

LECTURE NOTES ON COMMUNICATION TECHNOLOGIES

LECTURE NOTES ON COMMUNICATION TECHNOLOGIES with a spray of Signals and Systems

Marco Luise

University of Pisa, Dipartimento di Ingegneria dell'Informazione



Una pubblicazione dell'UNIVERSITÀ DI PISA

PREFAZIONE



”But surpassing all stupendous inventions, what sublimity of mind was his who dreamed of finding means to communicate his deepest thoughts to any other person, though distant by mighty intervals of place and time?”

— *Galileo Galilei, Dialogo sopra i due massimi sistemi del mondo*

Florence 1632, trans. by Albert Van Helden

Indice

Prefazione	v
1 1-2-3 of Signals and Systems	1
1.1 Basics of Fourier analysis of analog signals	2
1.1.1 Periodic Signals and the Fourier Series	2
1.1.2 Non-periodic Signals and the Fourier Transform	5
1.1.3 Bandlimited Signals	7
1.1.4 Dirac's delta function	9
1.2 Linear Filtering	10
1.2.1 Systems and Signals	10
1.2.2 Time- and Frequency-Characterization of LTI Systems	11
1.3 Filtering of Random Signals	14
1.3.1 Basics of random signals	14
1.3.2 Expectation, Autocorrelation function, and Power Spectral Density	14
1.4 Bandpass Signals and Systems	19
1.4.1 Baseband equivalent of a bandpass signal	19
1.4.2 The I-Q modulator	20
1.4.3 The I-Q demodulator	21
1.5 Fourier analysis of digital signals	22
1.5.1 Analog and Digital signals	22
	vii

1.5.2	FT of Digital signals	22
1.5.3	Filtering and Interpolation of digital signal	25
1.5.4	The sampling theorem	28
2	Basics of Wireless Communications Engineering	31
2.1	Baseband Transmission of Digital Signals	32
2.1.1	NRZ and bandlimited digital signals	32
2.1.2	Signals and spectra	33
2.1.3	Detection of a digital signal on the AWGN channel	35
2.2	Modulation, Demodulation and Modem Architecture	40
2.2.1	Linear I/Q digital modulation	40
2.2.2	Signal constellations	40
2.2.3	Demodulation of Linear I/Q modulated signals	42
2.2.4	Architecture of DSP-based Data Demodulators	44
2.3	Multiple Access for Wireless Communications	45
2.3.1	Multiplexing and Multiple Access	47
2.3.2	Multiplexing with Orthogonal Signals: FDM, TDM, and CDM	47
2.3.3	FDMA, TDMA, and CDMA	53
2.3.4	Efficiency of Multiple Access Techniques	58
2.4	Cellular Radio	62
2.4.1	Description of a cellular network	63
2.4.2	Cell Planning	65
2.4.3	Summing up...	71
2.5	CDMA and Universal Frequency Reuse	71
2.6	Fractional Frequency Reuse	72
3	Technologies for the "Last Kilometer" (Last Mile)	73
3.1	Multicarrier meets Information Theory: DSL Technologies and Shannon Capacity of the Additive Colored Gaussian Channel	74
3.1.1	xDSL system architecture	74
3.1.2	xDSL Frequency Plan and Communication Channel Model	75
3.2	Shannon Capacity of the ACGN Channel	78
3.2.1	Power allocation and the Water Filling Criterion	81
3.2.2	Why is "water-filling" called like that?	83
3.3	DMT (Discrete MultiTone) Modulation Format	84
3.4	Fiber-Based UltraBroadBand (UBB) Connectivity fo the Access Network	87
3.4.1	The Migration from Copper to Fiber in the Last Kilometer	87
3.4.2	FTTH Passive Optical Networks	87
3.4.3	The dominant technology for FTTH: ITU's Gigabit PON (G-PON)	90
	Bibliografia	93

CAPITOLO 1

1-2-3 OF SIGNALS AND SYSTEMS



“Le système Chappe: The world’s first (wireless) telegraph network developed by Napoleon, carrying digitally-encoded text messages across 19h-Century France via optical signals relayed by a network of repeating stations extended from Lyon to Venice”

—*Optical telegraph station next to the Rohan Castle in Saverne, France, renovated in 1998*

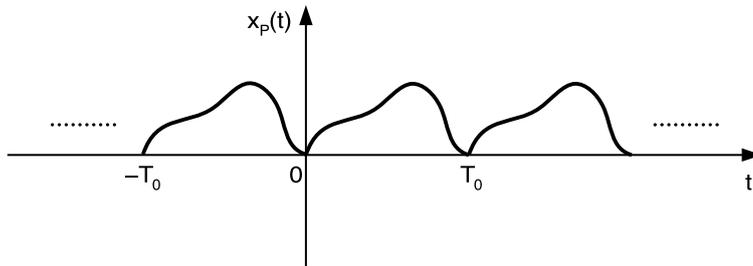


Figure 1.1 Example of a periodic signal

The motto of this chapter is: back to the basics. Its aim is in fact reviewing the main concepts related to the nature and processing of analog and digital signals, something that the experienced reader may either skip with no harm or read with renovated pleasure.

1.1 Basics of Fourier analysis of analog signals

The “Swiss Knife” of every communications engineer dealing with (wireless) systems design is *Fourier analysis*. We do not pretend to perform a comprehensive review of such a huge and fundamental topic. We just want to re-state here the main results and settle a notation concerning the analysis of time-continuous and time-discrete signals that will be the foundation of many, many concepts and tools we will extensively use in the next Chapters.

1.1.1 Periodic Signals and the Fourier Series

We’ve been taught back in primary schools that white light is a combination of all colors. This is a very first example of Fourier analysis. To be a little bit more specific, we know that every periodic signal $x_p(t)$ (i.e., such that $x_p(t) = x_p(t + T_0)$ for some $T_0 > 0$ that is called *repetition period* as in Fig. 1.1) can be decomposed into a sum of simpler periodic signals, namely, sinusoids as follows:

$$x_p(t) = A_0 + A_1 \cos(2\pi f_0 t + \theta_1) + A_2 \cos(2\pi 2f_0 t + \theta_2) + \dots \\ + A_k \cos(2\pi k f_0 t + \theta_k) + \dots \quad (1.1)$$

Apart from the constant, DC value, A_0 , it is seen that the sinusoids oscillate at frequencies $k f_0$, the so-called *harmonic frequencies*, that are integer multiples of the *fundamental* or *repetition* frequency $f_0 = 1/T_0$. The k -th component of the expansion (1.1) bears an amplitude A_k , and a phase θ_k . Whilst the value of the oscillation frequencies are always the same for any T_0 -periodic signal, the specific values of A_k and θ_k do depend on the shape of the actual $x_p(t)$ under analysis. The sequence of coefficients A_k is called the *amplitude spectrum* of $x_p(t)$, and the sequence of the phases θ_k is the *phase spectrum*. Knowledge of the amplitude and phase spectra is equivalent to knowledge of the signal itself, since based on that knowledge we can from (1.1) *synthesize* back the signal in the time domain. This is why (1.1) is called the *synthesis* equation. And this is why the popular musical instrument MiniMoog of the glorious 70’s was called a *synthesizer*: it simply implemented (1.1) by



Figura 1.2 The MiniMoog

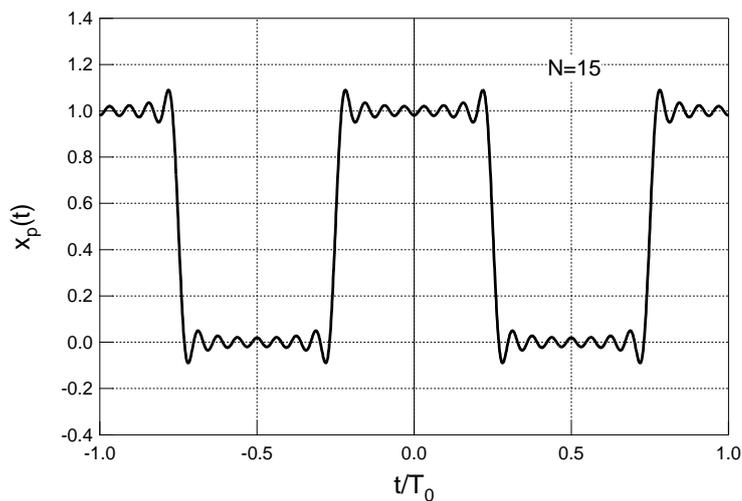


Figura 1.3 Synthesis of a rectangular pulse train with a finite number $N=15$ of sinusoidal components

heroic low-cost analog hardware to generate (arbitrary) periodic waveforms to be used in musical compositions as a replacement of naturally-generated sounds.

Equation (1.1) is the simplest form of a *Fourier series* for a periodic signal. In a sense, the pitfall of such representation is that exact synthesis of the periodic signal theoretically requires an *infinite* number of components. Nonetheless, such a representation can be used in the practice by truncating the series to a (small) number of significant components only, just as the MiniMoog used to do. Figure 1.3 shows how a periodic rectangular pulse train can be synthesized by the superposition of a finite number of elementary sinusoidal components, according to (1.1). Of course, given a certain waveform that we intend to synthesize, the problem is: what are the correct values of A_k and θ_k to be used in our synthesis equation? Giving a response to this question means *analyzing* signal $x_p(t)$ by

means of a proper *analysis equation* that we are to find. This is most easily done by resorting to a complex-number representation of the Fourier series. The key to such representation is Euler's formula for the cosine function:

$$A_k \cos(2\pi k f_0 t + \theta_k) = \frac{1}{2} [A_k \exp(j2\pi k f_0 t) \cdot \exp(j\theta) + A_k \exp(-j2\pi k f_0 t) \cdot \exp(-j\theta)] \quad (1.2)$$

The real-valued oscillating function is decomposed as the sum of two *rotating vectors* on the complex plane. The first one, $\exp(j2\pi k f_0 t)$ rotates counterclockwise with a frequency f_0 cycles/s (Hz), and the second one, $\exp(-j2\pi k f_0 t)$, rotates (counterclockwise) at the *negative frequency* $-f_0$ Hz. The sum of the two complex rotating vectors gives just the real-valued sinusoidal oscillation we started from. This complex decomposition entails the introduction of (complex) signal components with negative frequencies. The amplitude and phase spectra of the sinusoid are collapsed into a single complex-valued coefficient $X_k = A_k \exp[j\theta_k]$ (k positive) that is called the *Fourier coefficient* of $x_p(t)$. Elaborating (1.1) with (1.2), we get the following expression of the Fourier series containing the complex Fourier coefficients X_k :

$$x_p(t) = \sum_{k=-\infty}^{\infty} X_k \exp(j2\pi k f_0 t) \quad (1.3)$$

that is exactly equivalent to the real-valued form (1.1). Finding the amplitude and phase spectra A_k and θ_k is tantamount to finding the k -th Fourier coefficient X_k . With some effort, it is found that the *analysis equation* we were looking for is relatively simple:

$$X_k = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x_p(t) \exp(-j2\pi k f_0 t) dt \quad (1.4)$$

Esempio 1.1

Let us analyze the *pulse train* $x_p(t)$ we tried to synthesize in Fig. 1.3. Its Fourier coefficient is given by

$$X_k = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x_p(t) \exp(-j2\pi k f_0 t) dt = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x_p(t) \cos(2\pi k f_0 t) dt \quad (1.5)$$

where we have exploited the even-symmetry of our waveform. Now, considering that $x_p(t)$ is piecewise-constant, we also have

$$X_k = \frac{1}{T_0} \int_{-T_0/4}^{T_0/4} \cos(2\pi k f_0 t) dt = \frac{1}{T_0} \frac{\sin(2\pi k f_0 t) \Big|_{-T_0/4}^{T_0/4}}{2\pi k f_0} = \begin{cases} 1/2 & k = 0 \\ 0 & k = 2m, m \neq 0 \\ \frac{(-1)^m}{\pi k} & k = 2m + 1 \end{cases} \quad (1.6)$$

The Fourier coefficients X_k turns out to be real-valued (due to the even symmetry of $x_p(t)$); the resulting amplitude line spectrum of $x_p(t)$ is shown in Fig. 1.4.

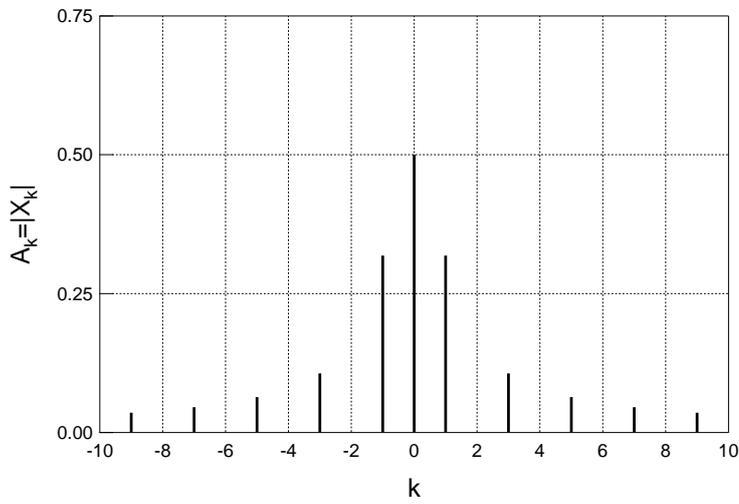


Figure 1.4 Amplitude spectrum of a rectangular pulse train

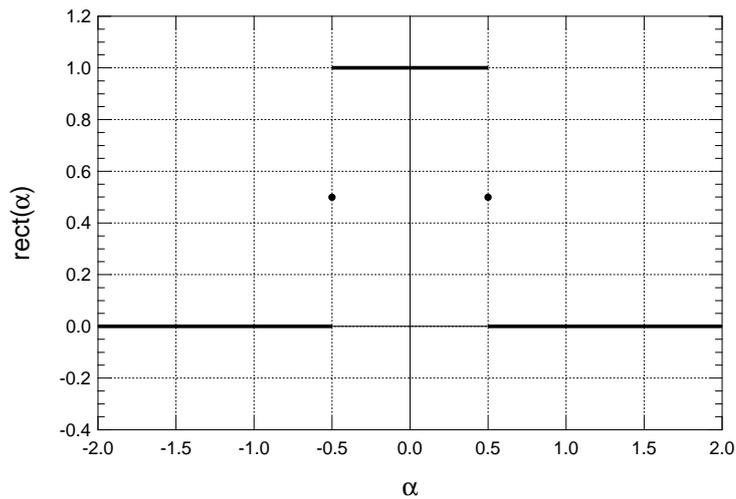


Figure 1.5 rect signal

1.1.2 Non-periodic Signals and the Fourier Transform

What was said until now was only applicable to *periodic* signals. What happens with impulsive signals $x(t)$ that are not periodic? Figure 1.5 shows a *rectangular pulse* that we define as follows:

$$\text{rect}(t/T) \triangleq \begin{cases} 1 & |t| \leq T/2 \\ 1/2 & |t| = T/2 \\ 0 & \text{elsewhere} \end{cases} \quad (1.7)$$

The question is: is this kind of signal amenable to Fourier analysis/synthesis? The answer

is (of course) positive, and how to do it can be found quite easily if we think of a non-periodic signal as a periodic signal with *infinitely-long* repetition period. The fundamental frequency thus becomes vanishingly small (infinitesimal), and so two components in the frequency spectrum of the signal that were previously separated by $\Delta f = (k + 1)f_0 - kf_0 = f_0$ now become infinitesimally close to each other. The frequency spectrum of the signal that used to be *discrete* (the line spectrum of Fig. 1.4 with a frequency “quantum” given by the fundamental frequency f_0) now becomes *continuous*. The relevant analysis equation turns out to be now

$$X(f) = \int_{t=-\infty}^{\infty} x(t) \exp(-j2\pi ft) dt \quad (1.8)$$

where the frequency f that appears as the argument of this complex-valued quantity $X(f)$ takes all *real* values with continuity.

This counterpart of the former Fourier coefficient is called the *Fourier Transform* (FT) of $x(t)$ and bears the same meaning as X_k . It has an amplitude $|X(f)|$ and a phase $\angle X(f)$, so that we still speak of (continuous) amplitude and phase spectra, respectively. Since the spectrum is now continuous, the synthesis equation cannot be a series any more, rather it is expressed in the form of a *Fourier Integral*:

$$x(t) = \int_{f=-\infty}^{\infty} X(f) \exp(j2\pi ft) dt \quad (1.9)$$

This relation is also called the *Inverse Fourier Transform* (IFT) of $X(f)$. The physical meaning that we can attach to the pair of relations (1.8)-(1.9) is the same as with the Fourier coefficient-series pair: the IFT is a synthesis equation that tells us how to build our own signal starting from a set of simpler components (the complex-valued sinusoids), and the FT tells us the specific values of the amplitude and phase of each sinusoid that has to be used in our synthesis procedure to build a specific signal. We can show easily that the FT $X(f)$ of a real-valued $x(t)$ (that we will denote at times $\mathcal{F}[x(t)]$) has a particular kind of symmetry that is called *Hermitian*: $X(-f) = X^*(f)$, i.e, $|X(-f)| = |X(f)|$, and $\angle X(-f) = -\angle X(f)$.

We will not waste any precious space in describing the many features of the FT as a tool for signal design and analysis in communications engineering. We just want here to recall some elementary results about FT theory that will be used in many places in the Chapters to follow. For instance, it is an easy exercise for the reader to show that the FT of the time-shifted version $x(t - t_0)$ of the signal $x(t)$ is

$$\mathcal{F}[x(t - t_0)] = X(f) \exp(-j2\pi ft_0) \quad (1.10)$$

so that the amplitude spectrum of the signal is left unchanged, and the phase spectrum is modified by a term proportional to the frequency of each component. Similarly, it is easy to show what happens if we perform an operation of radio-frequency *modulation* on the signal $x(t)$ as follows:

$$x_{RF}(t) = x(t) \cos(2\pi f_0 t) = \frac{x(t) \exp(j2\pi f_0 t) + x(t) \exp(-j2\pi f_0 t)}{2} \quad (1.11)$$

where f_0 is the *carrier frequency*. The corresponding modification of the FT is

$$X_{RF}(f) = \frac{X(f + f_0) + X(f - f_0)}{2} \quad (1.12)$$

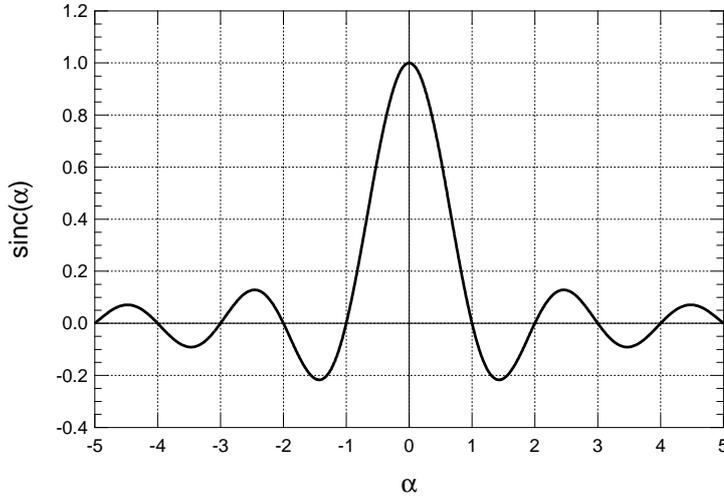


Figura 1.6 sinc function

that is, a frequency-shift of each of the frequency components the modulating signal is made of.

Esempio 1.2

Let us find the FT of the rect function $x(t)$ in (1.7):

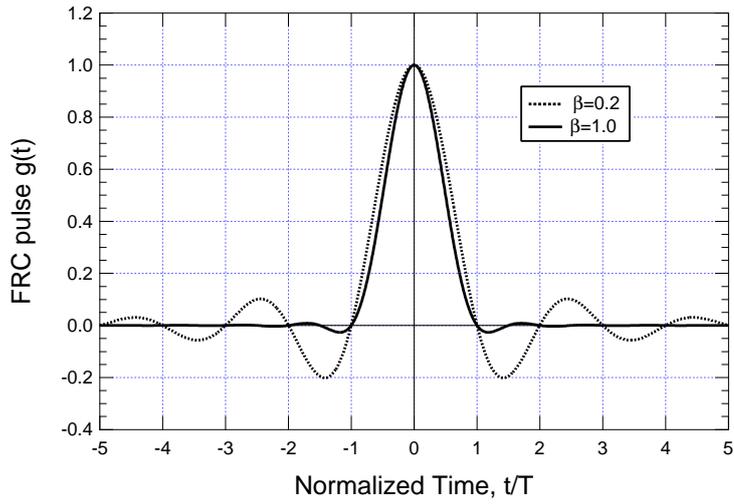
$$\begin{aligned} X(f) &= \int_{-\infty}^{+\infty} \text{rect}(t) \exp(-j2\pi ft) dt = \int_{-T/2}^{+T/2} \exp(-j2\pi ft) dt \\ &= \frac{\sin(\pi fT)}{\pi f} = T \frac{\sin(\pi fT)}{\pi fT} = T \text{sinc}(fT) \end{aligned} \quad (1.13)$$

We have introduced here a new identifier for a special waveform that we will extensively use in the following: the so-called sinc function (represented in Fig.1.6) that we define as $\text{sinc}(\alpha) = \sin(\pi\alpha)/(\pi\alpha)$.

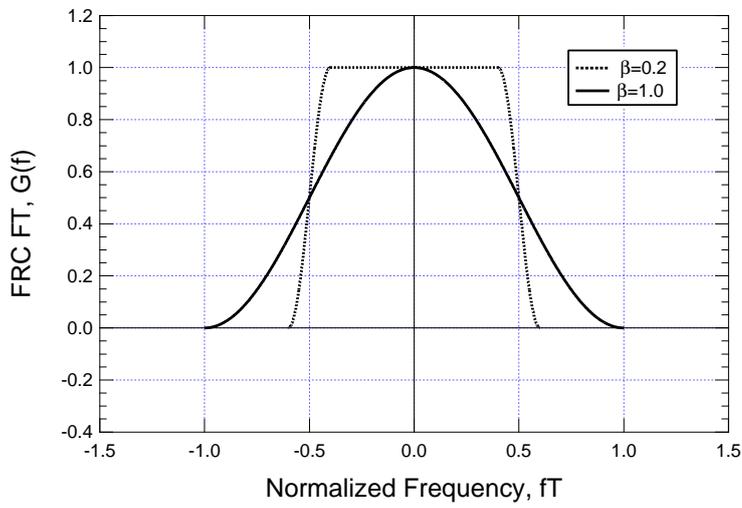
1.1.3 Bandlimited Signals

We will say that a signal is *bandlimited* when it has an amplitude spectrum $|X(f)|$ that is confined into a limited frequency band $[-B, B]$. Such limitation may hold exactly (*strictly* bandlimited signal) or to a good approximation, in the sense that out of the interval $[-B, B]$ the signal components, though not exactly null, are so small as to be considered negligible. An example of a (strictly) bandlimited signal is the popular *Frequency Raised Cosine* (FRC)pulse (or *Nyquist's pulse*) given by

$$g_N(t) = \text{sinc}(t/T) \frac{\cos(\beta\pi t/T)}{1 - (2\beta t/T)^2}$$



(a)



(b)

Figure 1.7 Waveform (a) and Fourier Transform (b) of the FRC pulse

$$G_N(f) = \begin{cases} T & |f| < (1 - \beta)/2T \\ \frac{T}{2} \left\{ 1 + \cos \left[\frac{\pi T}{\beta} \left(|f| - \frac{1-\beta}{2T} \right) \right] \right\} & (1 - \beta)/2T \leq |f| \leq (1 + \beta)/2T \\ 0 & \text{elsewhere} \end{cases} \quad (1.14)$$

and whose spectrum/waveform are represented in Fig. 1.7. Here, the bandwidth is $B = (1 + \beta)/2T$, and β , $0 \leq \beta \leq 1$ is a parameter that regulates the signal bandwidth and that is called *roll-off factor*. The value $1/2T$ is the so-called *Nyquist frequency*, corresponding to the minimum pulse bandwidth when $\beta = 0$.

1.1.4 Dirac's delta function

A peculiar signal that we will often use in the following chapters is Dirac's *delta function* $\delta(t)$. Its name is a little bit defying, since $\delta(t)$ is not actually a signal in the classical sense. We may speak of a *generalized signal* whose definition and existence is only justified through an integral property. Dirac's delta is in fact defined through the so-called *sampling* or *sifting* property:

$$\delta(t) : \int_{-\infty}^{-\infty} x(t)\delta(t)dt = x(0) \quad (1.15)$$

where $x(t)$ is any ordinary signal with no discontinuity at $t = 0$. From (1.15) we immediately have as a particular case

$$\int_{-\infty}^{-\infty} \delta(t)dt = 1 \quad (1.16)$$

so we may say that the Dirac's function has "unit area". It is easily argued that no ordinary function with property (1.15) exists, but this new mathematical entity proves useful in system theory and linear filtering, as we'll see in a while.

The standard heuristic representation of the delta function (also called *unit impulse*), whose rigorous treatment is found within the so-called *distributions theory*, can be obtained with the aid of a sequence of functions. Assume we have a rect function with duration $T = 2\varepsilon$ and amplitude $A = 1/2\varepsilon$ as represented in Fig. 1.8(a). The "area" of this signal is 1, irrespective of ε . Assume now that this pulse is made shorter and shorter (and consequently, taller and taller) keeping its unit area but becoming thinner and thinner, as suggested in Fig. 1.8(a). The limit of this pulse is a heuristic representation of $\delta(t)$: something whose time width is null, but whose amplitude is infinite, so that its area is unitary. This is what is symbolically depicted in Fig. 1.8(b) as the standard representation of a delta "function". Of course, such a signal does not exist in the ordinary sense.

The definition of $\delta(t)$ that follows our heuristic representation is

$$\delta(t) \triangleq \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \text{rect} \left(\frac{t}{2\varepsilon} \right) \quad (1.17)$$

This relation not only gives an idea about how the delta function "looks like", but can also be used in the practice, provided that i) $\delta(t)$ appears under an integral operator (as in its definition (1.15)), and ii) the limit in (1.17) is moved *outside* the integral operator, i.e., it is computed *subsequently* to the computation of the integral. The reader may verify (1.16) using this new definition. It is also easy to show that the definite integral of $\delta(t)$ on finite intervals of the kind $\int_b^a \delta(t)dt$ gives a value equal to 1 when the instant $t = 0$ lies within (a, b) , otherwise it gives 0.

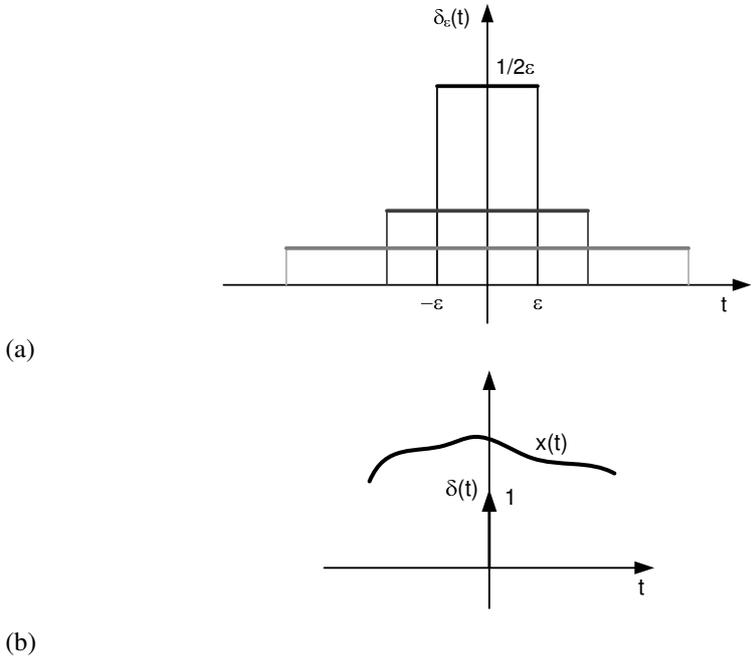


Figure 1.8 Definition of Dirac's function (a), and its symbolic representation (b)

Dirac's delta is also peculiar as far as its FT $\Delta(f)$ is concerned. First, the problem of finding the FT of $\delta(t)$ is well-posed since the FT (1.8) is an integral operator. Second, its computation is trivial, according to (1.15):

$$\Delta(f) = \int_{t=-\infty}^{\infty} \delta(t) \exp(-j2\pi ft) dt = \exp(-j2\pi ft)|_{t=0} = 1 \quad (1.18)$$

The FT of the unit impulse is thus *constant* on all frequencies, with no bandlimitation whatever.

1.2 Linear Filtering

In the following Chapters, we will familiarize with a number of signal processing functions that are implemented in a digital data receiver for wireless communications.

1.2.1 Systems and Signals

The simplest and most fundamental of such operations is perhaps *filtering*. In its simplest realization, filtering means designing a device, an electronic circuit, a piece of software or, in a word, a *system* that changes an *input* signal $x(t)$ into an *output* signal $y(t)$ according to some processing criteria. Our notation to indicate this will be

$$y(t) = \mathcal{T}[x(\alpha); t] \quad (1.19)$$

where \mathcal{T} is an operator representing the signal processing function performed by the system, and where we indicate that the processing depends in general on the *whole* input waveform

$x(\alpha)$ and on time as well. The simplest family of system are the *linear filters* that obey the *superposition rule*. Assume that we know that

$$y_1(t) = \mathcal{T}[x_1(\alpha); t] \quad , \quad y_2(t) = \mathcal{T}[x_2(\alpha); t] \quad (1.20)$$

and that we build up a signal $x(t)$ as a weighted superposition (i.e., a linear combination) of x_1 and x_2 as $x(t) = \alpha_1 x_1(t) + \alpha_2 x_2(t)$. The system is a *linear filter* iff

$$y(t) = \mathcal{T}[x(\alpha); t] = \mathcal{T}[\alpha_1 x_1(\alpha) + \alpha_2 x_2(\alpha); t] = \alpha_1 y_1(t) + \alpha_2 y_2(t) \quad (1.21)$$

This means that the output $y(t)$ can be obtained as a linear combination with the same coefficients α_1 and α_2 of the “single” outputs of the system to the single inputs x_1 and x_2 .

In addition to the property of linearity many filters used in the practice are also *time-invariant*. This means that their behavior does not change with time. Specifically, if we know that $y(t) = \mathcal{T}[x(\alpha); t]$, and we later submit to the system a time-shifted version of the same signal, namely, $x_{TS}(t) \triangleq x(t - t_0)$, we expect that the filter output $y_{TS}(t)$ be just the same waveform that we had earlier, modified only by the same time shift we introduced on the input:

$$y_{TS}(t) = \mathcal{T}[x_{TS}(\alpha); t] = \mathcal{T}[x(\alpha - t_0); t] = y(t - t_0) \quad (1.22)$$

Linear, Time-Invariant (LTI) systems are particularly simple to design and analyze, nonetheless they prove useful in many instances of signal processing, and especially for the detection of known signals embedded into noise.

1.2.2 Time- and Frequency-Characterization of LTI Systems

The properties of any LTI filter are completely characterized by the knowledge of a particular signal that is associated to the system, namely, its *impulse response*. The impulse response of an LTI system is just the system output in response to a $\delta(t)$ input:

$$h(t) \triangleq \mathcal{T}[\delta(\alpha); t] \quad (1.23)$$

In particular, it can be easily shown that the response (output) of an LTI to a generic input $x(t)$ can be found as

$$y(t) = \int_{-\infty}^{\infty} x(\alpha) h(t - \alpha) d\alpha = x(t) \otimes h(t) \quad (1.24)$$

This kind of “mixing” of the two signals x and h to give y deserves a specific name and notation: it is called the *aperiodic convolution* between the two signals, and is denoted by the symbol \otimes as in (1.24). The convolution is a symmetric, associative, and distributive operator, as can be easily proved. The transformation carried out by an LTI system on its input signal $x(t)$ to give the output signal $y(t)$ is often symbolized in a graphical form as in Fig. 1.9 where the impulse response of the system is explicitly indicated.

Fourier analysis of a signal reveals a fundamental tool also in the characterization of the properties of an LTI system. Everything revolves around a cardinal property of convolution. Starting back from the constituent relation (1.24) of an LTI system, and taking the FT of both sides of the equation, we find:

$$Y(f) = X(f) \cdot H(f) \quad (1.25)$$

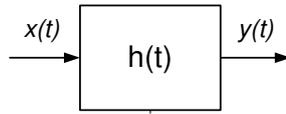


Figure 1.9 Graphical representation of the transformation effected by an LTI system

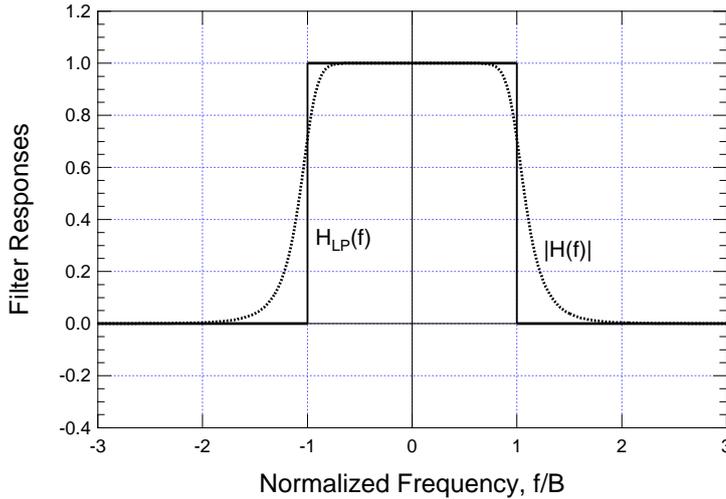


Figure 1.10 Amplitude response of lowpass filters

where $H(f)$ is the FT of the impulse response $h(t)$. It is seen that a (complicated) operation of convolution in the time domain has changed into a (simple) operation of product between two transforms in the frequency domain. This suggests that often the analysis and design of an LTI system is much simpler in the frequency domain than in the time domain due to the simpler operation that the system implicitly carries out on the FT of the input signal. The quantity $H(f)$ is called the *frequency response* of the filter, and is an alternative means to provide a full characterization of the behavior of the system. In particular, it can be shown that the response of the system to a purely sinusoidal input $x(t) = \cos(2\pi f_0 t)$ is just another sinusoidal signal at the same frequency in the form

$$x(t) = \cos(2\pi f_0 t) \Rightarrow y(t) = |H(f_0)| \cdot \cos(2\pi f_0 t + \angle H(f_0)) \quad (1.26)$$

The effect of the filter on the sinusoidal signal is an amplitude change by a factor equal to $|H(f_0)|$ (the *amplitude response* of the system at the frequency f_0), and a phase shift by $\angle H(f_0)$ (the *phase response*). The complex-valued version of (3.1) is

$$x(t) = \exp(j2\pi f_0 t) \Rightarrow y(t) = H(f_0) \cdot \exp(j2\pi f_0 t) \quad (1.27)$$

If we change the frequency of the sinusoidal signal, the amplitude and phase responses change, and so our system responds to different components at different frequencies in a different way. This is why $H(f_0)$ is called the frequency response, and also suggests that in general the system bears a *selective* (i.e., unequal) behavior with respect to frequency. Different components in the spectrum of a signals are treated differently. Some may pass substantially unaltered, others may be blocked altogether. An example of such behavior is

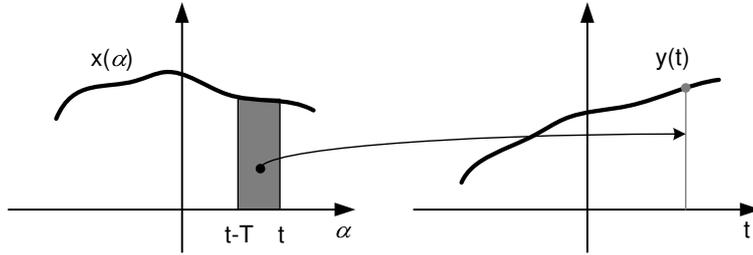


Figura 1.11 Sliding-window Integrator

given by the *low-pass* ideal filter, whose frequency response $H_{LP}(f)$ is represented in Fig. 1.10 (solid line). The “low” frequency components are passed unaltered (they are multiplied by $H_{LP}(f) = 1$), whilst higher-frequency components, in particular those outside the passband B are blocked. The same is substantially true for the system whose amplitude response $|H(f)|$ is also represented in Fig. 1.10 (dashed line). We can not call this an “ideal” filter, but it is nonetheless a good approximation of the ideal filter we have just mentioned, and it is easily realizable.

Esempio 1.3

We intend to find the impulse and the frequency response of a *sliding window* integrator. This LTI system produces an output whose value at a specific instant is the value of the integral of the input signal on a T -wide integration window immediately leading that instant, as is shown in Fig. 1.11:

$$h(t) = \frac{1}{T} \int_{t-T}^t x(\alpha) d\alpha \quad (1.28)$$

We leave to the reader the proof that (1.31) defines indeed an LTI. Assuming this, the impulse response is by definition the output of the system when the input is $\delta(t)$ and is given by

$$y(t) = \frac{1}{T} \int_{t-T}^t \delta(\alpha) d\alpha \quad (1.29)$$

As we know by the properties of Dirac’s δ function, the integral above is either equal to 1 if the instant $\alpha = 0$ (where $\delta(\alpha)$ is applied) is within the integration interval, otherwise it is equal to 0. This means that the output value is going to be 1 whenever $t - T \leq 0 < t$, that is to say, $0 < t \leq T$, and is going to be 0 outside this interval. The impulse response we seek for is a causal rectangular pulse that we can cast into the form

$$h(t) = \frac{1}{T} \text{rect} \left(\frac{t - T/2}{T} \right) \quad (1.30)$$

The frequency response is trivially the Fourier transform of $h(t)$, namely,

$$H(f) = \text{sinc}(fT) \exp(-j\pi fT) \quad (1.31)$$

1.3 Filtering of Random Signals

Random signals or, in the parlance of probability theory, *random processes* play a central role in communications theory. Not only a random process is the standard mathematical representation of all kinds of noise, interference, and disturbance of any sort that may afflict a signal to be detected. A random process is also the representation for the information-bearing signal itself: were the signal known in advance to the receiver, no need of sending it out by the transmitter would arise. The randomness of such signal is just related to the quantity of information that it contains: the more random it looks to the receiver *before* being received, the larger quantity of information it conveys *after* it is actually received.

1.3.1 Basics of random signals

A random signal (process) is a function of time whose shape is not (exactly) known in advance. The value at time t of a deterministic signal $w(t)$ like those encountered until now in the book is exactly known for every possible t , either because we have a mathematical representation of such signal, or because we have recorded it and stored it in a computer file. On the contrary, we can say that the value at a certain time t of a random signal is a *random variable*. The random process is thus the collection of all random variables at all times t , that we denote with the same notation as an ordinary deterministic signal, $w(t)$.

So we do not know in advance the value of the signal for each value of t . Rather, we only know *statistical properties* of such values. When we observe a random signal, what we get is a *realization* of all such random variables time after time, and, after our observation, we are left with a deterministic signal that could not be predicted in advance. The statistical description that we need to completely characterize the properties of such signal may appear to be the probability density function (pdf) $f_w(a; t)$ of the random variable $w(t)$ at time t , that allows to compute probabilities of any kind on $w(t)$. That is not the case indeed. For such signal it may be needed to know, just to make an example, the probability that $w(t) \leq 0$ and, together with this, that after a certain time τ , the probability that $w(t+\tau) \geq 0$. This *joint* probability cannot be computed from $f_w(a; t)$. What we need here is a *second-order* joint pdf $f_w(a_1, a_2; t, t + \tau)$. But then, why stopping just at the second order? The conclusion is that the complete characterization of a random process requires the knowledge of the *class* of joint pdf's of order K of the form

$$f_w(a_1, a_2, \dots, a_K; t, t + \tau_1, \dots, t + \tau_{K-1}) \quad (1.32)$$

for *each* (arbitrary large) K . This piece of knowledge is clearly very difficult to get, apart from simple cases. One such exception is that of *Gaussian processes*, for which the pdf's (1.32) are (jointly) Gaussian for any K .

1.3.2 Expectation, Autocorrelation function, and Power Spectral Density

In the practice, a very partial knowledge of the statistical properties of a random process may be sufficient to solve many problems in communications engineering. The simplest statistical property of a random process is its *expectation* function or simply *mean value*:

$$\eta_w(t) \triangleq \int_{-\infty}^{\infty} a f_w(a; t) da \quad (1.33)$$

This function represents, time instant by time instant, the most likely value that the process is going to assume before it is actually observed. In particular, the relation (1.33) is often

summarized into the following simplified notation:

$$\eta_w(t) = E\{w(t)\} \quad (1.34)$$

where we introduced the linear operator $E\{\cdot\}$ called *Expectation* whose definition is self-evident. In many examples of random signals that are dealt with in communications systems (be them information-bearing or just noise), we will see that $\eta_w(t)$ is a constant, and very very often $\eta_w(t) \equiv \eta = 0$.

Esempio 1.4

We are given the random process

$$w(t) = A \cos(2\pi f_0 t + \Theta) \quad (1.35)$$

where A and f_0 are known, while Θ is a uniform random variable in the interval $[-\pi/2, \pi/2]$. Let us find first the average value $\eta_w(t)$. According to the expectation theorem, instead of using $f_w(a; t)$ to perform the expectation we need, we can use the statistics of the parameter Θ the process depends on:

$$\begin{aligned} \eta_w(t) &= E\{w(t)\} = E\{A \cos(2\pi f_0 t + \Theta)\} \\ &= \int_{-\infty}^{-\infty} A \cos(2\pi f_0 t + \theta) f_{\Theta}(\theta) d\theta = A \int_{-\pi/2}^{\pi/2} \cos(2\pi f_0 t + \theta) \frac{1}{\pi} d\theta \\ &= -\frac{2A}{\pi} \sin(2\pi f_0 t) \end{aligned} \quad (1.36)$$

The average value $\eta_w(t)$ of $w(t)$ is a *first-order* statistical quantity, or first-order statistics. It is computed with a first-order pdf, and gives information of the statistical behavior of our random signal when observed at a *single* time instant. Another example of first-order statistics is the average instantaneous power of the process

$$P_w(t) \triangleq E\{w^2(t)\} = \int_{-\infty}^{-\infty} a^2 f_w(a; t) da \quad (1.37)$$

The main example of a *second* order statistics is the *autocorrelation function*

$$\begin{aligned} R_w(t, \tau) &\triangleq E\{w(t)w(t - \tau)\} \\ &= \int_{a_1=-\infty}^{-\infty} \int_{a_2=-\infty}^{-\infty} a_1 \cdot a_2 f_w(a_1, a_2; t, t - \tau) da_1 da_2 \end{aligned} \quad (1.38)$$

It is computed as the *statistical correlation* between the two random variables $w(t)$ and $w(t - \tau)$ that the random process takes at the two instants t and $t - \tau$, respectively. The autocorrelation function plays a fundamental role in the spectral analysis of a random signal, just as it does for deterministic signals. For deterministic signals, we know that it is a function of the delay τ only, whereas, according to (1.38), $R(t, \tau)$ is a function of both t and τ . We introduce therefore the notion of a wide-sense *stationary* (WSS) process. Assume that

$$R_w(t, \tau) = R_w(\tau) \quad , \quad \eta_w(t) \equiv \eta_w \quad (1.39)$$

We may say that the properties of the random process are always the same, irrespective of the time instant that we choose as our time origin: the average value does not depend on time, and the autocorrelation function depend only on the *time shift* between the two time instant $t_1 = t$ and $t_2 = t - \tau$ that we consider on our signal. The instantaneous power of a WSS process is constant in time and has the following relation with the autocorrelation function:

$$P_w(t) \equiv P_w = R_w(0) \quad (1.40)$$

For WSS processes, we may define a *power spectral density* (psd) function as the FT of the autocorrelation function:

$$S_w(f) \triangleq \int_{-\infty}^{+\infty} R_w(\tau) \exp(-j2\pi f\tau) d\tau = 2 \int_0^{+\infty} R_w(\tau) \cos(2\pi f\tau) d\tau \quad (1.41)$$

where the second equality comes from the fact (easy to show) that $R_w(\tau) = R_w(-\tau)$. The psd function indicates how the power associated to a certain signal is distributed on the different component of the frequency spectrum, and the total signal power can be found as

$$P_w = \int_{-\infty}^{+\infty} S_w(f) df \quad (1.42)$$

As with any random variable, the power of a WSS process can be evaluated as $P_w = \sigma_w^2 + \eta_w^2$.

Esempio 1.5

Take back into consideration the process we defined in Example 4, but assume that Θ is uniform in $[-\pi, \pi)$. The average value is now easily found to be 0, irrespective of time. We may wonder whether the process is WSS. To possibly verify this, we compute the autocorrelation function $R_w(t, \tau) = E\{w(t)w(t - \tau)\}$ by virtue of the expectation theorem:

$$\begin{aligned} R_w(t, \tau) &= E\{\cos(2\pi f_0 t + \Theta) \cos(2\pi f_0(t - \tau) + \Theta)\} \\ &= \int_{-\infty}^{+\infty} A \cos(2\pi f_0 t + \theta) A \cos(2\pi f_0(t - \tau) + \theta) f_{\Theta}(\theta) d\theta \\ &= \frac{A^2}{2\pi} \int_{-\pi}^{\pi} \cos(2\pi f_0 t + \theta) \cos(2\pi f_0(t - \tau) + \theta) d\theta \end{aligned} \quad (1.43)$$

Using trigonometric formulas, it is easy to show that

$$R_w(t, \tau) = \frac{A^2}{2} \cos(2\pi f_0 \tau) = R_w(\tau) \quad (1.44)$$

that does *not* depend on t . Our suspicion about the wide-sense stationarity of the process was well founded indeed!

A special example of WSS random signal is the *white noise* process. White light is the one made of all of the colors, we've already mentioned. Color means wavelength, and wavelength means frequency. A white noise process is zero-mean, and such that its psd is *constant throughout the whole frequency range*:

$$S_w(f) = N_0/2 \quad \forall f \quad (1.45)$$

This is of course an idealization, since no such signal may exist: its associated power is clearly infinite (recall (1.53)). But it is a good model for a random signal that has a much wider bandwidth than that of another “reference” signal or of a system under consideration. This is often the case for the electronic noise (thermal noise, Johnson noise) of active and passive devices, for the “ambient” noise detected by an antennas and so on. In such examples, the noise is generated by the superposition of a multitude of independent elementary components. A well-known result in probability (the central limit theorem) states that the outcome of such superposition is a Gaussian process: any set of random variables $w(t_1), w(t_2), \dots, w(t_N)$ obtained after observation of the signal at the time instants t_1, t_2, \dots, t_N has a multivariate Gaussian pdf.

Filtering is a fundamental operation for random signals, just as it is for deterministic waveforms. What happens in particular when a process $w(t)$ is filtered with an LTI system to obtain (the random process) $n(t)$? It is easy to understand how to derive the mean value function. From (1.33) we see that the computation of such function is obtained through *linear* operations: the “Expectation” operator is linear. We can invoke thus a “commutative” property of linear operators and say that

$$\begin{aligned}\eta_n(t) &= \text{E}\{n(t)\} = \text{E}\{\mathcal{T}[w(\alpha); t]\} = \mathcal{T}[\text{E}\{w(\alpha)\}; t] \\ &= \mathcal{T}[\eta_w(\alpha); t] = \eta_w(t) \otimes h(t)\end{aligned}\quad (1.46)$$

If $w(t)$ is zero-mean, $n(t)$ will always be zero-mean as well, irrespective of the particular LTI filter we may consider.

Esempio 1.6

Take back into consideration the parametric process of Example 4

$$w(t) = A \cos(2\pi f_0 t + \Theta) \quad (1.47)$$

where Θ is a uniform random variable in the interval $[-\pi/2, \pi/2]$. Its average value was found to be

$$\eta_w(t) = -\frac{2A}{\pi} \sin(2\pi f_0 t) \quad (1.48)$$

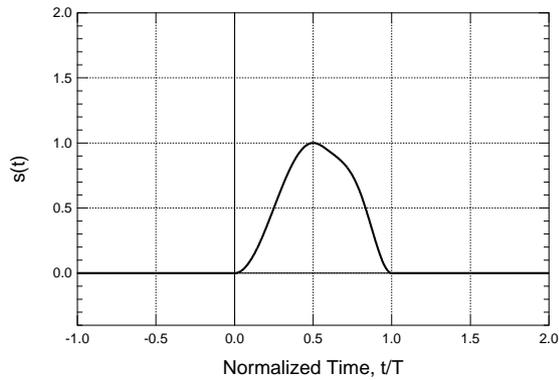
What if $w(t)$ is filtered by the sliding-window integrator of Example 3? Let us call $n(t)$ the result of such filtering. From (1.46) we know that $\eta_n(t) = \eta_w(t) \otimes h(t)$. But $\eta_w(t)$ is *sinusoidal* with time, so that the result of filtering can be easily evaluated through the notion of frequency response of the filter, as in (3.1):

$$\begin{aligned}\eta_n(t) &= -\frac{2A}{\pi} |H(f_0)| \sin(2\pi f_0 t + \angle H(f_0)) \\ &= -\frac{2A \text{sinc}(f_0 T)}{\pi} \sin(2\pi f_0 t + \pi f_0 T)\end{aligned}\quad (1.49)$$

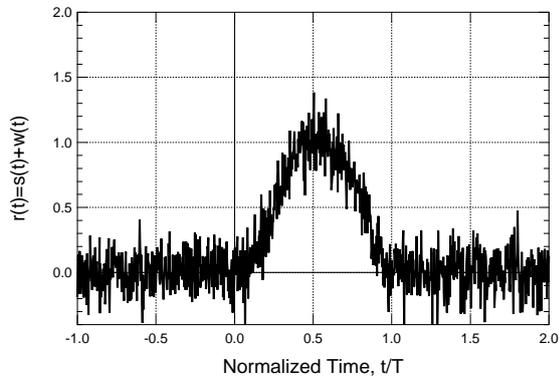
Filtering a WSS random process $w(t)$ is a special case that is easily characterized if we stick to a simple characterization of the output process $n(t)$. We already know from (1.46) how to compute the output average value. The output psd function is also easy to compute since, as happens with deterministic signals,

$$S_n(f) = |H(f)|^2 \cdot S_w(f) \quad (1.50)$$

where $|H(f)|^2$ is the *power response* of the LTI filter.



(a)



(b)

Figura 1.12 Time-limited signal $s(t)$ (a) and its noisy version $r(t)$ (b)

Esempio 1.7

A certain known signal $s(t)$ is sent out on a wireless radio channel. The receiver collects such signal corrupted by Additive White Gaussian Noise (AWGN) $w(t)$ with psd $N_0/2$ (a sketch of a possible noisy signal is in Fig. 1.12 (b)). The received signal processes $r(t) = s(t) + w(t)$ with an LTI system (filter) to get a filtered signal $y(t) = x(t) + n(t)$, where x and n are the filtered signal and noise components, respectively. Assume also that $s(t)$ is time-limited within the interval $[0, T]$ as in Fig. 1.12 (a), and that the reception filter has impulse response

$$h(t) = \frac{1}{E_s} s(T - t) \tag{1.51}$$

i.e., a reversed and time-delayed (to be causal) version of the transmitted pulse, scaled by the factor $E_s = \int s^2(t)dt$ that is, by the energy of the time-limited signal s . The output of the receive filter at time T is:

$$y(T) = x(T) + n(T) = \frac{1}{E_s} \int_{-\infty}^{+\infty} s(\alpha) s(T - (T - \alpha)) d\alpha + N = 1 + N \tag{1.52}$$

where N is a zero-mean Gaussian random variable with variance

$$\begin{aligned}\sigma_N^2 &= \int_{-\infty}^{+\infty} S_n(f) df = \int_{-\infty}^{+\infty} \frac{N_0}{2} |H(f)|^2 df \\ &= \frac{N_0}{2} \frac{1}{E_s^2} \int_{-\infty}^{+\infty} |S(f)|^2 df = \frac{1}{2E_s/N_0}\end{aligned}\quad (1.53)$$

We can compute now the *signal to noise ratio* (SNR) as the ratio between the squared signal component (signal power) and the variance of the noise component (noise power):

$$\text{SNR} \triangleq \frac{x^2(T)}{\sigma_N^2} = \frac{1}{(2E_s/N_0)^{-1}} = \frac{2E_s}{N_0}\quad (1.54)$$

It can be shown that (1.54) is the *best* SNR that can be attained upon filtering of $r(t)$ with *any* LTI system. The shape of $h(t)$ as in 1.51 is the “best match” to the received signal, and so this special filter is called the *matched filter*

1.4 Bandpass Signals and Systems

1.4.1 Baseband equivalent of a bandpass signal

The general form of a sinusoidal bandpass signal at frequency f_0 is

$$x_{BP}(t) = A \cos(2\pi f_0 t + \vartheta)\quad (1.55)$$

where A is the amplitude of the signal and ϑ its phase. An alternative formulation of (1.55) is

$$\begin{aligned}x_{BP}(t) &= A \cos(\vartheta) \cos(2\pi f_0 t) - A \sin(\vartheta) \sin(2\pi f_0 t) \\ &= x_I \cos(2\pi f_0 t) - x_Q \sin(2\pi f_0 t) \\ &= \Re\{(x_I + jx_Q) \exp(j2\pi f_0 t)\} \\ &= \Re\{x_{BB} \exp(j2\pi f_0 t)\}\end{aligned}\quad (1.56)$$

where we have introduced other quantities than the amplitude and phase of the sinusoid, that will be used extensively in the following. First, we defined the *In-phase/Quadrature components* $x_I = A \cos(\vartheta)$ and $x_Q = A \sin(\vartheta)$ as the “projections” of the sinusoid along the two main quadrature carriers at frequency f_0 , namely, $\cos(2\pi f_0 t)$ and $-\sin(2\pi f_0 t)$, respectively. Also, we introduced the complex-valued notation of the *baseband equivalent* of our sinusoidal signal $x_{BB} = x_I + jx_Q$. The amplitude of x_{BB} is the amplitude of our sinusoid, and the phase of x_{BB} is its initial phase, $x_{BB} = A \exp(j\vartheta)$. We summarize all this in the easy-to-remember visual representation given in Fig. 1.13

The spectrum of a sinusoid is monochromatic, i.e., it contains only one component at the frequency f_0 (and its twin at $-f_0$ if we use complex-valued FTs). What happens if by virtue of some kind of *modulation* the amplitude and/or phase of $x_{BP}(t)$ are made slowly

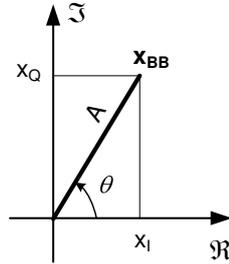


Figure 1.13 Visual representation of I/Q components and amplitude/phase of a baseband signal

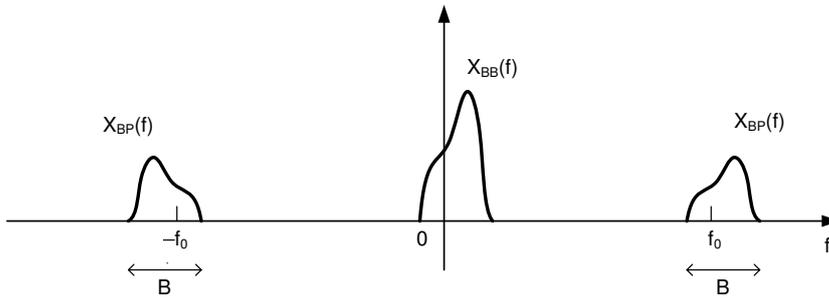


Figure 1.14 Samples of Bandpass and Baseband-equivalent spectra

varying in time? We can write

$$\begin{aligned}
 x_{BP}(t) &= A(t) \cos(2\pi f_0 t + \vartheta(t)) \\
 &= A(t) \cos(\vartheta(t)) \cos(2\pi f_0 t) - A(t) \sin(\vartheta(t)) \sin(2\pi f_0 t) \\
 &= x_I(t) \cos(2\pi f_0 t) - x_Q(t) \sin(2\pi f_0 t) \\
 &= \Re\{(x_I(t) + jx_Q(t)) \exp(j2\pi f_0 t)\} \\
 &= \Re\{x_{BB}(t) \exp(j2\pi f_0 t)\}
 \end{aligned}
 \tag{1.57}$$

The quantities (I/Q components, baseband equivalent) that we mentioned above are now (slowly) varying in time, where “slowly” is to be intended “on a time scale much larger than $1/f_0$ ”. It turns out that the spectrum of $x_{BP}(t)$ as in (1.57) is no longer monochromatic, but it is concentrated around the *carrier frequency* f_0 . The *passband* of such spectrum is $B \ll f_0$, and so the signal is *quasi-monochromatic* or *bandpass*. On the contrary, $x_{BB}(t)$ has a spectrum that is confined to baseband, with a bandwidth B much smaller than f_0 . Since the signal is complex valued, the spectrum of $x_{BB}(t)$ will not bear any Hermitian symmetry around 0. Figure 1.14 shows fictional examples of spectra of a bandpass, modulated, quasi-monochromatic signal, as well as its (non-Hermitian-symmetric) baseband equivalent.

1.4.2 The I-Q modulator

So the bandpass signal, once the carrier frequency is known, is completely specified by its baseband equivalent (also called *complex envelope*) $x_{BB}(t)$. Equation (1.57) tells us that $x_{BP}(t) = x_I(t) \cos(2\pi f_0 t) - x_Q(t) \sin(2\pi f_0 t)$. This is not only a mathematical

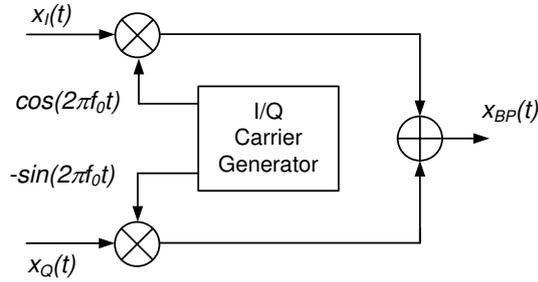


Figure 1.15 General architecture of an I-Q modulator

representation, but it also corresponds to the architecture of the so-called *I-Q modulator* that is used in the practice to generate an arbitrary bandpass modulated signal from its I-Q components as in Fig. 1.15. In the Chapters to follow, we will systematically adopt the complex-valued notation of baseband equivalent signals. To make notation shorter, we will drop the subscripts $_{BP}$ or $_{BB}$ to denote bandpass or baseband signals, respectively, but we will implicitly assume, unless otherwise stated, that all signals are complex envelopes.

1.4.3 The I-Q demodulator

The usual way of modulating a bandpass signal with a digital data stream is to encode the digital information into either $x_I(t)$, or $x_Q(t)$, or both. Once we have done so, and we send the bandpass modulated signal on a physical medium (radio, copper, fiber), the receiver needs to reconstruct either $x_I(t)$, or $x_Q(t)$, or both, to recover (*demodulate*) the digital data. The simplest way to do this starts from the general expression of the bandpass signal:

$$\begin{aligned} x_{BP}(t) &= \Re\{x_{BB}(t) \exp(j2\pi f_0 t)\} \\ &= \frac{x_{BB}(t) \exp(j2\pi f_0 t) + x_{BB}^*(t) \exp(-j2\pi f_0 t)}{2} \end{aligned} \quad (1.58)$$

From this we have,

$$x_{BP}(t) \cdot 2 \exp(-j2\pi f_0 t) = x_{BB}(t) + x_{BB}^*(t) \exp(-j2\pi \cdot 2f_0 t) \quad (1.59)$$

and we see that such (complex-valued) signal contains two components: the first one is just the one we intend to get, and the second is something unwanted and centered at the frequency $-2f_0$. What we have to do to get rid of the latter and keep the former is processing this signal with a *lowpass* filter whose bandwidth is just that of $x_{BB}(t)$ to remove the double-frequency components at $2f_0$. The result of this reasoning is simple:

$$x_{BB}(t) = x_I(t) + jx_Q(t) = \{x_{BP}(t) \cdot 2 \exp(-j2\pi f_0 t)\} \otimes h_{LP}(t) \quad (1.60)$$

where $h_{LP}(t)$ is the impulse response of the lowpass filter. Again, (1.60) is not just mathematics, but it is the outline of the so-called *I-Q demodulator* represented in Fig. 1.16 that is implemented in the vast majority of modern radio receivers. It is seen that the product of the received bandpass signal $x_{BP}(t)$ with the complex oscillation $2 \exp(-j2\pi f_0 t)$ is implemented as a pair of real products between the former and the real-imaginary components of the latter, and the lowpass filter is applied to both components as well.

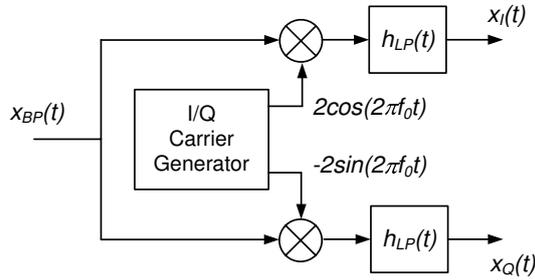


Figure 1.16 General architecture of an I-Q demodulator

1.5 Fourier analysis of digital signals

1.5.1 Analog and Digital signals

In the previous section we have reviewed the main concepts and results in Fourier analysis of time-continuous (*analog*) signals. The same results can be extended to *digital* signals, or, properly speaking, time-discrete ones. The most immediate way of generating a digital signal is using an analog-to-digital converter (ADC), or, in the parlance of signal analysis, performing the operation of *sampling*. Sampling an analog signal $x(t)$ with a certain sampling rate f_s samples/s (or simply Hz) or equivalently a sampling period $T_s = 1/f_s$ means extracting from $x(t)$ a sequence of *samples* $x[n]$ such as

$$x[n] = x(nT_s) \tag{1.61}$$

The square brackets indicates the the time index n they enclose is discrete, as opposed to continuous time t that is usually enclosed into round brackets. The value at time n of $x[n]$ is real-valued, and so its representation theoretically requires an infinite number of digits. In the practice, the ADC represents each value as an integer on a fixed (finite) number of binary digits. This introduces a (small) representation error: what we get out of the ADC is actually $x_q[n] = x[n] + q[n]$, where x_q is the *quantized* version of $x[n]$, and $q[n]$ is the *quantization noise*. When the number of bits in the digital representation of $x_q[n]$ is large (say, larger than 16), the quantization noise can be safely disregarded.

1.5.2 FT of Digital signals

Sampling and the ADC are the foundation of Digital Signal Processing (DSP). DSP techniques are again heavily based on Fourier analysis, so that we have to review the basics of Fourier transforms for time-discrete signals.

Generalizing the notions already introduced for analog signals, it can be easily shown that a non-periodic sequence $x[n]$ can be Fourier-decomposed as

$$x[n] = \frac{1}{f_s} \int_{-f_s/2}^{+f_s/2} \overline{X}(f) \exp(j2\pi n f / f_s) df \tag{1.62}$$

where $\overline{X}(f)$ is the FT of the sequence $x[n]$. Equation (1.62) is apparently a *synthesis* equation much similar to (1.8) for analog signals. The corresponding *analysis* equation that

gives the FT $\bar{X}(f)$ is

$$\bar{X}(f) = \sum_{n=-\infty}^{+\infty} x[n] \exp(-j2\pi n f / f_s) \quad (1.63)$$

A fundamental difference exists between the FT of analog and of digital signals. It is seen from (1.62) that the digital signal can be synthesized from a *finite* interval of continuous frequency components, whilst the analog signal requires component at *all* frequencies on the real axis. The reason for this is that the FT $\bar{X}(f)$ of a sequence is a *periodic* function of frequency, with a period equal to the sampling frequency f_s . Thus, the only independent components of $x[n]$ actually lie on an f_s -wide frequency interval, and no more components than those are required in the synthesis. This has also something to do with the property of sampled sinusoids. The sequence extracted by sampling a sinusoidal signal $x_1(t)$ at frequency f_1 is

$$x_1[n] = x_1(n/f_s) = \cos(2\pi n f_1 / f_s) \quad (1.64)$$

The ratio f_1/f_s is sometimes indicated with F_1 and is called the *normalized* frequency. Assume now we have a second sinusoidal signal $x_2(t)$ at the frequency $f_1 + f_s$. $x_2(t)$ is clearly (much) different from $x_1(t)$. After sampling $x_2(t)$ we get

$$x_2[n] = \cos(2\pi n f_2 / f_s) = \cos(2\pi n f_1 / f_s + 2\pi n) = \cos(2\pi n f_1 / f_s) = x_1[n] \quad (1.65)$$

After sampling, the two previously different sinusoidal signals look exactly the same! This means that in the digital domain, there can be no more independent sinusoidal components to synthesize a signal than those into a f_s -wide “base” interval: a component outside that interval is actually the “image” of another one that lies into the base interval at a distance equal to an integer multiple of f_s .

Esempio 1.8

Assume that we sample a time-continuous exponential signal $x(t) = \exp(-t/\alpha)u(t)$ with a sampling interval $T = s$. What we get is

$$x[n] = x(nT_s) = \exp(-nT/\alpha) = a^n u[n] \quad (1.66)$$

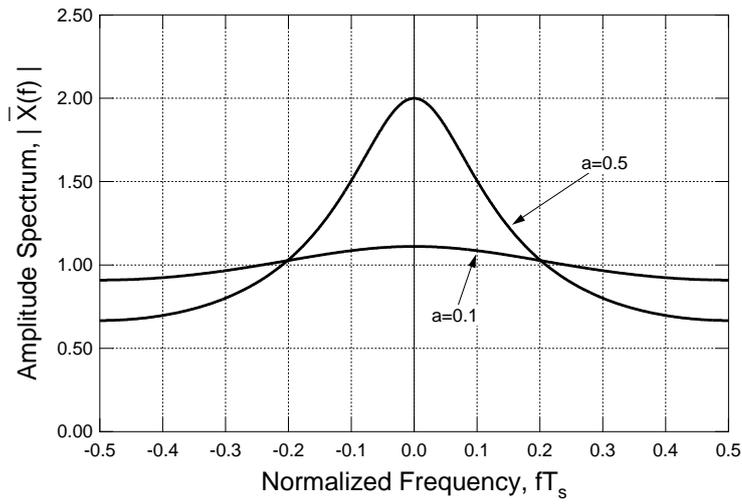
where $a \triangleq \exp(-T/\alpha) < 1$ is a real constant depending on the time constant of the signal and on the sampling frequency. The FT of the resulting sequence is

$$\begin{aligned} \bar{X}(f) &= \sum x[n] \exp(-j2\pi n f T_s) = \sum_{n=0}^{\infty} a^n \exp(-j2\pi n f T_s) \\ &= \sum_{n=0}^{\infty} [a \exp(-j2\pi f T_s)]^n = \frac{1}{1 - a \exp(-j2\pi f T_s)} \end{aligned} \quad (1.67)$$

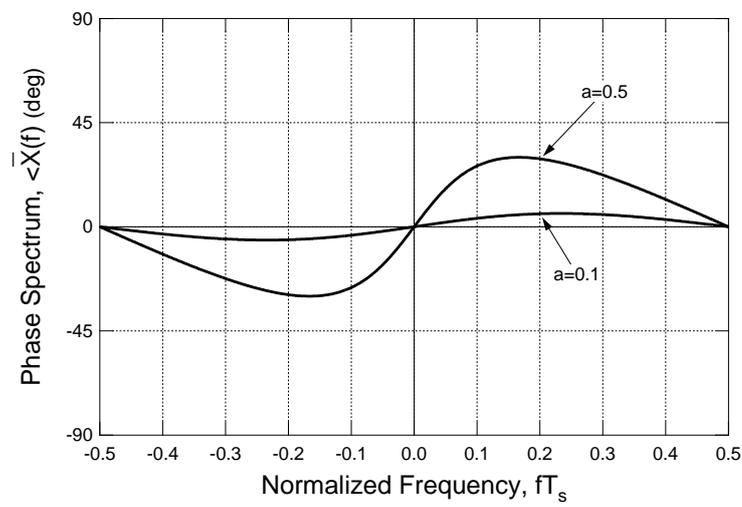
where certainty about the convergence of the series comes from the condition

$$|a \exp(-j2\pi f T_s)| = |a| < 1$$

The amplitude and phase spectra of $x[n]$ as resulting from (1.67) are shown in Fig. 1.17. Note that they are shown across a frequency span equal to $\pm f_s/2$ since the two functions are *periodic* with period f_s .



(a)



(b)

Figure 1.17 Amplitude (a) and phase (b) spectrum of the exponential sequence

Many of the properties that we already mentioned for the FTs of analog signals hold true for FT of sequences as well, with minor modifications. We refer in particular to Hermitian symmetry, delay and modulation theorems, and so on.

Esempio 1.9

In the domain of analog signals, special attention was devoted to the definition and properties of Dirac's delta function (Sect. 1.1.4). Is there something similar in the digital domain? The answer is simpler than expected. While $\delta(t)$ was a very special signal, its time-discrete counterpart is just the ordinary sequence

$$\delta[n] \triangleq \begin{cases} 1 & n = 0 \\ 0 & \text{elsewhere} \end{cases} \quad (1.68)$$

Its FT is clearly

$$\bar{\Delta}(f) = \sum_{n=-\infty}^{+\infty} \delta[n] \exp(-j2\pi n f / f_s) = 1 \quad (1.69)$$

just like the FT of $\delta(t)$.

If our sequence $x[n]$ comes from sampling of an analog signal $x(t)$ with FT $X(f)$, a fundamental question is: what is the relation between the FT $X(f)$ of the signal we start from, and the FT of the resulting sequence $x[n]$? The answer is called *Poisson's relation* and reads

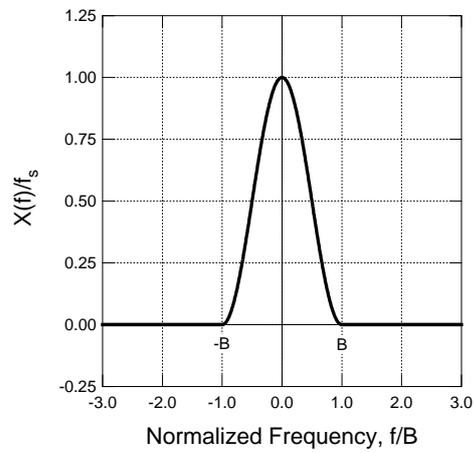
$$\bar{X}(f) = f_s \sum_{k=-\infty}^{+\infty} X(f - kf_s) \quad (1.70)$$

This equation tells us that the FT of $x[n]$ is the superposition of an infinite series of repetitions of the original FT of the analog signal, with a repetition period that is equal to the sampling frequency f_s . This of course gives a periodic FT $\bar{X}(f)$ as that of any digital signal. Notice that there might be a sort of "interference" between adjacent repetitions of the original spectrum that is called *aliasing*. An example of aliasing due to sampling is shown in Fig. 1.18 (b) that shows the FT $\bar{X}(f)$ of the sequence $x[n]$ obtained after sampling the analog signal $x(t)$ whose FT $X(f)$ is shown in Fig. 1.18 (a). Such superposition of adjacent spectra does not occur only if the original signal $x(t)$ is bandlimited into the band B , and the sampling frequency is larger than $2B$. The condition $f_s > 2B$ is called the *Nyquist's condition*. When it is verified, there's no aliasing in the sampled signal spectrum. Compare in this respect Fig. 1.18 (a) showing again $\bar{X}(f)$ resulting from the sampling of the signal in Fig. 1.18 (a), this time meeting Nyquist's condition. In particular, the main replica with $k = 0$ in Poisson's relation (1.70) that lies in the main interval $[-f_s/2, f_s/2)$ of the FT is a perfect replica (apart from an immaterial scale factor) of the original spectrum $X(f)$.

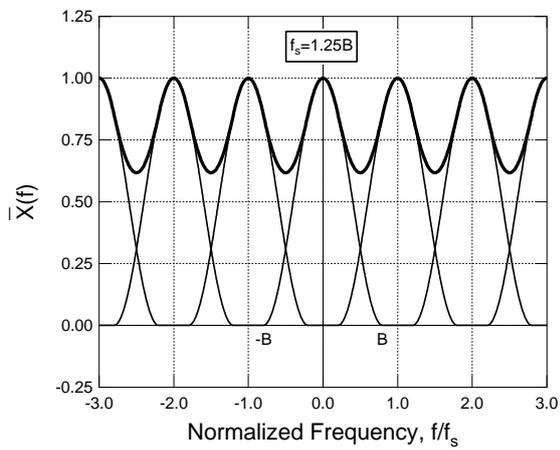
1.5.3 Filtering and Interpolation of digital signal

Needless to say, the notion of an LTI system translates nicely and neatly into the digital domain as well. Using a notation similar to the one we introduced for analog systems, a digital LTI filter is identified by an impulse response

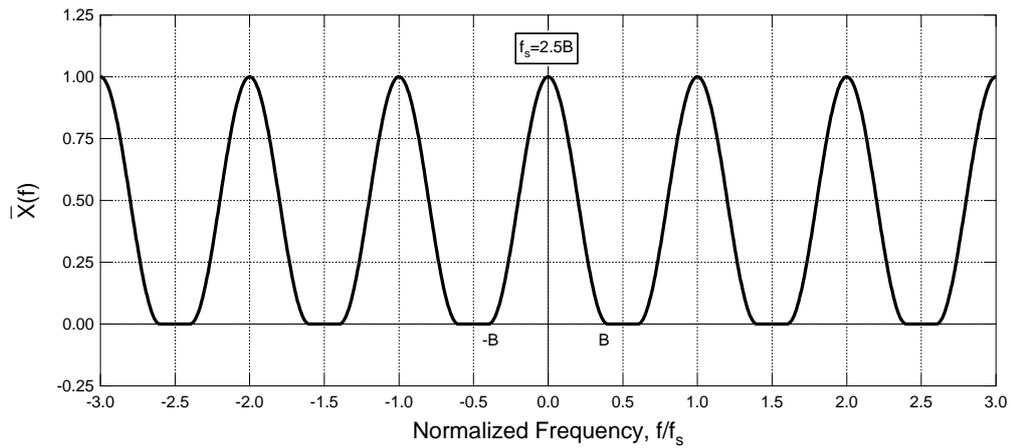
$$h[n] \triangleq \mathcal{T}[\delta[n]; n] \quad (1.71)$$



(a)



(b)



(c)

Figure 1.18 Spectrum of an analog signal (a), of the sampled digital signal with aliasing (b), and of the sampled digital signal without aliasing (c)

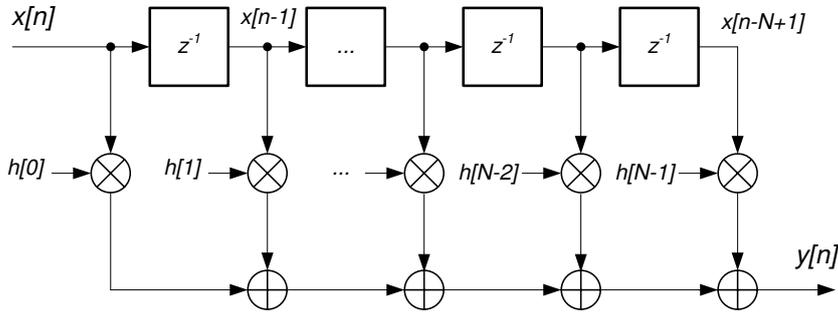


Figure 1.19 Implementation of an FIR filter

and its operation is described by the time-discrete aperiodic convolution between the input signal $x[n]$ and such impulse response $h[n]$:

$$y[n] = x[n] \otimes h[n] \triangleq \sum_{m=-\infty}^{+\infty} x[m] \cdot h[n-m] = \sum_{m=-\infty}^{+\infty} h[m] \cdot x[n-m] \quad (1.72)$$

The system may also have a frequency response

$$\bar{H}(f) \triangleq \sum_{n=-\infty}^{+\infty} h[n] \exp(-j2\pi n f / f_s) \quad (1.73)$$

so that the frequency-domain input-output relationship still is

$$\bar{Y}(f) = \bar{X}(f)\bar{H}(f) \quad (1.74)$$

If the impulse response $h[n]$ of the LTI system has a finite number of samples different from zero, the filter is *Finite Impulse Response* (FIR), otherwise it is IIR (*Infinite Impulse Response*). The operations to be computed to implement an FIR filter can be represented as in Fig. 1.19, where the blocks labeled z^{-1} implement the delay of their input sequence by one sample (unit-delay element), and where we have assumed that $h[n] = 0$ when $n < 0$ or $n \geq N$.

Once we have turned an analog signal into a digital sequence via sampling, and after we have possibly performed some digital processing on such (digital) signal (even simple storage on a digital medium such as a Compact Disc or a flash memory stick), it may be desired to reconstruct an analog signal from the resulting (retrieved) sequence. This reverse-sampling operation is called *interpolation* and in the practice it is implemented by a *Digital to Analog Converter* (DAC). The general form of an interpolated signal $\hat{x}(t)$ is

$$\hat{x}(t) = \sum_{n=-\infty}^{+\infty} x[n] \cdot p(t - nT_s) \quad (1.75)$$

where $x[n]$ is the sequence being interpolated, and $p(t)$ is the pulse shape that is specific of a particular interpolator. If $p(t)$ is the rectangular pulse in Fig. 1.20 (a), then we have a zero-hold interpolator that basically produces a sample-and-hold signal. If on the contrary $p(t)$ is the triangular pulse in Fig. 1.20 (b) we get a linear interpolator that joins consecutive

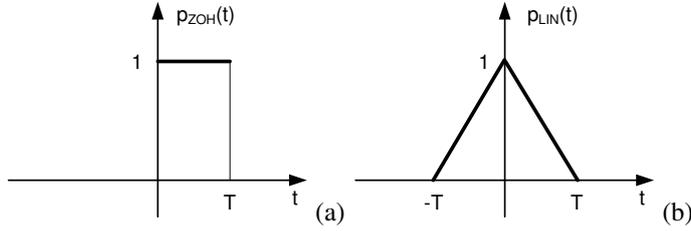


Figure 1.20 Interpolating pulses: (a) ZOH interpolator; (b) linear interpolator

values of the signal samples with straight line segments to produce the interpolated analog signal $\hat{x}(t)$. The frequency-domain counterpart of (1.75) is very simple:

$$\hat{X}(f) = P(f)\bar{X}(f) \tag{1.76}$$

Notice that both $\hat{X}(f)$ and $P(f)$ are FTs of analog signals, whilst $\bar{X}(f)$ is the FT of the sequence $x[n]$ to be interpolated.

1.5.4 The sampling theorem

A rather fundamental question about signal sampling comes immediately to one’s mind. Once an analog signal is sampled, and its samples are all collected and, say, stored, is it possible to fully recover such signal with no loss? At first sight the response is NO, since when converting a signal from time-continuous to time-discrete all that is in between samples appears to have been lost for ever. BUT... a glance in the frequency domain may give more hope. If the signal is bandlimited to B and we meet the Nyquist’s condition $f_s \geq 2B$, we already know that we have no aliasing, and we “see” an undistorted replica of the spectrum of the analog signal in the spectrum of our sequence (the replica with $k = 0$ in the Poisson formula (1.70)). The real issue is how to recover such replica and get back to the analog domain. The answer is relatively simple: we are to use an appropriate interpolator that preserves the replica with $k = 0$ while canceling all of the others. Figure 1.21 explains that (the reference is again the analog signal spectrum shown in Fig. 1.18 (a)): we need an interpolator whose FT $P(f)$ is flat within the frequency interval $[-f_s/2, f_s/2)$ that contains the main replica with $k = 0$, and zero outside that band. Also, it has to compensate the factor f_s in Poisson’s relation. In a word, we have to choose

$$P(f) = \frac{1}{f_s} \text{rect} \left(\frac{f}{f_s} \right) = T_s \text{rect}(fT_s) \tag{1.77}$$

Under Nyquist’s condition and using this interpolator, it is apparent that $\hat{X}(f) = X(f)$, so that we can say that the issue of reconstructing the sampled signal is now solved. The interpolating pulse that corresponds to such $P(f)$ is trivially $p(t) = \text{sinc}(t/T_s)$, so that the relevant interpolation formula is

$$\hat{x}(t) = \sum_{n=-\infty}^{+\infty} x[n] \cdot \text{sinc} \left(\frac{t - nT_s}{T_s} \right) \tag{1.78}$$

that is called the *cardinal interpolator*. Since we know that $\hat{X}(f) = X(f)$, it is also clear that $\hat{x}(t) = x(t)$

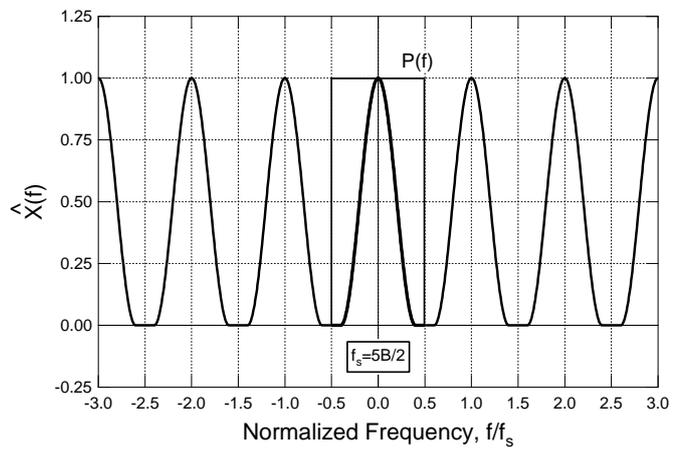


Figure 1.21 Frequency-domain interpretation of cardinal interpolation

CAPITOLO 2

TECHNOLOGIES FOR THE "LAST KILOMETER" (LAST MILE)



“Ahead of the Pack”

—The dreaded “Last Kilometer” banner of an Italian cycling race

In logistics, cycling races, and... communications, the last kilometers is the final segment that brings (goods, racer, bits) to its destination. The means to cover the last kilometer

has to be fast and reliable, so that special technologies are developed for the purpose. In the Internet, the main technologies of the last kilometer (or last mile in the USA) are the so-called xDSL for conventional telephone lines and FTTx for newly-installed fiber cables - the subject of this last chapter.

2.1 Multicarrier meets Information Theory: DSL Technologies and Shannon Capacity of the Additive Colored Gaussian Channel

The technology that is currently most widespread to cover efficiently and at low cost the *last kilometer* of the network and thus to provide "broadband" (i.e., high capacity) Internet connection to the residential subscriber, is a family of standards called xDSL, where *DSL* stands for Digital Subscriber Line, and *x* identifies two variants within the family, namely, ADSL and VDSL in order of commercial availability. The main features of xDSL will be described in the following sections, with particular emphasis on the information-theoretic aspects related to the DMT format (Discrete Multi Tone).

2.1.1 xDSL system architecture

xDSL access network technologies are intended to provide Internet connectivity to residential end-users, exploiting the pre-existing twisted pair wires belonging to the old (analog) telephone network. Such connection covers as already mentioned the last kilometer - the average distance between the subscriber's (fixed) location and the closest point-of-presence (POP) of an operator's transport network: a central exchange office or a cabinet on the street. In particular, the requirement is being able to provide a bit-rate up to 100 Mbit/s while at the same time guaranteeing the traditional telephone service (POTS, Plain Old Telephone Service).

The most popular and oldest version of the xDSL family is *Asymmetric DSL*, developed in the late 90's, meaning that the downstream (from the network to the end-user) has larger capacity than the upstream (from the end-user to the network). The overall architecture of xDSL is shown in Fig. 3.1. Here we assume that the user terminal connected to the Internet is a WiFi access point/router to implement a local wireless/wired area network, coexisting with a traditional telephone bearing an analog connection. Focusing on the upstream, the two connections are coupled through a *splitter*, actually a low-pass filter which selects the lower part of the spectrum for the analog connection and a high pass filter which selects the high-frequency band for the digital connection. The latter makes use of a remote ADSL transmission unit (ATU-R, ADSL Transmission Unit-Remote side, in other words, the *modem*) which generates the digitally modulated analog signal to be coupled with the telephone analog signal. The two signals are sent on a twisted pair to the cabinet/central office/POP, where the telephone signal is decoupled (via a splitter on the central side) and routed towards the telephone network (PSTN, Public Switched Telephone Network), while the digitally modulated signal is demodulated by a twin ADSL modem (ATU-C, Central-side ATU) and sent to the broadband digital transport network. In particular, the streams coming from the different ATU-Cs belonging to different users are multiplexed by the DSLAM (Digital Subscriber Line Access Multiplexer) into a single high-rate stream to be sent out onto the transport fiber. In addition to the central DSLAM, the necessary modification to the preexisting infrastructure lies in the addition of the ATU-R and ATU-C devices to be installed at the subscriber's site and at the POP, respectively.

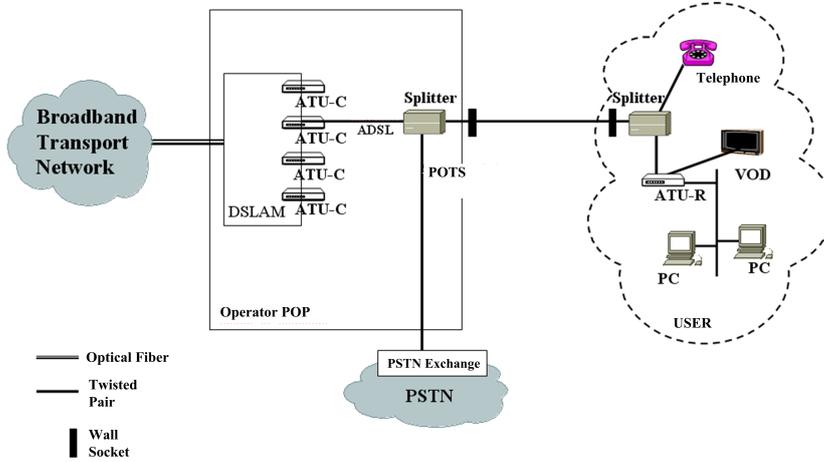


Figure 2.1 Architecture of an xDSL link

2.1.2 xDSL Frequency Plan and Communication Channel Model

Achieving hundred-Mbit/s capacity on a twisted pair represented at the time of development of xDSL technology a real technological challenge. It is pretty clear that, whatever the digital coding and modulation is going to be, a bit rate of tens of MBit/s requires a signal bandwidth of, at least, a few MHz, much, much wider than the 4kHz typical of POTS signals.

The starting requirement of xDSL is that concurrent provision of POTS and Internet connection must be guaranteed - this is actually easy to achieve and is obtained through the FDM/FDMA approach depicted in Fig. 3.2: POTS continues using its own preexisting baseband, with no change whatsoever, while the digital service is allocated onto a higher-frequency band, separated from POTS frequencies by a suited *guard band*. The digital connection is Frequency-Division Duplex (FDD), with the upstream being allocated to a low-frequency sub-band and the downstream allocated to a high-frequency, wider-bandwidth sub-band.

But, the main issues related to the technological challenge mentioned above are related to the very physical nature of the physical medium, namely: i) the frequency response of the twisted pair, and ii) the dominant noise affecting the digital signal. Concerning the frequency response, the twisted pair is actually a *transmission line* whose characteristics depends on a few factors: its length, the section of the conducting wires and so on. On the typical length of 1 km and on a bandwidth as wide as 1 MHz, the response is widely variable with frequency, creating a relevant issue of *frequency selectivity* for the digital signal. A simple model of the amplitude response of the twisted pair is

$$|H(f; L)|_{dB} = -\alpha \cdot L \cdot \sqrt{f} \tag{2.1}$$

where L is the length of the link, and α is the usual proportionality constant (also) depending on the section of the wires. Equation (3.1) show an exponential decay of the received power with the length of the cable, as well as a marked low-pass behavior with frequency. Frequency selectivity is negligible only over a few kHz bandwidth (just the one of POTS)

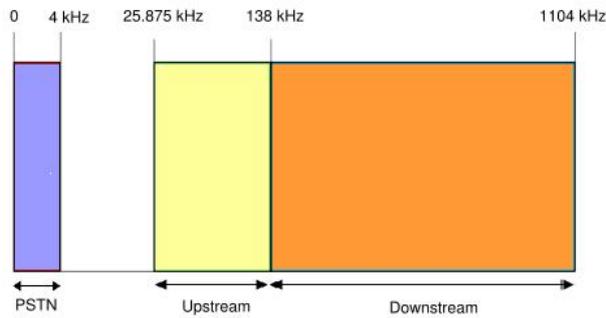


Figure 2.2 Frequency allocation of POTS/ADSL Upstream/ADSL Downstream



Figure 2.3 Unshielded Twisted Pair (UTP) cable aka Binder

and becomes essential across the wide bandwidth of xDSL services: 1 MHz of classical ADSL and 2 MHz of more modern ADSL2+, not to mention the 12 MHz of VDSL.

In addition to the bad frequency response, the twisted pair does not have any feature of protection against external interference (as opposed to more costly coaxial cables). So much so, that the limiting factor of the xDSL capacity turns out to be a well-known phenomenon dating back to the time of analog telephony: *crosstalk*. Crosstalk is the interference caused by a neighboring twisted pair carrying xDSL signals (the *interferer*) to another twisted pair (the *victim*) and it is caused by electromagnetic coupling of the two signals. This is why the pair of conductors of a telephone line are twisted: trying to avoid as much as possible this phenomenon by creating some degree of “spatial symmetry” seen by the interferer and thus diminishing the coupling. We will see in a while that crosstalk (abbreviated as XTalk) is very limited at the low frequencies of POTS, and this the reason why nobody cared about using coax instead of the twisted pair in the past.

The reason why XTalk arises stands in the way different pairs belonging to different subscribers are grouped together in their path to the cabinet/central. Figure 3.3 pictures an example of an underground cable (the so-called *binder*) of 25 twisted pairs coming from neighboring subscribers: the *unshielded* pairs are packed very close to each other, therefore they are prone to mutual interference, i.e., Xtalk.

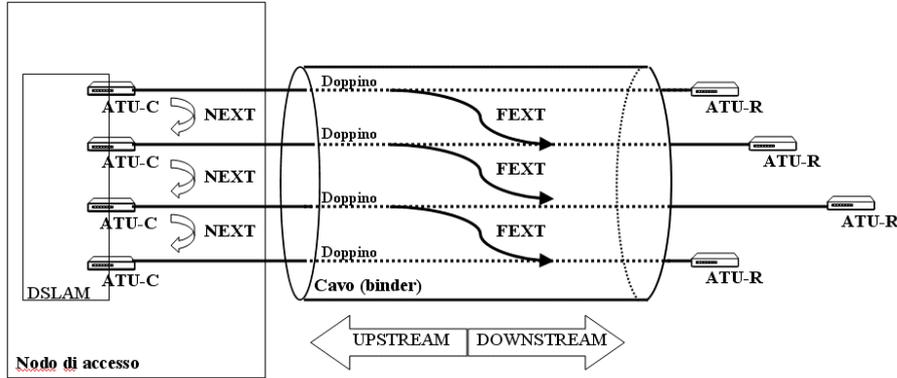


Figura 2.4 Upstream NEXT/FEXT

Further consideration of the issue reveals that we can identify *two* different sources of Xtalk: the so-called Far-End XTalk (FEXT) generated by transmitters that lie at the opposite (far) end of the link wrt the victim receiver, and the Near-End XTALK (NEXT) caused by transmitters that lie at the same (near) end of the link wrt the victim receiver, as is clearly shown in Fig. 3.4. It is also seen that the FEXT is generated by signals *co-propagating* with the victim (propagating in the same direction) whilst NEXT is generated by *counter-propagating* signals. Either kind of XTalk is, from the standpoint of the victim, a *random process*. The statistics of such processes are necessary to be able to find out the capacity limitation of the technology. To a good approximation, we can assume that the NEXT/FEXT is Gaussian by virtue of the central-limit theorem, since it is generated by a number of independent contributions coming from independent sources (the different subscribers' xDSL signals) that add-up when interfering with the victim's signal. In terms of spectral properties, it is easy understood that this kind of noiselike process cannot be white since it is caused by a non-white signals - it is a *colored* Gaussian noise, whose psd is well approximated by the following expression:

$$S_{NEXT}(f) \cong \alpha_{NEXT} \cdot f^{3/2} S_{xDSL}(f) \quad (2.2)$$

where k_{NEXT} is proportional to the number of active pairs in the binder, and $S_{xDSL}(f)$ is the psd of the (other users') xDSL digital signal. A similar model applies to the FEXT:

$$S_{FEXT}(f) \cong \alpha_{FEXT} \cdot f^2 \cdot L \cdot |H(f; L)|^2 S_{xDSL}(f) \quad (2.3)$$

where the main difference is that it also contains the amplitude response of the twisted pair $H(f; L)$ across the cable length L since the interferers come from the far end. At any frequency in the spectrum, it is seen that the intensity of the NEXT is larger than FEXT's. The physical reason is simple: the downstream signals that are generated by the ATU-C's are very strong in the binder section close to the Central (i.e., close to the upstream receiver), where at the same time the victim upstream signal is weak - therefore, the amount of interference collected there accounts for most of the NEXT, and its relative amplitude is relevant. On the contrary, in the same section of the binder where the upstream victim signal is weak, the interfering signals are weak as well since they (too) have traveled the whole connection, and their relative amplitude is modest. It is also seen that the NEXT, the most annoying noise component, has a substantially high-pass spectrum because of the

presence of the $f^{3/2}$ term in the spectrum. Such term is due to the dominant capacitive coupling effect between twisted pairs, with an intrinsically high-pass nature. The conclusion is that high-frequency components of the upstream signal are very much affected by NEXT - so much so, that in the FDD arrangement in Fig. 3.2 the upstream spectrum has been allocated to lower frequencies than the downstream's, also taking advantage of a smaller signal attenuation (see (3.1)).

Why is downstream not affected by NEXT as much as the upstream? Just because the "final" section of the twisted pairs of the victim user in the downstream is separated from the other pairs, since it enters the user's premise and travels *individually* for a few tens of meters. In such final path, where the victim downstream signal is weak, there is no possibility for the strong but physically separated near-end upstream signals to create interference - when they are grouped into the binder the interferer is not so strong any longer, the victim is not so weak any longer as well, and the NEXT is greatly reduced. On the contrary, in the cabinet/central office, the pairs are immediately collected into the binder at the output of the ATU-C's, and that is why the NEXT on the upstream, as described before, is much stronger.

From the discussion above, we also understand why xDSL services *are forced* to be *asymmetric*: bit-rate asymmetry is generated of course by asymmetric bandwidth allocation between upstream and downstream, and this is trivial. The reason why the two bands are asymmetric is just the presence of upstream NEXT: there is no point in trying to make the two bandwidths equal, because in so doing we decrease the capacity of the downlink, but we do not proportionally increase the capacity of the uplink, since we'd take in a further band of frequencies that is very noisy, and therefore (as we will see in detail later on) does not add very much to capacity.

All of the other impairments on the link, like environmental RF interference, impulsive noise, receiver noise, etc. do not add very much to the analysis above and can safely be neglected. As a consequence, all of our analysis above can be summarized into the following two main issue: i) frequency selectivity originated by the bad frequency response of the physical medium across the signal bandwidth, and ii) (strong) Additive Colored Gaussian Noise (ACGN) with a rising spectrum with frequency (we may call it.. blue noise). And such issues bring forward a further fundamental question: how can we use such transmission medium at best (that is, at the maximum bit rate that is possible)? In other words, for the channel modeling above, can we find the *Shannon capacity* of this frequency selective, ACGN link? The subject of the next section.

2.2 Shannon Capacity of the ACGN Channel

We already know about the celebrated Shannon capacity formula for the band-limited AWGN channel:

$$C = B \log_2(1 + SNR) = B \log_2\left(1 + \frac{P}{\sigma^2}\right) = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{ [bit/s]} \quad (2.4)$$

where B is the channel bandwidth, SNR is the receiver signal-to-noise ratio, P and σ^2 indicate the signal and noise power, respectively, and $S_w(f) = N_0/2$ is the two-sided psd of the white Gaussian noise (so that $\sigma^2 = N_0 B$). Unfortunately, (3.4) only holds for a non-selective channel with white noise. How can we possibly adapt it to our case? The simplified model of our xDSL uplink is in fact depicted in Fig. 3.5a, where the (colored)

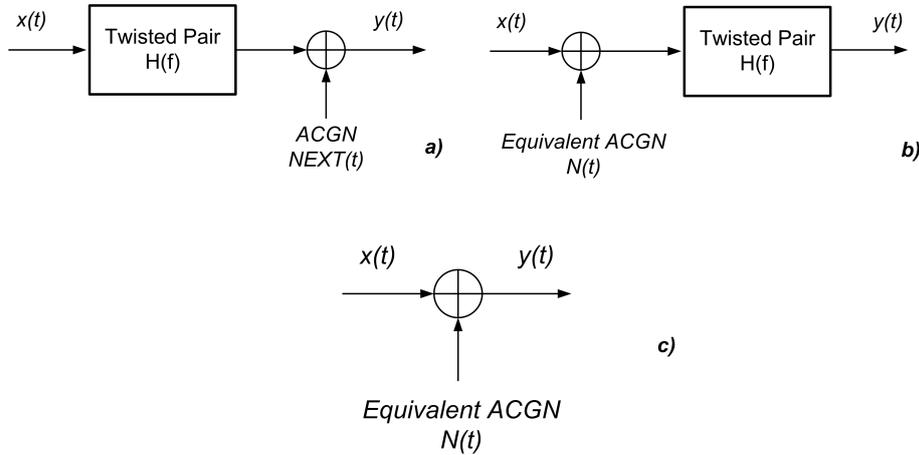


Figure 2.5 Frequency-Selective ACGN Channel and its Shannon-capacity equivalents

noise psd is the $S_{N_{NEXT}}(f)$ that we already know¹. We see that frequency selectivity is given by the frequency response $H(f)$, and the effect of NEXT is concentrated at the end of link - a reasonable approximation. Can we compute the Shannon capacity of this link?

The effect of selectivity can be easily accounted for. Since the noise is Gaussian, we can build an equivalent link model as in Fig. 3.5b, where the actual NEXT at the output of the twisted pair is replaced by an equivalent Gaussian noise $N(t)$ with psd $S_N(f)$ at its *input*. This can be done because Gaussian processes stay Gaussian when filtered, so that an input Gaussian noise gives an output noise component that is Gaussian. To make the two models equivalent, we have to enforce that

$$S_{N_{NEXT}}(f) = S_N(f) \cdot |H(f)|^2 \rightarrow S_N(f) = \frac{S_{N_{NEXT}}(f)}{|H(f)|^2} \quad (2.5)$$

The two models are completely end-to-end equivalent from any standpoint, so that their Shannon capacity is also the same - but what is the use of this transformation? The point is that now the filter representing the twisted pair is the last processing block in the link. If the response of the filter is *invertible* (as it happens for any physical medium), we can go on with our equivalent systems and come to the one in Fig. 3.5c where $H(f)$ has *disappeared*. We do this because the Shannon capacity of the original link is the same as that of this new equivalent one: the final channel filter is a deterministic invertible transformation not altering mutual information nor, therefore, capacity: the information transmission channel obtained observing either the input or the output of such filter bear the same Shannon capacity.

In conclusion, our final channel model is the one in Fig. 3.5c where we don't apparently have any frequency selectivity, and we have ACGN with psd $S_N(f)$ - selectivity is actually embedded into the new equivalent psd, making the noise spectrum even more "colored" than before. But still, the issue of finding Shannon capacity is not solved since we don't have white noise. How can we make the colored noise look any "whiter"?

The idea is simple: we can imagine to split the whole available bandwidth B into a number K of adjacent sub-bands all bearing the same bandwidth $\Delta f = B/K$, as

¹In reality, the analysis that we will perform is valid for any colored psd $S_{N_{NEXT}}$

sketched in Fig. 3.6. We can also imagine of splitting the input bit stream at the rate R_b into K sub-streams that are modulated onto K separate bandlimited digital signals $x_k(t)$, $k = 0, 1, \dots, K - 1$ each having the same bandwidth Δf and each allocated to the corresponding k -th sub-band $[k\Delta f, (k + 1)\Delta f)$, $k = 0, \dots, K - 1$. If K is sufficiently large, then each digital signals $x_k(t)$ "sees" within its own bandwidth a band-limited, sliced, bandpass version $N_k(t)$ of the total Gaussian noise $N(t)$ that is still Gaussian (since it is obtained after bandpass filtering of a Gaussian process) with a psd that in the narrow band $[k\Delta f, (k + 1)\Delta f)$ is substantially *flat*:

$$S_{N_k}(f) \simeq S_N(k\Delta f) = \frac{S_{NEXT}(k\Delta f)}{|H(k\Delta f)|^2}, \quad k\Delta f \leq f < (k + 1)\Delta f \quad (2.6)$$

The noise component $N_k(t)$ on the k -th sub-band is also statistically independent of the noise $N_m(t)$ on another sub-band $m \neq k$ since the two psd's of the two processes are non-overlapping in frequency and therefore the processes are uncorrelated (therefore independent since jointly Gaussian). The consequence of this is that the total Shannon capacity \mathcal{C} of the channel is the sum of the partial capacities \mathcal{C}_k of the parallel, independent channels on each sub-band. On the other hand, since the Gaussian noise in each subchannel is (approximately) white, it is easy to find \mathcal{C}_k by the usual AWGN formula:

$$\mathcal{C}_k = \Delta f \cdot \log_2(1 + SNR_k) = \Delta f \cdot \log_2\left(1 + \frac{P_k}{\sigma_{N_k}^2}\right) \quad (2.7)$$

If the subchannel bandwidth Δf is sufficiently small (in ADSL, $\Delta f = 4.3215$ kHz), we can (easily) compute the signal and noise power through their respective psd's, assuming they are constant across such band, so that

$$\begin{aligned} \mathcal{C}_k &= \Delta f \cdot \log_2\left(1 + \frac{P_k}{2 \cdot \Delta f \cdot S_N(k\Delta f)}\right) = \Delta f \cdot \log_2\left(1 + \frac{2 \cdot \Delta f \cdot S_X(k\Delta f)}{2 \cdot \Delta f \cdot S_N(k\Delta f)}\right) \\ &= \Delta f \cdot \log_2\left(1 + \frac{S_X(k\Delta f)|H(k\Delta f)|^2}{S_{NEXT}(k\Delta f)}\right) \text{ [bit/s]} \end{aligned} \quad (2.8)$$

The total Shannon capacity of the link is therefore

$$\begin{aligned} \mathcal{C} &= \sum_{k=0}^{K-1} \mathcal{C}_k = \\ &= \sum_{k=0}^{K-1} \Delta f \cdot \log_2\left(1 + \frac{P_k |H(k\Delta f)|^2}{2 S_{NEXT}(k\Delta f) \Delta f}\right) \\ &= \sum_{k=0}^{K-1} \Delta f \cdot \log_2\left(1 + \frac{S_X(k\Delta f) |H(k\Delta f)|^2}{S_{NEXT}(k\Delta f)}\right) \text{ [bit/s]} \end{aligned} \quad (2.9)$$

where $S_X(f)$ is the psd of the transmitted digital signal, completely under the control of the modem, and to be specified later on according to some criterion.

This expression is the starting point to derive on one side a general theoretical formula, and on the other to suggest a technology to be actually implemented in mass-market devices. Starting with theoretical derivations, if we let the number of sub-bands grow indefinitely

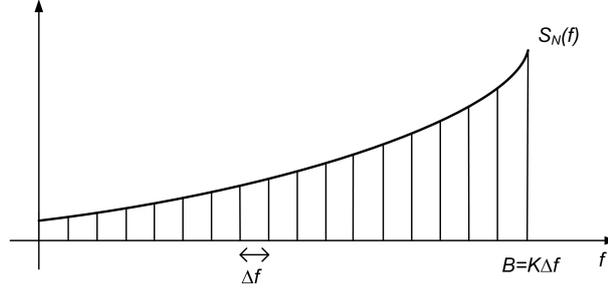


Figure 2.6 Computing Shannon Capacity of the ACGN Link

in (3.9), we get the general expression of Shannon capacity for the ACGN with frequency selectivity:

$$C_{ACGN} = \int_0^B \log_2 \left(1 + \frac{S_x(f)|H(f)|^2}{S_{NEXT}(f)} \right) df \quad (2.10)$$

where, still, the signal psd can be freely specified.

2.2.1 Power allocation and the Water Filling Criterion

The ACGN capacity formula (3.10) assume that the psd of the transmitted signal $S_x(f)$ is known or, in practical, discrete-sub-band version (3.7), that the individual power levels P_k are known. How can the modem find a proper “spectral shaping” or “power allocation” across the different sub-bands? We already know that lower-frequency sub-bands are in general less noisy than higher-frequency ones, so the temptation would be to allocate more power on those sub-bands bearing a higher level of noise just to balance the SNR across the whole xDSL bandwidth. Is this the correct approach?

We have first to understand what “correct” means in this context: rather than *correct* we should strive to find an *optimum* approach to allocate power - and in this context *optimum* means the one that can give us the maximum Internet connection speed $R_b \leq C$. In other words, the correct approach is that of finding the power allocation criterion that maximizes the Shannon capacity C of the link. When we speak of “allocation” what we intend is distributing a *finite* resource: the xDSL modem is allowed to transmit no more than a certain total signal power P_{tot} not to interfere with broadcast radio transmissions in the same band. Therefore, the power levels that we attribute to the various sub-band must be always such that $P_0 + P_1 + \dots + P_{K-1} = P_{tot}$. To sum up, we are faced with the following *constrained optimization* problem:

Find

$$P_k, \quad k = 0, \dots, K - 1$$

such that

$$C = \max_{P_0, P_1, \dots, P_{K-1}} \left[\Delta f \sum_{k=0}^{K-1} \log_2 \left(1 + \frac{P_k}{\sigma_k^2} \right) \right]$$

with the constraints

$$\sum_{k=0}^{K-1} P_k = P_{tot}$$

$$P_k \geq 0 \quad \forall k .$$

where we used the simplified notation $\sigma^2 \triangleq \sigma_{N_k}^2$. It is pretty clear that the problem makes no sense without the constant-total-power constraint - we could get in that case any arbitrarily high capacity by ever increasing the power levels allocated to the sub-bands.

The problem is solved by Lagrange multipliers method. Introducing the multiplier λ , the Lagrangian function to be maximized is

$$\mathcal{L}(P_0, P_1, \dots, P_{K-1}, \lambda) = \Delta f \sum_{k=0}^{K-1} \log_2 \left(1 + \frac{P_k}{\sigma_k^2} \right) + \lambda \left(\sum_{k=0}^{K-1} P_k - P_{tot} \right) \quad (2.11)$$

Differentiating wrt to P_k and equating to 0 we get

$$\frac{\Delta f \log_2 e}{1 + \frac{P_k}{\sigma_k^2}} \frac{1}{\sigma_k^2} + \lambda = 0 \quad , \quad k = 0, 1, \dots, K-1 \quad (2.12)$$

or

$$P_k + \sigma_k^2 = -\frac{\Delta f \log_2 e}{\lambda} \quad , \quad k = 0, 1, \dots, K-1 \quad (2.13)$$

computing the term-by-term summation on k of all these equations we also get

$$P_{tot} + \sigma_{tot}^2 = -K \frac{\Delta f \log_2 e}{\lambda} \quad \rightarrow \quad -\frac{\Delta f \log_2 e}{\lambda} = \frac{P_{tot} + \sigma_{tot}^2}{K} \quad (2.14)$$

where we used the constraint equation $\sum P_k = P_{tot}$ and where $\sigma_{tot}^2 \triangleq \sum \sigma_k^2$ represents the total noise power across all sub-bands - a value depending on the status of the channel and *not* depending on the particular power allocation. Using (3.14) into (3.13) we end up with

$$P_k + \sigma_k^2 = \bar{P} \quad , \quad \bar{P} \triangleq \frac{P_{tot} + \sigma_{tot}^2}{K} \quad (2.15)$$

with, again, \bar{P} a constant not depending on power allocation. The continuous-frequency version of this relation is (as the reader can easily verify)

$$S_x(f) + S_N(f) = \bar{S} \quad , \quad \bar{S} \triangleq \frac{1}{B} \int_0^B (S_x(f) + S_N(f)) df \quad (2.16)$$

These two final equation identify the criterion the modem has to use in allocating the power levels (i.e., finding the optimum spectral shaping of the transmitted signal) to maximize the ACGN capacity - it is called the *water-filling* criterion.

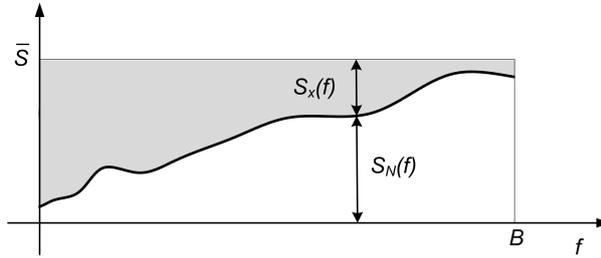


Figure 2.7 Representation of the “Water Filling” criterion

2.2.2 Why is “water-filling” called like that?

To understand why the solution to the previous problem has been named “water-filling” (WF), we start from the theoretical continuous-frequency solution (3.16) that we summarize hereafter:

$$\begin{cases} S_x(f) = \bar{S} - S_N(f) = \bar{S} - \frac{S_{NEXT}(f)}{|H(f)|^2} & S_N(f) < \bar{S} \\ S_x(f) = 0 & S_N(f) \geq \bar{S} \end{cases} \quad (2.17)$$

The frequency-flat (constant) value \bar{S} corresponds to the total signal+noise psd at the receiver: $S_{tot}(f) = S_x(f) + S_{W'}(f) = \bar{S}$, provided of course that the noise level on any subcarrier (frequency) does not exceed the noise level. The WF solution is represented in Fig. 3.7, and suggest what happens into an aquarium fish-tank, where the bottom is filled with sand (our colored noise level) and on top of that a certain quantity of *water* is poured to *fill* the tank until we come to a certain (flat) water-surface level. Frequency by frequency, the height of the water-level wrt the sand on the bottom represents the quantity of power that is allocated to that frequency, and the total amount of water-filled represents our P_{tot} (the shaded area in Fig. 3.7, whilst the total quantity of sand is σ_{tot}^2 . *tità di sabbia sul fondale* (area bianca nella figura).

It is pretty clear that the water level \bar{S} depends on the amount of power P_{tot} that can be expended - in some cases where the noise peaks on high frequencies, it may happen that the total water (power) quantity is not enough to cover all of the sand on the bottom, and some “dry islands” are left, as shown in Fig. 3.8. In such cases, the WF solution is clear: no power at all has to be allocated to those frequencies, i.e., the signals on the corresponding sub-bands in the finite-bands solution (3.15) has to be *switched off*, without wasting any signal power where the noise is too high².

As a final comment, we can say that WF is a counterintuitive criterion to allocate power: without solving the capacity maximization problem, a sensible heuristics might have been: let’s allocate in any sub-band a power level that is proportional to the relative noise level, so that each sub-band operates with the same value of SNR (i.e., $P_k = \overline{SNR} \sigma_k^2 \forall k$, $\overline{SNR} = P_{tot}/\sigma_{tot}^2$). Such solution would be labeled as fair, but it is *not* the optimal one.

²Since we are speaking here of the equivalent noise $N(t)$, this may mean either that the NEXT is very high (because we have all pairs active in the binder and we are considering high frequencies), and/or that the frequency response of the channel is very low (again on high frequencies, and when the length of the pair is relevant).

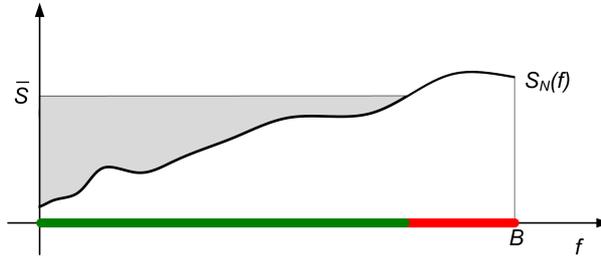


Figure 2.8 The “Island” effect in WF

2.3 DMT (Discrete MultiTone) Modulation Format

How can we actually design a digital signal format that implements all of the notions derived in the previous sections? The idea of splitting the whole bandwidth B into K separated sub-bands recalls the notion of a *multicarrier* format like the popular OFDM of wireless communications and broadcasting. Separation of the sub-bands is not implemented through strict bandlimitation of the signals $x_k(t)$; rather, it is realized via the usual orthogonality condition that we will discuss later on. In addition, to come as close as possible to Shannon’s capacity, a further step after power allocation is further needed.

Once power allocation is accomplished according to WF, like $P_k = \bar{P} - \sigma_k^2$, we have a widely variable level of operating SNR_k across the different sub-bands: typically, low-frequency sub-bands have a very good SNR (close to 40 dB) and high-frequency sub-bands have worst values (down to 10 dB and less). Therefore, the diverse sub-bands bear very different individual capacity

$$C_k = \Delta f \log_2 \left(1 + \frac{P_k}{\sigma_k^2} \right) \quad [\text{bit/s}] \quad (2.18)$$

When we split the overall modem bit rate R_b across the bandwidth and create the K modulated digital signals to be sent out, the different bit-rates on the different subcarriers will be widely different as well, and will have to obey *individually* the reliable-communication condition:

$$R_{b,k} \leq C_k, \quad \sum_{k=0}^{K-1} R_{b,k} = R_b \quad (2.19)$$

As is apparent, after *power* allocation is done, we have also to perform a function of *bit allocation*, i.e., distribution of the total bit-rate across the active subcarriers.

The particular flavor of the multicarrier technology that does all this and is at the base of xDSL format is called (Discrete MultiTone). We already know how to construct an efficient multicarrier modem: using FFT algorithms at the transmission and reception ends. This is what is also done in xDSL modems, which also includes the cyclic-prefix feature already examined for OFDM. The difference between OFDM and DMT lies however in a fundamental feature: in the xDSL link, the transmitter has also available a *return channel* from the receiver to provide frequency-by-frequency channel status information, something that OFDM formats in general do not have/exploit. Through the return channel we can make the modulation *Rate-Adaptive* (the so called RA-DMT), in the sense that the bit-rate is optimized against the available capacity via power/bit-rate (adaptive) allocation.

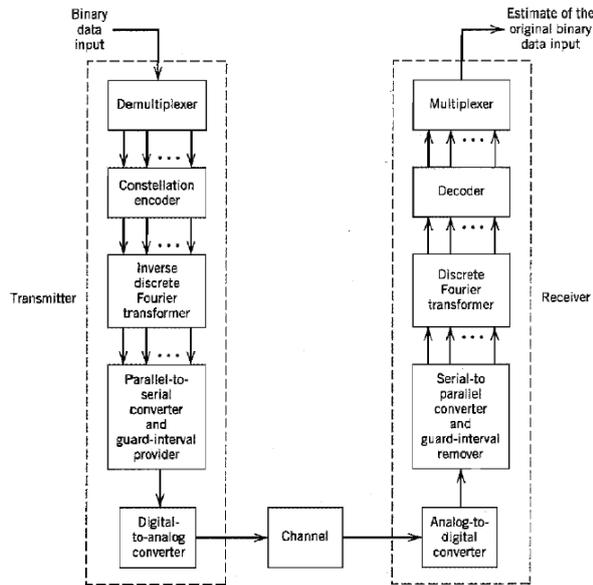


Figura 2.9 General Architecture of a DMT link

As already mentioned, separation of the sub-bands is achieved through carrier orthogonality. This means that the multicarrier symbol rate R_M is the same for all signals on any subcarrier, and it is equal to the subcarrier spacing: $R_M = \Delta f$. Therefore, the only way to allocate unevenly the bit-rate across the subcarriers is, on any subcarrier, to adapt and customize the constellation size, i.e., the number of points M_k in the constellation (16-QAM 64-QAM, 256-QAM, etc.) that is used onto subcarrier k : $R_{b,k} = R_M \cdot \log_2(M_k) = \Delta f \cdot N_{b,k}$. In xDSL, $\Delta f = R_M = 4.3125$ kHz and in the first commercially available version of ADSL $K = 256$, so that the overall bandwidth is $B = 1.104$ MHz and the downstream bit-rate is about 10 Mbit/s. The latest version is called ADSL2+ with $B = 2.208$ MHz and $R_b = 20$ Mbit/s, while the wideband version VDSL2 has $B = 12$ MHz and $R_b = 100$ Mbit/s on a shorter connection not exceeding 500 m.

RA-DMT needs accurate knowledge of the noise level on each subcarrier. To accomplish this, when the connection is started, the transmitter sends a preamble formatted with *uniform* power allocation on all carriers, and the receiver evaluates the signal-to-noise ratio on each subcarrier, sending such values back to the transmitter through the return channel. From the uniform-allocation SNRs, the transmitter derives the noise levels and performs power/bit allocation according to the WF criterion. The most “noisy” carriers, i.e. those with low SNR_k after power allocation, will use simple modulations with very few bits/symbol (BPSK, QPSK), or will be switched off altogether if the noise is exceedingly high. Vice versa, carriers with high SNR_k will use multi-bit/symbol constellation.

Just to make an example, DMT modulation for first-generation ADSL has a maximum of 32 subcarriers with a cyclic prefix of 5 upstream symbols and 256 subcarriers with a prefix cycle of 32 downstream symbols. Constellations can allocate up to a maximum of 15 bits per symbol ($M = 32768$ points !), with a special kind of channel encoding (trellis encoding with Viterbi detection). The general architecture of the RA-DMT link reflecting our description above is shown in Fig. 3.9.

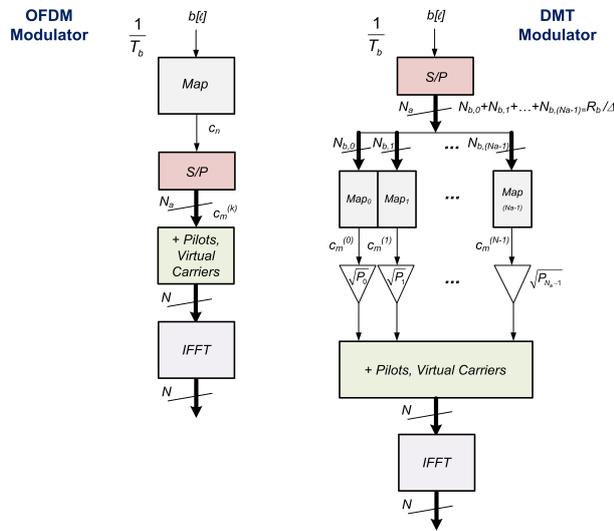


Figure 2.10 Comparison between OFDM and DMT

Contrary to what happens with OFDM, where the mapping is unique for all subcarriers (and can therefore be placed before the S/P converter), the “constellation encoder” (or mapper) block in DMT individually maps the input groups of bits to different QAM constellations according to the *bit allocation* outcome. In addition, the amplitude of the IDFT coefficients (which converts the frequency domain input to samples of the output time signal), will be different for each sub-band, according to the *power allocation* outcome - see Fig. 3.10.

Esempio 2.10

The ACGN Shannon capacity is

$$C_{ACGN} = \int_0^B \log_2 \left(1 + \frac{S_x(f)}{S_N(f)} \right) df \tag{2.20}$$

Can we derive from this (more) general expression the (simpler) formula for the AWGN channel ?

Concerning noise, we start by letting $S_N(f) = N_0/2$ into (3.20), but we also need to specify the signal spectral shaping $S_x(f)$ - this can be done by allocating our signal power according to the WF criterion:

$$S_x(f) = \bar{S} - N_0/2 \tag{2.21}$$

so that

$$C_{ACGN} = \int_0^B \log_2 \left(\frac{2\bar{S}}{N_0} \right) df = B \log_2 \left(\frac{2\bar{S}}{N_0} \right) \tag{2.22}$$

On the other hand,

$$\bar{S} = \frac{P_x + \sigma_N^2}{2B} = \frac{P_x + N_0B}{2B} \tag{2.23}$$

and therefore

$$C_{ACGN} = B \log_2 \left(\frac{2(P_x + N_0 B)}{2BN_0} \right) = B \log_2 \left(1 + \frac{P_x}{N_0 B} \right) \quad (2.24)$$

Q.E.D.

2.4 Fiber-Based UltraBroadBand (UBB) Connectivity fo the Access Network

Copper-based access networks, even with the most advanced versions of xDSL, namely VDSL2 with complicated DSP-based interference mitigation techniques (so-called *vectoring*) cannot break the 1Gbit/s barrier in terms of user bit-rate - a connection speed that is considered today (2022) UBB. The only way to provide such high-capacity is adopting optical technologies that, as already anticipated, are extending their reach beyond the traditional arena of backbone connectivity for the transport network to the last kilometer of access networks.

2.4.1 The Migration from Copper to Fiber in the Last Kilometer

Figure 3.11 shows from left to right the evolution of last-km technologies in chronological order of adoption and in increasing order of connection speed. In metropolitan areas encompassing a short distance between user terminals and the nearest cabinet/central office, ADSL is being replaced by a mixed copper/fiber technology, identified by the acronym FTTC, *Fiber To The Curb* or *Fiber To The Cabinet*, where the fiber connection is a bridge between the true operator PoP and a cabinet on the street shared by many users. The last hundreds of meters are covered by a broadband VDSL2 connections on copper, and we cannot label this architecture as truly UBB. Truly UBB technologies are based on the two rightmost technologies in Fig. 3.11, namely, FTTB *Fiber To The Building* or FTTH *Fiber To The Home*. In FTTB, the fiber comes to a residential building with a few to many subscriber properties, and the final connection to the end-user is just a so-called *vertical* connection from an optical network unit in the basement of the building up to the individual users' apartments. A vertical connection can be easily provided via Ethernet RJ cables and do not create any restrictions in term of user bit-rate. The ultimate UBB technology is clearly FTTH, where the *whole* access network is based on fiber, including vertical connections, so that the end-users has available a fiber connection right at the wall socket.

2.4.2 FTTH Passive Optical Networks

The dominant technology that is rapidly being adopted (especially in metropolitan areas) to provide FTTH UBB services is that of so-called Passive Optical Networks (PONs). A PON is a point-to-multipoint fiber network architecture to connect a group of end users with an Internet Point of Presence (PoP) belonging to a network operator (ISP, Internet Service Provider). The main feature of the PON, whose architecture is shown in Fig. 3.13, is that no active devices are used along the connection: the active devices (modems) only reside in the PoP and at the end users, while the network itself is made up of single-mode fiber segments and a series of passive optical splitters (an example is shown in Fig. 3.14). The optical

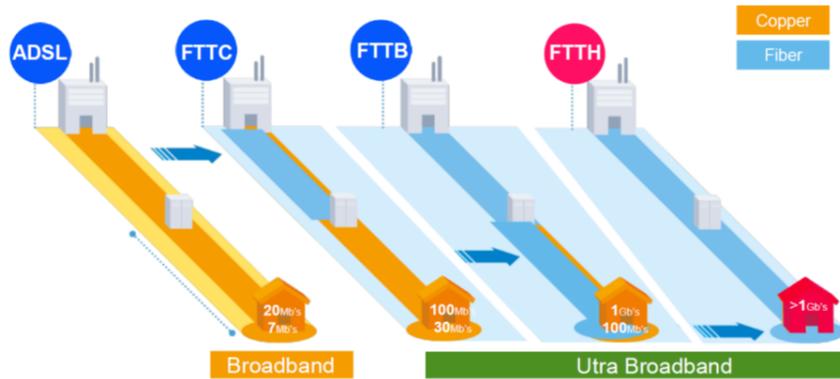


Figura 2.11 Evolution of Technologies for the Last kilometer

splitter is just a “fork” of fibers (from one into many) to propagate a single multiplexed digital optical signal towards a number of users, without any intermediate processing (Fig. 3.12). The PON inherits the high-speed performance of optical links, resulting in addition intrinsically not subject to interference (because based on fibre) and robust (because it is passive). Depending on the extension of the coverage area and on specifications about the users’ bit-rates, there may be several successive splitting stages to serve large groups of users - an example is shown in Fig. 3.15 which represents the architecture of a multi-stage PON. Figure 3.13 also shows the appropriate naming of PON devices: the ONTs, *Optical*

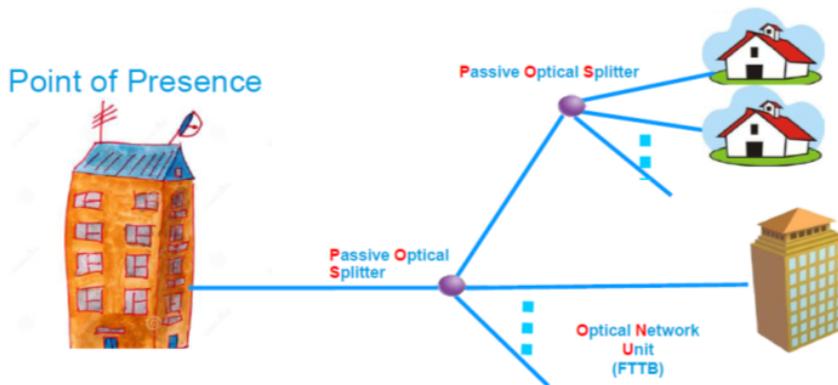


Figura 2.12 General Architecture of a PON for the provision of FTTH Services

Network Terminators (aka ONU, *Optical Network Units*), are the user-side modems, whilst the OLT, *Optical Line Termination*, is the analogue of the DSLAM of xDSL networks, i.e. the flow aggregator and the interface to the external IP-based transport network. It is also seen that the PON supports both communication directions on a single fiber using a

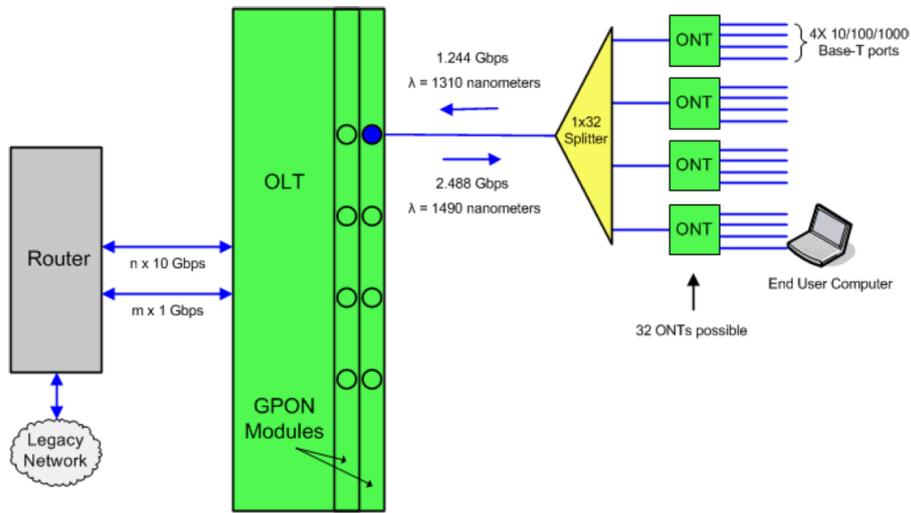


Figure 2.13 General Architecture of a PON for the provision of FTTH Services



Figure 2.14 1 x 4 Passive Optical Splitter (from <http://www.fs.com>)

Wavelength-Division Duplexing (WDD) approach, in which the downstream takes place in the third window and the upstream in the second window, respectively.

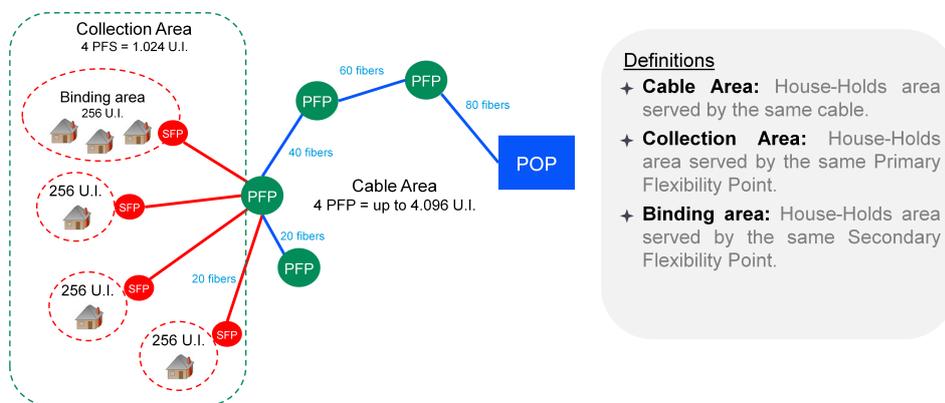


Figura 2.15 Multi-Stage PON Example (courtesy of OpenFiber S.p.A.) - POP=Point Of Presence, PFP=Primary Flexibility Point, SFP=Secondary Flexibility Point, U.I.=Unità Immobiliare (Real-Estate Unit)

2.4.3 The dominant technology for FTTH: ITU's Gigabit PON (G-PON)

All of the architectural schematics above already referred to the PON standard that is widely being adopted worldwide: ITU-T G.984 G-PON. From the standpoint of communication technologies, the G-PON is based on conventional IM/DD modems over standard G.652 single-mode fiber with WDD to provide full-duplex. Digital streams are protected with Reed-Solomon (255,239) block codes, and the nominal capacity offered to the end-user is 2.4 Gbit/s downstream and 1.2 or 2.4 Gbit/s upstream depending on implementation - that's why we speak of UBB. The maximum connection distance (from OLT to ONU) is 20 km, much longer than in copper-based ADSL networks. Multiple-access is accomplished via

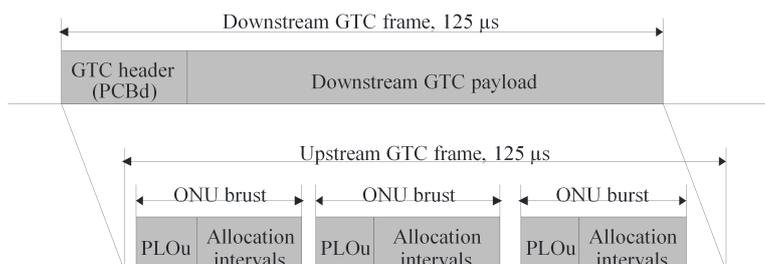


Figura 2.16 Formato delle trame G-PON Transmission Convergence layer (GTC) up/downstream

the a classic TDM/TDMA with a centrally formatted 125 μs frame downstream from the ONT, and multiple bursts of upstream TDMA data, still organized into an upstream 125 μs frame (3.16). The upstream synchronization of ONUs in the TDMA frame is very simple and effective and will not be treated here.

From what we have said until now it is pretty clear that the basic technologies of G-PON are quite consolidated; the most complex issue in the development of PON in particular and of FTTH in general is *economic* and resides in the large investments that are necessary to

replace the decades-old copper connections with new fiber-based one. After the deployment of G.PON architectures is accomplished, operators will have the opportunity to upgrade the capacity of the network without the need to further replace the physical medium. The roadmap of such development, that in some areas has already started, is shown in Figure 3.17. Broadly speaking, the penetration of FTTx technologies worldwide is highly variable from area to area within a country, and from country to country, creating further “divide” in the opportunity to access information services by the end-user.



Figura 2.17 Evolution Roadmap of PON Technologies

BIBLIOGRAFIA

1. M. Luise, G. Vitetta, "*Teoria dei Segnali*", 3a Edizione, Mc Graw Hill International Editions, 2009.
2. J. Proakis, M. Salehi, "*Digital Communications*", 5th Edition, Mc Graw Hill International Editions, 2018.
3. D. Applebaum, "*Probability and Information* ", 2nd Edition, Cambridge University Press, 1996.
4. A. Goldsmith, "*Wireless Communications*", Cambridge University Press, 2005.

