y}\} = \int_{-\infty}^{\infty} p_y(y)e^{j\omega y}dy = E\{e^{j\omega g(x)}\} = \int_{-\infty}^{\infty} p_x(x)e^{j\omega g(x)}dx$$

If the integral can be written as

$$[7.68]\ \int_{-\infty}^{\infty} h(y)e^{j\omega y}dy$$

then

$$[7.69]\ f_y(y)=h(y).$$

7.2.10 Independence

The random variables $\mathbf{x}_1$, …, $\mathbf{x}_n$ are mutually independent if the events $\{\mathbf{x}_1\leq x_1\}$, …, $\{\mathbf{x}_n\leq x_n\}$ are mutually independent as well. As a consequence:

$$[7.70]\ F(x_1, \ldots, x_n) = F(x_1)\cdots F(x_n)$$

$$[7.71]\ f(x_1, \ldots, x_n) = f(x_1)\cdots f(x_n)$$

The random variables can be independent per groups. The resulting relation is the following:

$$[7.72]\ f(x_1, \ldots, x_n) = f(x_1, \cdots, x_m)\cdots f(x_{m+1}\cdots x_n)$$

that can be generalized to the case of more than two groups.

7.2.11 Mean, variance and covariance

In case of two random variables **x** and **y** and a function of the two variables g(**x**,**y**) the r.v. **z**=g(**x**,**y**) can be derived. The expected value can assume the following expression:

$$[7.73]\ E\{\boldsymbol{z}\} = \int_{-\infty}^{\infty} z f_z(z) dz = E\{g(\boldsymbol{x},\boldsymbol{y})\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y) dx dy$$

Generalizing the [7.73] to the case of n random variables

$$[7.74]\ E\{\boldsymbol{y}\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\dots,x_n) f(x_1,\dots,x_n) dx_1 \cdots dx_n$$

The covariance can be calculated for pairs of random variables and for real variables can be expressed as:

$$[7.75]\ C_{ij}=E\{(x_i-\eta_i)(x_j-\eta_j)=E\{x_ix_j\}-E\{x_i\}E\{x_j\}$$

The variance of the single r.v. $\mathbf{x}_i$ can be obtained considering C_{ii}.

$$[7.76]\ \sigma_i^2 = C_{ii} = E\{|\boldsymbol{x}_i - \eta_i|^2\} = E\{|\boldsymbol{x}_i|^2\} - |E\{\boldsymbol{x}_i\}|^2$$

The random variables are uncorrelated if $C_{ij}=0$ for every $i\neq 0$. As a aconsequence if $\mathbf{x}=\mathbf{x}_1+\dots+\mathbf{x}_n$ then $\sigma_x^2 = \sigma_1^2 + \cdots + \sigma_n^2$.
Taking all the C_{ij} and creating a matrix $C_n=\{C_{ij}\}$, it is called covariance matrix. Instead, taking the R_{ij} and putting them in a matrix $R_n=\{R_{ij}\}=\{E\{x_ix_j\}\}$, it is called correlation matrix.
If the r.vs. $\mathbf{x}_1,\dots,\mathbf{x}_n$ are independent they are also uncorrelated.

7.2.12 Conditional densities

Extending the definition of conditional density for one variable to the case of several random variables, the conditional density of the random variables $\mathbf{x}_n,\dots,\mathbf{x}_{k+1}$, assuming the condition $\mathbf{x}_k,\dots,\mathbf{x}_1$, can be expressed as

$$[7.77]\ f(x_n,\cdots,x_{k+1}|x_k,\cdots,x_1) = \frac{f(x_1,\cdots,x_k,\cdots,x_n)}{f(x_1,\cdots,x_k)}$$

To obtain the probability distribution function it is necessary to integrate

$$[7.78]\ F(x_n,\cdots,x_{k+1}|x_k,\cdots,x_1) = \int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_{k+1}} f(\alpha_n,\cdots,\alpha_{k+1}|x_k,\cdots,x_1)\, d\alpha_{k+1}\cdots d\alpha_k$$

Utilizing [7.77] and [7.78] the conditional moments can be calculated.

7.2.13 Characteristic function

The characteristic function of a vector of random variables is defined as follows:

$$[7.79]\ \Phi(\Omega) = E\left\{e^{-j\Omega X'}\right\} = E\left\{e^{j\sum_i \omega_i x_i}\right\} = \boldsymbol{\Phi}(\mathrm{j}\Omega)$$

where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ and $\Omega=[\omega_1, \dots, \omega_n]$.

7.2.14 Complex random variables

The concepts introduced for real random variables can be extended to be applied to complex variables. A complex random variable can be expressed as **z**=**x**+j**y** where both **x** and **y** are real random variables. The statistical behaviour and the relative parameters of a series of complex random variables $\mathbf{z}_i=\mathbf{x}_i+\mathrm{j}\mathbf{y}_i$ can be achieved utilizing the joint distribution functions of the random variables $\mathbf{x}_i$ and $\mathbf{y}_i$ indicated with $f(x_1,y_1, x_2,y_2, \dots, x_n,y_n)$. The random variables z_i are statistically independent if

$$[7.80]\ f(x_1,y_1, x_2,y_2, \dots, x_n,y_n)= f_1(x_1,y_1)\, f_2(x_2,y_2) \cdots f_n(x_n,y_n)$$

7.3 GENERALITIES ON THE STOCHASTIC PROCESSES

7.3.1 Generalities and definitions

A **stochastic process** **x**(t) is a rule to assign a function **x**(t, ζ) to every outcome of an experiment ζ. A stochastic process is a family of functions of the time that depends on the parameter ζ. Every function is named **realization** or **sample**.
The domain of ζ are all the possible outcomes of the experiment S and the domain of t is the set of the real numbers.
The process can be continuous time if the time domain is continuous or discrete time if the time domain is the set of integer numbers.
The process can be with continuous states (**x**(t, ζ) assumes non countable values) or with discrete states (**x**(t, ζ) assumes countable values).
The dependance on ζ can be omitted in the notations so that the process can be indicated with just **x**(t).
A stochastic process can be interpreted as follows:

- if t and ζ are both variables, then **x**(t) is a set of functions **x**(t, ζ) and ζ outcome of the experiment;
- if ζ is fixed, then **x**(t) is a function of the time, which is a realization (sample) of the process;
- if t is fixed, then **x**(t) is a random variable showing the state of the process at the instant t;
- if both t and z are fixed, then **x**(t) is a number.

A classical example is represented by the voltage variation

$$[7.81]\ \mathbf{x}(t) = \mathbf{r}\cos(\omega t+\boldsymbol{\phi})$$

where the random quantities are the amplitude **r** and the phase **ϕ**. A sample is the full expression which includes the dependance on the events ζ:

$$[7.82]\ \mathbf{x}(t, \zeta_i) = \mathbf{r}\,(\zeta_i)\cos(\omega t+ \boldsymbol{\phi}(\zeta_i)).$$

A second example of great interest in the digital transmission systems used to characterize the format of the signals actually utilized is the following:

$$[7.83]\ s(t) = \sum_k a_k p(t - kT)$$

where p(t) is an energy waveform (typically known) and a_k are the random variables representing a discrete time stochastic process which determines the amplitude of the waveform. Such a kind of expression will be deeply approached in section 7.5.

7.3.2 Density and distribution probability functions of a stochastic process

A stochastic process is actually an infinite noncountable set of random variables, one for any different value of t. For a particular value of t a stochastic process **x**(t) is a r.v. with cumulative probability distribution function

$$[7.84]\ F(x, t)=P\{\mathbf{x}(t)\leq x\}$$

The function depends also on t in addition to x and, according to its definition, it represents the probability of the event $\{\mathbf{x}(t)\leq x\}$, which includes all the outcomes ζ such that for a certain instant t the values **x**(t, ζ_i) of the process are not greater than x. F(x, t) is named first order distribution function. The partial derivative with respect to x of the [7.84] defines the first order density function:

$$[7.85]\ f(x, t) = \frac{\partial F(x,t)}{\partial x}$$

The joint distribution function corresponds to the second order distribution function

$$[7.86]\ F(x_1,x_2;t_1,t_2) = P\{\mathbf{x}(t_1)\leq x_1;\ \mathbf{x}(t_2)\leq x_2)\}$$

with $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ r.v. As a consequence, the second order density function is

$$[7.87]\ f(x_1,x_2;t_1,t_2) = \frac{\partial^2 F(x_1,x_2;t_1,t_2)}{\partial x_1 \partial x_2}.$$

By definition and for the consistency condition the following relations are valid

$$[7.88]\ F(x_1;t_1) = F(x_1,\infty;t_1,t_2)$$

$$[7.89]\ f(x_1;t_1)= \int f(x_1,x_2;t_1,t_2)\ dx_2$$

while the distribution function of order n, indicated with $F(x_1,\ldots,x_n;t_1,\ \ldots,t_n)$, is the joint cumulative of the n r.vs. $x_1,\ldots,x_n$, and the probability density of order n is indicated with $f(x_1,\ldots,x_n;t_1,\ \ldots,t_n)$ and $f_{n-1}= \int f_n\ dx_n$.

7.3.3 Properties

To know the statistical characteristics of a real random process it is important the knowledge of the cumulative distribution function of order n, $F(x_1,\ \ldots,\ x_n;t_1,\ \ldots,\ t_n) = P\{\mathbf{x}(t_1)\leq x_1;\ \ldots,\ \mathbf{x}(t_n)\leq x_n)$, for every x_i, t_i and n. The knowledge of this function for every n allows to fully characterize the process. The joint statistics of two stochastic processes **x**(t) and **y**(t) can be determined from the joint distribution function of the random variables $\mathbf{x}(t_1)$, …, $\mathbf{x}(t_n)$, $\mathbf{y}(t_1')$, …, $\mathbf{y}(t_n')$.
A complex stochastic process $\mathbf{z}(t) = \mathbf{x}(t) + j\mathbf{y}(t)$ is described by the joint distribution function of the component processes **x**(t) and **y**(t).

7.3.4 Moments and related properties

For some applications only some statistical quantities are important, as already verified for the random variables.
The mean η(t) is the expected value of the random variable **x**(t)

$$[7.90]\ \eta(t) = E\{x(t)\} = \int_{-\infty}^{\infty} x f(x,t) dx$$

The autocorrelation $R(t_1,t_2)$ is the expected value of the product $\mathbf{x}(t_1)\mathbf{x}(t_2)$

$$[7.91]\ R(t_1,t_2) = E\{x(t_1)x(t_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1,x_2;t_1,t_2) dx_1 dx_2$$

where x_1, x_2 are the values that the corresponding random variables can assume and the integral is calculated over the area in which the two random variables $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ can assume the values x_1 and x_2. If $R(t_1,t_2)$ is calculated on the diagonal $t_1=t_2=t$ the average power is achieved

$$[7.92]\ P(t)=R(t,t)=E\{\mathbf{x}^2(t)\}$$

The autocovariance function $C(t_1,t_2)$ of a real process **x**(t) corresponds to the covariance of the r.vs. $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$

$$[7.93]\ C(t_1,t_2) =E\{(x(t_1)-\eta(t_1))(x(t_2)-\eta(t_2))\} =R(t_1,t_2)- \eta(t_1)\eta(t_2)$$

and the value on the diagonal C(t,t), that means $t=t_1=t_2$, corresponds to the variance of the process **x**(t).

$$[7.94]\ C(t,t)=E\{|\mathbf{x}(t)-\eta(t)|^2\} = \int|\mathbf{x}(t)- \eta(t)|^2 f(x;t)dx=\sigma^2(t)=P(t)-|\eta(t)|^2$$

The complex process is characterized by the expression $\mathbf{z}(t)=\mathbf{x}(t)+j\mathbf{y}(t)$ and its statistical properties are defined in terms of the joint statistical parameters of the two component processes.
The autocorrelation of a complex process is expressed as

$$[7.95]\ R(t_1,t_2) \triangleq R_{xx}(t_1,t_2) = E\{\mathbf{x}(t_1)\mathbf{x}^*(t_2)\}=\iint x_1 x_2^* f_2(x_1,x_2;t_1,t_2) dx_1 dx_2$$

The autocovariance of a complex process is expressed as

[7.96] $C(t_1,t_2) \triangleq C_{xx}(t_1,t_2) = R(t_1,t_2) - \eta(t_1)\eta^*(t_2) = \iint (x_1 - E\{x_1\})(x_2 - E\{x_2\})^* f_2(x_1,x_2;t_1,t_2)dx_1dx_2$

If $C(t_1,t_2)$ is equal to zero for $t_1 \neq t_2$ the process is named White Process. If also $\eta=0$ then it is called White Noise.
The autocorrelation function has the following properties:

- $R(t_2,t_1)=E\{\mathbf{x}(t_2)\mathbf{x}^*(t_1)\}=R^*(t_1,t_2)$;
- $P(t)=R(t,t)=E\{|\mathbf{x}(t)|^2\}\geq 0$ is the average power of the process;
- $R(t_1,t_2)$ is positive definite function but also starting from a function $R(t_1,t_2)$ positive definite it is possible to find a process $\mathbf{x}(t)$ having an autocorrelation function equal to $R(t_1,t_2)$.

As concerns the autocovariance of the process, $C(t_1,t_2)$, it can be defined also as the correlation function of the centered stochastic process $\underline{\mathbf{x}}(t)=\mathbf{x}(t)-\eta(t)$.
The correlation coefficient of the process $\mathbf{x}(t)$ can be defined as the ratio:

[7.97] $$r(t_1,t_2) = \frac{C(t_1,t_2)}{\sqrt{C(t_1,t_1)C(t_2,t_2)}}$$

The correlation coefficient can be also considered the autocovariance of the normalized process

$$\breve{\boldsymbol{x}}(t) = \frac{x(t)}{\sqrt{C(t,t)}}.$$

In case of two processes the joint statistical characteristics can be defined in terms of the cross correlation function

[7.98] $R_{xy}(t_1,t_2)=E\{\mathbf{x}(t_1)\mathbf{y}^*(t_2)\}$

and of the cross covariance function

[7.99] $C_{xy}(t_1,t_2) = R_{xy}(t_1,t_2) - \eta_x(t_1)\eta_y^*(t_2)$.

The two processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are called mutually orthogonal if $R_{xy}(t_1,t_2)=0$ for every t_1 and t_2. Instead, they are uncorrelated if $C_{xy}(t_1,t_2)=0$ for every t_1 and t_2. For a white noise two values $\mathbf{n}(t_i)$ and $\mathbf{n}(t_j)$ are uncorrelated for every t_i and $t_j \neq t_i$.

7.3.5 Discrete time stochastic processes

The introduced concepts and definitions of statistical mean, autocorrelation function, autocovariance function, cross correlation function and cross covarianve function can be easily applied and extended to the discrete time processes expressed as x(n), with n integer number.
As concerns mean, autocorrelation and autocovariance, similarly to the expressions of the r.vs.

- Mean $\eta(n)=E\{x(n)\}$
- Autocorrelation $R(n_1,n_2)=E\{\mathbf{x}(n_1)\mathbf{x}^*(n_2)\}$
- Autocovariance $C(n_1,n_2)=R(n_1,n_2) - \eta(n_1)\eta^*(n_2)$

In addition, $\mathbf{x}(n)$ is white if $C(n_1,n_2)=q(n_1)\delta(n_1-n_2)$ with $q(n)=R(n,n)$
If the discrete process is the result of the sampling of a continuous process x(t) with sampling period T such that $x(n)=x(t)|_{t=nT}$, then:

- ➔ $\eta(n)=\eta(t)|_{t=nT}$
- ➔ $R(n_1,n_2)=R(t_1,t_2)|_{t1=n1T,\ t2=n2T}$

Of course, in order to get the mean function of the continuous process starting from the mean function of the discrete process the sampling period T must be chosen respecting the Nyquist sampling theorem.

7.3.6 Stationary processes

A stochastic process is named strict sense stationary (SSS) if its statistical properties don't vary as a consequence of a shift of the time domain. In other words the processes $\mathbf{x}(t)$ and $\mathbf{x}(t+c)$ have the same statistical functions independently on the value of c.

Two processes **x**(t) and **y**(t) are named jointly stationary if the joint statistical properties of **x**(t) and **y**(t) don't vary as a consequence of a shift of the time domain, which means that are the same of **x**(t+c) and **y**(t+c) independently on the value of c. A complex process **z**(t) = **x**(t) + j**y**(t) is stationary if the two processes **x**(t) and **y**(t) are jointly stationary.
In general a process is SSS if the n^{th} order probability density function

$$[7.100]\ f(x_1, \ldots, x_n;t_1, \ldots, t_n) = f(x_1, \ldots, x_n;t_1+t_0, \ldots, t_n+t_0)\ \forall t_0\ \forall n$$

if n=1 f(x;t)=f(x).
For stationary processes at least of order 2, the second order statistics must respect the following relation:

$$[7.101]\ f(x_1,x_2;t_1,t_2) = f(x_1\ x_2;t_1+t_0, t_2+t_0)= f(x_1,x_2;\tau)\ \text{ with } \tau =t_1-t_2$$

Definitively, the probability density function depends only on τ which is the difference between t_1 and t_2, for every t_0. Moreover, also the joint probability density function of the random variables **x**(t) and **x**(t+τ) depends only on τ.
A stochastic process is named wide sense stationary (WSS) if the mean E{**x**(t)}=η is constant (and also the variance is constant) and the autocorrelation function

$$[7.102]\ R_x(\tau)= E\{\mathbf{x}(t+\tau)\mathbf{x}^*(t)\}$$

depends only on τ (and also autocovariance). Sometimes the autocorrelation function can be expressed in the symmetric form: E{**x**(t+τ/2)**x***(t-τ/2))}.
For τ=0 R(0)=E{|**x**|^2}; thus, the average **power of a stationary process doesn't depend on** τ and is equal to R(0). A SSS process is also a WSS process but, in general, it is not true the viceversa.
The autocovariance function of a WSS process assumes the expression

$$[7.103]\ C(\tau)=R(\tau)-|\eta|^2$$

and depends only on τ too, while the correlation coefficient can be expressed as

$$[7.104]\ r(\tau)=C(\tau)/C(0)$$

Two processes **x**(t) and **y**(t) are **jointly stationary in wide sense** if their cross correlation function depends only on τ

$$[7.105]\ R_{xy}(t_1,t_2)=R_{xy}(\tau)=E\{x(t+\tau)y^*(t)\}$$

and thus the cross covariance function is expressed as

$$[7.106]\ C_{xy}(\tau)=R_{xy}(\tau) - \eta_x\eta_y^*.$$

Finally, if the process **x**(t) is stationary in wide sense and is also a white noise then

$$[7.107]\ C(\tau)=q\delta(\tau)$$

The autocorrelation and cross correlation functions of stationary processes must respect the following properties:

- $R_x(-\tau)=R_x^*(\tau)$ (Hermitian symmetry, valid also for the autocovariance function)
- $R_{xy}(-\tau)=R_{yx}^*(\tau)$

In case of a complex process **z**(t)=**x**(t)+j**y**(t) with **x**(t) and **y**(t) real, then

- $R_z(\tau)=[R_{xx}(\tau)+R_{yy}(\tau)]-j[R_{xy}(\tau)-R_{yx}(\tau)]$
- $|R_{xy}(\tau)|^2\leq R_x(0)R_y(0)=P_xP_y$

If **x**(t)=**y**(t) for a stationary process at least in wide sense, then $|R_x(\tau)|\leq R_x(0)$
Still in the hypoeteses of stationarity the relations can be demonstrated:
$\eta= E\{\bar{x}\}$, $P_x = E\{P_{xx}\}$, $R_{xx}(\tau) = E\{R_{xx}(\tau)\}$, $K_{xx}(\tau) = E\{K_{xx}(\tau)\}$,
where $\bar{x}$, P_{xx}, $R_{xx}(\tau)$ and $K_{xx}(\tau)$ are the quantities relative to the generic single realization of the process, i.e. the time mean value, the time mean power, the autocorrelation function and the time autocovariance function.

7.3.7 Cyclostationary processes

A process is named **cyclostationary** in **strict sense** (or periodically stationary) if the statistical properties are invariant with respect to a time shift which is an integer multiple of a constant interval T named **period**:

$$[7.108]\ F(x_1, \ldots, x_n;t_1, \ldots, t_n) = F(x_1, \ldots, x_n;t_1+mT, \ldots, t_n+mT)\ \forall m\ \forall n$$

A cyclostazionary process is NOT stazionary because the invariance is respected only for multiples of T and not for any t_0.
The discrete time process obtained sampling the cyclostazionary process with period T is stazionary.
Stationary and cyclostationary processes are linked by the following property:
if the process **x**(t) is cyclostazionary in strict sense and α is random variable characterized by uniform distribution in (0,T) and independent on **x**(t), then the process $\underline{\mathbf{x}}(t)=\mathbf{x}(t-\alpha)$ is stazionary in strict sense and

$$[7.109]\ \bar{F}(x_1, \cdots, x_n; t_1, \cdots, t_n) = \frac{1}{T}\int_0^T F(x_1, \cdots, x_n; t_1 - \alpha, \cdots, t_n - \alpha)d\alpha$$

A random process **x**(t) is defined cyclostazionary in wide sense if

- the mean $\eta(t+nT)= \eta(t)$ is periodic of the same period T characterizing the process itself,
- the autocorrelation function $R_x(t+mT+\tau,t+mT)=R_x(t+\tau,t)$ is a function of two variables t and τ and is periodic in the variable t with the same period T characterizing the process itself.

If a process is strict sense cyclostationary then

$$[7.110]\ f(x;t+mT)=f(x;t) \text{ and } f(x_1,x_2;t_1+mT, t_2+mT)=f(x_1,x_2;t_1,t_2)$$

as a consequence it is also cyclostationary in wide sense.
Two processes **x**(t) and **y**(t), which are cyclostationary at least in wide sense are jointly cyclostazionary with period T if

$$[7.111]\ R_{xy}(t+\tau,t)=E\{x(t+\tau)y^*(t)\}$$

is a periodic function of period T.
Since mean and autocorrelation are periodic they can be represented by the Fourier series. As concerns the mean the expression is the following:

$$[7.112]\ \eta(t) = \sum_{n=-\infty}^{\infty} \eta_n e^{j2\pi\frac{nt}{T}}$$

with the coefficients of the Fourier series η_n having the following expression:

$$[7.113]\ \eta_n = \int_{n=-T/2}^{n=T/2} \eta(t)e^{-j2\pi\frac{nt}{T}}\,dt$$

As concerns the autocorrelation the expression of the Fourier series is the following:

$$[7.114]\ R_x(t+\tau,t) = R_x(t,\tau) = \sum_{n=-\infty}^{\infty} R_x^{(n/T)} e^{j2\pi\frac{nt}{T}}$$

$$[7.115]\ R_x^{(n/T)}(\tau) = \frac{1}{T}\int_{n=-T/2}^{n=T/2} \eta(t)e^{-j2\pi\frac{nt}{T}}\,dt$$

being $R_x^{(n/T)}$ the cyclic correlation function of order n of the cyclostazionary process.
A wide sense stationary process can be considered as a particular case of cyclostationary process with $R_x^{(n/T)}(t)=0\ \forall n\neq 0$. Thus, $R_x^{(0)}(\tau)=R_x(\tau)$ is the correlation function of the stationary process in wide sense.

7.3.8 Ergodic processes

It can be very important to carry out the statistical properties of a random process leveraging on data extracted from the available ones that must necessarily come from the realizations (samples)

of the process itself. In fact, such data are deterministic and thus measurable. As an example, to evaluate the mean of a process x(t) it is necessary to observe several realizations (samples) simultaneously and to achieve the statistical mean in the following way

$$[7.116]\ \eta(t) \cong \frac{1}{n}\sum_i x(t,\zeta_i)$$

but taking into account that actually infinite realizations (samples) would be necessary. In case just one realization (sample) is available, it is possible to calculate only the temporal mean over the entire time axis with respect to it.
An interesting subset of random processes is that in which the statistical means are equal to the time means, so that with just one or a few realizations it is possible to get the statistical properties. In general, the time mean cannot be determined (time domain ∞), but of course also in this case an approximation can be well accepted. If the mean is constant, then it is possible to estimate the moments starting from the realizations.

7.3.9 Spectral theory

The concept of power spectrum for random signals can be interpreted as a transformation of averages (statistical parameters such as mean and autocorrelation). The transformation (intended as the mathematical operator) is deterministic but the quantities that are the subject of transformation are actually the expected value of relative parameters. Only stationary processes will be dealt with.
The power spectrum or power spectral density of a stationary process (WSS) **x**(t) is defined as the Fourier transform of the autocorrelation function:

$$[7.117]\ S(\omega) = \int_{-\infty}^{\infty} R(\tau)e^{-j\omega\tau}d\tau$$

S(ω) is real in ω because R(-τ)=R*(τ) and if the process **x**(t) is real, then R(τ) is real and even and S(ω) is real and also even.
Applying the usual properties of the Fourier transform, and in particular the inverse formulation, it is possible to carry out the autocorrelation function starting from the power spectrum:

$$[7.118]\ R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{j\omega\tau}d\omega$$

The power of the process can be easily calculated by means of:

$$[7.119]\ R(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)d\omega$$

In case of two processes the cross power spectrum can be obtained from the Fourier transform of the cross correlation function $R_{xy}(\tau)=E\{x(t+\tau)y^*(t)\}$

$$[7.120]\ S_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau)e^{-j\omega\tau}d\tau$$

while the cross correlation function can be achieved by means of the inversion formula:

$$[7.121]\ R_{xy}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xy}(\omega)e^{j\omega\tau}d\omega$$

The cross power spectrum density function $S_{xy}(\omega)$, in general, is complex even in the case the two processes **x**(t) and **y**(t) are real; moreover, since $R_{xy}(-\tau)=R_{yx}^*(\tau)$, then $S_{xy}(\omega)=S_{yx}^*(\omega)$.
Wiener Kinchine theorem
Considering a realization (sample) of a stochastic process x(t) it is possible to calculate the Fourier transform truncated in the interval (-T/2,T/2):

$$[7.122]\ X_T(\omega) = \int_{-T/2}^{T/2} x(t)e^{-j\omega t}dt$$

$X_T(\omega)$ is the spectral density function relative to the selected realization (sample) and is another stochastic process depending on the variable ω. Once ω is fixed $X_T(\omega)$ is a random variable. The Wiener Kinchine theorem can be expressed as:

$$[7.123]\ S_x(\omega) = \lim_{T\to\infty} \frac{1}{T} E\{|X_T(\omega)|^2\}$$

Definitively, the power spectral density of the process is the result of the average over all the possible spectral densities of all the realizations (samples) of the process. Moreover, the following relation is valid:

$$[7.124]\ S_x(\omega)=F\{R_x(\tau)\}$$

Spectral theory for cyclostationary processes

In the case of a cyclostationary process the power spectral density function $S_x(\omega)$ is defined as the Fourier transform of the autocorrelation function of order 0 relative to the process:

$$[7.125]\ S_x(\omega) = S_x^{(0)}(\omega) = F\left\{R_x^{(0)}(\tau)\right\} = \int_{-\infty}^{\infty} R_x^{(0)}(\tau) e^{-j\omega\tau} d\tau$$

Spectral theory for discrete time processes

Also in the case of discrete random processes the power spectrum can be defined as follows:

$$[7.126]\ W(\omega) = \sum_{m=-\infty}^{\infty} R_x(m) e^{-j\omega mT}$$

where $R_x(m)$ is the autocorrelation function of the process **x**(n).

The [7.126] shows that W(ω) is periodic with 2Ω=2π/T as the period. The inverse formula gives the autocorrelation function from the power spectrum:

$$[7.127]\ R_x(m) = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} W(\omega) e^{j\omega mT} d\omega$$

The power of the discrete time process can be expressed as:

$$[7.128]\ R_x(0) = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} W(\omega) d\omega$$

In the case of sequences the zeta transform is often used. Applying the z transform to the autocorrelation function of the discrete time process, it is possible to obtain:

$$[7.129]\ W_x(z) = \sum_{m=-\infty}^{\infty} R_x(m) z^{-m}$$

and the relation $S_x(z)\big|_{z=e^{-j\omega T}} = W(\omega)$ is valid.

The constant T is assumed to be the sampling period of a continouos process to obtain $x(n)=x(t)|_{t=nT}$. As a consequence $R_x(m)= R_x(\tau)|_{\tau=mT}$. The relationship between the spectrum of the continuous process and the spectrum of the discrete time (sampled) process can be achieved thanks to the Poisson formula:

$$[7.130]\ W(\omega) = \frac{1}{T} \sum_{n=-\infty}^{\infty} S(2\pi f + 2n\Omega), \quad \Omega = \frac{\pi}{T}$$

Definitively, the spectrum of the discrete time process is the sum of replicas shifted in frequency of the spectrum of the time continuous process. As a consequence, it is necessary that the value of T satisfies the requirements of the Nyquist theorem so that aliasing among the different replicas are avoided.

7.3.10 Transformations of stochastic processes through systems

Considering a stochastic process **x**(t) and a transformation T[·] which can be applied to the process **x**(t), it is possible to create a correspondence between each sample (realization) **x**(t, ζ_i) and a function **y**(t, ζ_i) which actually will be the samples (realizations) of a new stochastic process **y**(t). The relation can be expressed as:

$$[7.131]\ \mathbf{y}(t)= T[\mathbf{x}(t)]$$

As a matter of fact, it is possible to consider a system, fully characterized by the rule (transformation, operator) T, in which the process **y**(t) is the output to an input **x**(t).
The system is deterministic if it operates only on the time variable t and ζ is considered as a parameter. The consequence is that if two samples (realizations) of **x**(t), for example **x**(t, ζ_1) and **x**(t, ζ_2), are identical as a function of t, then also the samples of **y**(t) corresponding to the two inputs, **y**(t, ζ_1) and **y**(t, ζ_2), are identical in t. Alternatively, the system is stochastic if the transformation T operates on both variables so that even if **x**(t, ζ_1) and **x**(t, ζ_2) are identical the two outputs, **y**(t, ζ_1) and **y**(t, ζ_2), may be different.
If the system is considered as composed of physical components or characterised by equations, instead of as input-output as above, the classification can depend on the deterministic or stochastic nature of the components or of the coefficients of the equations.
Practical systems are all deterministic and such will be considered hereinafter. As a consequence, the statistical properties of the output process in general will be expressed in terms of the statistical properties of the input process.

Linear systems

A significant subset of systems are those for which the linearity property is respected, which implies that the transformation is linear. In general, the input-output relation is the following:

$$[7.132]\ \boldsymbol{y}(t) = \int_{-\infty}^{\infty} h(t,u)\boldsymbol{x}(u)du$$

where h(t,u) is the nucleous of the transformation and depends on t. The operator (integral) is deterministic.
An interesting case concerns the time invariant transformations for which h(t,u)=h(t-u) so that the expression of the output signal is:

$$[7.133]\ \boldsymbol{y}(t) = \int_{-\infty}^{\infty} h(t-u)\boldsymbol{x}(u)du = \int_{-\infty}^{\infty} h(\tau)\boldsymbol{x}(t-\tau)du = h(t) * x(t)$$

which is the typical convolution between the input and the impulse response h(t)=T[δ(t)].
For linear systems the stationarity properties are kept from the input to the output. If the input process **x**(t) is strict sense stationary also the output process **y**(t) is strict sense stationary. If the input process **x**(t) is wide sense stationary, the processes **x**(t) and **y**(t) are jointly WSS.
The mean of the transformation of the stochastic process **x**(t) is equal to the transformation of the mean of the stochastic process **x**(t):

$$[7.134]\ E\{T[\mathbf{x}(t)]\} = T\{E[\mathbf{x}(t)]\}$$

or, equivalently, $\eta_y(t)=T[\eta_x(t)]$.
For stationary processes it is possible to demonstrate that

$$[7.135]\ R_y(\tau)=R_x(\tau) \ast r_h(-\tau)$$

with

$$[7.136]\ r_h(\tau) = \int_{-\infty}^{\infty} h\left(t+\frac{\tau}{2}\right) h\left(t-\frac{\tau}{2}\right) dt$$

and that, utilising the Wiener Kinchnine theorem,

$$[7.137]\ S_y(f)=|H(f)|^2 S_x(f)$$

Moreover,

$$[7.138]\ \eta_y=H(0)\eta_x$$

with

$$[7.139]\ H(0) = \int_{-\infty}^{\infty} h(\tau)d\tau$$

7.3.11 The Gaussian process

Two r.v. **x** and **y** are jointly normal if

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}}\exp\left\{-\frac{1}{2\left(1-\rho_{12}^2\right)}\left[\frac{(x-\eta_1)^2}{\sigma_1^2}-\rho_{12}\frac{(x-\eta_1)(y-\eta_2)}{\sigma_1\sigma_2}+\frac{(y-\eta_2)^2}{\sigma_2^2}\right]\right\}$$

with η_i means, σ_i standard deviation, ρ_{12} cross correlation coefficient.
The marginal density functions are

$$f_x(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\eta_1)^2}{2\sigma_1^2}}\quad f_y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\eta_2)^2}{2\sigma_2^2}}$$

In the case of n variables

$$f(X)=\frac{1}{(2\pi)^{n/2}\sqrt{\Delta}}\exp\left\{-\frac{1}{2}(X-\Theta)^T C^{-1}(X-\Theta)\right\}$$

C is the covariance matrix of the n r.v., $\Delta=\det\mathbf{C}$ and $\Theta=[\eta_i, \ldots, \eta_n]^T$. If the r.v. are uncorrelated C is diagonal.
Alternative definition: the n random variables are jointly Gaussian if and only if for every value of the coefficients $a_1, \ldots, a_n$ the random variable $\mathbf{y}=a_1\mathbf{x}_1, \ldots, a_n\mathbf{x}_n=a^T\mathbf{X}$ has a Gaussian probability density.

Characteristics of the gaussian random processes

Considering a set of time instants $t_1, \ldots, t_n$ the joint density function of order n is expessed as

$$f(X)=\frac{1}{(2\pi)^{n/2}\sqrt{\Delta}}\exp\left\{-\frac{1}{2}\left(X-\Theta(t_1,...,t_n)\right)^T C_x^{-1}(t_1,...,t_n)\left(X-\Theta(t_1,...,t_n)\right)\right\}$$

where $\Theta=[\eta_x(t_1), \ldots, \eta_x(t_n)]^T$ and $\mathbf{C}_x(t_1, \ldots, t_n)$ is the covariance matrix which contains the elements $C_{x,ij}=\mathrm{cov}(\mathbf{x}_i,\mathbf{x}_j)=\mathrm{cov}(\mathbf{x}(t_i),x(t_j))=C_{x,ij}(t_i,t_j)$.
Properties

- $\eta_x(t)$ and $C_x(t_i,t_j)$ for every pair (t_i,t_j) allow to evaluate the joint density functions of the process for any order n
- If the process is clyclostationary in wide sense it is clyclostationary in strict sense as well
- Two processes are jointly gaussian if their probability density of order n $f_{xy}(x_1, \ldots, x_n;y_1, \ldots, y_n)$ is of gaussian type

If $\mathbf{C}_x(t_i, t_j) = \sigma^2(t_i)\,\delta(t_i-t_j)$ the Gaussian process is named *white* or *white noise*, in general non stationary. If $\mathbf{C}_x(t_i, t_j) = N_0\delta(t_i-t_j)$ the process is *stationary white Gaussian*.

Linear transformation between Gaussian processes

If a linear transformation is applied to a normal process **x**(t) also the process **y**(t) is normal. If two processes in input to two linear systems are jointly Gaussian, the relative outputs are jointly gaussian as well.

Stationary white process

$$[7.140]\quad S_x(f)=F\{R_x(\tau)\}=F\{N_0\delta(\tau)\}=N_0\quad \forall f$$

with N_0 (Watt/Hz) power spectral density of the stationary white process, which provides a limited value:

$$[7.141]\quad P=\int_{-\infty}^{\infty}S_x(f)df=\int_{-\infty}^{\infty}N_0 df$$

7.3.12 Random signals and their sources

Sometimes signals were mentioned as vehicle of the information: in this respect, nevertheless, it is necessary to point out that they are ***random signals***, because of course the deterministic ones, implying that they are known over the whole time evolution, cannot carry any information. Often in the telecommunication operations signals with time evolution that cannot be characterized as

deterministic, as those ones recalled in the previous chapters, are utilized; then signals of random type are considered and dealt with according a statistical approach. It is necessary to consider all the possible signals with non zero probability to be observed, because in general each of them considered as stand alone is not sufficiently representative. The aim of this section is to introduce some brief note on the information flow sources, also for the best understanding, in stochastic sense, of the information content as well as of the random signals characteristics, object of the analysis performed in the next sections.
First of all, the elementary transmission of an information can be identified with the choice of the value of a real ***random variable***, hereinafter indicated with the notation ***r.v.*** The ***source*** is named ***analogue*** if the r.v., Z, is continuous, implying that its generic ***determination***, z, can assume any of the non countable values within a known interval, usually finite; instead, it is named ***digital***, if the r.v., Z_q, is discrete, meaning that its determination can assume any of the symbols which constitute a set, $\{z_q\}$ with q=1, 2,.., M, discrete, countable and finite.
Referring to the digital case, the considered elementary transmission can be reduced to the choice of a symbol, z_q, among the M possible different elements which, as a whole, compose a M-nary alphabet. To such an event the ***quantity of information*** is associated:

$$[7.142]\ I(z_q) \hat{=} \log_2 \frac{1}{P(z_q)} ,$$

where $P(z_q)$ is the probability of z_q. The occurrence of one of the two binary digits in condition of equal probability [P(0)=P(1)=½] corresponds thus to the information quantity unit, named ***bit***. If each symbol has the same probability to occur, that means:

$$[7.143]\ P(z_q) = \frac{1}{M} ,$$

and if in the $M = 2^b$ b is an integer, the quantity of information, identical, of each symbol corresponds to the number b of the binary digits that can be associated to it.
Considering the evolution of a source, the ***information flow*** is considered. In the analogue case very often a time continuous flow is obtained: then, from time to time a sample function x(t) can be generated, which constitutes the ***realization*** (***sample***) of a real ***continuous stochastic process***, indicated with X(t), which includes all the possible realizations; the process, for each generic assay time instant *t* is reduced to the continuous r.v. X_t=X(*t*), as the one previously introduced. In the Figura 7.1 some realizations of a continuous process are shown, highlighting the determinations of the r.v. X_t, in the instant *t*, as well as the ones of the r.v. $X_{t+\tau}$, in the instant $t+\tau$.

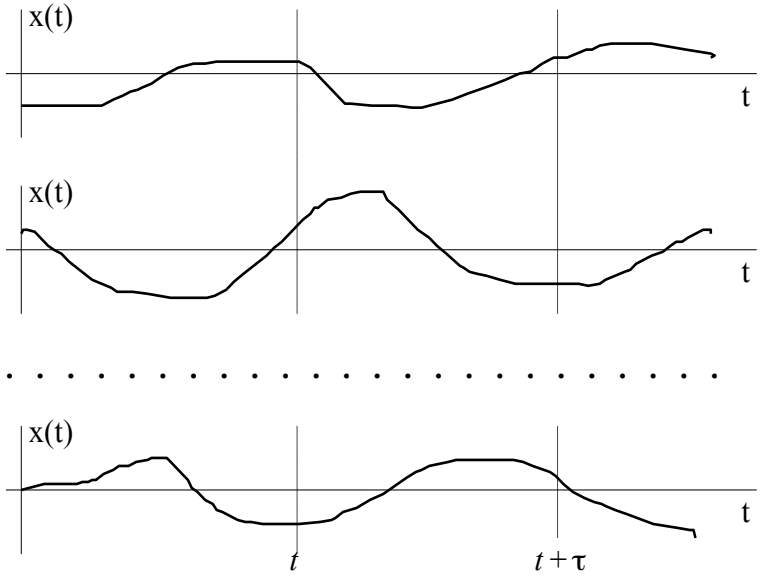


Figura 7.1 - Example of realizations of a continuous process.

Also the case of discrete flow of analogue information can occur, considering for example a source that, in each case, generates a sequence, z(n), that constitutes the ***realization*** of a real ***discrete stochastic process***, indicated with Z(n), that includes all the possible realizations; the process for each assay value k of the integer variable n is reduced to the continuous r.v. Z_k, as well as the one already previously introduced; definitively, the process is a sequence of continuous r.v.
The information flow of a digital source corresponds always to a real discrete process, $Z_q(n)$, which includes as realizations all the digital flows that can be generated from time to time, $z_q(n)$; nevertheless, differently from the immediately previous case, a ***discrete stochastic process with discrete values*** occurs, considering that for each generic assay integer value, k, it is reduced to a discrete r.v. Z_{qk}, which, as already mentioned, has one of the M possible symbols of the set $\{z_q\}$ as determination. The type of considered process is therefore a sequence of discrete r.v.
The random signals that carry the information in the telecommunication systems are obtained from the sources after proper conversion processes of the physical nature of the information flows: from that generic with which it is created to the electric one. Such processing operations are implemented by ***trasducers***, equipment that are not involved in the transmission and thus hereinafter will be ignored, being inside the sources. If the source is analogue, the waveform transmitted randomly, i.e. an ***analogue signal***, has a random evolution in the time domain similar to the one of the considered information flow: as a consequence, it is the realization of a continuous process too. Instead, if the source is digital, in that respect a ***digital signal*** is obtained, with a variety of random evolution, but anyway always associated to the considered information digital flow: neglecting the theoretical case of sampled digital signal, in practice still the realization of a continuous process is obtained, even if dependent on the discrete process with discrete values.
Definitively, all the possible physical signals having probability different from zero to cross a transmission system, but that can always be generated by sources of the same type, as a whole constitute a real ***stochastic process***, of which the single waveform, x(t), is the ***realization*** or ***sample***.

7.3.13 Characterization of continuous processes

In a real continuous process X(t), at every generic assay intant, t, a continuous r.v., $X_t=X(t)$, corresponds and the instant values, $x_t = x(t)$, of all the possible realizations are the ones that the r.v. X_t can assume. The distribution of the values x_t is characterized through a function:

$$p(x_t\,;t),$$

named ***distribution density*** or even ***probability density*** at the instant t; in fact, the value of $p(x_t;t)dx_t$ expresses the probability that the X_t assumes at the time t a value between x_t and x_t+dx_t.
More in general, the knownledge of the statistical interdependence of a set of n r.v., X_1, X_2,.., X_n, is necessary, obtained in correspondence of the assay instants t_1, t_2, .., t_n, all arbitrary. Then, the ***joint distribution density*** or also ***joint probability density*** of order n is utilised:

$$p_n = p_n\,(x_1, x_2, .., x_n; t_1, t_2, .., t_n);$$

for example, $p_2dx_1dx_2$ represents the probability that at the instants t_1 and t_2 the values assumed by X_1 and X_2 can respectively span between x_1 and x_1+dx_1 and between x_2 and x_2+dx_2. The knowledge of the function p_n is sufficient in order to know all the functions of lower order; in fact, the relation is:

$$[7.144]\quad p_{n-1} = \int p_n\,dx_n\ .$$

Usually the r.v. X_n at the instant t_n is conditioned by the possible knowledge of the r.v. extracted in correspondence of the previous instants $t_{n-1},.., t_2,t_1$; then, the ***conditioned distribution density*** or in alternative the ***conditioned probability density*** of order 1 and with n-1 conditionings is defined:

$$p_{|n-1} = p_{|n-1}(x_n;t_n|x_{n-1},..,x_2,x_1;t_{n-1},..,t_2,t_1);$$

the value of $p_{|n-1}dx_n$ expresses the probability that at the instant t_n the X_n assumes a value spanning between x_n and x_n+dx_n, under the conditioning of the values $x_{n-1},..,x_2,x_1$. Then, the following relation is obtained:

$$[7.145]\ p_n = p_{|n-1}p_{n-1}(x_{n-1},..,x_2,x_1;t_{n-1},..,t_2,t_1),$$

and reiterating the [7.145] itself:

$$[7.146]\ p_n= p_{|n-1}p_{|n-2}p_{|n-3}....p_{|1}p(x_1;t_1).$$

To fully characterize a process not always it is necessary to know the high order probability densities, as shown hereinafter.
A particularly simple case is that one in which the hypothesis of full statistical independence among all the r.v. X_t occurs, in correspondence of different assay instants: then, all the conditioned functions are reduced to the first order distribution and the [7.146] implies that the order n joint probability density is expressed simply as the product of n functions of the first order.
For particular stochastic processes, named first order ***Markoff processes***, the statistical dependence among the different values of X_t is limited to the contiguous value, implying that for the conditioned functions of any order the following property is valid:

$$[7.147]\ p_{|k-1} = p_{|1}(x_k;\ t_k \mid x_{k-1};t_{k-1});$$

from the [7.146] then the result is:

$$[7.148]\ p_n= p_{|1}(x_n;t_n \mid x_{n-1},\ t_{n-1})\cdots p_{|1}(x_2;t_2|\ x_1;t_1)p(x_1;t_1).$$

The consequence is that the processes of the considered type are fully described by the second order probability density (see [7.144]). In the Markoff processes of order n the description is complete if the probalility density of order n+1 is known.
Limiting the knowledge to the probability density of the second order, it is possible the evaluation of some quantities of a generic process, useful to synthetically represent at least its most important characteristics. Hereinafter this limitation will be considered.
The following notation for the statistical mean is adopted:

$$E\{F_n\}=\int_n F_n(v_1,v_2,..,v_n)p_n(v_1,v_2,..,v_n)dv_1\,dv_2..dv_n$$

where F_n is a function of the r.v. $V_1,V_2,..,V_n$ and the integral is of order n. At the generic assay instant t_1, considering the r.v. $X_1=X(t_1)$, the ***statistical mean value of the process*** (or mathematical hope), the ***mean statistical power of the process*** and the ***variance of the process*** are defined, respectively by means of the:

$$[7.149]\ E\{x_1\}=\int x_1 p(x_1;t_1)dx_1 \ \hat{=}\ \eta(t_1)\ ,$$

$$[7.150]\ E\{|x_1|^2\}=\int |x_1|^2\ p(x_1;t_1)dx_1 \ \hat{=}\ P(t_1)\ ,$$

$$[7.151]\ E\{|x_1- E\{x_1\}|^2\}=\int |x_1-E(x_1)|^2\ p(x_1;t_1)dx_1 \ \hat{=}\ \sigma^2(t_1)\ ,$$

which in general result to depend on the assay instant t_1.
Then the following property is valid:

$$[7.152]\ \sigma^2(t_1) = P(t_1) - |\eta(t_1)|^2\ .$$

In correspondence of the pair of generic assay instants t_1 and t_2 and by means of the statistical cross correlation between the r.v. $X_1 = X(t_1)$ and $X_2 = X(t_2)$, expressed by the probability density of the second order, the ***autocorrelation function of the process*** and the ***autocovariance function of the process*** are defined, respectively by means of the relations(1):

(1) The notation with the symbol of the conjugation operation, useless in the considered case of real process, will be useful later on when broadening to complex processes.

$$[7.153]\ \mathrm{E}\left\{x_1 x_2^*\right\} = \iint x_1 x_2^* p_2(x_1,x_2;t_1,t_2)\,dx_1\,dx_2 \ \hat{=}\ R(t_1,t_2)\,,$$

$$[7.154]\ \mathrm{E}\left\{\left[x_1-\mathrm{E}\{x_1\}\right]\left[x_2-\mathrm{E}\{x_2\}\right]^*\right\} = \iint \left(x_1-\mathrm{E}\{x_1\}\right)\left(x_2-\mathrm{E}\{x_2\}\right)^* p_2(x_1,x_2;t_1,t_2)\,dx_1\,dx_2 \ \hat{=}\ K(t_1,t_2);$$

moreover, the following property is valid:

$$[7.155]\ K(t_1,t_2) = R(t_1,t_2)-\eta(t_1)\eta^*(t_2),$$

and it is immediate to verify that for $t_2=t_1$ the following relations are obtained:

$$[7.156]\ R(t_1,t_1) = P(t_1)\ ,\quad K(t_1,t_1) = \sigma^2(t_1)\ ,$$

and thus the [7.155] in such a case is the same of the [7.152].
Note that all the introduced quantities are characteristics of the process, but in general are functions of one or both the independent time assay variables, t_1 and t_2. Furthermore, it is meaningful to observe that, thanks to the properties [7.155] and [7.156], all the considered quantities result to be known or can be figured out starting just from the knowledge of the two quantities:

$$\eta(t_1)\quad ,\quad R(t_1,t_2),$$

which thus play a fundamental role for the synthetic characterization, limited to the second order, of the process.

7.3.14 Stationary continuous processes

In the operational reality of the transmission signals, either useful or undesirable that they may be, they are realizations of stochastic processes. If very prolonged observations are considered the signal characteristics in the statistical domain (in particular, the mean value and the autocorrelation function) are not unchanged over time: just refer to the finite duration of the communications, as well as the possibility of signals, in theory all energy ones, transiting over the same connection in instants even immediately close among one another, generated by very different sources (for example either speech or data source in the sound band); on the other hand the design and the validation of the equipment and of the means which constitute the transmission systems can be based on the average performance observable in several intervals of short duration. This justifies the common practice to extrapolate, on the whole time axis, the signal characteristics, such as they are determined by statistical evaluations performed on limited time periods, as long as the duration is sufficient; in other words, the hypothesis of time invariance of the random nature of the signals in transit is often acceptable. In conclusion, it is reasonable to focus the attention on that particular category of processes, referred to as the ***continuous stationary stochastic processes***, that regardless of the characteristics which will be pointed out below, have power random signals as realizations.
In stationary continuous processes all the probability density functions are invariant with respect to a time shift; in particular, therefore, also the just above introduced synthetic quantities enjoy such a property. The most interesting stochastic processes for telecommunications, because they form a class that includes the previous ones, but it is much broader, are those ***weakly stationary*** or ***stationary in the broad sense***, for which the considered property is valid in reference to the first and second functions order; in other words, considering $t_1=t+\tau$ and $t_2=t$ with any τ, a probability density of the generic instant value $x_1 = x(t+\tau)$ independent on time:

$$[7.157]\ p(x_1;t_1) = p(x_2;t_2) = p(x_1) = p(x_2),$$

and a joint probability density of the pair of instant values $x_1=x(t+\tau)$ and $x_2=x(t)$, which depends on τ, are obtained:

$$[7.158]\ p_2(x_1,x_2;t_1,t_2) = p_2(x_1, x_2;\tau).$$

In the frame of the practice so far followed, to limit the characterization of the processes at the second order, thus, the difference between the actual stationarity and that one in broad sense cannot be realised: therefore, hereinafter such difference will be neglected, but the reader shall remind what carried out for the processes of the stricter class is valid also for the processes stationary in broad sense.

In a stationary process the statistical mean value, the power and the variance results to be constant; respectively the following equalities are obtained:

$$[7.159]\ \mathrm{E}\{x(t)\} = \int x_1 p(x_1)\,dx_1 \triangleq \eta_x\ ,$$

$$[7.160]\ \mathrm{E}\{|x(t)|^2\} = \int |x_1|^2\, p(x_1)\,dx_1 \triangleq P_x\ ,$$

$$[7.161]\ \mathrm{E}\{|x(t)-\mathrm{E}\{x(t)\}|^2\} = \int |x_1-\eta|^2\, p(x_1)\,dx_1 \triangleq \sigma_x^2\ .$$

Moreover, the autocorrelation and autocovariance functions depend on the difference τ between two assay instants and not on their absolute value, i.e. the relation is[1]:

$$[7.162]\ \mathrm{E}\left\{x(t+\tau)x^*(t)\right\} = \iint x_1 x_2^* p_2(x_1,x_2;\tau)\,dx_1\,dx_2 \triangleq R_{xx}(\tau)\ ,$$

$$[7.163]\ \mathrm{E}\left\{\left[x(t+\tau)-\mathrm{E}\left\{x(t+\tau)\right\}\right]\left[x(t)-\mathrm{E}\left\{x(t)\right\}\right]^*\right\} = \iint (x_1-\eta)(x_2-\eta)^* p_2(x_1,x_2;\tau)\,dx_1\,dx_2 \triangleq K_{xx}(\tau),$$

where both $R_{xx}(\tau)$ and $K_{xx}(\tau)$ are hermitian functions:

$$[7.164]\ R_{xx}(\tau) = R_{xx}^*(-\tau)\ ,\qquad K_{xx}(\tau) = K_{xx}^*(-\tau)\ ;$$

then, the following properties are valid:

$$[7.165]\ R_{xx}(\tau) = K_{xx}(\tau) + |\eta_x|^2\ ,\ |R_{xx}(\tau)| \le R_{xx}(0) = P_x\ ,\ |K_{xx}(\tau)| \le \sigma_x^2,$$

$$[7.166]\ R_{xx}(0) = P_x = \sigma_x^2 + |\eta_x|^2 = K_{xx}(0) + |\eta_x|^2\ .$$

Still in the hypothesis of stationarity the following relations can be demonstrated:

$$[7.167]\ \eta_x = \mathrm{E}\{\overline{x}\}\quad ,\quad P_x = \mathrm{E}\{\mathcal{P}_{xx}\}\quad ,\quad R_{xx}(\tau) = \mathrm{E}\{R_{xx}(\tau)\}\quad ,\quad K_{xx}(\tau) = \mathrm{E}\{K_{xx}(\tau)\}\ ,$$

where $\overline{x}$, $\mathcal{P}_{xx}$, $R_{xx}(\tau)$ and $K_{xx}(\tau)$ are the respective quantities relative to the generic single realization of the process, in particular the time mean value, the time mean power, the time autocorrelation function and the time autocovariance function.

It is important to observe that, utilizing the [7.165] and [7.166], all the considered quantities result to be known or they can be figured out starting from just the knowledge of the statistical mean value and of the autocorrelation function:

$$\eta_x\ ,\quad R_{xx}(\tau)\ ,$$

which thus play a fundamental role for the synthetic characterization of a stationary process.

Keeping the hypothesis of stationarity, it is possible to define the ***power spectral density*** or also ***power spectrum*** of the process, by means of the relation of Wiener-Khintchine:

$$[7.168]\ P_x(f) \triangleq F\{R_{xx}(\tau)\}\ ;$$

the following expression is obtained:

$$[7.169]\ P_x(f) = \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left\{|X_T(f)|^2\right\}\ ,$$

where $X_T(f)$ indicates the Fourier transform of the generic realization truncated in the finite observation interval T. Note that the function $P_x(f)$ is real by definition, being $R_{xx}(\tau)$ hermitian,

---

[1] The notation with the symbol of the conjugation operation, useless in the considered case of real process, will be useful later on when broadening to complex processes.

and it is non negative thanks to the [7.169]; it becomes even in the case, very frequent, of real process. Utilizing the [7.165] and the [7.168] the following expression is carried out:

$$[7.170]\quad P_x(f) = F\{K_{xx}(\tau) + |\eta_x|^2\} = F\{K_{xx}(\tau)\} + |\eta_x|^2\delta(f),$$

which highlights the existence of a discrete spectral component in the origin, linked to the square of the modulus of the statistical mean value. Calculating the inverse Fourier transform of the [7.168] in $\tau = 0$ and reminding the [7.166] the following result is obtained:

$$[7.171]\quad P_x = \int P_x(f)df\ .$$

Thanks to the Wiener-Khintchine relation, to characterize a stationary process its statistical mean value, η_x, and its power spectral density, $P_x(f)$, are sufficient. Assuming to consider this latter function, the remarks developed with respect to the spectral extension of the deterministic signals are still valid: in the case of stationary physical processes it is thus always possible to adopt the name of ***base band processes*** and ***shifted band processes***, using respectively the conditions $f_M >> f_m$ and $f_M < 2f_m$ where f_m and f_M are the extremes of the monolateral bandwidth of the process.
A particularly interesting subset of stationary processes is that one of the ***ergodic processes***, for which the single realization, x(t), observed over the whole time axis, contains all the statistical properties of the process, so that all the average time quantities converge, with probability which tends to one, to the corresponding statistical means. Indicating always with $\overline{x}$, $\mathcal{P}_{xx}$, $R_{xx}(\tau)$ and $K_{xx}(\tau)$ respectively the mean value, the power and the autocorrelation and autocovariance functions of the process, the following equalities are obtained:

$$[7.172]\quad \overline{x} = \eta_x \quad , \quad \mathcal{P}_{xx} = P_x \quad , \quad R_{xx}(\tau) = R_{xx}(\tau) \quad , \quad K_{xx}(\tau) = K_{xx}(\tau)\ .$$

Then, the knowledge of the whole time evolution of just one realization allows to determine the synthetic quantities of the stationary ergodic process to which it belongs to.

7.3.15 Cross correlation for stationary processes

Let's consider a pair of real continuous processes, X(t) and Y(t), jointly stationary, meaning that both the probability densities of the first order are time independent (see [7.157]) and the joint probability density of the pair of intant values $x_1=x(t+\tau)$ and $y_2=y(t)$ depends on the difference $\tau=t_1-t_2$ between the assay instants and not on their absolute values. In addition to the characteristic quantities already introduced for each process (hereinafter labeled by adding the single subscript x or y), similarly with the [7.162] and [7.163] the ***cross correlation function*** and the mutual ***covariance function*** of the pair of processes [1] are defined and result both to be function of τ:

$$[7.173]\ E\{x(t+\tau)y^*(t)\} \hat{=} R_{xy}(\tau)\ ,$$

$$[7.174]\ E\{[x(t+\tau)-\eta_x][y(t)-\eta_y]^*\} \hat{=} K_{xy}(\tau)\ ,$$

where η_x and η_y are the statistical mean values; exchanging between each other the subscripts in the [7.173] and [7.174], the following relations (with the variable change $z=t+\tau$) can be demonstrated:

$$[7.175]\ R_{yx}(\tau) = R^*_{xy}(-\tau) \quad , \quad K_{yx}(\tau) = K^*_{xy}(-\tau).$$

The following relation is valid between the functions:

$$[7.176]\ R_{xy}(\tau) = K_{xy}(\tau) + \eta_x \eta^*_y;$$

moreover, the following properties are valid:

---

[1] The notation with the symbol of the conjugation operation, useless in the considered case of real process, will be useful later on when broadening to complex processes.

$$[7.177]\ |R_{xy}(\tau)|^2 \le R_{xx}(0)R_{yy}(0) = P_x P_y\ ,$$

$$[7.178]\ |K_{xy}(\tau)|^2 \le K_{xx}(0)K_{yy}(0) = \sigma_x^2\ \sigma_y^2,$$

having indicated with P_x and P_y and with σ_x^2 and σ_y^2 respectively the powers and the variances of the processes.
Furthermore, defined the ***power mutual spectral density*** of the processes

$$[7.179]\ P_{xy}(f) \hat{=} F\{R_{xy}(\tau)\},$$

the relation

$$[7.180]\ P_{xy}(f) = \lim_{T\to\infty} \frac{1}{T} E\left\{X_T(f) Y_T^*(f)\right\}$$

follows, where $X_T(f)$ and $Y_T(f)$ are the Fourier transforms of the generic realizations truncated in the finite observation interval T; exchanging the subscript between each other, the following relation is obtained:

$$[7.181]\ P_{xy}^*(f) = P_{yx}(f) = F\{R_{yx}(\tau)\}.$$

From the [7.180] it is possible to note that the power mutual spectral density is different from zero only at the frequencies for which both the power spectral densities of the considered processes X(t) and Y(t) are different fro zero.
It is possible to observe that in the case of processes jointly stationary and both real, the cross correlation and covariance functions are always real functions, while the power mutual spectral density in general is a hermitian function of the frequency.
Two processes jointly stationary are ***incoherent*** (**orthogonal**) if the cross correlation function is zero for any τ, i.e. the following relation is valid:

$$[7.182]\ R_{xy}(\tau) \equiv 0;$$

utilizing the [7.179] in such a case the result is that also the power mutual spectral density is equal to zero for any frequency:

$$[7.183]\ P_{xy}(f) \equiv 0\ , \qquad \text{se:}\ \ R_{xy}(\tau) \equiv 0\ .$$

Instead, if the covariance function is zero for any τ:

$$[7.184]\ K_{xy}(\tau) = R_{xy}(\tau) - \eta_x \eta_y^* \equiv 0\ ,$$

the two processes are named ***uncorrelated***; still from the [7.179] then the power mutual spectral density results to consist in just one discrete component in the origin:

$$[7.185]\ P_{xy}(f) = \eta_x \eta_y^* \delta(f)\ ,\ \text{se:}\ K_{xy}(\tau) \equiv 0\ .$$

If the value in $\tau = 0$ of the cross correlation function is zero, i.e. the following relation is valid:

$$[7.186]\ R_{xy}(0) = E\{x(t)\, y^*(t)\} = 0,$$

The two processes in the same generic assay instant, correspond to a pair of orthogonal r.v. $X(t)$ and $Y(t)$; in such a case (see [7.181]) the result is:

$$[7.187]\ \int P_{xy}(f)\,df = \int \Re\left\{P_{xy}(f)\right\} df = 0\ , \qquad \text{se:}\ \ R_{xy}(0) = 0.$$

Two stationary continuous processes ***statistically independent***, for which the joint probability densities of various order are reduced to the products of the joint densities of each process considered separately:

$$[7.188]\ p_{n+m}(x_1, x_2, .., x_n, y_1, y_2, .., y_m) = p_n(x_1, x_2, .., x_n)\, p_m(y_1, y_2, .., y_m),$$

are always uncorrelated. If at least one has the mean value equal to zero, they are also incoherent.

7.3.16 Process sum and complex process

Let's consider the continuous process, A(t), with realizations a(t) = x(t) + y(t) given by the sum of the ones of two real jointly stationary processes, and apply the [7.159] and [7.162]. Considering the linearity of the operator statistical mean, E{•}, the mean statistical value results to be:

$$[7.189]\ \eta_a = E\{x(t)+y(t)\} = E\{x(t)\}+E\{y(t)\} = \eta_x+\eta_y.$$

Then, the autocorrelation function is obtained:

$$[7.190]\ R_{aa}(\tau)=E\{[x(t+\tau)+y(t+\tau)][x(t)+y(t)]^*\}=E\{x(t+\tau)x^*(t)\}+$$
$$E\{y(t+\tau)y^*(t)\}+E\{x(t+\tau)y^*(t)\}+E\{y(t+\tau)x^*(t)\} = R_{xx}(\tau)+R_{yy}(\tau)+R_{xy}(\tau)+R_{yx}(\tau)\ .$$

The considered summation process is thus stationary too. Moreover, thanks to the [7.169] and [7.181] the following relation is obtained:

$$[7.191]\ P_a(f) = P_x(f)+P_y(f)+2\Re\{P_{xy}(f)\}\ .$$

The incoherence (orthogonality) of the two addend processes (see [7.182] and [7.183]) is then a sufficient condition in order that the processes X(t) and Y(t) can be summed up in power, as well as in power spectral density:

$$[7.192]\ P_a = P_x+P_y\ ,\quad P_a(f) = P_x(f)+P_y(f),\quad \text{se: } R_{xy}(\tau) \equiv 0\ .$$

Moreover, the rule of the simple sum in power is valid also in the case the two real processes correspond to two orthogonal r.v., as can be demonstrated integrating the [7.191] and applying the [7.187]; thus, if the cross correlation function is equal to zero in the origin the processes are named ***additive processes in power***:

$$[7.193]\ P_a = P_x+P_y\ ,\quad \text{if: } R_{xy}(0) = \int P_{xy}(f)df=\int \Re\left\{P_{xy}(f)\right\}df=0\ .$$

On the other hand, remind that in general the power spectra cannot be summed, implying that $P_a(f)\neq P_x(f)+P_y(f)$.
Finally, note that the statistical independence between the two addend processes, which then results to be uncorrelated, allows to perform the simple sum of the powers and of the spectra only if at least one of the two processes has the mean value equal to zero (see [7.183] and [7.184]).
A second interesting case, for which the linearity of the statistical mean operator can be still utilized, is the one of the complex process, B(t), with realizations b(t)=x(t)+jy(t) where X(t) and Y(t) are still two real processes jointly stationary.
Applying the [7.159] and [7.162], the statistical mean value can be carried out:

$$[7.194]\ \eta_b = E\{x(t)+jy(t)\} = E\{x(t)\}+jE\{y(t)\} = \eta_x+j\eta_y.$$

Then, the autocorrelation function is obtained:

$$[7.195]\ R_{bb}(\tau) = E\{[x(t+\tau)+jy(t+\tau)][x(t)+jy(t)]^*\} = E\{x(t+\tau)x^*(t)\}+E\{y(t+\tau)y^*(t)\}-$$
$$jE\{x(t+\tau)y^*(t)\}+jE\{y(t+\tau)x^*(t)\}=R_{xx}(\tau)+R_{yy}(\tau)-j[R_{xy}(\tau)-R_{yx}(\tau)].$$

Thus, also the considered complex process is stationary. Moreover, thanks to the [7.168], [7.179] and [7.181] the following relation is valid:

$$[7.196]\ P_b(f) = P_x(f)+P_y(f)+2\Im\{P_{xy}(f)\}.$$

Considering $\tau=0$ in the [7.177] and [7.195], the result is that the power of the considered complex process is always the sum of the powers of the two processes that are composed:

$$[7.197]\ P_b = P_x + P_y;$$

in case they are (orthogonal) incoherent (see [7.183]) the condition is sufficient in order that the simple relation concerning the power spectral densities is valid:

$$[7.198]\quad P_b(f) = P_x(f)+P_y(f), \quad \text{if:} \quad R_{xy}(\tau) \equiv 0 \ .$$

7.3.17 Real stationary dicrete processes

7.3.17.1 Real stationary discrete processes with continous values

Considering what above recalled, it is simple to examine the case of a real discrete process, indicated with Z(n), constituted by a sequence of infinite real constinuous, Z_k, characterized by the probability density functions of various order concerning the determinations $z_k=z(k)$. In correspondence of two generic integer assay variables, k_1 and k_2, of the integer variable n, with the same formalism of the [7.149] - [7.156] the ***statistical mean value***, $\eta(k_1)$, the ***statistical mean power***, $P(k_1)$, the ***variance***, $\sigma^2(k_1)$, as well as the ***autocorrelation function***, $R(k_1,k_2)$, and the ***autocovariance function of the discrete process***, $K(k_1,k_2)$ are obtained.
In the case of interest, of the ***stationary discrete processes*** or ***stationary in broad sense***, the probability density of the first order is independent on k and that one of the second order is invariant with respect to a generic shift. As a consequence, the statistical mean value, the power and the variance result to be constant, independent on the integer assay variable, while the autocorrelation and autocovariance functions depend on the integer difference, $\nu=k_1-k_2$, of the two assay variables and not on their absolute values; then, the following expressions are carried out:

$$[7.199]\ E\{z(k)\} = \int z_k p(z_k)\,dz_k \ \hat{=}\ \eta_z,$$

$$[7.200]\ E\{z^2(k)\} = \int z_k^2 p(z_k)\,dz_k \ \hat{=}\ P_z,$$

$$[7.201]\ E\{[z(k)-E\{z(k)\}]^2\} = \int (z_k-\eta_z)^2\, p(z_k)\,dz_k \ \hat{=}\ \sigma_z^2 \ ,$$

$$[7.202]\ E\{z(k+\nu)\,z(k)\} = \iint z_{k+\nu} z_k p_2(z_{k+\nu},z_k;\nu)\,dz_{k+\nu}\,dz_k \ \hat{=}\ R_{zz}(\nu) \ ,$$

$$[7.203]\ E\{[z(k+\nu)-\eta_z][z(k)-\eta_z]\} = \iint (z_{k+\nu}-\eta_z)(z_k-\eta_z) p_2(z_{k+\nu},z_k;\nu)\,dz_{k+\nu}\,dz_k \ \hat{=}\ K_{zz}(\nu),$$

where both the quantities $R_{zz}(\nu)$ and $K_{zz}(\nu)$ are even functions and the following properties are valid:

$$[7.204]\quad R_{zz}(\nu) = K_{zz}(\nu)+\sigma_z^2, \qquad |R_{zz}(\nu)| \le R_{zz}(0) = P_z \ , \quad |K_{zz}(\nu)| \le K_{zz}(0) = \sigma_z^2 \ ,$$

$$[7.205]\ R_{zz}(0) = P_z = \sigma_z^2 + \eta_z^2 = K_{zz}(0) + \eta_z^2 \ .$$

As in the case of the continuous process, the statistical mean value and the autocorrelation function are sufficient to characterize a discrete stationary process.
Considering a pair of real discrete processes, Z(n) and V(n), both jointly stationary, in addition to the synthetic quantities above introduced for each process, and hereinafter labeled by adding the single subscript z or v, the ***cross correlation function*** and the mutual ***covariance function*** of the pair of processes are defined:

$$[7.206]\ E\{z(k+\nu)v(k)\} \ \hat{=}\ R_{zv}(\nu) \ ,$$

$$[7.207]\ E\{[z(k+\nu)-\eta_z][v(k)-\eta_v]\} \ \hat{=}\ K_{zv}(\nu) \ ,$$

where η_z and η_v are the statistical mean values; exchanging the subscripts between each other in the [7.206] and [7.207], the following relations can be demonstrated:

$$[7.208]\ R_{vz}(\nu) = R_{zv}(-\nu) \ , \quad K_{vz}(\nu) = K_{zv}(-\nu) \ .$$

Between the two functions the following relation exists:

$$[7.209]\ R_{zv}(\nu) = K_{zv}(\nu) + \eta_z\eta_v;$$

moreover, the following properties are valid:

$$[7.210]\ R_{zv}^2(\nu) \le R_{zz}(0)R_{vv}(0) = P_z P_v\ ,$$

$$[7.211]\ K_{zv}^2(\nu) \le K_{zz}(0)K_{vv}(0) = \sigma_z^2\ \sigma_v^2\ ,$$

having indicated with P_z and P_v and with σ_z^2 and σ_v^2 respectively the powers and the variances of the two real processes.
Two jointly stationary discrete processes are named ***incoherent*** (***orthogonal***) if the cross correlation function is zero for any ν, i.e the following equality is respected:

$$[7.212]\ R_{zv}(\nu) \equiv 0.$$

Instead, if all the values of the covariance function are equal to zero:

$$[7.213]\ K_{zv}(\nu) = R_{zv}(\nu) - \eta_z\eta_v \equiv 0\ ,$$

the two processes are named ***uncorrelated***. Two stationary discrete processes statistically independent are always uncorrelated; if at least one has mean value equal to zero they are also incoherent (orthogonal).

7.3.17.2 Stationary real discrete processes with discrete values

All the presented expressions are applicable also to a ***discrete process with discrete values***, indicated with $Z_q(n)$, which corresponds to a sequence of infinite discrete r.v., Z_{qk}, able to assume one of the identified values of a discrete set, $\{z_q\}$, of finite cardinality M. Since, still in the stationary case and with limitation just at the second order, the discrete r.v. are characterized by the ***probabilities***:

$$[7.214]\ P(z_{q1}; k_1) = P(z_{q2}; k_2) = P(z_q) \hat{=} \pi_q, \quad q = 1, 2, .., M,$$

independent on the integer assay variable, and by the ***joint probabilities***:

$$[7.215]\ P_2(z_{i1}, z_{j2}; k_1, k_2) = P_2(x_i, x_j; \nu) \hat{=} \pi_{ij}(\nu) \quad , \quad i, j = 1, 2, .., M\ ,$$

which depend on the integer difference $\nu = k_1 - k_2$, the operator statistical mean must assume the specific forms:

$$[7.216]\ E\{F_1\} \hat{=} \sum_{q=1}^{M} F_1 \pi_q,\ E\{F_2\} \hat{=} \sum_{i=1}^{M}\sum_{j=1}^{M} F_2 \pi_{ij}(\nu),$$

where F_1 is a function of the discrete r.v. Z_{qk} and F_2 is in general a function of both the discrete r.v. Z_{i1} and Z_{j2}.

7.4 CYCLOSTATIONARY STOCHASTIC PROCESSES

7.4.1 First and second order cyclostationary processes

In many practical cases in a continuous process, S(t), an additive periodic component can be present(1), and therefore of deterministic type, as for example in the case the following realization occurs:

$$[7.217]\ s(t) = x(t) + A_c\cos(\omega_c t),$$

where x(t) is the realization of a real stationary process, X(t), while A_c and $\omega_c = 2\pi f_c$ are known constants. In other cases, a hidden periodicity can occur, not appearing as an addend, as for example in the case of the following realization:

$$[7.218]\ s(t) = x(t)\cos(\omega_c t),$$

(1) In some processes deterministic components with multiple periodicity, with periods not commensurable among one another, can be present.

but that appears when non linear transformations are applied on the process.
If from the calculation of the autocorrelation function of a process an expression periodical in t is achieved, i.e. $R(t,\tau)=R(t-kT_0,\tau)$ is obtained, the name ***cyclostationary stochastic process*** is used. Furthermore, it is possible to distinguish between cyclostationary process of the first order, if the mentioned periodicity occurs also in the statistical mean value, of the second order, if the mentioned periodicity doesn't occur in the statistical mean value. For example, applying the [7.149] and [7.153] to the [7.218], the following expressions are obtained:

$$[7.219]\ \eta_s(t) = E\{x(t)\cos(\omega_c t)\}=E\{x(t)\}\cos(\omega_c t) = \eta_x\cos(\omega_c t),$$

$$[7.220]\ R_{ss}(t,\tau)=E\{x(t+\tau)\cos[\omega_c(t+\tau)]x(t)\cos(\omega_c t)\}= E\{x(t+\tau)x(t)\}\cos[\omega_c(t+\tau)]\cos(\omega_c t)=\frac{1}{2}R_{xx}(\tau)\cos(\omega_c\tau)+\frac{1}{2}R_{xx}(\tau)\cos[\omega_c(2t+\tau)],$$

which identify the process as cyclostationary of the first or of the second order, respectively for η_x different from zero or equal to zero.
In the mentioned cases of cyclostationarity, usually the analysis can be performed as in the cases of the stationary processes, adopting the expedient to consider to add on the time axis a random shift, z, and to calculate the statistical mean also on such an independent r.v., with additional hypotheses that it has distribution density, $p(z)=1/T_0$, uniform in an interval equal to the period of cyclostationarity. With such a procedure the result is to remove the dependance on t in the considered synthetic quantities (see section 7.4.3.1)(2), which result to be:

$$[7.221]\ \eta_s = E\{\eta_s(t+z)\}= \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} \eta_s(t)\,dt\ ,$$

$$[7.222]\ R_{ss}(\tau) = E\{R_{ss}(t+z,\tau)\}= \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} R_{ss}(t,\tau)\,dt\ ,$$

meaning that they can be calculated as average values in the period of the respective functions periodical in the independent variable t.
Applying the above mentioned procedure, the considerations previously developed on the power spectra and on cross correlation can be applicable to the cyclostationary processes. If they are of the first order, in the power spectrum some discrete components appear, $P_{ci}\delta(f-f_{ci})$, associated to the frequencies f_{ci} of the periodicities; instead, the power spectral density is a usual continuous function for $f\neq 0$, in the case of cyclostationary processes of the second order.

7.4.2 Processes represented by means of the complex envelope

Let's consider a real process, X(t), represented by means of the complex envelope, i.e. with realization in the form:

$$[7.223]\ x(t)= \Re\left\{\underline{x}(t)e^{j\omega_c t}\right\}= x_c(t)\cos(\omega_c t)-x_s(t)\sin(\omega_c t)\ ,$$

where $x_c(t)$ and $x_s(t)$ are the realizations of two real processes in base band, $X_c(t)$ and $X_s(t)$, with power spectral densities $P_c(f)$ and $P_s(f)$ limited within $f_{cM}=f_{sM}<f_c=\omega_c/2\pi$.
The process complex envelope, $\underline{X}(t)$, with realization $\underline{x}(t)=x_c(t)+jx_s(t)$, is stationary if the two component processes are jointly stationary, while it is cyclostationary if at least one of the two is such. After the calculation of the time mean (if any) after having added a random shift [in the case one or both processes, $X_c(t)$ and $X_s(t)$, are cliclostationary], the autocorrelation function $R_{\underline{x}\underline{x}}(\tau)$, of the process $\underline{X}(t)$ results to be (see [7.195]):

(2) Nevertheless, also the correlation between spectral components is removed, introducing the possibility to make errors in the evaluation of the power spectra in some infrequent particular cases.

$$[7.224]\ R_{\underline{x}\underline{x}}(\tau)= R_{cc}(\tau) + R_{ss}(\tau) - j[R_{cs}(\tau)–R_{sc}(\tau)]\ ,$$

with obvious meaning of the symbols in the second part. Then, the rule of the simple sum in power is applied (see [7.197]), achieving the relation:

$$[7.225]\ P_{\underline{x}}= P_c+P_s,$$

where with $P_{\underline{x}}$, P_c and P_s the respective powers of the processes $\underline{X}(t)$, $X_c(t)$ and $X_s(t)$ are indicated. Applying the Fourier transform to the [7.224] the power spectral density of the complex envelope process is obtained (see [7.196]):

$$[7.226]\ P_{\underline{x}}(f)= P_c(f)+P_s(f)+2\Im\{P_{cs}(f)\}.$$

Back to the real process X(t), starting from the [7.223] and applying the [7.190] its autocorrelation function can be calculated:

$$[7.227]\ R_{xx}(t,\tau)=\frac{1}{2}[R_{cc}(\tau)+R_{ss}(\tau)]\cos(\omega_c\tau)+\frac{1}{2}[R_{cs}(\tau)-R_{sc}(\tau)]\sin(\omega_c\tau)+\frac{1}{2}[R_{cc}(\tau)-R_{ss}(\tau)]\cos[\omega_c(2t+\tau)]-\frac{1}{2}[R_{cs}(\tau)+R_{sc}(\tau)]\sin[\omega_c(2t+\tau)]\ .$$

The presence in the last two addends in the second member of the expression [7.227] highlights that in general the considered process is cyclostationary, also in the case both processes in base band $X_c(t)$ and $X_s(t)$ are stationary. The periodical dependence on the time t of the autocorrelation function can be removed, as already mentioned, by means of integration over a period (see [7.222]): then, the $R_{xx}(\tau)$ results to be composed only of the first two addends of the [7.227]:

$$[7.228]\ R_{xx}(\tau) = \frac{1}{2}[R_{cc}(\tau)+R_{ss}(\tau)]\cos(\omega_c\tau) + \frac{1}{2}[R_{cs}(\tau)-R_{sc}(\tau)]\sin(\omega_c\tau).$$

Thus, for $\tau=0$, the relation between the powers is:

$$[7.229]\ P_x = \frac{1}{2}(P_c+P_s) = \frac{1}{2}P_{\underline{x}}.$$

Transforming the [7.228] and taking into account the [7.181], the following expression of the power spectral density of the process of the process X(t) is achieved:

$$[7.230]\ P_x(f) = \frac{1}{4}[P_c(f-f_c)+P_s(f-f_c)+P_c(f+f_c)+P_s(f+f_c)]-\frac{1}{2}\Im\{\underline{P}_{cs}(f-f_c)-\underline{P}_{cs}(f+f_c)\};$$

reminding the [7.226], then, the relation, shown in Figura 7.2 is obtained:

$$[7.231]\ P_x(f) = \frac{1}{4}[P_{\underline{x}}(f-f_c)+P_{\underline{x}}(-f-f_c)].$$

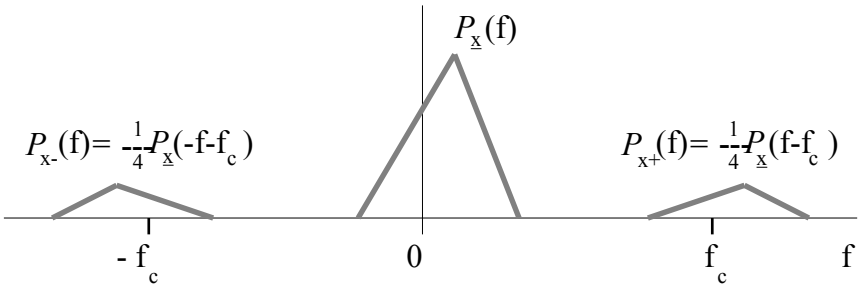


Figura 7.2 – Relation between the power spectral density of a cyclostationary real process not in base band and the one of its complex envelope, $P_{\underline{x}}(f)$.

It is interesting to note that the process X(t), usually cyclostationary, becomes stationary if, and only if, both the base band processes are stationary and moreover the following relations are satisfied:

$$[7.232]\ R_{cc}(\tau) = R_{ss}(\tau)\ ,$$

$$[7.233]\ R_{cs}(\tau) = -R_{sc}(\tau)\ .$$

Putting τ=0 in the [7.233] and [7.175], $R_{cs}(0) = 0$ is obtained; as a consequence in the considered case the r.v. $X_c(t)$ and $X_s(t)$, at the same generic instant t, must be orthogonal (see [7.186]), implying that the total power of the process is the sum of the powers of the base band processes.

7.4.3 Stationary process not in base band

Let's consider a real stationary process X(t), with realization x(t) which is not of base band type. Then, let's refer to a process $\hat{X}(t)$, having as realization the Hilbert transform $\hat{x}(t)$, and to the processes $X_+(t)$ and $X_-(t)$, with respective realizations $x_+(t)=[x(t)+j\hat{x}(t)]/2$ and $x_-(t)=[x(t)-j\hat{x}(t)]/2$; all the processes result to have statistical mean value equal to zero, as well as the process X(t). As by definition the spectra of all the truncated realizations $x_{+T}(t)$ and $x_{-T}(t)$ are different from zero only on the frequency half axes respectively positive and negative, the power spectral density of the processes $X_+(t)$ and $X_-(t)$,

$$[7.234]\ P_{x+}(f) = F\{R_{x+x+}(\tau)\}\ ,\quad P_{x-}(f) = F\{R_{x-x-}(\tau)\}\ ,$$

enjoy the same property and, moreover, the power mutual spectral density results to be equal to zero (see [7.179]), i.e. $P_{x+x-}(f)\equiv 0$ is obtained, and the processes $X_+(t)$ and $X_-(t)$ are incoherent. Then, from the $x(t) = x_+(t)+x_-(t)$ and $\hat{x}(t) = -j[x_+(t)-x_-(t)]$ the following relation is obtained:

$$[7.235]\ P_x(f) = P_{\hat{x}}(f) = P_{x+}(f)+P_{x-}(f),$$

from which applying the inverse transform the similar relations between the autocorrelation functions of the two processes are derived:

$$[7.236]\ R_{xx}(\tau) = R_{\hat{x}\hat{x}}(\tau) = R_{x+x+}(\tau)+R_{x-X-}(\tau)\ .$$

Moreover, for the cross correlation functions and their transforms the following relations are carried out:

$$[7.237]\ R_{x\hat{x}}(\tau) = -R_{\hat{x}x}(\tau) = j[R_{x+x+}(\tau)-R_{x-x-}(\tau)]\ ,$$

$$[7.238]\ P_{x\hat{x}}(f) = P_{\hat{x}x}(f) = j[P_{x+}(f)-P_{x-}(f)]\ .$$

From the [7.236] and [7.237] it is possible to note that $R_{\hat{x}x}(\tau)$ is the Hilbert transform of the $R_{xx}(\tau)$.

Indicating as usual with P_x the power of the process X(t), still for the orthogonality (incoherence) between the processes $X_+(t)$ and $X_-(t)$ the following relation is obtained:

$$[7.239]\ R_{x+x+}(0) = R_{x-X-}(0) = \frac{1}{2}P_x\ ,$$

from which the following equalities are derived:

$$[7.240]\ R_{x\hat{x}}(0) = R_{\hat{x}x}(0) = 0;$$

then, the processes X(t) and $\hat{X}(t)$ result to be additive in power:

$$[7.241]\ E\{x(t)\,\hat{x}(t)\} = \int \Re\{P_{x\hat{x}}(f)\}df = 0\ .$$

Fixing an arbitrary frequency f_c, as long as within the bandwidth of the process X(t), let's then consider the process in base band $\underline{X}(t)$, with realization $\underline{x}(t)=x_c(t)+jx_s(t)$ corresponding to the complex envelope of x(t). The components in phase and in quadrature are real, of the type in base band and related to the realizations of the processes X(t) and $\hat{X}(t)$ by the relations

$$x_c(t)=x(t)\cos(\omega_c t)+\hat{x}(t)\sin(\omega_c t) \text{ and } x_s(t)=-x(t)\sin(\omega_c t)+\hat{x}(t)\cos(\omega_c t).$$

Calculating from these last relations the autocorrelation functions the following relations are obtained:

$$[7.242]\ R_{ii}(t,\tau) = \frac{1}{2}[R_{xx}(\tau)+ R_{\hat{x}\hat{x}}(\tau)]\cos(\omega_c\tau) - \frac{1}{2}[R_{x\hat{x}}(\tau) - R_{\hat{x}x}(\tau)]\sin(\omega_c\tau) \pm \frac{1}{2}[R_{xx}(\tau)- R_{\hat{x}\hat{x}}(\tau)]\cos[\omega_c(2t+\tau)]\pm\frac{1}{2}[R_{x\hat{x}}(\tau)+R_{\hat{x}x}(\tau)]\sin[\omega_c(2t+\tau)]\ ,$$

where the choice of the upper or of the lower sign is made respectively for i=c or i=s. Thanks to the [7.236] and [7.237], which make the periodical addends in t equal to zero, finally the following relations are obtained:

$$[7.243]\ R_{cc}(\tau) = R_{ss}(\tau) = R_{xx}(\tau)\cos(\omega_c\tau)+ R_{\hat{x}x}(\tau)\sin(\omega_c\tau).$$

As already highlighted in the previuous section, being the process X(t) stationary in broad sense, also the processes in base band $X_c(t)$ and $X_s(t)$ result to be stationary in broad sense, with identical autocorrelation functions (see also [7.232]) and with statistical mean value equal to zero.
Putting $\tau = 0$ in the [7.243] and reminding the [7.240] the result is:

$$[7.244]\ P_c = P_s = P_x\ .$$

It is possible to note that the power is not divided on the two processes having as realizations the real part and the imaginery part of the complex envelope, but it appears identical on each of them.
By means of the [7.236] and [7.237], the [7.243] can assume the expression:

$$[7.245]\ R_{cc}(\tau) = R_{ss}(\tau) = R_{x+x+}(\tau)e^{-j\omega_c\tau}+R_{x-x-}(\tau)e^{j\omega_c\tau}\ .$$

Then, applying the Fourier transform the power spectral densities of the base band processes are achieved:

$$[7.246]\ P_c(f) = P_s(f) = P_{x+}(f+f_c)+P_{x-}(f-f_c)\ .$$

The cross correlation functions, $R_{cs}(\tau)$ and $R_{sc}(\tau)$, of the processes $X_c(t)$ and $X_s(t)$ can be carried out from the already mentioned expressions of the realizations as a function of x(t) and $\hat{x}(t)$, obtaining the relations:

$$[7.247]\ R_{cs}(\tau)=-R_{sc}(\tau)=R_{xx}(\tau)\sin(\omega_c\tau)+ R_{x\hat{x}}(\tau)\cos(\omega_c\tau) = j[R_{x+x+}(\tau)e^{-j\omega_c\tau}-R_{x-x-}(\tau)e^{j\omega_c\tau}].$$

Therefore, the [7.233] results to be confirmed, as well as the:

$$[7.248]\ R_{cs}(0) = R_{sc}(0) = E\{x_c(t)x_s(t)\}= \int \Re\{P_{cs}(f)\}df=0\ ,$$

which highlights that the component processes in phase and in quadrature can be added from power point of view.
Then, the respective power mutual spectral densities, $P_{cs}(f)$ and $P_{sc}(f)$ are obtained:

$$[7.249]\ P_{cs}(f) = -P_{sc}(f) = j[P_{x+}(f+f_c) - P_{x-}(f-f_c)]\ ,$$

which, being $P_x(f)$ real, are purely imaginery. The processes $X_c(t)$ and $X_s(t)$ in general are not incoherent (orthogonal); on the basis of the property $P_{x+}(f)=P_{x-}(-f)$, then, it is possible to demonstrate that the occurrence of the particular relation of symmetry with respect to f_c:

$$[7.250]\ P_{x+}(f+f_c) = P_{x+}(f_c-f)\ ,$$

is sufficient condition in order that the processes $X_c(t)$ and $X_s(t)$ are incoherent (orthogonal).
Thanks to the [7.245] and [7.247], the autocorrelation function (see [7.224]) of the process complex envelope, which results to be stationary in broad sense and with statistical mean value equal to zero, assumes the particular expression:

$$[7.251]\ R_{\underline{x}\underline{x}}(\tau)= 2[R_{cc}(\tau)-jR_{cs}(\tau)] = 4R_{x+x+}(\tau)e^{-j\omega_c\tau}\ ;$$

thus, the power spectral density is obtained:

$$[7.252]\ P_{\underline{x}}(f)= 2[P_c(f) - jP_{cs}(f)] = 4P_{x+}(f+f_c)\ .$$

In the second term of the previous equation it is possible to note that, being $P_{cs}(f)$ purely imaginery, $P_{\underline{x}}(f)$ is anyway expressed as the sum of two real functions, according to the definition itself of power spectral density, always real and non negative. Then, in the [7.252] the already highlighted relation [7.231] between the spectra of the processes $\underline{X}(t)$ and $X(t)$ can be recognized. Integrating the [7.252] and reminding the [7.244] finally the relations between the powers are obtained:

$$[7.253]\ P_x = P_c = P_s = \frac{1}{2} P_{\underline{x}} .$$

7.4.3.1 Appendix 1. Synthetic parameters in the cyclostationary processes

In the cyclostationary processes of the second order the autocorrelation function results to be periodic, meaning that:

$$[7.254]\ R(t-T_0, \tau) = R(t, \tau) .$$

If a random shift, z, is introduced on the origin of the time axis and this variable is assumed to be statistically independent with uniform distribution of the probability density over one period, which therefore results to be:

$$[7.255]\ p(z) = \frac{1}{T_0} ,$$

the statistical mean value of the [7.254] with respect to z provides:

$$[7.256]\ E\{R(t+z, \tau)\} = \int_{-T_0/2}^{T_0/2} R(t+z,\tau)p(z)\,dz = \frac{1}{T_0} \int_{t-T_0/2}^{t+T_0/2} R(x,\tau)\,dx ,$$

where the change of the variable x=t+z is introduced.
Being the integrand function periodic (see [7.254]) any shift of the integration extremes is allowed in the last member of the [7.256]; in particular, it is possible to put t=0, then obtaining:

$$[7.257]\ E\{R(t+z, \tau)\} = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} R(t,\tau)\,dt = R(\tau).$$

The procedure leads to remove the dependence on time in the function $R(t, \tau)$ without any other hypothesis in addition to the periodicity of this latter. Thus, it is applicable also to the statistical mean value of a cyclostationary process of the first order, for wchich:

$$[7.258]\ \eta(t-T_0) = \eta(t) ,$$

is obtained and as a consequence the result is:

$$[7.259]\ E\{\eta(t+z)\} = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} \eta(t)\,dt = \eta.$$

7.5 PROCESSES REPRESENTED BY MEANS OF TIME SERIES

7.5.1 Real processes with random factors

Often the processes are represented by means of an infinite time series of known energy signals, each of which including a continuous r.v. The most common example is offered by the type of a real process with random factors, X(t), with realization which appears in the expression:

$$[7.260]\ x(t) = \sum_k z_k p(t-kT) ,$$

with k integer variable and T constant, and where:

⇒ $z_k=z(k)$ are the determinations of the continuous r.v., Z_k, which with their sequence, Z(n), consitute a stationary real discrete process;

⇒ p(t) is a real function, practically time limited with practical duration T_p even much greater than T, having energy spectrum $\mathcal{E}_{pp}(f)$ strictly limited in bandwidth and in energy:

$$[7.261]\ \mathcal{E}_{pp}=\int p^2(t)dt=\int E_{pp}(f)df=C_{pp}(0)\ ,$$

where $C_{pp}(\tau)$ is the time autocorrelation function of p(t), having the energy spectrum $\mathcal{E}_{pp}(f)$ as Fourier transform.

First of all, the stationary discrete real process Z(n) results to be characterised (see section 7.3.17) by means of the statistical mean value, $E\{z(k)\}=\eta_z$, and of its autocorrelation function, $E\{z(k+\nu)z(k)\}=R_{zz}(\nu)$. Then, the autocovariance function, $K_{zz}(\nu)=R_{zz}(\nu)-\eta_z^2$ and the power, $P_z=R_{zz}(0)=\sigma_z^2+\eta_z^2$, where σ_z^2 is the variance of the discrete process, are obtained.

Putting in the [7.260] $z_k=z_{ek}+\eta_z$ (Z_{ek} indicates the correspondent ***fair*** r.v., i.e. with the same stochastic characterization but having the mean value equal to zero), immediately the continuous process X(t) results to include the periodical addend, with period T:

$$[7.262]\ x_0(t)=\eta_z rep_T\{p(t)\}=\eta_z\sum_k p(t-kT)\ ,$$

and it is in general cyclostationary, of the first order if $\eta_z\neq 0$.

As shown in section 7.5.4.1, starting from the [7.260] and applying the [7.153] and [7.222] the statistical mean value and the autocorrelation function (not depending on time) of the cyclostationary continuos process X(t) are determined:

$$[7.263]\ \eta_x=\frac{1}{T}\eta_z\int p(t)dt\ ,$$

$$[7.264]\ R_{xx}(\tau)=\frac{1}{T}\sum_\nu R_{zz}(\nu)C_{pp}(\tau-\nu T)=\frac{1}{T}\eta_z^2\, rep_T\{C_{pp}(\tau)\}+\frac{1}{T}\sum_\nu K_{zz}(\nu)C_{pp}(\tau-\nu T).$$

Concerning the mean value of the continuous process it is possible to observe that it becomes equal to zero if the mean value of the discrete process is zero, that is if $\eta_z=0$ occurs; moreover, it is possible to obtain a zero mean value also for $\eta_z\neq 0$, with opportune choice of the waveform p(t) such that its integral is equal to zero.

Applying the Wiener-Khintchine relation to the [7.264] (see [7.168]), the following expression of the power spectral density of the continuous process is obtained:

$$[7.265]\ P_x(f)=\frac{1}{T}\sigma_z^2\sum_k c_k\delta(f-kf_0)+\frac{1}{T}\mathcal{E}_{pp}(f)\sum_\nu K_{zz}(\nu)e^{-j2\pi\nu fT}=\frac{1}{T^2}\sigma_z^2\sum_k\mathcal{E}_{pp}(kf_0)\delta(f-kf_0)+\frac{1}{T}\mathcal{E}_{pp}(f)\left[\sigma_z^2+2\sum_{\nu=1}^{\infty}K_{zz}(\nu)\cos(2\pi\nu fT)\right].$$

It is possible to note that in the spectrum of the process there are:

⇒ a series of discrete components at frequencies multiple of $f_0=1/T$, associated to the presence of the periodical addend $x_0(t)$ in the process (see [7.262]), with magnitude proportional to the square of the statistical mean value of the discrete process Z(n);

⇒ a continuous part, with evolution which depends both on the evolution of the energy spectrum of the deterministic waveform p(t), and on the function between squared brackets in the [7.265], periodic with period $f_0=1/T$; such a function results to be the sum of a constant, equal to the variance of the discrete process Z(n) and of a variable related to the autocovariance of the same process for the values $\nu\neq 0$.

Thanks to the stationarity of the discrete process, the statistical average energy of the generic addend of the series [7.260], separately considered, is invariant with respect to k, implying that it is a characteristic of the continuous process:

$$[7.266]\ E_{xT} \triangleq E\{\mathcal{E}_{kk}\} = E\left\{\int z_k^2 p^2(t-kT)dt\right\} = P_z \mathcal{E}_{pp}.$$

Hereinafter, two cases will be considered, assuming the hypotesis that the r.v. Z_k are orthogonal, or that the addend energy signals, which constitute the periodical form $rep_T\{p(t)\}$, are orthogonal; for both it is possible to demonstrate that the power of the process X(t) can be expressed by means of the simple relation:

$$[7.267]\ P_x = \frac{E_{xT}}{T},$$

implying that the power can be calculated immediately as the ratio between the common value, E_{xT}, of the statistical mean energy of the elements of the series and the value, T, of its time rate.

7.5.1.1 Case of orthogonality of the r.v. of the discrete process

If the r.v. Z_k of the discrete process are all orthogonal among one another, meaning that the autocorrelation function of the process $R_{zz}(\nu)$ is equal to zero for every $\nu \neq 0$, while for $\nu=0$ $R_{zz}(0)=P_z=\sigma_z^2$ ($\eta_z=0$) is obtained, the [7.264] becomes:

$$[7.268]\ R_{xx}(\tau)=\frac{1}{T}R_{zz}(0)C_{pp}(\tau) = \frac{1}{T}P_z C_{pp}(\tau);$$

for $\tau=0$ and taking into account the [7.261] and [7.266], the power of the process X(t) assumes the expression:

$$[7.269]\ P_x = R_{xx}(0)=\frac{1}{T}P_z C_{pp}(0)=\frac{1}{T}P_z \mathcal{E}_{pp}= \frac{E_{xT}}{T} \quad , \quad \text{for } R_{zz}(\nu) = 0 \text{ con } \nu \neq 0 .$$

Moreover, applying the Fourier transform to the [7.268] the result is:

$$[7.270]\ P_x(f) = \frac{1}{T} P_z \mathcal{E}_{pp}(f),$$

implying that the evolution of the power spectrum of the process has no discrete components and is proportional to that one of the energy spectrum of the known function p(t).

7.5.1.2 Case of orthogonality of the waveforms shifted in time

If all the deterministic functions p(t-kT), shifted in time, are orthogonal among one another, meaning that:

$$[7.271]\ \int p(t-kT)p(t-hT)dt = C_{pp}[(h-k)T] = C_{pp}(\nu T) = \begin{cases} E_{pp}, & \nu=0 \\ 0, & \nu \neq 0 \end{cases},$$

the [7.264] for $\tau=0$ provides, regardless the statistical dependence of the r.v. Z_k:

$$[7.272]\ P_x = R_{xx}(0) = \frac{1}{T} R_{zz}(0)\, \mathcal{E}_{pp} = \frac{1}{T} P_z \mathcal{E}_{pp} = \frac{E_{xT}}{T} \quad , \quad \text{for } C_{pp}(\nu T) = 0 \text{ con } \nu \neq 0 .$$

It is important to point out that, thanks to the mentioned orthogonality of the shifted waveforms, the power of the process is not affected by the presence of the cross correlation among the r.v. Z_k of the discrete process. Instead, the cross correlation affects the evolution of the power spectral density, for which the general expression [7.265] is valid.

7.5.2 Sampled processes in base band

What presented in the previous section can be applied to the stationary continuous real processes with realization represented by means of samples. Considering a process in base band X(t) with maximum frequency f_M and choosing a generic sampling period $T_{ca} \leq 1/2f_M$, the generic realization can be represented by means of a time series:

$$[7.273]\ x(t) = \sum_k c_k \operatorname{sinc}\left(\frac{t}{T_{ca}} - k\right),$$

with k integer variable, and where:

- the samples $c_k=x(kT_{ca})$ are the determinations of the continuous r.v. C_k which with their sequence constitute C(n), a stationary discrete real process;
- $\operatorname{sinc}(t/T_{ca})$ is the known sampling function, with energy and energy spectrum given respectively by the:

$$[7.274]\ \mathcal{E}_{ss} = T_{ca}\ ,$$

$$[7.275]\ \mathcal{E}_{ss}(f) = T_{ca}^2 \operatorname{rect}(fT_{ca})\ ,$$

and that respects the property of orthogonality of the shifted waveforms:

$$[7.276]\ C_{ss}(\nu T_{ca}) = 0\ ,\quad \text{for } \nu \neq 0\ .$$

Taking into account that the integral of the function $\operatorname{sinc}(t/T_{ca})$ is equal to T_{ca}, on the basis of the [7.263] immediately the result is:

$$[7.277]\ \eta_x = \eta_c,$$

meaning that the statistical mean value, η_x, of the continuous process corresponds to that one, η_c, of the discrete process constituted by the sequence of the random samples.
Utilizing the results of the previous section and noting that the [7.275] limits the spectrum in the frequency slot $(-1/2T_{ca};1/2T_{ca})$, the power spectral density of the continuous process assumes the expression:

$$[7.278]\ P_x(f) = \eta_c^2\ \delta(f) + T_{ca}\operatorname{rect}(fT_{ca})\left[\sigma_c^2 + 2\sum_{\nu=1}^{\infty} K_{cc}(\nu)\cos(2\pi\nu f T_{ca})\right],$$

where σ_c^2 and $K_{cc}(\nu)$ are respectively the variance and the autocovariance of the discrete process C(n), constituted by the sequence of the r.v. having as determinations the random samples taken at interval T_{ca}. Thanks to the [7.276], the power of the continuous process can be simply obtained dividing by T_{ca} the statistical mean energy of the generic element of the time series with which the process itself is expressed; then, from the [7.266] and from the [7.274] the following relation is obtained:

$$[7.279]\ P_x = \frac{1}{T_{ca}} P_c \mathcal{E}_{ss} = P_c = \sigma_c^2 + \eta_c^2$$

being P_c the power of the discrete process.
Taking into account that the representation by means of samples keeps unaltered the waveform of the signal varying the sampling interval, as long as it respects the disequality $T_c \leq 1/2f_M$, the following remarks can be obtained, based on the invariance of the synthetic characteristics of the continuous process with realization [7.273], in particular of η_x, P_x and of $P_x(f)$ for which the [7.277], [7.278] and [7.279] are achieved.
On the basis of the [7.277] and [7.279], the sequence of the random samples keeps unaltered the quantities of the first order, i.e. η_c, σ_c^2 and P_c, varying the sampling period T_{ca}; instead, the statistical mean energy, $E_{xTca}=P_xT_{ca}$, of the generic element of the time series results to be different. Taking into account the [7.278], also the autocovariance function, $K_{cc}(\nu)$, results to be different for every $\nu \neq 0$. In particular, being by hypothesis $P_x(f)=0$ for $f>f_M$, the different random samples are found out to be uncorrelated (i.e. $K_{cc}(\nu)=0$ for $\nu \neq 0$) if, and only if, the continuous process has a constant spectral density from $f_m=0_+$ to $f_M=B$:

$$[7.280]\ P_x(f) = \frac{P}{2f_M}\,\mathrm{rect}\left(\frac{f}{2f_M}\right)$$

and moreover the sampling period assumes the maximum value: $T_{ca}=1/2f_M=1/2B$, that is equal to the Nyquist interval. In such particular case the continuous process is named a ***base band white process***.

7.5.3 Complex processes with random factors

Let's consider a cyclostationary real continuous process not in base band and the relative representative complex envelope process $\underline{X}(t)$, with realization (see [7.223]) which is supposed to be expressed in terms of the time series:

$$[7.281]\ \underline{x}(t)= x_c(t)+jx_s(t) = \sum_k(a_k+jb_k)p(t - kT),$$

where as usual k is an integer variable and:

- ⇨ a_k and b_k are the determinations of the continuous r.v., A_k and B_k, which constitute with their sequences, A(n) and B(n), a pair of stationary discrete real processes;
- ⇨ p(t) is a real function, practically limited with practical duration T_p even much greater than T, having energy spectrum $\mathcal{E}_{pp}(f)$ strictly limited in base band.

Starting from the [7.226] and with the same symbols and procedures used to achieve the [7.265] the autocorrelation function and the power spectral density of the continuous process $\underline{X}(t)$ are obtained:

$$[7.282]\ R_{\underline{x}\underline{x}}(\tau) = \frac{1}{T}\sum_\nu\{[R_{aa}(\nu)+R_{bb}(\nu)]C_{pp}(\tau-\nu T)-jR_{ab}(\nu)[C_{pp}(\tau-\nu T)-C_{pp}(\tau+\nu T)]\}$$

$$[7.283]\ P_{\underline{x}}(f)=\frac{1}{T^2}(\eta_a{}^2+\eta_b{}^2)\sum_k\mathcal{E}_{pp}(kf_0)\delta(f-nf_0)+\frac{1}{T}\mathcal{E}_{pp}(f)(\sigma_a{}^2+\sigma_b{}^2)+\frac{2}{T}\mathcal{E}_{pp}(f)\{\sum_{\nu=1}^{\infty}[K_{aa}(\nu)+K_{bb}(\nu)]\cos(2\pi fT)-\sum_\nu K_{ab}(\nu)\sin(2\pi\nu fT)\},$$

in which also the covariance, $K_{ab}(\nu)$, of the discrete processes appears. Instead, to evaluate the power $P_{\underline{x}}$ of the complex envelope process the statistical dependence between the two discrete processes has no impact, as it is possible to note from the [7.282] for $\tau=0$ taking into account that the autocorrelation, $C_{pp}(u)$, is an even function of the argument u; this is consistent with the relation of simple sum (see [7.229]) between the power $P_{\underline{x}}$ and those, P_c and P_s, of the two real processes $X_c(t)$ and $X_s(t)$.

Reminding the [7.266], the statistical mean energy of the generic element of the time series [7.281] assumes the expression:

$$[7.284]\ E_{\underline{x}T}=E\left\{\int\left(a_k^2+b_k^2\right)p^2(t-kT)\,dt\right\} = (P_a+P_b)\mathcal{E}_{pp} = (\sigma_a{}^2+\sigma_b{}^2+\eta_a{}^2+\eta_b{}^2)\,\mathcal{E}_{pp}.$$

Both in the case the r.v. A_k and B_k are orthogonal ($R_{aa}(\nu)=R_{bb}(\nu)=0$ for every $\nu\neq0$), and in the case the shifted waveforms are orthogonal (see [7.271]) although the r.v. are correlated, thus, the power of the considered process is obtained:

$$[7.285]\ P_{\underline{x}} = \frac{E_{\underline{x}T}}{T} = \frac{1}{T}(P_a+P_b)\mathcal{E}_{pp}\ .$$

The developed analysis can be utilized in the case of a continuous process, in general cyclostationary, strictly limited in a bandwidth B and with minimum frequency $f_m\neq0$. Adopting for the realization the representation with double sequence of samples and assuming $A_k=C_{ck}$, $B_k=C_{sk}$, $T=1/B$ and $p(t)=\mathrm{sinc}(t/T)$, and thus also $\mathcal{E}_{pp}=T$ (see the second of the [7.274]), from the [7.285] the power of the representative complex envelope process is obtained:

$$[7.286]\ P_{\underline{x}} = \sigma_{Cc}{}^2+\sigma_{Cs}{}^2+\eta_{Cc}{}^2+\eta_{Cs}{}^2,$$

with obvious meaning of the symbols. Thanks to the [7.229], finally the power of the considered process is obtained, which is simply one half of that provided by the [7.286]. The reader may approach the task to demonstrate that if the process is stationary the statistical mean values of the random samples are equal to zero ($\eta_{Cc}=\eta_{Cs}=0$) and their variances are:

$$[7.287]\ \sigma_{Cc}^2=\sigma_{Cs}^2=(1/2)\ P_{\underline{x}} = P_x\ .$$

in line with the [7.244].

7.5.4 Process sum of real processes with random factors

Let's consider a particular type of continuous process X(t), with realization which can be represented with the expression:

$$[7.288]\ x(t) = \sum_k\left[a_k p_1(t-kT)+b_k p_2(t-kT)\right],$$

with k integer variable and where:

- ⇨ a_k and b_k are the determinations of the continuous r.v., A_k and B_k, which constitute with their sequences, A(n) and B(n), a pair of stationary discrete real processes;
- ⇨ $p_1(t)$ and $p_2(t)$ are real functions, with energies $\mathcal{E}_{11}$ and $\mathcal{E}_{22}$, practically limited in time with practical duration even much greater than T, and that respect the condition:

$$[7.289]\ \int p_i(t-kT)\ p_j(t-hT)dt=C_{ij}[(h-k)T]= C_{ij}(\nu T)=0, \text{ for } \nu=0 \text{ and } i\neq j \text{ and for } \nu\neq 0.$$

With the same symbols and procedure used to achieve the [7.282] the autocorrelation function of the considered constinuous process is obtained:

$$[7.290]\ R_{xx}(\tau)=\frac{1}{T}\sum_\nu\left\{R_{aa}(\nu)C_{11}(\tau-\nu T)+R_{bb}(\nu)C_{22}(\tau-\nu T)\right\}+\frac{1}{T}\sum_\nu R_{ab}(\nu)\left[C_{12}(\tau-\nu T)+C_{21}(\tau+\nu T)\right].$$

Leaving to the reader the task to calculate the power spectral density and to demonstrate that it is impacted by the cross correlation between the discrete processes, here let's just verify that instead the power of the continuous process is independent on $R_{ab}(\nu)$. In fact, the [7.290] provides for $\tau=0$:

$$[7.291]\ P_x= R_{xx}(0) = \frac{1}{T}\left[R_{aa}(0)\,\mathcal{E}_{11}+R_{bb}(0)\mathcal{E}_{22}\right] = \frac{1}{T}(P_a\mathcal{E}_{11}+P_b\mathcal{E}_{22}) = \frac{E_{xT}}{T}\ ,$$

where P_a and P_b are as usual the powers of the discrete processes and E_{xT} is the statistical mean energy of the generic element of the time series, as results from the general expression:

$$[7.292]\ E_{xT} =E\left\{\int\left[a_k p_1(t-kT)+b_k p_2(t-kT)\right]^2 dt\right\} = P_a\mathcal{E}_{11} + P_b\mathcal{E}_{22}+ 2R_{ab}(0)C_{12}(0)\ ,$$

which is reduced to:

$$[7.293]\ E_{xT} = P_a\mathcal{E}_{11} + P_b\mathcal{E}_{22},$$

because the two known waveforms are orthogonal (see [7.289]).
What above presented, and in particular the simple calculation of the power of the continuous process by means of the statistical mean energy of the generic element of the time series, is applicable to the broader case of process with N added realizations, that is:

$$[7.294]\ x(t)=\sum_{i=1}^{N}\sum_k a_{ik}p_i(t-kT),$$

only if the [7.289] is verified. Then, the result is:

$$[7.295]\ P_x= \frac{E_{xT}}{T} = \frac{1}{T}\sum_{i=1}^{N} P_{ai}E_{ii}.$$

7.5.4.1 Appendix 2. Analysis of a real process with random factors

Applying the [7.149] to the cyclostationary process [7.260] and taking into account that only the variables Z_k are random, the statistical mean value, dependent on time, is obtained:

$$[7.296]\ \eta_x(t) = E\{\sum_k z_k p(t-kT)\} = \sum_k E\{z_k\} p(t-kT) = \eta_z \sum_k p(t-kT).$$

Thus, it is exactly the periodic addend, of period T, highlighting in the [7.262], present in every realization. The dependence on time of the [7.296] can be removed applying the [7.221]; changing the variable x = t-kT the result is:

$$[7.297]\ \eta_x = \frac{1}{T}\int_{-T/2}^{T/2} \eta_x(t)\,dt = \frac{1}{T}\,\eta_z \sum_k \int_{-kT-T/2}^{-kT+T/2} p(x)\,dx = \frac{1}{T}\eta_z \int p(t)\,dt\,.$$

Applying the [7.153] and still taking into account that only the variables A_k are random, the autocorrelation function of the considered cyclostationary process is obtained:

$$[7.298]\ R_{xx}(t,\tau) = E\{[\sum_k z_k p(t-kT)][\sum_h z_h p(t-hT+\tau)]\} = \sum_k \sum_h E\{z_k z_h\} p(t-kT) p(t-hT+\tau)$$

Replacing the integer variable ν=h-k and utilizing the [7.202] then the result is:

$$[7.299]\ R_{xx}(t,\tau) = \sum_k \sum_\nu E\{z_k z_{k+\nu}\} p(t-kT) p(t-kT+\tau-\nu T) = \sum_\nu R_{zz}(\nu) \sum_k p(t-kT) p(t-kT+\tau-\nu T),$$

which is periodic, of period T, in the variable t.
The dependence on t of the autocorrelation function can be removed applying the [7.222]; in fact, changing the variable x=t-kT the result is:

$$[7.300]\ R_{xx}(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} R_{xx}(t,\tau)\,dt = \frac{1}{T}\sum R_{zz}(\nu) \sum_k \int_{-kT-T/2}^{-kT+T/2} p(x) p(x+\tau-\nu T)\,dx = \frac{1}{T}\sum R_{zz}(\nu) C_{pp}(\tau-\nu T),$$

where $C_{pp}(\tau)$ is the time autocorrelation function of the waveform p(t).

7.6 GAUSSIAN CONTINUOUS PROCESSES

7.6.1 Gaussian continuous processes: the noise

The ***Gaussian continuous processes*** are very interesting because this cathegory respects the important property to allow the full statistical knowkedge just on the basis of the knowledge of the probability density function of the second order. In particular, in case of stationarity, that as above stated is always in strict sense, the probability density function of the first order of a real Gaussian continuous process has the expression, independent on time:

$$[7.301]\ p_g(x) \hat{=} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\eta)^2}{2\sigma^2}},$$

where η is the statistical mean value, $\sigma^2 = R(0) - \eta^2$ is the variance and $R(\tau)$ is the autocorrelation function of the process; the probability density of the second order is an exponential function too, similar to the [7.301], which results completely determined when $R(\tau)$ is known. In the case, very frequent, of stationary Gaussian continuous process with zero mean (η=0) is thus sufficient just the knowledge of the autocorrelation function, or of the power spectral density (see [7.168]).
The sum of independent Gaussian processes is still a Gaussian process, with the mean value resulting to be the sum of the mean values and variance resulting to be the sum of the variances. Moreover, it is possible to demonstrate (theorem of the central limit) that the sum of a number n of random signals belonging to arbitrary processes, but independent, usually tends to approximate a Gaussian process when n tends to infinite. Then, the usual characterization of a stochastic process as gaussian results to be acceptable: a typical example in the field of telecommunications is that one of the characterization of the set of plenty of undesirable random additive signals, that cannot be associated to the useful signal and due to a large number of independent causes, by means of a real Gaussian continuous process which is named ***Gaussian Noise*** (GN). In the remainder of the section this latter will be dealt with, but remind that what carried out is applicable to any process of the same type, both undesired and useful.

In the telecommunications a stationary process representative of a Gaussian noise and assumed with zero mean, is usually characterized by means of its power spectral density, $N(f)$. Once such a function is identified, the autocorrelation function:

$$[7.302]\ R_{nn}(\tau) = F^{-1}\{N(f)\}\ ,$$

and the power P_n, which, considering that the mean value is equal to zero, results to be equal to the variance σ_n^2 of the process, can be immediately calculated; utilizing the [7.171]:

$$[7.303]\ P_n=\sigma_n^2= \int N(f)\,df$$

is achieved. In general, it is possible to note that larger the frequency interval within which $N(f)$ is different from zero greater the noise power, or variance.
Often, the case in which $N(f)=N_0$ is considered, meaning that the power spectral density is invariant with respect to the frequency, and it is named ***White Gaussian Noise*** (WGN); otherwise, the noise is generically named ***colored***. From the [7.302] with $N(f)$ constant the autocorrelation function of the white gaussian noise results to be:

$$[7.304]\ R_{nn0}(\tau) = N_0\delta(\tau) = \frac{1}{2}N_0\delta(\tau),$$

where with

$$[7.305]\ N_0=2N_0$$

usually the power of the white gaussian noise per unit of band is indicated, that can be calculated from the [7.303] with the integral limited from -1 Hz to 1 Hz.
It is possible to note from the [7.303] that if the spectral density of the noise were exactly constant over the whole frequency axis, the power would be infinite, which is physically absurd; in practice, the attribute of white noise is actually used meaning that the spectral density is uniform only within a limited bandwidth, but sufficiently extended to largely include that one actually occupied by the useful spectrum to which usually the noise spectrum is added. Instead, referring to a useful gaussian continuous process, the attribute white implies that the power spectral density is constant in its limited bandwidth and zero anywhere else (see section 7.5.2).
Referring to the time domain, a generic realization $n(t)$ of the noise is a symmetrical signal in theory unlimited (see [7.301]); nevertheless, in practice it is possible to define a peak value of the noise, n_p, as that value with a very low probalility, P_p, to be overcome by the absolute value of $n(t)$:

$$[7.306]\ P_p \hat{=}\ 2\int_{n_p}^{\infty} p_g(n)\,dn\ .$$

Utilizing the [7.301], the previous expression can be transformed to the following expression:

$$[7.307]\ P_p= \mathrm{erfc}(x) = \mathrm{erfc}\left(\frac{n_p}{\sqrt{2}\sigma_n}\right)\ ,$$

where the ***error complementary function*** was used:

$$[7.308]\ \mathrm{erfc}(x) \hat{=}\ \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-y^2}\,dy\ ,$$

and:

$$[7.309]\ n_p= x\sqrt{2}\,\sigma_n$$

was considered. Then, if for example $P_p=10^{-5}$ is assumed, resulting from the [7.307] $x=3.12$, from the [7.309] $n_p=4{,}46\sigma_n$ is calculated, from which the peak factor at 10^{-5} of the gaussian noise is calculated:

$$[7.310]\ F_{pn}= \frac{n_p}{\sigma_n} = 4{,}46,$$

that means $F_{pn}(dB) = 20\log F_{pn}= 13$ dB.

7.6.2 Stationary gaussian noise not in base band

Let's consider a stationary Gaussian noise, N(t), with power spectral density, $N(f)$, different from zero only for $0<f_m<|f|<f_M$ (often f_m is much greater than zero and the process is of shifted band type). Choosing arbitrarily a frequency f_c within the bandwidth of the process, the representation (see [7.223]):

$$[7.311]\ n(t) = n_c(t)\cos(\omega_c t)-n_s(t)\sin(\omega_c t)$$

can be adopted for the generic realization. Starting from the [7.311], the following expressions, identical, of the autocorrelation functions of the base band processes, $N_c(t)$ and $N_s(t)$, which have as realizations $n_c(t)$ and $n_s(t)$, can be obtained (see [7.243]):

$$[7.312]\ R_{ncc}(\tau) = R_{nss}(\tau) = R_{nn}(\tau)\cos(\omega_c t)+ R_{\hat{n}n}(\tau)\sin(\omega_c t)\ ,$$

where $R_{nn}(\tau)$ e $R_{\hat{n}n}(\tau)$ are respectively the autocorrelation function of the noise and the cross correlation function obtained between the Hilbert transform of the noise and the noise itself. The processes $N_c(t)$ and $N_s(t)$ result to be stationary Gaussian; then, putting $\tau=0$ in the [7.312] the following relations between the powers are obtained:

$$[7.313]\ \sigma^2_{nc} = \sigma^2_{ns} = \sigma^2_n;$$

it is possible to note that, contrary to intuition, the shifted band noise power is not divided on the base band processes, but it is present, with the same amount, on both of them. As usual, the reason is the particular relation between the signal and its representative complex envelope.
The identical power spectral densities of the base band noise processes assume the expression (see [7.246]):

$$[7.314]\ N_c(f)=N_s(f)=N_+(f+f_c)+N_-(f-f_c).$$

It is possible to note that the evolution of such functions depends not only on the evolution of $N(f)$, but also on the value of f_c, as shown in Figura 7.3; in fact, if f_c corresponds to one of the extremes of the bandwidth the noises in base band have the maximum extension, from –B to B (see Figura 7.3.b,c), while the extension is minimum, from -B/2 to B/2, in the case f_c is chosen at the center of the bandwidth (see Figura 7.3.d).
A particular, but very frequent, case is that one in which the Gaussian noise is white, with power spectral density constant in band, N_0, and zero anywhere else. If f_c is chosen at the center of the bandwidth, the [7.314] becomes:

$$[7.315]\ N_c(f) = N_s(f) = 2N_0\,\mathrm{rect}(f/B) = N_0\,\mathrm{rect}(f/B)\ ;$$

the relation [7.313] between the powers of the processes assumes moreover the particular expression:

$$[7.316]\ \sigma^2_{nc} = \sigma^2_{ns} = 2N_0B = N_0B\ .$$

Since the symmetry condition of the half spectrum $N_+(f)$ with respect to f_c (see [7.250]) results to be satisfied, the two noises in base band results to be incoherent (orthogonal) and, being gaussian, statistically independent.

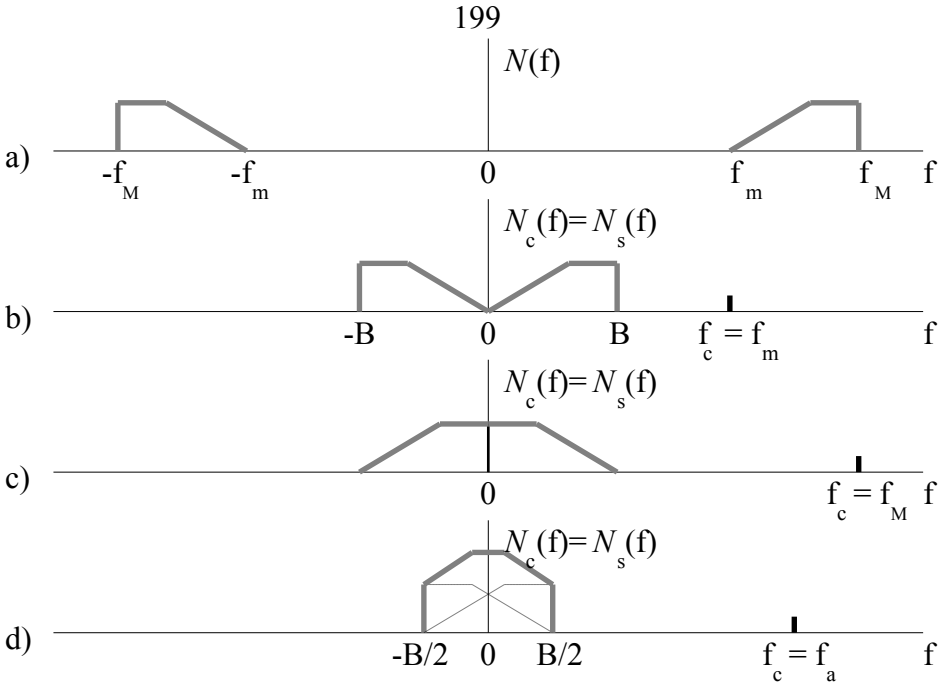


Figura 7.3 – Power spectral density of a gaussian noise not in base band (a) and of the relative base band processes, with f_c at the bounds of the bandwidth (b,c) and at the center (d).

7.6.3 White gaussian noise in the signal space

Let's consider a physical stationary process, X(t), with real realizations, x(t), which can be considered simultaneously limited in duration T and in bandwidth B.
Adopting for the generic realization of the process a representation by means of a base, with good approximation the following development is obtained:

$$[7.317]\ x(t) = \sum_{k=1}^{N} x_k \psi_k(t),$$

by means of which the realizations of the continuous process X(t) are put in correspondence with those ones of the discrete stationary process, X(n), constituted by a countable sequence of N≥2BT r.v. X_k. In the signal space identified by the set of unit vectors $\{\boldsymbol{\psi}_k\}$ associated to the set $\{\psi_k(t)\}$ of N orthonormal functions, each realization x(t) of the process can be thus represented with a vector **x** having as coordinates the particular values $x_1, x_2, .., x_N$ extracted for the variables X_k. Moreover, the choice of the $\psi_k(t)$ to obtain a development like [7.317] with the r.v. uncorrelated (Karhunen-Loeve development) is impossible.
In the case of the realization of a white Gaussian noise, N(t), with zero mean and with power per unit band, N_0, equal to twice the power spectral density, $\mathcal{N}_0$, uniform over the whole frequency axis, just within the time definition interval of the base $\{\psi_k(t)\}$ the following relation is valid:

$$[7.318]\ n(t) = \lim_{N\to\infty} \sum_{k=1}^{N} n_k \psi_k(t),$$

in which for every base $\{\psi_k(t)\}$, the r.v. N_k, which have determinations n_k, have all (see section 7.6.3.1) zero mean value and identical variance:

$$[7.319]\ E\{n_k\} = 0,$$

$$[7.320]\ E\{n_k^2\} = \mathcal{N}_0 = \frac{1}{2} N_0.$$

Moreover, the considered r.v. are all Gaussian and statistically independent among one another.

In the signal space with dimension N tending to infinite, a realization of the considered noise can be thus represented with a vector, **n**, having components n_k each with probability density function of the first order (see [7.301]):

$$[7.321]\ p(n_k) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{n_k^2}{N_0}} .$$

Thanks to the statistical independence, the resulting vector **n** has thus the probability density function of order N tending to infinite:

$$[7.322]\ p_{N\to\infty}(\mathbf{n}) = \prod_{k=1}^{N} p(n_k) = (\pi N_0)^{-N/2} e^{-\frac{|\mathbf{n}|^2}{N_0}} .$$

7.6.3.1 Appendix 3. The components of the gaussian noise

In the signal space identified by any real base $\{\psi_k(t)\}$, defined in the interval (-T/2;T/2), for a Gaussian noise the representation [7.318] is adopted, where the determinations n_k of the r.v. N_k are

$$[7.323]\ n_k = \int_{-T/2}^{T/2} n(t)\psi_k(t)dt .$$

Taking into account the linearity of the [7.323] and that the statistical mean value of the noise is zero, that means $E\{n(t)\}=0$, the following relation is achieved:

$$[7.324]\ E\{n_k\} = \int_{-T/2}^{T/2} E\{n(t)\}\psi_k(t)dt = 0.$$

The components of the noise are thus all zero mean random variables.
Still from the [7.323] the following relation is obtained:

$$[7.325]\ E\{n_k n_h\} = \int_{-T/2}^{T/2}\int_{-T/2}^{T/2} E\{n(x)n(y)\}\psi_k(x)\psi_h(y)dx\,dy = \int_{-T/2}^{T/2}\int_{-T/2}^{T/2} R_{nn}(y-x)\psi_k(x)\psi_h(y)dx\,dy .$$

In the case the noise is white, with constant spectral density N_0 in a bandwidth extended from the origin up to the maximum frequency f_M, implying that (see [7.283]):

$$[7.326]\ N(f) = N_0\ \mathrm{rect}(f/2f_M)$$

is valid, applying the inverse Fourier transform the autocorrelation function of the noise is obtained:

$$[7.327]\ R_{nn0}(\tau) = N_0 2f_M \mathrm{sinc}(2f_M\tau);$$

making f_M tend to infinite, then (see [7.304])

$$[7.328]\ \ R_{nn0}(\tau) = N_0\delta(\tau) = \frac{1}{2}N_0\delta(\tau),\ \ f_M \to \infty$$

is achieved. Introducing the [7.328] in the [7.325] thus

$$[7.329]\ E\{n_k n_h\} = N_0 \int_{-T/2}^{T/2}\int_{-T/2}^{T/2} \delta(y-x)\psi_k(x)\psi_h(y)dx\,dy = N_0 \int_{-T/2}^{T/2} \psi_k(t)\psi_h(t)dt = \begin{cases} N_0 = N_0/2 , & h=k \\ 0 , & h \neq k \end{cases}$$

is obtained. Thus, the components of the white noise with bandwidth tending to infinite are all incoherent random variables and, being Gaussian, statistically independent; moreover, they all have the same variance:

$$[7.330]\ E\{n_k^2\} = N_0 = \frac{1}{2}N_0.$$

7.7 MARKOV PROCESSES

The Markov processes are characterized by the property of each random variable extracted at the instant t_n to depend just on the variable extracted at the instant t_{n-1}.
They can be characterized both by a set of continuous and discrete states. In such latter case the are named also Markov *chains*. In this case if also the time variable belongs to a discrete set the dependence is always and just relative to one unit of time before.
The stated properties can be expressed by means of the following analytic relations:

$$[7.331]\ p[\mathbf{x}(t_{n+1})=\mathbf{x}_{n+1}|\mathbf{x}(t_n)=\mathbf{x}_n, \mathbf{x}(t_{n-1})=\mathbf{x}_{n-1}, \ldots, \mathbf{x}(t_1)=x_1] = p[\mathbf{x}(t_{n+1})=\mathbf{x}_{n+1}|\mathbf{x}(t_n)=\mathbf{x}_n]$$

where $t_1<t_2< \ldots <t_n<t_{n+1}$.
The birth and death process, the random walk and the renewal process are particular examples of Markov process.
Before starting to deal with the Markov processes separating the discrete time and the continuous time cases, some properties of general interest are presented.

7.7.1 Properties of the Markov processes

Property 1
Applying to the [7.331] the law of the chain

$$[7.332]\ p(x_1, \ldots, x_n)=p(x_n|x_{n-1}, \ldots, x_1) \cdots p(x_2|x_1)p(x_1)$$

the following relation is obtained:

$$[7.333]\ p(x_1, \ldots, x_n)=p(x_n|x_{n-1})p(x_{n-1}|x_{n-2}) \cdots p(x_2|x_1)p(x_1)$$

If viceversa the [7.333] is true for every n then the process $\mathbf{x}_n$ is Markovian because (see [7.332])

$$[7.334]\ p\left(x_n|x_{n-1},...,x_1\right)=\frac{p\left(x_1,...,x_{n-1},x_n\right)}{p\left(x_1,...,x_{n-1}\right)}=p\left(x_n|x_{n-1}\right)$$

Property 2
From the property 1 the following equality is carried out

$$[7.335]\ E\{x_n|x_{n-1}, \ldots, x_1\}= E\{x_n|x_{n-1}\}$$

Property 3
A Markov process keeps to be a Markov process if the time axis is reversed

$$[7.336]\ p(x_n|x_{n+1}, \ldots, x_{n+k})= p(x_n|x_{n+1})$$

Demonstration: the left side of the [7.336] is equal to

$$[7.337]\ \frac{p\left(x_n,x_{n+1},...,x_{n+k}\right)}{p\left(x_{n+1},...,x_{n+k}\right)}=\frac{p\left(x_{n+1}|x_n\right)}{p\left(x_{n+1}\right)}p\left(x_n\right)$$

and since

$$[7.338]\ p(x_{n+1}|x_n)p(x_n)= p(x_n,x_{n+1})= p(x_n|x_{n+1})\ p(x_{n+1})$$

the [7.336] results to be verified.

7.7.2 Time discrete Markov process

A sequence of random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ constitutes a discrete time Markov chain if for every n (n=1, 2, …) and any possible value that the random variables can assume ($i_1<i_2<\ldots<i_n$) the following condition occurs:

$$[7.339]\ p[\mathbf{x}_n=j\ |\mathbf{x}_{n-1}=i_{n-1}, \ldots, \mathbf{x}_1=i_1] = p[\mathbf{x}_n=j\ |\ \mathbf{x}_{n-1}=i_{n-1}]$$

where the p[·] is the transition probability.
Thus, consistently with the definition of this class of processes, the hystory of the evolution is completely contained in the current state. Obviously, to rebuild the succession of the states the knowledge of the initial probability is necessary. Therefore, given the initial probability and the probabilities of transition it is possible to univocally establish the probability to stay in the different states at the instant n.
If the transition probability doesn't depend on n the Markov chain is *homogeneous*, that means:

$$[7.340]\ p_{ij} = p[\mathbf{x}_n=j \mid \mathbf{x}_{n-1}=i].$$

In this way the probability of transition at m steps can be defined as

$$[7.341]\ p_{ij}^{(m)} \triangleq p\left[\mathbf{x}_{n+m} = j \middle| \mathbf{x}_n = i\right].$$

Also the recursive expression can be carried out:

$$[7.342]\ p_{ij}^{(m)} = \sum_k p_{ik}^{(m-1)} p_{kj} \qquad m=2, 3, \ldots .$$

Note that in general a homogeneous Markov chain is not stationary. Anyway, it tends to be stationary for $n\to\infty$.
The [7.342] simply show that to pass from the state i to the state j in m steps it is necessary to pass from the state i to the generic state k in m-1 steps and then from the state k to the state j in one step. The resulting probability is the product of the two probabilities because this last two events are independent. Summing the probabilities for all the possible k the $p_{ij}^{(m)}$ is obtained.
A Markov chain in which each state can be reached from any other state is named *irreducible*. In terms of transition probability this situation is expressed in the existence of an integer m_0 such that $p_{ij}^{(m_0)} > 0$.
Let's indicate with A the set of all the states of the Markov chain. A subset of the states A_1 is named *closed* if it is not possible any transition between one of any state of A_1 to any of the states of A_1^c (with A_1^c the complement of A_1). If A_1 is composed of just one state, this is named *absorbent* state (that corresponds to $p_{ii}=1$).
If A is closed and doesn't contain any closed subset, then an irreducible Markov chain is obtained. In the opposite case it is named *reducible*.
Let's consider the probability to return once to the state j after n steps:

$$[7.343]\ p_j^{(n)} \triangleq p\left[\text{ritorno a j in n passi}\right]$$

In general, the probability to return to the state j will be:

$$[7.344]\ p_j = \sum_{n=1}^{\infty} p_j^{(n)}$$

If $p_j = 1$ the state is named *recurring*, otherwise if $p_j < 1$ it is maned *transient*. Moreover, if it is possible to return to the state j only at the steps $\gamma, 2\gamma, 3\gamma, \ldots$ the state is named *periodic* with period γ. If $\gamma=1$ the state is *not periodic*.
Considering the states for which $p_j=1$ the recurrency mean time can be defined:

$$[7.345]\ M_j \triangleq \sum_{n=1}^{\infty} n p_j^{(n)}$$

If $M_j=\infty$ the state j is named with *zero recurrency*, if $M_j<\infty$ it is named with *non zero recurrency*.
Let's define the probability to find the system in the state j at the n^{th} step:

$$[7.346]\ \pi_i^{(n)} \triangleq p\left[\mathbf{x}_n = j\right]$$

Two theorems are introduced, without demonstration.

Theorem 1:
The states of an irreducible Markov chain are all transient or all with non zero recurrency or all with zero recurrency. In the case they are periodic they have all the same period γ.
Theorem 2:
For a homogeneous irreducible and not periodic Markov chain the limit probabilities

$$[7.347]\ \pi_i = \lim_{n \to \infty} \pi_i^{(n)}$$

always exist and are independent on the probability distribution of the initial state.
Moreover the two options can occur:

- ⇒ *All the states are transient or with zero recurrency and thus $\pi_j=0$ for every j;*
- ⇒ *All the states are with non zero recurrency and thus $\pi_j>0$ for every j; in this case it is possible to carry out $\pi_j=1/M_j$. The quantities π_j result univocally determined by the equation system*

$$[7.348]\ \sum_i \pi_i = 1$$

$$[7.349]\ \pi_j \triangleq \sum_i \pi_i p_{ij}$$

Example
Let's consider the finite state diagram shown in Figura 7.4 as representation of a Markov chain.

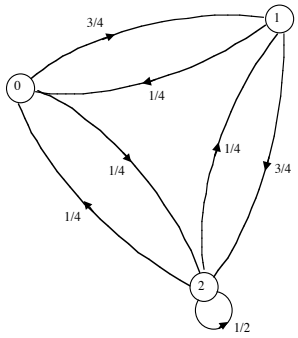


Figura 7.4 – Finite state diagram of a Markov chain

The transition probability matrix

$$[7.350]\ \mathbf{P}=[p_{ij}]$$

and the probability vector

$$[7.351]\ \boldsymbol{\pi}=[\pi_0, \pi_1, \pi_2, \ldots]$$

can be defined. In this way the [7.349] becomes

$$[7.352]\ \boldsymbol{\pi} = \boldsymbol{\pi}\,\mathbf{P}$$

In the considered example:

$$[7.353]\ \mathrm{P} = \begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

while the [7.352] becomes

$$[7.354]\quad \begin{aligned} \pi_0 &= 0\pi_0 + 0.25\pi_1 + 0.25\pi_2 \\ \pi_1 &= 0.75\pi_0 + 0\pi_1 + 0.25\pi_2 \\ \pi_2 &= 0.25\pi_0 + 0.75\pi_1 + 0.5\pi_2 \end{aligned}$$

As the first of the [7.354] is linearly dependent on the other two it is necessary to introduce the [7.348] that in this case is

$$[7.355]\ \pi_0+\pi_1+\pi_2 = 1$$

Solving the system

$$\pi_0=0.2$$

$$[7.356] \qquad \pi_1=7/25=0.28$$

$$\pi_2=13/25=0.52$$

In this way the state equilibrium probabilities are achieved.
Instead, if the dynamic behaviour of the system is interesting, the $\pi_j^{(n)}$, that are the probabilities that the system is in the state j in the instant n, must be figured out.
Let's define the vector

$$[7.357]\ \pi^{(n)} \triangleq \left[\pi_0^{(n)},\pi_1^{(n)},\pi_2^{(n)},...\right]$$

Using the definition of transition probability and the definition it is possible to calculate $\pi^{(1)}$ in terms of **P** of the initial distribution $\pi^{(0)}$

$$[7.358]\ \pi^{(1)} = \pi^{(0)}\ \mathbf{P}$$

Similarly

$$[7.359]\ \pi^{(2)} = \pi^{(1)}\ \mathbf{P} = [\pi^{(0)}\ \mathbf{P}]\ \mathbf{P} = \pi^{(0)}\ \mathbf{P^2}$$

Generalizing

$$[7.360]\ \pi^{(n)} = \pi^{(n-1)}\ \mathbf{P} = \pi^{(0)}\ \mathbf{P^n}$$

The steady state probability is given by

$$\pi = \lim_{n\to\infty}\ \pi^{(n)} = \lim_{n\to\infty}\ \pi^{(n-1)}\mathbf{P}$$

and thus

$\pi= \pi P$

Returning to the example, with the hypotesis that at the beginning the system is in the state 0 with probability 1, that means

$\pi^{(0)} = [1, 0, 0]$

the sequence of the probabilities is provided in Tabella 7.1.

Tabella 7.1

n	0	1	2	3	4	…	∞
$\pi_0^{(n)}$	1	0	0.25	0.187	0.203		0.2
$\pi_1^{(n)}$	0	0.75	0.062	0.359	0.254		0.28
$\pi_2^{(n)}$	0	0.25	0.688	0.454	0.543		0.52

In the case the vector of the initial probabilities is

$\pi^{(0)} = [0, 1, 0]$

the sequence of the probabilities is provided in Tabella 7.2.

Tabella 7.2

n	0	1	2	3	4	…	∞
$\pi_0^{(n)}$	0	0.25	0.187	0.203	0.199		0.2
$\pi_1^{(n)}$	1	0	0.375	0.25	0.289		0.28
$\pi_2^{(n)}$	0	0.75	0.438	0.547	0.512		0.52

Instead, if the vector of the initial probabilities is

$\pi^{(0)} = [0, 0, 1]$

the sequence of the probabilities is provided in Tabella 7.3.

Tabella 7.3

n	0	1	2	3	4	…	∞
$\pi_0^{(n)}$	0	0.25	0.187	0.203	0.199		0.2
$\pi_1^{(n)}$	0	0.25	0.313	0.266	0.285		0.28
$\pi_2^{(n)}$	1	0.5	0.5	0.531	0.516		0.52

It is possible to note that the limit probabilities are independent on the initial state.
Generalizing the homogeneous definition of the transiiton probability [7.340] it is possible to define:

$$[7.361]\ p_{ij}(m,n) \triangleq P\left[\mathbf{x}_n = j \middle| \mathbf{x}_m = i\right]$$

which provides the probability that the system is in the state j at the step n having been passed from the state i at the step m≤n. It is possible to state that in intermediate instant q the system is passed from intemediate states k.

$$[7.362]\ p_{ij}(m,n) = \sum_k P\left[\mathbf{x}_n = j, \mathbf{x}_q = k \middle| \mathbf{x}_m = i\right]$$

The [7.361] can be rewritten as:

$$[7.363]\ p_{ij}(m,n) = \sum_k P\left[\mathbf{x}_q = k \middle| \mathbf{x}_m = i\right] P\left[\mathbf{x}_n = j \middle| \mathbf{x}_m = i, \mathbf{x}_q = k\right]$$

but for the properties of the Markov chains

$$[7.364]\ P[\mathbf{x}_n=j|\mathbf{x}_m=i, \mathbf{x}_q=k] = P[\mathbf{x}_n=j|\mathbf{x}_q=k]$$

Applying to the [7.362] the following relation is obtained

$$[7.365]\ p_{ij}(m,n) = \sum_k p_{ik}(m,q) p_{kj}(q,n)$$

known as Chapman-Kolmogorov equation for a discrete time Markov chain.

7.7.3 Continuous time Markov processes

The continuous time Markov chains are characterized by having the time instants in which the system can change state belonging to a continuous set.
For every integer n and for every sequence $t_1, t_2, \ldots, t_{n+1}$ such that $t_1 < t_2 < \ldots < t_{n+1}$

$$[7.366]\ p[\mathbf{x}(t_{n+1})=j \,|\mathbf{x}(t_1)=i_1, \ldots, \mathbf{x}(t_n)=i_n] = p[\mathbf{x}(t_{n+1})=j \mid \mathbf{x}(t_n)=i_n]$$

The probability of transition as a function of time can be defined as follows:

$$[7.367]\ p_{ij}(s,t) \triangleq P\left[X(t) = j \middle| X(s) = i\right]$$

with X(t) state at time t>s. If three next instants s≤u≤t are considered, to pass from the state i at the time s to the state j at the time t the process shall have to pass from an intermediate state k at time u. In this way the Chapman-Kolmogorov equation for a time continuous Markov chain is obtained:

$$[7.368]\quad p_{ij}(s,t) = \sum_{k} p_{ik}(s,u)\, p_{kj}(u,t)$$

8 IMPERFECT TRANSMISSION

8.1 INTRODUCTION

The previous chapters have introduced the basic concepts of deterministic transmission of signals, in the cases of ideal and perfect conditions, and the mathematical instruments to characterize the random processes. Real transmission usually is neither ideal nor perfect and moreover some impairments occur due to non ideal behaviour of equipment and propagation means (transmission channel). The theory of random processes is useful to characterize such phenomena because the received signal is unknown and the impairments are intrinsically random.
In this chapter the basic principles of the non perfect connection will be introduced and the methodology to take into account the impairments in the analysis of both linear and non linear circuits will be approached.
Finally, the disturbs independent on the signals, and in particular the Gaussian noise, will be dealt with, including the noise equivalent bandwidth and the signal to noise ratio.

8.2 IMPERFECT CONNECTION

8.2.1 Impairments in the electric connection

Starting from the analysis developed in section 5.1.1, in which the ideal electric connection has been approached (see Figure 8.1), it is possible to identify different kind of impairments related to the cases of non ideal final user (case A), non ideal generator (case B) or both non ideal (case C). To generalize the situation the concept of section of the connection must be considered, as depicted in Figure 8.2: everything staying (before) on the left of the section belongs to the generator (equivalent source) while everything staying (after) on the right belongs to the final user (equivalent destination).

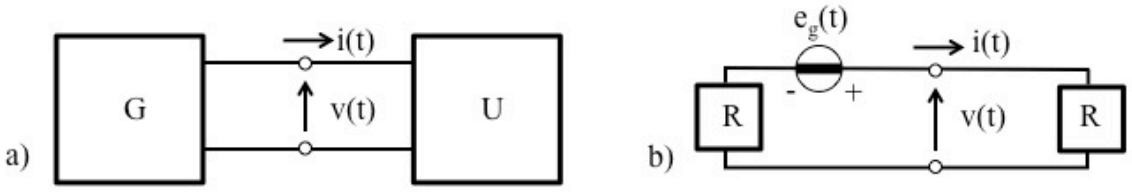


Figure 8.1: Ideal connection

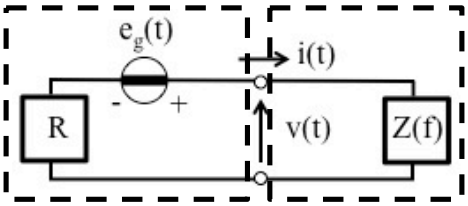


Figure 8.2: Concept of section of the connection

As concerns the above mentioned case A (Figure 8.2), the source can be represented with an ideal generator of electromotive force $e_g(t)$ having a constant resistance R connected in series while the user is assumed to be represented by a passive LTI bipole with impedance $Z(f)\neq R$ and as a consequence a reflection coefficient $\rho_u(f)\neq 0$. In this situation there is a direct signal, occurring in the section, which assumes the same expression of the signal in case of no reflection $x_d(t) = \frac{1}{2\sqrt{R}} e_g(t)$, demonstrating that the not ideal user doesn't impact it, it is the same in case of ideal connection and is also faithful. Then, there is a reflected signal which derives from the convolution between the input (the direct signal) and the pulse reflected response:

$$[8.1]\ x_r(t)=h_{ru}(t) \ast x_d(t)=[F^{-1}\{\rho_u(f)\}] \ast e_g(t),$$

which fully includes the consequences of the not perfect matching ($\rho_u(f)\neq 0$), it increases if the difference with the ideal case increases, it is related to the loss of instantaneous power given to the

user and is not proportional to the useful signal although dependent on it. Definitively, $x_r(t)$ is considered an undesired effect but with origin dependent on the signal.
In the case B (non ideal generator) it is possible to represent the non ideal source by means of an active LTI bipole with an internal impedance $Z_g(f)\neq R$ (reference resistance) and to characterize it considering a direct signal $x_g(t)$ generated with no reflection from the final user ($\rho_u(f)=0$), as shown in Figure 8.3. The reflection coefficient of the generator bipole, observed by the user considering $e_g(t)=0$ is (see also [5.9]):

$$[8.2]\ \varrho_u(f) = \frac{Z_g(f)-R}{Z_g(f)+R}.$$

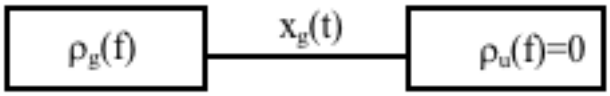


Figure 8.3: Non ideal source and perfectly matched final user

It is evident that the transfer is not perfect because it is not possible to evaluate $x_g(t)$ univocally as faithful signal generated by the source due to the not constant internal resistance, unless the transformation of the signal in electromagnetic format is known.
In the case C, depicted in Figure 8.4, both user and generator are not ideal. As a consequence, both $x_d(t)$ and $x_r(t)$ are dependent on both $\rho_g(f)$ and $\rho_u(f)$ (while in [8.1] the direct signal was independent on $\rho_u(f)$), because the reflected signal by U is not fully absorbed by G due to the reflection also on this side of the connection.

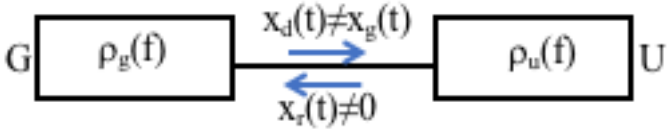


Figure 8.4: Both generator and user non ideal

Definitively, in the case both the generator and the user are not ideal the connection is not perfect and the following impairments occur:

- $x_g(t)$ is not faithful
- $x_d(t)$ is dependent on $\rho_g(f)$ and $\rho_u(f)$
- $x_r(t)\neq 0$

In practice, in the useful bandwidth the targets are $|\rho_g(f)|<<1$ and $|\rho_u(f)|<<1$ in order to try to avoid reflections or at least to limit the negative effects. Hereinafter, as simplyfing hypothesis, all the connection will be assumed to be perfectly matched.
Summarizing, in Table 8.1 the main concepts above introduced are reported.

Table 8.1: Imperfect connection

Case	*Condition*	*Impairment*
Ideal connection	$Z_g(f)=Z_u(f)=R$	Matching, no reflected signals
Ideal generator, non ideal user	$Z_g(f)=R$, $Z_u(f)\neq R$	Direct component faithful, reflected component not faithful
Non ideal source, ideal user	$Z_g(f)\neq R$, $Z_u(f)=R$	No reflected signal, direct component not faithful
Both source and user not ideal	$Z_g(f)\neq R$, $Z_u(f)\neq R$	Both direct and reflected signals not faithful

8.2.2 Additive unexpected effect in output

To analyse and characterize the behaviour of systems usually each equipment or transmission channel is considered as a block with an input and an output. Hereinafter the characterization of the behaviour of imperfect equipment will be approached utilizing a methodology based on the realization of an equivalent block diagram in which the impairment occurrence and its consequences on the system input-output behaviour can be separated from an ideal system. Thus, the characterization of the impairment can be associated to distinct blocks or functions.
In real systems the impairments to the transmission occur inside the equipment or along the transmission means. Therefore, the usual situation is that represented in Figure 8.5a in which a real base band transmission channel matched at the two ends is depicted. To carry out an analytical expression of the impairment it is possible to consider an ideal quadrupole (represented in Figure 8.5b) with transfer function $H_0(f)=g_0e^{-j2_\pi ft_0}$ (see [5.88]) and then figuring out the difference of the outputs. If x(t) is the input signal sample of a real stationary (or cyclostationary) process with power P_x, $x_0(t) = g_0x(t-t_0)$ is the expected faithful signal in output from Q_0 (idealized channel), r(t) is the received signal, sample of a stationary (or cyclostationary) process with power P_r and $r(t)\neq x_0(t)$ due to the impairments of the real channel. The deviation

$$[8.3]\ \varepsilon(t) = r(t) - x_0(t) = r(t) - g_0x(t-t_0)$$

is characterized as *additive unexpected effect* at the output. ε(t) depends on the single sample (thus not fully meaningful) and not necessarily it is harmful to transfer information because it contains a faithful addend.
Example $r(t) = g_1x(t-t_0)$ ➔ $\varepsilon(t) = (g_1 - g_0)x(t-t_0)$ faithful

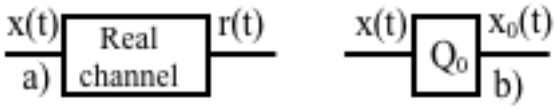


Figure 8.5: Real channel and idealized channel

To assess the imperfect transmission, disregarding the single sample (realization), to avoid that the unexpected effects contains a faithful component, it is opportune to express the received signal in a different way

$$[8.4]\ r(t) = x_u(t) + d(t)$$

with the faithful signal

$$[8.5]\ x_u(t) = g_ux(t-t_u)$$

and the other term

$$[8.6]\ d(t) = r(t) - x_u(t)$$

are defined to satisfy the requirement that t_u is such that the cross correlation between the received signal and the faithful addend $R_{rxu}(0)$ is maximum and g_u is such that the cross correlation between the undesired disturb process and the signal $x_u(t)$ $R_{dxu}(0)=0$. As a consequence of the two just provided relations on the cross correlations, $x_u(t)$ can be properly indicated as the useful signal received because due to the maximum correlation with r(t) there are no components in d(t), which is the disturb. In addition, still due to the $R_{dxu}(0)=0$ the power of the received signal is the sum of the powers of the useful signal and of the disturb

$$[8.7]\ P_r = P_u + P_d$$

and

$$[8.8]\ \int S_{dx_u}(f)df = 0,$$

being $S_{dxu}(f)$ mutual power spectral density. In general, the last two equations may not be valid due to the separation between events dependent and not dependent on the signal.
If the undesired event occurs at the output of a channel affected by impairments such as $d_i(t)$, sample of the process $D_i(t)$ with zero mean, statistically *independent* on the input signal x(t), $R_{dixu}(\tau)=0$ is always valid.
$D_i(t)$ is orthogonal also to any other disturbing process *dependent* on the signal with sample $d_d(t)$. Then:

$$[8.9]\ d(t)= r(t) - x_u(t) = d_d(t) + d_i(t)$$

$$[8.10]\ P_r = P_u + P_{dd} + P_{di}$$

Figure 8.6 pictorially shows how the different sets of effects are nested among one another.

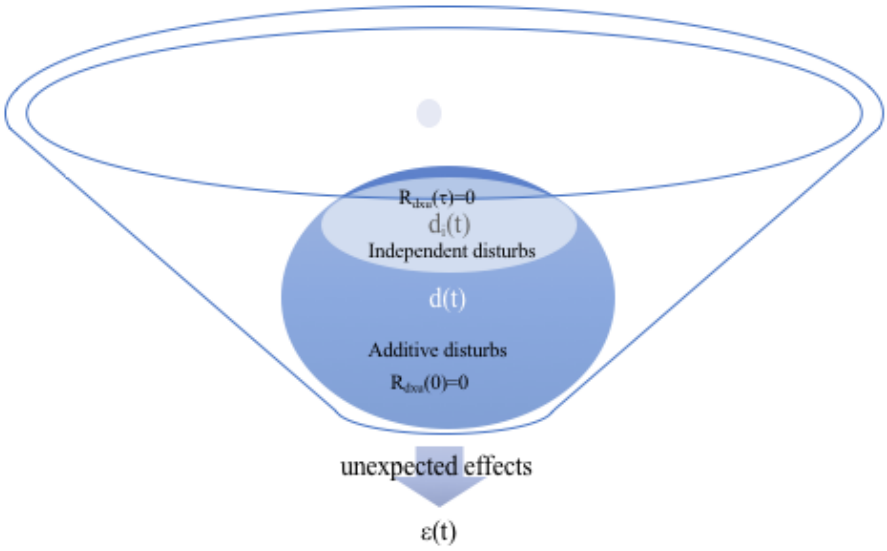


Figure 8.6: Set of unexpected effects containing additive disturbs containing independent disturbs

It is worth to consider an equivalent representation of an imperfect channel starting from [8.9]. In that respect the scheme of Figure 8.7 can be drawn.

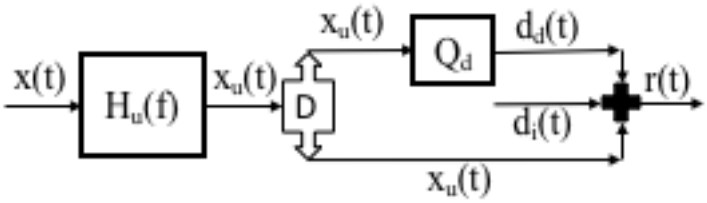


Figure 8.7: Equivalent scheme of channel affected by additive impairments

It is possible to note that the received signal results to be composed of three contributions that are simply summed:

1. the useful signal, output of an ideal quadrupole Q_u with transfer function

$$[8.11]\ H_u(f) = g_u e^{-j2\pi f t_u};$$

2. the independent disturb d_i;
3. the disturb due to distorsion d_d, dependent on the input signal, is the output of a distorting quadrupole Q_d which has at the input the useful signal.

In [8.11] g_u and t_u satisfy the conditions on the correlations [$R_{rxu}(0)$ is maximum, $R_{dxu}(0)=0$] and are different from the values of the ideal function. The element D is a diplexer.
It is worth to remark that $d_d(t)$ can be compensated if the processing of Q_d is known and reversible while $d_i(t)$ can be just limited but not removed not being able to carry out any information on it not being correlated with the input signal ($R_{d_i x_u}(\tau) = 0$).

8.2.3 Imperfect transmission with linear time variant channel

The quadrupole in base band which represents the channel can be *linear* and *matched* but, for example in radio channels, with pulse response h(u,t) variable with t randomly and statistically independent on the input signal x(t).
Applying the Fourier transform to h(u,t) with respect to u (time referred to the excitation instant of the Dirac pulse) [h(u,t) ⇔ H(f,t)] the result is the transfer function H(f,t) variable with the current time t both on the modulus and on the argument. The consequences are:

- ➔ Random deviation with respect to the ideal transfer fuction $H_0(f) = h_0 e^{-j2\pi f t_0}$
- ➔ Linear time variant distortion

due to unexpected additive effect dependent on the input signal but with random dependence statistically independent on the input process.
In this scenario it is worth to analyse the case of unexpected effect in channel with slowly variable attenuation. Thus, the main hypotheses are:

- hypotesis 1 The evolution of the impairment in the time variable linear channel is slow
- hypotesis 2 The two phenomena (signal and impairments) are statistically independent

In the adopted model the dependences on the two variables f and t is supposed to be separable so that it is possible to write

$$[8.12]\ H(f,t) = H_c(f) \bullet g_s(t)$$

where $g_s(t)$ is real, adimensional and positive function, sample of a process independent on the signal characterized by very narrow bandwidth, while $H_c(f)$ is assumed to have perfect behaviour up to f_M.

$$[8.13]\ H_c(f) \cong H_0(f) = g_0 e^{-j2\pi f t_0} \text{ for } |f|<f_M$$

g_0 and t_0 real constants of the ideal expected quadrupole.
The output will be the convolution between the pulse response (inverse Fourier transform of $H_c(f)$) and the input signal, with $|f|<f_M$. To this aim it is possible to consider

$$[8.14]\ h'(u,t)=g_s(t)F^{-1}\{H_0(f)\}=g_s(t)\ g_0\delta(u-t_0)$$

so that finally the output received signal will have the expression:

$$[8.15]\ r(t) \cong\ g_s(t)g_0[\delta(t-t_0) \ast x(t)]=g_s(t)g_0x(t-t_0)=g_s(t)x_0(t)$$

if $x_0(t)=g_0s(t-t_0)$ is the expected signal, the unexpected effect is (see [8.3]):

$$[8.16]\ \varepsilon(t)=r(t)-x_0(t) \cong\ [g_s(t)-1] \bullet x_0(t)$$

r(t) and ε(t) non stationary due to $g_s(t)$ which introduces dependence on t.
The [8.16] can be represented with the scheme of Figure 8.8a.

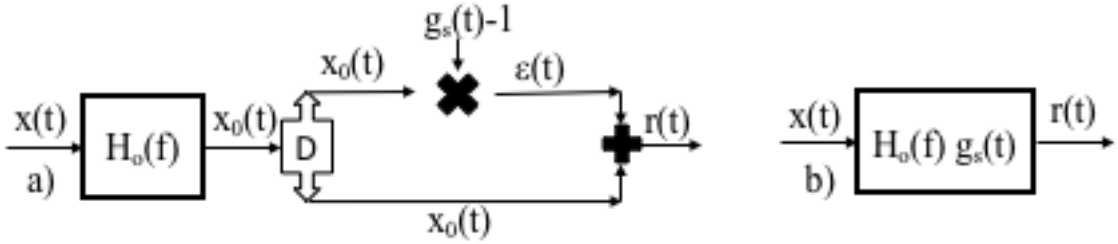


Figure 8.8: Equivalent scheme of linear channel with slowly variable attenuation

The concept of attenuation can be introduced. In fact, considering that the channel characteristics vary slowly, that the process with sample $g_s(t)$ is statistically independent on the one represented by x(t), applying the operator E{•} to the process x(t), it is possible to define the cross correlation function between the input and output processes:

$$[8.17]\ R_{rx}(t,t) \cong\ g_s(t)E\{g_0x(t+t-t_0)x^*(t)\}=g_s(t)g_0R_{xx}(t-t_0)$$

for $\tau=t_0$ it is max and its value is $R_{rxM}(t) \cong g_s(t)g_0P_x$ which coincides with the upper bound $\sqrt{P_r(t)P_x}$ and $P_r(t)$ is the slowly varying power of the received process, given by:

[8.18] $P_r(t)=R_{rr}(t,0) \cong E\{g_0x(t-t_0)\ g_0x^*(t-t_0)\}=g_s^2(t)\ g_0^2R_{xx}(0)=g_s^2(t)\ g_0^2P_x$

In conclusions, r(t) is just the useful received signal, because the considered phenomenon doesn't introduce disturb but removes the stationarity of the received process [$g_s(t)$ introduces dependence on t]. The scheme of Figure 8.8a can be simplified in the scheme of Figure 8.8b which represents the [8.15] from which it is possible to characterize the channel by means of a time variable attenuation:

$$[8.19]\ A(t) = \frac{1}{|H_0(f)g_s(t)|^2} = A_0A_s(t) = \frac{P_x}{P_x(t)}$$

in which it is possible to identify two contributions, the former is the expected attenuation with no variability [$g_s(t)=1$]

$$[8.20]\ A_0 = \frac{1}{|H_0(f)|^2} = \frac{1}{g_0}$$

while the latter is the supplementary attenuation

$$[8.21]\ A_s(t) = \frac{1}{g_s^2(t)}$$

Almost every time the channel behaviour introduces a time variable attenuation, it is quite easy to estimate the actual evolution of the function $g_s(t)$ after having opportunely processed r(t) at the output of the channel. Being the estimation independent on the expected signal it is possible to evaluate $A_s(t)$ instant by instant. Thus, it is easy to compensate the variable attenuation by means of a complementary quadrupole placed in cascade to the channel, characterized as linear, time variant and with variable transfer function and gain which can be expressed respectively as:

$$[8.22]\ \overline{H}(f,t) = \frac{1}{g_s(t)} \quad \text{for } f_m \le |f| \le f_M,$$

$$[8.23]\ G_{AGC}(f,t) = g_s^2(t) \quad \text{for } f_m \le |f| \le f_M.$$

Both functions are independent on the frequency within the useful bandwidth. Figure 8.9 shows the general scheme of the quadrupole utilized for the time variable compensation of the attenuation. It is called amplifier with automatic gain control (AGC). It is important that all the quadrupoles in the chain are linear or are frequency converters and it can be placed everywhere in the chain after the attenuator.

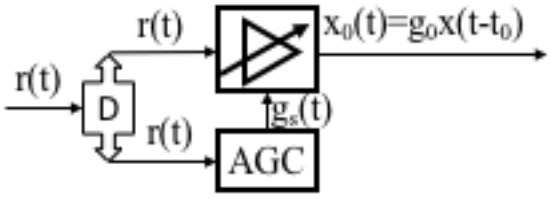


Figure 8.9: Quadrupole for Automatic Gain Control (AGC)

8.2.4 Imperfect transmission with linear time invariant channel

Time invariant linear distorsion

In the case of a real transmission channel that can be represented by means of a LTI base band quadrupole, characterized by diffusion parameters, which are a function of the frequency, and both the sections are matched, the only important function is the transfer one, which is the transmittance in the information transfer direction:

[8.24] $H(f)=g(f)e^{-j\Gamma(f)}$, with $g(f)=|H(f)|$, $\Gamma(f)=-\arg H(f)$

Then, being x(t) the sample of the input process, r(t) the sample of the output process and h(t) the channel pulse response, as usual the input output relation is given by the convolution:

$$[8.25]\ r(t) = h(t) * x(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau$$

If g(f) or Γ(f) don't satisfy the conditions for the ideal transfer in all the bandwidth of the input process, as a consequence r(t) is not faithful and the *time invariant linear distorsion* arises.
In Table 8.2 the possible effects and the consequences of the occurrence of time invariant linear distorsion are respectively listed, highlighting that in the former case the effects is irreversible while in the latter it is possible to restore the signal. Figure 8.10 shows the functional scheme of the occurrence of the two effects separately. The left block corresponds to the bandwidth cut, due to the presence of a band pass filter, while the right block introduces the in band linear distorsion.

Table 8.2: Effects and consequences of the time invariant linear distorsion

Two effects	
Partial suppression of the power spectrum of the input process	*Modification without suppression*
Distorsion due to bandwidth cut	In band linear distorsion
Irreversible	Possible restore the signal

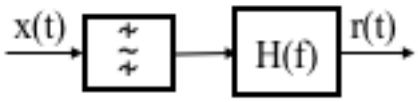


Figure 8.10: Block diagram highlighting the separation of the two effects of the time invariant linear distorsion

In the above mentioned conditions (real base band channel with imperfect transfer function) to assess in band linear distorsion two functions must be considered:

$$[8.26]\ \frac{g(f)}{g_0}[dB] = 20log_{10}\frac{g(f)}{g_0}$$

$$[8.27]\ \gamma_0(f) = \Gamma(f) - 2\pi f t_0$$

represented in Figure 8.11.

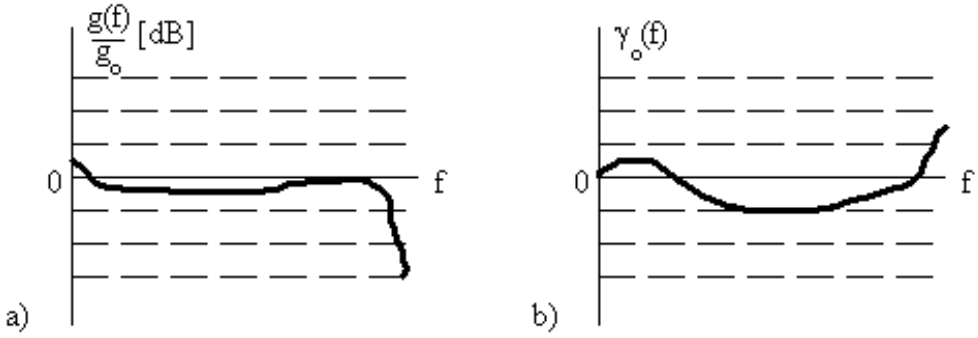


Figure 8.11: Example of evolution of the two functions characterizing the linear distorsion

The [8.26] and [8.27] contain two constants, respectively g_0 and t_0, that are referred to the transfer function $H_0(f)$ of the ideal quadrupole which represents the reference for the imperfect one within the bandwidth of the input process. Since $H_0(f)$ characterizes the ideal channel its output represents just the expected signal $x_0(t)=g_0x(t-t_0)$.

The imperfection of the channel is analytically represented by the difference between H(f) and $H_0(f)$. An important parameter to assess this difference is the relative deviation of the transfer function:

$$[8.28]\ H_\varepsilon(f) = \frac{H(f)-H_0(f)}{H_0(f)} = \frac{g(f)e^{-j\gamma_0(f)}}{g_0} - 1$$

Considering that the quadrupole representing the channel is physically feasible, both H(f) and $H_0(f)$ are Hermitian. As a consequence also $H_\varepsilon(f)$ is Hermitian. Inverting the [8.28] the result is:

$$[8.29]\ H(f)=H_0(f) + H_e(f)H_0(f).$$

Then, applying the inverse Fourier transform the pulse response of LTI channel becomes:

$$[8.30]\ h(t)=F^{-1}\{H_0(f)\} + F^{-1}\{H_\varepsilon(f)H_0(f)\}=g_0\delta(t-t_0) + h_\varepsilon(t) * g_0\delta(t-t_0);$$

Since the expected signal is $x_0(t)=g_0x(t-t_0)$, then the output signal can be carried out

$$[8.31]\ r(t)=g_0x(t-t_0) + h_\varepsilon(t) * g_0x(t-t_0)=x_0(t)+h_\varepsilon(t) * x_0(t)$$

where

$$[8.32]\ \varepsilon(t)=h_\varepsilon(t) * x_0(t)=F^{-1}\{H_\varepsilon(f)\} * x_0(t)$$

is the additive unexpected effect in output due to in band distorsion. Figure 8.12 shows the scheme derived from the [8.31]. It is possible to note that the quadrupole Q_ε, which introduces the unexpected effect, is LTI and its transfer function is $H_\varepsilon(f)$. Since it is hermitian and the input signal is real, then ε(t) is hermitian too.

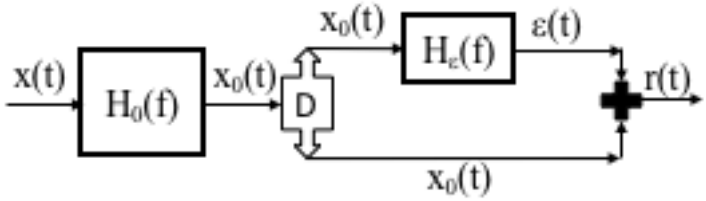


Figure 8.12: Representation of the imperfect linear time invariant channel

Applying the Fourier transform to [8.32] the result is that the spectrum of ε(t) is entirely contained in the band of x(t). It is important to remark that ε(t) may contain a faithful contribution proportional to $x_0(t)$ and for this reason it can't be considered a disturb.

The compensation of the distorsion can be achieved also with a complementary quadrupole, with respect to the imperfect one, placed in cascade. Such a methodology is called equalization of time invariant linear distorting channels, characterized by a transfer function H(f) independent on time, and is implemented by means of a quadrupole with the transfer function:

$$[8.33]\ \bar{H}(f) = \frac{1}{H(f)}\bar{g}e^{-j2\pi f\bar{t}} \text{ for } |f_m|\leq f\leq |f_M|$$

remembering that the cascade of LTI (or frequency converters) quadrupoles provides a total transfer function as the result of the product of all the transfer functions in the chain.

$\bar{g}$ and $\bar{t}$ are real and positive constants and in particular $\bar{t}$ must be sufficiently large to guarantee causality.

The equalization can be implemented also in case of time variant channels characterized by transfer functions H(f,t) dependent on time. In such a case it is called adaptive equalization.

For the physical feasibility $H(f)\neq 0$ in the useful band (in practise G(f) not too small) to avoid the presence of poles.

The equalization doesn't have to be necessarily located close to the distorting quadrupole because in the product of the transfer functions the order is not important, as long as the components are linear.

8.2.5 Imperfect transmission with non linear channel

Devices with non linear input output relation, for example amplifiers, can have a linear behaviour for small input signals, as in many cases happens. In some cases it is not possible to work with small signals, for example in presence of transmission means (in particular radio channels) characterized by high attenuation which needs to be compensated. Overdimensioning the amplifier to let it work in linear zone is not always possible due to technical/economic reasons.

Static non linear distorsion

Considering a generic r(t)=f[x(t),t] as input output characteristic function of a non linear device in a certain bandwidth and assuming on the basis of a restrictive hypotesis that

$$[8.34]\ r=f_g[x],$$

with $f_g[x]$ non linear function of x, independent on t, a more general representation of that one for small signals is obtained, called model of quadrupole with *static non linearity*, or *memoryless*. In Figure 8.13a a characteristic function of the type [8.34] is drawn. The part of the curve approximating the straight tangent line is representative of the behaviour of the non linear device for small signals, thus assuming an almost linear behaviour with gain:

$$[8.35]\ G_1=g_1^2, \quad g_1=(d/dx)f_g[x]|_{x=0}$$

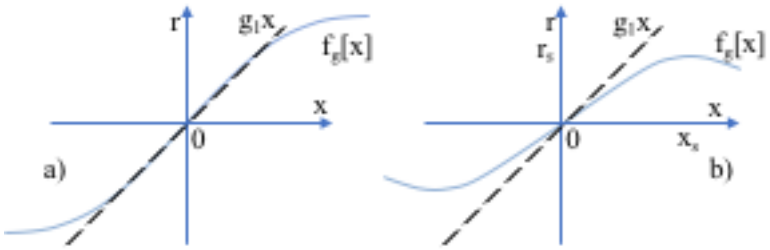


Figure 8.13: Non linear characteristic functions

In the linear zone we can assume that the output signal is faithful to the input one $r(t)\approx g_1x(t)$ but to let the circuits work in linear manner they must be overdimensioned accepting as coutrepart a low amplification efficiency.

Seldom they must work in conditions in which the derivative of the characteristic function is zero around the point of *saturation*, as indicated in Figure 8.13b.

As a consequence the output signal is different from the expected signal

$$[8.36]\ x_0(t)=g_1x(t)$$

and the *static non linear distorsion* occurs, generating at the output an additive unexpected effect due to non linear static distorsion that can be analytically expressed as

$$[8.37]\ \varepsilon(t)=f_g[x]-g_1x(t)$$

which is dependent on the input signal.

Analysis methodology of static non linearity

The [8.34] can be developed in power series with initial value in the origin:

$$[8.38]\ r = f_g[x] = \sum_{n=1}^{\infty} g_n x^n$$

where the coefficients g_n are real, constant and proportional to the n-order derivative of the static characteristic function, for x=0.

Since $x_0(t)=g_1x(t)$ is the expected signal, the unexpected effect becomes

$$[8.39]\ \varepsilon(t) = r(t) - x_0(t) = \sum_{n=2}^{\infty} g_n x^n$$

By means of $\gamma_n=g_n/g_1^n$ for $n\geq 2$

$$[8.40]\ \varepsilon(t) = \sum_{n=2}^{\infty} \gamma_n x_0^n(t) = f_\gamma[x_0(t)]$$

The single contributions $\varepsilon_n(t)=g_n x^n(t)$ of order n show a different dependence varying x(t): 1 dB of relative variation of P_x implies n dB variations in the power of $\varepsilon_n(t)$.

Once the coefficients g_n are known, the number of contributions to consider in the series development depends on P_x (usually $n\leq 5$ is ok). For small P_x, the behaviour is that occurring for small signals and $\varepsilon_n(t)$ can be neglected.

The spectrum $S_\varepsilon(f)$ of the unexpected process with sample $\varepsilon(t)$ can be carried out applying the Fourier transform to the autocorrelation function of $\varepsilon(t)$.

In case x(t) is in base band with maximum frequency f_M, $S_\varepsilon(f)$ will be extended up to nf_M (n max number of non negligible contributions). In case x(t) is in shifted band with f_M the upper bound of the monolateral bandwidth large B and centered on f_a, if $B<<f_a$ and n is small, then $S_\varepsilon(f)$ is concentrated around the origin and at the frequencies nf_a.

The consequence is that the unexpected effect due to non linear distorsion *extends also to other signals which use adjacent bandwidths.*

The [8.39] suggests the scheme of Figure 8.14 with the quadrupole Q_ε having non linear static characteristic function of the [8.40] and producing the unexpected effect.

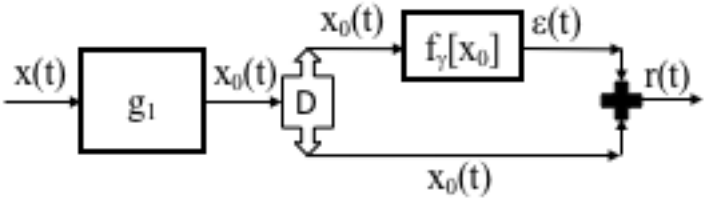


Figure 8.14: Equivalent representation of static non linear channel

Often in the unexpected effect a faithful addend proportional to $x_0(t)=g_1x(t)$ is present; then, the useful output signal is $x_u(t)=g_ux(t)$ with $g_u<g_1$ (compression).

Impairment reduction of the static non linearity

To reduce the effects of the non linear static distorsion different techniques can be applied.

Back off: reduction of the input power to work in the linear zone, but the output power decreases too (greater the number n of the unexpected contributions $\varepsilon(t)$ and more efficient the technique is).

Perfect selective filter: very selective in the useful band, removes the out of band components which can be extended much more than the useful band. The distorsion is reduced and other communications are not distrurbed. Using this technique, in case the input signal is in narrow shifted band with information on the instant phase and the expected signal is harmonic with $i(t)=A_c$ while starting from [8.40] the unexpected contributions are

$$[8.41]\ \varepsilon_n(t) = g_n s_m^n(t) = g_n A_c^n \cos^n[2\pi f_c t + \varphi(t)]\ ,\ \text{for } n \geq 2$$

It is possible to demonstrate that for n even all the spectral components are out of the useful band while for n odd one addend is in faithful form and others with external components. Filtering is useful to make perfect the transmission regardless the presence of static non linearity.

Combination of two twin imperfect quadrupoles: excited through a three port circuit which equally divides the input signal power on two signals, equal or inverted; the output distorted signals from the quadrupoles algebrically added unless a factor $1/\sqrt{2}$, with same or opposite sign, as shown in Figure 8.15, according to the following expressions:

$$[8.42]\ x_1(t) = \frac{1}{\sqrt{2}}x(t),\ x_2(t) = \pm\frac{1}{\sqrt{2}}x(t)$$

$$[8.43]\ r(t) = \frac{1}{\sqrt{2}}r_1(t) \pm \frac{1}{\sqrt{2}}r_2(t)$$

With the quadrupoles Q working with small signals, the scheme shows a linear behaviour, with gain equal to that of each of the component quadrupoles. Each one works with signals, both in input and in output, that are 3 dB below those in input and output of the whole scheme. With the same useful output power the quadrupoles operate with a back-off greater than 3 dB.

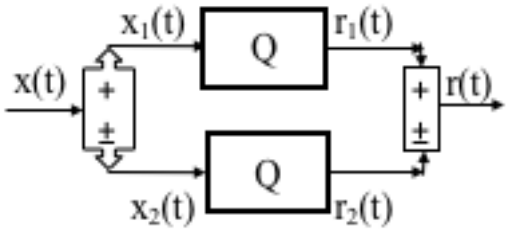


Figure 8.15: Pair of identical non linear quadrupoles

For both the choices of the signs, a first advantage: the unexpected effects of the 2nd order are reduced more than 3 dB, those of the 3rd order reduced of more than 6 dB, and so on for the effects of even higher order.
With the sign -, configuration of the quadrupoles in opposition of phase ("push pull"), the result is:

$$[8.44]\ r(t)=g_1x(t)+(1/2)g_3x^3(t)+\cdots+2^{(1-n)/2}g_nx^n(t),\ \text{for n odd},$$

from which the advantage of the back-off greater than 3 dB and the cancellation of the unexpected effects of even order are evident.
Thus, to reduce distorsion it is possible to use a pair of high power amplifiers with high efficiency even if characterized by meaningful non linearity of even order (that are removed).
Compensation of the static non linearity by means of predistorsion
To obtain a linear end-to-end behaviour almost up to the saturation it is possible to put in cascade before the distorting quadrupole an opportune quadrupole that can be represented with the non linear static model. The two functions

$$[8.45]\ r=f[x_p] \text{ and } x_p=p[x]$$

are the non linear static characteristics of a quadrupole Q and of the complementary one, P = $\bar{Q}$ *predistorting* quadrupole, inserted before in cascade to compensate the distorsion of the first one; moreover, g_1 is the gain of Q with small signals; to achieve and end-to-end linear behaviour of Q more extended of the one with small signals, the proportionality relation must be imposed:

$$[8.46]\ r = f\{p[x]\} = g_1x.$$

The proportionality constant is equal to the gain to small signals of the quadrupole Q, in order to have a small signal gain of the predistorting device equal to 1; in this way it works with output power close to the input one, i.e. not high and the quadrupole Q in condition very similar to that without compensation. Defining the inverse static characteristic of the predistorting quadrupole P as:

$$[8.47]\ x = p^{-1}[x_p]$$

from [8.46] it is possible to get:

$$[8.48]\ p^{-1}[x_p] = x = r/g_1=(1/g_1)f[x_p]$$

Since $x_p = p[x]$ must be monodrome, the compensation with predistorsion can be implemented only for values of x_p which allow to not reach the saturation of $r = f[x_p]$.

8.3 Imperfect transmission with independent disturbs

8.3.1 Independent (on the signal and among one another) disturbs

The real transmission through an imperfect channel is characterized also by undesired signals, which have origin not related to the useful signal and orthogonal with respect to it. They are independent on the useful signal and among one another. The physical origin is in phenomena (both intrinsic and extrinsic) arising in the quadrupoles which (in cascade) compose the channel. Usually, they have very small power spectral densities.

If the attenuation $A_A(f)$ of the transmission means is high, as a consequence on the receiving side very small useful powers are present. Then, just the disturbs at the output of the transmission means and in the receiving next quadrupoles are considered, until the useful powers are recovered to sufficient levels to be allowed to neglect the input of further disturbing contributions.
Every disturb independent on the useful signal can be observed only at the output of the quadrupole in which it is generated.
Regardless the nature and the localization of the disturbs, they can be represented by means of the equivalent scheme of Figure 8.16:

Figure 8.16: Equivalent representation of real channel with independent disturbs

Both the transmission means and the quadrupole Q are considered with no impairments while in the connection section additive source with realization $d_T(t)=\Sigma_j d_j(t)$, where $d_j(t)$ belongs to a process statistically independent on the others and on the useful one, is present. Also good matching in the connection sections (as usually occurs) is assumed.
As concerns the disturbs having intrinsic nature, in the transmission means and in the quadrupole Q, due to the physical nature of dielectric and conductors, of the components and of the used electronic devices, countless and very small additive undesired signals are generated, having the following characteristics:

- statistically independent among one another and
- not related to the useful signal in transit.

Due to the central limit theorem this set of phenomena generates gaussian disturbing processes with zero mean, **gaussian noise**, indicated with n(t). Two cathegories of noise can be identified:

1. *Intrinsic gaussian noise* of the transmission means $n_{Ai}(t)$ and
2. *equivalent gaussian noise* of the quadrupole $n_Q(t)$.

They are composed of the following effects:

- thermal noise (Johnson effect),
- quantic noise (electrical current fuctuation effect),
- scintillation noise (flicker effect),
- splitting (distribution) noise,
- eccess noise, etc.

the effect of the thermal noise in a resistor at the absolute temperature T (in Kelvin degrees) closed on a matched bipole generates a process with constant power spectral density

$$[8.49]\ N = \frac{1}{2}KT,$$

where $K = 1.38 \cdot 10^{-23}$ [J/K] is the Boltzmann constant.
As concerns the disturbs having extrinsic nature, they are introduced inside the transmission means and can be:

- *interferences*, associated to the invasiveness of electromagnetic signals which are generated in other systems or links different from the one considered;
  - Modeled with a single equivalent additive sample, indicated with $d_I(t)$
- *atmospheric and non natural disturbs*, which arises out of the telecommunication world due to natural shocks and to transient phenomena in electrical machines and plants, included the energy distribution lines.
  - represented by means of two equivalent additive samples, indicated with $d_p(t)$ (isolated causes, pulse disturbs) and $n_{Ae}(t)$ (continuous process with gaussian statistic, extrinsic gaussian noise of the transmission means, total effect due to a great number of causes uncorrelated among one another, all of them negligible as a single).

Thus, analytically:

$$[8.50]\ d_T(t)=\Sigma_j d_j(t) = d_A(t) + n_Q(t),$$

where the first addend

$$[8.51]\ d_A(t) = n_{Ai}(t) + n_{Ae}(t) + d_I(t) + d_p(t),$$

is due to the transmission means, while the second $n_Q(t)$ is due to the receiving quadrupole.

8.3.2 Reduction of the effects of independent disturbs

The spectral density relative to the sample $d_A(t)$ usually is much smaller than the useful one within its band, but the spectral extension is much greater. As a consequence the disturbing power can be of the same order of magnitude or even greater than the useful power (as well as the peak values of the respective samples).

Filtering and limitation in base band

If the useful process is in base band and only the peak values s_p and the extremes of the band (f_m and f_M) are known, $d_A(t)$ can be *reduced* utilizing a perfect filter with the bandpass equal to the bandwidth of the useful process, in the event followed by a non linear quadrupole, as shown in Figure 8.17.

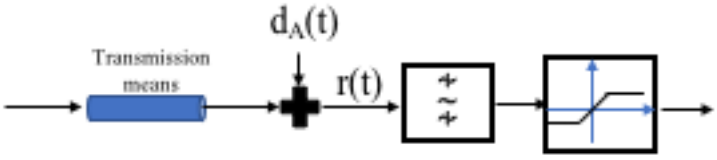


Figure 8.17: Use of filter and amplitude limiter to reduce independent disturbs

The filter can be very effective to reduce disturbs because they usually have very large bandwidth (gaussian) compared with that of the useful process. Instead, the filter has a little impact on pulse like disturbs because their spectral contents are practically all within the useful band, since the spectrum is decreasing with the frequency especially if f_m is small and f_M is large. Notwithstanding the filtering, the instant values signal+disturb can be much greater than the peak value s_p with no disturbs; the working condition of the non linear devices with small signals may be no longer valid.

Using an amplitude limiter (strongly not linear) with a characteristic function shown in Figure 8.18 the impairment can be removed with a limiting threshold A_s almost equal or slightly greater than s_p, keeping the limiter characteristic function linear up to A_s. The limitation reduces the bandwidth within the useful band but can introduce contributions out of the band. In this case further new filtering in the useful band is necessary.

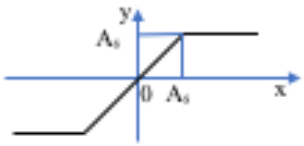


Figure 8.18: Characteristic function of an amplitude limiter

Filtering in shifted band

If the useful process is in shifted band (only the extremes are known), $d_A(t)$ can be *reduced* with a band pass filter.

Very often the reference frequency f_c is much greater than B so that only filters with large transition bandwidths are feasible.

Example: f_c = 10 GHz and B = 20 MHz, filter with Q_F=100. It is possible to remove the spectrum of the disturb out of a band $B + 2\Delta f_t = 20 + 200 = 220$ MHz (>10B).

$$\Delta f_t = f_t/Q_t = 10000000000/100=100000000=100MHz$$

In the case of the optical fibers f_c is so high that the use of the filter is not suggested; also because there is only the intrinsic gaussian noise, $d_A(t) \approx n_{Ai}(t)$, with very small spectral density.

The supereterodine receiver

A further reduction of the effects of the disturb can be obtained with a second filtering stage after having lowered the frequency.

Example: passing from f_c = 10 GHz to f_{cu} = 100 MHz and useful bandwidth B = 20 MHz, filter with Q_F = 100, the spectrum of the disturb out of $B + 2\Delta f_t = 20 + 2 = 22$ MHz is eliminated, greater than the useful one of just the 10%.

The technique is used in channels with very narrow shifted band (non optical), so that the quadrupole Q shown in Figure 8.16 is the cascade of linear amplifiers and a converter with band pass filters both in input, at f_c, and in output, at $f_{IF}=f_{cu}$ much smaller than f_c. This system, shown in Figure 8.19 is called supereterodine receiver. The frequency $f_{IF} << f_c$ and also greater than $3f_M$ of the modulating signal is called intermediate frequency.

For shifted band signals equivalent transfer functions in base band and complex envelopes are considered.

Q can be represented, as shown in Figure 8.20, by means of a quadrupole in base band with gain $\underline{G}(f)$, very selective, assuming constant $G_0=g_0^2$ value within the useful base band and values much smaller than G_0 just out of it.

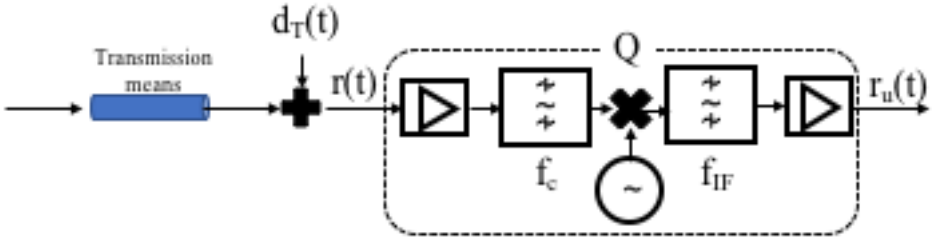


Figure 8.19: Equivalent representation of real linear channel in very narrow shifted band with additive disturb and supereterodine receiver

Figure 8.20: Equivalent scheme in base band of real linear channel in very narrow shifted band with additive disturb and selectivity provided by the supereterodine receiver

Indicating with $s_m(t)$ the complex envelope referred to f_c of the useful sample input to $\underline{G}$ and with $d_T(t)$ the complex envelope referred to f_c of the disturbing sample input to $\underline{G}$,

$$[8.52]\ r(t) = s_m(t)+d_T(t)$$

is the total complex envelope received by $\underline{G}$, while

$$[8.53]\ r_u(t) = g_0[s_m(t)+d(t)]$$

is the complex envelope of the output signal referred to the frequency f_{IF}.

Excluding g_0 the useful part is unchanged, while the disturbing one, d(t), is greatly reduced due to the cut of the band caused by the strong selectivity of the function $\underline{G}(f)$.

Using the supeterodine technique the output power and peak value of the additive disturb are so small that it is useless an amplitude limiter, which is instead inserted if the useful signal has constant envelope because it allows to keep constant the characteristic notwithstanding the disturb.

A further advantage is the partition of the amplification gain on more stages of the receiver at different frequencies.

As a consequence it is possible to use low gain (and low noise) amplifiers at high frequencies, where it is difficult to obtain high gains, and equipment at lower frequencies with much higher gains at even lower costs.

8.3.3 System additive gaussian noise

The disturbing contributions of gaussian nature (both intrinsic and extrinsic) concern

- the *transmission means*, characterised as one additive source of gaussian noise with sample $n_A(t) = n_{Ai}(t) + n_{Ae}(t)$ and power spectral density $N_A(f)$, and
- the *quadrupole*, equivalent gaussian noise, with sample $n_Q(t)$ and power spectral density $N_Q(f)$.

In total:

$$[8.54]\ n_S(t) = n_A(t) + n_Q(t),$$

sample of one additive gaussian process in the considered section of the channel, *system additive gaussian noise* (AGN = Additive Gaussian Noise) or *system noise* of the linear channel.
The two contributions are independent, therefore the noise power spectral density is:

$$[8.55]\ N_S(f) = N_A(f) + N_Q(f).$$

$N_A(f)$ at the output of the transmission means with attenuation $A_A(f)=1/G_A(f)$ varies case by case, but often the means is assumed to be in thermal equilibrium with the environment, at a temperature by convention T_0=290 K.
Assuming the input port of the transmission means connected to a matched resistor (at T_0), the bipole that can be observed from the output port provides the monolateral noise spectral density:

$$[8.56]\ N_{A0} = N_0 = KT_0,$$

otherwise connecting the output, by means of a generic pass band filter, to another resistor, matched and at temperature T_0, due to the N=(1/2)KT an exchange of energy at non zero balance would occur, causing the absurd of a heat exchange between two bodies at the same temperature.
The disturb at the output of the transmission means in thermal equilibrium with the environment doesn't depend on $A_A(f)$, which instead impacts the spectral density of the useful process in the same section.
The noise power at the output of the transmission means in a limited band usually assumes very small absolute values: for example in 1 MHz the power is $N_0 = KT_0 \cdot 10^6 = 4 \cdot 10^{-15}$ W; therefore, if the signal is strongly attenuated (due to great $A_A(f)$) it is not a problem.
$N_Q(f)$ depends on the particular physical realization of the quadrupole. If it is LTI, with gain $G_Q(f)$ and matched at both the ports, although not fed in input, provides at the output an additive gaussian process with zero mean, with power spectral density $N_{int}(f)$.

8.4 Spectral density and power analysis of a transmission system

So far the unexpected effects, disturbs and noise have been characterised in terms of statistical models, location (real or equivalent) of occurrence and nature. Very often it is important to assess quantitatively the noise power also with respect to the useful power and not only in absolute sense.

8.4.1 Parameters of the single quadrupole and of the cascade (spectral density and power)

Considering the quadrupole of the scheme in Figure 8.21 the input and output sections are assumed to be matched, that implies to have no reflected signals and x(t) and y(t) are assumed to be stationary stochastic processes at least in wide sense. Moreover:

$$x(t) \Leftrightarrow S_x(f) \Leftrightarrow P_x \quad \text{and} \quad y(t) \Leftrightarrow S_y(f) \Leftrightarrow P_y$$

with S(f) the respective spectral densities and with P the respective average powers.

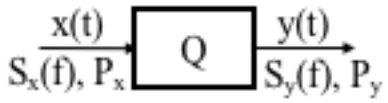


Figure 8.21: Generic LTI quadrupole with matched input and output

The gain function of the quadrupole is

$$[8.57]\ G_Q(f) = \frac{S_y(f)}{S_x(f)}$$

while, if y(t) is the faithful replica of the signal x(t), the gain of the quadrupole becomes

$$[8.58]\ G_Q = \frac{P_y}{P_x}$$

If, as in this case, the quadrupole Q is linear it can characterised by means of the complex transfer function $H_Q(f)$, the relation $S_y(f) = |H_Q(f)|^2 S_x(f)$ is valid and as a consequence $G_Q(f) = |H_Q(f)|^2$.
In case of a cascade of quadrupoles (for example two as shown in Figure 8.22), Q_1 and Q_2 are characterized by the transfer functions $H_{Q1}(f)$ and $H_{Q2}(f)$, the gain function of the cascade quadrupoles of Q_1 and Q_2 is

$$[8.59]\ G_Q(f) = |H_{Q1}(f)|^2 |H_{Q2}(f)|^2 = G_{Q1}(f) G_{Q2}(f).$$

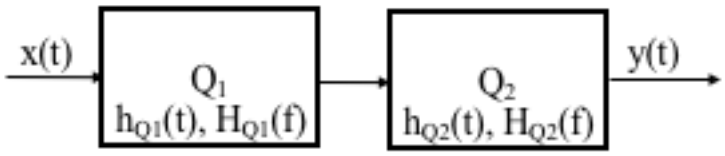


Figure 8.22: Cascade of quadrupoles

The transfer function of the equivalent quadrupole Q $H_Q(f)=H_{Q1}(f)H_{Q2}(f)$, can be easily generalized for M quadrupoles in cascade.
If the quadrupole Q distorts the input signal, the output signal y(t) will be the sum of the faithful signal to x(t), indicated with $x_0(t)$ and the unexpected effect in output ε(t) i.e.:

$$[8.60]\ y(t) = x_0(t) + \varepsilon(t).$$

If the unexpected effect at the output is an additive disturb and thus the signals are uncorrelated, then

$$[8.61]\ P_y = P_{x0} + P_\varepsilon$$

where P_{x0} is the faithful component power (useful), P_ε is the power of the unexpected effect which behaves as a disturb. On this basis, it is possible to define the gain of the quadrupole as

$$[8.62]\ G_Q = \frac{P_{x_0}}{P_x}$$

and the signal-disturb ratio:

$$[8.63]\ SNR_\varepsilon = \frac{P_{x_0}}{P_\varepsilon}$$

The unexpected effect ε(t) must be always considered in the analysis of quadrupoles first of all when they distort the input signal and the relative power must be taken into account in calculating the power of the output signal. The signal-disturb ratio is the most important parameter in the design of a physical telecommunication link.

8.4.2 Power analysis of noisy quadrupoles

Power analysis of a single noisy quadrupole

If the quadrupole **Q** is **noisy**, even without any input signal, at the output a signal characterised as a random process is observed, uncorrelated with the input signal, which depends on the electronic components utilized to build Q.

Indicating with d(t) the disturb at the output, with x(t) the input signal, the output signal will be $y(t) = x_0(t) + d(t)$, where $x_0(t)$ is faithful to the input.

The power of the output signal is $P_y = P_{x0} + P_d$. The gain of the quadrupole is

$$[8.64]\ G_Q = \frac{P_{x_0}}{P_x}$$

while the signal-disturb ratio is

$$[8.65]\ SNR_d = \frac{P_{x_0}}{P_d}$$

Last case: signal **x(t)** at the input **affected by disturb** $x(t) = x_u(t) + d_i(t)$.

If Q is linear (and the unexpected effect is neglected) for the principle of superimposition of effects

$$[8.66]\ y(t) = x_0(t) + d_{i0}(t) + d(t)$$

$x_0(t)$ faithful to x(t), $d_{i0}(t)$ caused by the input disturb $d_i(t)$, d(t) disturb introduced by Q.

The disturbs are independent among one another and with respect to the faithful component of the output signal. Therefore

$$[8.67]\ P_y=P_{x0}+P_{di0}+P_d$$

Thus:

- gain of the quadrupole $G_Q = \frac{P_{x_0}}{P_x}$, ratio between the average powers that takes into account the behaviour (amplification or attenuation) of the quadrupole on the useful signal at the input of Q;
- signal to disturb ratio at the input $SNR_i = \frac{P_x}{P_{d_i}}$;
- signal to disturb ratio at the output $SNR_o = \frac{P_{x_0}}{P_{d_{io}}+P_d}$.

If Q is non linear and is not working at the small signals or Q is linear but distorts, the disturbs dependent on the input signal (classified as unexpected effect) must be taken into account.

Noisy linear quadrupoles

If the transmission system is composed of **noisy linear quadrupoles**, the relation between the mean powers at the input and at the output is analysed.

If the quadrupoles are non distorting the transfer function is:

$$[8.68]\ H_Q(f) = \begin{cases} H_Q e^{-j2\pi f t_0} & f_m \le |f| \le f_M \\ 0 & otherwise \end{cases}$$

where $H_Q = (G_Q)^{1/2}$ and t_0 is a constant. If the quadrupole is distorting, the distorsion can be taken into account introducing the unexpected effect.

Equivalent representation of noisy quadrupoles

Considering a noisy quadrupole Q with no input signal, as shown in Figure 8.23A, d(t) is the disturb observed at the output and P_d is the power of the disturb, supposed stationary at least in wide sense.

Q can be represented as cascade of a NON NOISY quadrupole Q with a noise source $n_I(t)$ in input such that the power of the output noise process of Q is equal to P_d. Thus, if G_Q is the gain of the quadrupole, the power of the noise source at the input of the non noisy quadrupole Q must be:

[8.69] $N_I = \frac{P_d}{G_Q}$.

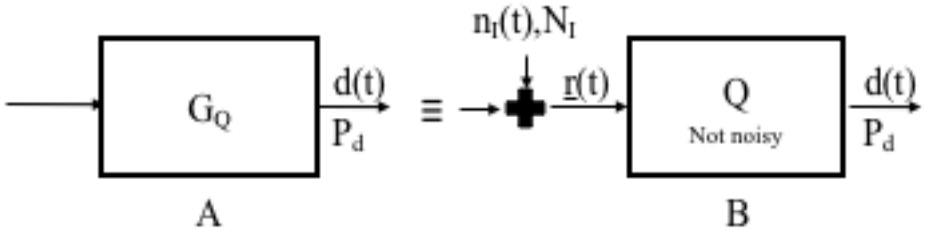


Figure 8.23: Noisy quadrupole represented as a cascade of a non noisy quadrupole and a noise source at the input

Noise equivalent bandwidth

Considering a non noisy quadrupole Q, with transfer function $H_Q(f)$, having at the input a signal n(t) white noise process with bilateral spectral density $S_N(f) = N_0/2$, the output signal (process) y(t) has spectral density $S_y(f) = |H_Q(f)|^2 N_0/2$. The power of the output process is:

[8.70] $P_y = \int_{-\infty}^{\infty} |H_Q(f)|^2 \frac{N_0}{2} df$

If $H_Q(f)$ is like in [8.68] for the output power

[8.71] $P_y = G_Q N_0/2 \cdot 2B = G_Q N_0 B$

is obtained, in which $B = f_M - f_m$ is the monolateral bandwidth of the quadrupole.
If $H_Q(f)$ is generic the goal is to get an expression like [8.71]:

[8.72] $P_y = \int_{-\infty}^{\infty} |H_Q(f)|^2 \frac{N_0}{2} df = N_0 G B_{eq}$

with G and B_{eq} to be defined to statisfy the equality. $G = G_0$ is assumed, equal to the value of the squared modulus of $H_Q(f)$ at the center of the band. Thus, the monolateral bandwidth B_{eq} must be determined. Since

[8.73] $\int_{-\infty}^{\infty} |H_Q(f)|^2 \frac{N_0}{2} df = \frac{N_0}{2} 2 G B_{eq}$

then

[8.74] $B_{eq} = \frac{1}{2G_0} \int_{-\infty}^{\infty} |H_Q(f)|^2 df$

In general, if B_{eq} is assigned and P_d is the power of the disturb at the output of a quadrupole with gain G_Q, it is possible to introduce a noise temperature of the quadrupole T_Q such that the power of the disturb introduced by Q at its output is:

[8.75] $P_d = kT_Q G_Q B_{eq}$.

Noise and system temperatures for a cascade of quadrupoles

Matching among subsystems composing the transmission system is assumed. In the case of the cascade of three quadrupoles, as shown in Figure 8.24:

- η_i (i=1,2,3) noise powers produced inside the quadrupoles,
- N_i total noise powers in the i^{th} section of the link,
- η_0 power of the disturbing process coming from outside measured within the band of the receiver.

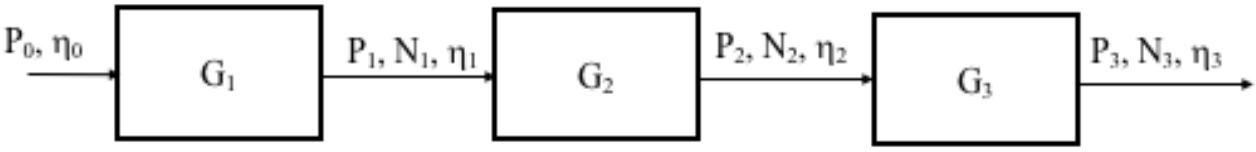


Figure 8.24: Cascade of noisy quadrupoles

Assuming non distorting quadrupoles, in the last section of the chain, the ratio between power of the useful signal P_3 and of the total disturb N_3 at the output of the last device can be indicated as:

$$[8.76]\ SNR_3 = \frac{P_3}{N_3}$$

If not distorting quadrupoles at least in the interested band, they can be represented with gains G_i, i=1,2,3 and the following relations are valid:

$$[8.77]\ P_3 = G_1G_2G_3P_0,\ N_3 = \eta_0G_1G_2G_3 + \eta_1G_2G_3 + \eta_2G_3 + \eta_3.$$

Replacing the [8.77] in the [8.76] and simplifying:

$$[8.78]\ SNR_3 = \frac{P_0}{\eta_0+\frac{\eta_1}{G_1}+\frac{\eta_2}{G_1G_2}+\frac{\eta_3}{G_1G_2G_3}} = \frac{P_0}{\eta_0+V_I}$$

the noise power V_I is the one introduced in total by the cascade of the quadrupoles applied at the input of the first quadrupole

$$[8.79]\ V_I = V_1 + \frac{V_2}{G_1} + \frac{V_3}{G_1G_2}$$

with V_i, i = 1, 2, 3 noise power of the equivalent noise source put at the input of the i^{th} quadrupole. Supposing to having defined an equivalent bandwidth common to the whole chain, that in case of quadrupoles with [8.68] is the bandwidth of the transmission system $B = f_M - f_m$, it is possible to express the V_i in terms of a noise temperature T_i:

$$[8.80]\ V_i = \frac{\eta_i}{G_i} = kT_iB_{eq}, i = 1,2,3.$$

Replacing [8.80] in the [8.79] and assuming that V_I can be expressed in terms of the temperature T_I ($V_I = kT_IB_{eq}$) the formula which relates T_I to the noise temperatures T_i is obtained

$$[8.81]\ T_I = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1G_2}$$

and can be generalised to the case of L linear quadrupoles in cascade as

$$[8.82]\ T_I = T_1 + \sum_{n=2}^{L}\frac{T_n}{\prod_{k=1}^{n-1}G_k}$$

Since the noise power at the input of the equivalent non noisy quadrupole with $G_Q = G_1G_2G_3$ is equal to $\eta_0 + V_I$, the system temperature T_s is defined considering that:

$$[8.83]\ ½kB_{eq}T_0 + ½kB_{eq}T_I = ½kB_{eq}(T_0 + T_I) = ½kT_sB_{eq}$$

from which

$$[8.84]\ T_s = T_0 + T_I.$$

Noise figure of a cascade of quadrupoles

The noise figure is an alternative parameter to take into account and to assess the noisy behaviour of a single or a chain of devices. As shown in Figure 8.25 the noisy quadrupole Q, characterized by a constant gain G_Q in band, has at its input a signal affected by disturb with power η_0 (within the bandwidth) and introduces a disturb at the output with power P_d.

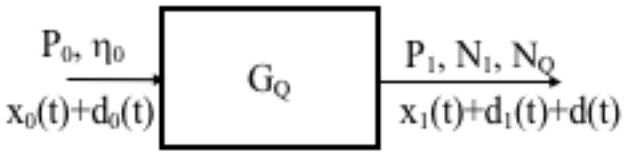


Figure 8.25: Noisy quadrupole with disturb plus signal in input

The noise figure F is defined as the ratio between the noise power at the output of the quadrupole and the noise power at its input. It takes into account the noise power increase due to Q supposed

noisy. F is also defined as the ratio between SNR at the input of the quadrupole and SNR at the output of the noisy quadrupole and by definition F≥1.
Using the second definition, it quantifies the degradation experienced by the input SNR due to the disturb introduced by the quadrupole. As a consequence:

$$[8.85]\ F = \frac{P_0}{\eta_0}\frac{1}{\frac{P_1}{N_1+\eta_1}}$$

with:

- P_1 power of the useful signal at the output of Q,
- N_1 power of the disturb associated to the input disturb and
- N_Q power of the internal disturb of Q.

F can be alternatively expressed as:

$$[8.86]\ F = \frac{N_1+N_Q}{N_1} = 1 + \frac{N_Q}{N_1} \geq 1$$

which confirms that F is the ratio between the total power of the disturb at the output of Q when it is noisy and the noise power occurring at the output if Q is NOT noisy and the only contribution of disturb is due to the input disturb.
Starting from [8.86], since $N_1 = G_Q\eta_0$, the total noise power at the output is:

$$[8.87]\ N_{out} = N_1 + N_Q = FN_1 = FG_Q\eta_0 = FG_QN_0B_{eq} = FkT_0G_QB_{eq},$$

Relation between the noise figure of Q and its noise temperature
Starting from [8.86], since $P_1 = G_QP_0$ and $N_1 = G_Q\eta_0$:

$$[8.88]\ F = \frac{\eta_0G_Q+N_Q}{\eta_0G_Q} = 1 + \frac{N_Q}{\eta_0G_Q}.$$

Since $V_Q = N_Q/G_Q$ is the power of the disturb produced by Q and applied at its input, then $V_Q=kT_QB_{eq}$.
Considering that $\eta_0 = kT_0B_{eq}$, replacing in [8.88] then:

$$[8.89]\ F = 1 + \frac{T_Q}{T_0}$$

and inverting:

$$[8.90]\ T_Q=T_0(F-1).$$

T_0 is the reference noise temperature to evaluate the noise figure.
Indicating with F_i the noise figures of the quadrupoles of the cascade, starting from [8.81] and utilising [8.90]:

$$[8.91]\ F^{(3)} - 1 = F_1 - 1 + \frac{F_2-1}{G_1} + \frac{F_3-1}{G_1G_2}$$

generalised to the case of L quadrupoles in cascade as

$$[8.92]\ F^{(L)} - 1 = F_1 - 1 + \sum_{n=2}^{L}\frac{F_n-1}{\prod_{k=1}^{n-1}G_k}$$

On the basis of the definition, the noise power at the output of the cascade of quadrupoles is:

$$[8.93]\ N_{Tot} = kT_0B_{eq}G_QF,$$

G_Q total gain of the cascade of quadrupoles.
Example of application of the analysis of the transmission system in terms of power
Both [8.81] and [8.91] show that the predominant contribution to the noise power is related to the noise temperature (or to the noise figure) of the first stage, i.e. T_1 (F_1), thus to the disturb produced by the first quadrupole of the cascade. As a consequence, the noise temperature (or the noise figure) of the first stage must be very small to avoid degradations impossible to compensate.

The transmission means can introduce disturbs (η_0) going in input to the receiving front-end after which there is the demodulator (DEM). The performance of the demodulator depend on the SNR ratio at its input. Thus, with reference to Figure 8.26, SNR is calculated after R and before DEM.
R amplifies and shifts in frequency the received signal whitin the band of DEM (with supereterodine receiver composed of at least three subsystems). It is important to design the cascade which realizes R to not introducing excessive degradations on the SNR at the input of R.

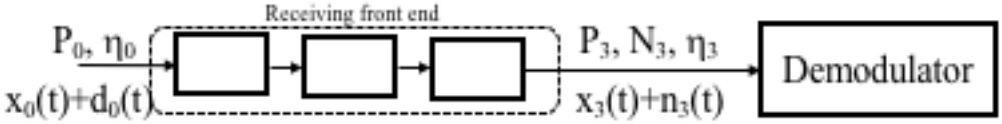


Figure 8.26:

The degradations can be due to disturbs introduced by quadrupoles and to elements of R which distort the signal creating an unexpected effect that can be observed at the output of the quadrupole. Greater the SNR ratio at the input of DEM, better are the performance.

Signal-disturb ratio at the output of the chain

Starting from [8.78], dividing numerator and denominator by P_0 the result is:

$$[8.94]\ SNR_3 = \frac{1}{\frac{\eta_0}{P_0}+\frac{\eta_1}{P_0G_1}+\frac{\eta_2}{P_0G_1G_2}+\frac{\eta_3}{P_0G_1G_2G_3}}$$

Since $snr_0 = P_0/\eta_0$ is the ratio SNR at the input of the cascade of quadrupoles and $snr_i = P_0G_1G_2...G_i/\eta_i$, i = 1,2,3 is the signal disturb ratio at the output of the i^{th} quadrupole calculated considering only the disturb introduced by the i^{th} quadrupole, the folllowing relation is obtained

$$[8.95]\ (SNR_3)^{-1} = (snr_0)^{-1} + (snr_1)^{-1} + (snr_2)^{-1} + (snr_3)^{-1}$$

the total signal disturb ratio depends on the signal disturb ratios at the output of each quadrupole each separately evaluated and respect to the only component of disturb introduced by the considered quadrupole.
From [8.95] it is evident that SNR_3 is limited by the lower value of snr_i (if $snr_1 << snr_i$ for i = 2,3 then $SNR_3 \cong snr_1$ is obtained). As a consequence it is important to care to not having subsystems in the cascade with a ratio snr_i such to compromise the performance of all the chain.

8.4.3 Sensitivity of the receiver

The disturbs produced by the components of the receiver and by the transmission means can't be removed because they are uncorrelated with the transmitted signal.
To allow the receiver to extract the information contained in the signal with the desired degree of reliability a minimum SNR must be guaranteed (signal over the level of the disturb).
The sensitivity (S) of the receiver relative to the considered disturb is the *minimum mean power detectable by the receiver*. If the mean power of the total disturb, indicated with N, is constant

$$[8.96]\ S = SNR_{min} \cdot N.$$

If the signal arrives with power below the defined S, the processing subsystem inside the receiving equipment doesn't guarantee the reliability degree required in recovering the information, implying that it is not able to exploit correctly its functionality because of the disturb.
The assessment of the minimum detectable power is fundamental and qualifies the whole link. As a matter of fact *lower S* implies *better behaviour of the receiver.*
If the power N of the disturb contains only the power of the thermal noise the sensitivity is defined only with respect to just the thermal noise. If in N the contributions of interference due to other transmitters or phenomena are included, the ratio signal/(noise+interference) is defined, (SINR) and the sensitivity is defined with respect to the SINR.
Considering a value of thermal noise, with the interference the sensitivity S grows (performance are worse).

9 BASIC PRINCIPLES OF ANALOGUE MODULATION AND DEMODULATION

9.1 Modulation of harmonic signals

The signals, of pass band type, that carry the information throughout the systems are represented as

$$[9.1]\ s(t) = \Re\{e^{j2\pi f_0 t}\}$$

with s(t) product of a complex signal x(t) with a complex signal of type $e^{j2\pi f_0 t}$. The complex exponential signal is deeply modified by means of this operation that is called **Modulation**. In general:

- x(t) usually in base band, complex envelope of the pass band signal, *modulating signal*
- $e^{j2\pi f_0 t}$=cos(2πf$_0$t)+jsin(2πf$_0$t) complex sum of two real harmonic signals, *carrier signal*
- f$_0$ *carrier frequency*

x(t) modifies the amplitude and phase characteristics of the complex signal $e^{j2\pi f_0 t}$. In fact, the complex envelope of s(t) can be expressed as

$$[9.2]\ x(t) = I(t) + jQ(t)$$

with I(t) and Q(t) two real modulating signals. Replacing [9.2] in the [9.1] it is possible to obtain:

$$[9.3]\ s(t) = I(t)\cos(2\pi f_0 t) - Q(t)\sin(2\pi f_0 t)$$

Transforming I(t)=R(t)cos(φ(t)) and Q(t)=R(t)sin(φ(t)) and replacing in [9.3] the following relation is obtained:

$$[9.4]\ s(t) = R(t)\cos(2\pi f_0 t + \varphi(t)),$$

in which $R(t) = \sqrt{I^2(t) + Q^2(t)}$ is the amplitude function, $\varphi(t) = \text{arctg}\left\{\frac{Q(t)}{I(t)}\right\}$ is the phase function. The function arctg{x} allows to evaluate the phase of the signal from the knowledge of I(t) and Q(t). The system which realizes this function is also called phase detector.
The expression [9.4] shows that the modulated harmonic signal is obtained varying the amplitude and/or the phase of the harmonic carrier.

9.1.1 General schemes of transmitters for modulated harmonic signals

The circuits to realise [9.3] and [9.4] are shown in Figure 9.1 (A and B). The subsystem to generate the modulating signal x(t) (either the pair (I(t),Q(t)) or the pair (R(t), ϕ(t))) and the subsystem to implement the modulation of the carrier signal of harmonic type are well separated.
The transformation which associates either I(t),Q(t) or R(t), ϕ(t) to the input signal must be REVERSIBLE.
Circuit A: *harmonic modulator* with carriers in phase and quadrature. Real carriers modulated by:

- I(t) (modulating signal for the carrier in phase) and
- Q(t) (modulating signal for the carrier in quadrature).

Circuit B: transmitter *AM-PM* (Amplitude Modulation - Phase Modulation).
The harmonic signal is modulated varying the amplitude (of the harmonic signal itself) at the output of the phase modulator with phase (or frequency) characteristics altered by the phase signal ϕ(t). The subsystem for the phase modulation is based on an oscillator which modifies its oscillation characteristics (phase or frequency) proportionally to the modulating signal ϕ(t).
In both schemes the circuit to implement the product is a *multiplier* or *mixer*.
The scheme A is more utilised in practice. The scheme to implement the complementary receiving scheme can be easily obtained.
The scheme to obtain the signal in [9.3] is shown in Figure 9.2. Starting from the received signal r(t), the receiver must carry out the components I(t) and Q(t). The two low pass filters remove the harmonic modulated components centered around 2f$_0$. The latter are generated when the input signal to the receiver enters the mixer relative to the carrier in phase thus generating a new modulated signal like:

$$[9.5]\ s(t)\cos(2\pi f_0 t) = [I(t)\cos(2\pi f_0 t) - Q(t)\sin(2\pi f_0 t)]\cos(2\pi f_0 t) =$$
$$= I(t)[\cos(0) + \cos(2\pi 2 f_0 t)] - Q(t)[\sin(0) - \sin(2\pi 2 f_0 t)]$$

The low pass filter removes the components around the frequency $2f_0$.

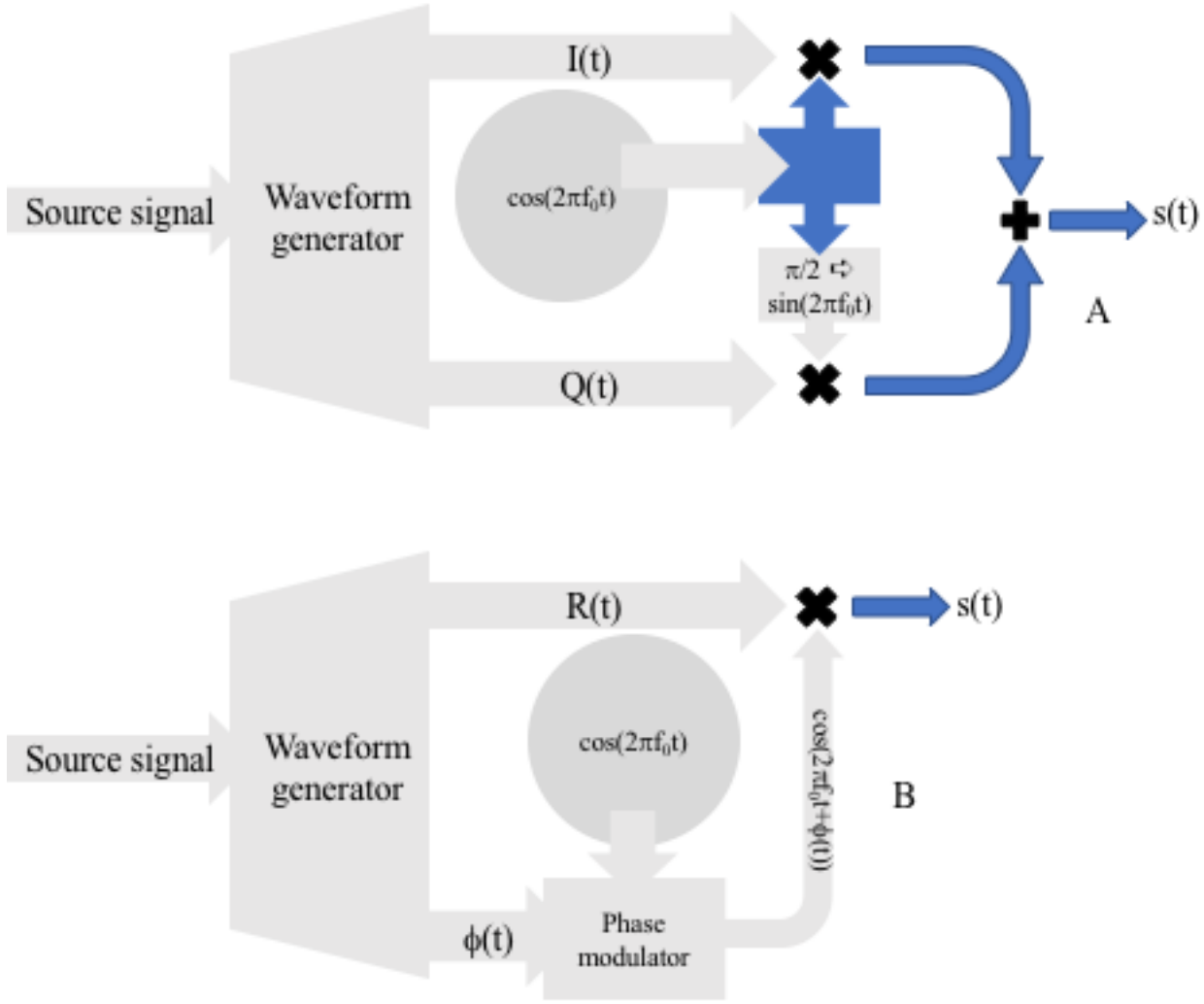


Figure 9.1: General schemes of transmitter for signals in [9.3] and [9.4]

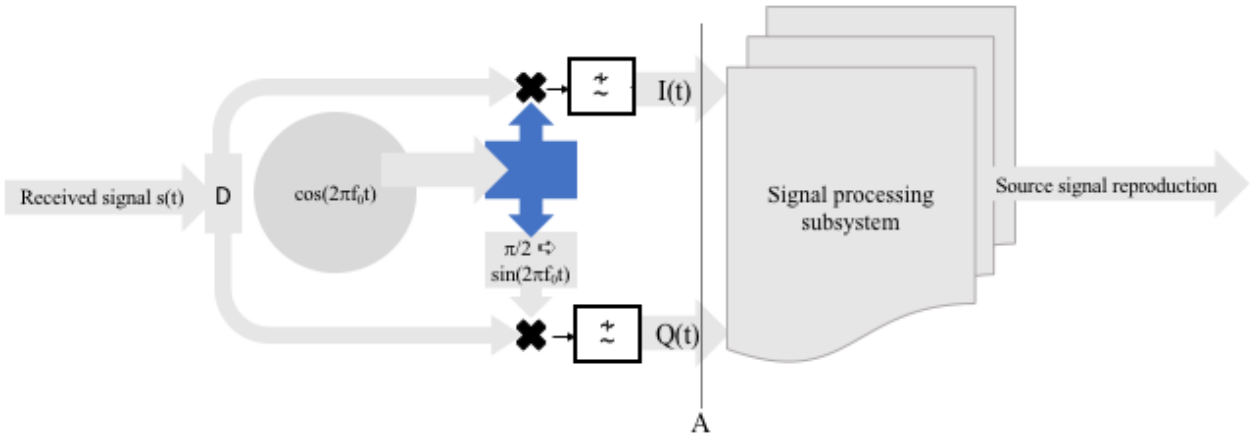


Figure 9.2: General scheme of receiver for signal in [9.3]

Similar considerations are valid in the case of mixer on the quadrature branch. The performance of the subsystem in charge to process the signals I(t) and Q(t) depend on the useful signal-disturb ratio measured at the section A.

To obtain the second member of [9.5] the carrier signals in the receiver must be identical to those ones used by the transmitter to generate the modulated harmonic signal. Without the *synchronization condition*, problems arise in regenerating the signal.

Example: harmonic carriers in the mixers in transmission with phase shift θ with respect to those ones in the receiver. The signal at the ouput of the transmitter, in the time reference of the receiver:

$$[9.6]\ s(t) = I(t)\cos(2\pi f_0 t + \theta) - Q(t)\sin(2\pi f_0 t + \theta).$$

The signal at the output of the low pass filter on the branch in phase:

$$[9.7]\ I'(t) = I(t)\cos(\theta) - Q(t)\sin(\theta)$$

I'(t) contains the desired contribution I(t) weighed by cos(θ) and the term Q(t)sin(θ) represents a disturb generated by the signal trasmitted on the carrier in quadrature. If $\theta = \pi/2$, then $I'(t) = -Q(t)$ implying that the received signal on the in phase branch is the same of that transmitted on the branch in quadrature but with opposite sign.

This kind of inconvenient can't be compensated if θ is unknown. The knowledge of θ allows to synchronise the receiver so that the product in the receiver is realized with the signals $\cos(2\pi f_0 t + \theta)$, $\sin(2\pi f_0 t + \theta)$ aligned in time with those ones of the transmitter. Similar issues and solutions can occur on the branch in quadrature.

The situation is more complex when the harmonic carriers of the transmitter differ in phase and frequency from those ones of the receiver. To avoid interference between the two transmissions in phase and in quadrature, before executing the product in the receiver, the carrier signals in the receiver must be synchronized with those ones of the transmitter (subsystems for synchronization not shown in the scheme of Figure 9.2 are used). If time references of receiver and transmitter are aligned, then the receiver operates in *coherent* way.

The scheme with circuits necessary for the synchronization recovery, which generate again the carrier signals in phase and quadrature with correct time alignment θ, is shown in Figure 9.3. The signals generated again by means of opportune processing of the received signal s(t) inside the subsystem is indicated as *synchronization circuit.*

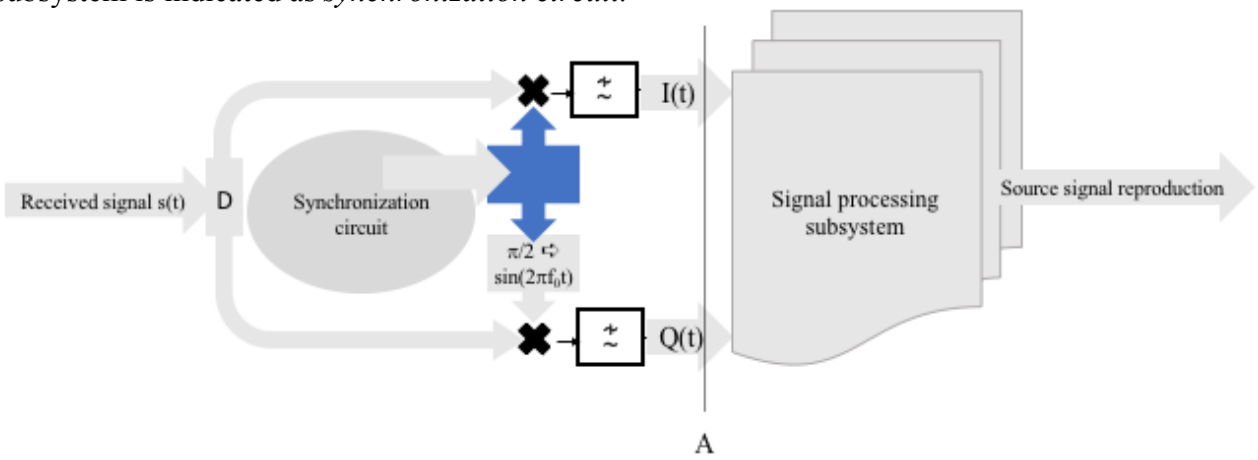


Figure 9.3: Receiver with synchronization circuit

9.2 Harmonic modulation with analogue modulating signal

The modulation schemes based on harmonic carrier can be divided in *entity* modulations (for example AM) and *angle* modulations (either the phase or the frequency are modified).

9.2.1 Amplitude modulations family

Classical amplitude modulation

For the classical amplitude modulation (AM), the complex envelope relative to the modulating signal is real:

[9.8] $x(t) = A_c[1 + m(t)] = I(t) + jQ(t)$, and $Q(t) = 0$,

and the AM signal takes the expression:

[9.9] $s(t) = A_c[1 + m(t)]\cos(2\pi f_0 t)$

A_c constant, $m(t)=k_a s_p(t)$ real modulating signal (contains the information), k_a constant such that $|m(t)|\leq 1 \ \forall t$. The general scheme of AM modulator is shown in Figure 9.4.

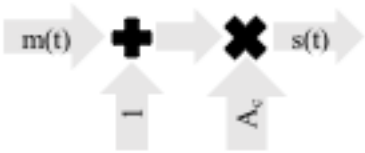


Figure 9.4: General scheme of AM modulator

The AM signal has harmonic component coherent with the one of the oscillator generating the modulated signal, used in the receiver to re-generate the carrier reference necessary for the correct demodulation. With modulating signal (process) characterized by zero mean stationary, the power spectrum of the AM modulated signal is:

[9.10] $S_{AM}(f)=(A_c^2/4)[\delta(f-f_0)+ \delta(f+f_0)+S_M(f-f_0)+S_M(f+f_0)]$,

where $S_M(f)$ is the power spectrum of the modulating signal. $S_{AM}(f)$ has two discrete components due to the presence of the harmonic component in the transmitted signal. In fact, a part of the power is used to send a residual part of carrier useful for synchronization circuits in the receivers but not carrying any information (which is contained in m(t)). The average power of the AM signal is:

[9.11] $P_{AM} = \frac{A_c^2}{2} + \frac{A_c^2}{2} P_M$

which is the integral of the spectrum [9.10]. P_M is the mean power of the modulating signal m(t). The power *efficiency* of the modulation is defined as:

[9.12] $\eta_{AM} = \frac{P_M}{P_x} = \frac{P_M}{1+P_M} < 1$

where P_x is the power of the complex envelope of the AM signal.
The peak power is:

[9.13] $P_{AM}^{peak} = \frac{A_c^2}{2}(1 + max\{m(t)\})^2$

Figure 9.5A shows the spectrum of the signal m(t) while Figure 9.5B shows the power spectrum of the AM signal. B is the band of the signal m(t) and 2B is the band of the modulated signal.

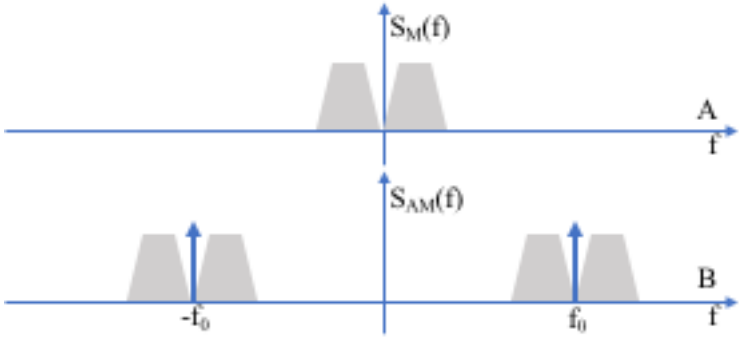


Figure 9.5: Spectrum of m(t) (A) and power spectrum of the AM signal

Double side band with suppressed carrier modulation

The signal with double side band with suppressed carrier (DSB-SC: Double Side Band - Suppressed Carrier) is obtained from [9.9] suppressing the term of the carrier achieving:

$$[9.14]\ s_{DSB\text{-}SC}(t) = A_c m(t)\cos(2\pi f_0 t)$$

The power spectrum of the signal $s_{DSB\text{-}SC}(t)$ is:

$$[9.15]\ S_{AM}(f)=(A_c^2/4)[S_M(f-f_0)+S_M(f+f_0)]$$

Of course the power efficiency of the DSB-SC signal is equal to 1, because all the transmitted power is used to send m(t).
The starting real signal had a bandwidth B, the modulated signal has always band equal to 2B and this could be an inconvenient if the physical channel over which the signal must be sent has some bandwidth constraints.
It is mandatory to have $f_0 > 2f_M$, being f_M the maximum frequency of the signal, necessary to avoid overlapping of the base band signal with the shifted band one.

Single side band with suppressed carrier modulation

To solve the inconvenient to take up the double of the band, signals with single side band with suppressed carrier are used (SSB-SC: Single Side Band - Suppressed Carrier).
A real modulating base band signal m(t) is used and the complex envelope is considered:

$$[9.16]\ x(t) = A_c(m(t) \pm j\hat{m}(t))$$

which is the analytic signal associated to m(t). Then the SSB-SC signal has the expression:

$$[9.17]\ S_{SSB\text{-}SC}(t) = A_c[m(t)\cos(2\pi f_0 t) \pm \hat{m}(t)\sin(2\pi f_0 t)]$$

in which $\hat{m}(t) = m(t) * h_H(t)$ is the Hilbert transform of m(t) and $h_H(t) = 1/\pi t$ nucleous of the Hilbert transform. The Fourier transform of the Hilbert nucleous is an ideal phase shifter of 90°, i.e. $H_H(f)=-j$ if $f>0$, $H_H(f)=j$ if $f<0$. Depending on the sign in [9.16] the suppression of one of the two subbands of the real signal m(t) is got.
With sign – the left side band is suppressed and the signal SSB with upper side band [USSB-SC (U: upper)] is obtained. With sign + the signal LSSB-SC (L: lower) with lower side band is obtained. Figure 9.6 shows the spectrum of an USSB-SC signal.
The mean power is:

$$[9.18]\ P_{SSB} = E\{s^2(t)\} = (A_c^2/2)\,E\{m^2(t) + \hat{m}^2(t)\} = A_c^2 P_M > P_{AM}$$

A SSB signal occupies half of the band of a DSB-SC signal with the same P_M utilized.

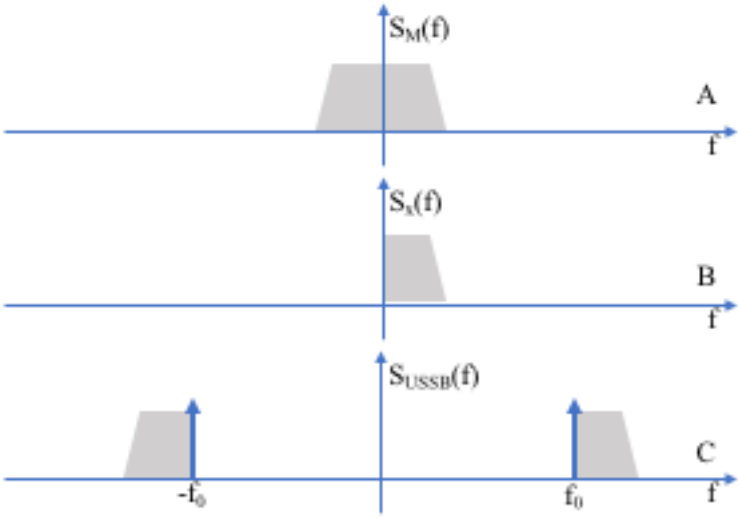


Figure 9.6: Spectrum of a SSB signal

9.2.2 Angle modulation: phase (PM) and frequency (FM)

Phase and frequency modulations are particular case of angle modulations. For both modulations the complex envelope of the modulating signal is:

$$[9.19]\ x(t) = A_c e^{j_\theta(t)}$$

The modulus of the complex envelope is constant and equal to $R(t)=A_c$; the phase of the harmonic carrier $\phi(t)=2\pi f_0 t+\Theta(t)$ results to be a linear function of the modulating signal $\Theta(t)$.
Starting from [9.1] and using [9.19] the expression of the generic angle modulated signal is:

$$[9.20]\ s(t) = A_c \cos(2\pi f_0 t + \Theta(t)).$$

Indicating with m(t) the signal containing the information

- the angle signal is phase modulated if: $\Theta(t) = D_p m(t)$
- the angle signal is frequency modulated if: $\Theta(t)=D_f \int_{-\infty}^{t} m(s)ds$

with D_p and D_f constants.
The instant frequency of the angle modulated signal is:

$$[9.21]\ f_i(t) = \frac{1}{2\pi}\frac{d}{dt}\phi(t) = f_c + \frac{1}{2\pi}\frac{d}{dt}\Theta(t)$$

The instant frequency of the frequency modulated signal:

$$[9.22]\ f_i(t) = f_0 + \frac{1}{2\pi}D_f m(t).$$

Thus, from [9.22] it is possible to carry out that the instant frequency of the harmonic signal frequency modulated varies as a function of m(t). Moreover, the frequency deviation of the signal with respect to the carrier frequency can be defined as:

$$[9.23]\ \Delta f(t) = f_i(t) - f_0 = \frac{1}{2\pi}\frac{d}{dt}\Theta(t)$$

and the peak frequency deviation, ΔF, can be defined as:

$$[9.24]\ \Delta F = max\left\{\frac{1}{2\pi}\frac{d}{dt}\Theta(t)\right\}$$

For FM signals the frequency peak deviation depends on the signal peak value m(t):

$$[9.25]\ \Delta F = \frac{1}{2\pi}D_f max\{m(t)\}$$

An increase of the amplitude of m(t) implies an increase of peak deviation of the signal frequency and thus the FM signal bandwidth enlarges but the transmitted signal power DOESN'T increase (always proportional to $A_c^2/2$).
For entity (amplitude) modulated signals a variation of the modulating signal power affects the power of the transmitted signal but DOESN'T increase the bandwidth occupied by the modulated signal.
For PM type signals the peak phase deviation is

$$[9.26]\ \Delta\theta=\max\{\Theta(t)\}$$

which is related to the peak value of m(t) by

$$[9.27]\ \Delta\theta = D_p \max\{m(t)\}$$

Two important parameters can be defined:
1. Phase modulation index: $\beta_p = \Delta\theta$;
2. Frequency modulation index: $\beta_f = \Delta F/B$, in which B is the modulating signal bandwidth.
In the case the modulating signal is a pure sinusoid instead of B the frequency f_m of the sinusoid is considered.

Spectrum of angle modulated signals

$$[9.28]\ S_{PM,FM}(f)=(A_c^2/4)\ \{S_x(f-f_0)+S_x^*(-f-f_0)\}$$

with $S_x(f)$ power spectrum of the complex envelope (Fourier transform of the correlation function of the process $x(t) = A_c e^{j\phi(t)}$), that is not simply to calculate.
The spectral analysis of angle modulated signals is usually limited to those that can be Fourier transformed. The spectrum of the signal S(f):

[9.29] $S(f)=(A_c/2)\{X(f-f_0)+X^*(-f-f_0)\}$

with X(f) Fourier transform of the complex envelope x(t):

[9.30] $X(f) = F\{A_c e^{j_\Theta(t)}\}$

Also X(f) is not easy to be calculated because it should be developed for the specific Θ(t).
It is interesting if m(t) is sinusoidal like $m_p(t) = A_m\sin(2\pi f_m t)$. Indicating with $\beta= D_p A_m$, the phase function is $\Theta(t) = \beta\sin(2\pi f_m t)$.
For the FM with sinusoidal modulating: $m_f(t) = A_m\cos(2\pi f_m t)$, the frequency peak deviation is

[9.31] $\Delta F = D_f A_m/2\pi.$

The complex envelope of the modulated signal with sinusoidal m(t) is:

[9.32] $x(t) = A_c e^{j\beta sin(2\pi f_m t)}$

which is a periodic signal with period $T_m = 1/f_m$, which can be developed in Fourier series

[9.33] $x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn2\pi f_m t}$

in which it is possible to demonstrate that

[9.34] $c_n = \frac{A_c}{T_m}\int_{-\frac{T_m}{2}}^{\frac{T_m}{2}} e^{j\beta sin(2\pi f_m t)} e^{-jn2\pi f_m t} dt = A_c J_n(\beta)$

with $J_n(x)$ Bessel function of the first type of order n, that can't be expressed in closed form. Calculating the Fourier transform of [9.33] the (striped) spectrum of x(t) is achieved:

[9.35] $X(f) = \sum_{n=-\infty}^{\infty} c_n \delta(f - nf_m) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta)\, \delta(f - nf_m)$

Drawing the evolution of the spectrum X(f) varying β and f_m it is evident that the bandwidth of an angle modulated signal depends on β and on f_m.

Carson formula

It was found out that the 98% of the total power of the FM or PM signal is contained in the band:

[9.36] $B_T = 2(\beta+1)B$

with β modulation index (phase or frequency), B band of the modulating signal which is equal to f_m when the modulating signal is sinusoidal.
The Carson rule provides and indication accurate enough to assess the bandwidth occupied by a PM or FM signal even with not sinusoidal modulating signal as a function of B (bandwidth of the modulating signal).

Narrow band angle modulation

When the modulating signal takes very small values (ex. $|\Theta(t)| < 0.2$) developing in series the complex envelope x(t) and stopping at the first order it is possible to obtain:

[9.37] $x(t) \cong A_c(1 + j\Theta(t)) = A_c + jA_c\Theta(t) = I(t) + jQ(t)$

from which:

[9.38] $s(t) = A_c \cos(2\pi f_0 t) - A_c\Theta(t)\sin(2\pi f_0 t)$

The signal is very similar to a signal type AM with non suppressed carrier in which the modulating signal is on the branch in quadrature instead that on the in phase one.
Considering the [9.38] the Fourier spectrum of a narrrow band angle modulated signal can be calculated:

[9.39] $S(f)=(A_c/2)[\delta(f-f_0)+\delta(f+f_0)+j(Q(f-f_0)-Q(f+f_0))]$

Q(f) Fourier transform of Θ(t) and

- $Q(f) = D_p M(f)$ if the signal is PM type or
- $Q(f) = D_f M(f)/j2\pi f$ if the signal is FM type.

9.3 Performance analysis of harmonic modulation systems with analogue signals

To evaluate performance of the different modulation schemes the utilized parameter is the signal-noise ratio at the output of the respective receivers as a function of the signal-noise ratio at the input of the receiver as a function of different significant parameters.
Hypothesis 1: coherent receivers.
Hypothesis 2: transmission channel introducing no distorsion and the only additive disturb is white Gaussian with spectral bilateral density $N_0/2$.
At the input of the receiver the signal has the expression:

$$[9.40]\ r(t) = s(t) + n(t),$$

Utilising the complex envelope both for the useful signal and for the disturb, also r(t) can be represented with its complex envelope $x_r(t)$, sum of the two complex envelopes:

$$[9.41]\ x_r(t) = I(t) + n_I(t) + j(Q(t) + n_Q(t))$$

in which $n_I(t)$ and $n_Q(t)$ are the components in phase and in quadrature associated to the complex envelope of the disturb process.
To compare performance of harmonic signals modulated with analogue signals:

1. The signal to noise ratio at the output of the receiver assuming that at the input there is the modulated signal + gaussian additive noise [hypothesis: received useful signal power P_s (in r(t)=s(t)+n(t) equal to the one of the transmitted signal) is the same in all the cases].
2. The signal to noise ratio at the output is expressed as a function of the ratio between P_s and the noise power in the base band of the modulating signal that is transmitted [signal-disturb ratio in base band (BB)]

$$[9.42]\ SNR_{BB} = \frac{P_s}{N_0 B}$$

3. Indicating with B_T the pass band signal bandwidth, the signal-noise ratio at the input of the receiver evaluated in the band B_T of the input signal of course is equal to:

$$[9.43]\ SNR_{in} = \frac{P_s}{N_0 B_T}$$

The signal-disturb ratio at the input can be always related to the ratio [9.42]:

$$[\ 9.44]\ SNR_{in} = \frac{P_s}{N_0 B_T} = \frac{P_s}{N_0 B_T}\frac{B}{B} = SNR_{BB}\frac{B}{B_T}$$

9.3.1 Performance of AM systems

Let's consider the demodulation system for AM signals based on the coherent scheme of Figure 9.2. Starting from the complex envelope of r(t) the signal noise ratio at the output is:

$$[9.45]\ SNR_{out}^{(AM)} = \frac{A_c^2 P_M}{2N_0 B}$$

in which P_M is the power of the modulating signal. Since the power of the signal at the input of the receiver is

$$[9.46]\ P_s=(A_c^2/2)(1+P_M)$$

the signal-noise ratio at the input is:

$$[9.47]\ SNR_{in}^{(AM)} = \frac{\frac{A_c}{2}(1+P_M)}{2N_0 B}$$

so that the ratio between [9.45] and [9.47] is:

$$[9.48]\ \frac{SNR_{out}^{(AM)}}{SNR_{in}^{(AM)}} = \frac{2P_M}{1+P_M}$$

Starting from [9.48] it is possible to achieve:

$$[9.49]\ \frac{SNR_{out}^{(AM)}}{SNR_{BB}^{(AM)}} = \frac{P_M}{1+P_M}$$

In the case of a DSB-SC system, repeating all the passages the result is

$$[9.50]\ \frac{SNR_{out}^{(DSB-SC)}}{SNR_{in}^{(DSB-SC)}} = 2$$

and

$$[9.51]\ \frac{SNR_{out}^{(DSB-SC)}}{SNR_{BB}^{(DSB-SC)}} = 1$$

Thus, the performance of a DSB-SC system are the same of the base band system but with double required bandwidth for the transmission in shifted band (in fact $B_T = 2B$).
For USSB or LSSB systems the result is

$$[9.52]\ SNR_{out}^{(USSB)} = \frac{A_c^2 P_M}{N_0 B}$$

considering that the occupied bandwidth is just B. Since the power of the modulated signal is $P_s=A_c^2P_M$:

$$[9.53]\ \frac{SNR_{out}^{(USSB)}}{SNR_{in}^{(USSB)}} = 1$$

and

$$[9.54]\ \frac{SNR_{out}^{(USSB)}}{SNR_{BB}^{(USSB)}} = 1$$

which shows that a SSB system is equivalent to a base band system both as concerns the performance in terms of signal-disturb ratio and for the occupied bandwidth ($B_T = B$).
Moreover, DSB-SC and SSB are equivalent from signal-disturb at the output point of view but the SSB occupies one half of the bandwidth.

9.3.2 Signal-noise ratio for PM and FM systems

The assessment of the signal-disturb ratio at the output of the receiver for PM and FM signals is difficult. Working with coherent detector of the type in Figure 9.2 once obtained again the components in phase and in quadrature of $x_r(t)$ in [9.41] the modulating signal is recovered using an angle detector which gives back a function:

- $K\angle x_r(t)$ for PM signal;
- $Kd/dt(\angle x_r(t))$ for FM signal.

where K is the gain of the angle detector. As an example the output of the angle detector can be obtained calculating $\arctan\{\Im\{x_r(t)\}/\Re\{x_r(t)\}\}$. Reminding the expression of $x_r(t)$ it is possible to immediately observe that without disturb the angle detector gives back exactly $\phi(t)$.

PM signals analysis

If the signal-noise ratio at the input of the receiver is high it is easily possible to demonstrate that the signal-noise ratio at the output can be approximated as:

$$[9.55]\ SNR_{out}^{(PM)} = \frac{A_c^2 D_p^2 P_M}{2N_0 B}$$

Reminding that $D_p = \beta_p/V_p$, β_p modulation index, V_p peak value of m(t), [9.55] becomes:

$$[9.56]\ SNR_{out}^{(PM)} = \frac{A_c^2 \beta_p^2 P_M / V_p^2}{2N_0 B}$$

The signal-noise ratio at the input for the PM case is:

$$[9.57]\ SNR_{in}^{(PM)} = \frac{A_c^2/2}{2(N_0/2)B_T} = \frac{A_c^2}{2N_0B_T}$$

with $B_T = 2(\beta_p + 1)B$ which is the Carson relation specialized to the PM case.
The ratio between [9.56] and [9.57], using the expression of di B_T is:

$$[9.58]\ \frac{SNR_{out}^{(PM)}}{SNR_{in}^{(PM)}} = 2\beta_p^2(\beta_p + 1)\frac{P_M}{V_p^2}$$

while

$$[9.59]\ \frac{SNR_{out}^{(PM)}}{SNR_{BB}^{(PM)}} = \beta_p^2\frac{P_M}{V_p^2}$$

from which it is possible to note that the improvement of signal-noise ratio with respect to the base band signal is directly proportional to the phase modulation index β_p, but growing β_p the phase deviation can go out of the main interval $[-\pi,\pi)$. In this event, a circuit of phase unwrapping must be used to get at the output the actual value of the phase and not the wrapped one in the main interval.
To limit the phase deviation in $[-\pi,\pi)$, the mandatory condition is $D_p m(t) = \beta_p m(t)/V_p < \pi$. If the modulating signal is sinusoidal the max improvement of the ratio is 6.9 dB.
The obtained expressions are valid only when the signal-noise ratio at the output is high. In fact, increasing β_p also the occupied bandwidth by the signal grows and thus the bandwidth of the input filter, the power of the noise grows and thus, with the same received power, the signal-noise ratio at the input decreases and the above formulas are no longer valid.

FM signals analysis

The FM detector gives back a signal proportional to the derivative of the phase of the complex envelope of the received signal.
With high signal-noise ratio at the input:

$$[9.60]\ SNR_{out}^{(FM)} = \frac{3A_c^2\beta_f^2P_M/V_p^2}{2N_0B}$$

Signal-noise ratio at the input:

$$[9.61]\ SNR_{in}^{(FM)} = \frac{P_s}{N_0/2B_T} = \frac{A_c^2}{4N_0(\beta_f+1)B}$$

Ratio between the two SNR:

$$[9.62]\ \frac{SNR_{out}^{(FM)}}{SNR_{in}^{(FM)}} = 6\beta_f^2(\beta_f + 1)\frac{P_M}{V_p^2}$$

with β_f frequency modulation index, V_p peak value of the modulating signal m(t).
Since the power of the FM signal is always $P_s = A_c^2/2$:

$$[9.63]\ \frac{SNR_{out}^{(FM)}}{SNR_{BB}^{(FM)}} = 3\beta_f^2\frac{P_M}{V_p^2}$$

If the modulating signal is sinusoidal the result is $(P_M/V_p^2) = 1/2$ and as a consequence:

$$[9.64]\ \frac{SNR_{out}^{(FM)}}{SNR_{BB}^{(FM)}} = \frac{3}{2}\beta_f^2$$

The performance of a FM system can improve increasing β_f. As in the PM case, if β_f increases then also the bandwidth of the signal grows and thus, with the same received power, the signal disturb ratio at the input decreases and the above formulas are no longer valid.
For high values of β_f the minimum value of the signal-disturb ratio at the input for which the relations are valid must increase (threshold effect).

If the modulating signal is sinusoidal, the relation between base band and output signal-disturb ratios is:

$$[9.65]\ SNR_{out}^{(FM)} = \frac{\beta_f^2 SNR_{BB}^{(FM)}}{1+\left(\frac{12}{\pi}\beta_f\right)SNR_{BB}^{(FM)} e^{-\frac{1}{2(\beta_f+1)}SNR_{BB}^{(FM)}}}$$

The curves of Figure 9.7 show that the performance of a FM system can be much better than those of the base band system.
For β_f= 5 and $SNR_{BB}^{(FM)}$ = 25 dB the performance of the FM sytem are of 15.7 dB better than those of the base band system but the occupied bandwidth by the transmitted signal is larger.
Inside a transmitter or receiver for FM signals some modifications can be applied to allow to further improve the performance of FM systems.

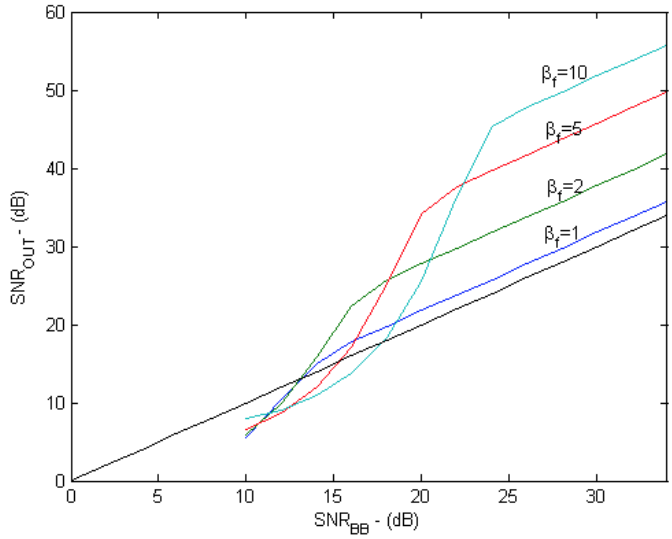


Figure 9.7: Performance of FM receiver with Gaussian noise

9.4 Carrier synchronization and Phased Locked Loop

9.4.1 Carrier recovery track and hold

Synchronization is one of the most critical functions of a communication system with coherent receiver. To some extent, it is the basis of a synchronous communication system.

The receiver needs to estimate and compensate for frequency and phase differences between a received signal's carrier wave and the receiver's local oscillator for the purpose of coherent demodulation, regardless the communication system is analog or digital.

The demodulation process, whose basic principles are introduced in the previous sections, to be perfectly and successfully exploited needs the information concerning the exact phase value to be recovered with an error as small as possible.

To reach this goal a circuit, based on a feedback loop system and named Phase Locked Loop (PLL), is implemented. The scope of the circuit is to keep coherence between a reference signal at frequency f_i and the output signal at frequency f_0, comparing the phases.

A phase-locked loop is a feedback system composed of a voltage controlled oscillator (VCO) and a phase comparator connected so that the oscillator keeps a constant phase angle relative to a reference signal. Phase-locked loops can be used, for example, to generate stable output high frequency signals from a fixed low-frequency signal.

In order to work properly the signal used in a PLL device shall be cleaned from the modulation. This result can be achieved either transmitting a pilot frequency without modulation or deriving a signal without modulation directly from the received signal. This signal is used as an input for the PLL device.

PLL is often used in carrier synchronization. It is a closed-loop control system consisting of:

- Phase detector (PD) a nonlinear device which generates at the output the phase difference between two oscillating input signals V_i and V_0;
- Voltage-controlled oscillator (VCO) nonlinear device which adjusts the oscillator frequency based on this phase difference with the aim to eliminate it so that at steady state, the output frequency will be exactly very close to the input frequency (ideally the same);
- Loop filter (LF);
- A feedback interconnection.

The general scheme of a PLL is shown in Figure 9.8 while the main equations are shown below:

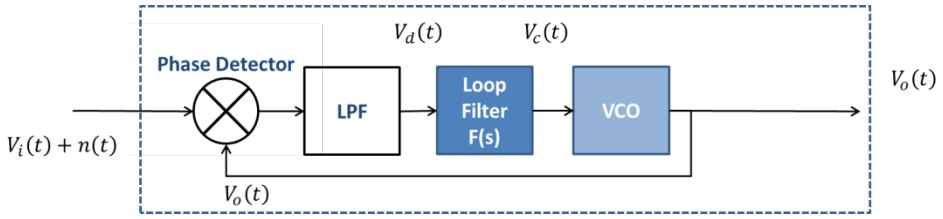


Figure 9.8: PLL (Phase Locked Loop) Scheme

$$[9.66]\ V_i(t) = V_i \sin(2\pi f_i t + \phi_i)$$

$$[9.67]\ V_o(t) = V_o \cos\left(2\pi f_0 t + \hat{\phi}_o\right)$$

$$[9.68]\ V_d(t) = K_d \sin(2\pi \Delta f t + \Delta\phi) \approx K_d(2\pi \Delta f t + \Delta\phi)$$

$$[9.69]\ V_c(t) = V_d(t) * h_{F(s)}(t)$$

The phase detector receives as input the reference signal and the output of the VCO and provides as output an error signal which is used as the input of the VCO, after a filtering stage which is not always applied.

In case of coherent demodulation Δf and Δϕ must be estimated and brought to 0 by the PLL device. The output of VCO synthetizes a frequency which is different from the input frequency. Via a mixer and a low pass filter a tension V_d is generated, which is proportional to the phase depending on $2\pi\Delta f t + \Delta\phi$. This term can drive, after the loop filter, the VCO to synthetize a frequency and phase closer to the input one.

In Figure 9.9 a combination of the two types of errors, frequency and phase, are presented:

- the frequency error generates a rotation of the carrier vector on the Real/Imaginary plane;
- the phase error generates a deviation from the reference phase (in this case 45 degrees has been assumed for visual representation).

In Figure 9.10 the effect of a pure phase error is shown, assuming that $\Delta f = 0$ the thermal noise has been assumed -50 dB, while the overall noise includes the phase error.

In Figure 9.11 the effect of a pure frequency error is shown, assuming that $\Delta\phi = 0$ the thermal noise has been assumed -50 dB, while the overall noise includes the frequency error.

A possible way to operate for PLL is to reduce the phase dependence on time reducing Δf and then keeping to a minimum the phase error $\Delta\phi$, as shown in Figure 9.12.

It is important to note that once the frequency shift has been reduced to a minimum (let's assume 1 part over 100.000) the error can be considered a pure phase error and the linearized model of PLL is valid and usually adopted.

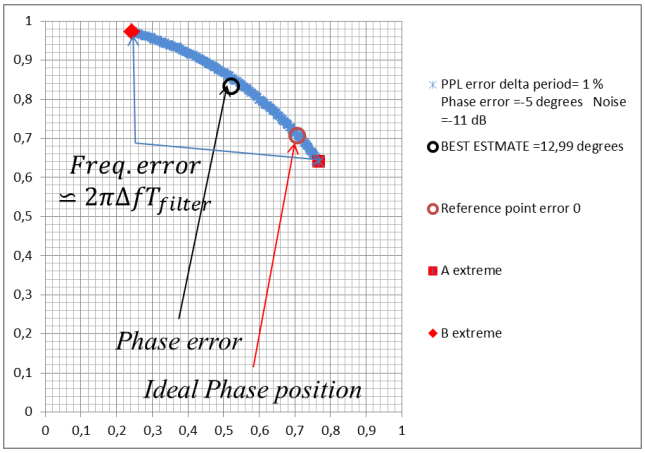


Figure 9.9: The graphics of the output errors of the VCO

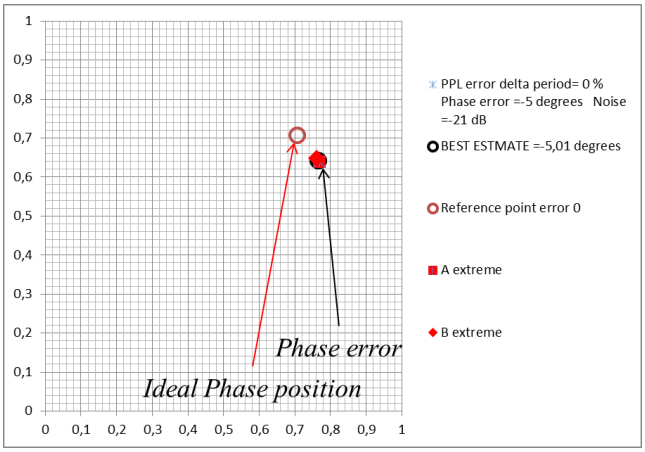


Figure 9.10: PLL pure Phase error

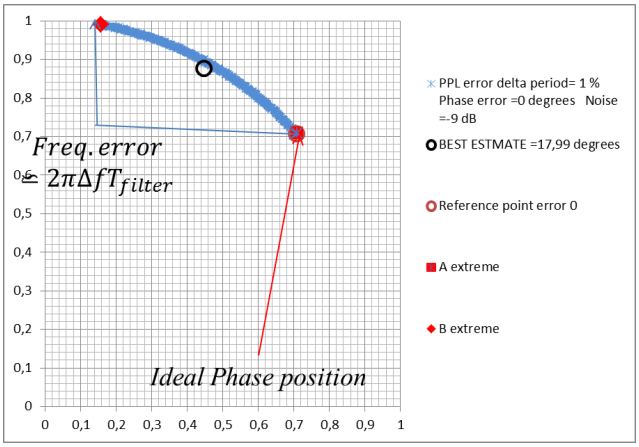


Figure 9.11: PLL Pure Frequency Error

It is evident that the linearization cannot solve the ambiguity depending on the tangent period which is π and not 2π. This kind of ambiguity shall be solved, taking into account additional information on the used modulation.

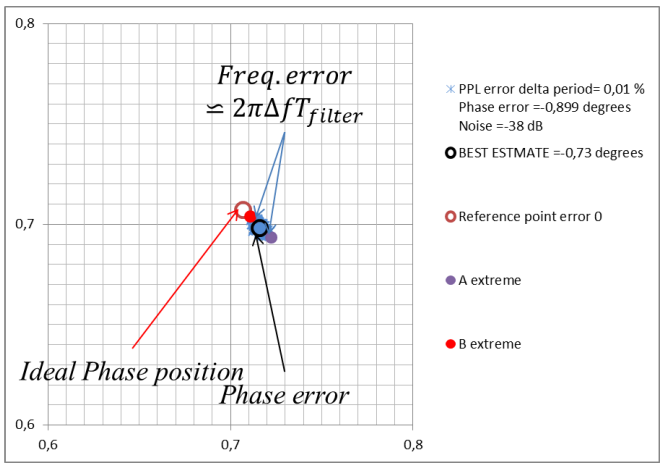


Figure 9.12: PLL error during the locked phase

9.4.2 Phase detection and VCO

The phase detection can be considered linear with phase for small errors: it compares the phase at each input and generates an error signal, $v_e(t)$, proportional to the phase difference between the two inputs. K_D is the gain of the phase detector (V/rad):

$$[9.70]\ v_e(t) = K_D\big(\phi_{out}(t) - \phi_{in}(t)\big)$$

Formerly it has been shown that:

$$[9.71]\ K_d \sin(2\pi\Delta f t + \Delta\phi) \approx K_d(2\pi\Delta f t + \Delta\phi)$$

The averaged transfer characteristic of such a phase detector is shown below. Note that in many implementations, the characteristic may be shifted up in voltage (single supply/single ended).

In PLL applications, the VCO is treated as a linear, time-invariant system. The excess phase of the VCO is the system output. Figure 9.13 shows a typical input output characteristic as a function of ϕ.

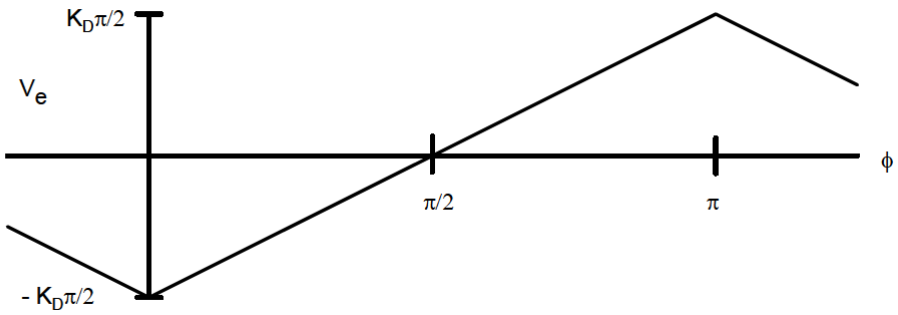


Figure 9.13: Phase detection capture range

Therefore, the mixer approximates for small angular errors a linear detector, which can be modeled, for performance and stability analysis, using the transfer function. If the phase difference is π/2, then the average or integrated output from the XOR-type phase detector will be zero (or v_{dd} for single supply, digital XOR). The slope of the characteristic in either case is K_D. The VCO oscillates at an angular frequency, ω_{out}. Its frequency is set to a nominal ω_0 when the control voltage is zero. Frequency is assumed to be linearly proportional to the control voltage with a gain coefficient K_0 or K_{VCO} (rad/s/v).

$$[9.72]\ \omega_{out} = \omega_0 + K_0 V_d$$

$$[9.73]\ \phi_o = \phi_{out} - \phi_{in} = K_0 \int_0^t v_d(t)dt$$

9.4.3 Linearised phase control loop and dynamic response

Assuming a linear approximation model, the block scheme in Figure 9.14 can represent the PLL.

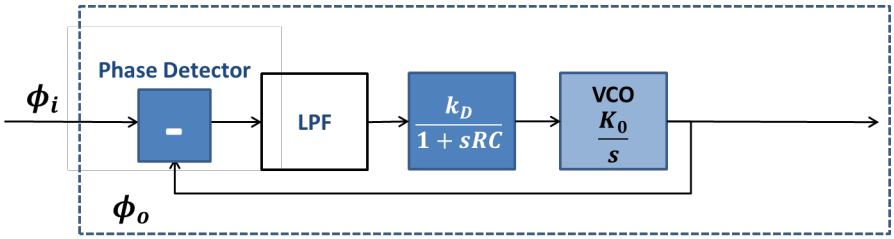


Figure 9.14: PLL linear approximation

$$[9.74]\ \phi_o = \frac{V_1}{K_d} = \frac{\omega_1 - \omega_0}{K_d K_0}$$

Thus, a high loop gain is beneficial for reducing phase errors.

The transfer function from reference phase input to VCO phase output, H(s), can be obtained as

$$[9.75]\ H(s) = \frac{\phi_{out}(s)}{\phi_{in}(s)} = \frac{K_d F(s) K_O/s}{1+K_d F(s) K_O/s} = \frac{K_d F(s) K_O}{s+K_d F(s) K_O}$$

The open loop transfer function, which can be used for verifying the Loop stability is given by:

$$[9.76]\ T(s) = \frac{K_d F(s) K_O}{s}$$

the sensitivity transfer function from the reference phase input to the phase error, S(s), is

$$[9.77]\ S(s) = \frac{\phi_d(s)}{\phi_{in}(s)} = \frac{1}{1+K_d F(s) K_O/s} = \frac{s}{s+K_d F(s) K_v}$$

If the loop filter frequency is lower than the crossover frequency, which you might want to do to attenuate the high frequency ripple from the phase detector, then the phase margin can become unacceptably small. If the loop gain is increased, $K_V = K_D K_O$, to reduce the residual phase error, the result is a smaller phase margin. Thus, there is a conflict between stability of the loop and minimizing the phase error. However, the loop can work if $\omega_1 > \omega_{crossover}$, but then it may result an insufficient filtering of the phase detector output.
Before this problem is fixed, let's look at the root locus and then the closed loop response of this PLL.

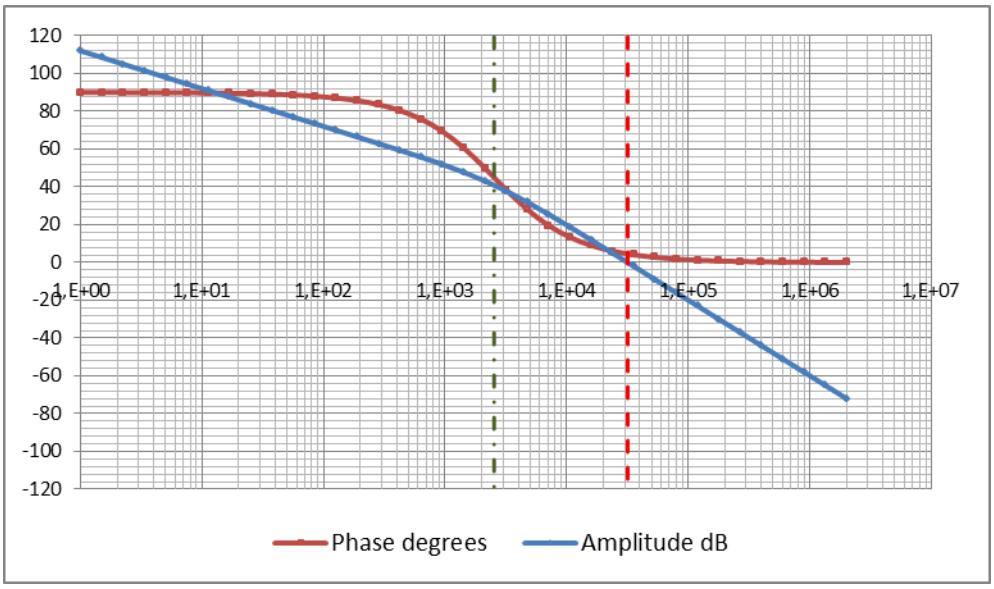


Figure 9.15: Bode characterization of the open loop of PLL, with $RC=3.96\text{x}10^{-4}$, $K_dK_0=400000$, $\omega_1=2525$ Hz

To evaluate the stability and the performance of the loop it is usually considered the dumping factor defined as:

$$[9.78]\ \zeta = \frac{1}{2}\sqrt{\frac{\omega_1}{K_0}} = 0.04$$

This form allows to use standard equations and normalized plots to describe the frequency and transient response of the system. A large K_o, which is useful for reducing phase error, leads to a small ζ, which is bad for stability and settling time.

10 TRANSMISSION MEANS

10.1 INTRODUCTION TO THE TRANSMISSION MEANS

The fundamental element of the transmission is the transmission means, within which the electromagnetic signal propagates for the distance r of the link with a speed usually not very different from the characteristic one of an e.m. (e.m.=electromagnetic) wave in the free space ($c=3\cdot10^8$ m/s); such a finite speed, although very large, is the main responsible of the delay in reception $t_0(r)$, growing with r. The transmission means are divided in two cathegories: **transmission lines**, or **physical carriers**, physical systems with uniform distribution along a longitudunal coordinate, which guiding the signal propagation up to its destibìnation cover physically the distance r of the link and therefore are also named ***physical carriers***; the ***radio links***, in which the signal radiated by a transmitting antenna freely propagates in the form of e.m. waves until it is captured by a receiving antenna at distance r, therefore also named ***radio carriers***. The physical carrier means have a structure constituted of two conductors isolated with dielectric or composed of just dielectrics: respectively **metallic pairs** and **optical fibers**.
Usually, the transmission means have passive nature and their behaviour is linear; in the ideal equivalent scheme, in which also the characteristic of time invariance is considered, in addition to the mentioned delay the attenuation of the carrier means is significant, function which grows with the distance r:

$$[10.1]\ A_0(r)=\frac{W_{xx}}{W_{yy}},$$

where W_{xx} and W_{yy} indicate the powers of the signals x(t) and y(t), respectively at the input and at the output of the means itself. If the materials of the physical carriers were ideal, thanks to the uniformity of the line the configuration of the guided e.m. wave would be the same in every section, but with a propagation delay; instead, the attenuation of the signal which characterizes the real means is then due to the ohmic losses in the materials. In the generic radio carrier means, the radiated e.m. wave distributes its e.m. power over a spherical surface with area ever increasing with the distance, so that the attenuation of the e.m. field in a far point can be as a first approximation assessed taking into account just the mentioned effect, linked to the free propagation phenomenon, implying to neglect the physical characteristics of the interested real space.
The transmisiosn means is very important with respect to the offered service by the whole link; sometimes the choice of the type is obliged, as for example in the case of transmission with mobile terminals, always performed via radio, but in most cases derives from economic considerations in a non ideal context.

10.1.1 Metallic pairs

Whenever the distances to cover between two fixed terminals are small, often a physical carrier means made of a pair of metallic conductors is utilized, isolated between each other by dielectric material, placed along track of the path established between the terminals. Usually the transversal section is constant, both in terms of geometry and materials: such a type of structure is named metallic pair uniform transmission line or simply ***metallic pair***. The propagation of the signal occurs in the dielectric, with plane e.m. wave, having both the electric field and the magnetic one always placed on the transversal plane, i.e. with TEM type mode (TEM=Transverse Electric and Magnetic); thus, the metal performs the guide function along the track of the carrier means.
The most common types of metallic lines are the ***symmetrical pair*** (see Figura 10.1a), usually named also ***two wire line***, and the ***coaxial pair*** (see Figura 10.1b). In the former case an open structure is obtained, with an e.m. field that is extended in theory in all the space; instead, the latter case belongs to the cathegory of the closed structures, in which the e.m. field is assumed practically present inside a hollow conductor, with a shield effect also with respect to other possible e.m. fields generated outside the pair by natural or man made phenomena not correlated

to the desired transmission. Aiming to the benefits deriving from the shielding, a ***shielded pair*** is used (see Figura 10.1c), one more example of closed structure.

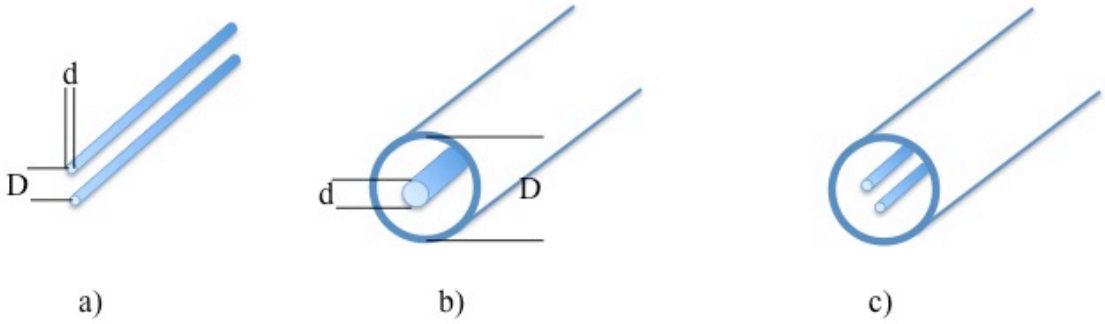


Figura 10.1: Metallic pairs, symmetric open type (a), coaxial (b) and symmetric closed (c).

The structural data of the pair are the same of the materials: the dielectric constant ε (or the relative one $\varepsilon_r=\varepsilon/\varepsilon_0$) and the magnetic permeability $\mu=\mu_0$ of the isolator [(1)] and the conductability σ and the magnetic permeability μ_m of the conductor, and those concerning the geomtery of the section such as the common diameter d and the distance between axes D of the wires in the symmetrical pair or the external diameter d of the internal conductor, the internal diameter D of the external conductor and its thickness s in the coaxial pair.
To guarantee uniformity, i.e. the invariance of the section of a two wire section, the conductors are sometimes plunged in a isolating plastic material (for example polyethylene, with $\varepsilon_r=2,26$). More usually, each metallic wire is wrapped by a plastic sheath with constant thickness and the constant distance between the axes of the elements of the pair is obtained interlacing and containing them in a further sheath; The spiral wrapping performs also the important function to protect against disturbs, obtained by compensation of the effects induced on the two conductors which at every half pass invert their relative position with respect to the disturbing e.m. fields. A typical value of the wire diameter is d=0.6 mm.
The centering of the internal conductor of a coaxial pair can be obtained as shown in Figura 10.2, with small dielectric disks (a), with an isolating plastic material wrapped as a spiral (b) or full so that the space between the two conductors is fully filled (c); assuming that the pitch of the disks or of the spiral is sufficiently short, the dielectric can be assumed homogeneous, with dielectric constant in between of the air and of the isolating material ones: thus, the value of ε_r spans from 1.1 up to 2.26, in the case of full polyethylene.

Figura 10.2: Longitudinal section of coaxial pair with small disks (a) and with spiral (b).

To the structural solutions considered in Figura 10.2 correspond, in the same order, transmission behaviours of lower quality, but as counterpart greater mechanical flexibility of the structure is acquired, especially in the case of full dielectric the conductor is realized like a copper sock (wires with very small diameter interweaved according to enwrapping spirals). In the case of isolation with small disks the typical values d=2,6 mm and D=9,5 mm are obtained. In the coaxial structures placed in operation, often putting together many pairs, adequate sheaths to protect the whole set are utilized, realizing a ***coaxial cable***.
The metallic pairs are used almost exclusively over short distances, typically shorter than 1 km.

10.1.2 Optical fibers

In the transmission between two fixed terminals at great distance, but in some case even with short distances, it is very frequent to use a physical carrier means based on a fully isolating structure, so

(1) The values in the free space are $\varepsilon_0=(36\pi)^{-1}10^{-9}$ [F/m] and $\mu_0=4\pi10^{-7}$ [H/m].

that there are no losses due to conductor materials. In the typical solution shown in Figura 10.3, named step index, the structure is constituted of two dielectric materials separated by a single cylindric surface, with circular section with radius a, which delimits the internal ***core*** from ***cladding*** which surrounds it. The first material has dielectric relative constant $\varepsilon_{r1}=\varepsilon_1/\varepsilon_0$ and rifraction index $n_1 = \sqrt{\varepsilon_{r1}}$, while the second has $\varepsilon_{r2}=\varepsilon_2/\varepsilon_0$ and $n_2 = \sqrt{\varepsilon_{r2}}$ with value slightly lower than n_1; both, usually made of silica glass opportunely but differently doped, have rifraction index of value around 1.4 and magnetic permeability $\mu_1=\mu_2=\mu_0$, equal to the free space one.

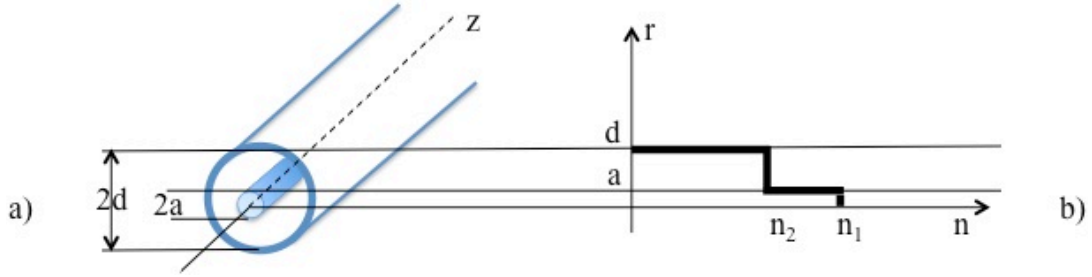


Figura 10.3: Optical fiber with section with circular symmetry (a), and theoretical profile of the rifraction index n along the radius (b).

The structure sketched in Figura 10.3, with invariant section in longitudinal sense and finite transversal extension of the cladding, is usually named ***optical fiber***, considering the very small trasversal dimensions (usually 2d=125 μm) and the use of frequencies very close to those of the visible spectrum. The optical fiber is an open structure, with e.m. field extended in theory in all the dielectric space, not homogeneous; nevertheless, in the conditions of utilization the guided e.m. is in practice mainly concentrated in the core region, so that in the cladding the values of the field can be neglected even at distances form the axis lower than the external radius d. In practice the structure is thus efficiently shielded.
An optical fiber can allow the propagation of multiple different configurations of e.m. waves, or guided modes, which being all non T.E.M. type, have each an own cut off frequency below which the propagation of the mode cannot be supported. In most of the applications, always for distances beyond one chilometer, in order to achieve the best performance, just the fundamental mode, which has the lowest cut off frequency, is active; chosen the excitation frequency f or the corresponding wavelenght λ=c/f, the desired **monomodal fiber** is obtained with reduction to the opportune value of the radius a of the core: typically, 2a≅10μm is achieved using wavelengths λ around 1 μm (1.3 or 1.55 μm).
With the same operating frequency, a structure working in multimodal mode, that is a ***multimodal fiber***, has a greater diameter of the core, around 2a≅60 μm, reducing, as counterpart of its worse transmission performance, the coupling problems between the fiber and terminals and among segments of fiber.
In the frame of the production process the optical fiber is sometimes covered by a dielectric sheath, opaque to the used frequencies, which other than ensuring an aven more complete shielding has the main task to allow a better manageability of the structure, otherwise very fragile; the external diameter is in this way doubled. In the structures placed in operation with bundling many fibers, adequate protective sheaths of the whole set are used, constituting an ***optical cable***.
The meaningful use of the optical fibers in the transmission is due to the possibility to get a very small value of the attenuation of the signal per unit lenght travelled in the means, thanks to the absence of conductor materials and to the low losses that the silica glass shows around the wavelenght of 1.5 μm, i.e. for frequencies extremely high (beyond 10^{14} Hz).

10.1.3 Radio path

In the case of free mobility of at least of one of the terminals or of its far location in the space surrounding the Earth, the transmission can be exploited only utilizing the free e.m. propagation. Such a modality is also very convenient in the transmission with broadcast of the information, i.e.

when the same signal must be transferred from a terminal toward several multiple fixed user terminals; it can also be used fruitfully for the transmission between two fixed points in alternative to the physical means, typically in particular orographic conditions or for the advantage offered by the quickness of the installation.
The basic structure of a ***radio path*** includes a pair of ***antennae*** and the free space in between (see Figura 10.4), supposed ideal as first approximation. The antennae are material systems with a port accessible by classical signals: exciting with an input signal x(t) a transmitting antenna, it is capable to radiate a spherical e.m. field which in every point of the free space has instant intensity of the electric and magnetic fields both proportional to the signal, but with a delay due to the distance; a receiving antenna hit by such an e.m. field is capable to capture the radiation, providing an output signal y(t) proportional to the mentioned instant intensities and, thus, faithful to the excitating signal x(t) of the path.

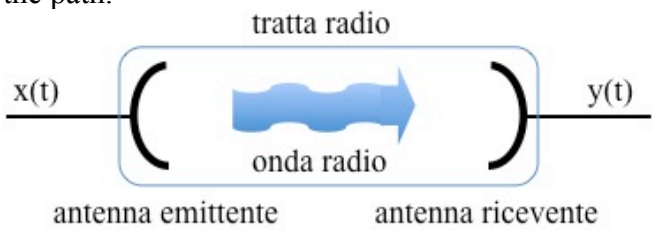


Figura 10.4: Basic structure of a radio path.

Almost always the antennae are essentially made of conductor materials, often with filiform (threadlike) structures, as the dipole antenna, or surface type, as the parabolic antenna. Often, other than passive they are reciprocal, so that the same antenna can operate both in transmission and in reception. There are several types of antennae with very different performance, essentially from the capability to radiate and capture with different effectiveness depending on their orientation with respect to the joining straight line of the pair of the antennae.
Considering that the radio paths use the same space, it is necessary to pay great attention to the interferences, to use such natural common and scarce resources at the best: one of the most used methods is the allocation to each radio carrier of a ***radiofrequency*** f_c, accurately identified within the spectrum of the radio applications extended from a few tens of kHz up to some tens of GHz, at the center of an interval wide B as limited as possible taking into account the application (usually $B \ll f_c$, even for more than one order of magnitude).

10.2 INTRODUZIONE AI MEZZI TRASMISSIVI REALI

Nel paragrafo 10.1 sono state individuate le due fondamentali categorie in cui possono essere distinti i mezzi trasmissivi, ossia le linee di trasmissione e le tratte radio, rispettivamente anche denominati portanti fisici e portanti radio. Si distinguono per il diverso meccanismo tramite il quale il segnale transita lungo il mezzo trasmissivo: si hanno infatti le due possibilità di propagazione e.m. guidata secondo la coordinata longitudinale che segna il percorso o di propagazione e.m. libera nello spazio interposto tra una coppia di antenne.
Nel seguito ci si propone di affrontare l'analisi dei mezzi trasmissivi impiegabili nel trasferimento elettromagnetico della informazione, tenendo in conto le imperfezioni che possono manifestarsi nei casi effettivi e le loro conseguenze sui segnali trasmessi. Si dà per scontata la conoscenza delle pertinenti nozioni basiche di elettromagnetismo, anche se non mancano alcuni richiami, giustificati per motivi di notazione e per comodità di disporre in modo immediato delle espressioni che esprimono le definizioni e le proprietà fondamentali. Saranno considerati nell'ordine i mezzi portanti fisici metallici, concentrando l'attenzione sulle strutture costituite da coppie di conduttori isolati da dielettrico (paragrafo 10.3), i mezzi portanti fisici ottici, costituiti da strutture composte esclusivamente di dielettrici (paragrafo 10.4), e infine le tratte radio (paragrafo 10.5).
Alcune precisazioni e dimostrazioni sono raccolte in una appendice (appendice A5), articolata in paragrafi.

10.3 MEZZI TRASMISSIVI METALLICI

10.3.1 Struttura delle coppie metalliche

Quando le distanze da coprire tra due terminali fissi siano modeste e i segnali abbiano banda utile non molto estesa, si ricorre spesso all'impiego di un mezzo portante fisico costituito da una coppia di conduttori metallici, isolati tra loro, posati lungo il tracciato del percorso stabilito tra i terminali. Di norma la sezione è invariata, sia nella geometria che nei materiali: un tale tipo di struttura è denominata linea di trasmissione uniforme a coppia metallica o più semplicemente *coppia metallica*. Nel seguito si presuppone che la struttura sia cilindrica lungo lo sviluppo longitudinale, in modo da potere adottare una coordinata rettilinea z; anche se il tracciato effettivo è spesso curvilineo, il suo raggio di curvatura è infatti sufficientemente grande da potere assumere che il comportamento e.m. della linea sia equivalente a quello che si avrebbe qualora essa fosse posata secondo una unica direzione.

Le versioni più comuni di linee metalliche sono la *coppia simmetrica* (vedi Figura 10.5a), comunemente denominata anche *linea bifilare* o *doppino*, e la *coppia coassiale* (vedi Figura 10.5b). Nel primo caso si ha un esempio di struttura aperta, con un campo e.m. (e.m. = elettromagnetico) che si estende in teoria in tutto lo spazio; il secondo caso appartiene invece alla categoria delle strutture chiuse, in cui il campo e.m. si assume confinato all'interno di un conduttore cavo, con effetto di schermo anche nei riguardi di eventuali altri campi e.m. generati all'esterno della coppia da fenomeni naturali o artificiali non correlati alla trasmissione desiderata; mirando ai benefici derivanti dalla schermatura, si impiega anche il *doppino schermato* (vedi Figura 10.5c), altro esempio di struttura chiusa.

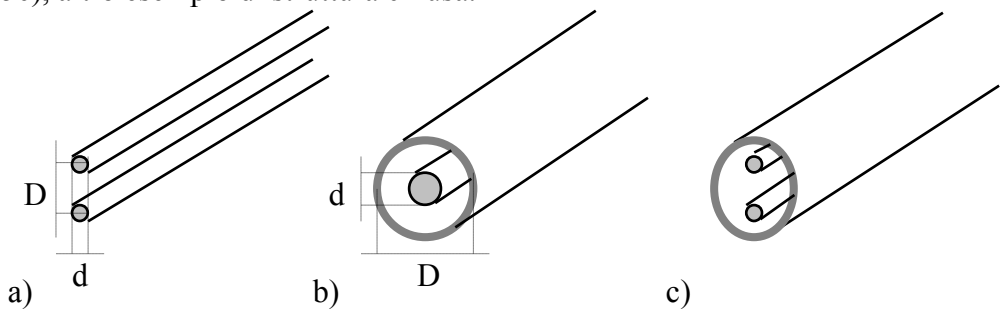


Figura 10.5: Coppie metalliche, di tipo simmetrico aperto (a), coassiale (b) e simmetrico chiuso (c)

I dati strutturali della coppia sono quelli dei materiali, ossia la costante dielettrica ε (o quella relativa $\varepsilon_r \triangleq \varepsilon/\varepsilon_0$) e la permeabilità magnetica $\mu = \mu_0$ dell'isolante[1] e la conducibilità σ e la permeabilità magnetica μ_m del conduttore, e quelli relativi alla geometria della sezione: il diametro comune d e l'interasse D dei fili nella coppia simmetrica oppure il diametro esterno d del conduttore interno, il diametro interno D del conduttore esterno e lo spessore s di quest'ultimo nella coppia coassiale.

Per garantire la uniformità, ossia la invarianza della sezione di una coppia bifilare, i conduttori sono a volte immersi in un materiale plastico isolante (ad esempio polietilene, con $\varepsilon_r = 2{,}26$). Più comunemente, sono circondati entrambi da una guaina a spessore costante di tale materiale e quindi intrecciati; il loro avvolgimento a spirale ha anche l'importante funzione di protezione nei riguardi dei disturbi, ottenuta per compensazione degli effetti indotti sui due conduttori che ad ogni semipasso invertono la loro posizione relativa ai campi e.m. disturbanti.

Il centraggio del conduttore interno di una coppia coassiale può essere ottenuto come mostrato in Figura 10.6, con dischetti di dielettrico (a), con un materiale plastico isolante avvolto a spirale (b) oppure pieno in modo da colmare interamente lo spazio tra i due conduttori (c); ammesso che il passo dei dischetti o della spirale sia sufficientemente breve, si può assumere che il dielettrico sia omogeneo, con costante dielettrica di valore intermedio tra quello dell'aria e del materiale isolante: si passa così dal un valore di ε_r attorno a 1,1 al valore di 2,26, nel caso del polietilene pieno.

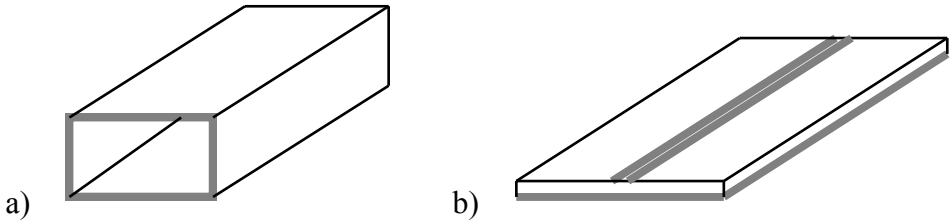
---

[1] I valori nel vuoto sono $\varepsilon_0 = (36\pi)^{-1}\, 10^{-9}$ [F/m] e $\mu_0 = 4\pi\, 10^{-7}$ [H/m].

a) b) c)

Figura 10.6: Sezione longitudinale di coppia coassiale con dischetti (a) e con spirale (b).

Alle soluzioni considerate corrispondono, nell'ordine, comportamenti elettromagnetici di minore qualità, ma in contropartita si acquisisce maggiore flessibilità meccanica della struttura, soprattutto se nel caso di dielettrico pieno il conduttore esterno viene realizzato in forma di calza di rame (fili di assai piccolo diametro intessuti secondo spirali controavvolgentesi).
Solo per brevi lunghezze (di norma minori di 10 m) e nel caso di trasmissione di segnali in banda traslata ad altissima frequenza (di norma dell'ordine di grandezza di 1 GHz = 10^9 Hz), si fa uso di *guide d'onda metalliche*, strutture chiuse con conduttore unico cavo, prevalentemente del tipo a sezione rettangolare come mostrato in Figura 10.7a; altre strutture guidanti di tipo aperto caratterizzate da costanti di attenuazione molto maggiori di quelle delle guide d'onda, come ad esempio la *microstriscia* (vedi Figura 10.7b), trovano impiego per lunghezze assai più brevi (di norma minori di10 cm). Per la scarsa o nulla rilevanza che tali strutture hanno nella funzione di mezzo trasmissivo (ben diverso è il loro ruolo come elementi di connessione nell'ambito degli impianti, degli apparati e addirittura dei circuiti integrati) nel seguito non verranno più prese in considerazione.



a) b)

Figura 10.7: Rectangular metallic waveguide (a) and microstrip (b) .

10.3.2 Grandezze caratteristiche delle coppie metalliche

10.3.2.1 Impedenza caratteristica e costante di propagazione

Le grandezze fondamentali atte a qualificare un tronco di linea di trasmissione a coppia metallica reale di lunghezza generica L possono essere desunte dall'esame del suo comportamento agli estremi. Assunto il comportamento perfetto, il mezzo considerato è rappresentabile tramite un quadripolo LTI simmetrico, passivo, adattato, caratterizzato quindi entro la banda utile del segnale trasmesso dalla funzione di trasferimento (vedi [5.95] e [5.96]):

$$[10.2]\ H(f, L) \equiv H_{21}(f, L) \equiv H_{12}(f, L) = e^{-\alpha_0 L}\, e^{-j 2\pi f L/v},\ \text{per } f_m \le |f| \le f_M;$$

il modulo della funzione risulta dipendere unicamente dalla lunghezza del tronco tramite la legge esponenziale, dove il fattore α_0 è la costante di attenuazione del mezzo perfetto, mentre l'argomento dipende invece proporzionalmente sia dalla lunghezza che dalla frequenza. La menzionata specificità è dovuta al meccanismo del trasferimento e.m. (e.m. = elettromagnetico) del segnale lungo il mezzo trasmissivo, che si ammette avvenga per *propagazione guidata* del solo modo fondamentale del tipo TEM (vedi appendice A5.1) secondo la coordinata longitudinale che segna il percorso.
Nel caso di una coppia metallica reale, una volta accertata la propagazione unimodale, il tronco di linea è ancora rappresentabile con un quadripolo LTI, simmetrico e passivo, ma deve essere caratterizzato, per ogni frequenza almeno in banda utile e con riferimento a una resistenza R reale e costante, tramite una trasmettenza identica nei due versi, $H_c(f, L) \equiv H_{21}(f, L) \equiv H_{12}(f, L)$, e una riflettenza comune alle due porte, $\rho_c(f, L) \equiv H_{11}(f, L) \equiv H_{22}(f, L)$, in generale non più nulla.
Facendo tendere L all'infinito, per effetto delle perdite $H_c(f, L)$ tende a zero e il coefficiente di riflessione in entrata viene a coincidere con la riflettenza $\rho_c(f, \infty)$, indipendentemente dal

particolare coefficiente di riflessione del carico in uscita; in tali condizioni si ricava dalla [5.28] la impedenza della linea infinitamente lunga, denominata *impedenza caratteristica* in quanto tipica della sola struttura:

$$[10.3]\ Z_c(f) = R\,\frac{1+\rho_c(f,\infty)}{1-\rho_c(f,\infty)}\ .$$

Assunto che il tronco di linea sia adattato ($\rho_c(f,\infty)\equiv 0$, ossia $Z_c(f)\equiv R$) e posta la sua funzione di trasferimento sotto la forma generica di esponenziale complesso, $H_c(f, L) \equiv e^{-\chi(f,L)}$, imponendo che per ogni frequenza il comportamento di due tronchi in cascata di lunghezza z e L-z, con z positivo minore di L, sia identico a quella dell'intero tronco, si ricava che nell'esponente si ha la separazione delle variabili f e L:

$$[10.4]\ H_c(f, L) = e^{-\gamma_c(f)L}\ ,$$

dove la grandezza complessa:

$$[10.5]\ \gamma_c(f) = -\,\frac{1}{L}\ln[H_c(f, L)],$$

indipendente dalla lunghezza L, è la *costante di propagazione* della coppia.
L'attenuazione di inserzione del tronco di linea adattata risulta dalla *costante di attenuazione* della coppia, $\alpha_c(f) \triangleq \Re\{\gamma_c(f)\}$, tramite la relazione:

$$[10.6]\ A_c(f, L) = |H_c(f, L)|^{-2} = e^{2\alpha_c(f)L}\ .$$

Nota la *costante di fase* della coppia, $\beta_c(f) \triangleq \Im\{\gamma_c(f)\}$, si ottengono poi il tempo di ritardo:

$$[10.7]\ t_{rc}(f, L) \triangleq -\,\frac{\arg\{H_c(f, L)\}}{2\pi f} = \frac{\beta_c(f)}{2\pi f}\,L,$$

e il tempo di ritardo di gruppo:

$$[10.8]\ t_c(f, L) \triangleq -\,\frac{1}{2\pi}\frac{d}{df}\left[\arg\{H_c(f, L)\}\right] = \frac{1}{2\pi}\,\frac{d\beta_c(f)}{df}\,L.$$

Si noti che i due ritardi sono entrambi proporzionali alla lunghezza L; inoltre essi coincidono se e solo se la costante di fase è proporzionale alla frequenza; allora risultano indipendenti dalla frequenza. Una struttura per cui non è soddisfatta tale condizione viene qualificata come un *mezzo trasmissivo dispersivo*.

10.3.2.2 Velocità di fase e velocità di gruppo

Dai rapporti tra la lunghezza del tronco e i tempi di ritardo dati dalle [10.7] e [10.8] risultano altre due grandezze indipendenti dalla lunghezza medesima e perciò tipiche della struttura: la *velocità di fase*:

$$[10.9]\ v\ (f) \triangleq \frac{L}{\tau_{rc}(f,L)} = \frac{2\pi f}{\beta_c(f)}\ ,$$

e la *velocità di gruppo*:

$$[10.10]\ v_c(f) \triangleq \frac{L}{\tau_c(f, L)} = 2\pi\left[\frac{d\beta_c(f)}{df}\right]^{-1}.$$

La velocità di gruppo nella coppia soddisfa sempre la condizione:

$$[10.11]\ v_c(f) \le c = \frac{1}{\sqrt{\mu_0\varepsilon_0}} \cong 3\cdot 10^8\ \text{m/s}\ ,$$

dove con c è indicata la velocità della propagazione e.m. nel vuoto.
Le due velocità ovviamente coincidono e sono indipendenti dalla frequenza nel caso di costante di fase proporzionale alla frequenza, ossia di mezzo trasmissivo non dispersivo; il comune valore costante si indica con v_c.

10.3.2.3 Lunghezza d'onda e lunghezza elettrica
La grandezza, ancora tipica della struttura, definita dal rapporto:

$$[10.12]\ \lambda_c(f) \triangleq \frac{2\pi}{\beta_c(f)},$$

prende il nome di *lunghezza d'onda* nella coppia. In regime armonico e a tempo costante essa misura di quanto occorre spostarsi lungo l'asse del mezzo trasmissivo affinché la fase dell'onda che si propaga abbia compiuto un angolo giro. Nel caso di mezzo trasmissivo non dispersivo, per il quale la costante di fase è proporzionale alla frequenza e si ha una unica velocità costante v_c, la lunghezza d'onda è inversamente proporzionale alla frequenza: $\lambda_c(f) = v_c/f$.
Con riferimento a un tronco di linea di lunghezza L, sono opportuni alcuni commenti sul prodotto:

$$[10.13]\ \beta_c(f)\, L = 2\pi\, \frac{L}{\lambda_c(f)} = 2\pi f\, \frac{L}{v_\beta(f)}\ ,$$

che si misura in radianti e viene denominato *lunghezza elettrica* del mezzo trasmissivo.
Nell'impiego di coppie metalliche il valore assoluto della lunghezza elettrica è più significativo di quello della lunghezza fisica. Se ad esempio si verifica $L \ll \lambda_c(f)$, tanto da comportare lunghezza elettrica molto minore di un radiante, il tronco di linea può essere considerato più alla stregua di una semplice connessione metallica, che non come un mezzo trasmissivo; tenuto conto che la velocità di fase è spesso molto elevata (non molto minore di $c \cong 3\cdot10^8$ m/s), quanto sopra può accadere anche per lunghezze fisiche dell'ordine del centinaio di metri, se le frequenze dei segnali da trasmettere sono dell'ordine del kHz. Si ha dunque a che fare davvero con un tronco di linea di trasmissione quando la lunghezza elettrica non risulti molto minore di un radiante; il tronco sarà poi elettricamente corto quando L sia dell'ordine di $\lambda_c(f)$ ed elettricamente lungo per $L \gg \lambda_c(f)$; per frequenze molto elevate (ordine del GHz) anche un tratto di un metro di coppia è dunque da considerare come un tronco di linea di trasmissione.

10.3.2.4 Parametri primari
Si usa spesso caratterizzare una coppia metallica tramite quattro grandezze fisiche reali positive, denominate *parametri primari* della linea; essi sono: la resistenza assiale per unità di lunghezza, *r* [Ω/m], la induttanza assiale per unità di lunghezza, *l* [H/m], la conduttanza di dispersione trasversale per unità di lunghezza, *g* [Ω^{-1}/m], e la capacità trasversale per unità di lunghezza, *c* [F/m]. Le grandezze considerate sono legate alle proprietà dei materiali impiegati e alla geometria della sezione; sono perciò di maggiore interesse per il progetto e la produzione della coppia.
Come si vedrà nell'immediato seguito, i valori assunti da *r* risultano dipendere dalla frequenza quando si superino alcuni kHz, mentre i rimanenti parametri primari possono ritenersi costanti, almeno in prima approssimazione.
Nel caso di coppie con conduttori ideali e con riferimento alla propagazione unimodale TEM (vedi appendice A5.1), il vettore $\mathbf{J}_s$ della densità di corrente superficiale (misurata in A/m) è diverso da zero solo sulla superficie di separazione tra conduttore e dielettrico ed è diretto assialmente. Passando al caso reale monomodale con metalli a conducibilità σ elevata, ma finita, quanto meno è alta la frequenza, tanto più le correnti hanno modo di scorrere anche all'interno dei conduttori: si ha allora, come grandezza fisica, il vettore $\mathbf{J}_v$ della densità di corrente volumetrica (misurata in A/m²), orientato come $\mathbf{J}_s$ del corrispondente caso ideale e con intensità J_v che diminuisce con legge esponenziale reale allontanandosi dalla superficie, verso l'interno dei conduttori, come mostrato nell'esempio in Figura 10.8.

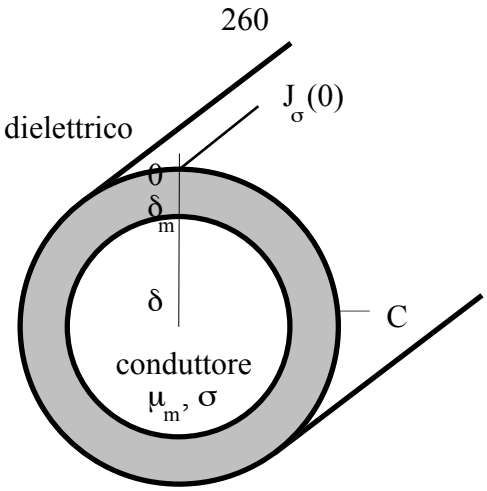


Figura 10.8: Distribuzione esponenziale della corrente volumetrica nel conduttore.

Il fenomeno dell'addensamento, sempre crescente all'aumentare della frequenza, delle correnti di conduzione verso le superfici che separano i conduttori dai dielettrici entro cui avviene la propagazione si denomina *effetto pellicolare*. Il decadimento della intensità della densità di corrente volumetrica avviene in modo che essa si riduce di un fattore $1/e \cong 0{,}37$ a una distanza dalla superficie, che prende il nome di *spessore di penetrazione*, data dalla:

$$[10.14]\ \delta_m \triangleq \frac{1}{\sqrt{\pi \mu_m \sigma f}}\ ,$$

dove f è la frequenza e σ e μ_m sono la conducibilità e la permeabilità magnetica del conduttore.
Nel caso più comune di conduttori in rame si hanno i valori $\sigma = 0{,}58\ 10^8$ [Ω^{-1}/m] e $\mu_m = \mu_0 = 4\ \pi\ 10^{-7}$ [H/m]; dalla [10.14] si ottiene allora: δ_m (rame) = $0{,}066/\sqrt{f}$ e, quindi, i seguenti valori mostrati in Tabella 10.1:

Tabella 10.1: Parametri primari di una coppia metallica

f [kHz]	1	3	6	10	30	60	100	300	600	1000
δ_m[mm]	2,09	1,21	0,85	0,66	0,38	0,27	0,21	0,12	0,09	0,07

La tabella mostra che alle basse frequenze, ad esempio al di sotto di 10 kHz, la penetrazione della corrente nei conduttori è tale che, per le usuali dimensioni nelle applicazioni di trasmissione (ingombri delle sezioni dei conduttori di norma inferiori al millimetro) si può ammettere la ipotesi di distribuzione uniforme della densità di corrente volumetrica sulla sezione, anche per materiali ad alta conducibilità come il rame. Tale ipotesi non è invece affatto valida alle alte frequenze: δ_m si riduce addirittura a 1 μm a 4,4 GHz! Comunque, già a 60 kHz è sufficiente penetrare di appena 0,27 mm all'interno di un conduttore di rame per trovare una densità di corrente pari a 0,37 volte quella in superficie (la riduzione è a un decimo per una penetrazione di 0,62 mm).
Alle basse frequenze, per cui di norma è trascurabile l'effetto pellicolare, la resistenza assiale unitaria di un singolo conduttore reale con sezione di area A e contorno di perimetro C risulta dalla nota relazione, ricavabile per $f \to 0$ (in corrente continua):

$$[10.15]\ r_{b1} = \frac{1}{\sigma A}\ .$$

Per frequenze sufficientemente alte (per cui è valida la $C\delta_m << A$), associando all'effetto pellicolare una sezione ridotta (vedi appendice A5.2) avente un'area equivalente in continua data dal prodotto del perimetro C per lo spessore δ_m, si ha invece:

$$[10.16]\ r_{a1}(f) = \frac{1}{\sigma C \delta_m(f)} = \frac{1}{C}\sqrt{\frac{\pi \mu_m}{\sigma}}\sqrt{f}\ ,$$

e tenuto conto della [10.15]:

$$[10.17]\ r_{a1}(f) = r_{b1}\frac{A}{C\delta_m(f)} = r_{b1}\frac{A}{C}\sqrt{\pi\mu_m\sigma}\sqrt{f} \gg r_{b1}.$$

In contropartita dell'incremento della resistenza assiale unitaria, sempre alle alte frequenze, l'effetto pellicolare consente una efficace schermatura delle strutture chiuse, giacché anche piccoli spessori del conduttore esterno possono bastare a ritenerle impenetrabili da parte di eventuali campi e.m. generati al di fuori della linea. In particolare la struttura della coppia coassiale, così come quella del doppino schermato, è proprio motivata dalla possibilità di racchiudere il segnale utile in un volume che risulta ben protetto dalle interferenze, sempre che lo spessore del metallo sia sufficientemente maggiore di δ_m. Si noti in proposito dalla [10.14] che la schermatura alle basse frequenze, altrimenti poco efficace, può essere migliorata ricorrendo a materiali conduttori ad alta permeabilità magnetica avvolti in spessore sottile sul conduttore di rame esterno.

10.3.3 Comportamento delle coppie metalliche

10.3.3.1 Condizioni di Heaviside

Come è stato evidenziato in precedenza, dal punto di vista della utilizzazione sono rilevanti la impedenza caratteristica, $Z_c(f)$, e la costante di propagazione, $\gamma_c(f)$, grandezze in generale complesse e dipendenti dalla frequenza, che vengono anche denominate *parametri secondari* della coppia. Volendo perseguire l'obiettivo di una trasmissione perfetta occorre dunque mirare ad ottenere che nella banda utile del segnale la linea reale abbia:

- impedenza caratteristica reale e costante, $Z_c(f) = R$,
- costante di attenuazione indipendente dalla frequenza, $\alpha_c(f) = cost$,
- costante di fase proporzionale alla frequenza, $\beta_c(f) = 2\pi f/v_c$, con $v_c = cost$.

Tra i parametri primari e quelli secondari valgono le note relazioni (vedi appendice A5.3):

$$[10.18]\ Z_c(f) = \sqrt{\frac{r + j\omega l}{g + j\omega c}},$$

$$[10.19]\ \gamma_c(f) = \alpha_c(f) + j\beta_c(f) = \sqrt{(r + j\omega l)(g + j\omega c)}.$$

A patto che la dipendenza dalla frequenza della resistenza assiale unitaria possa essere trascurata, ponendo $r(f) \cong r(0) = r_b$, in base alla osservazione della [10.18] appare evidente la possibilità di conseguire l'obiettivo di una impedenza caratteristica reale e costante. Imponendo che si abbia $Z_c = R$, ossia che il numeratore e il denominatore del radicando nella [10.18] soddisfino la relazione:

$$[10.20]\ r_b + j\omega l = R^2(g + j\omega c),$$

si ottengono le *condizioni di Heaviside* sui parametri primari:

$$[10.21]\ \frac{r_b}{g} = \frac{l}{c} = R^2 = Z_c^2.$$

Servendosi della [10.20] nella [10.19] si ricava immediatamente che una coppia per cui siano soddisfatte le condizioni di Heaviside, oltre ad avere impedenza caratteristica reale e costante, ha costanti di attenuazione e di fase date dalle:

$$[10.22]\ \alpha_H = \frac{r_b}{R} = \sqrt{r_b g},\ \beta_H(f) = \frac{j\omega l}{R} = 2\pi f\sqrt{lc}.$$

Si può dunque asserire che un tronco di lunghezza L di una coppia che soddisfi le condizioni di Heaviside e abbia parametri primari tutti indipendenti dalla frequenza in una stretta banda assegnata mostra un comportamento perfetto, in quanto le riflettenze sono nulle e la funzione di trasferimento risulta (vedi [10.4]):

$$[10.23]\ H_H(f) = e^{-\gamma_H(f)L} = e^{-\sqrt{r_b g}\,L}\, e^{-j2\pi f\sqrt{lc}\,L}\ .$$

Le considerazioni svolte ebbero in passato grande rilevanza nelle applicazioni della trasmissione telegrafica e telefonica a media e grande distanza su coppie simmetriche, con banda contenuta al di sotto di 4 kHz e conseguente ammissibilità pratica della ipotesi di indipendenza di *r* dalla frequenza. Il rispetto delle condizioni di Heaviside comportava comunque interventi di modifica sulle coppie simmetriche normalmente prodotte e installate a costi ottimizzati (1), con conseguente non trascurabile aumento del costo per unità di lunghezza. Con l'abbandono dell'impiego delle coppie metalliche per la trasmissione di segnali a banda stretta su distanze oltre qualche chilometro, le condizioni di Heaviside hanno perso rilevanza applicativa; infatti per tronchi di linea di lunghezza contenuta il comportamento imperfetto introduce effetti indesiderati, proporzionali a L, che risultano tollerabili.

10.3.3.2 Modelli asintotici a bassa e alta frequenza

Di norma è possibile scegliere, senza particolari aggravi di costo, dei materiali isolanti di qualità tale da rendere trascurabile la conduttanza di dispersione trasversale unitaria; le [10.18] e [10.19] assumono pertanto le forme approssimate:

$$[10.24]\ Z_c(f) \cong \sqrt{\frac{r + j\omega l}{j\omega c}}\ ,$$

$$[10.25]\ \gamma_c(f) \cong \sqrt{j\omega c\left(r + j\omega l\right)}\ .$$

Al tendere a zero della frequenza, si può assumere $r(f) \cong r(0) = r_b >> \omega l$, dove tenuto conto che i due conduttori della coppia, di sezione A_1 e A_2, sono percorsi in serie dalle correnti volumetriche si ha dalla [10.15]:

$$[10.26]\ r_b = r_{b1} + r_{b2} = \frac{1}{\sigma}\frac{A_1 + A_2}{A_1 A_2}\ .$$

Trascurando allora $j\omega l$ rispetto a *r* nelle [10.24] e [10.25], si perviene alle seguenti espressioni del modello asintotico a bassa frequenza:

$$[10.27]\ Z_b(f) \triangleq \sqrt{\frac{r_b}{j\omega c}} = (1 - j)\sqrt{\frac{r_b}{4\pi c}}\ \frac{1}{\sqrt{f}},$$

$$[10.28]\ \gamma_b(f)\ \triangleq \sqrt{j\omega c r_b} = (1 + j)\sqrt{\pi f c r_b}\ ,$$

e quindi

$$[10.29]\ \alpha_b(f) = \beta_b(f) = \sqrt{\pi c r_b}\ \sqrt{f}\ .$$

Le [10.27] e [10.28] sono perciò ben lungi dal soddisfare le condizioni di trasmissione perfetta.
Se invece si fa tendere la frequenza verso l'infinito, ammesso che risulti $r(f) \ll \omega l$ si può trascurare *r* rispetto a $j\omega l$ nella [10.24] e assumere nella [10.25] l'approssimazione:

$$[10.30]\ \sqrt{1+\frac{r(f)}{j\omega l}}\ \cong 1 - j\,\frac{r(f)}{2\omega l}.$$

Si giunge allora alle seguenti espressioni del modello asintotico ad alta frequenza:

$$[10.31]\ Z_a \triangleq \sqrt{\frac{l}{c}} = R,$$

---

(1) Di norma si ottiene $r/g >> l/c$, stante i valori molto piccoli assunti da *g*; per soddisfare le [10.20] [5.2.19] si ricorreva all'aumento di *l* ottenuto avvolgendo un filo ferromagnetico sui conduttori di rame (metodo Krarup) o inserendo in serie delle bobine a distanza periodica molto minore della lunghezza d'onda (metodo Pupin).

$$[10.32]\ \gamma_a(f) \triangleq \sqrt{-\omega^2 l c}\left[1-j\frac{r(f)}{2\omega l}\right],$$

e quindi

$$[10.33]\ \alpha_a(f) = \frac{1}{2}\frac{r(f)}{R}\ ,\ \beta_a(f) = 2\pi f\sqrt{l c}\,.$$

Si noti che tanto la impedenza caratteristica Z_a, che la costante di fase $\beta_a(f)$, assumono le medesime espressioni che avrebbero nel rispetto delle condizioni di Heaviside, senz'altro soddisfacenti le rispettive condizioni di trasmissione perfetta. Altrettanto non si può invece affermare per la costante di attenuazione, $\alpha_a(f)$. Tenuto conto che i due conduttori della coppia, di sezione A_1 e A_2 e di perimetro C_1 e C_2, sono in serie, dalle [10.16] e [10.17] si ricava infatti:

$$[10.34]\ r(f) \cong r_a(f) = r_{a1}(f) + r_{a2}(f) = \frac{1}{\sigma\delta_m}\left(\frac{1}{C_1}+\frac{1}{C_2}\right) = r_b k_r\sqrt{f}\ ,$$

dove r_b è dato dalla [10.26] e

$$[10.35]\ k_r \triangleq \frac{A_1A_2(C_1+C_2)}{C_1C_2(A_1+A_2)}\sqrt{\pi\mu_m\sigma}\ ,$$

è una costante dipendente dai materiali e dalla geometria della sezione; dalla prima delle [10.33] si ha dunque:

$$[10.36]\ \alpha_a(f) \cong \frac{1}{2}\frac{r_b}{R}k_r\sqrt{f}\,,$$

che mostra una dipendenza proporzionale con la radice della frequenza. Si noti che per le frequenze, non troppo basse, per cui la [10.34] è validamente applicabile, l'effetto pellicolare impone che si abbia $r_a(f) \gg r_b$; di conseguenza risulta:

$$[10.37]\ k_r\sqrt{f} \gg 1.$$

10.3.3.3 Comportamento a banda larga

È interessante analizzare i legami che intercorrono tra i due modelli asintotici, a bassa e alta frequenza. Si hanno le relazioni:

$$[10.38]\ \frac{Z_b(f)}{Z_a} = (1-j)\sqrt{\frac{r_b}{4\pi c}}\frac{1}{\sqrt{f}}\sqrt{\frac{c}{l}} = (1-j)\sqrt{\frac{r_b}{4\pi l}}\frac{1}{\sqrt{f}},$$

$$[10.39]\ \frac{\alpha_b(f)}{\alpha_a(f)} = \sqrt{\pi c r_b}\sqrt{f}\sqrt{\frac{l}{c}}\frac{2}{k_r r_b}\frac{1}{\sqrt{f}} = \frac{1}{k_r}\sqrt{\frac{4\pi l}{r_b}},$$

$$[10.40]\ \frac{\beta_b(f)}{\beta_a(f)} = \sqrt{\pi c r_b}\sqrt{f}\frac{1}{2\pi f\sqrt{l c}} = \sqrt{\frac{r_b}{4\pi l}}\frac{1}{\sqrt{f}}\,.$$

Posto:

$$[10.41]\ f_x \triangleq \frac{r_b}{4\pi l}\,,$$

dalle precedenti si ricavano le più semplici forme:

$$[10.42]\ Z_b(f) = (1-j)\sqrt{\frac{f_x}{f}}Z_a,$$

$$[10.43]\ \alpha_b(f) = \frac{1}{k_r}\frac{1}{\sqrt{f_x}}\,\alpha_a(f),$$

$$[10.44]\ \beta_b(f) = \sqrt{\frac{f_x}{f}}\,\beta_a(f).$$

Dalla [10.44] si deduce che per $f = f_x$ si ha il comune valore $\beta_b(f_x) = \beta_a(f_x) = X$, per cui la grandezza f_x definita tramite la [10.41] risulta essere la frequenza di incrocio delle costanti di fase dei due modelli. Nella Figura 10.9 sono tracciati con linea sottile gli andamenti di $|Z_c|/R$, α/X e β/X forniti dai due modelli asintotici al variare della frequenza normalizzata f/f_x; sono poi riportati a tratto pieno gli andamenti delle medesime funzioni ricavabili dalle [10.18] e [10.19] con la $r(f)$ effettiva.

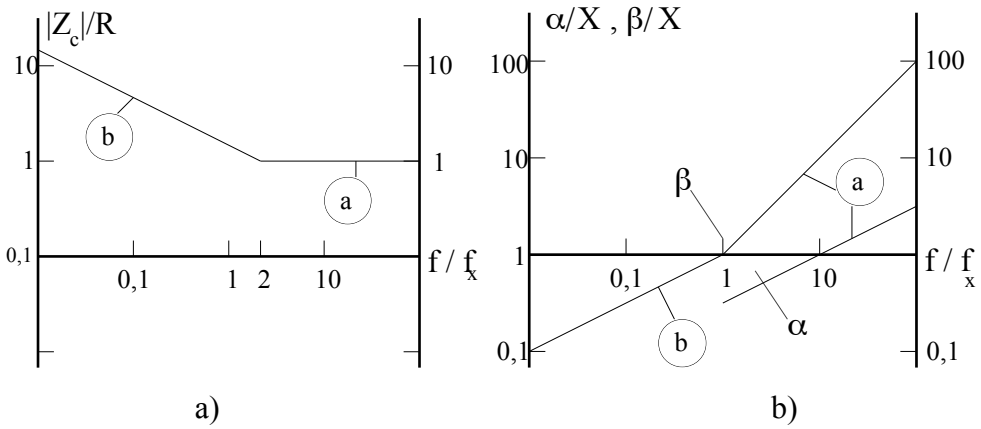


Figura 10.9: Andamenti asintotici (a tratto sottile) ed effettivi (a tratto pieno) del modulo della impedenza caratteristica (a) e delle costanti di attenuazione e di fase (b).

Si osserva in conclusione quanto segue:

- i modelli asintotici in bassa e alta frequenza sono attendibilmente impiegabili solo quando la frequenza f_x risulti nettamente esterna alla banda utile del segnale da trasmettere;
- il comportamento delle coppie metalliche alle alte frequenze è decisamente più prossimo a quello perfetto, ma comporta attenuazioni maggiori;
- nel campo di validità dei modelli la costante di attenuazione appare comunque variabile con legge proporzionale alla radice quadrata della frequenza e, grazie alla [10.41], si constata che la crescita alle frequenze elevate avviene con un fattore costante minore di quello alle frequenze basse;
- il comportamento a banda larga delle coppie metalliche, con la frequenza f_x interna alla banda utile del segnale da trasmettere, è marcatamente imperfetto.

10.3.4 Mezzi trasmissivi con coppie simmetriche

10.3.4.1 Coppia simmetrica

Si abbia una coppia simmetrica con diametro dei fili d, distanza tra gli assi D, conduttori in rame con $\sigma = 0{,}58\ 10^8\ [\Omega^{-1}/m]$ e $\mu_m = \mu_0 = 4\pi\ 10^{-7}$ [H/m], e isolante di costante dielettrica relativa $\varepsilon_r = \varepsilon/\varepsilon_0$, dove $\varepsilon_0 = (36\pi)^{-1}10^{-9}$ [F/m].

Dato che per le sezioni e i perimetri dei due conduttori si hanno le:

$$[10.45]\ A_1 = A_2 = \frac{\pi d^2}{4},\ C_1 = C_2 = \pi d,$$

dalle [10.26] e [10.34] si ottiene:

$$[10.46]\ r_b = \frac{1}{\sigma}\frac{8}{\pi d^2};$$

$$[10.47]\ r_a(f) = \frac{1}{\sigma}\sqrt{\pi\mu_m\sigma}\ \sqrt{f}\ \frac{2}{\pi d} = 2\sqrt{\frac{\mu_0}{\pi\sigma}}\ \frac{1}{d}\sqrt{f}\ ;$$

ammesso che sia D»d/2, si hanno poi le seguenti espressioni approssimate degli altri due parametri primari significativi (si rammenta che *g* può essere trascurato):

$$[10.48]\ l \cong \frac{\mu_o}{\pi}\ \ln(2D/d),$$

$$[10.49]\ c \cong \frac{\pi\varepsilon_0\varepsilon_r}{\ln(2D/d)}.$$

Da tali ultime espressioni risulta l'impedenza caratteristica ad alta frequenza (vedi [10.31]):

$$[10.50]\ Z_a = \sqrt{\frac{l}{c}} \cong \frac{1}{\pi}\ \sqrt{\frac{\mu_0}{\varepsilon_0}}\ \frac{1}{\sqrt{\varepsilon_r}}\ln\left(\frac{2D}{d}\right) = \frac{120}{\sqrt{\varepsilon_r}}\ln\left(\frac{2D}{d}\right)\ ;$$

si calcola infine dalla [10.41] la frequenza di incrocio:

$$[10.51]\ f_x = \frac{r_b}{4\pi l} \cong \frac{2}{\pi\sigma\mu_0 d^2\ln(2D/d)}.$$

Scegliendo a titolo di esempio i valori 2D/d = 5 [ln(2D/d) = 1,61] e ε_r=2,26 (polietilene), per alcuni valori molto comuni del diametro d si calcolano i valori mostrati in Tabella 10.2 della resistenza assiale unitaria in continua, della induttanza assiale unitaria, della capacità trasversale unitaria, della impedenza caratteristica ad alta frequenza e della frequenza di incrocio.

Tabella 10.2: Parametri caratteristici di una coppia simmetrica

d [mm]	0,4	0,6	0,9
r_b [Ω/km]	275	122	44
l [mH/km]	0,64		
c [nF/km]	40		
Z_a [Ω]	128		
f_x [kHz]	33,8	15,0	6,7

I dati dell'ultima riga mostrano come nella più diffusa applicazione della coppia simmetrica, ossia per il collegamento di terminali con trasmissione di segnali entro la banda 300 - 3.400 Hz, tipica della telefonia analogica, sia ammissibile il modello asintotico a bassa frequenza, con parametri primari indipendenti dalla frequenza.
A partire dalla [10.42] e con i valori sopra considerati si hanno allora gli andamenti del modulo della impedenza caratteristica a bassa frequenza mostrati in Figura 10.10, sensibilmente variabili nella banda telefonica attorno a 600 Ω. A partire dalla [10.29] si hanno poi gli andamenti della costante di attenuazione e della costante di fase mostrati in Figura 10.11. Da essi si deduce che è sufficiente contenere opportunamente la lunghezza del tratto di coppia entro alcune centinaia di metri per poterlo considerare elettricamente corto, ossia con effetti praticamente trascurabili a riguardo della risposta in fase.

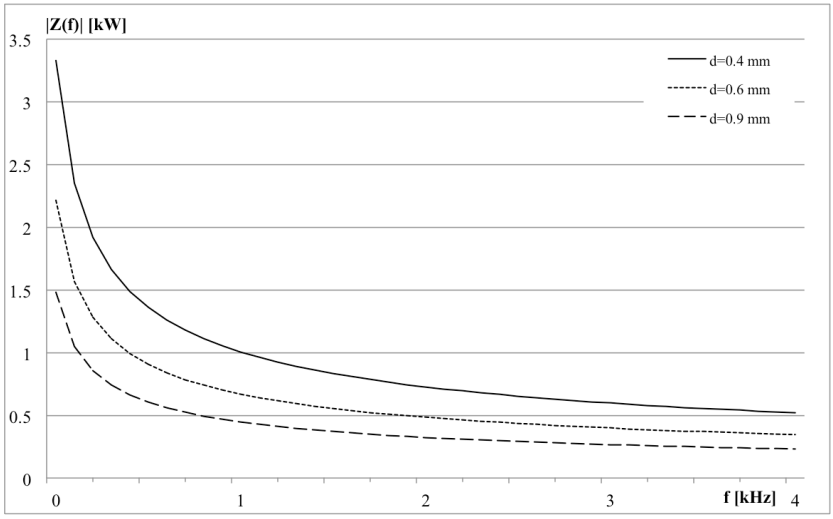


Figura 10.10: Esempi di andamenti del modulo della impedenza caratteristica di coppie simmetriche in bassa frequenza.

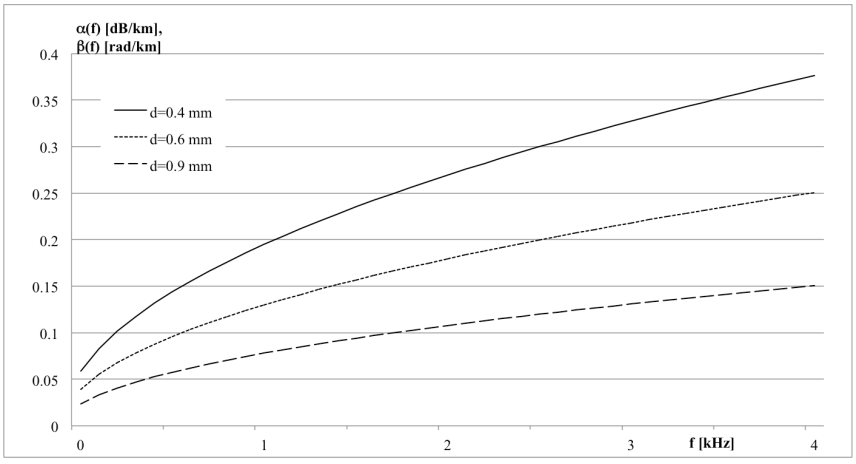


Figura 10.11: Esempi di andamenti costante di attenuazione e della costante di fase di coppie simmetriche in bassa frequenza.

Nei casi in cui il tratto di coppia venga usato per distanze nettamente maggiori di quelle considerate oppure per la trasmissione di segnali con estremo di banda ben superiore a quello della banda telefonica, è necessario considerarne il comportamento senza assumere approssimazioni asintotiche.

10.3.4.2 Cavi a coppie simmetriche

Molto frequentemente accade che numerosi collegamenti in coppia simmetrica abbiano in comune una parte del loro tragitto. Allo scopo di ridurre i costi di posa e installazione delle linee e di proteggere maggiormente il mezzo trasmissivo, si adottava nella parte comune del percorso la multiplazione di spazio delle coppie simmetriche, raccogliendole in una struttura cava, appunto denominata *cavo a coppie simmetriche*.

La maggiore parte dei cavi a coppie simmetriche era destinata a percorsi cittadini, prevedendo prevalentemente una posa interrata. Nel passato più lontano si impiegavano cavi contenenti moltissime coppie (alcune migliaia) per i collegamenti di giunzione tra nodi di commutazione urbana; in tempi più recenti i cavi, costituiti da alcune centinaia di coppie, venivano posati solo per disporre dei collegamenti tra le centrali di commutazione e i terminali di utente: la struttura di uno di tali cavi è mostrata in Figura 10.12. In esso 10 coppie simmetriche con d = 0,4 mm e isolate con polietilene compatto vengono cordate, ossia avvolte a spirale, a formare un sottogruppo; 10 sottogruppi cordati formano poi un gruppo di 100 coppie e 8 di tali gruppi vengono raccolti in un unico cavo a 800 coppie.

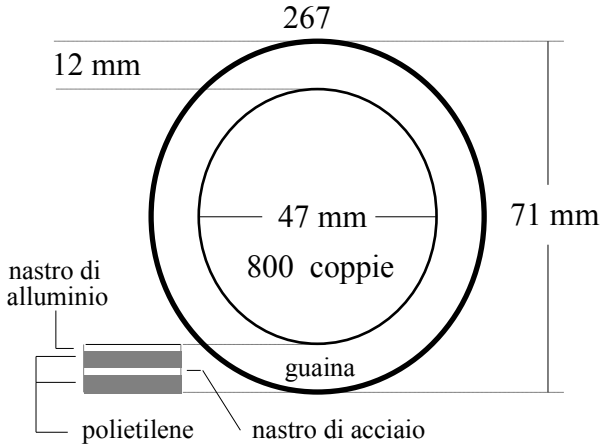


Figura 10.12: Cavo a 800 coppie simmetriche.

È da rimarcare che nell'affasciamento di più conduttori si ha l'alterazione della capacità trasversale per unità di lunghezza rispetto al valore riscontrabile nella coppia isolata, nel senso di un suo aumento. Ad esempio nel caso di diametro dei fili d = 0,4 mm si ottiene $c = 50$ nF/km.

10.3.5 Mezzi trasmissivi con coppie coassiali

10.3.5.1 Coppia coassiale

Si abbia una coppia coassiale con diametro esterno d del conduttore interno e diametro interno D del conduttore esterno, entrambi in rame con $\sigma = 0{,}58\ 10^8$ [Ω^{-1}/m] e $\mu_m = \mu_0 = 4\pi\ 10^{-7}$ [H/m], e isolante di costante dielettrica relativa $\varepsilon_r = \varepsilon/\varepsilon_0$ Il maggiore costo di tale struttura chiusa rispetto a quella bifilare può essere giustificato solo dalle migliori prestazioni, ottenibili grazie alla schermatura: è allora necessario che lo spessore di penetrazione δ_m sia minore dello spessore s del conduttore esterno, condizione che alla luce della [10.14] può essere soddisfatta con piccoli valori di quest'ultimo a patto che la coppia sia impiegata per la trasmissione di segnali con frequenza minima superiore a qualche decina di kHz. Nel seguito si farà pertanto riferimento alla resistenza assiale unitaria ad alta frequenza $r_a(f)$, espressa dalla [10.34].
Dato che per le sezioni e i perimetri dei due conduttori si hanno le:

$$[10.52]\ A_1 = \frac{\pi d^2}{4}\ ,\ A_2 \cong \pi s D\ ,\ C_1 = \pi d\ ,\ C_2 = \pi D,$$

dalla [10.34] si ottiene:

$$[10.53]\ r_a(f) = \frac{1}{\sigma}\sqrt{\pi\mu_m\sigma}\ \sqrt{f}\ \frac{d+D}{\pi d D} = \sqrt{\frac{\mu_o}{\pi\sigma}}\ \frac{1+D/d}{D}\sqrt{f}\ ,$$

mentre il valore in continua risulta dalla [10.26]:

$$[10.54]\ r_b \cong \frac{1}{\sigma}\ \frac{d^2+4sD}{\pi d^2 sD} = \frac{1}{\sigma}\ \frac{4\left(1+d^2/4sD\right)}{\pi d^2};$$

si hanno poi le seguenti espressioni approssimate degli altri due parametri primari significativi (si rammenta che g può essere trascurato):

$$[10.55]\ l \cong \frac{\mu_o}{2\pi}\ln(D/d),$$

$$[10.56]\ c \cong \frac{2\pi\varepsilon_0\varepsilon_r}{\ln\left(D/d\right)}.$$

Da tali ultime espressioni risulta l'impedenza caratteristica ad alta frequenza (vedi [10.31]):

$$[10.57]\ Z_a = \sqrt{\frac{l}{c}} \cong \frac{1}{2\pi}\sqrt{\frac{\mu_0}{\varepsilon_0}}\ \frac{1}{\sqrt{\varepsilon_r}}\ \ln(D/d) = \frac{60}{\sqrt{\varepsilon_r}}\ \ln(D/d);$$

si calcola infine dalla [10.41] la frequenza di incrocio:

$$[10.58]\ f_x = \frac{r_b}{4\pi l} \cong \frac{2\left(1+d^2/4sD\right)}{\pi\sigma\mu_0 d^2 \ln\left(D/d\right)}.$$

Rammentando le [10.33], si ricavano le costanti di attenuazione e di fase del modello asintotico valido per $f \gg f_x$:

$$[10.59]\ \alpha_a(f) = \sqrt{\frac{\pi\varepsilon_0}{\sigma}}\ \frac{1+D/d}{D\ln\left(D/d\right)}\ \sqrt{\varepsilon_r}\ \sqrt{f},$$

$$[10.60]\ \beta_a(f) = 2\pi f\sqrt{\mu_0\varepsilon_0}\ \sqrt{\varepsilon_r} = \sqrt{\varepsilon_r}\ \beta_0(f),$$

essendo $\beta_0(f) \triangleq 2\pi f/c$ la costante di fase intrinseca del vuoto.
Si lascia al lettore la dimostrazione che la costante di attenuazione al variare del rapporto D/d risulta minima per il valore ottimo:

$$[10.61]\ (D/d)_0 \cong 3{,}6\ ,$$

che viene pertanto di norma adottato nella costruzione della coppia coassiale. Per le [10.57] e [10.59] si possono allora adoperare le formule pratiche:

$$[10.62]\ Z_a \cong \frac{77}{\sqrt{\varepsilon_r}}\ ,$$

$$[10.63]\ \alpha_a(f) \cong \frac{2{,}48}{D}\sqrt{\varepsilon_r}\ \sqrt{f}\ \ 10^{-9}[\text{Nep/m}] = \frac{2{,}15}{D}\sqrt{\varepsilon_r}\ \sqrt{f}\ \ 10^{-8}[\text{dB/m}].$$

Dalla [10.63] risulta inoltre che una coppia coassiale costruita con l'intento di minimizzare le perdite ha la struttura mostrata in Figura 10.6a; infatti con il centraggio del conduttore interno tramite dischetti dielettrici si riescono a ottenere valori della costante dielettrica relativa equivalente attorno a 1,1, che sono assai prossimi al valore minimo unitario. Tale tipo di coppia, ottima per la trasmissione a grande distanza di segnali del tipo in banda base, ha allora una impedenza caratteristica $Z_a \cong 75\ \Omega$.
La coppia coassiale che è stata maggiormente impiegata nella trasmissione a grande distanza, del tipo a dischetti dielettrici, ha i parametri strutturali:

$$d = 2{,}6\text{ mm},\ D = 9{,}5\text{ mm},\ s = 0{,}254\text{ mm},\ \varepsilon_r \cong 1{,}07;$$

sono state anche usate le coppie, sempre con dischetti dielettrici, con parametri strutturali:

$$d = 1{,}2\text{ mm},\ D = 4{,}4\text{ mm},\ s = 0{,}180\text{ mm},\ \varepsilon_r \cong 1{,}15,$$
$$d = 0{,}7\text{ mm},\ D = 2{,}9\text{ mm},\ s = 0{,}100\text{ mm},\ \varepsilon_r \cong 1{,}40,$$

denominate rispettivamente "mini-coassiale" e "micro-coassiale". Il lettore può verificare che i valori scelti soddisfano sostanzialmente la [10.61]; applicando le [10.54] - [10.58] risultano i valori mostrati in Tabella 10.3.

Tabella 10.3: Parametri caratteristici di una coppia coassiale

d [mm] - D [mm]	0,7 - 2,9	1,2 - 4,4	2,6 - 9,5
r_b [Ω/km]	64	22	5,5
l [mH/km]	0,28	0,26	0,26
c [nF/km]	55	49	46
Z_a [Ω]	72	73	75
f_x [kHz]	17,8	6,8	1,7

I dati dell'ultima riga mostrano che il buon comportamento tipico del modello asintotico ad alta frequenza, con impedenza caratteristica quasi costante e costante di fase praticamente proporzionale alla frequenza, è ottenibile in effetti purché i segnali trasmessi nelle coppie coassiali considerate abbiano spettri con frequenza minima sufficientemente maggiore della frequenza di incrocio. Si osserva inoltre che se tale condizione non è rispettata gli spessori del conduttore esterno non hanno sufficiente effetto di schermo.
Nelle applicazioni su percorsi brevi, tanto per segnali in banda base (ad esempio in reti locali), quanto per segnali in banda traslata (tipicamente per collegamenti tra antenne e apparati), si preferisce che la coppia coassiale goda di buona flessibilità meccanica, offerta dalla struttura mostrata in Figura 10.6c, con dielettrico pieno; assumendo che sia polietilene, che ha $\varepsilon_r = 2{,}26$, si ha in tale caso dalla [10.62] una impedenza caratteristica $Z_a \cong 50\ \Omega$. Quando i segnali siano in banda traslata a frequenza molto alta (ordine del GHz), per contenere il forte aumento della costante di attenuazione (vedi [10.63]) può risultare conveniente la tendenza ad aumentare la dimensione della sezione, sempre nel rispetto della [10.61], sopportando il maggiore costo per unità di lunghezza che ne deriva. In tale prospettiva occorre tuttavia tenere presente che è comunque opportuno rimanere in regime di propagazione unimodale, soddisfando la condizione:

$$[10.64]\ f_t = \frac{c}{\lambda_t} \cong \frac{2c}{\pi\left(d+D\right)\sqrt{\varepsilon_r}} > f\,,$$

dove f_t è la frequenza di taglio del primo modo superiore nel cavo coassiale. Si noti che con i valori sopra considerati si ottiene $fD < 10^8$, ossia D < 10 mm per f=10 GHz.

10.3.5.2 Cavi a coppie coassiali

Nel passato la coppia coassiale è stata largamente impiegata come mezzo trasmissivo per collegamenti a grande distanza, su lunghezze anche di alcune migliaia di chilometri, con segnali del tipo in banda base sia analogici che numerici. Per la esigenza di disporre di almeno due coppie, da dedicare ai due versi delle comunicazioni tipicamente bidirezionali come la telefonia, ma soprattutto per aumentare la capacità di trasporto in direttrici ad alto traffico era prassi comune raccogliere una pluralità di coppie coassiali all'interno di un struttura cava, appunto denominata *cavo a coppie coassiali* o semplicemente *cavo coassiale*.
I cavi coassiali trovavano largo impiego per disporre di collegamenti di giunzione tra i nodi principali delle reti nazionali, oltre che per realizzare collegamenti internazionali, continentali e transoceanici. Nelle direttrici nazionali a maggiore traffico la multiplazione di spazio, che inizialmente riuniva quattro coppie coassiali, si è spinta fino a raccogliere una ventina di esse. Un esempio tipico di un cavo di tale tipo, per posa interrata, è quello mostrato in Figura 10.13, contenente 16 coppie isolate a dischetti con dimensioni d=2,6 mm e D=9,5 mm. Si può notare come la struttura sia adeguatamente protetta, oltre che da un rivestimento plastico, da una guaina in alluminio corrugato che assicura buone proprietà meccaniche pur consentendo una certa flessibilità. Negli interstizi tra le coppie coassiali e le nastrature erano spesso alloggiate anche alcune coppie simmetriche cordate assieme due a due, denominate bicoppie.

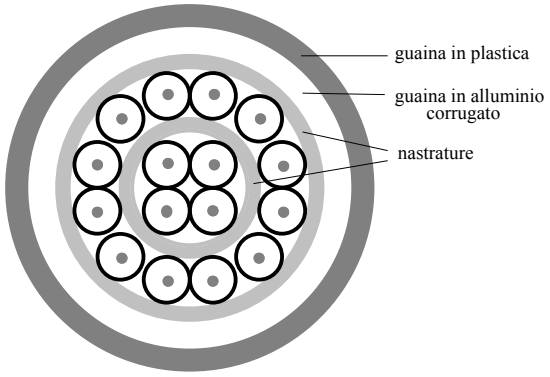


Figura 10.13: Cavo a 16 coppie coassiali per installazione sotterranea.

10.4 MEZZI TRASMISSIVI OTTICI

10.4.1 Struttura delle fibre ottiche

Nella trasmissione a distanza di segnali con banda utile molto estesa è assai frequente l'impiego di un mezzo portante fisico basato su una struttura interamente isolante. Nella soluzione tipica mostrata in Figura 10.14, a salto di indice ("step index"), essa è costituita da due materiali dielettrici separati da una unica superficie cilindrica, a sezione circolare di raggio a, che delimita il *nucleo* interno ("core"), con costante dielettrica relativa $\varepsilon_{r1}=\varepsilon_1/\varepsilon_0$ e indice di rifrazione $n_1=\sqrt{\varepsilon_{r1}}$, dal *mantello* che lo circonda ("cladding"), con costante dielettrica relativa $\varepsilon_{r2}=\varepsilon_2/\varepsilon_0$ e indice di rifrazione $n_2=\sqrt{\varepsilon_{r2}}$ di valore poco minore di n_1. L'effetto di guida del campo e.m. da parte della superficie di discontinuità è ottenibile anche con differenze assai piccole degli indici di rifrazione, caso in cui la struttura è denominata *guida d'onda dielettrica lieve* uniforme; si fa nel seguito esclusivo riferimento alla differenza relativa degli indici di rifrazione:

$$[10.65]\ \Delta \triangleq \frac{n_1^2 - n_2^2}{2\,n_2^2} \cong \frac{n_1 - n_2}{n_2} << 1,$$

con valori pratici non superiori a 1%. Entrambi i materiali hanno permeabilità magnetica, $\mu_1=\mu_2=\mu_0=4\pi\,10^{-7}$ [H/m], uguale a quella del vuoto.

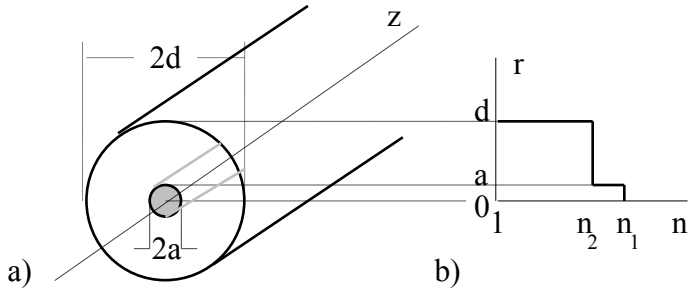


Figura 10.14: Fibra ottica con sezione a simmetria circolare (a), e profilo teorico dell'indice di rifrazionenlungo il raggio (b).

La struttura schematizzata in Figura 10.14, con sezione invariante in senso longitudinale ed estensione trasversale finita del mantello, è usualmente denominata *fibra ottica*, date le sue piccolissime dimensioni trasversali (di norma 2d = 125 μm) e l'uso di frequenze prossime a quelle dello spettro visibile. Nell'ambito del processo di produzione la fibra ottica viene ricoperta con un rivestimento dielettrico, opaco alle frequenze impiegate, che oltre ad assicurare una completa schermatura alle frequenze interessate ha il compito di consentire una migliore maneggevolezza del mezzo trasmissivo, altrimenti molto fragile; il diametro esterno risulta così raddoppiato.
La fibra ottica a salto di indice di rifrazione viene in effetti realizzata con un profile non discontinuo, ma raccordato in modo continuo tra i valori n_1 e n_2 entro un breve intervallo centrato sul raggio teorico a del nucleo. Il comportamento non è tuttavia molto discosto da quello deducibile con il modello teorico a salto di indice. Gli approfondimenti sugli effetti di particolari profili dell'indice di rifrazione, sempre del tipo con transizione quasi brusca tra nucleo e mantello oppure con profilo graduale (in inglese "graded index"), esulano dallo scopo delle presenti note.
Nonostante che il percorso effettivo della fibra ottica sia in generale curvilineo, si assume in seguito che la struttura sia cilindrica lungo il suo sviluppo assiale, in modo da potere adottare una coordinata rettilinea z: i raggi di curvatura dell'asse sono infatti abbastanza grandi da potere assumere che il comportamento e.m. della struttura sia equivalente a quello che si avrebbe qualora essa fosse posata lungo una retta.

10.4.2 Grandezze caratteristiche delle fibre ottiche

10.4.2.1 Modi guidati

La fibra ottica è una struttura aperta, con campo e.m. che si estende in teoria in tutto lo spazio dielettrico, non omogeneo. Tuttavia nelle condizioni di impiego il campo e.m. guidato è in pratica prevalentemente concentrato nella regione del nucleo, di raggio a, in modo che nel mantello i valori del campo risultano trascurabili a distanze dall'asse minori del raggio, d, della superficie cilindrica che lo delimita; ciò rende lecita la ipotesi di mantello indefinitamente esteso in senso trasversale.

Una fibra ottica ideale con mantello praticamente illimitato può consentire la propagazione di una pluralità di modi guidati, a patto che la frequenza sia maggiore delle frequenze di taglio di questi ultimi. A tale proposito è usuale fare riferimento alla *frequenza normalizzata*, definita per una struttura nota dalla:

$$[10.66]\ V \triangleq a\,\omega\sqrt{\mu_0\left(\varepsilon_1-\varepsilon_2\right)} = a\,k\sqrt{n_1^2 - n_2^2} = a\,k\,n_2\sqrt{2\Delta}\,,$$

dove k è la costante di fase intrinseca (numero d'onda) del vuoto,

$$[10.67]\ k \triangleq \omega\sqrt{\mu_0\varepsilon_0} = \frac{2\pi f}{c} = \frac{2\pi}{\lambda}\,,$$

essendo $\lambda = c/f$ la lunghezza d'onda nel vuoto ($c \cong 3\cdot 10^8$ m/s). Utilizzando la [10.66] si ha il vantaggio che le frequenze di taglio normalizzate, V_{ti}, dei modi guidati sono espresse tramite numeri noti: in particolare il modo fondamentale ha frequenza di taglio nulla, mentre per il primo modo superiore si ha $V_{t1} = 2{,}405$ (vedi appendice A5.4).

In generale in un tronco di lunghezza L di fibra ottica che operi alla frequenza f in regime multimodale occorre fare riferimento, oltre che alla *costante di fase* $\beta_0(f)$ del modo fondamentale, a tutte le *costanti di fase*, $\beta_i(f)$, degli altri m modi superiori che si possono propagare nella struttura priva di perdite, ossia per cui le frequenze di taglio normalizzate V_{ti} risultano minori di V. Ad ogni frequenza non inferiore a quella di taglio ciascuna delle grandezze considerate risulta compresa entro i limiti espressi mediante la relazione (vedi [A5.4.1]):

$$[10.68]\ k(f)n_2 \le \beta_i(f) \le k(f)n_1, \qquad i = 0,1,2,..,m\,,$$

dove kn_2 e kn_1 sono le costanti di fase intrinseche del dielettrico illimitato rispettivamente con indice di rifrazione n_2 e n_1. Per ciascuno dei modi il primo segno di uguaglianza nella [10.68] si verifica al taglio, ossia si ha $\beta_i(f_{ti}) = k(f_{ti})\,n_2$; il secondo si ha per $f\to\infty$, ossia per $f \gg f_{ti}$ si ha in pratica $\beta_i(f) \cong k(f)n_1$.

Nel caso di fibra ottica reale la propagazione del modo i-esimo può essere caratterizzata con buona approssimazione tramite la *costante di propagazione* complessa:

$$[10.69]\ \gamma_i(f) = \alpha_i(f) + j\,\beta_i(f),\ i = 0,1,2,..,m,$$

assumendo che la costante di fase $\beta_i(f) \triangleq \Im\{\gamma_i(f)\}$ sia quella del corrispondente modo della struttura idealizzata e dove la *costante di attenuazione* $\alpha_i(f) \triangleq \Re\{\gamma_i(f)\}$ tiene conto delle reali condizioni.

10.4.2.2 Costante di fase normalizzata e ritardo di gruppo specifico

Si osservi che in una fibra ottica a piccolo salto di indice, in cui sono assai prossimi i due indici di rifrazione (vedi [10.65]), dalla limitazione [10.68] risulta che l'andamento della costante di fase $\beta_i(f)$ del generico modo guidato si discosta assai poco da quello della costante di fase intrinseca kn_2. Le differenze, per quanto molte piccole, vengono tuttavia esaltate dai valori molto grandi delle *lunghezze elettriche* $\beta_i(f)L$, calcolabili in prima approssimazione per mezzo della $\beta_i L \cong kn_2 L = 2\pi n_2 L/\lambda$ con L enormemente maggiore di λ.

Allo scopo di rendere più evidenti gli scostamenti sopra menzionati si ricorre alla *costante di fase normalizzata*, definita tramite la espressione (vedi [A5.4.2]):

$$[10.70]\ b_i \triangleq \frac{\beta_i^2 - k^2 n_2^2}{k^2 n_1^2 - k^2 n_2^2} \cong \frac{\beta_i - k n_2}{k n_1 - k n_2}\ ,\ \text{per } i = 0,1,2,..,m;$$

si ottiene infatti per ogni modo una diversa funzione $b_i(f)$, che grazie alla [10.68] varia dal valore 0 alla frequenza di taglio al valore 1 per $f \gg f_{ti}$. Risolvendo la [10.70] rispetto a β_i si ottiene la seguente relazione:

$$[10.71]\ \beta_i \cong k n_2 + k b_i (n_1 - n_2),\ i = 0,1,2,..,m,$$

che permette di risalire alla costante di fase una volta nota quella normalizzata.
La deviazione dall'andamento non lineare della costante di fase è messa in evidenza tramite il *ritardo di gruppo specifico* (per unità di lunghezza) dello i-esimo modo guidato:

$$[10.72]\ t_i \triangleq \frac{d\beta_i}{d\omega} = \frac{1}{c}\frac{d\beta_i}{dk},\ i = 0,1,2,..,m.$$

In generale gli indici di rifrazione n_1 e n_2 dei dielettrici sono variabili con la frequenza, ma si può spesso ammettere che la variazione con la frequenza della loro differenza sia trascurabile; sotto tale ipotesi e con l'approssimazione $dV/dk \cong V/k$, derivando la [10.71] si ricava la seguente espressione:

$$[10.73]\ t_i \cong \frac{1}{c}\frac{d(k n_2)}{dk} + \frac{n_1 - n_2}{c}\frac{d(V b_i)}{dV},\ i = 0,1,2,..,m\ ,$$

dove la funzione $d(Vb_i)/dV$ assume per una stessa V valori diversi in dipendenza dal modo (vedi fig. A5.2.b), ma comunque compresi tra 0 e 2, tendendo a 1 per $V \gg V_{ti}$.
In base alla [10.73] e considerando anzitutto ciascun modo da solo, si osserva che il relativo ritardo di gruppo specifico non risulta costante con la frequenza: una fibra ottica è dunque soggetta comunque alla *dispersione intramodale*, anche se operante in regime monomodale, con il solo modo fondamentale. Si nota successivamente che i ritardi di gruppo specifici $t_i(f_c)$ dei vari modi a una stessa frequenza portante $f_c = c/\lambda_c$ non sono uguali tra loro: ciò conduce ad affermare che una fibra ottica che operi in regime multimodale mostra anche una *dispersione intermodale*.
Il primo addendo del secondo termine nella [10.73] evidenzia un effetto di *dispersione del materiale*, che prende origine dalla identica dipendenza dalla frequenza (e quindi da k) degli indici di rifrazione n_1 e n_2 dei dielettrici; il secondo addendo esplicita un secondo effetto di *dispersione di guida*, tipico della propagazione guidata con modi non TEM. Si noti che quest'ultimo è l'unico responsabile della dispersione intermodale, poiché il primo effetto agisce ugualmente su tutti i modi.

10.4.2.3 Costante di attenuazione

Il generico modo in una fibra ottica reale è caratterizzato, oltre che dalla costante di propagazione, da una costante di attenuazione:

$$[10.74]\ \alpha_i \triangleq -\frac{1}{2}\ln\left(\frac{P_i - P_{pi}}{P_i}\right) \cong \frac{1}{2}\frac{P_{pi}}{P_i}\ ,$$

che tiene conto della assai piccola quota P_{pi} di potenza perduta in un percorso di lunghezza unitaria in ragione della potenza iniziale P_i dell'onda e.m.. La potenza perduta P_{pi} è dovuta a trasformazione di energia e.m. guidata che si converte in calore, con corrispondente *attenuazione per assorbimento*, oppure che si converte in energia e.m. di modi irradianti, con corrispondente *attenuazione per irradiazione*. In entrambi i casi si hanno sia cause intrinseche, legate ai materiali dielettrici, che estrinseche, connesse a impurità indesiderate nei materiali o a irregolarità nella geometria della struttura (principalmente la non rettilineità dell'asse).
Ammettendo la medesima qualità nei due dielettrici di una fibra geometricamente perfetta, si perviene a costanti di attenuazione che alla frequenza assegnata assumono sostanzialmente il

medesimo valore $\alpha(f)$, indipendentemente dai modi, tranne che per un sensibile aumento nell'intorno delle frequenze di taglio(1). In una buona struttura reale, con mantello in vetro di silice assai puro e nucleo costituito dal medesimo materiale, ma opportunamente drogato allo scopo di ottenere la desiderata differenza dell'indice di rifrazione, la costante di attenuazione ha un andamento come quello mostrato nell'esempio in Figura 10.15, che palesa la convenienza ad applicare le guide dielettriche considerate impiegando segnali con lunghezze d'onda nell'intervallo spettrale del vicino infrarosso (frequenze dell'ordine di $3\ 10^{14}$ Hz).

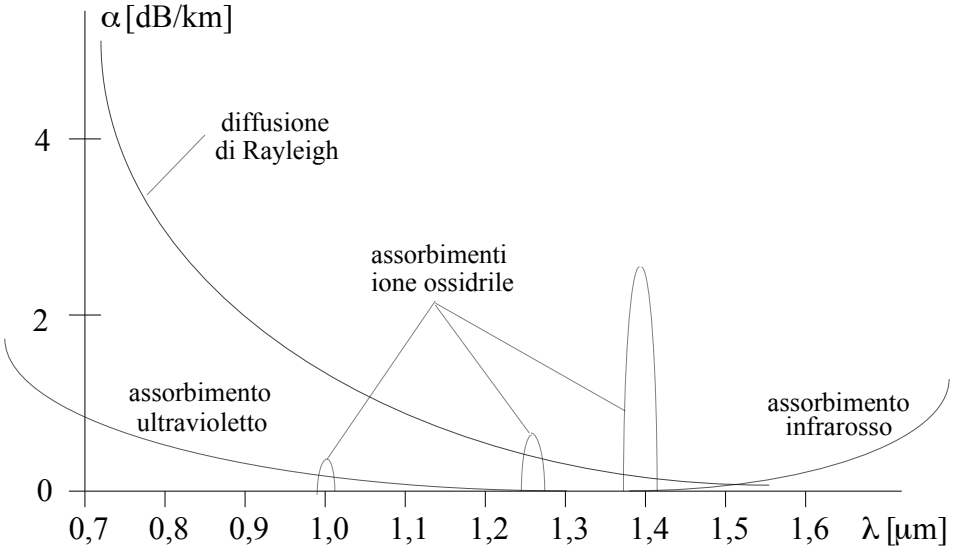


Figura 10.15: Costante di attenuazione in una fibra di vetro di silice.

La Figura 10.15 tiene conto delle seguenti cause di attenuazione intrinseca:

- diffusione di Rayleigh con conversione in energia e.m. irradiante, secondo una legge proporzionale a λ^{-4}, dovuta a fluttuazioni microscopiche dell'indice di rifrazione,
- assorbimento nell'infrarosso con conversione in calore, dovuto a vibrazioni molecolari,
- assorbimento nell'ultravioletto con conversione in calore, dovuto a transizioni elettroniche;

tiene inoltre conto dell'attenuazione estrinseca per assorbimento con conversione in calore, dovuto a presenza di impurità (solo ione ossidrile OH^-).

Nella fase iniziale delle applicazioni nelle telecomunicazioni le fibre ottiche erano prodotte con maggiore attenuazione estrinseca per impurità indesiderate, con effetto più rilevante proprio in corrispondenza dei valori più bassi della attenuazione intrinseca. Fu allora scelto di operare in corrispondenza della lunghezza d'onda portante $\lambda_c \cong 0{,}85\ \mu m$, che si usa denominare 1ª finestra, anche per la disponibilità di sorgenti ottiche direttamente funzionanti attorno a tale valore. In seguito ai decisi miglioramenti tecnologici dei processi produttivi le applicazioni si sono orientate prima verso la lunghezza d'onda $\lambda_c \cong 1{,}3\ \mu m$ e poi $\lambda_c \cong 1{,}55\ \mu m$, rispettivamente denominate 2ª finestra e 3ª finestra: infatti in una fibra ottica attuale di buona qualità a tali valori corrispondono rispettivamente un minimo relativo e il minimo assoluto di $\alpha(\lambda)$ (vedi Figura 10.15).

10.4.2.4 Modello approssimato con l'ottica geometrica

Limitatamente al caso di una fibra ottica a salto di indice di rifrazione con rapporto a/λ sufficientemente grande da permettere la propagazione di un numero molto elevato di modi ($V \gg V_{ti} = 2{,}405$), si può ricorrere a un modello approssimato ottenibile applicando l'ottica geometrica.

(1)In prossimità del taglio il campo si estende in quota considerevole nel mantello e il diametro esterno finito di quest'ultimo non rende più accettabile l'ipotesi di identica perdita intrinseca nei due dielettrici.

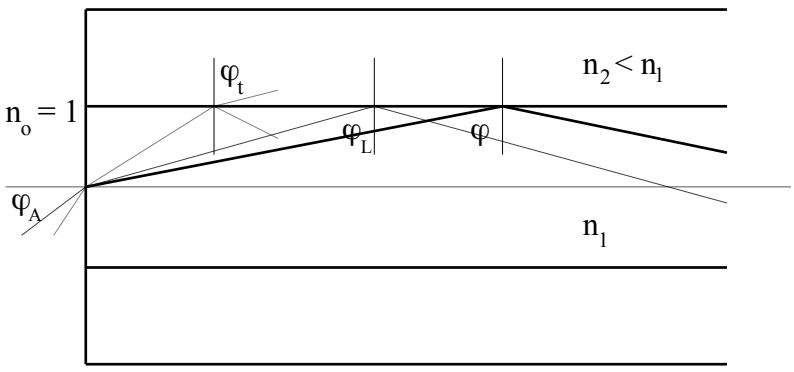


Figura 10.16: Raggi in una fibra ottica a salto di indice.

Con riferimento alla Figura 10.16, un raggio all'interno del nucleo subisce una riflessione totale sulla superficie di separazione tra nucleo e mantello, se l'angolo di incidenza φ è maggiore dell'angolo limite φ_L, in corrispondenza del quale l'angolo φ_t del raggio rifratto nel mantello assume il valore $\pi/2$. Dalla legge di rifrazione di Snell:

$$[10.75]\ n_1 \sin(\varphi) = n_2 \sin(\varphi_t)\ ,$$

per $\varphi_t = \pi/2$ si ottiene per definizione:

$$[10.76]\ \varphi_L \triangleq \arcsin\left(\frac{n_2}{n_1}\right);$$

si ha quindi la propagazione guidata, con assenza di rifrazione nel mantello, se si considerano solo i raggi con $\varphi > \varphi_L$, corrispondenti a quelli a monte della sezione iniziale della fibra ottica con angolo di incidenza minore del valore φ_A, ottenibile tramite la:

$$[10.77]\ \sin(\varphi_A) = n_1 \sin\left(\frac{\pi}{2} - \varphi_L\right) = n_1 \cos(\varphi_L)\ .$$

L'angolo φ_A è denominato *angolo di accettazione* della fibra; il seno di tale grandezza:

$$[10.78]\ \sin(\varphi_A) = n_1 \cos\left[\arcsin\left(\frac{n_2}{n_1}\right)\right] = \sqrt{n_1^2 - n_2^2} \triangleq \text{N.A.}\ ,$$

è denominato *apertura numerica* (N.A. = Numerical Aperture). Si noti che la [10.66] può essere posta nella forma:

$$[10.79]\ V = \frac{2\pi}{\lambda} a\ \text{N.A.}$$

Accettando l'approssimazione dell'ottica geometrica, valida al limite per $\lambda \to 0$, l'insieme dei modi guidati in numero finito molto grande risulta sostituito dall'insieme continuo dei raggi guidati, con $\varphi > \varphi_L$. Il modello inoltre non tiene conto della propagazione guidata a frequenze molto lontane da quella di taglio: infatti solo in tale condizione si tende all'annullamento del campo e.m. nel mantello, che corrisponde all'assenza di raggi rifratti.

10.4.3 Comportamento delle fibre ottiche a salto di indice

10.4.3.1 Comportamento in regime multimodale

Per caratterizzare il comportamento di un tratto di lunghezza L di guida dielettrica reale in regime di propagazione multimodale, ossia di un tronco di *fibra ottica multimodo*, si ricorre alle costanti di propagazione [10.69], ma si deve anche valutare come il segnale di eccitazione nella sezione iniziale si ripartisce negli m+1 modi, così come questi ultimi concorrono alla formazione del segnale in uscita, nella sezione finale al termine del tronco.

Con riferimento allo schema equivalente mostrato in Figura 10.17, in cui al mezzo ottico multimodale corrisponde una pluralità di diverse linee di trasmissione, si hanno di norma delle perdite di accoppiamento agli estremi, che possono essere valutate globalmente tramite una riflettenza $H_{11m}(f, L)$ e il coefficiente di riflessione del carico, e una trasmettenza complessiva:

$$[10.80]\ H_m(f,L) = \sum_{i=0}^{m} \xi_i e^{-\alpha_i(f)L} e^{-j\beta_i(f)L} ,$$

dove i coefficienti ξ_i dipendono dalle effettive condizioni agli estremi.

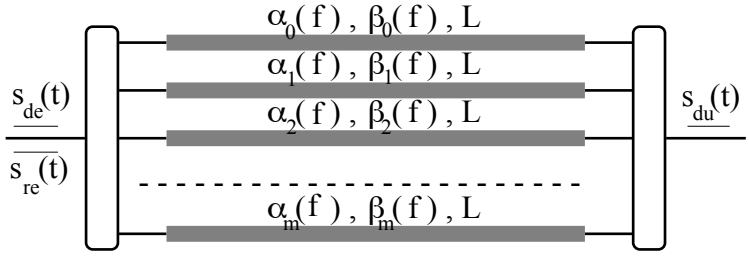


Figura 10.17: Schema equivalente di un tronco di fibra ottica multimodo.

Dato che la trasmissione in una fibra ottica multimodo a salto di indice interessa una banda monolatera relativa, B/f_c, estremamente piccola attorno a una precisata frequenza portante $f_c = c/\lambda_c$ di valore estremamente elevato, per $f \cong f_c$ è lecito in prima approssimazione assumere ξ_i costanti, trascurare la dipendenza dalla frequenza delle costanti di attenuazione ponendo:

$$[10.81]\ \alpha_i(f) \cong \alpha_i(f_c) \triangleq \alpha_{ic} ,$$

e ignorare la dispersione intramodale linearizzando le costanti di fase:

$$[10.82]\ \beta_i(f) \cong \beta_{ic} + 2\pi(f - f_c)\, t_{ic} ,$$

dove si è posto $\beta_i(f_c) \triangleq \beta_{ic}$ e si è indicato con $t_{ic} \triangleq t_i(f_c)$ il ritardo di gruppo specifico dello i-esimo modo guidato, calcolato in f_c.

Posto $\zeta_i \triangleq \xi_i e^{-\alpha_{ic}L} e^{-j\beta_{ic}L}$, la risposta impulsiva trasmessa del quadripolo equivalente in banda base (vedi [5.94]) a partire da quello rappresentato dalla [10.80] diviene allora:

$$[10.83]\ \underline{h}_m(t) \cong F^{-1}\left\{\sum_{i=0}^{m} \zeta_i e^{-j2\pi f t_{ic} L}\right\} = \sum_{i=0}^{m} \zeta_i \delta(t - t_{ic}L), \text{ per } |f| \le B ,$$

che nonostante le semplificazioni introdotte rivela un comportamento imperfetto, causato dalla dispersione intermodale: con riferimento anche alla Figura 10.17, esso è dovuto alla combinazione all'estremo del tronco di fibra di più segnali che, pure avendo la medesima origine, hanno viaggiato con diversi ritardi di gruppo $t_{ic}L$; dalla [10.83] si ottiene infatti l'inviluppo complesso in uscita:

$$[10.84]\ \underline{s}_{du}(t) = \underline{h}_m(t) * \underline{s}_{de}(t) = \sum_{i=0}^{m} \zeta_i\, s_{de}(t - t_{ic}L) .$$

Il difetto evidenziato nella [10.84] è valutabile qualitativamente tramite il valore massimo delle differenze (vedi [10.73]):

$$[10.85]\ \Delta t_{ij}(f_c) \triangleq (t_{ic} - t_{jc})L \cong \frac{n_1 - n_2}{c} L \left[\left.\frac{d(Vb_i)}{dV}\right|_{V_c} - \left.\frac{d(Vb_j)}{dV}\right|_{V_c} \right],$$

tra i ritardi di gruppo degli m modi che si possono efficacemente propagare alla frequenza normalizzata V_c, ottenibile dalla f_c tramite la [10.66]. In base agli andamenti delle funzioni $d(Vb_i)/dV$ mostrati in figura A5.2, il massimo valore teorico della [10.85] risulterebbe $\Delta t_M = 2\,(n_1 - n_2)L/c$; se però si escludono quei pochi modi più veloci per cui si verifica $d(Vb_i)/dV < 1$ per

$V_{ti} \cong V_c$, dato che per essi l'attenuazione da valutare in prossimità del taglio risulta nettamente maggiore di quella degli altri modi, il valore massimo della [10.85] si ottiene in pratica dalla differenza tra il ritardo di gruppo del modo più lento di ordine elevato, per cui $d(Vb_i)/dV \cong 2$, e quello del modo più veloce che sia lontano dal taglio, ossia il ritardo del modo fondamentale per cui si ha $d(Vb_0)/dV \cong 1$; si assume dunque (vedi [10.65]):

$$[10.86]\ \Delta t_M \cong \frac{n_1 - n_2}{c} L(2 - 1) \cong \frac{n_2}{c} L\Delta .$$

Indicati con $t_{r1L} \triangleq n_1 L/c$ e $t_{r2L} \triangleq n_2 L/c$ i tempi di ritardo intrinseci dei due materiali dielettrici sul percorso di lunghezza L, risulta $\Delta t_M \cong t_{r1L} - t_{r2L}$, da cui non è però lecito dedurre che t_{r1L} e t_{r2L} siano il massimo e minimo ritardo di gruppo nel tronco di fibra.
Il difetto evidenziato dalla [10.84] limita l'impiego delle fibre multimodo a salto di indice al caso di tronchi di lunghezza L abbastanza breve (dell'ordine di 100 m) da soddisfare la condizione $\Delta t_M B \ll 1$, ossia:

$$[10.87]\ LB << \frac{c}{n_2 \Delta},$$

in modo che per le frequenze in banda utile possano essere quasi trascurate le differenze tra i ritardi di gruppo, ossia l'effetto della dispersione intermodale risulti contenuto in limiti accettabili.
Ricorrendo al modello approssimato con l'ottica geometrica e trascurando in modo analogo la dispersione del materiale, si ricava per il generico raggio guidato con angolo di incidenza $\varphi > \varphi_L$ il tempo di ritardo nel tronco di lunghezza L, coincidente per ipotesi con il ritardo di gruppo:

$$[10.88]\ t_{rL}(\varphi) = \frac{n_1}{c \sin(\varphi)} L;$$

considerando il percorso più lungo ($\varphi = \varphi_L$) e quello più breve ($\varphi = \pi/2$), si ha allora servendosi della [10.76] la differenza massima:

$$[10.89]\ \Delta t_M = \frac{n_1^2}{c\, n_2} L - \frac{n_1}{c} L \cong \frac{n_1}{n_2} \frac{n_2}{c} L\Delta,$$

praticamente uguale alla [10.86]. Il risultato deriva dal fatto che l'approssimazione con l'ottica geometrica non tiene conto della propagazione in prossimità della frequenza di taglio e ciò corrisponde all'avere considerato nella [10.86] solo i modi per cui si ha $d(Vb_i)/dV \geq 1$.

10.4.3.2 Comportamento in regime monomodale

Il comportamento di un tronco di guida dielettrica di lunghezza L diviene decisamente molto soddisfacente anche per notevole larghezza di banda monolatera utile B (dell'ordine del GHz), a patto di operare in regime di propagazione unimodale(1), ossia con frequenze ($V < V_{t1} = 2{,}405$) che diano luogo alla propagazione del solo modo fondamentale. Si ha infatti in tale caso la scomparsa del dannoso effetto di dispersione intermodale, che altrimenti comporterebbe una severa limitazione del prodotto LB (vedi [10.87]), con la conseguente possibilità di estendere molto la lunghezza L (anche fino a 100 km). Diviene allora rilevante la dispersione intramodale, denominata anche *dispersione cromatica*, legata allo scostamento dalla linearità del ritardo di gruppo specifico (vedi [10.73] per i = 0) e perciò bene evidenziata tramite la derivata non nulla di quest'ultimo:

$$[10.90]\ s \triangleq \frac{dt_i}{d\omega} = \frac{1}{c^2} \frac{d^2\beta}{dk^2} \cong \frac{1}{c^2} \frac{d^2(kn_2)}{dk^2} + \frac{n_1 - n_2}{c^2 k} V \frac{d^2(Vb)}{dV^2} ,$$

calcolata con le medesime approssimazioni che hanno condotto alla [10.73].

---

(1) In seguito nella costante di propagazione e nelle grandezze ad essa legate viene omesso il pedice indicativo del modo fondamentale, non necessario in regime monomodale.

Grazie alla unimodalità, nello schema equivalente generale in Figura 10.17 resta una unica linea di trasmissione; si hanno ancora di norma delle perdite di accoppiamento agli estremi, che possono essere valutate globalmente tramite una riflettenza $H_{11,0}(f, L)$ e il coefficiente di riflessione del carico, mentre la trasmettenza del tronco di *fibra ottica monomodo* si semplifica nella:

$$[10.91]\ H_0(f, L) = \xi_0\, e^{-\alpha(f)L}\, e^{-j\beta(f)L},$$

dove il coefficiente ξ_0 dipende dalle effettive condizioni agli estremi.
Dato che la trasmissione interessa una banda monolatera relativa B/f_c che risulta comunque molto piccola dato il valore estremamente elevato della frequenza portante $f_c = c/\lambda_c$, per f attorno a f_c ma con $|f - f_c|$ non del tutto trascurabile rispetto a f_c è lecito ignorare la debole dipendenza dalla frequenza della costante di attenuazione (vedi Figura 10.15) ponendo:

$$[10.92]\ \alpha(f) \cong \alpha(f_c) \triangleq \alpha,$$

ma non si può trascurare la dispersione intramodale. La costante di fase del modo fondamentale è comunque rappresentabile con buona approssimazione tramite lo sviluppo in serie di potenze di punto iniziale f_c, troncato al secondo ordine:

$$[10.93]\ \beta(f) \cong \beta_c + 2\pi(f - f_c)t_c + \frac{1}{2}[2\pi(f - f_c)]^2 s_c, \text{ per } |f - f_c| \le B/2,$$

dove $\beta_c \triangleq \beta(f_c)$, $t_c \triangleq t(f_c)$ e $s_c \triangleq s(f_c)$.
Posto $\zeta_0 \triangleq \xi_0 e^{-\alpha L}\, e^{-j\beta_c L}$, il quadripolo equivalente in banda base (vedi [5.94]), a partire da quello rappresentato dalla [10.91], può essere scomposto in due elementi in cascata, come mostrato in Figura 10.18, ottenendo un quadripolo ideale con trasmettenza:

$$[10.94]\ H_0(f,L) = \zeta_0\, e^{-j2\pi f t_c L},$$

avente ritardo di gruppo costante $t_c L$ e attenuazione anch'essa costante:

$$[10.95]\ A_0(f, L) = |H_0(f,L)|^{-2} = |\xi_0|^{-2}\, e^{2\alpha L},$$

seguito da un quadripolo con trasmettenza avente attenuazione unitaria, ma non perfetto in argomento a causa della dispersione intramodale cui corrisponde un valore s_c in generale non nullo:

$$[10.96]\ H_s(f, L) = e^{\mp j(2\pi f \Delta t_s)^2},$$

dove la scelta del segno è determinata dal segno di s_c e si è posto

$$[10.97]\ \Delta t_s \triangleq \sqrt{\frac{|s_c| L}{2}}.$$

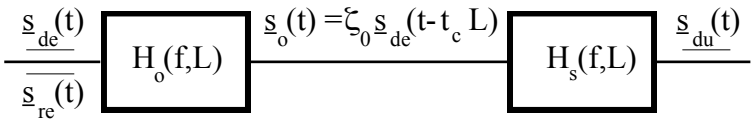


Figura 10.18: Schema equivalente di un tronco di fibra ottica multimodo.

Ammesso che il valore assoluto dell'esponente nella [10.96] risulti minore di 0,5, ossia che sia soddisfatta la condizione:

$$[10.98]\ 2\pi B \Delta t_s < 1,$$??B⇔f??

è lecita l'approssimazione $e^{jx^2} \cong 1 + j2[1 - \cos(x)]$, in modo che la risposta impulsiva del quadripolo non perfetto risulta:

$$[10.99]\ \underline{h}_s(t) \cong F^{-1}\left\{1 \mp j2\left[1 - \cos\left(2\pi f \Delta t_s\right)\right]\right\} = \delta(t) \mp j\left[2\delta(t) - \delta\left(t + \Delta t_s\right) - \delta\left(t - \Delta t_s\right)\right]$$

Indicato con $\underline{s}_0(t) \triangleq \zeta_0 \underline{s}_{de}(t - t_c L)$ l'inviluppo complesso in entrata, si ottiene in uscita:

$$[10.100]\ \underline{s}_{du}(t) = \underline{h}_s(t) * \underline{s}_0(t) = \underline{s}_0(t) \mp j[2\underline{s}_0(t) - \underline{s}_0(t+\Delta t_s) - \underline{s}_0(t-\Delta t_s)],$$

che al di là della deformazione messa in evidenza dal segnale additivo entro parentesi quadra, nel caso di durata limitata di $\underline{s}_0(t)$ rivela un effetto di incremento della durata del segnale in uscita della quantità $2\Delta t_s$, che nell'ambito di validità del modello deve rispettare la [10.98].

10.4.4 Mezzi trasmissivi con fibra ottica monomodo

10.4.4.1 Fibra ottica monomodo

Nella trasmissione di segnali a banda larga tra punti fissi a grande distanza è senz'altro preferibile l'impiego della fibra ottica in regime unimodale, che soffre solo della limitazione comportata dalla dispersione intramodale o cromatica (vedi [10.90]). Per ottenere una bassa attenuazione (vedi [10.95]) si scelgono fibre con dielettrici di elevata qualità: di norma il mantello è in vetro di silice (SiO_2 amorfo) assai puro, con indice di rifrazione $n_2 \cong 1{,}45$ nel vicino infrarosso, e il nucleo è costituito dal medesimo materiale, opportunamente drogato con composti di boro, fluoro, germanio o fosforo allo scopo di ottenere la desiderata differenza relativa degli indici di rifrazione. Operando nella 2ª finestra ($\lambda_c \cong 1.310$ nm) oppure nella 3ª finestra ($\lambda_c \cong 1.550$ nm) si ottengono assai bassi valori della costante di attenuazione α, rispettivamente attorno a 0,35 dB/km e 0,25 dB/km; a causa dell'andamento non rettilineo dell'asse reale della fibra ottica, con raggio di curvatura aleatorio continuamente variabile, la costante di attenuazione può subire incrementi attorno a 0,1 dB/km.

Il rispetto della condizione di unimodalità della propagazione in una fibra ottica a salto di indice, espressa tramite la:

$$[10.101]\ V < V_{t1} = 2{,}405,$$

dove V_{t1} è la frequenza di taglio normalizzata del primo modo superiore, comporta che il raggio del nucleo rispetti la limitazione (vedi [10.66] e [10.67]):

$$[10.102]\ a < \frac{V_{t1}}{k n_2 \sqrt{2\Delta}} \cong 0{,}19 \frac{\lambda}{\sqrt{\Delta}},$$

avendo assunto che il materiale del mantello sia vetro di silice. Con valori della differenza relativa degli indici di rifrazione, Δ, attorno a 3‰ il raggio del nucleo di una fibra ottica monomodo operante in 2ª o 3ª finestra risulta allora attorno a 4 - 5 μm. Il diametro esterno del nucleo è tipicamente di 125 μm. Scelto il valore di λ_c, in 2ª o 3ª finestra, il proporzionamento della fibra ottica a salto di indice risulta in base alle seguenti considerazioni sulla dispersione intramodale o cromatica.

In luogo della grandezza s definita nella [10.90] è di sovente impiegato il *coefficiente di dispersione cromatica* della fibra ottica monomodale:

$$[10.103]\ D(\lambda) \triangleq \frac{dt_i}{d\lambda} = D_{n2}(\lambda) + D_g(\lambda) = -\frac{2\pi c}{\lambda^2} s,$$

semplicemente legato a s come mostrato nell'ultimo termine (vedi appendice A5.5). I due addendi nella espressione [10.90] divengono allora:

$$[10.104]\ D_{n2}(\lambda) \triangleq -\frac{\lambda}{c}\frac{d^2 n_2}{d\lambda^2},\ D_g(\lambda) \triangleq -\frac{n_1 - n_2}{c\lambda} V \frac{d^2(Vb)}{dV^2},$$

e sono rispettivamente denominati *coefficiente di dispersione cromatica di materiale* e *coefficiente di dispersione cromatica di guida*. Sempre nel caso di vetro di silice, il primo coefficiente è nullo attorno a 1.270 nm, è negativo per λ inferiore e positivo per λ superiore, come mostrato dalla curva tratteggiata in Figura 10.19; il secondo coefficiente, che dipende (vedi [10.65] e [10.66]) dalla differenza relativa degli indici di rifrazione, Δ, e dal raggio del nucleo, a, risulta invece

sempre negativo in condizione di unimodalità, come appare anche dalla curva a tratto e punto in Figura 10.19. I valori dei due contributi sono tali da condurre alla esistenza di un punto di zero λ_0, in corrispondenza del quale risulta $D_{n2}(\lambda_0) = - D_g(\lambda_0)$, ossia:

$$[10.105]\ D(\lambda_0) = -s(\lambda_0) = 0,$$

a patto che la lunghezza d'onda sia maggiore di 1.270 nm.

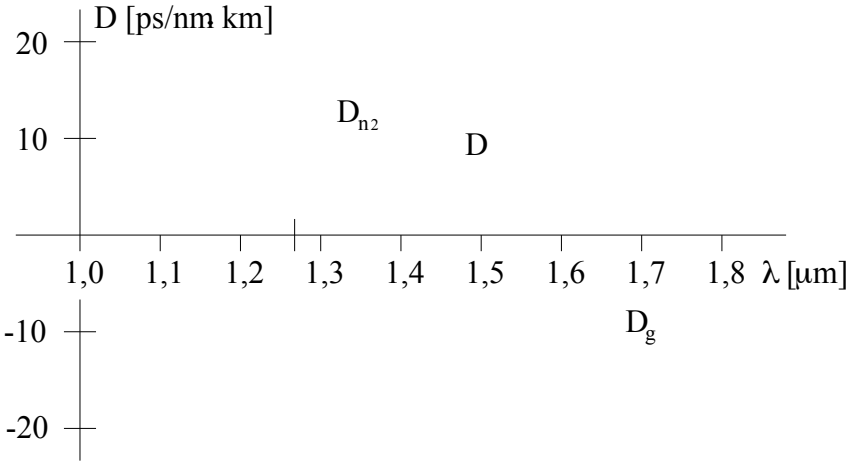


Figura 10.19: Coefficiente di dispersione cromatica in una fibra di vetro di silice e contributi di dispersione di materiale (a tratteggio) e di guida (a tratto e punto).

Poiché le espressioni riguardanti la dispersione sono state ottenute introducendo delle approssimazioni, i piccoli errori che esse comportano non permettono di ricavare con esattezza il punto di zero λ_0. Con metodi di analisi più approfonditi, che tengono anche conto del reale profilo continuo dell'indice di rifrazione nel passaggio dal nucleo al mantello, è comunque possibile proporzionare la fibra ottica, ossia stabilire i valori di Δe di a, in modo da ottenere il soddisfacimento della [10.80], che comporta l'assenza dei considerati effetti di dispersione del secondo ordine. Nel caso di impiego della fibra ottica in 3ª finestra, per ottenere che λ_0 si sposti attorno a 1.550 nm si considerano spesso profili di indice opportuni, ottenendo strutture denominate *fibre ottiche a dispersione spostata* o fibre ottiche DS (Dispersion Shifted).
Una fibra ottica ben progettata per una lunghezza d'onda $\lambda_c \cong \lambda_0$ dà luogo a valori di $D(\lambda_c)$ minori di 3 ps/nm km, che in 3ª finestra conducono a valori di s_c inferiori a $0{,}4\ 10^{-26}\ s^2/m$. Dalla [10.97] con L = 100 km si ottiene allora $\Delta t_s < 14$ ps, che risulta in pratica trascurabile anche a fronte del brevissimo tempo di bit T = 400 ps, corrispondente al ritmo binario di 2,5 Gbit/s.

10.4.4.2 Cavi a fibre ottiche monomodo

10.4.4.3 Mezzi trasmissivi con fibra ottica multimodo

10.4.4.4 Fibra ottica multimodo

proporzionamento del nucleo

10.4.4.5 Cavi a fibre ottiche multimodo

10.5 MEZZI TRASMISSIVI REALI CON PORTANTE RADIO

qualcosa di simile a:

Quando le distanze da coprire tra due terminali fissi siano contenute e i segnali abbiano banda utile non molto estesa, si ricorre spesso all'impiego di un mezzo portante fisico costituito da una coppia di conduttori metallici, isolati tra loro, posati lungo il tracciato del percorso stabilito tra i terminali. Di norma la sezione è invariata, sia nella geometria che nei materiali: un tale tipo di struttura è denominata linea di trasmissione uniforme a coppia metallica o più semplicemente *coppia metallica*. Nel seguito si presuppone che la struttura sia cilindrica lungo lo sviluppo longitudinale, in modo da potere adottare una coordinata rettilinea z; anche se il tracciato effettivo è spesso curvilineo, il suo raggio di curvatura è infatti sufficientemente grande da potere assumere che il comportamento e.m. della linea sia equivalente a quello che si avrebbe qualora essa fosse posata secondo una unica direzione.
Assunto il comportamento perfetto, il mezzo considerato è rappresentabile tramite un quadripolo LTI simmetrico, passivo, adattato, caratterizzato quindi entro la banda utile del segnale trasmesso dalla funzione di trasferimento (vedi [3.4.10]):

[4.4.1] $H(f) = S_{21}(f) = S_{12}(f) = g(L)\, e^{-j2\pi fL/v}$, per $f_m \le |f| \le f_M$,

dove in particolare il modulo della funzione di trasferimento risulta dipendere dalla lunghezza L del mezzo tramite la legge di proporzionalità inversa:

[4.4.2] $g(L) = g(L_1)\, \frac{r_1}{r}$,

essendo L_1 una distanza di riferimento (ad esempio $r_1 = 1$ km). La menzionata specificità è dovuta al meccanismo della trasferimento e.m. (e.m. = elettromagnetico) del segnale lungo il mezzo trasmissivo, che si ammette avvenga per *propagazione libera* di una onda e.m. sferica (vedi appendice A4.5), nello spazio costituito da un dielettrico omogeneo, isotropo e privo di perdite.