

# Mobile communications



## Information theory and digital transmission (elements)

**Politecnico di Torino**  
**2006**



# Elements of Information Theory

## Discrete Information Source

Source symbol – a random variable  $X$  assuming values  $x_1, x_2, x_3, \dots$

$x_i$  belongs to an alphabet  $\Xi \equiv \{\xi_1, \xi_2, \dots, \xi_M\}$

$P_X(x_i) = P(x_i)$  probability that the outcome of  $X$  is  $x_i$

The sequence of symbols  $x_1, x_2, x_3, \dots, x_k$

shows a probability of occurrence  $P(x_1, x_2, \dots, x_k)$

2



The data generated by a discrete can be described by making use of a discrete random variable. The random variable assumes with a given probability values belonging to an alphabet. The alphabet and the probability associated to the r.v. outcomes characterise the source.

A source is said to be *stationary* if:

$$P(x_1, x_2, \dots, x_k) = P(x_{1+h}, x_{2+h}, \dots, x_{k+h})$$

For any integer  $k$  and  $h$ .

The source generates independent symbols if:

$$P(x_1, x_2, \dots, x_k) = P(x_1) \cdot P(x_2) \cdot \dots \cdot P(x_k)$$

# Information and Entropy

The information content associated to a single event (symbol) is defined as:

$$I(x_i) = -\log_2 P(x_i) = \log_2 \frac{1}{P(x_i)} \quad (\text{bit})$$

The information contained in two independent events:

$$I(x_i, x_j) = I(x_i) + I(x_j) \quad (\text{bit})$$

The entropy of a random variable is defined as:

$$H(X) = -\sum_{i=1}^M P(x_i) \cdot \log_2 [P(x_i)] = E[I(X)] \quad (\text{bit/symbol})$$

3



The information carried by the  $i$ -th symbol is  $\geq 0$  and coincides to 0 just in case the occurrence probability of that symbol is 1:

$$P(x_i) = 1$$

Being  $M$  the possible outcomes (the possible symbols), they are identified in a binary representation, by *words* of

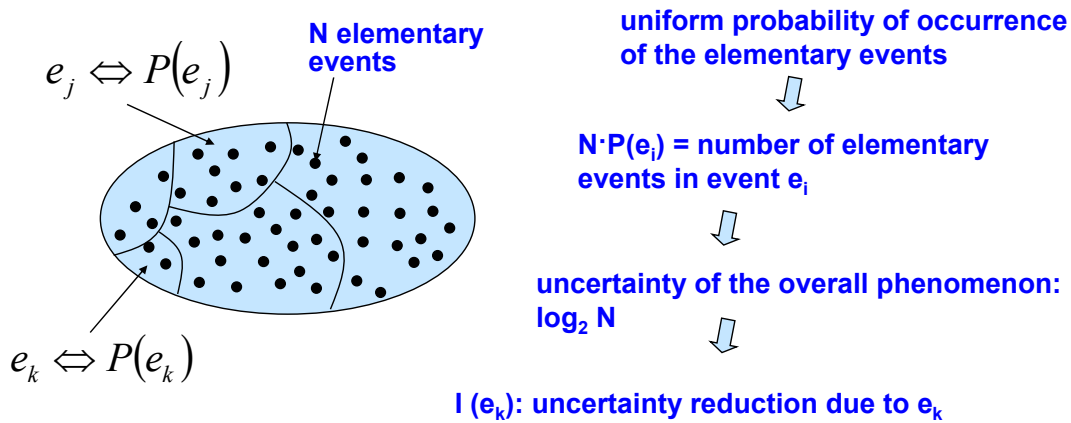
$$\log_2 M$$

bits. It can be intuitively said that the lower is the outcome probability, the higher is the information carried by that event. The information is defined as a logarithmic function of the outcome probability. The information associated to two independent events of probability  $P(x_i) \cdot P(x_j)$ , is:

$$I(x_i, x_j) = -\log_2 [P(x_i) \cdot P(x_j)] = I(x_i) + I(x_j)$$

The entropy is tied to the average content of information generated by a source according to an alphabet of  $M$  symbols.

## Information through an example



$$I(e_k) = \log_2 N - \log_2 [N \cdot P(e_k)] = -\log_2 [P(e_k)]$$

4

TELECOM LAB  
ITALIA

The information represents the reduction of uncertainty deriving from the occurrence of a given event. In the example of the above figure, the occurrence of the event  $e_k$  reduces the uncertainty in relation to the overall phenomenon

Through these considerations it is clear how the information tied to a given event is increasing with the decreasing of the outcome probability of the event itself.

## Some properties of entropy

**limits**  $0 \leq H(X) \leq \log_2 M \Rightarrow$  Maximum value reached with uniform distribution:  $P(x_i)=1/M; i=1,2,...M$

**convexity**  $H(\alpha X + \beta Y) \geq \alpha H(X) + \beta H(Y)$

$$H(X, Y) = - \sum_{x,y} P(x, y) \log_2 P(x, y) = -E[\log_2 P(x, y)]$$

**example**

$$\left. \begin{array}{l} x_1 \rightarrow P(x_1) = 0.5 \\ x_2 \rightarrow P(x_2) = 0.5 \end{array} \right\} H(X) = 1$$

$$\left. \begin{array}{l} x_1 \rightarrow P(x_1) = 0.1 \\ x_2 \rightarrow P(x_2) = 0.9 \end{array} \right\} H(X) = 0.469$$

5



The entropy of a random variable  $X$  is a symmetric function. It reaches the maximum value when the random variable is characterised by a uniform distribution.

The entropy represents the mean level of un-determination generated by the source (it is expressed in bits due to the binary base chosen for its representation)

The joint entropy represents the un-certainty about the two joint variables  $(X, Y)$ . In case the two variables are independent:

$$H(X, Y) = H(X) + H(Y)$$

The expression may be extended to any number of independent random variables.

## Some properties of entropy (I)

### conditional entropy

$$H(X/Y) = -\sum_{x,y} P(x,y) \cdot \log_2 [P(x/y)] = -E[\log_2 P(X/Y)]$$

$$H(X,Y) = -\sum_{x,y} P(x,y) \log_2 P(x,y)$$

$\uparrow$   
 $P(y) \cdot P(x/y)$   
 $\downarrow$

$$H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y)$$

The conditional entropy represents the uncertainty of X conditioned to the knowledge of Y. The relationship arising between  $H(X,Y)$ ;  $H(X)$  and  $H(X/Y)$ , states that the uncertainty of (X,Y) is the sum of the uncertainty of X and the uncertainty of Y given X.

## Source coding

average entropy per symbol:

$$\bar{H} = \lim_{k \rightarrow \infty} \frac{1}{k} H(X_0, X_2, \dots, X_k)$$

average # of bits used to code symbols:

$$\bar{n} = \sum_{i=1}^M P(x_i) n_i$$

For stationary and independent sources, some source coding exists such that:

$$H(X) \leq \bar{n} \leq H(X) + 1$$

← Shannon theorem

$\bar{n}$  may approach indefinitely  $H(X)$  through appropriate source codes

coding efficiency:  $\frac{H(X)}{\bar{n}}$

7



A source generates un-limited sequences of symbols. It is suitable to define the mean number of bits per symbol, through a limit function (figure). For an independent source, symbols are independent hence:

$$\bar{H} = \lim_{k \rightarrow \infty} \frac{1}{k} H(X_0, X_2, \dots, X_k) = H(X)$$

The source can be encoded by making use of a code of fixed length. In this case, if

$$\Xi \equiv \{\xi_1, \xi_2, \dots, \xi_M\}$$

is the alphabet used, each symbol must be represented by a binary vector whose length is at least  $\log_2 M$  ( $\log_2 M$  is the minimum number of bits per symbol needed to encode the source of information).

Conversely, a variable length encoding can be adopted. For instance, one can use  $n_i$  bits to encode the symbol  $x_i$ , where  $n_i$  is tied in some way to the probability  $P(x_i)$  of occurrence of  $x_i$ .

If the source symbol are independent and identically distributed, then the efficiency of the source encoding is measured by making use of:

$$\bar{n} = \sum_{i=1}^M n_i P(x_i)$$

An optimal source encoding is the one that minimises the mean number of bits per symbol.

## Examples

Let the alphabet be:

$A, B, C, D, E$

with probability  
of occurrence:

$0.6, 0.1, 0.1, 0.1, 0.1$

fixed length coding

variable length coding

$A$  000

0

$B$  001

100

$C$  010

101

$D$  011

110

$E$  100

111

$\bar{n} = 3$

$\bar{n} = 1.8$

Shannon limit  $H(X) = 1.77$

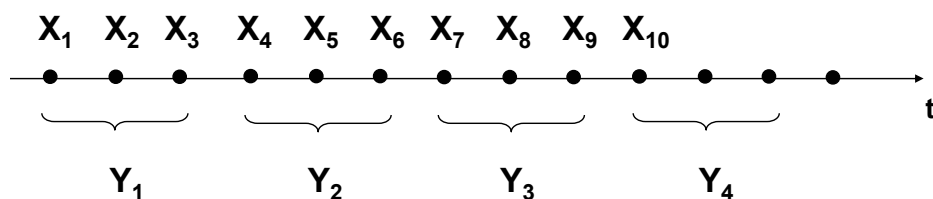
coding efficiency  $\eta = \frac{1.77}{\bar{n}}$

The example shows how the use of a non-uniform source encoding can reduce the mean number of necessary bits.

The length of the word encoding the single symbol must be a function of the occurrence probability of the symbol itself.

# Approaching the Shannon limit through block coding

X independent source



Source symbols are grouped in blocks of  $m$  symbols each and coded with  $\bar{n}_m$  bit in average per symbol

Shannon theorem

applied to the block coding  $\Rightarrow H(Y) \leq \bar{n}_m \leq H(Y) + 1$

with  $\Rightarrow H(Y) = m \cdot H(X)$

9

$$H(X) \leq \bar{n} \leq H(X) + \frac{1}{m}$$

TELECOM LAB  
ITALIA

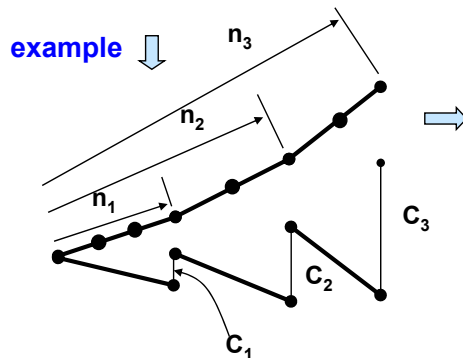
The Shannon limit can be indefinitely approached if a block encoding is adopted.

# Optimal encoding

Binary code:

for an alphabet of M symbol, each one coded with a word of  $n_i$  bit ( $i=1,2,\dots,M$ )

.....the code existence condition is:  $\sum_i c_i \cdot 2^{-n_i} \leq 1$



$$c_3 + c_2 \cdot 2^{(n_3 - n_2)} + c_1 \cdot 2^{(n_3 - n_1)} \leq 2^{n_3}$$

$c_i$  : number of symbol coded with a code of depth  $n_i$

10

TELECOM LAB  
ITALIA

As said, aiming at an optimal source encoding means to adopt a code tied to the occurrence probability of the symbols.

If a binary tree is used to describe the code, the condition of the figure guarantees the existence of the code.

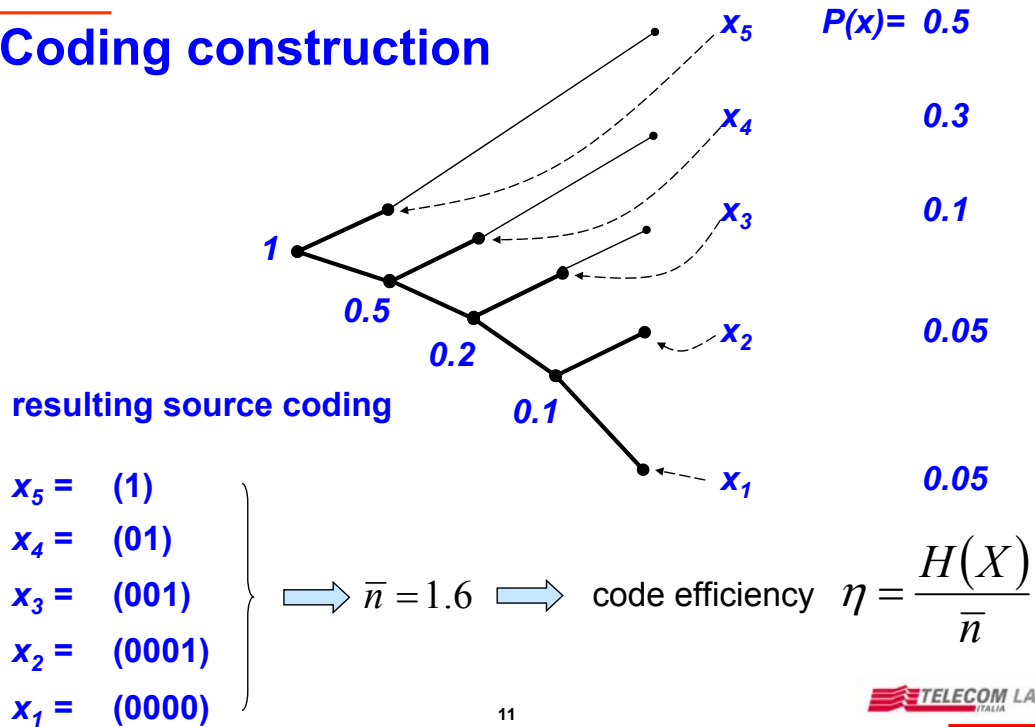
For each code of depth  $n_i$ ,  $c_i$  symbols are encoded by using  $c_i$  words among the  $2^{n_i}$  available (*leaves*) at that depth level.

The number of leaves  $c_i$  utilised at the depth  $n_i$  inhibits the use of:

$$c_i \cdot 2^{(n_m - n_i)}$$

*leaves* of depth  $n_m$  ( $n_3$  in the figure).

## Coding construction

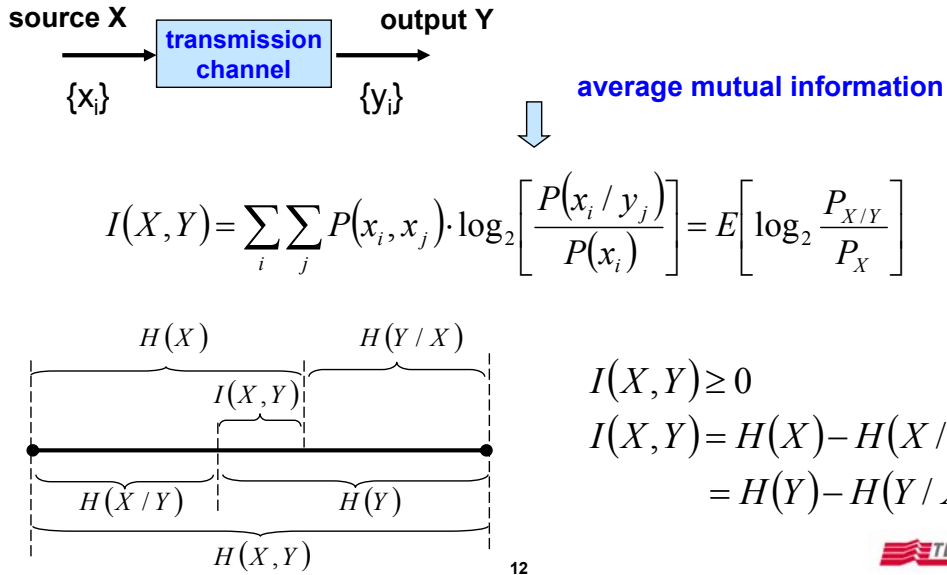


An optimal source code is the one minimising the mean number of bits per symbol needed to represent the source itself.

The method reported in the figure (Huffman) allows to construct an optimal encoding. It is based on the association of the leaves of a binary tree to the symbols, performed with the criterion of assigning the higher depth to the less likely symbols.

The procedure continues until the tree root is reached. The symbols are arranged in the decreasing order of occurrence probabilities.

## Average mutual information



The figure represents a discrete transmission system. The *mutual information* arising between two symbols  $x_i$  and  $y_j$  respectively *transmitted* (input) and *received* (output) through the system, is defined as the measure of the information transferred by the channel when, being  $x_i$  transmitted,  $y_j$  is received.

$I(x_i / y_j)$  is the measure of the uncertainty remaining on  $x_i$  being  $y_j$  the received symbol.

$$I(x_i, y_j) = I(x_i) - I(x_i / y_j) = -\log_2 P(x_i) + \log_2 P(x_i / y_j) = \log_2 \left[ \frac{P(x_i / y_j)}{P(x_i)} \right]$$

If no error arise on the channel, then:

$$I(x_i / y_j) = 0 \quad \text{in fact : } P(x_i / y_j) = 1; \text{ from which : } I(x_i, y_j) = I(x_i)$$

If  $y_j$  is independent from  $x_i$  we have:  $I(x_i, y_j) = 0$ . By using the expression:

$$I(x_i / y_j) = I(x_i) - I(x_i, y_j)$$

The mean value of the mutual information  $I(X, Y)$  in function of the entropy of  $X$  and of the entropy of  $X$  conditioned to  $Y$ :

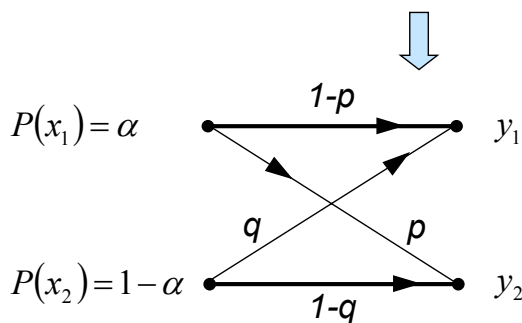
$$H(X / Y) = H(X) - I(X, Y)$$

The latter expression indicates that the mean information per symbol transferred through the channel equals the average source entropy diminished by the conditional entropy  $H(X/Y)$ . Being  $I(X, Y) \geq 0$ , we have:  $H(X) \geq H(X/Y)$  that is, the conditioning effect does not increase the entropy. In other words, the reception of  $Y$  contributes in any case to reduce the uncertainty on the symbols generated at the input of the channel.

# Channel capacity

**Channel capacity:**  $C = \max_{P(x_i)} I(X, Y)$

**Example – binary channel**



**Discrete channels:**

- binary (symmetric-non symmetric)
- binary with cancellation
- generic symmetric
- symmetric with cancellation and error

**$p, q$  error probabilities**

The objective is of course determining the input  $X$  given  $Y$ .

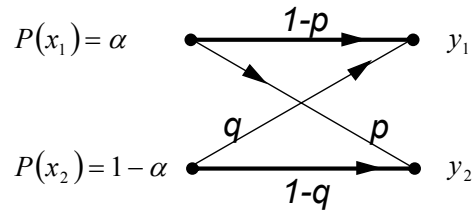
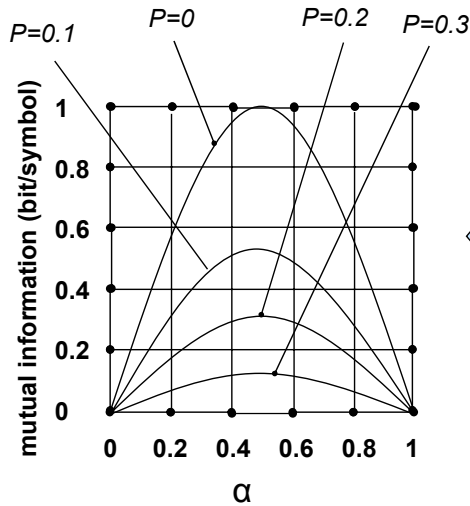
This lead to minimise  $H(X/Y)$  namely the uncertainty on  $X$  having received  $Y$ .

Being  $H(X/Y) = H(X) - I(X, Y)$  it would be needed to maximise  $I(X, Y)$ , namely the mean information per symbol transported by the channel.

However, it is impossible increasing the mutual information through operations applied once the signal has been received (this is an intuitive statement).

Let's assume that an alphabet has been defined for the input symbols and an alphabet has been defined for the output symbols. Given the transmission channel, the mutual information is just function of the symbol emission probability of the source. The channel capacity is just function of this probability.

## Binary channel



$$C = \max_{\alpha} [H(Y) - H(Y/X)]$$

with  $p = q$ :

$$C = 1 + p \log_2 p + (1-p) \log_2 (1-p)$$

In the example above, the channel capacity is calculated in function of  $\alpha$ .

$$P(y_1) = \alpha(1-p) + (1-\alpha)q; \quad P(y_2) = 1 - [\alpha(1-p) + (1-\alpha)q]$$

from which:

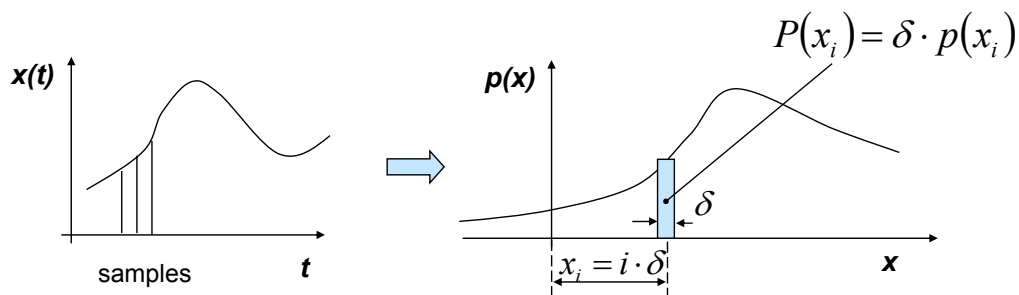
$$H(Y) = -P(y_1) \log_2 P(y_1) - P(y_2) \log_2 P(y_2)$$

$$\begin{aligned} H(Y/X) &= -\sum_i \sum_j P(x_i, y_j) \cdot \log_2 P(y_j/x_i) = \\ &= -\sum_i P(x_i) \sum_j P(y_j/x_i) \log_2 P(y_j/x_i) \end{aligned}$$

The derivative of  $I(X,Y)$  is then put to 0 and the related value of  $\alpha$  is calculated (function convexity):

$$\frac{\partial I(X,Y)}{\partial \alpha} = 0$$

## Sources and channels – the continuous case



**absolute entropy:**  $H_{abs}(X) = \lim_{\delta \rightarrow 0} \left[ - \sum_i p(x_i) \cdot \delta \cdot \log_2(p(x_i) \cdot \delta) \right]$

**developing  $\log_2$  and substituting  $\Sigma$  with  $\int$  absolute entropy may be written:**

$$H_{abs}(X) = H(X) + H_R(X) \text{ where: } \begin{cases} H(X) = - \int_{-\infty}^{+\infty} p(x) \log_2 p(x) dx \text{ (bit/sample)} \\ H_R(X) = - \lim_{\delta \rightarrow 0} \log_2 \delta \text{ (bit/sample)} \end{cases}$$

In the case of a continuous source of information, the sampling theorem must be applied to reduce the generated information to a sequence of discrete signals.

Let  $x(t)$  and  $B$  respectively the signal and the bandwidth generated by the source. The sampling theorem allows to represent exhaustively the signal through samples performed with a period of  $1/2B$ .

The sample values are described by a random variable  $X$  characterised by a defined probability distribution  $p(x)$ . The process is assumed to be:

- stationary ( $p(x)$  does not depend on time)
- without memory

The random variable  $X$  is approximate through a discrete random variable assuming the following values:

$$x_i \equiv i \cdot \delta; \quad i = 0, \pm 1, \pm 2, \pm 3, \dots \text{ with } P(x_i) \cong p(x_i) \cdot \delta$$

If we look at the expression calculated for the absolute entropy (figure),  $H_R(X)$  represents the reference entropy of the random variable  $X$ . with  $\delta \rightarrow 0$   $H_R(X)$  goes to infinity. The random variable  $X$  is in fact continuous and the samples take an infinite number of values.

$H(X)$  is the differential entropy (it is the differentiating element among different sources). Note that  $H(X)$  can assume positive or negative values in function of  $p(x)$ .

# Max entropy for a limited power source

$p(x)$  that maximises relative entropy  $H(X)$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

Gaussian function with

- average value = 0
- variance  $\equiv$  average source power

corresponding to the relative entropy:

$$H(X) = \frac{1}{2} \log_2(2\pi e \sigma^2)$$

16

TELECOM LAB  
ITALIA

To calculate the function  $p(x)$  that maximises the entropy previously defined, the Lagrange multipliers method is used. The latter is based on the fact that the maximum value of  $H(x)$  results from both the (unknown)  $p(x)$  maximising the entropy and from a function including the constraints the  $p(x)$  must obey to. If  $S$  is the mean power generated by the source, these constraints are:

$$S = \int_{-\infty}^{+\infty} x^2 p(x) dx \quad \text{and} \quad \int_{-\infty}^{+\infty} p(x) dx = 1$$

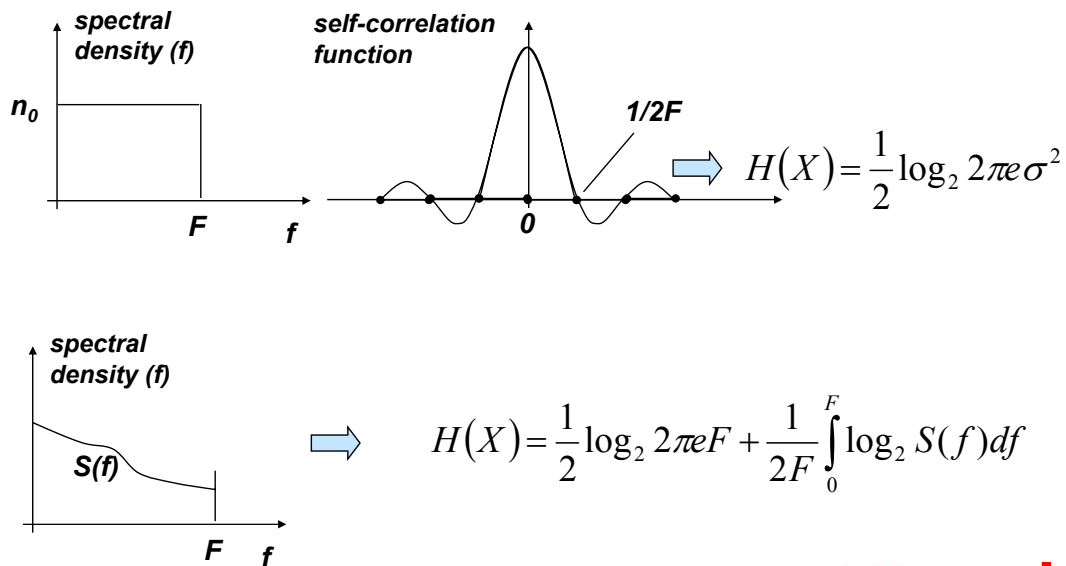
To obtain the maximum value of the entropy, the distribution  $p(x)$  obtained in this way is to be used in the formula giving the relative entropy.

Note that the relative entropy of a source does not change in case the source distribution  $p(x)$  has a non null mean  $\mu$  (translation of  $\mu$  with respect to 0). From the definition of entropy it is in fact apparent that the latter is invariant with respect to translations. But the mean-translated source utilises a transmission power that is greater than the zero-mean source:

$$E[x^2] = \sigma^2 + \mu^2$$

On the signal transmission side, it is then essential to use zero-mean functions.

## Entropy vs. spectrum



17

TELECOM LAB  
ITALIA

The maximum value of the relative entropy has been calculated under the assumption that the samples of the original function were independent. It can be shown that the independence condition holds just when the function has a uniform spectral density  $n_0$  (figure). In fact, in this case, the self-correlation function of the original function vanishes in the multiples of the sampling time  $t = i \cdot 1/2F$  if  $F$  is the band of the process (figure).

Whenever the gaussian process does not a uniform spectrum, the relative entropy becomes:

$$H(X) = \frac{1}{2} \log_2 2\pi e F + \frac{1}{2F} \int_0^F \log_2 S(f) df$$

The expression assumes its minimum value when  $S(f)$  becomes:

$S(f) = n_0$  in fact  $\because n_0 \cdot F = \sigma^2$   
From the above remarks, it can be noted that the gaussian process  $S(f)$  (with mean power  $S$ ) that maximises the relative entropy must have a uniform spectral density.

# Mutual information and channel capacity

average mutual information:

$$I(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{X,Y}(x, y) \log_2 \frac{p_X(x/y)}{p_X(x)} dx dy$$

$\swarrow$   
 $\frac{p_Y(y/x)}{p_Y(y)}$

relation between mutual information, relative entropy, conditional entropy:

$$I(X, Y) = H(Y) - H(Y/X) = H(X) - H(X/Y)$$

18



The transfer of the information  $x(t)$  generated by the source on a continuous channel generates as channel output, a signal  $y(t)$  whose two-way association with  $x(t)$  must be established by the receiving side.

The average mutual information maintained by  $Y$ , being  $X$  the transmitted signal has already been considered in the discrete case.

The relationship between  $I(X, Y)$ ,  $H(X)$ ,  $H(X/Y)$  and those equivalent conditioned to  $X$  are obtained by introducing the *information* (definition):

$$I(x, y) = I(x) - I(x/y)$$

Note that  $I(X, Y)$  is, as in the discrete case, the difference between two entropies. With reference to the absolute entropy, one can argue that the mutual information  $I(x, y)$  eliminates this divergence.

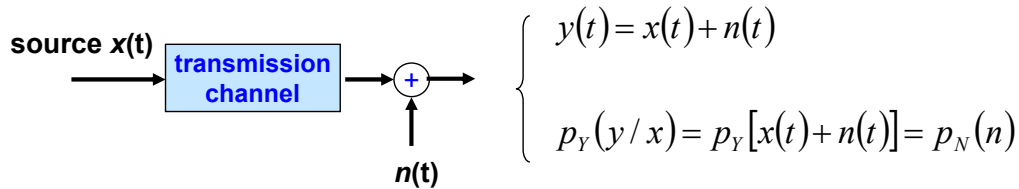
By integrating over  $x$  and  $y$  the previous expression becomes:

$$I(X, Y) = H(X) - H(X/Y)$$

where:

$$H(X/Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{XY}(x, y) \log_2 p_Y(y/x) dx dy = - \int_{-\infty}^{+\infty} p_X(x) \int_{-\infty}^{+\infty} p_Y(y/x) \log_2 p_Y(y/x) dx dy$$

## Mutual information and channel capacity (II)



channel capacity:

$$C = \max_{p(x)} [I(X, Y)] \quad \text{bit/sample}$$

19



Now the channel and its characteristics are considered. It is assumed that the noise introduced by the channel is of additive nature. This means that the channel noise is independent from the input signal. If  $p_N(n)$  is the probability density function of the additive noise, one has:

$$H(Y/X) = - \int_{-\infty}^{+\infty} p_N(n) \cdot \log_2 p_N(n) dn$$

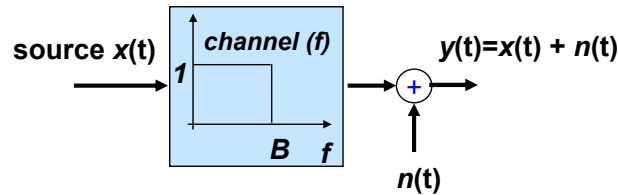
Hence, with reference to the relationship arising between the mutual entropy, the relative entropy and the conditional entropy, the following is obtained:

$$I(X, Y) = H(Y) - H(N)$$

The channel capacity, defined as the maximum of the information that can be transferred is calculated similarly to the discrete case:

$$C = \max_{p(x)} [H(Y) - H(N)] = \max_{p(x)} [H(Y)] - H(N)$$

## Channel with additive gaussian noise



**channel capacity:**  $C = \frac{1}{2} \log_2 2\pi e(S + N) - \frac{1}{2} \log_2 2\pi eN = \frac{1}{2} \log_2 \left( 1 + \frac{S}{N} \right)$  (bit/sample)

**pace of sampling: 2B**



$$C = B \log_2 \left( 1 + \frac{S}{N} \right) \text{ (bit/s)}$$

**Hartley-Shannon law**

20

**TELECOM LAB**  
ITALIA

It is assumed that no distortion is introduced by the channel in the available spectrum  $B$  and the amplitude of the signal is maintained (the attenuation is recovered through an ideal amplifier). In addition, the additive noise is supposed to be gaussian between  $0$  and  $B$ , with zero mean value and average power  $S = E[X^2]$  fixed.

Since signal and noise are independent, the average power at the output of the channel is:

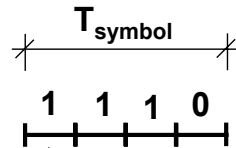
$$E[Y^2] = E[X^2] + E[N^2] = S + N$$

The maximum value of  $H(Y)$  calculated over  $p(x)$  is the obtained when  $p_Y(y)$  is gaussian with zero mean and spectrum uniformly distributed (independent samples).

Hence,  $p(x)$  must be gaussian so that, together with the additional gaussian noise, on the receiving side, a function  $p(y)$  is produced which is gaussian by turn.

The values of  $C$  in terms of [bits/sample] and [bit/s] are those indicated in the figure.

# Spectral efficiency



$$T_{bit} = \frac{1}{R} = \frac{T_{symbol}}{m} = \frac{1}{mR_{symbol}}$$

M symbols  
m bits per symbol;  $m = \log_2 M$

bit rate  $R = \frac{m}{T_{symbol}} \text{ (bit/s)}$

**Spectral efficiency:**  $\frac{1}{BT_{bit}} \left[ \frac{\text{bit/s}}{\text{Hz}} \right]$

B: channel bandwidth  
 $1/T_{bit}$ : bit rate

21



As said, the information bits are transmitted on the channel by making use of an alphabet composed by M configurations of waveforms (different phases, amplitudes,..) each configuration bears m bits, where:

$$M = 2^m$$

The figure represents the elementary liaisons incurring between the bit rate and the symbol period  $T_{symbol}$ .

To increase the bit rate R, the bandwidth assigned to the channel must be increased (Shannon).

A good measure of the bit transmission efficiency is given by the spectral efficiency, which represents the number of bits per second that can be transmitted per each Hz of the assigned bandwidth.

## Spectral efficiency (II)

limiting channel capacity with  
mean signal power  $S$   
( $n_0$  noise spectral power  
density)

$$\lim_{B \rightarrow \infty} C = \lim_{B \rightarrow \infty} \left[ B \cdot \log_2 \left( 1 + \frac{S}{N} \right) \right] = \frac{1}{\ln 2} \cdot \frac{S}{n_0}$$

energy  $E$  needed to transmit  
a single bit

$$E_b = n_0 \cdot \ln 2 = n_0 \cdot 0.693 \Rightarrow \frac{E_b}{n_0} = 0.693 \Rightarrow \frac{E_b}{n_0} (dB) = -1.6 dB$$

22

TELECOM LAB  
ITALIA

The Shannon theorem doesn't say anything on the encoding scheme that has to be adopted to reach a given channel capacity.

With a fixed signal power  $S$ , the channel capacity grows with the growth of the bandwidth  $B$  assigned to the channel. In spite of that,  $C$  is limited. In fact, while  $B$  tends to infinity,  $C$  tends to a limited value (figure).

To find the limiting value, an un-defined form has to be resolved (De L'Hospital). Note that the average noise power can be expressed as the product of the noise spectral density by the band  $B$ .

$$N = n_0 \cdot B$$

The above limit is then:

$$C_{\infty} = \lim_{B \rightarrow \infty} \left[ \frac{\log \left( 1 + \frac{S}{n_0 B} \right)}{1/B} \right] = \lim_{B \rightarrow \infty} \frac{1}{1 + \frac{S}{n_0 B}} \cdot \ln e \cdot \left( \frac{S}{n_0} \right) = \frac{S}{n_0} \cdot \frac{1}{\ln 2}$$

We've seen that the Shannon formula holds for an error-free transmission (or assuming an unlimited possibility of adopting error correction codes); additive gaussian noise and non correlated to the signal; a distortion-free (amplitude and phase) channel.

The above limit says that, in order to obtain a given transmission capability with a low spectral efficiency ( $C/B \rightarrow 0$ ) a minimum value of  $E_b/n_0$  must be guaranteed.

## Spectral efficiency (III)

Communication efficiency

$$\beta = \frac{S}{N} \cdot \frac{B}{C} = \frac{S}{n_0 C} = \frac{E_b}{n_0}$$

Shannon formula

$$\frac{C}{B} = \log_2 \left( 1 + \beta \cdot \frac{C}{B} \right) = \log_2 \left( 1 + \frac{S}{n_0 C} \cdot \frac{C}{B} \right)$$

- $S/n_0$  tied to the error probability
- $C/B$ =spectral efficiency  $1/(B \cdot T_{\text{bit}})$  previously introduced, with  $R=C$  (ideal case)

23

TELECOM LAB  
ITALIA

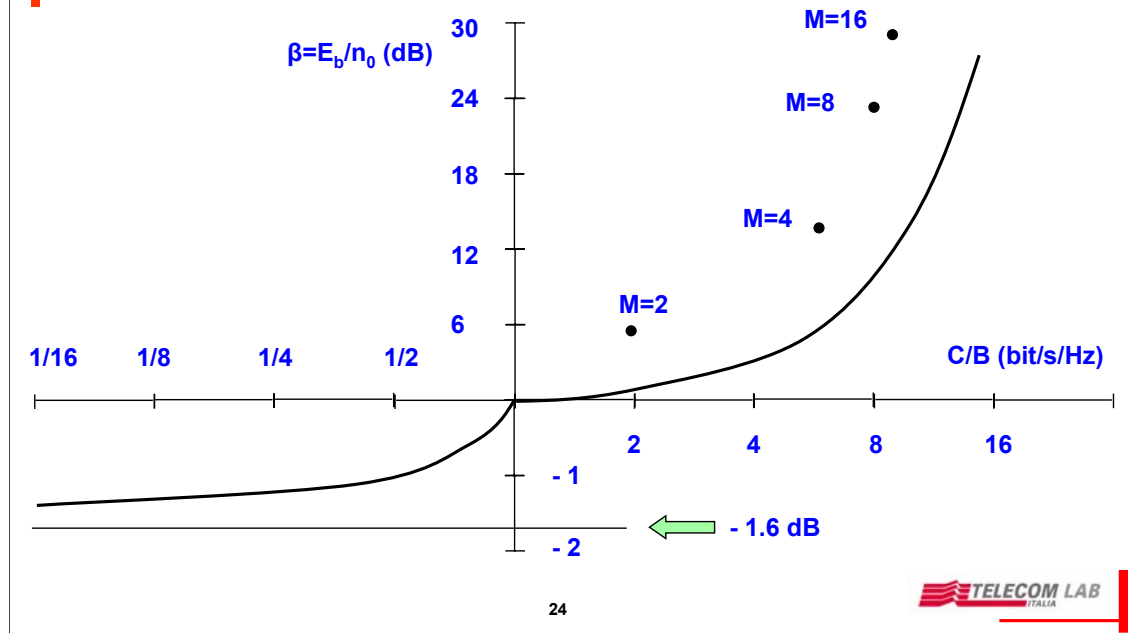
A good measure of the communication efficiency in a real system is represented by the product of  $S/N$  by  $B/C$ . The former factor measures the rate between the mean power of the signal and the mean power of the noise which is necessary to guarantee a given error probability. The second factor measures the number of Hz needed for each bit/s. The communication efficiency is then calculated as indicated in the figure.

In other terms, the communication efficiency is the ratio between the necessary (received) energy per bit and the spectral noise density.

The minimum value of this energy has been obtained in the case when there are no band limitations ( $\beta=0.693$ ) and using the expression of Shannon for the limit capacity  $C$ .

Still through the Shannon formula, the two ratios  $S/N_0$  and  $C/B$  can be put in relation. The Shannon limit is approached through suitable modulation techniques that allow to transmit several bits per symbol (the most suitable modulation technique depends on the value of  $\beta$ ).

## Spectral efficiency (IV)



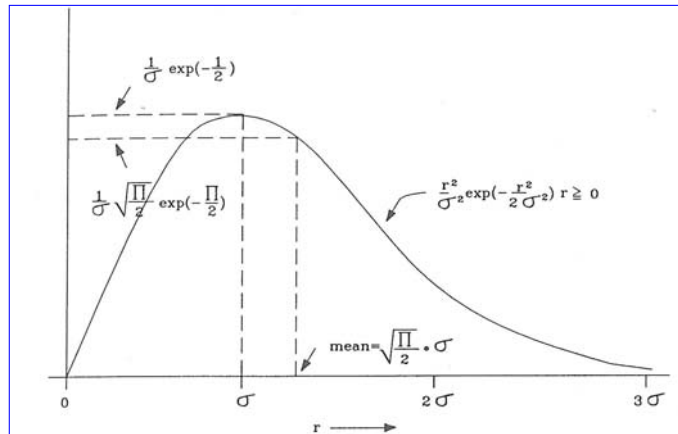
The figure represents the formula obtained in the previous figure. The value of  $S/N_0$  (hence the error probability) has been fixed. The upper arc compares the theoretical value of Shannon (the curve) with the values obtained through some modulation schemes, defined by the value of  $M$  ( $M$  is the number of modulation levels, able to represent  $m = \log M$  bits per symbol).

The lower arc represents (in another scale) the ratio  $E_b/n_0$  when no band limits exists. (the value tends to the limiting value of  $-1.6$  dB).

The upper arc shows that for great values of the ratio  $C/B$ , the increase of 1 bit/Hz requires an increase of about 3 dB of the ratio  $E_b/n_0$ .

# Spectral efficiency

Rayleigh distribution (dynamic conditions):



$$p(\rho / m) = 2\rho \cdot e^{-\rho^2}$$

$$\sigma = 4m_d^2 / \pi$$

$$\rho = r / \sqrt{\sigma}$$

25

TELECOM LAB  
ITALIA

The radio channel induces on the carrier amplitude, frequency and phase modifications. These modifications are limiting its capability of transporting information.

The interference effect produced through the frequency re-use is measured by the parameter C/I (*Channel to Interference*) which measures the ratio between two powers: that of the useful channel and that of the *interference*.

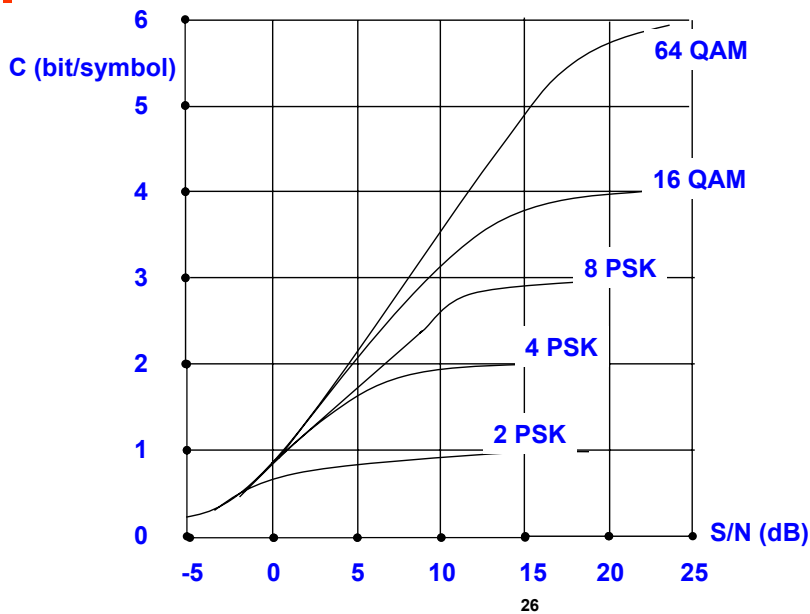
When steady state conditions hold, the noise on the channel is a white noise with gaussian distribution.

Besides, the channel is subject to the attenuation undergone by the electromagnetic signal.

In dynamic conditions, the Rayleigh attenuation (short term) is added (see e1- *propagation models*).

The Shannon formula represents an upper limit for the channel capacity. The dynamic channel behaviour is exactly the most important limiting factor.

## Discrete input and continuous output



In the real case, the source is a discrete source which generates a discrete symbol which is by turn transformed in e.g. the phase of a continuous function. This is transmitted and received as a continuous waveform.

A discrete input is the translated in a continuous output.

In this case, the joint probability of generating a symbol  $x_i$  as input function and of receiving a symbol  $y$  as output is:

$$P(x_i, y) = P(x_i) p_{Y/X}(y/x_i)$$

And the mutual information is calculated as:

$$I(X, Y) = \sum_{i=1}^N P(x_i) \int_y p_{Y/X}(y/x_i) \cdot \log_2 \frac{p_{Y/X}(y/x_i)}{\sum_{i=1}^N P(x_i) p_{Y/X}(y/x_i)} dy$$

## Transmission-modulation

$s(t)$  defined over  $(0,T)$  with finite energy

$$E_s = \int_0^T |s(t)|^2 dt$$

orthogonal signals



$$\psi_i(t) : \int_0^T \psi_i(t) \cdot \psi_j(t) dt \rightarrow \delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

each  $s(t)$  may be defined through a linear combination of  $\psi_i(t); i = 1, 2, \dots, M$



$$s(t) = \sum_{k=1}^M s_k \cdot \psi_k(t)$$



$$s_k = \int_0^T s(t) \cdot \psi_k(t) dt$$

$s_k, k=1, 2, \dots, M$  are the coefficient and may be calculated

27

TELECOM LAB  
ITALIA

The transmission objective is to allow the receiver to recognise the signal  $s_i(t)$  transmitted by the source of information in the period  $T$ . the signal is chosen from time to time among the  $N$  waveforms corresponding to the symbols the source is generating with a given probability.

The digital modulation techniques associate exactly the digital sequences forming a symbol with a specific waveform. The binary modulation occurs when the modulator associates the 2 binary symbols to two different waveforms.

If  $k$  bits are modulated at once in the time  $T$ , the modulator applies an alphabet of  $N = 2^k$  waveforms: in this case an  $N$ -level modulation takes place.

The transmission is said to be coherent if the generated signals are exactly known by the receiving side (e.g. phase, amplitude). For a non-coherent modulation this knowledge is not ensured.

To start, the generic signal  $s(t)$  defined in the interval  $(0,T)$  is represented in some way (e.g. geometrical or vectorial representation).  $s(t)$  may be represented through a complex function. It is assumed that  $s(t)$  has a finite energy.

Let's consider a set of  $M$  signals

$$\psi_i(t); i = 1, \dots, M$$

The signals are orthogonal in that they obey the conditions reported in the figure. each signal with a finite energy in  $(0,T)$  can be represented through a combination of linear signals (figure). In this case the set of orthogonal signals is said to be *complete*.

## Orthonormal representation

a generic signal  $s(t)$ : 
$$s(t) = \sum_{i=1}^M s_i \psi_i(t) + \varepsilon(t) \rightarrow E_\varepsilon = \int_0^T \varepsilon^2(t) dt$$

value of coefficients minimising the error energy:

$$\min_{\{s_1, s_2, \dots, s_N\}} [E_\varepsilon] \rightarrow s_i = \int_0^T s(t) \cdot \psi_i(t) dt$$

with error energy = 0:

$$E_s = \int_0^T |s(t)|^2 dt = \int_0^T \sum_{i=1}^M |s_i|^2 \cdot |\psi_i(t)|^2 dt = \sum_{i=1}^M |s_i|^2$$

28

TELECOM LAB  
ITALIA

In any case, given a set of orthogonal functions  $\psi_i(t)$  is possible to represent any signal  $s(t)$  with a given error  $\varepsilon(t)$ .

If we calculate the coefficients minimising the error energy, the same coefficients  $s_i$  calculated for the case of no errors, are obtained. This implies that  $s(t)$  can be fully represented through linear combinations of  $\psi_i(t)$ .

Note that the signal energy  $s(t)$  is given by the sum of the squares of the modules of  $s_i$ .

Each signal in  $(0, T)$  can then be represented by a vector with  $M$  components  $\{s_1, s_2, s_3, \dots, s_M\}$ .

If  $s_i(t)$  with  $i=1, 2, \dots, N$  are the signals to be represented, in  $(0, T)$ , for each  $s_i(t)$  holds:

$$s_i(t) = \sum_{k=1}^M s_{ik} \cdot \psi_k(t)$$

Let  $\vec{s}_i$  be the vector representing  $s_i(t)$ .

The energy of the generic signal is: vale:

$$E_{s_i} = \sum_{k=1}^M |s_{ik}|^2$$

## Correlation and square distance

scalar product:  $(\bar{s}_i, \bar{s}_j) = \sum_{k=1}^M s_{ik} \cdot s_{jk}$

using the definition of  $s_{ik} \rightarrow (\bar{s}_i, \bar{s}_j) = \int_0^T s_i(t) \cdot s_j(t) dt$

correlation coefficient between  $\bar{s}_i$  and  $\bar{s}_j$   $\rho_{\bar{s}_i \bar{s}_j} = \frac{(\bar{s}_i, \bar{s}_j)}{\sqrt{E_{s_i}} \sqrt{E_{s_j}}} = \frac{\int_0^T s_i(t) s_j(t) dt}{\sqrt{E_{s_i}} \sqrt{E_{s_j}}}$

square distance between  $\bar{s}_i$  and  $\bar{s}_j$   $\Downarrow$

$$d_{\bar{s}_i \bar{s}_j}^2 = \int_0^T [s_i(t) - s_j(t)]^2 dt = E_{s_i} + E_{s_j} - 2\rho_{\bar{s}_i \bar{s}_j} \sqrt{E_{s_i} \cdot E_{s_j}}$$

29



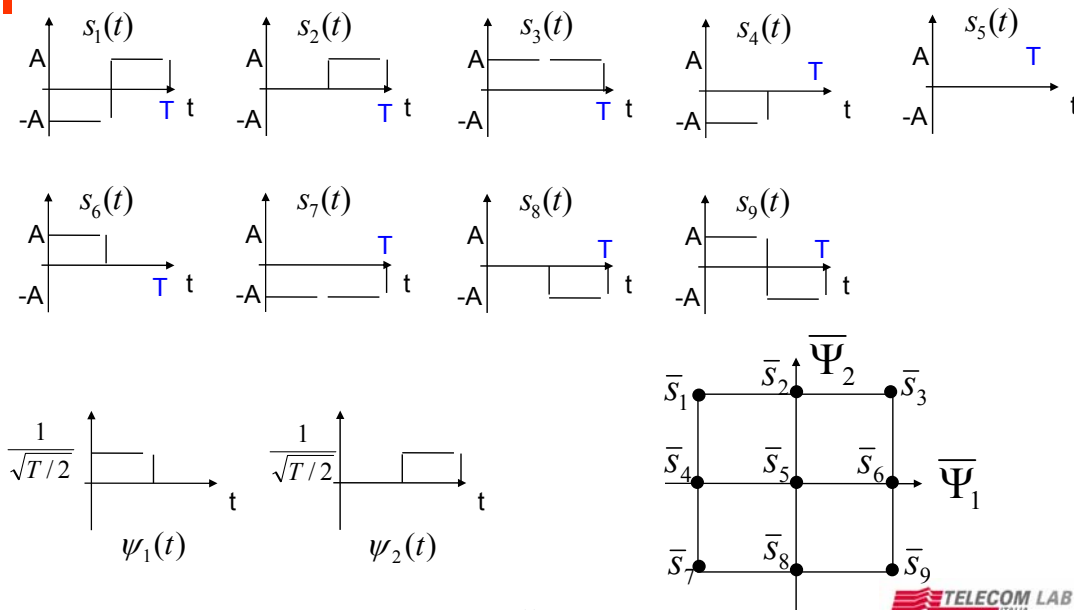
The scalar product of two arrays  $\bar{s}_i$  and  $\bar{s}_j$  is the sum of the product of their homologous components. It corresponds to the product of two of the signals described previously and described by  $s_i(t)$  and  $s_j(t)$ , being the product integrated between 0 and T (crossed correlation). The scalar product is proportional to the cosine of the angle between the two arrays.

The correlation coefficient between  $\bar{s}_i$  and  $\bar{s}_j$  is defined according to the above product, normalised over the square roots of the energies of the two signals.

The quadratic distance between  $\bar{s}_i$  and  $\bar{s}_j$  in the real case can be expressed in function of the only energies of the signals given by  $\bar{s}_i$  and  $\bar{s}_j$ .

The above expressions hold for real signals.

## An example



30

The figure represents a set of 9 signals defined in  $(0, T)$ .

In this case we have a modulation bearing 9 combinations per symbol ( $3^2$ ).

All the signals can be represented through just two orthogonal functions:

$$\psi_1(t); \psi_2(t) \rightarrow \int_0^T \psi_1(t) \cdot \psi_2(t) dt = 0$$

The array representation on the plane of the coordinates

$$\bar{\Psi}_1, \bar{\Psi}_2$$

Characterises each function through an array (a point) whose components represent respectively the coefficients of  $\psi_1(t)$  and  $\psi_2(t)$

This representation is named *constellation*.

Note that the orthogonal functions  $\psi_1(t)$  and  $\psi_2(t)$  show an unitary energy:

$$\int_0^T \psi_i^2(t) dt = 1$$

It is easy to calculate the function coefficient of the functions  $\psi_1(t)$  and  $\psi_2(t)$  which give rise to the various waveforms generated by the source. For instance, for  $s_2(t)$  we get:

$$s_{21} = 0; \quad s_{22} = A\sqrt{T/2}$$

# Gram-Schmidt algorithm

how to build-up a set of orthogonal functions

$$\psi_i(t); \quad i = 1, \dots, M$$

representing a set of independent functions

$$s_i(t); \quad i = 1, \dots, N$$

**Step 1**      $\psi'_1(t) = s_1(t); \quad \psi_1(t) = \frac{\psi'_1(t)}{\sqrt{E_{\psi'_1}}}$

·  
·  
·

**Step k**      $\psi'_k(t) = s_k(t) - \sum_{i=1}^{k-1} (\bar{s}_k, \bar{\Psi}_i) \cdot \psi_i(t); \quad \psi_k(t) = \frac{\psi'_k(t)}{\sqrt{E_{\psi'_k}}}$

$$\psi_1(t) \text{ and } \psi_2(t) \text{ are orthogonal} \rightarrow (\bar{\Psi}_1, \bar{\Psi}_2) = (\bar{s}_2, \bar{\Psi}_1) - (\bar{s}_2, \bar{\Psi}_1) \cdot (\bar{\Psi}_1, \bar{\Psi}_1) = 0$$

31



Starting from set of functions  $s_i(t); \quad i = 1, \dots, N$  a set of orthogonal functions can be constructed  $\psi_k(t)$  able to generate the original functions

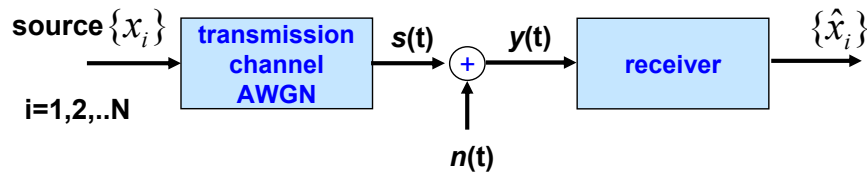
The Gram-Schmidt algorithm runs through N steps. The set of N functions can be constructed in general through M ( $M \leq N$ ) orthogonal functions. If this is possible, the algorithm applied over N steps will lead to define signals that are linearly dependent on those previously generated through the same algorithm. These signals are discarded.

In words, the elementary array  $\bar{\Psi}$  is calculated by imposing that the signal  $s_k(t)$  is the sum of its projections (scalar products) over  $\psi_k(t)$

$$\bar{\Psi}_1, \bar{\Psi}_2, \dots, \bar{\Psi}_{k-1}$$

The number M of functions  $\psi_k(t)$  needed to represent all the signals  $s_i(t); \quad i = 1, \dots, N$  is said *dimension* of the set of signals.

## Optimal decision



**additive noise function: the sum of**

- a noise portion which can be represented through the base of orthogonal functions  $\psi_k(t)$
- a noise portion  $n'(t)$  which cannot be embedded in that representation

$$n(t) = \sum_{k=1}^M n_k \cdot \psi_k(t) + n'(t)$$

**$n_k$ : a gaussian variable with average=0 and variance**

$$\sigma^2 = E\{n_k^2\} = n_0 / 2$$

32

TELECOM LAB  
ITALIA

Let's suppose that the following is verified in the communication channel of the figure:

- no distortion is introduced by the channel
- the noise is additive, i.e. independent from the input signal
- the noise can be represented by a gaussian statistic with zero-mean and bilateral power density equal to  $n_0/2$  where  $n_0$  is the spectral density of the noise and is given by  $k \cdot T$  [W/Hz] ( $k$ : Boltzman constant;  $T$ : absolute temperature).

The noise can be decomposed according to the scheme of figure. the component  $n'(t)$  cannot be represented with the base  $\psi_k(t)$  hence is orthogonal to all the signals generated by the source. For this reason, just the noise component which can be represented through the base  $\psi_k(t)$  is affecting the result on the receiving side, where the receiver correlates the received signal with all the source signals  $s_i(t)$ .

The components  $n_k$  of the noise are random variables with zero mean and variance:

$$\sigma^2 = \frac{n_0}{2}$$

(the proof is omitted). Moreover, they are independent.

## The binary case

received signal:  $y(t) = s_i(t) + n(t); \quad i = 1, 2$

if  $y(t) = s_0(t) + n(t) \rightarrow$  the components of  $y(t)$  are :

$$y_k = s_{0k} + n_k$$
$$E[y_k] = s_{0k}$$

if  $y(t) = s_1(t) + n(t) \rightarrow$  the components of  $y(t)$  are :

$$y_k = s_{1k} + n_k$$
$$E[y_k] = s_{1k}$$

33



The joint distribution of the  $y_k$ s conditioned to the event that the source has transmitted  $s_i(t)$  is the product of the marginal gaussian distributions (see the figure):

$$p(\bar{y} / s_i(t)) = \frac{1}{(\sqrt{2\pi})^M \cdot \sigma^M} \cdot \exp\left(\frac{-1}{2\sigma^2} \sum_{k=1}^M |y_k - s_{ik}|^2\right)$$

On the basis of these probabilities, the receiver must assume a decision.

In particular, it is needed to decide if, being  $y(t)$  the received signal, the source transmission has been either  $s_0(t)$  or  $s_1(t)$ . This *a posteriori* probability can be calculated in function of the above conditional probability.

## Optimal decision for the binary case

decision for  $\{x_0\}$ :  $p(s_0(t)/\bar{y}) > p(s_1(t)/\bar{y})$

decision for  $\{x_1\}$ :  $p(s_1(t)/\bar{y}) > p(s_0(t)/\bar{y})$

maximum likelihood criterion

receiver decision for  $\{x_0\}$ :



$$\int_0^T y(t) \cdot s_0(t) dt - \frac{E_0}{2} > \int_0^T y(t) \cdot s_1(t) dt - \frac{E_1}{2}$$

34

TELECOM LAB  
ITALIA

The *a posteriori* probabilities

$$p(s_0(t)/\bar{y}) \text{ e } p(s_1(t)/\bar{y})$$

Can be calculated (Bayes theorem) in function of the *a priori* probabilities:

$$p(\bar{y}/s_0(t)) \text{ e } p(\bar{y}/s_1(t))$$

For instance, for  $s_0(t)$  the following is obtained:

$$p(s_0(t)/\bar{y}) = \frac{p(\bar{y}/s_0(t)) \cdot P(x_0)}{p(\bar{y})}$$

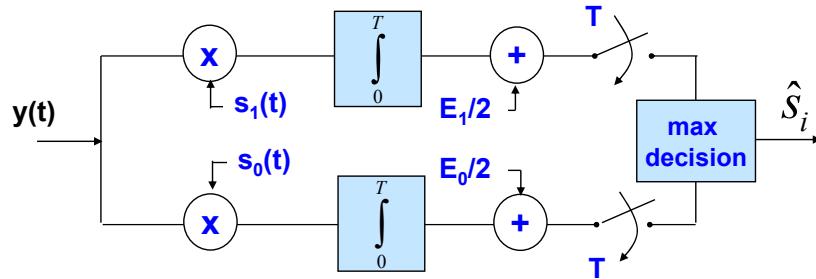
The above expression is used to calculate the decision criterion in function of the *a priori* probability. The receiver decides for  $\{x_0\}$  if:

$$\frac{p(s_0(t)/\bar{y})}{p(s_1(t)/\bar{y})} = \frac{p(\bar{y}/s_0(t)) \cdot P(x_0)}{p(\bar{y}/s_1(t)) \cdot P(x_1)} > 1$$

which becomes, when  $P(x_0) = P(x_1) = 1/2$ :

$$p(\bar{y}/s_0(t)) > p(\bar{y}/s_1(t))$$

## The binary receiver (correlation mechanism)



receiver decision for  $\{x_0\}$ :

$$\Rightarrow \int_0^T y(t) \cdot s_0(t) dt - \frac{E_0}{2} > \int_0^T y(t) \cdot s_1(t) dt - \frac{E_1}{2}$$

35

TELECOM LAB  
ITALIA

Starting from the case of uniform probabilities and reminding that in this case the decision is assumed on the basis of the criterion:

$$p(\bar{y} / s_0(t)) > p(\bar{y} / s_1(t))$$

with reference to the joint distributions considered for  $y$ , the criterion can be written as:

$$\sum_{k=1}^M |y_k - s_{0k}|^2 < \sum_{k=1}^M |y_k - s_{1k}|^2$$

Developing the squares and keeping in mind that:

$$\sum_{k=1}^M y_k s_{ik} = \int_0^T y(t) s_i(t) dt$$

The criterion becomes that indicated in the figure:

$$\int_0^T y(t) \cdot s_0(t) dt - \frac{E_0}{2} > \int_0^T y(t) \cdot s_1(t) dt - \frac{E_1}{2}$$

The scheme refers to the receiver in case of a binary modulation of the kind described and realises exactly the decision criterion.

## Error probability

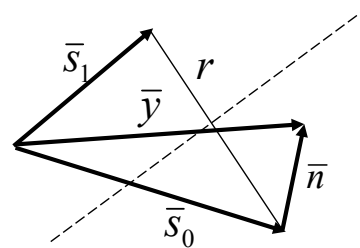
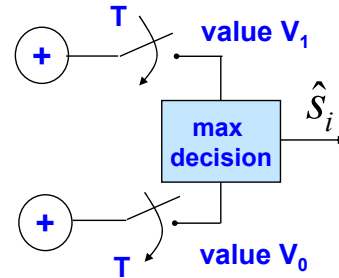
In case the source has transmitted  $s_0(t)$

error probability  $P_e = pr(V_1 - V_0) > 0$

$$P_e = Q \left[ \sqrt{\frac{E_1 + E_0 - 2\rho\sqrt{E_1 E_0}}{2n_0}} \right]$$

two signals with the same energy

$$P_e = Q \left[ \sqrt{\frac{E_b(1-\rho)}{n_0}} \right] \quad E_b = \frac{(E_0 + E_1)}{2 \log_2 N}$$



TELECOM LAB  
ITALIA

36

The calculus of the error probability utilises the results already obtained in the previous steps.

The two decision variables in case  $s_0(t)$  is transmitted, are represented in an array form through two arrays. The distance is:  $\bar{s}_0$  ed  $\bar{s}_1$

$$d^2_{\bar{s}_0, \bar{s}_1} = \int_0^T [s_0(t) - s_1(t)]^2 dt = E_{s_0} + E_{s_1} - 2\rho\sqrt{E_{s_0} \cdot E_{s_1}}$$

Since  $s_0(t)$  is the transmitted signal, there is a  $n$  error on the receiving side when the resulting array falls in the domain formed by all the points having a distance from  $s_1$  lower with respect to that from  $s_0$ :

$$d_{\bar{y}, \bar{s}_1} < d_{\bar{y}, \bar{s}_0}$$

This happens whenever the array projected over the line  $r$  depicted in figure overcomes the distance:

$$d^2_{\bar{s}_0, \bar{s}_1} / 2$$

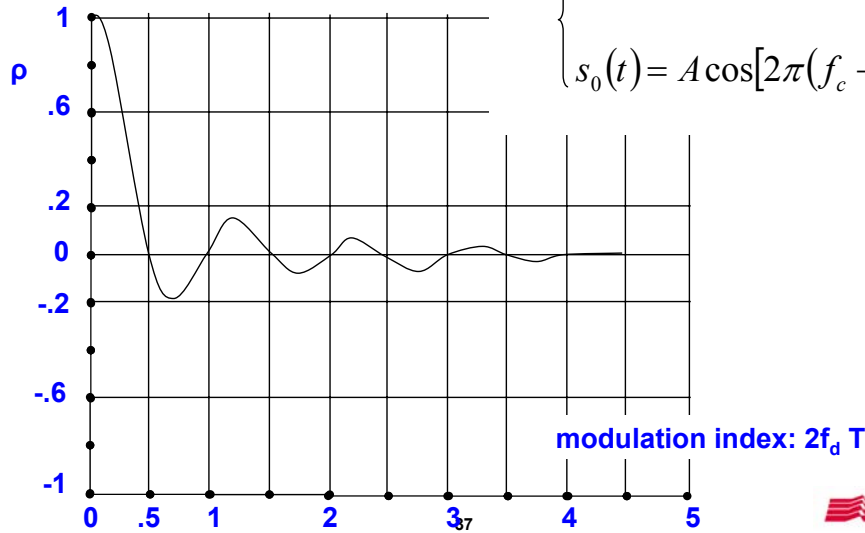
Being the noise a gaussian random variable with zero mean and variance  $\sigma^2 = n_0/2$ , the formula in figure holds for the error probability. In this formula:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{z^2}{2}} dz$$

## An example of binary signal: FSK

Two frequencies located on the sides of a central frequency  $f_c$  are used:

$$\begin{cases} s_1(t) = A \cos[2\pi(f_c + f_d)t] \\ s_0(t) = A \cos[2\pi(f_c - f_d)t] \end{cases}$$



This kind of modulation utilises signals of the same energy with different frequencies. Being a binary modulation, for the error probability, we have:

$$P_e = Q\left[\sqrt{\frac{E_b(1-\rho)}{n_0}}\right]$$

The correlation coefficient is:

$$\rho = \frac{1}{E} \int_0^T A \cos[2\pi(f_c + f_d)t] \cdot A \cos[2\pi(f_c - f_d)t] dt$$

and it can be approximated by:

$$\rho = \frac{\sin 4\pi f_d T}{4\pi f_d T}$$

The behaviour is reported in figure in function of the ratio between the maximum excursion of the frequency and the frequency of the symbol  $1/T$  (modulation index). The modulation index results  $2f_d T$ .

The error probability assumes its minimum value with a modulation index equal to 0.7. The intersections of the curve with the abscissa represent the points where two source functions are orthogonal ( $\rho=0$ ).