

Appunti di Comunicazioni Elettriche

Capitolo 7

Trasmissione numerica - Parte I

Introduzione ai segnali di tipo numerico.....	1
Schema generale di un sistema di trasmissione binaria	2
Definizione del filtro adattato.....	4
Scelta della soglia.....	7
Codifica antipodale e codifica ortogonale.....	10
Problema dell'interferenza di intersimbolo.....	15
<i>Forme d'onda di Nyquist</i>	19
<i>Roll off delle forme d'onda ad intersimbolo nullo</i>	22
Forma d'onda da trasmettere	22
<i>Cenni sulla formazione delle forme d'onda in trasmissione</i>	24
Estrazione del timing.....	27
<i>Codifica NRZ e codifica RZ</i>	28
<i>Codifica RZ di tipo antipodale</i>	32
<i>Schema del dispositivo estrattore del timing</i>	33
<i>Diagramma ad occhio</i>	34
<i>Generazione degli impulsi di campionamento (schema alternativo)</i>	35
<i>Sollecitazioni sul filtro passa-banda e corrispondenti codifiche di linea</i> 37	
Codice AMI.....	37
Codice HDB-3.....	38
Codice Manchester	39
Criterio generale di dimensionamento del sistema	40
<i>Rapporto S/N a valle del campionatore</i>	41
<i>Rapporto S/N a valle del filtro in ricezione</i>	46
<i>Potenza ricevuta dal mezzo trasmissivo</i>	48
<i>Osservazione: caso della codifica ortogonale</i>	49
Esempio numerico: trasmissione numerica del segnale TV a colori	51
Sistema numerico multitratta su cavo	56
<i>Tratte rigenerative</i>	57

INTRODUZIONE AI SEGNALE DI TIPO NUMERICO

I **segnali di tipo numerico** sono costituiti da sequenze di *segnali elementari*, temporalmente distinti tra loro, ciascuno di durata T . In linea teorica, si possono immaginare quante si vogliono forme d'onda cui far corrispondere i messaggi elementari che si vogliono rappresentare (**discretizzazione sulla forme**), oppure, usando una sola forma d'onda, *si possono scegliere per essa ampiezze diverse a seconda dell'informazione da trasmettere* (**discretizzazione sulle ampiezze**). Questo è proprio il caso della **trasmissione numerica**: volendo ad esempio trasmettere le 10 cifre decimali, i segnali elementari avranno la stessa forma d'onda, ad esempio rettangolare, ma 10 ampiezze diverse, ognuna

corrispondente ad una cifra. Nel caso della trasmissione binaria (che è quella di nostro interesse in quanto è impiegata nei sistemi elettronici), le ampiezze possibili sono solo 2.

In generale, comunque, la trasmissione di un simbolo piuttosto che un altro corrisponde ad una scelta tra un numero finito M di possibilità, dove M sono anche i simboli dell'alfabeto sorgente che si vuole trasmettere ($M=2$ nel sistema binario oppure $M=10$ nel sistema decimale). In ricezione, ricevendo la sequenza di segnali elementari, eventualmente attenuata dal mezzo trasmissivo e deformata dai disturbi, si tratterà di "interpretarla", stabilendo quale è stato il simbolo trasmesso in ciascun intervallo T . Si tratta, perciò, di fare ancora una volta una scelta tra le M possibilità note a priori. Nel caso binario, la cosa è più semplice, in quanto le possibilità sono solo due: in questo, c'è una notevole somiglianza (che sarà ancora più chiara nei prossimi discorsi) con quanto accade nel sistema radar, dove, in ricezione, la scelta è tra la presenza o l'assenza del bersaglio.

Una operazione di decisione è tipica nella ricezione dei segnali numerici: da essa consegue che il segnale fornito in uscita dal ricevitore può essere corretto oppure parzialmente errato (nel caso qualche segnale elementare sia stato interpretato erroneamente). E' intuitivo comprendere, come nel caso del radar, che si commetterà un errore solo quando il disturbo supera un certo valore: per esempio, nel caso di trasmissione binaria in cui all'uno logico è associata una tensione $+V$ e allo zero logico una tensione $-V$, l'errore si potrà commettere solo quando il disturbo è tale da trasformare una ampiezza positiva in negativa o viceversa.

Fatte queste premesse, sottolineiamo alcune differenze tra un sistema di trasmissione analogico ed un sistema di trasmissione numerico:

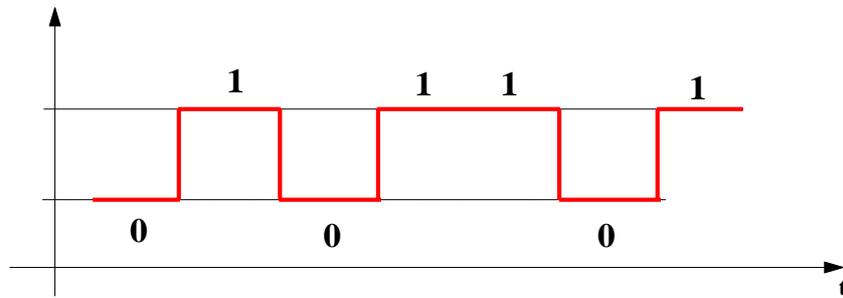
- una prima differenza è che, mentre nella trasmissione analogica è necessario utilizzare apparecchiature ottimizzate per il tipo di segnale da trasmettere (segnale telefonico, segnale musicale, segnale televisivo¹ e così via), nella trasmissione digitale ciò non è necessario: i segnali numerici hanno tutti la stessa forma d'onda (a parte, ovviamente, la sequenza dei simboli), per cui è possibile trasmettere un qualunque segnale numerico senza preoccuparsi di cosa i simboli numerici rappresentano;
- una seconda differenza riguarda gli errori: così come nella trasmissione analogica non è possibile far arrivare al ricevitore la stessa forma d'onda trasmessa, allo stesso modo nella trasmissione digitale non si può mai raggiungere la certezza di non commettere errori; a fronte di questo, però, ci sono due considerazioni: in primo luogo, si può fare in modo, nella trasmissione digitale, che la probabilità di errore sia talmente bassa che l'effetto degli errori diventi praticamente invisibile²; in secondo luogo, mentre nel caso analogico, per aumentare il rapporto S/N di 10 dB, bisogna aumentare di 10 dB la potenza del segnale (o ridurre di 10 dB quella del rumore), nel caso digitale questo non è necessario.

SCHEMA GENERALE DI UN SISTEMA DI TRASMISSIONE BINARIA

A questo punto, ci concentriamo sulla **trasmissione binaria**. Il modo più semplice di trasmettere un segnale binario, cioè una sequenza di 1 e 0 logici, è quello di associare a ciascun simbolo un impulso rettangolare di tensione, che avrà ampiezza V_0 nel caso dello 0 logico e V_1 nel caso dell' 1 logico:

¹ Pensiamo, per esempio, alla diversa banda necessaria per la trasmissione dei segnali elencati: abbiamo bisogno di almeno 3.4 kHz per il segnale telefonico, di almeno 15 kHz per il segnale musicale e di almeno 5 MHz per il segnale televisivo in formato PAL.

² I sistemi reali di trasmissione numerica non sbagliano praticamente mai se sono "disturbati" soltanto dal rumore termico ed elettronico generato dalle apparecchiature elettroniche utilizzate. Sono invece molto più sensibili ad altri disturbi, tipicamente quelli provenienti da apparecchiature meccaniche in movimento (si pensi alle vecchie centrali telefoniche).



Nel caso più banale, si può pensare di trasmettere una tensione nulla per lo 0 ed una tensione di un valore $+V$ (ad esempio $+5V$) per l'1.

In ricezione, non interessa ricostruire la forma d'onda trasmessa, ma serve riottenere, con la maggiore fedeltà possibile, la sequenza di simboli trasmessa. Quindi, così come nel radar, anche nella trasmissione numerica bisogna effettuare, al terminale ricevente, una decisione binaria. Le possibilità sono le seguenti:

- se è stato trasmesso il simbolo logico 0 e si è scelto di associare a tale simbolo un livello nullo di tensione, in ricezione giunge solo rumore:

$$s_R(t) = n(t)$$

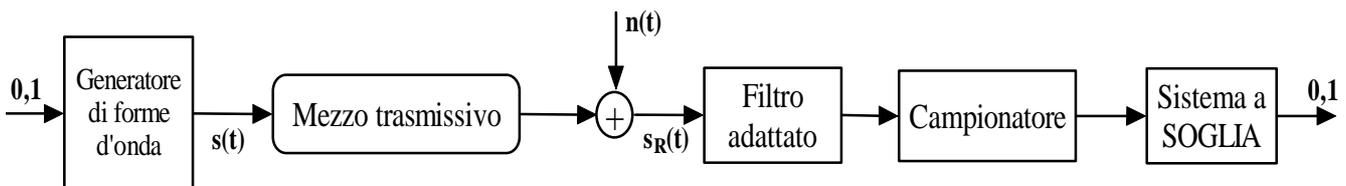
In questo caso, il sistema deve essere in grado di capire che il segnale ricevuto è solo rumore;

- se invece è stato trasmesso il simbolo logico 1 e si è scelto di associare ad esso una certa forma d'onda elementare $s(t)$ di durata T (ad esempio un impulso rettangolare di durata T e di una ampiezza prefissata), questa giunge al ricevitore, attenuata di una certa costante k e ritardata (di una quantità nota); la differenza col radar, però, è nel fatto che le forme d'onda in ricezione, se ci sono, si possono presentare solo ad intervalli di tempo regolari (mentre nel radar gli echi tornano indietro con ritardi legati alla posizione del bersaglio e quindi non determinabili a priori); possiamo perciò scrivere che, in questo caso, in ricezione si ha

$$s_R(t) = ks(t - nT) + n(t)$$

A parte la differenza circa il ritardo con cui giunge l'eventuale forma d'onda, si può procedere in maniera del tutto analoga a quelle seguita nel caso radar: al fine di minimizzare la probabilità di errore, bisogna fare in modo che, almeno in un certo istante (che indicheremo ancora t_m e che varia con le forme d'onda scelte per la trasmissione), il segnale sia quanto più grande possibile rispetto ai valori che può assumere il rumore.

Possiamo a questo punto tracciare uno schema generale di un sistema di trasmissione binaria:



La sequenza binaria da trasmettere va in ingresso ad un dispositivo che ha il compito di associare ad essa le corrispondenti forme d'onda da trasmettere. La sequenza di tali forme d'onda costituisce il segnale $s(t)$ che viene inviato sul mezzo trasmissivo. All'uscita di quest'ultimo, giunge un segnale,

inevitabilmente attenuato, con sovrapposto un rumore $n(t)$ (che supponiamo essere, come al solito, bianco e con distribuzione gaussiana delle ampiezze). La composizione di tali segnali rappresenta il segnale $s_R(t)$ che va in qualche modo elaborato in ricezione. La prima elaborazione è compiuta da un filtro, progettato con la stessa filosofia che, nel caso del segnale radar, ha portato alla definizione del **filtro adattato** come filtro ottimo³. Nel paragrafo successivo vedremo la cosa in modo analitico.

All'uscita del filtro adattato otteniamo dunque un segnale che, in istanti di tempo distanti T secondi uno dall'altro, presenta un massimo. Questi massimi vengono misurati (mediante un dispositivo che quindi è semplicemente un **campionatore** con periodo di campionamento T) e, in base a tali misure, il successivo sistema a soglia decide se il simbolo trasmesso è 1 oppure 0: se la tensione elettrica misurata in ingresso è maggiore di un livello prefissato, il sistema a soglia fornisce in uscita il valore 1, mentre in caso contrario fornisce 0.

E' importante fare una osservazione. Lo schema tracciato nell'ultima figura, tipico di un sistema di trasmissione binaria, può essere in effetti distinto in due parti, una binaria ed una analogica: infatti, noi abbiamo simboli binari solo prima del generatore di forme d'onda e dopo il sistema a soglia, mentre, nel resto dello schema, avendo a che fare con forme d'onda continue nel tempo, siamo in presenza di un normale sistema analogico. In altre parole, quindi, il generatore di forme d'onda ed il sistema a soglia effettuano, rispettivamente, il passaggio da digitale ad analogico ($D \rightarrow A$) e da analogico a digitale ($A \rightarrow D$). In altre parole ancora, essi fanno da interfaccia tra il sistema binario ed il sistema analogico usato per la trasmissione.

DEFINIZIONE DEL FILTRO ADATTATO

Faremo adesso dei discorsi estremamente simili a quelli fatti per il sistema radar: lo scopo è di far vedere che, anche nel caso della trasmissione binaria, il filtro adattato in ricezione è particolarmente utile.

Così come nel radar, il filtro deve essere tale da fornire, per il segnale, il massimo valore possibile in corrispondenza di un certo istante t_m ; di conseguenza, anche in questo caso, *una definizione sensata del rapporto S/N in ingresso al sistema a soglia prevede il confronto tra la potenza di picco del segnale utile e la potenza media di rumore:*

$$\frac{S}{N}_{\text{soglia}} = \frac{P_{S,\text{picco}}}{P_{N,\text{media}}}$$

Dobbiamo massimizzare questo rapporto.

Cominciamo dal rumore: se $H(f)$ è la funzione di trasferimento del filtro di ricezione, possiamo valutare la potenza di rumore come il valor medio dei quadrati della tensione di rumore, per cui

$$P_{N,\text{media}} = \int_{-\infty}^{+\infty} h_n(f) |H(f)|^2 df$$

dove $h_n(f)$ è la densità spettrale di potenza bilatera del rumore bianco sovrapposto al segnale.

³ Ricordiamo che il filtro adattato è ottimo nel senso che fornisce il massimo rapporto S/N all'uscita a parità di energia per simbolo (cioè a parità di potenza media del segnale trasmesso). Una condizione affinché il filtro sia ottimo è che il rumore sia additivo bianco.

Questa formula fornisce anche la varianza della variabile casuale n_{uC} ottenuta campionando il rumore all'istante di lettura. La funzione densità di probabilità di questa variabile casuale ha la seguente espressione:

$$f(n_{uC}) = \frac{1}{\sqrt{2\pi P_{N,media}}} e^{-\frac{n_{uC}^2}{2P_{N,media}}}$$

Passiamo al calcolo di $P_{S,picco}$: indichiamo con $s_R(t)$ la forma d'onda ricevuta, dove l'ampiezza a è per il momento incognita. Se $S_R(f)$ è lo spettro (bilatero) del segnale ricevuto, lo spettro (sempre bilatero) del segnale in uscita dal filtro è

$$S_u(f) = H(f)S_R(f)$$

L'andamento nel tempo di questo segnale si ottiene antitrasformando:

$$s_u(t) = \int_{-\infty}^{+\infty} S_u(f) e^{j\omega t} df = \int_{-\infty}^{+\infty} H(f) S_R(f) e^{-j\omega \tau} df$$

A noi interessa la potenza di picco di questo segnale: supponendo allora che tale picco venga raggiunto nell'istante t_m e considerando l'ampiezza pari ad a , possiamo scrivere che

$$P_{S,picco} = a^2 |s_u(t_m)|^2 = a^2 \left| \int_{-\infty}^{+\infty} H(f) S_R(f) e^{j\omega t_m} df \right|^2$$

Possiamo dunque formalizzare l'espressione del rapporto S/N:

$$\left. \frac{S}{N} \right|_{soglia} = \frac{a^2 \left| \int_{-\infty}^{+\infty} H(f) S_R(f) e^{j\omega t_m} df \right|^2}{\int_{-\infty}^{+\infty} h_n(f) |H(f)|^2 df}$$

A questo punto, osserviamo che l'energia del segnale ricevuto è

$$E_R = a^2 \int_{-\infty}^{+\infty} |S_R(f)|^2 df$$

Esplicitando da qui il valore di a^2 e sostituendo nell'espressione del rapporto S/N, otteniamo quanto segue:

$$\left. \frac{S}{N} \right|_{soglia} = \frac{\left| \int_{-\infty}^{+\infty} H(f) S_R(f) e^{j\omega t_m} df \right|^2}{\int_{-\infty}^{+\infty} h_n(f) |H(f)|^2 df} \frac{E_R}{\int_{-\infty}^{+\infty} |S_R(f)|^2 df}$$

Come ci aspettavamo il rapporto S/N risulta direttamente proporzionale, a parità di $h_n(f)$, di $S_R(f)$ e di $H(f)$, all'energia del segnale ricevuto.

Come ulteriore passaggio, consideriamo adesso che $S_u(f) = H(f)S_R(f)$ e quindi anche che

$$|S_u(f)|^2 = |H(f)|^2 |S_R(f)|^2 \longrightarrow |S_R(f)|^2 = \frac{|S_u(f)|^2}{|H(f)|^2}$$

Sostituendo nell'espressione del rapporto S/N, otteniamo

$$\frac{S}{N}\Big|_{\text{soglia}} = \frac{\left| \int_{-\infty}^{+\infty} S_u(f) e^{j\omega t_m} df \right|^2}{\int_{-\infty}^{+\infty} h_n(f) |H(f)|^2 df} \frac{E_R}{\int_{-\infty}^{+\infty} |S_u(f)|^2 \frac{1}{|H(f)|^2} df}$$

Portando a primo membro l'energia E_R del segnale ricevuto, otteniamo una quantità che è caratteristica del sistema considerato. Tale quantità, per una fissata $h_n(f)$, dipende esclusivamente da $H(f)$ e da $S_R(f)$, o, in modo equivalente, da $H(f)$ e da $S_U(f)$. Di conseguenza, la formula appena ricavata consente di determinare la funzione di trasferimento $H(f)$ ottima ai fini delle prestazioni del sistema.

I casi che si possono presentare (in fase di dimensionamento) sono due:

- nel primo caso, si fissa la forma d'onda $s_R(t)$ in ingresso al ricevitore: questo è il caso che esamineremo in questo paragrafo e che si verifica quando risulta conveniente generare in trasmissione un certo tipo di forma d'onda;
- nel secondo caso, invece, si fissa la forma d'onda $s_u(t)$ che si vuole ottenere in uscita dal ricevitore: questo caso sarà considerato successivamente, quando il requisito fondamentale sarà la necessità di ottenere, in uscita dal filtro di ricezione, forme d'onda cosiddette *ad intersimbolo nullo*.

Consideriamo il caso in cui sia stata fissata la forma d'onda $s_R(t)$ (ad esempio, una forma d'onda approssimativamente rettangolare). Ovviamente, si intende fissata la forma dei segnali in ricezione e non l'ampiezza e quindi l'energia, che come visto è proporzionale ad a^2 e dipende inoltre dall'attenuazione.

Il nostro obiettivo è dunque ottimizzare il sistema per una prefissata $s_R(t)$ (intesa come tipo di forma d'onda). Così come nel radar, l'ottimizzazione consiste nel rendere massimo il rapporto tra il rapporto S/N e l'energia E_R :

$$\frac{S}{N}\Big|_{\text{soglia}} \frac{1}{E_R} = \frac{\left| \int_{-\infty}^{+\infty} H(f) S_R(f) e^{j\omega t_m} df \right|^2}{\frac{1}{2} \int_{-\infty}^{+\infty} h_n(f) |H(f)|^2 df} \frac{1}{\int_{-\infty}^{+\infty} |S_R(f)|^2 df}$$

Tutto sta a scegliere una funzione di trasferimento opportuna: una volta trovato il massimo valore del suddetto rapporto, è possibile determinare, per ogni valore desiderato di $\frac{S}{N}\Big|_{\text{soglia}}$, la corrispondente energia E_R , la corrispondente ampiezza a , in definitiva, l'energia E_T da trasmettere.

Per semplicità, supponiamo che il rumore sia bianco, per cui

$$\frac{\left| \frac{S}{N} \right|_{\text{soglia}}}{E_R} = \frac{\left| \int_{-\infty}^{+\infty} H(f) S_R(f) e^{j\omega t_m} df \right|^2}{\frac{h_n}{2} \int_{-\infty}^{+\infty} |H(f)|^2 df \int_{-\infty}^{+\infty} |S_R(f)|^2 df} \cdot 1$$

Applichiamo nuovamente la **disuguaglianza di Schwarz**: ripetendo gli stessi identici discorsi fatti nel caso del radar, si trova che

$$\frac{\left| \frac{S}{N} \right|_{\text{soglia}}}{E_R} \leq \frac{1}{h_n/2}$$

dove la h_n è la densità spettrale monolatera di rumore ed il segno di uguaglianza (cioè la condizione ottima) si ottiene quando la funzione di trasferimento del filtro assume l'espressione tipica del filtro adattato in presenza di rumore bianco:

$$\boxed{H(f) = k S_R^*(f)}$$

Ancora una volta, quindi, *se la funzione di trasferimento del filtro è proporzionale al complesso coniugato dello spettro del segnale ricevuto, le prestazioni sono quelle ottimali e non dipendono dalla forma del segnale ricevuto, ma solo dalla sua energia.* Un filtro così fatto è un **filtro adattato**.

Questo è dunque il procedimento che, fissata la forma d'onda $s_R(t)$, porta alla ottimizzazione del sistema tramite una ottimizzazione del filtro.

SCelta DELLA SOGLIA

Così come abbiamo fatto per il sistema radar, un secondo passo importante, dopo la definizione del filtro in ricezione, è quello di capire dove va sistemata la **soglia**, analizzando le caratteristiche del segnale in ricezione. Queste caratteristiche non possono che essere statistiche, in quanto ci sono due motivi di casualità: in primo luogo, c'è la casualità legata al fatto che il ricevitore non può sapere, per ogni intervallo di tempo, se sta ricevendo 1 oppure 0; in secondo luogo, c'è la casualità legata al fatto che il rumore non è descrivibile in termini deterministici.

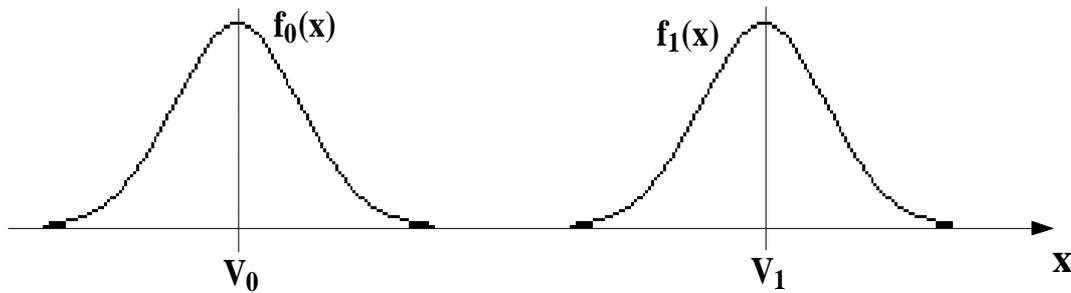
Dobbiamo dunque capire come sia fatta la funzione densità di probabilità del segnale che giunge al sistema a soglia. Nel caso del sistema radar, avevamo dedotto che le possibilità erano due: in presenza di solo rumore, si aveva una *distribuzione di Rayleigh*, mentre, in presenza di eco, si aveva una *distribuzione di Rice*. La differenza era dovuta alla presenza del demodulatore ad involuppo: in un caso (assenza di eco), l'ingresso al demodulatore era costituito da rumore passa-banda, mentre nell'altro (presenza di eco), era costituito da rumore passa-banda sommato al segnale.

Nel caso della trasmissione binaria, invece, stiamo per il momento supponendo che non ci sia stata modulazione in trasmissione, per cui non c'è nemmeno una demodulazione in ricezione, il che, ovviamente, semplifica le cose.

In ingresso al filtro abbiamo un rumore termico gaussiano bianco: dopo il filtraggio, questo rumore non è più bianco, ma rimane comunque gaussiano in base ad una nota proprietà dei sistemi

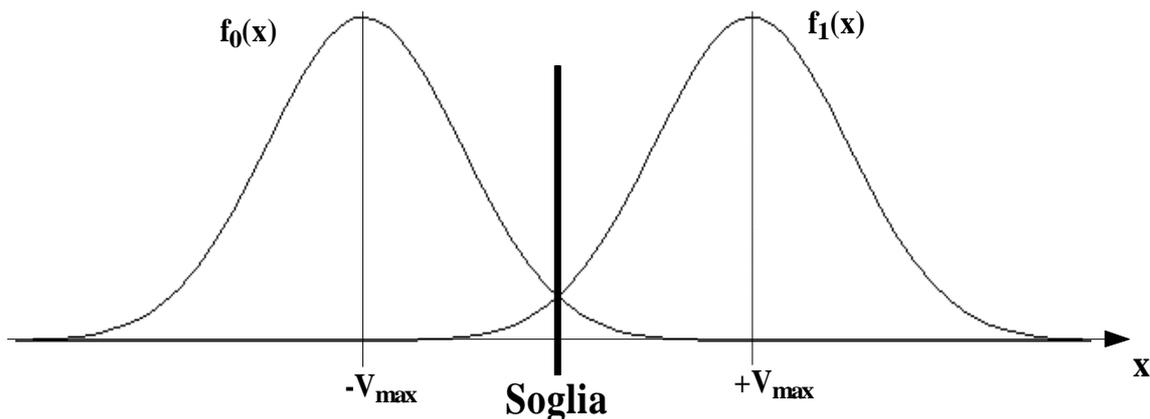
lineari. Quindi, in uscita dal filtro, avremo un rumore gaussiano la cui funzione densità di probabilità è centrata sul valore assunto dal segnale⁴. Le possibilità sono dunque due:

- se è stato trasmesso un 1 e se V_1 è il valore del segnale corrispondente ad 1, la campana gaussiana sarà centrata su tale valore;
- se invece è stato trasmesso uno 0 e V_0 è il valore corrispondente del segnale, la campana gaussiana sarà centrata su tale valore⁵.



La funzione $f_0(x)$ è la densità di probabilità del segnale, in uscita dal campionario, condizionata all'evento 0 trasmesso; la funzione $f_1(x)$ è invece la densità di probabilità dello stesso segnale, condizionata però all'evento 1 trasmesso. Entrambe le funzioni corrispondono a delle gaussiane tanto più larghe e schiacciate⁶ quanto maggiore è la varianza del rumore sovrapposto al segnale utile. Si tratta inoltre di gaussiane identiche (in quanto il rumore che si somma al segnale è comunque una variabile casuale a valor medio nullo), salvo ovviamente il valore medio, che differisce a seconda del simbolo che è stato trasmesso.

Evidentemente, nella figura abbiamo considerato il caso ideale in cui le due campane non presentano praticamente alcuna sovrapposizione. Questo dipende, ovviamente, oltre che dalla varianza del rumore, dalla distanza tra i livelli V_0 e V_1 di tensione. *Per chiari motivi tecnologici, questi due valori non possono essere eccessivamente distanziati, per cui c'è sempre una parziale sovrapposizione.* La situazione è dunque del tipo indicato nella figura seguente:



⁴ Si intende ovviamente il valore massimo del segnale in uscita dal filtro (ossia il valore del quale il successivo sistema a soglia farà uso per la decisione): in assenza di rumore, l'uscita del filtro è un segnale che, nell'istante di campionamento, vale V_0 se è stato trasmesso 0 e V_1 se è stato trasmesso 1.

⁵ In realtà, nel caso più semplice, il segnale è nullo quando è trasmesso lo 0. Tuttavia, per semplici motivi di analogia formale con il caso in cui è stato trasmesso 1, conviene considerare un generico valore V_0 .

⁶ Ricordiamo che l'area sottesa dalla gaussiana deve valere sempre 1, per cui, quanto più la campana è larga rispetto al valor medio, tanto più è schiacciata e viceversa.

Una volta misurata, nell'istante prefissato di lettura, la tensione in uscita dal campionatore, si tratta di decidere se è stato trasmesso 1 oppure 0. Si pone allora una soglia S : se la tensione è superiore alla soglia, si decide per l'1 trasmesso, mentre in caso contrario si decide per lo 0 trasmesso. E' chiaro che, in presenza di un rumore a statistica gaussiana, non si può mai avere una probabilità nulla di sbagliare, perché sappiamo che la gaussiana tende solo asintoticamente a zero.

Indicata, dunque, con S la soglia, valutiamo la probabilità di errore nei due casi:

- supponiamo che sia stato trasmesso 0, per cui la curva cui fare riferimento è $f_0(x)$: in questo caso, il sistema sbaglia se il valore x del segnale nell'istante di lettura è superiore alla soglia S , per cui deduciamo che la probabilità di errore vale

$$p_{\epsilon 0} = P(\epsilon | 0) = P(x > S) = \int_S^{+\infty} f_0(x) dx$$

In base a questa formula, $p_{\epsilon 0}$ corrisponde all'area sottesa da $f_0(x)$ tra la soglia e $+\infty$;

- supponiamo adesso che sia stato trasmesso 1, per cui dobbiamo fare riferimento a $f_1(x)$: in questo caso, il sistema sbaglia se il valore x del segnale nell'istante di lettura è inferiore alla soglia S , per cui la probabilità di errore vale

$$p_{\epsilon 1} = P(\epsilon | 1) = P(x < S) = \int_{-\infty}^S f_1(x) dx$$

In base a questa formula, $p_{\epsilon 1}$ corrisponde all'area sottesa da $f_1(x)$ tra $-\infty$ e la soglia.

Si tratta, adesso, di scegliere il valore della soglia S . A tal fine, dobbiamo valutare la probabilità di errore complessiva del sistema: tale probabilità, applicando il noto *teorema delle probabilità totali*, vale

$$p(\epsilon) = p(0T)P(\epsilon | 0) + p(1T)P(\epsilon | 1)$$

dove $p(1T)$ e $p(0T)$ sono, rispettivamente, la probabilità di aver trasmesso 1 e la probabilità di avere trasmesso 0. Tali probabilità, in generale, sono uguali (pari quindi a 0.5), in quanto una generica sequenza binaria avrà uno stesso numero di 0 e di 1: ponendo allora $p_T = p(1T) = p(0T)$, abbiamo che

$$p(\epsilon) = p_T [P(\epsilon | 0) + P(\epsilon | 1)] = 0.5 \cdot [P(\epsilon | 0) + P(\epsilon | 1)]$$

A questo punto, proprio perché i simboli 0 ed 1 sono equiprobabili, è chiaro che conviene porre la soglia esattamente nel punto in cui si intersecano le curve corrispondenti a $f_0(x)$ e $f_1(x)$, il che avviene al centro tra i valori V_0 e V_1 :

$$S = \frac{V_0 + V_1}{2}$$

Così facendo, anche le singole probabilità di errore $P(\epsilon | 0)$ e $P(\epsilon | 1)$ sono uguali. Ponendole entrambe uguali a $P(\epsilon | T)$, concludiamo dunque che

$$\boxed{p(\epsilon) = 0.5 \cdot 2 \cdot P(\epsilon | T) = P(\epsilon | T)}$$

Abbiamo cioè concluso che la probabilità di errore del sistema è pari alla probabilità di sbagliare il singolo bit.

Le cose cambiano, ovviamente, se i simboli 0 e 1 del messaggio trasmesso non sono equiprobabili: per esempio, se è più probabile il simbolo 1 (quindi $p(1T) > p(0T)$), allora ci converrà porre la soglia più lontana dal valore V_1 , in modo che $P(\epsilon|1) < P(\epsilon|0)$.

A questo punto, è interessante osservare l'analogia tra la probabilità di errore in un sistema di trasmissione numerica ed il rapporto segnale/rumore in un sistema di trasmissione analogica. Possiamo in particolare considerare, come sistema analogico, proprio il sistema radar, che, come visto, presenta forti analogie con il sistema numerico almeno per quanto riguarda la fase conclusiva di decisione: in quel caso, per ottenere un buon funzionamento del sistema in termini di risoluzione radiometrica, era importante il rapporto S/N in ingresso al sistema a soglia, mentre in questo caso, per ottenere un buon funzionamento del sistema in termini di fedeltà della sequenza in uscita, è importante minimizzare la probabilità di errore.

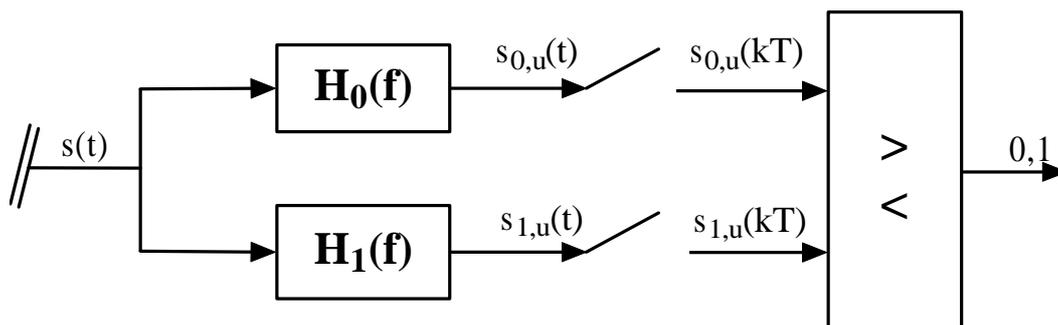
Potremmo d'altra parte ragionare anche in questo caso in termini di rapporto S/N in ingresso al sistema a soglia, proprio perché la natura binaria del sistema viene fuori tipicamente all'uscita del sistema a soglia. Per quanto riguarda il segnale, ci interessa il picco che il segnale raggiunge, ad opera del filtro adattato, nell'istante di lettura e in corrispondenza della trasmissione di un 1 (nella trasmissione dello 0, il segnale è nullo per le ipotesi fatte fino ad ora): possiamo perciò ragionare in termini deterministici e dire che ci interessa il valore V_1 . Per quanto riguarda, invece, il rumore, dobbiamo necessariamente ragionare in termini statistici: il parametro del rumore che meglio quantifica l'influenza del rumore stesso sul funzionamento del sistema è ovviamente la deviazione standard σ_n , che rappresenta lo scostamento del segnale, dovuto appunto al rumore, rispetto al valore atteso V_1 .

In definitiva, quindi, il rapporto cui potremmo fare riferimento, per il dimensionamento del sistema, è il rapporto V_1 / σ_n , che andrebbe perciò reso massimo. In termini di rapporto S/N vero e proprio, dovremmo invece far riferimento al rapporto $V_1^2 / V_{N,eff}^2$ tra la potenza di picco del segnale e la potenza media del rumore (rappresentata dal quadrato del valore efficace del rumore stesso).

CODIFICA ANTIPODALE E CODIFICA ORTOGONALE

A partire da questo paragrafo, vogliamo capire quali criteri vanno seguiti per la scelta delle forme d'onda da associare ai simboli logici 1 e 0.

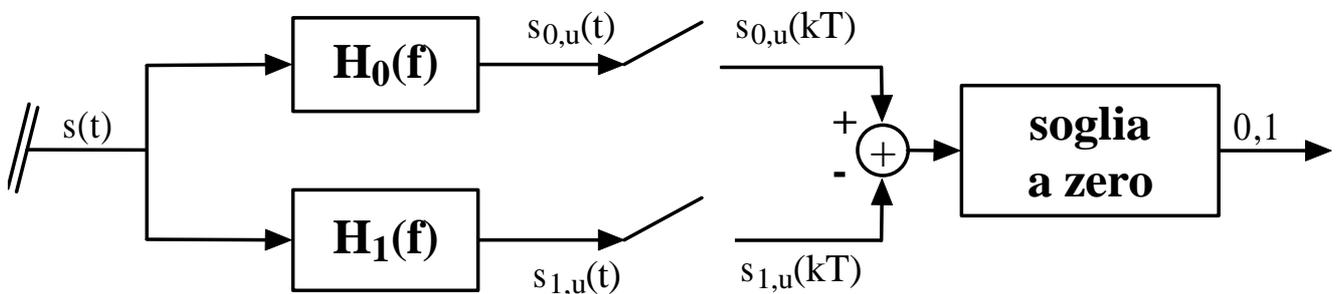
Il modo più intuitivo di procedere è quello di associare una forma d'onda $s_0(t)$ alla trasmissione di uno 0 e una diversa forma d'onda $s_1(t)$ alla trasmissione di un 1. Se si adotta questa scelta, a prescindere da come siano fatte le due forme d'onda, il primo problema che si pone è sul filtro in ricezione, che potrà essere adattato ad una sola delle due forme d'onda. E' sufficiente, allora, disporre due distinti filtri, uno adattato a $s_0(t)$ e l'altro adattato a $s_1(t)$:



L'adattamento dei due filtri è tale, quindi, che ognuno di essi massimizzi la propria uscita, nell'istante di lettura, quando riceve in ingresso il "proprio" segnale, ossia $s_0(t)$ per $H_0(f)$ e $s_1(t)$ per $H_1(f)$. Confrontando semplicemente le due uscite, basterà fare riferimento a quella di valore maggiore per decidere se è stato trasmesso 1 oppure 0:

- se $s_{0,u}(kT) > s_{1,u}(kT)$, allora il simbolo trasmesso è 0;
- se $s_{0,u}(kT) < s_{1,u}(kT)$, allora il simbolo trasmesso è 1;

In modo del tutto equivalente, se $s_{0,u}(kT) - s_{1,u}(kT) > 0$, allora il simbolo trasmesso è 0, altrimenti il simbolo trasmesso è 1. Questo significa che, una volta fatta la differenza tra i due segnali campionati, l'organo decisore può semplicemente essere un comparatore, ossia un **sistema a soglia zero**:



Il problema successivo è quindi quello di scegliere le due forme d'onda $s_0(t)$ e $s_1(t)$ in modo che lo schema appena tracciato funzioni nel modo migliore possibile. Ci sono allora almeno due criteri possibili.

Un primo criterio consiste nel fare in modo che, quando l'uscita di un filtro è massima, l'altra è invece nulla:

- avendo trasmesso 0, deve risultare $s_{0,u}(kT) = [s_{0,u}(t)]_{MAX}$ e $s_{1,u}(kT) = 0$;
- avendo trasmesso 1, deve risultare $s_{1,u}(kT) = [s_{1,u}(t)]_{MAX}$ e $s_{0,u}(kT) = 0$

Si parla, in questo caso, di **codifica ortogonale**. Il motivo di questo nome è semplice: sia $s(t)$ il segnale che giunge dal mezzo trasmissivo; questo segnale va in ingresso ai due filtri adattati, ciascun dei quali, per quanto sappiamo dallo studio dei sistemi lineari tempo-invarianti, effettua semplicemente la convoluzione di $s(t)$ con la propria risposta all'impulso:

$$u(t) = x(t) * h(t)$$

Nel caso del filtro adattato, abbiamo allora che

$$H_0(f) = S_0^*(f) \longrightarrow s_{0,u}(t_m) = \int_{-\infty}^{+\infty} s(\alpha) s_0(\alpha) d\alpha$$

$$H_1(f) = S_1^*(f) \longrightarrow s_{1,u}(t_m) = \int_{-\infty}^{+\infty} s(\alpha) s_1(\alpha) d\alpha$$

Supponiamo allora che sia stato trasmesso lo 0, per cui $s(t)=s_0(t)$: l'uscita del primo filtro deve essere massima, mentre l'altra deve essere nulla, per cui la condizione che le due forme d'onda devono rispettare è

$$\int_{-\infty}^{+\infty} s_0(\alpha)s_1(\alpha)d\alpha = 0$$

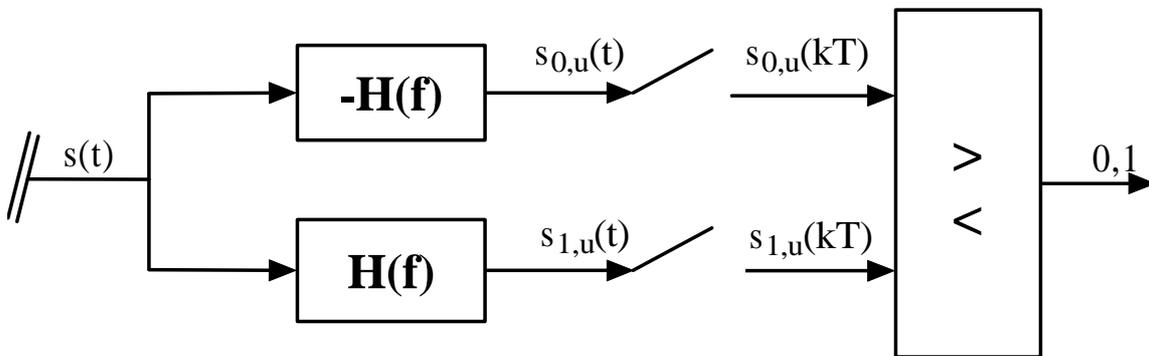
Questa è la definizione di **segnali ortogonali** tra loro.

La stessa condizione si ottiene, ovviamente, nel caso di 1 trasmesso, imponendo che sia nulla l'uscita del primo filtro.

Si può fare, però, qualcosa di ancora migliore: *si può fare in modo che, quando una uscita è massima in positivo, l'altra sia massima in negativo.* Questo tipo di codifica di linea è detta **codifica antipodale**.

E' molto facile ottenere una codifica antipodale: basta usare la stessa forma d'onda per la trasmissione di 1 e 0, cambiata però di segno. Per esempio, se la forma d'onda corrispondente ad 1 è un rettangolo di ampiezza +V, quella associata a 0 sarà lo stesso rettangolo, ma di ampiezza -V.

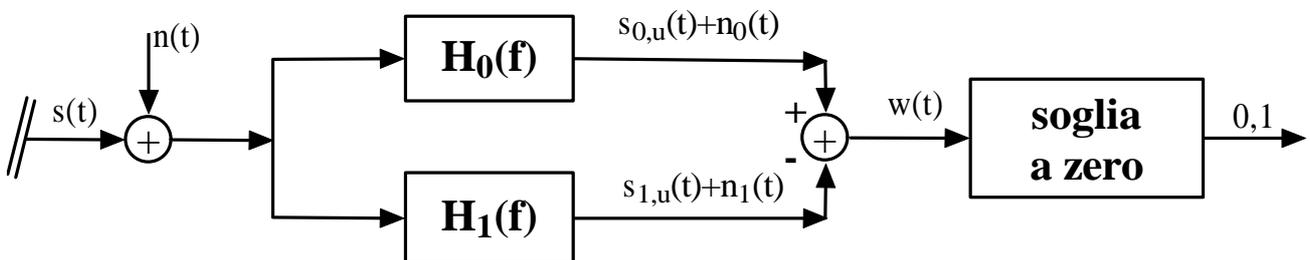
E' chiaro che, con questa scelta, i due filtri in ricezione differiranno semplicemente per il segno:



Quando il simbolo è 0, si trasmette $-s(t)$ e si ottiene la massima uscita positiva dal ramo superiore e la massima uscita negativa dal ramo inferiore; il contrario si ottiene invece quando il simbolo è 1, ossia trasmettendo $s(t)$.

Questa codifica riduce la probabilità di errore. Per rendercene conto, facciamo il discorso seguente.

Consideriamo per prima la codifica ortogonale, includendo questa volta il rumore:



(per semplicità, abbiamo trascurato di indicare il campionamento all'uscita dei filtri).

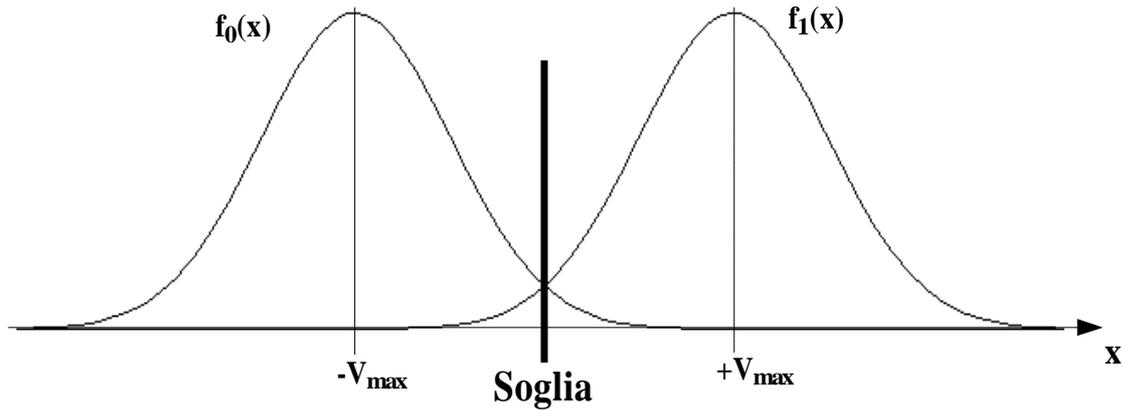
Nel caso in cui sia stato trasmesso $s_0(t)$, l'uscita del primo filtro è massima, mentre quella del secondo è nulla. Quindi, il segnale in ingresso al comparatore è

$$w(t) = [s_{0,u}(t)]_{MAX} + n_0(t) - n_1(t) = V_{MAX} + n_0(t) - n_1(t)$$

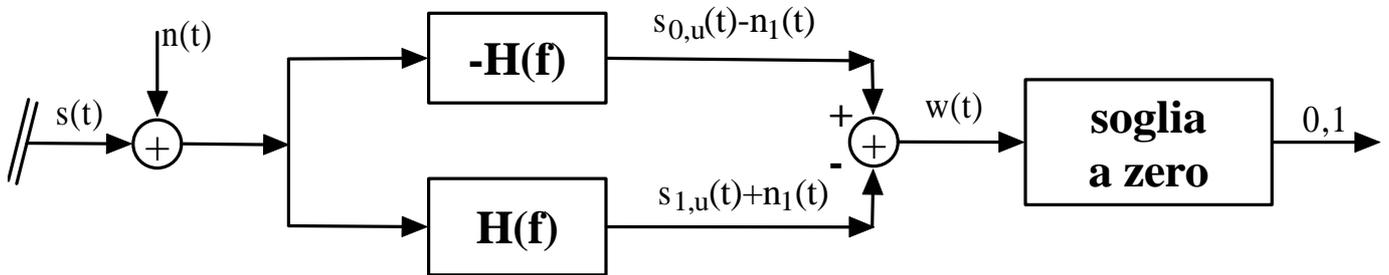
Se invece è stato trasmesso $s_1(t)$, il segnale diventa

$$w(t) = n_0(t) - [s_{1,u}(t)]_{MAX} - n_1(t) = -V_{MAX} + n_0(t) - n_1(t)$$

In entrambi i casi, il rumore sovrapposto è $n_0(t)-n_1(t)$, mentre il valore di picco del segnale è, in modulo, lo stesso. Dato che il successivo campionatore deve misurare il valore massimo del segnale in uscita dal filtro, possiamo allora affermare che la distribuzione statistica del segnale in uscita dal campionatore è quella illustrata nella figura seguente:



La distanza tra le campane, centrate sui valori $+V_{max}$ e $-V_{max}$.
 Passiamo alla codifica antipodale:



Nel caso in cui sia stato trasmesso $s_0(t)$, l'uscita del primo filtro è al massimo positivo, mentre l'uscita del secondo è al massimo negativo, per cui il segnale in ingresso al campionatore è

$$w(t) = [s_{0,u}(t)]_{MAX} - n_1(t) - [-s_{1,u}(t)]_{MAX} - n_1(t) = 2V_{MAX} - 2n_1(t)$$

Il rumore che esce dai due filtri è ovviamente lo stesso tranne un cambiamento di segno, ed è ovvio che sia così, in quanto i due filtri hanno la stessa funzione di trasferimento, salvo il segno, e ricevono in ingresso lo stesso rumore.

Se invece è stato trasmesso $s_1(t)$, il segnale è

$$w(t) = [-s_{1,u}(t)]_{MAX} - n_1(t) - [s_{1,u}(t)]_{MAX} - n_1(t) = -2V_{MAX} - 2n_1(t)$$

In questo caso, quindi, il rumore sovrapposto ha la stessa entità di prima, ma le due campane sono centrate su $-2V_{max}$ e $+2V_{max}$, ossia sono a distanza doppia rispetto a prima. Questo fa sì che

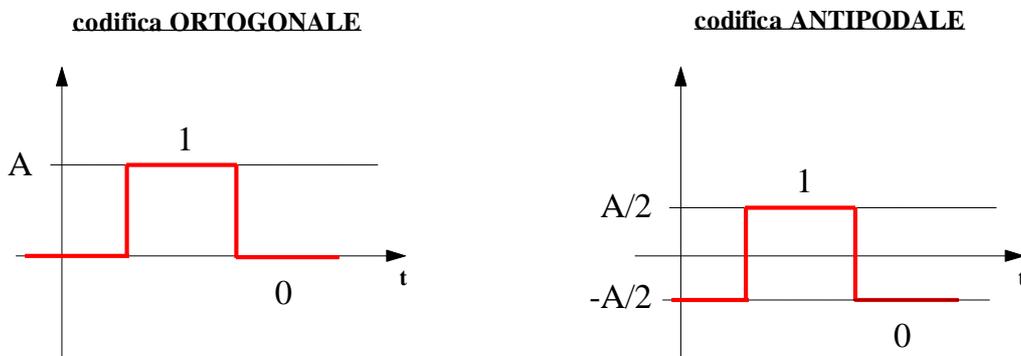
diminuisca l'area sottesa dalle code delle gaussiane delimitate dalla soglia, ossia che diminuisca la probabilità di errore.

In definitiva, la codifica antipodale risulta sicuramente preferibile, rispetto alla codifica ortogonale, dal punto di vista della minimizzazione della probabilità di errore. Ovviamente, questo vale nell'ipotesi che, in entrambi i casi, il picco del segnale in uscita dal filtro sia V_{MAX} o $-V_{MAX}$.

Questo non è l'unico criterio valido per la scelta della codifica: è anche importante valutare quanta potenza serve trasmettere nel caso di una o dell'altra codifica.

Infatti, è abbastanza intuitivo⁷ comprendere che si può ottenere la stessa probabilità di errore, con la codifica ortogonale e con quella antipodale, scegliendo diversamente i valori massimi e minimi in uscita dai filtri: proprio perché il fattore determinante è la distanza tra le "campane gaussiane", se si imposta una codifica antipodale dove i valori di uscita sono $+V_{max}$ e $-V_{max}$, si ottiene una probabilità di errore esattamente pari a quella che ottiene con una codifica ortogonale dove i valori di uscita sono 0 e $2V_{max}$.

In altre parole, considerando ad esempio la sequenza 10 (che si legge uno-zero), si ottiene la stessa probabilità di errore trasmettendo le seguenti forme d'onda:



Le due forme d'onda vanno però confrontate dal punto di vista della potenza di picco e della potenza media. Per quanto riguarda la *potenza di picco*, è ovvio che nel caso della codifica antipodale è minore: il valore di picco è metà rispetto a quello della codifica ortogonale, per cui la potenza di picco è 1/4, ossia 6dB inferiore:

$$\begin{aligned} \text{ortogonale} &\rightarrow V_{\text{picco}} \rightarrow P_{\text{picco}} = V_{\text{picco}}^2 \\ \text{antipodale} &\rightarrow V_{\text{picco}}/2 \rightarrow P_{\text{picco}} = V_{\text{picco}}^2 / 4 \end{aligned}$$

D'altra parte, mentre nella codifica ortogonale non dobbiamo trasmettere nulla quando il simbolo è 0, nella codifica antipodale dobbiamo trasmettere invece un rettangolo di ampiezza $-A/2$, il che va evidentemente a svantaggio della *potenza media*: infatti, si trova infatti che la potenza media associata alla codifica ortogonale è $V_{\text{picco,ort}}^2 / 2$, mentre quella associata alla codifica antipodale è $V_{\text{picco,ant}}^2$ (doppia della precedente). Se i due valori di picco sono uguali, la codifica ortogonale consente dunque un risparmio di 3dB di potenza media; al contrario, se il picco nell'antipodale è metà di quello nella ortogonale, come supposto prima, la situazione si inverte: la potenza media della codifica antipodale diventa metà dell'altra, per cui è la codifica antipodale che ci fa risparmiare 3dB di potenza media.

⁷ Si può anche fare una verifica analitica

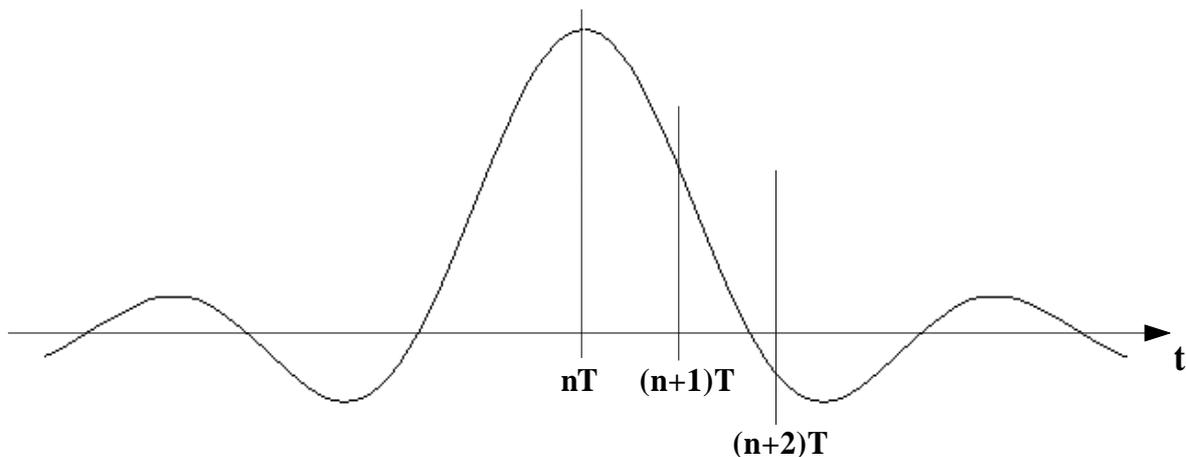
Il motivo del risparmio è nel fatto che la forma d'onda nel caso ortogonale è pari a quella nel caso antipodale, cui però è sovrapposta una continua che avrà una certa potenza. Quindi, *con la codifica antipodale possiamo risparmiare la spesa di inviare in linea una potenza media costante* (tensione continua), che peraltro non porta alcuna informazione ed è del tutto inutile per il funzionamento del sistema di trasmissione.

La conclusione è dunque che, scelti opportunamente i picchi (in modo da garantire la stessa probabilità di errore finale), la codifica antipodale è sempre preferibile.

PROBLEMA DELL'INTERFERENZA DI INTERSIMBOLO

Nel paragrafo precedente abbiamo esaminato alcuni semplici criteri con cui scegliere le forme d'onda elementari $s_0(t)$ ed $s_1(t)$ da associare ai simboli 0 ed 1 e da inviare all'ingresso del filtro adattato: un primo criterio è quello della ortogonalità, mentre l'altro è quello di adottare forme d'onda identiche a meno del segno. Un altro importantissimo criterio con cui effettuare questa scelta riguarda il problema dell'*interferenza di intersimbolo*: abbiamo visto che, in corrispondenza di ciascun segnale elementare ricevuto in ingresso, il filtro adattato genera una certa forma d'onda, la quale, nell'istante di lettura, presenta il massimo valore possibile, in modo da facilitare il funzionamento del successivo del sistema a soglia; tuttavia, tale forma d'onda non si esaurisce dopo un tempo T , per cui va a sommarsi, sia pure con ampiezza ridotta, sia alle forme d'onda corrispondenti ai simboli successivi sia a quelle corrispondenti ai simboli precedenti, andando perciò ad influenzare gli altri campionamenti. Questo è appunto il concetto dell'*interferenza tra simboli successivi* o **interferenza intersimbolo**.

Consideriamo un esempio concreto. Supponiamo che la forma d'onda in uscita dal filtro adattato e corrispondente alla trasmissione di un 1 sia quella indicata nella figura seguente:



Come si vede, essa presenta il massimo valore nell'istante nT , che quindi sarà l'istante di lettura, ma presenta anche dei massimi secondari in istanti successivi.

Supponiamo allora che il successivo simbolo trasmesso sia 0, per cui la forma d'onda trasmessa è nulla (supponendo una codifica ortogonale). Il campionario andrà a leggere il valore del segnale nell'istante $(n+1)T$ e troverà una tensione non nulla (come dovrebbe invece essere), dovuta alla composizione delle forme d'onda dei simboli precedenti e successivi (quelle più influenti sono ovviamente relative al simbolo immediatamente precedente ed a quello immediatamente successivo): se il valore di tale tensione è superiore alla soglia, il decisore interpreterà il simbolo trasmesso come un 1, commettendo perciò un errore.

Non solo, ma se il simbolo trasmesso fosse stato effettivamente un 1, la composizione dei massimi secondari corrispondenti alle varie forme d'onda potrebbe portare ad errori sui simboli

successivi, così come anche la composizione dei minimi, che potrebbero portare il segnale al di sotto della soglia quando invece dovrebbe rimanere al di sopra per indicare che è stato trasmesso un 1.

E' chiaro che questa situazione è intollerabile, per cui bisogna porvi rimedio.

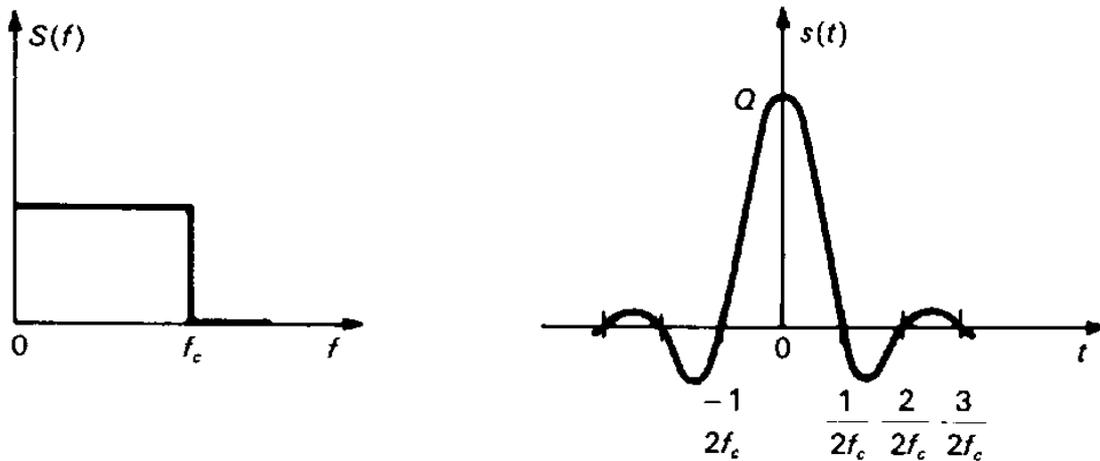
Un rimedio molto semplice potrebbe essere il seguente: dato che un sistema reale è causale, basterebbe tener conto degli ultimi simboli trasmessi e cancellare di volta in volta l'interferenza. In altre parole, una volta accertato quale sia il simbolo trasmesso, basta generare localmente la corrispondente forma d'onda e sottrarla a quella ricevuta. Questo, però, funzionerebbe se non ci fossero errori nell'interpretazione: in presenza di anche un solo errore, si perderebbe completamente la sequenza giusta.

Il problema, dunque, si può porre nei seguenti termini: *oltre a sfruttare al massimo l'energia della singola forma d'onda (giungendo così al concetto del filtro adattato), è necessario ottenere, in uscita dal filtro, delle forme d'onda con la minore interferenza reciproca possibile.* Vediamo allora quali possono essere queste forme d'onda.

Un primo esempio potrebbe essere la funzione Seno Cardinale indicata prima, avente la seguente espressione analitica:

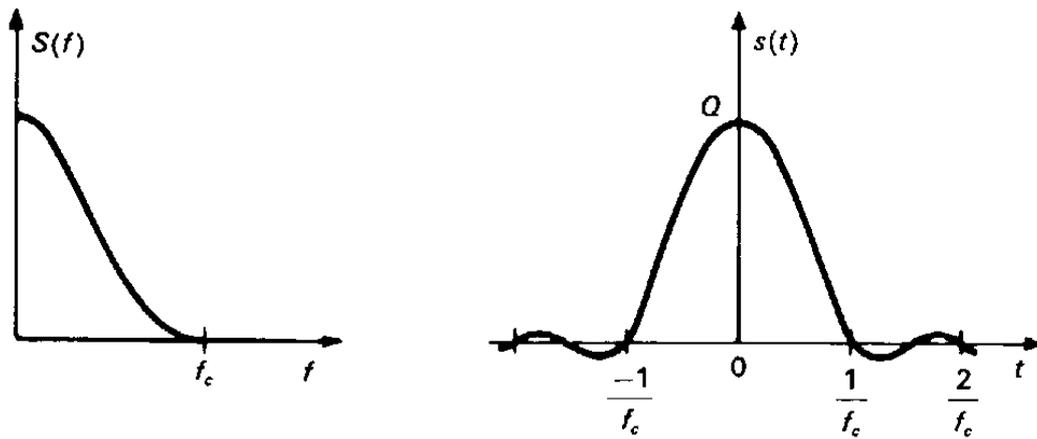
$$s(t) = Q \frac{\sin(2\pi f_c t)}{2\pi f_c t}$$

Nella figura seguente vengono riportati la forma d'onda e il corrispondente spettro (che deve racchiudere un'area pari a Q):



Le oscillazioni laterali che compaiono in questa forma d'onda decrescono lentamente allontanandosi dall'impulso principale. Precisamente, esse variano inversamente al tempo, cioè con $1/|t|$. Lo spettro è costante sull'intervallo $[0, f_c]$. Il valore della banda f_c di questo segnale può essere scelto imponendo che la forma d'onda sia "praticamente" contenuta nell'intervallo T : quanto maggiore è f_c , tanto maggiore è l'attenuazione. Si verifica, infatti, che l'ampiezza delle oscillazioni diventa $1/10$ dell'ampiezza massima nell'istante per cui $2f_c|t| > 3$, mentre diventa $1/100$ nell'istante per cui $2f_c|t| > 32$.

Per ottenere un decremento ancora più rapido delle oscillazioni laterali si può pensare ad un segnale il cui spettro, anziché essere costante nella banda f_c , abbia una transizione più graduale dal valore massimo al valore 0. Per esempio, nella figura seguente è riportato uno spettro a forma di \cos^2 e la corrispondente forma d'onda:



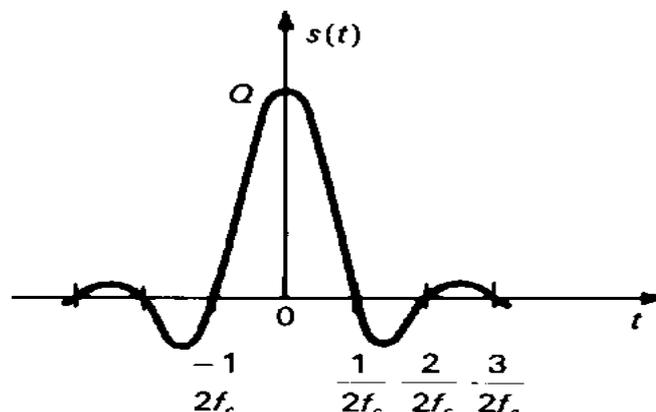
L'espressione analitica di questa forma d'onda, che per semplicità non riportiamo, indica che, per $|t|$ abbastanza più grande di $1/2f_c$, l'ampiezza delle oscillazioni tende ad essere inversamente proporzionale a $|t|^3$ anziché a $|t|$ come nel caso precedente. Da qui consegue che già dopo il primo zero (che si verifica per $|t|=1/f_c$) l'ampiezza delle oscillazioni è meno di $1/10$ rispetto all'ampiezza massima.

Osservazione

I due esempi appena descritti mostrano una cosa interessante: fissato un certo ritmo di trasmissione $1/T$ dei segnali elementari e fissata anche la massima interferenza tollerabile tra forme d'onda adiacenti, è possibile individuare un tipo di forma d'onda caratterizzata dalla minima banda f_c possibile. Questo è il motivo per cui, *volendo occupare la minima banda possibile, conviene trasmettere tutti i segnali elementari con una stessa forma d'onda distinguendoli uno dall'altro per la diversa ampiezza.*

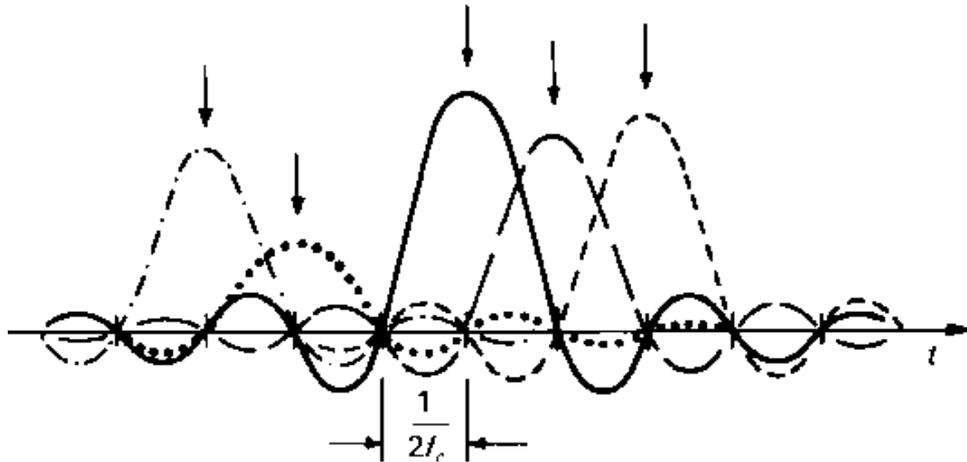
Le due forme d'onda precedentemente illustrate hanno dunque la caratteristica di esaurirsi in modo più o meno veloce, garantendo così una interferenza, tra forme d'onda adiacenti, tanto minore quanto maggiore è l'attenuazione nel tempo. D'altra parte, però, si può fare una osservazione importante: se l'unico parametro variabile, per ciascuna forma d'onda elementare, non è la forma ma l'ampiezza, *non è in effetti necessario che le successive forme d'onda si esauriscano praticamente nell'intervallo; al contrario, è sufficiente che esse, pur sovrapponendosi nel tempo, consentano comunque la determinazione precisa delle rispettive ampiezze.*

Per comprendere questa considerazioni, basta considerare nuovamente la forma d'onda Seno Cardinale prima illustrata:



Si osserva che gli zeri di questa forma d'onda sono distanziati, dall'impulso principale, di multipli di $1/2f_c$. Allora, se si fa in modo che, ad ogni intervallo $T=1/2f_c$, venga fuori dal filtro una forma d'onda come quella di figura, i vari simboli, anche se sovrapposti, risultano perfettamente distinguibili, in quanto, effettuando la lettura (o *campionamento*) della forma d'onda complessiva negli istanti multipli di $1/2f_c$, soltanto l'impulso corrispondente al simbolo voluto è diverso da zero.

La figura seguente mostra il risultato di una simile scelta, indicando 5 diverse forme d'onda elementari e la loro sovrapposizione:

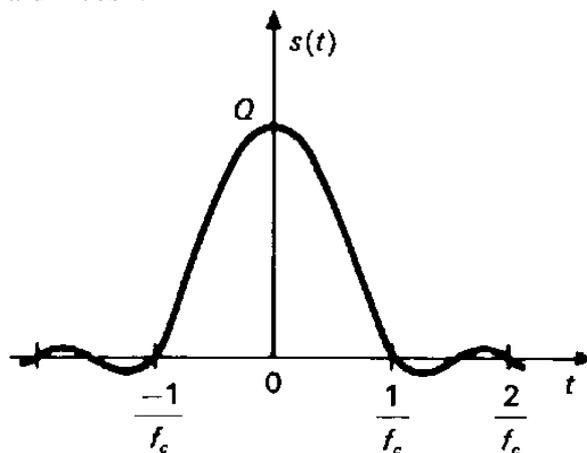


Effettuando la lettura negli istanti indicati dalle freccette, si ottiene sempre e solo l'ampiezza massima della forma d'onda associata al simbolo di interesse.

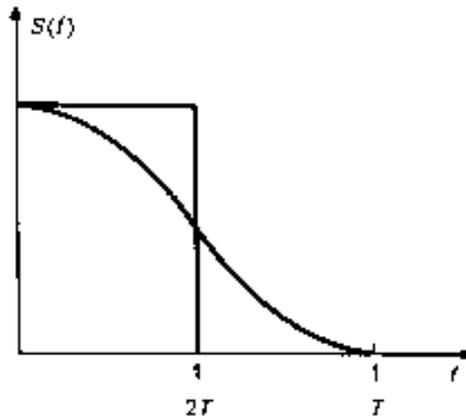
Forme d'onda come queste si dicono **ad intersimbolo nullo**: esse hanno il valore massimo nel proprio istante di campionamento e valgono invece zero negli altri istanti di campionamento, in modo da non interferire con le forme d'onda successive. Si parla anche di forme d'onda **a zeri equidistanti**, proprio perché gli istanti in cui valgono zeri, corrispondendo agli istanti di campionamento, sono equidistanti tra loro.

Questa soluzione presenta diversi vantaggi (tra i quali quello di consentire la massima velocità di trasmissione compatibile con una certa banda assegnata), ma è molto critica da un altro punto di vista: *è sufficiente che ci sia un piccolo errore nell'istante di campionamento, perché il contributo degli impulsi adiacenti a quello considerato diventi molto importante.*

A questa soluzione critica se ne possono sostituire delle altre, che sono meno critiche a prezzo però di una maggiore occupazione di banda (a parità di velocità). Per esempio, riprendiamo la forma d'onda il cui spettro ha forma di \cos^2 :



Anche in questo caso, gli zeri risultano equidistanti tra loro e dall'impulso principale, per cui anche questa forma d'onda è ad intersimbolo nullo. La distanza tra uno zero e l'altro vale $1/f_c$, il che significa che la trasmissione deve eventualmente avvenire scegliendo $f_c=1/T$. Nel caso precedente, bisognava invece prendere $f_c=1/2T$, ossia una banda metà di quella considerata in questo caso:



Forme d'onda di Nyquist

Le due forme d'onda considerate prima non sono frutto di tentativi, ma sono deducibili da proprietà generali, alle quali si giunge durante la ricerca degli spettri delle forme d'onda che presentano zeri equidistanti tra loro e dall'impulso principale. Vediamo allora come si procede.

Consideriamo una forma d'onda generica $s(t)$. Supponiamo che essa presenti degli zeri equidistanti: analiticamente, se vogliamo che gli zeri siano negli istanti $t = \pm kt_0$, con $k=1,2,\dots$, dobbiamo imporre che

$$s(\pm kt_0) = 0 \quad k = 1, 2, \dots$$

Se adesso indichiamo con $S(f)$ lo spettro (bilatero) di questo segnale, possiamo esprimere $s(t)$ come antitrasformata di Fourier:

$$s(t) = \int_{-\infty}^{+\infty} S(f) e^{j2\pi ft} df$$

Imponendo la condizione sugli zeri equidistanti, questa diventa

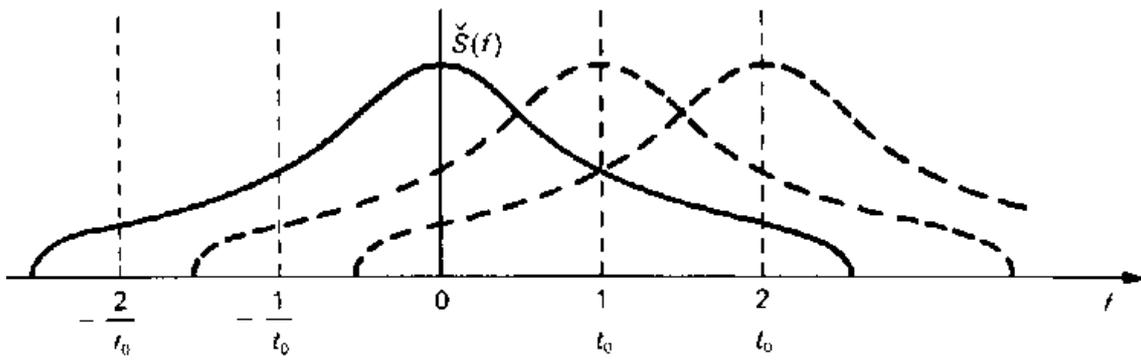
$$s(\pm kt_0) = \int_{-\infty}^{+\infty} S(f) e^{\pm j2\pi fkt_0} df = 0 \quad (*)$$

La condizione così ottenuta è una condizione sullo spettro di $s(t)$. Vogliamo allora renderci conto di cosa comporti questa condizione.

Supponiamo allora di avere un segnale $X(f)$ il cui spettro è ottenuto ripetendo $S(f)$ con periodo pari a $1/t_0$:

$$X(f) = \sum_{k=-\infty}^{+\infty} S\left(f - \frac{k}{t_0}\right)$$

Una operazione di questo tipo è illustrata nella figura seguente:



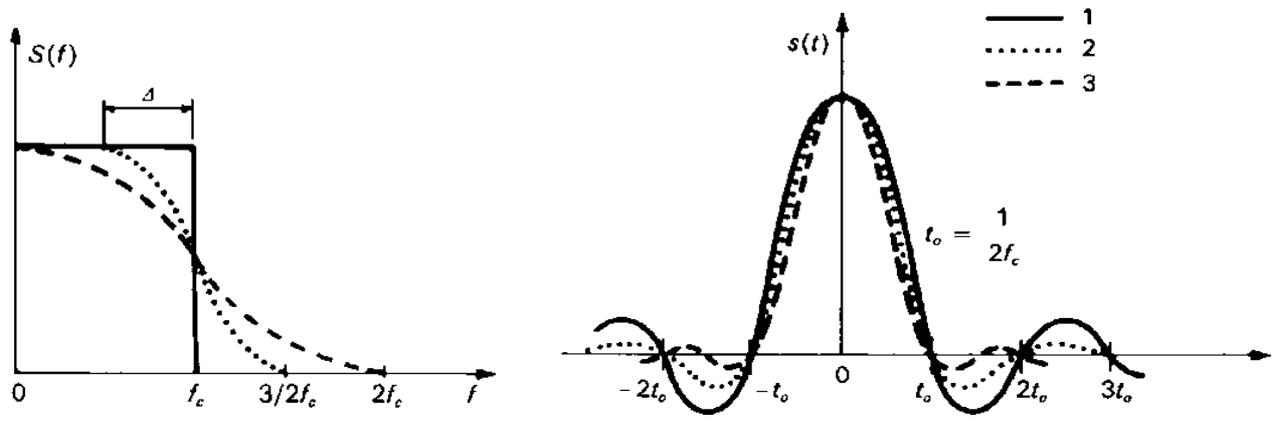
Trattandosi di una funzione periodica, $X(f)$ è esprimibile tramite una serie di Fourier; si tratterebbe di calcolare i coefficienti dello sviluppo. Tuttavia, se imponiamo che $S(f)$ soddisfi alla condizione indicata dalla (*), si trova che i termini dello sviluppo sono tutti nulli tranne quello costante. In altre parole, quindi, se $S(f)$ rispetta la condizione (*), $X(f)$ è una funzione costante:

$$X(f) = \sum_{k=-\infty}^{+\infty} S\left(f - \frac{k}{t_0}\right) = K$$

D'altra parte, $X(f)$ è una ripetizione periodica di $S(f)$, per cui $S(f)$ potrà banalmente essere a sua volta costante, ovviamente nell'intervallo $1/t_0$:

$$S(f) = K$$

Quindi, una soluzione del problema è quella indicata con 1 (linea a tratto continuo) nella figura seguente, dove sono indicati sia lo spettro sia la corrispondente forma d'onda:



Abbiamo cioè ottenuto ancora una volta la funzione Seno Cardinale.

Il valore della costante K si ricava facilmente: l'area sottesa da $S(f)$, che per la formula di antitrasformazione corrisponde a $s(0)$, è pari all'area sottesa, nell'intervallo $1/t_0$, dalla retta corrispondente a K : quindi

$$s(0) = K \cdot \frac{1}{t_0} \longrightarrow K = s(0)t_0$$

Quello trovato, però, non è l'unico spettro $S(f)$ che soddisfa la (*) e cioè la condizione di zeri equidistanti nel tempo. Ci sono infatti infinite altre funzioni $S(f)$ che, ripetute con periodo $1/t_0$, danno luogo ad una costante K . In base ai discorsi fatti, si tratta di tutte le funzioni tali che

$$\sum_{k=-\infty}^{+\infty} S\left(f - \frac{k}{t_0}\right) = K = s(0)t_0$$

Naturalmente, è sufficiente che questa condizione sia soddisfatta nel periodo $1/t_0$ (siamo nel dominio della frequenza), per cui ad esempio si può imporre che valga nell'intervallo $\left[-\frac{1}{2t_0}, +\frac{1}{2t_0}\right]$:

$$\sum_{k=-\infty}^{+\infty} S\left(f - \frac{k}{t_0}\right) = K = s(0)t_0 \quad -\frac{1}{2t_0} \leq f \leq \frac{1}{2t_0}$$

Dato che siamo interessati a spettri con banda limitata (in quanto è limitata la banda a disposizione sul mezzo trasmissivo), consideriamo solo funzioni $S(f)$ limitate tra $-1/t_0$ e $1/t_0$: se indichiamo con $f_0=1/t_0$ la frequenza di ripetizione degli impulsi, $S(f)$ deve essere limitato tra $\pm f_0$.

In definitiva, lo spettro che ricerchiamo deve godere delle seguenti proprietà:

$$\begin{aligned} \sum_{k=-\infty}^{+\infty} S\left(f - \frac{k}{t_0}\right) &= s(0)t_0 & -\frac{f_0}{2} \leq f \leq \frac{f_0}{2} \\ S(f) &= 0 & f \leq -f_0 \cap f \geq f_0 \end{aligned}$$

Le forme d'onda il cui spettro soddisfa queste condizioni appartengono alla cosiddetta **famiglia di Nyquist**. La funzione più semplice, come si è visto prima, è il Seno Cardinale. Proprio dallo spettro rettangolare del Seno Cardinale si ottengono altri spettri che fanno al caso nostro: basta aggiungere a tale spettro una funzione dispari intorno alle frequenze $\pm f_0/2$ (e ovviamente pari intorno a $f=0$, in modo da avere forme d'onda reali). Alcuni esempi di funzioni del genere sono riportati nella stessa figura di prima: si tratta di spettri con transizione sinusoidale, comprendenti anche il caso estremo in cui l'intervallo di transizione occupa tutta l'estensione dello spettro. Tra queste funzioni, c'è anche quella con spettro a forma di \cos^2 di cui si è parlato prima.

All'aumentare della transizione dello spettro, le ondulazioni temporali, che precedono e seguono l'impulso principale, vanno diminuendo di importanza. In questo modo, si ha a disposizione un insieme di forme d'onda, tutte con gli zeri nei punti desiderati, tra cui si può scegliere adottando un adeguato compromesso tra criticità della soluzione e banda occupata:

- lo spettro rettangolare è evidentemente quello con la banda minima possibile;
- lo spettro a forma di \cos^2 dà invece luogo ad un impulso la cui ampiezza, per $|t| > t_0$, non supera mai $1/10$ dell'ampiezza dell'impulso principale: questo significa che, per segnali che tollerano una interferenza intersimbolo di $1/10$, come proprio i segnali numerici, non ha importanza l'esatta sincronizzazione tra la durata dell'impulso e il periodo di ripetizione degli impulsi.

Roll off delle forme d'onda ad intersimbolo nullo

C'è un parametro che quantifica lo smorzamento dello spettro del segnale considerato: si definisce, infatti, **roll off** (simbolo: δ) il rapporto l'intervallo di frequenza $\Delta f = \frac{f_c}{2} - f_x$ (dove f_x è la frequenza a partire dalla quale comincia lo smorzamento) e la frequenza $\frac{f_c}{2}$ (metà della frequenza di cifra):

$$\delta = \frac{\Delta f}{\frac{f_c}{2}} = \frac{\frac{f_c}{2} - f_x}{\frac{f_c}{2}}$$

Nel caso della funzione Seno Cardinale, lo spettro ha transizione netta dal valore massimo al valore minimo, il che significa che $f_x = \frac{f_c}{2}$, ossia che $\Delta f=0$ e quindi anche $\delta=0$. Nel caso, invece, del Coseno rialzato (linea tratteggiata nell'ultima figura riportata), risulta $f_x=0$ e quindi $\delta=1$.

Il roll off, in base alle considerazioni fatte prima, quantifica anche la banda occupata dal segnale considerato: infatti, è evidente che tale banda è

$$B = \frac{f_c}{2} + \Delta f = \frac{f_c}{2} + \frac{f_c}{2} \delta = \frac{f_c}{2} (\delta + 1)$$

Come già detto prima, *all'aumentare dell'roll off, ossia al diminuire dello smorzamento dello spettro, aumenta la banda utilizzata e quindi aumenta la velocità di decadimento a zero della funzione nel tempo*. Ovviamente, si ottengono sempre forme d'onda che tendono asintoticamente a zero, per cui sono di durata teoricamente infinita; d'altra parte, però, tali forme d'onda possono essere ritenute nulle quando il loro valore scende al di sotto del livello del rumore.

FORMA D'ONDA DA TRASMETTERE

Nei paragrafi precedenti abbiamo dunque fissato un basilare vincolo nel nostro sistema di trasmissione numerico: abbiamo infatti stabilito che, all'uscita dal filtro in ricezione, è necessario ottenere forme d'onda a intersimbolo nullo. Il secondo vincolo da soddisfare è evidentemente quello di garantire che l'intero sistema funzioni, con una accettabile probabilità di errore, con la minima potenza trasmessa. Ci sono allora vari modi di procedere:

- il primo modo è quello di fissare il tipo di filtro in ricezione: in questo caso, come si vedrà tra un attimo, risulta automaticamente fissata la *fisionomia* del segnale che dobbiamo trasmettere sul mezzo trasmissivo in modo che giunga al ricevitore e garantisca le prestazioni desiderate;
- un secondo modo è invece quello di scegliere la forma d'onda $s_t(t)$ più comoda per la trasmissione: in questo caso, va quindi determinato il filtro ottimo in ricezione, che produca cioè le prestazioni ottime del sistema compatibilmente con l'ottenimento dell'intersimbolo nullo in uscita;

- un ultimo modo è, infine, quello di fissare sia $s_t(t)$ sia la forma d'onda $s_u(t)$ in uscita dal filtro, dopo di che quest'ultimo resta determinato a priori; in particolare conviene effettuare l'analisi per diverse coppie $s_t(t)$, $s_u(t)$ in modo da disporre di opportune possibilità di scelta.

In questo paragrafo ci occupiamo del primo modo di procedere, mentre successivamente saranno esaminati gli altri modi.

Supponiamo, dunque, di aver stabilito quale debba essere la forma d'onda in uscita dal filtro di ricezione: dovrà trattarsi di una forma d'onda ad intersimbolo nullo e con un assegnato *roll off* (ad esempio $\delta=0$ se vogliamo una forma d'onda del tipo $\sin(t)/t$). In secondo luogo, al fine di massimizzare il rapporto S/N in uscita dal filtro e quindi dal successivo campionatore, supponiamo di fissare anche il filtro, scegliendo un **filtro adattato** al segnale ricevuto. Allora, *se è fissata la forma d'onda in uscita ed è fissato il filtro, risulta automaticamente fissata la forma d'onda in ingresso al filtro stesso*. In realtà, è facile accorgersi che è fissato solo il modulo dello spettro di tale forme d'onda, mentre rimane ancora da fissare la fase.

Sia $S_R(f) = |S_R(f)|e^{j(S_R(f))}$ lo spettro della forma d'onda in ingresso al filtro: essendo quest'ultimo, per ipotesi, adattato, avrà una funzione di trasferimento

$$H(f) = |S_R(f)|e^{-j(S_R(f))}$$

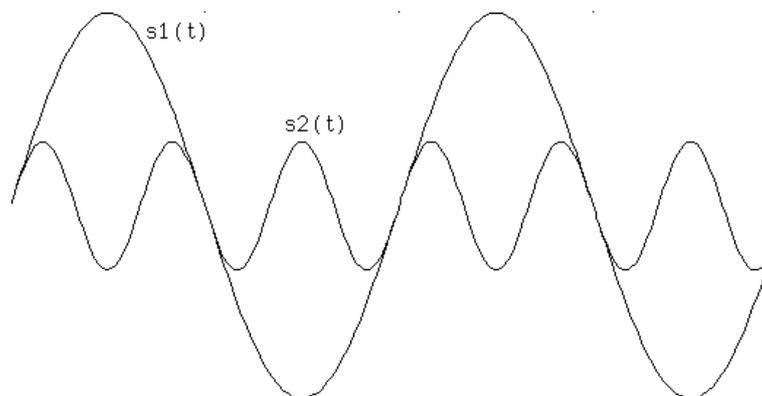
Il prodotto tra le due funzioni rappresenta lo spettro della forma d'onda in uscita dal filtro: si tratta evidentemente di

$$X(f) = |S_R(f)|^2$$

Su questo spettro dobbiamo imporre il vincolo che corrisponda ad una forma d'onda ad intersimbolo nullo. Di conseguenza, *tale vincolo fissa solo il modulo dello spettro del segnale da ricevere*. Non ci sono invece vincoli sulla fase di $S_R(f)$ ed è ovvio che sia così: quale che sia questa fase, ci pensa comunque il filtro adattato ad eliminarla.

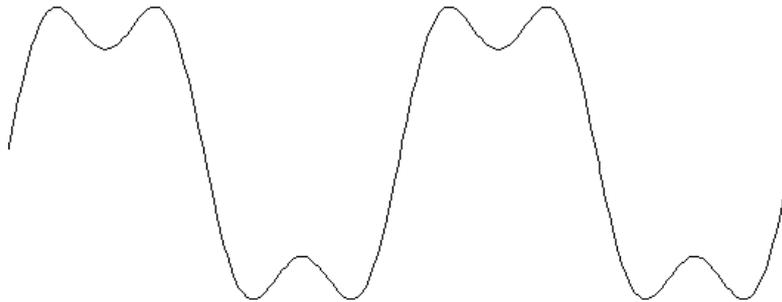
Abbiamo quindi la totale libertà di scelta sulla fase di $S_R(f)$. Come spendiamo questo grado di libertà? E' ovvio che lo spendiamo in modo da ottenere un segnale particolarmente facile da "maneggiare". *Conviene allora scegliere $e^{j(S_R(f))}$ in modo che la forma d'onda corrispondente abbia il fattore di picco⁸ più piccolo possibile.*

Per comprendere il significato di questa scelta, consideriamo due sinusoidi, una di ampiezza A_1 e frequenza f_1 e l'altra di ampiezza $A_1/3$ e frequenza $3f_1$:



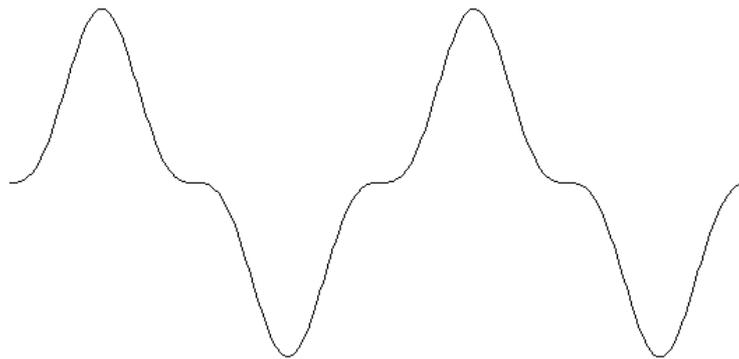
⁸ Il **fattore di picco** di un segnale è il rapporto tra la potenza di picco e la potenza media.

Supponiamo di sommare le due forme d'onda: la potenza media della somma è sempre la stessa, quale che sia lo sfasamento tra le due, proprio perché la potenza media di ciascuna di esse è la stessa. Ciò che cambia, invece, al variare dello sfasamento è il valore di picco e quindi la potenza di picco. Se le due sinusoidi sono in fase, come nella figura precedente, allora la loro composizione da origine alla seguente forma d'onda (sempre a valor medio nullo):



Ad esempio, se le due sinusoidi hanno ampiezza 1V e 3V, il valore di picco della loro somma risulta essere 2.8271V.

Se, invece, le due sinusoidi sono sfasate di 180°, la forma d'onda risultante è la seguente (sempre a valor medio nullo):

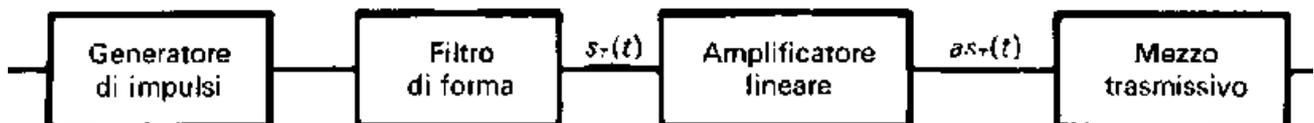


Il valore di picco risulta essere 3.9965V. La potenza di picco è quindi decisamente maggiore rispetto a prima, a parità di potenza media: il fattore di picco è aumentato.

In definitiva, variando la fase relativa delle componenti spettrali, si modifica la forma della funzione nel tempo, modificando così il fattore di picco. Il principio cui attenersi, come detto, è quello di minimizzare il fattore di picco, in modo da evitare la saturazione degli amplificatori.

Cenni sulla formazione delle forme d'onda in trasmissione

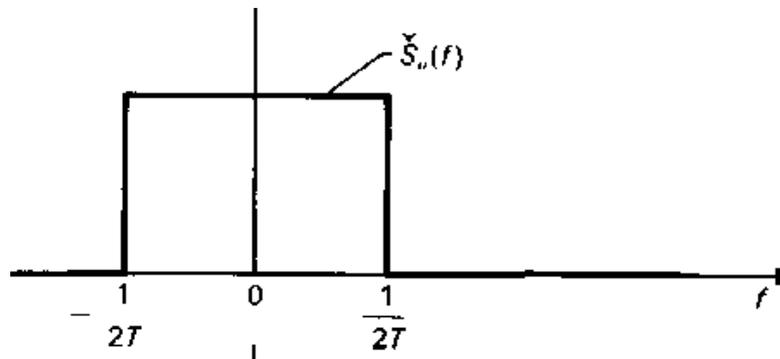
E' opportuno soffermarsi brevemente sul problema di come generare prefissate forme d'onda in trasmissione. Un metodo molto comodo è quello di utilizzare lo schema indicato nella figura seguente:



Le forme d'onda desiderate vengono prodotte da un apposito filtro (**filtro di forma**) comandato da opportuni impulsi. Se gli impulsi di comando fossero impulsi ideali, allora la funzione di trasferimento del filtro di forma sarebbe semplicemente uguale allo spettro $S_T(f)$ del segnale desiderato.

Nella pratica, gli impulsi non sono ideali, per cui hanno una certa durata, anche perché solo così essi possono avere una certa energia. Generalmente, si usano degli impulsi rettangolari, prodotti a loro volta da *dispositivi* funzionanti come interruttori comandati. A questo proposito, dato che, spesso, è importante generare le forme d'onda desiderate con una certa efficienza, bisogna tener presente che i suddetti *dispositivi* sono generalmente caratterizzati (e limitati) da una certa potenza di picco: di conseguenza, *conviene che la durata dell'impulso di comando sia convenientemente elevata, al fine di ottenere la massima energia possibile per il segnale $s_T(t)$.*

Per verificare questa affermazione, consideriamo ad esempio il caso in cui lo spettro $S_u(f)$ che si vuole ottenere in uscita dal filtro di ricezione è uno spettro costante (quindi $\delta=0$) in una banda $1/2T$, come indicato nella figura seguente:



Questo stesso spettro, per le considerazioni fatte in precedenza ipotizzando un mezzo trasmissivo ideale (cioè con attenuazione costante in banda) e rumore bianco, deve venir fuori dal trasmettitore.

Supponiamo adesso che il filtro di forma sia comandato da impulsi rettangolari di durata τ (con $\tau \leq T$, dove T è il periodo di cifra). Lo spettro di questi impulsi rettangolari è notoriamente

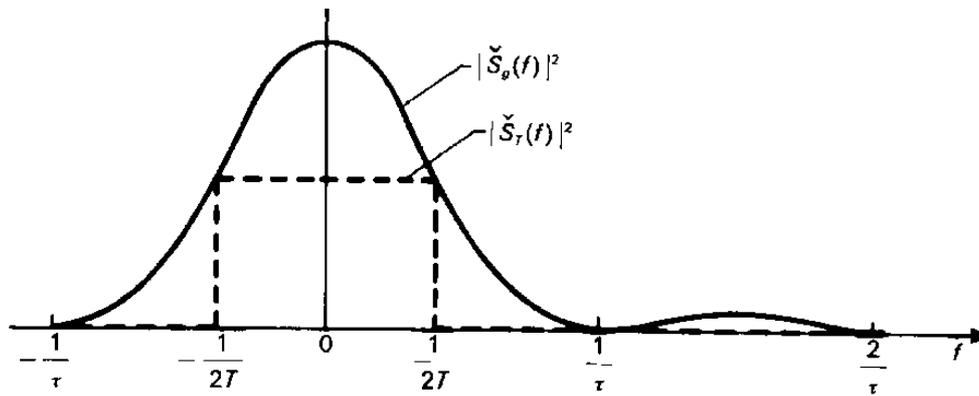
$$S_g(f) = a\tau \frac{\sin(\pi f\tau)}{(\pi f\tau)}$$

cui corrisponde uno spettro di energia nella forma

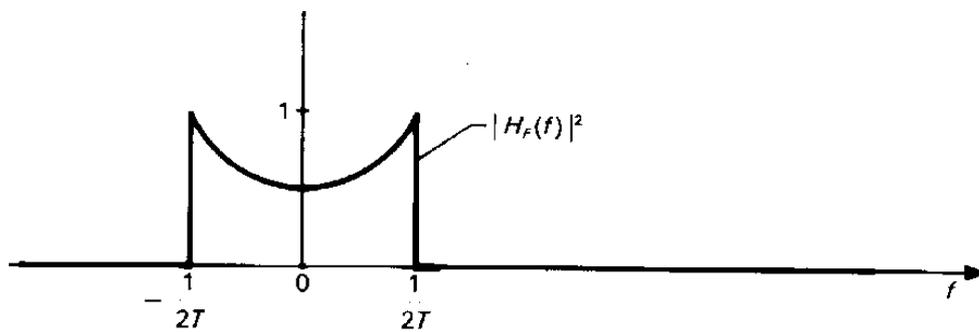
$$|S_g(f)|^2 = a^2 \tau^2 \frac{\sin^2(\pi f\tau)}{(\pi f\tau)^2}$$

avendo indicato con a l'ampiezza degli impulsi.

Questo spettro di energia è riportato a tratto pieno nella figura seguente, dove viene confrontato con lo spettro di energia $|S_T(f)|^2$ che si vuole in trasmissione (uguale allo spettro in uscita dal filtro di ricezione):



Se il filtro di forma deve trasformare $|S_g(f)|^2$ in $|S_u(f)|^2$, la sua funzione di trasferimento $H_F(f)$ dovrà avere quadrato del modulo del tipo indicato nella figura seguente:



Così facendo, infatti, il filtro attenua le componenti di $|S_g(f)|^2$ con energia esuberante e ottiene lo spettro $|S_u(f)|^2$ desiderato.

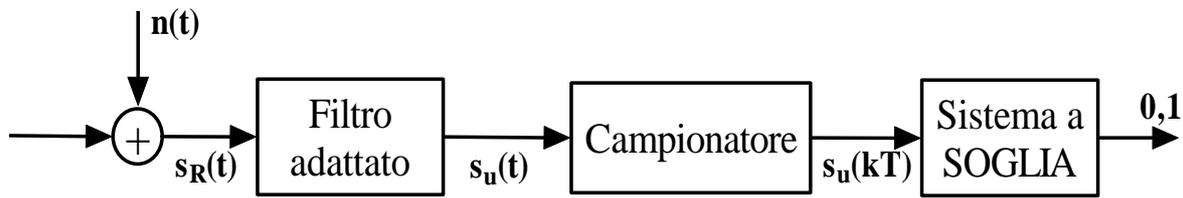
Possiamo a questo punto calcolare l'energia che esce dal filtro, cioè l'energia trasmessa:

$$E_T = \dots(?)\dots = a^2 \tau^2 \frac{\sin^2\left(\pi \frac{1}{2T} \tau\right)}{\left(\pi \frac{1}{2T} \tau\right)^2} \frac{1}{T} = a^2 T \frac{\sin^2\left(\pi \frac{1}{2T} \tau\right)}{\left(\frac{\pi}{2}\right)^2}$$

Questa energia è sempre crescente al crescere di τ fino a $\tau=T$, da cui troviamo conferma di quanto detto prima: ai fini dell'efficienza, è opportuno usare impulsi di comando della massima durata possibile (pari a T).

ESTRAZIONE DEL TIMING

Fino ad ora, abbiamo sempre supposto di avere la possibilità, dopo il filtro di ricezione, di campionare il segnale $s_u(t)$ nell'istante in cui esso è massimo:

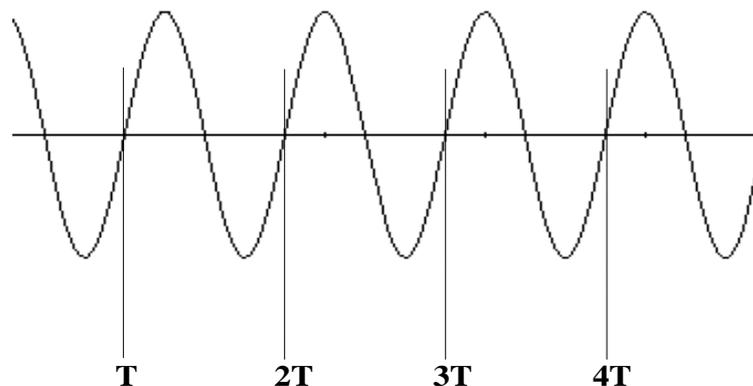


Gli istanti di campionamento, per le considerazioni dei paragrafi precedenti, sono a distanza T uno dall'altro, dove T è il **periodo di cifra** o anche periodo di ripetizione degli impulsi. La frequenza di campionamento, detta **frequenza di cifra**, è pari all'inverso del periodo di cifra:

$$f_c = \frac{1}{T}$$

Dobbiamo allora trovare il modo di far funzionare il campionatore con questa frequenza: se visualizziamo il campionatore come un semplice interruttore, dobbiamo fare in modo che esso si chiuda ogni T secondi.

Il modo più semplice è quello di utilizzare un oscillatore che oscilli alla frequenza f_c desiderata: in corrispondenza degli istanti in cui l'oscillazione va a zero e riparte con pendenza positiva, l'oscillatore deve inviare degli impulsi che pilotano l'interruttore:

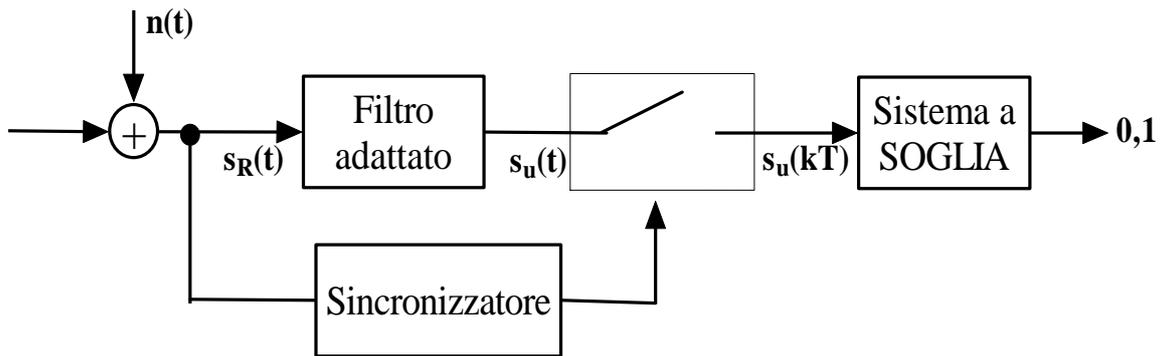


Si tratta però di un meccanismo fallimentare: infatti, bisogna tener presente che anche in trasmissione c'è un oscillatore, che serve a generare i vari impulsi; mentre, teoricamente, i due oscillatori (in trasmissione ed in ricezione), dovrebbero oscillare entrambi a f_c , in pratica essi oscilleranno ad una frequenza sicuramente diversa. Non solo, ma anche supponendo che i due oscillatori siano sincroni in frequenza, dovrebbero esserlo anche in fase, allo scopo di identificare, all'interno di ciascun periodo T , l'istante preciso di campionamento: infatti, se le sinusoidi in ricezione ed in trasmissione sono sfasate, gli istanti di campionamento saranno senz'altro diversi da quelli teorici ottimali.

Avere due oscillatori, uno in trasmissione e uno in ricezione, sincroni sia in frequenza sia in fase è irrealizzabile.

La soluzione più ragionevole è quella di disporre di un oscillatore autonomo in ricezione, il quale però estragga

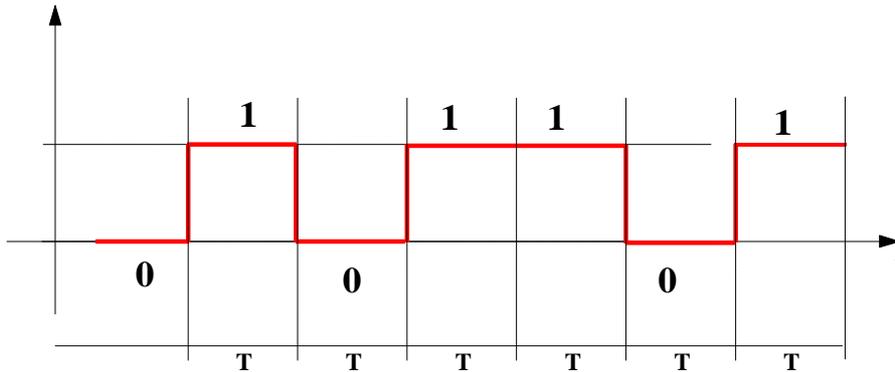
l'informazione di sincronizzazione direttamente dal segnale ricevuto:



Il blocco indicato in figura come "sincronizzatore" (o **estrattore del timing**) serve appunto a ricavare direttamente da $s_R(t)$ le informazioni necessarie alla sincronizzazione del campionatore. I prossimi discorsi sono dunque dedicati a capire come estrarre queste informazioni.

Codifica NRZ e codifica RZ

Supponiamo che le forme d'onda ricevute (praticamente uguali a quelle trasmesse, salvo l'inevitabile attenuazione) siano degli impulsi rettangolari. In particolare, supponiamo di adottare forme d'onda che, in corrispondenza del simbolo 1, mantengono il livello alto per tutto il periodo di cifra T:



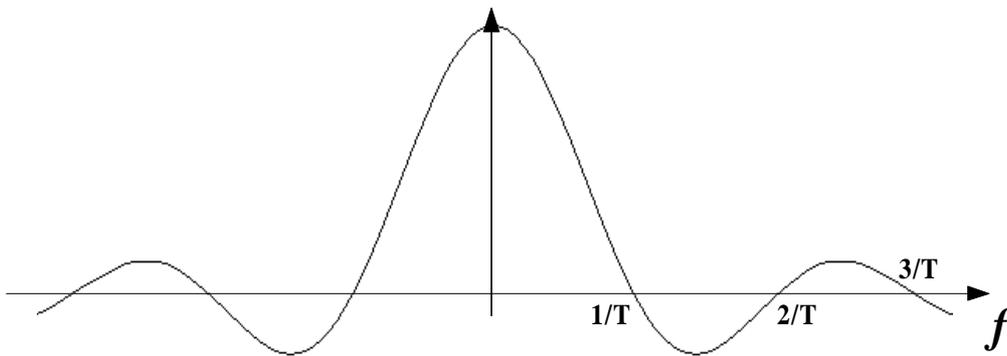
Una codifica di linea di questo tipo prende il nome di **codifica con non ritorno a zero** (brevemente *codifica NRZ*). In particolare, si tratta di una **codifica NRZ di tipo ortogonale**, visto che non viene trasmesso nulla per lo 0 e rettangoli di una certa ampiezza per gli 1.

Il vantaggio di una codifica NRZ è che, fissato il valore di valore di picco, l'energia della forma d'onda relativa ad un simbolo è la massima possibile (il che va a vantaggio del filtro adattato, la cui uscita è proporzionale all'energia della forma d'onda ricevuta).

In una forma d'onda come quella disegnata poco fa, dove è contenuta l'informazione di temporizzazione? E' evidentemente contenuta nelle transizioni dal livello basso al livello alto e viceversa, che avvengono tutte in istanti ben precisi, distanziati tra loro di T o multipli di T nel caso ci siano sequenze di più simboli uguali.

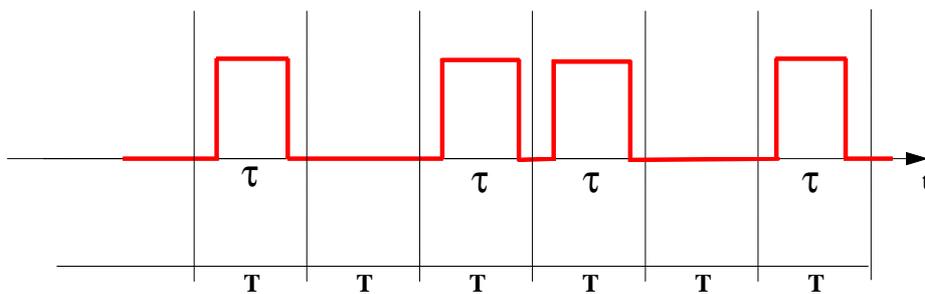
In termini di spettro, ci si può allora aspettare che lo spettro di quel segnale contenga una componente armonica a frequenza $1/T$, cioè alla frequenza di cifra f_c . Allora, per isolare questa sinusoide (al fine di utilizzarla per pilotare il campionatore) basterebbe mandare il segnale in ingresso ad un filtro passa-banda molto stretto centrato attorno a f_c .

In realtà, questo meccanismo non funziona, in quanto il segnale considerato non potrà mai avere una componente a frequenza $1/T$: infatti, lo spettro di ciascun rettangolo che compone il segnale, preso a se stante, è a forma di Seno Cardinale, per cui, se T è la durata del rettangolo, il suddetto spettro risulta nullo proprio in corrispondenza della frequenza $1/T$ e dei suoi multipli.

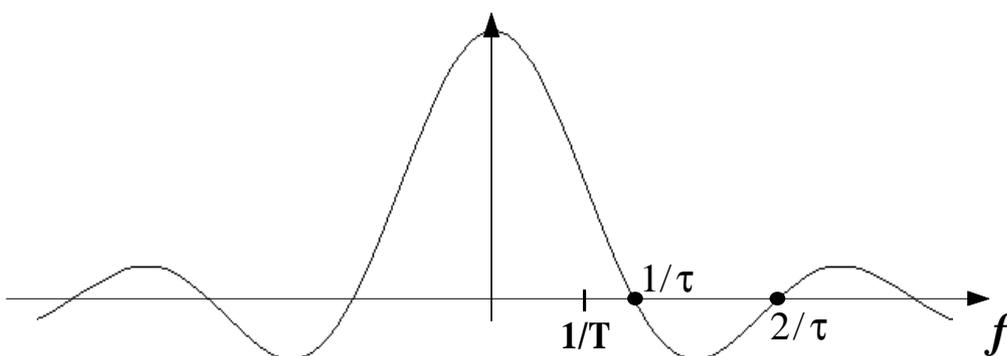


Di conseguenza, quello che il filtro passa-banda tira fuori sono le componenti spettrali adiacenti alla frequenza $1/T$, dalle quali non si può ricavare alcuna informazione.

Per ovviare a questo inconveniente, bisognerebbe fare in modo che lo spettro della singola forma d'onda non si annulli in corrispondenza di $1/T$ e questo lo si può ottenere, nel modo più semplice possibile, usando forme d'onda rettangolari di durata τ inferiore a T :



Così facendo, lo spettro di ciascun rettangolo si annulla in $1/\tau$ e suoi multipli, mentre invece non si annulla in $1/T$ (se $\tau < T$, allora $1/\tau > 1/T$):



Naturalmente, questo vale per lo spettro del singolo rettangolo. Lo spettro della forma d'onda complessiva è la composizione dei singoli spettri, per cui la componente spettrale a $1/T$ può esserci oppure no, a seconda di come si combinano i vari spettri, che corrispondono a forme d'onda uguali ma traslate temporalmente le une rispetto alle altre. Analizziamo allora la cosa da un punto di vista analitico.

Consideriamo una forma d'onda $s(t)$ rettangolare, di durata τ , centrata nell'istante $t=0$: il corrispondente spettro ha notoriamente espressione

$$s(t) \xrightarrow{\text{Fourier}} S(f) = \frac{\sin(\pi f \tau)}{\pi f \tau}$$

Traslando il rettangolo di nT periodi (con n generico), il corrispondente spettro, introducendo l'operatore ritardo, è

$$s(t - nT) \xrightarrow{\text{Fourier}} S(f)e^{-j\omega nT} = \frac{\sin(\pi f \tau)}{\pi f \tau} e^{-j\omega nT}$$

Alla frequenza $1/T$, questo spettro ha ampiezza

$$p_n = [S(f)e^{-j2\pi f nT}]_{f=1/T} = T \frac{\sin\left(\pi \frac{1}{T} \tau\right)}{\pi \tau} e^{-j2\pi \frac{1}{T} nT} = T \frac{\sin\left(\pi \frac{1}{T} \tau\right)}{\pi \tau} e^{-j2\pi n} = T \frac{\sin\left(\pi \frac{1}{T} \tau\right)}{\pi \tau} \neq 0$$

dove abbiamo tenuto conto del fatto che $e^{-j2\pi n} = 1$ (basta sviluppare l'esponenziale).

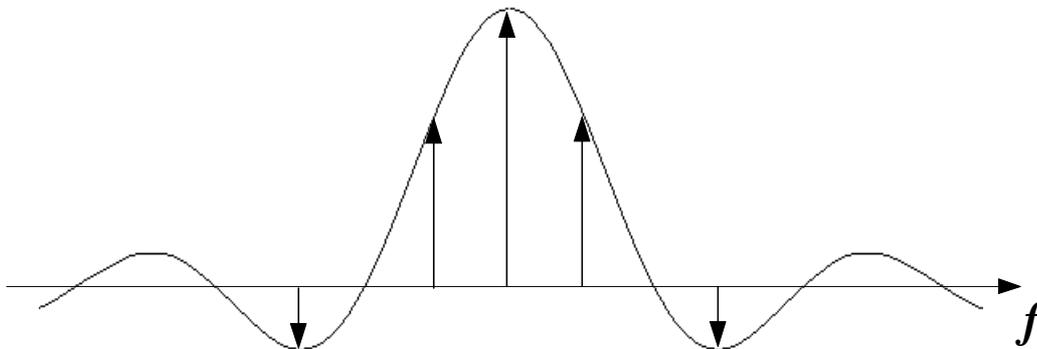
Quindi, la generica forma d'onda ritardata di nT ha uno spettro la cui ampiezza, a frequenza $1/T$, è data dal coefficiente p_n . Quindi, ogni spettro contribuisce con un proprio peso p_n all'ampiezza della componente ad $1/T$, come detto prima.

Allora, se supponiamo di prendere la singola forma d'onda $s(t)$ e di ripeterla periodicamente nel tempo, otteniamo il seguente segnale $x(t)$:

$$x(t) = \sum s(t - nT) \xrightarrow{\text{Fourier}} X(f) = \sum S(f)e^{-j\omega nT} = \sum \frac{\sin(\pi f \tau)}{\pi f \tau} e^{-j\omega nT}$$

Questo segnale, essendo periodico, ha spettro $X(f)$ a righe (si tratta cioè di una successione di impulsi). In particolare, per quanto trovato poco fa, se $S(f)$ non è nullo a frequenza $1/T$ ed ai suoi multipli, anche lo spettro $X(f)$ presenta le componenti a frequenza $1/T$ ed ai suoi multipli. In questo caso, quindi, sarebbe possibile isolare la riga a frequenza $1/T$.

Ovviamente, il nostro segnale non è perfettamente periodico, in quanto, in generale, la sequenza di 1 e 0 è varia. Di conseguenza, il suo spettro non potrà essere solo a righe, pur essendo queste presenti nello spettro, ma dovrà necessariamente presentare anche una parte continua:



Si ha un impulso a frequenza nulla in quanto il valor medio del segnale non è zero (ricordiamo che il valore medio del segnale corrispondente all'ampiezza della componente armonica a frequenza nulla, ossia all'ampiezza della continua). Si hanno poi delle righe di una certa ampiezza (da

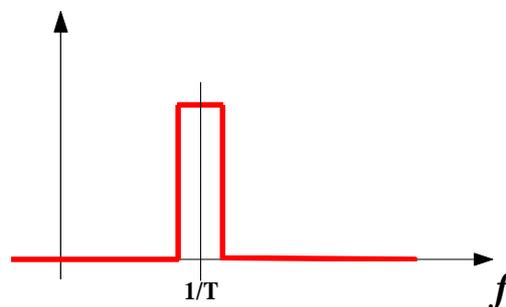
determinare) a frequenza $1/T$ e alle sue *armoniche* (cioè ai suoi multipli). Infine, sovrapposta a queste righe si ha una parte continua di spettro dovuta appunto al fatto che non c'è regolarità nella ripetizione delle forme d'onda (i fronti di salita e i fronti di discesa non sono equidistanti tra di loro).

Tutta questa codifica di linea prende il nome di **codifica con ritorno a zero** (brevemente codifica RZ), proprio per indicare il fatto che, in ciascun intervallo T , il livello alto non viene mantenuto per tutto T ma solo per $\tau < T$. Essa presenta vantaggi e svantaggi:

- il vantaggio, come visto, è la presenza della riga a frequenza $1/T$: usando un filtro passa-banda molto stretto centrato attorno a questa frequenza, si ottiene in uscita una forma d'onda abbastanza simile ad una sinusoidale a frequenza $1/T$ (sarà una sinusoidale con sovrapposte delle ulteriori fluttuazioni sinusoidali, simili a quelle dovute al rumore);
- uno svantaggio è invece relativo all'energia: infatti, ogni forma d'onda dura $\tau < T$, per cui, a parità di valore di picco, porta con sé meno energia di quanta ne porterebbe se durasse proprio T (come nel caso della codifica NRZ);
- un altro svantaggio è relativo alla banda: infatti, la banda occupata da un impulso di durata τ è sicuramente maggiore di quella occupata da un impulso di durata $T > \tau$.

Questi due svantaggi sconsigliano l'uso di forme d'onda del tipo appena descritto, a meno che non si possa fare altrimenti.

Facciamo adesso qualche osservazione a proposito dell'operazione di filtraggio passa-banda: abbiamo detto che l'operazione di filtraggio serve per estrarre il **timing** (cioè gli istanti di campionamento) sganciandosi dai fastidi generati dal rumore sovrapposto. Questo è il motivo per cui è importante usare un filtro molto stretto attorno alla frequenza $1/T$:



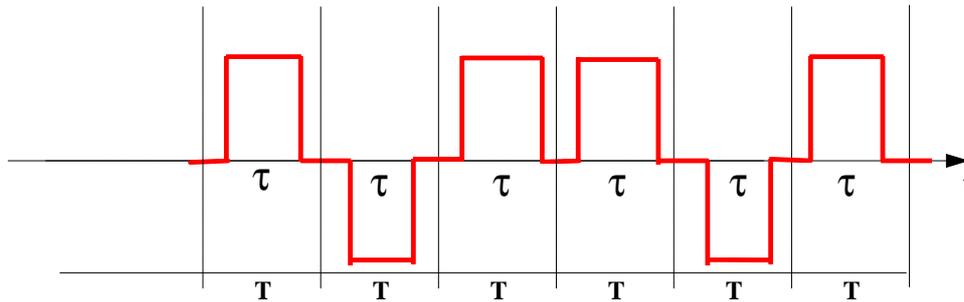
Oltre al rumore, ci sono da considerare le componenti spettrali del segnale adiacenti alla frequenza $1/T$: tali componenti rappresentano il comportamento pseudo-casuale del segnale, dipendente dalla successione di 1 e 0 che è stata trasmessa. *L'effetto di queste componenti crea, sulla componente di interesse a $1/T$, lo stesso "fastidio" di un rumore sovrapposto ad una portante modulata in FM*: infatti, sappiamo che il rumore passa-banda può sempre essere rappresentato come una sinusoidale, isofrequenziale con la portante, modulato in ampiezza e in fase in modo casuale. Questo significa che la fase della portante risulta sporcata dal rumore e quindi che gli istanti di campionamento (che coincidono con gli istanti di attraversamento dello zero) sono spostati nel tempo.

Ancora una volta, quindi, deduciamo l'importanza di avere un filtro molto stretto, in modo da filtrare quanto più è possibile non solo il rumore, ma anche la parte continua dello spettro del segnale.

Ovviamente, per avere un filtro passa-banda molto stretto bisogna realizzare un sistema con una risposta all'impulso $h(t)$ (pari all'antitrasformata della funzione di trasferimento) che si smorzi molto lentamente nel tempo.

Codifica RZ di tipo antipodale

Supponiamo adesso di adottare una **codifica RZ di tipo antipodale**, per cui associamo un impulso positivo per 1 ed un impulso negativo, di uguale ampiezza, per lo 0:

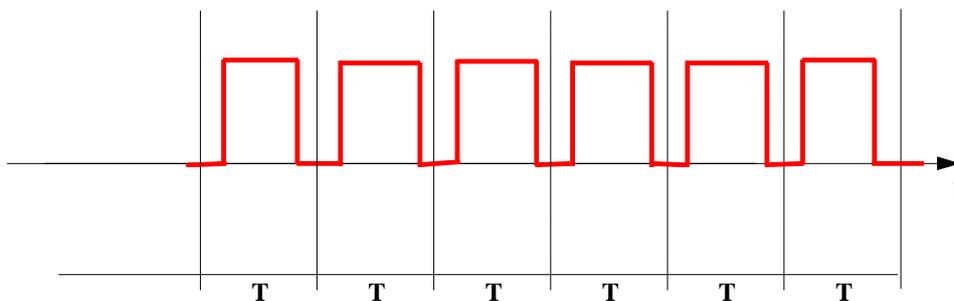


In questo caso, supponendo (come avviene nella realtà) che i simboli 1 e 0 siano equiprobabili, *il fatto che la singola forma d'onda non vada a zero a frequenza 1/T non è assolutamente sufficiente a garantire la presenza della riga a frequenza 1/T del segnale complessivo. Vediamo perché.*

Sia $S(f)$ lo spettro del singolo impulso positivo: supponiamo che esso sia non nullo a frequenza $1/T$, per cui $S(1/T) \neq 0$. Lo spettro del singolo impulso negativo sarà $-S(f)$ e anch'esso sarà ovviamente non nullo a frequenza $1/T$. Allora, se i simboli sono equiprobabili, si ha, in media, la stessa probabilità di avere $S(1/T)$ oppure $-S(1/T)$. Di conseguenza, effettuando la media temporale (operazione che, in frequenza, corrisponde al filtraggio passa-banda), non si ottiene nulla in uscita, in quanto la riga dello spettro viene cancellata quando si effettua la media sui simboli.

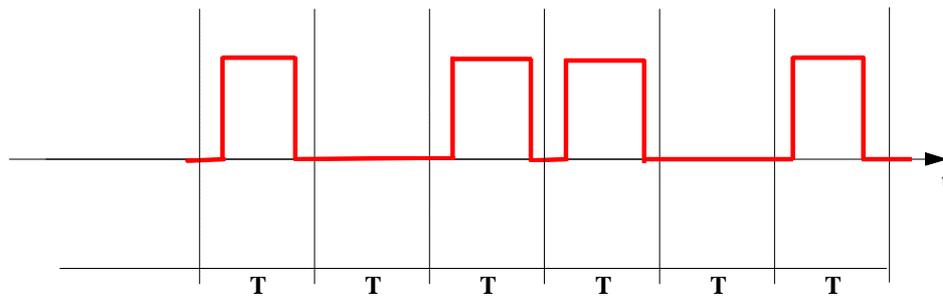
Quindi, *la codifica RZ di tipo antipodale presenta il problema per cui, con pari probabilità, le forme d'onda possono essere positive o negative.*

Si può, d'altra parte, ovviare molto facilmente a questo problema. Considerata la forma d'onda ricevuta, la si raddrizza a doppia semionda, in modo che gli impulsi negativi diventino positivi:



Così facendo, otteniamo un segnale perfettamente periodico, che quindi presenta sicuramente la riga a $1/T$. Non solo, ma questa riga non presenta altre componenti di segnale adiacenti, proprio perché l'intero spettro è a righe, per cui l'esito del filtraggio passa-banda fornisce una sinusoide abbastanza pulita. Ovviamente, non sarà perfettamente pulita in quanto c'è da considerare sempre il contributo del rumore.

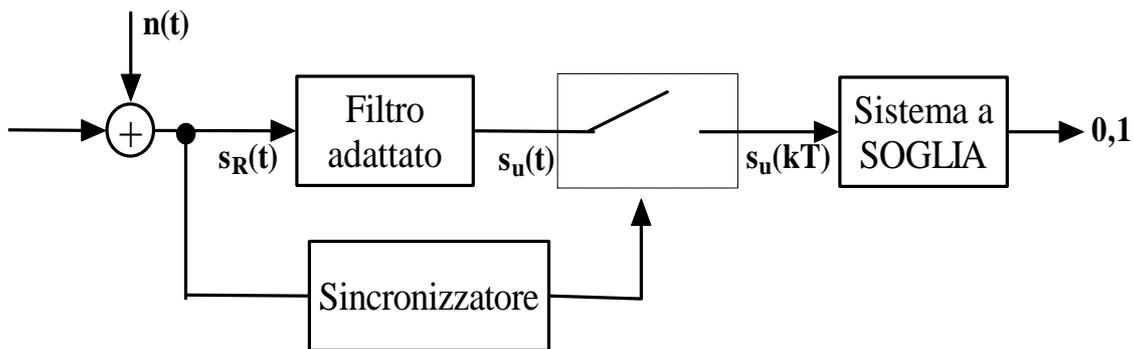
In effetti, anziché raddrizzare a doppia semionda, potremmo semplicemente raddrizzare a singola semionda, il che equivale semplicemente ad azzerare gli impulsi negativi:



In questo caso, si presentano al filtro passa-banda solo gli impulsi positivi, il che equivale a presentare uno spettro in parte continuo e in parte a righe, con la riga a $1/T$ sicuramente presente. E' chiaro che il raddrizzamento a doppia semionda è preferibile, ma, in alternativa, potrebbe andare bene anche il raddrizzamento a singola semionda.

Schema del dispositivo estrattore del timing

A questo punto, ci chiediamo con quale schema di principio si possa ricavare, dal segnale ricevuto, un segnale di comando per il campionatore. Dobbiamo cioè sintetizzare il blocco che nella figura seguente è indicato come *sincronizzatore (estrattore del timing)*:



In base a quanto detto nel paragrafo precedente, il primo passo è quello di raddrizzare (a doppia semionda o, in alternativa, a singola semionda) il segnale ricevuto $s_R(t)$. Il segnale raddrizzato presenta sicuramente la componente a frequenza $1/T$, più o meno sporcata da eventuali componenti spettrali adiacenti del segnale stesso e dall'immane rumore.

Effettuando, quindi, un successivo filtraggio passa-banda molto stretto, si ottiene un segnale molto prossimo ad una sinusoida a frequenza $1/T$. Questa sinusoida è tale che gli istanti corrispondenti agli attraversamenti dello zero con pendenza positiva coincidano con gli istanti di campionamento. Allora, mandando in ingresso la sinusoida ad un trigger, questo fornisce in uscita degli impulsi molto stretti in corrispondenza proprio degli istanti di campionamento.

Questi impulsi (che andranno a comandare il campionatore) saranno dunque perfettamente sincroni con la frequenza di cifra f_s e saranno anche in relazione di fase fissa rispetto alle forme d'onda ricevute e quindi anche alle forme d'onda in uscita dal filtro di ricezione (cioè le forme d'onda da campionare).

Lo schema di principio del *sincronizzatore* è dunque il seguente:



In realtà, c'è un problema da affrontare: non è garantito che la suddetta relazione di fase sia tale da far combaciare perfettamente gli impulsi di campionamento con i massimi delle forme d'onda in uscita dal filtro. Il motivo è che il segnale ed i sincronismi percorrono strade diverse, per cui è *necessario prevedere necessariamente uno sfasamento temporale tra le forme d'onda di comando e quelle in uscita dal filtro di ricezione.*

In linea teorica, si potrebbe pensare di mandare gli impulsi del trigger in un *ritardatore temporale*, regolabile dall'operatore: questo però significa applicare ad una forma d'onda a banda larga (quanto più gli impulsi sono stretti nel tempo, tanto maggiore è la banda occupata in frequenza) un operatore che richiede una ampiezza della funzione di trasferimento costante ed una fase rigorosamente rettilinea.

Al contrario, è molto più semplice applicare un ritardo variabile ad una sinusoide, come quella in uscita dal filtro passa-banda, in quanto, in questo caso, diventa importante soltanto la caratteristica del ritardatore alla frequenza della sinusoide stessa⁹.

In definitiva, quindi, per spostare la posizione relativa tra gli impulsi di comando e le forme d'onda che escono dal filtro di ricezione, si inserisce uno *sfasatore variabile* in cascata al filtro passa-banda (dal quale esce una sinusoide) e prima del trigger:

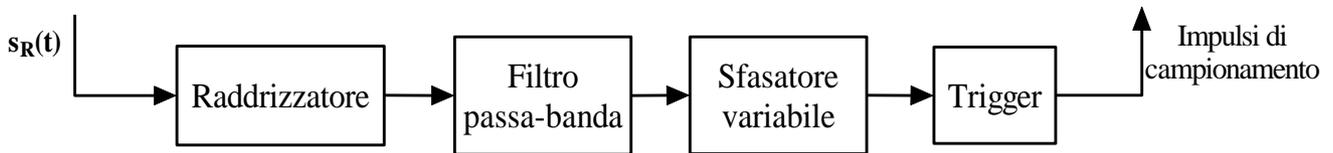


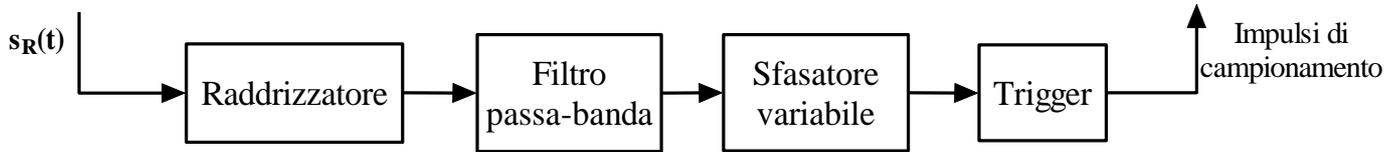
Diagramma ad occhio

A questo punto, si tratta di regolare lo sfasamento in modo che il campionamento avvenga sempre negli istanti di massimo. Si usa a tale scopo il cosiddetto **diagramma ad occhio**.....

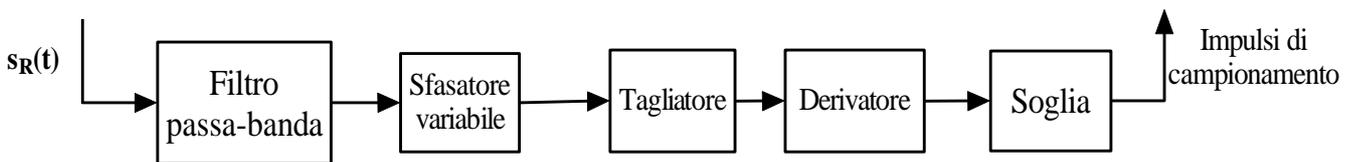
⁹ Ricordiamo che una sinusoide è l'unico segnale per il quale uno sfasamento equivale ad una traslazione temporale: per questo motivo, basta conoscere lo sfasamento introdotto dal ritardatore alla frequenza considerato per valutare la corrispondente traslazione nel tempo.

Generazione degli impulsi di campionamento (schema alternativo)

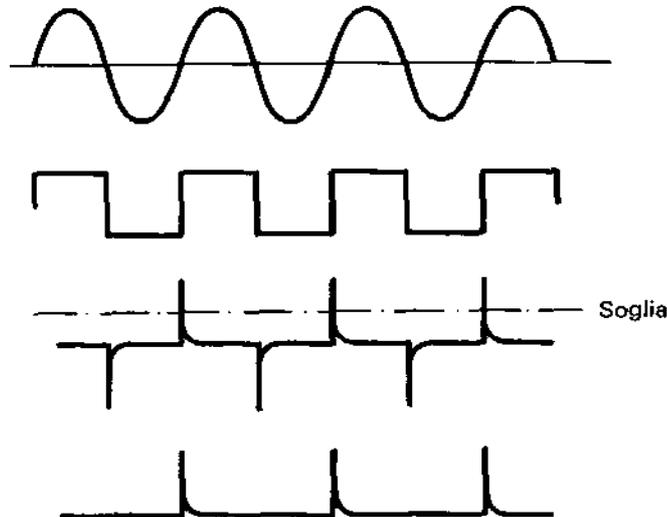
Riprendiamo lo schema precedentemente tracciato per ottenere gli impulsi di campionamento:



Questo schema vale solo quando il singolo segnale ricevuto contiene una riga alla frequenza di cifra $1/T$. Uno schema alternativo, nell'ipotesi che tutto il segnale contenga la suddetta riga a $1/T$, potrebbe anche essere il seguente:



In pratica, al posto del trigger, abbiamo usato una cascata di 3 dispositivi, il cui funzionamento è illustrato nella figura seguente, dove vengono riportati gli andamenti temporali dei vari segnali coinvolti:



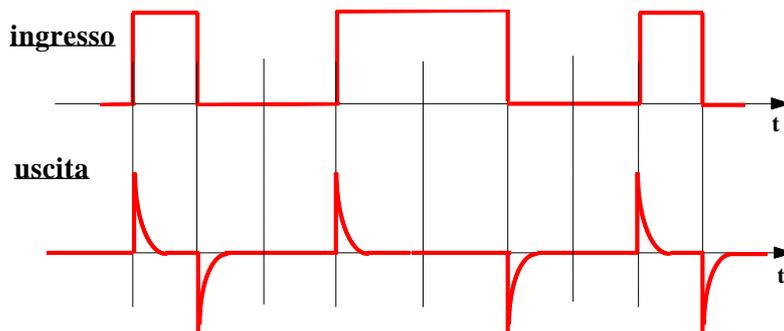
Come già abbiamo visto, all'uscita dal *filtro passa-banda* si ottiene un segnale molto prossimo ad una sinusoide a frequenza $1/T$. Questa sinusoide viene mandata in ingresso ad un *tagliatore* (che, ad esempio, è uno stadio amplificatore dove i transistor vanno in saturazione), il quale squadra semplicemente la sinusoide. L'onda quadra così ottenuta viene inviata in ingresso ad un *derivatore*, che genera degli impulsi molto stretti in corrispondenza dei fronti di salita e di discesa dell'onda quadra stessa. Si ottiene così un segnale ad impulsi sia positivi sia negativi, cioè un segnale antipodale (nel senso che per ogni forma d'onda positiva ce n'è una negativa). D'altra parte, solo gli impulsi positivi corrispondono agli istanti di campionamento (in quanto corrispondono agli attraversamenti dello zero, da parte della sinusoide, con pendenza positiva), per cui vanno azzerati

gli impulsi negativi, ossia serve un raddrizzamento a singola semionda, che in questo caso può essere fatto semplicemente con un *sistema a soglia*.

Nel caso in cui il singolo segnale ricevuto (associato al singolo simbolo) non contenga la riga a $1/T$, è evidente un fatto: volendo ottenere comunque, da questo segnale, una sinusoide a frequenza $1/T$, è necessario adottare un operatore non lineare¹⁰. Un possibile modo di procedere, in questo caso, è quello rappresentato dello schema seguente, in cui l'elemento non lineare è il **raddrizzatore a doppia semionda**:

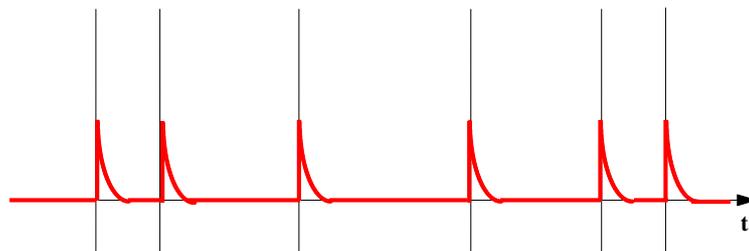


Abbiamo già avuto modo di dire che le informazioni circa il sincronismo sono contenute nelle transizioni del segnale $s_R(t)$ dal livello alto al livello basso e viceversa. Il modo più semplice di tener conto di queste transizioni è quello di inviare il segnale direttamente in ingresso ad un *derivatore*, il quale, così come visto prima, genera degli impulsi molto stretti in corrispondenza delle transizioni:



(si è supposto, in questa figura, che il segnale ricevuto avesse una codifica NRZ di tipo ortogonale).

Si ottiene così nuovamente una forma d'onda RZ di tipo antipodale. Raddrizzando, allora, gli impulsi negativi, si ottiene il seguente segnale:



L'effetto del raddrizzamento a doppia semionda è quello di ottenere un segnale maggiormente regolare di quello in uscita dal derivatore. Tale segnale, quindi, pur non essendo perfettamente periodico, si presta ad un successivo filtraggio passa-banda per l'estrazione della sinusoide a frequenza $1/T$. Dopo di che, lo schema procede come nel caso precedente.

¹⁰ Infatti, se applicassimo un operatore lineare ad un segnale contenente un certo numero di componenti spettrali, l'uscita avrebbe le stesse componenti spettrali e non potrebbe mai averne di nuove. Viceversa, se usassimo, ad esempio, un operatore quadratico, comparirebbero in uscita sia le componenti presenti in ingresso sia i noti prodotti di intermodulazione (si ricordi la questione della distorsione armonica).

Sollecitazioni sul filtro passa-banda e corrispondenti codifiche di linea

Consideriamo nuovamente il caso in cui il segnale ricevuto $s_R(t)$ non contiene la riga a frequenza $1/T$, per cui tale riga, in base a quanto visto poco fa, può essere ottenuta derivando, raddrizzando e filtrando il segnale stesso. Per situazioni come questa, c'è da fare una importante considerazione: supponiamo che la sequenza di simboli che è stata trasmessa presenti un discreto numero di simboli consecutivi uguali, come ad esempio

1 0 0 0 0 0 0 1 0

La presenza di k zeri (oppure k uno) consecutivi fa sì che la forma d'onda ricevuta non presenti transizioni per un intervallo di tempo pari ad kT , il che comporta, dopo le operazioni di derivazione e di raddrizzamento, la mancanza di impulsi per lo stesso intervallo di tempo. Questo significa che il segnale in ingresso al filtro passa-banda è nullo per kT secondi, per cui il filtro, in pratica, non viene sollecitato durante tale intervallo di tempo. Ne deriva che l'oscillazione prodotta dal filtro tende sia ad attenuarsi sia a cambiare la propria fase¹¹. Questa situazione è inaccettabile, in quanto abbiamo visto quanto importante sia la precisione degli istanti di campionamento. Si pone perciò il problema di evitare che il filtro rimanga privo di sollecitazioni per troppo tempo.

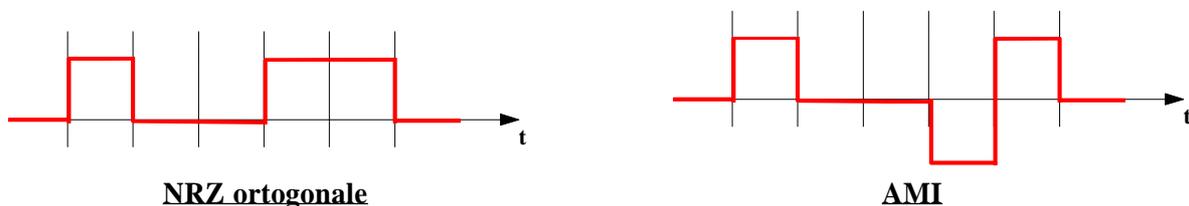
L'unica possibilità è quella di modificare in modo opportuno la sequenza di simboli in trasmissione: da un lato, non bisogna ovviamente alterare il contenuto informativo della sequenza, ma, dall'altro, bisogna evitare sequenze troppo lunghe di simboli uguali consecutivi. A questo scopo sono dedicate alcune particolari **codifiche di linea** illustrate nei paragrafi che seguono.

Codice AMI

L'acronimo **AMI** sta per *Alternated Marked Inversion*. Il concetto su cui si basa questa codifica è duplice:

- in primo luogo, si trasmette una determinata forma d'onda (di durata T) per il simbolo 1 e non si trasmette niente per lo 0: in questo senso, si tratta di una *codifica NRZ di tipo ortogonale*;
- la differenza, però, con una codifica ortogonale è che la forma d'onda relativa ad 1 viene trasmessa alternativamente positiva e negativa.

Per capire il meccanismo, consideriamo la sequenza 1 0 0 1 1. Nella figura seguente sono illustrate la codifica NRZ di tipo ortogonale e la codifica AMI:



Dopo il primo 1, cui corrisponde un rettangolo positivo, mentre la codifica ortogonale prevede anche per il successivo 1 un rettangolo positivo, la codifica AMI prevede un rettangolo negativo, seguito poi da uno positivo per l'1 successivo.

¹¹ Il principio è generale: fin quando ci sono delle sollecitazioni regolari e frequenti, il filtro passa-banda fornisce una sinusoide che mantiene la coerenza con la sequenza di ingresso; se, però, il filtro viene lasciato oscillare senza sollecitazioni, la fase varia.

E' chiaro che una scelta di questo tipo aumenta la complessità, ma è anche evidente che, in presenza di una sequenza di simboli 1 consecutivi, sono sempre presenti delle discontinuità a distanza T, per cui il filtro passa-banda è continuamente sollecitato.

Oltre a semplificare la temporizzazione, questo codice ha un altro grande pregio: infatti, esso ha valor medio nullo, il che significa che il suo spettro non contiene la riga a frequenza 0; questo significa che possiamo tranquillamente trasmettere, insieme al segnale vero e proprio, anche l'alimentazione¹², cioè una componente continua di ampiezza opportuna. In ricezione, proprio perché la riga a frequenza 0 corrisponde solo all'alimentazione, quest'ultima può essere facilmente separata dal segnale mediante un trasformatore di disaccoppiamento.

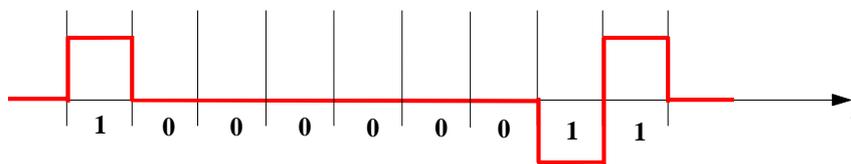
Codice HDB-3

Il codice AMI, mentre affronta con successo il caso in cui ci sono più 1 consecutivi nella sequenza binaria trasmessa, non affronta il caso, del tutto equivalente, in cui si sono più 0 consecutivi. Questo caso è invece affrontato da una evoluzione del codice AMI, detta **codice HDB-3** (che sta per *High Density Binary Code*): esso si pone, come obiettivo, quello di evitare che per più di 3 periodi di cifra successivi non arrivino impulsi al filtro passa-basso. In altri termini, si evita che ci siano, nella sequenza binaria trasmessa, più di tre 0 consecutivi.

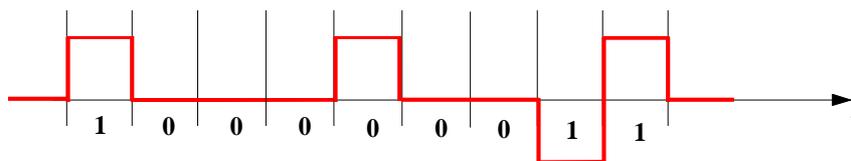
Facciamo un esempio concreto, supponendo che la sequenza binaria da trasmettere sia la seguente:

1 0 0 0 0 0 0 1 1

Se adottassimo semplicemente la codifica AMI, avremmo la seguente forma d'onda trasmessa:



Con questa forma d'onda, il filtro passa-banda non riceve impulsi per 6 periodi di cifra, il che, come già detto in precedenza, può provocare una inaccettabile variazione della fase (oltre che dell'ampiezza) della sinusoide in uscita. Allora, il codice HDB-3 è tale da inviare un impulso dopo tre 0 consecutivi. C'è ovviamente da decidere la polarità (positivo o negativo) dell'impulso: se usassimo un impulso negativo, esso rispetterebbe la regola dell'alternanza di segno prevista dal codice AMI, per cui il ricevitore interpreterebbe l'impulso come un 1. Di conseguenza, per far capire al ricevitore che l'impulso è *fittizio*, bisogna mandarlo positivo, così da ottenere la seguente forma d'onda:



¹² Si parla, in questo caso, di **telealimentazione**.

In definitiva, quindi, la polarità degli **impulsi fittizi** (corrispondenti cioè ai simboli 0) va scelta in modo che tali impulsi violino la regola dell'alternanza di segno, in modo che il ricevitore possa interpretarli appunto come fittizi e non come rappresentativi del simbolo 1.

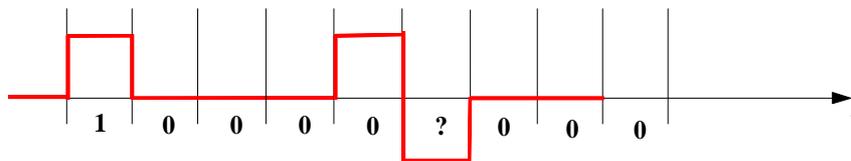
Ovviamente, perché un meccanismo di questo tipo funzioni, il sistema deve avere memoria degli ultimi 4 simboli ricevuti: infatti, quando riceve l'impulso fittizio, per decidere se esso sia effettivamente fittizio o meno deve necessariamente confrontarne il segno con l'impulso precedente, che è arrivato 4 simboli prima: se il segno è concorde, l'impulso è fittizio, altrimenti l'impulso corrisponde sicuramente ad un 1.

Anche questo modo di procedere ha però un inconveniente, che è quello di variare localmente il valor medio nel caso di una sequenza molto lunga di 0. Questo impedisce la trasmissione dell'alimentazione, in base alle considerazioni fatte prima, per cui bisogna perfezionare ulteriormente il procedimento.

Consideriamo, per esempio, la sequenza seguente:

1 0 0 0 0 0 0 0 0 1

Abbiamo otto 0 consecutivi, per cui è necessario inserire degli impulsi fittizi. Il primo impulso va inviato dopo tre 0 consecutivi e deve violare l'alternanza di segno con l'impulso precedente, per cui siamo nella stessa condizione del caso precedente. Al fine, però, di lasciare invariato il valore medio locale, deve essere trasmesso subito dopo un ulteriore impulso, sempre fittizio, di polarità opposta al precedente. La forma d'onda diventa perciò la seguente:

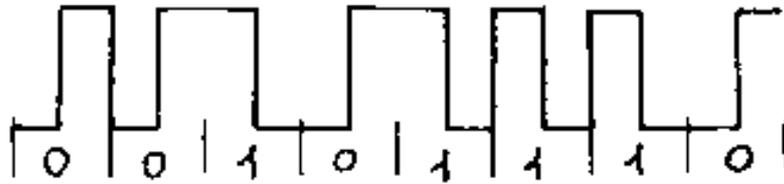


A questo punto, però, si pone un problema: dato che il secondo impulso fittizio rispetta l'alternanza di segno con l'ultimo impulso corrispondente ad 1, c'è una ambiguità, nel senso che il ricevitore non è in grado di decidere se esso sia fittizio (corrispondente perciò ad uno 0) o meno (corrispondente ad un 1). L'unico modo per decidere è quello di aspettare l'impulso successivo: se questo ulteriore impulso è negativo, allora anche il 6° simbolo è 0, mentre, in caso contrario, si dovrà necessariamente optare per un 1.

Codice Manchester

Un altro importante codice di linea è il cosiddetto **codice Manchester**, che è il codice RZ più utilizzato. Essendo un codice RZ, esso prevede che, in ciascun periodo di cifra T , il livello alto venga mantenuto per un tempo $\tau < T$. In particolare, in questo caso il livello alto è mantenuto per metà del periodo e il simbolo corrispondente viene individuato a seconda che la transizione del segnale, in corrispondenza di $T/2$, sia in salita o in discesa: se la transizione è in salita, allora il simbolo trasmesso è 0, mentre invece, se la transizione è in discesa, allora il simbolo trasmesso è 1.

La figura seguente mostra un esempio di codifica Manchester:



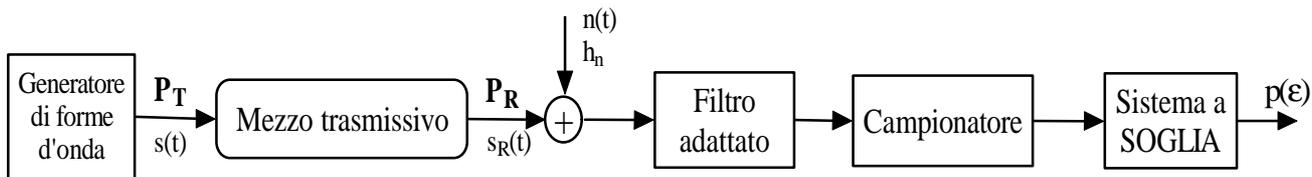
Il pregio di questa codifica è evidente: almeno ogni mezzo periodo T è presente una transizione, che quindi può essere usata per la sincronizzazione. Derivando il segnale, si ottengono impulsi distanziati al più di T .

Concludiamo ricordando che, nella realtà, nonostante le codifiche impiegate, gli istanti di campionamento non potranno essere mai quelli desiderati: sarà sempre presente, quindi, un jitter del quale bisogna tener conto quando si valutano le prestazioni del sistema.

CRITERIO GENERALE DI DIMENSIONAMENTO DEL SISTEMA

Così come nei paragrafi precedenti, effettueremo il dimensionamento di un sistema di trasmissione numerico supponendo che il rumore sovrapposto al segnale sia solo rumore termico, cioè rumore bianco con distribuzione gaussiana delle ampiezze.

Il nostro obiettivo è quello di determinare la potenza P_T da trasmettere per ottenere un sistema di trasmissione con le prestazioni desiderate.



Quando ci siamo occupati del dimensionamento di **sistemi di trasmissione analogici**, abbiamo determinato la suddetta potenza da trasmettere sulla base di due vincoli di progetto:

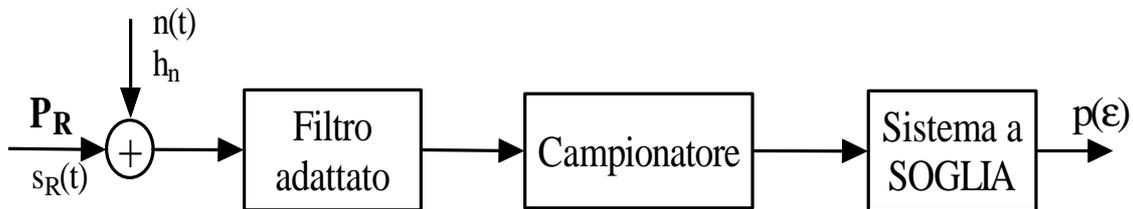
- in primo luogo, la banda passante B , ossia la banda con cui il segnale considerato può essere trasmesso;
- in secondo luogo, la qualità con la quale questo segnale deve essere ricevuto: essa è quantificata dal rapporto segnale-rumore in ingresso al ricevitore cui è destinato il segnale.

Nel caso di un sistema di trasmissione numerico, ci sono due specifiche del tutto equivalenti a quelle appena citate e valide per il caso analogico:

- in primo luogo, la velocità di trasmissione dei bit, quantificata dalla **frequenza di cifra f_S** , ossia dal numero di simboli che giunge al ricevitore nell'unità di tempo; la frequenza di cifra è dunque l'analogo della banda B nel sistema analogico;
- in secondo luogo, anche in questo caso la qualità con cui l'informazione giunge al ricevitore, quantificata però dalla **probabilità di errore $p(\epsilon)$** , intesa come la probabilità che il generico

bit ricevuto risulta diverso da quello trasmesso; $p(\epsilon)$ è dunque l'analogo del rapporto S/N nel sistema analogico¹³.

Come si deduce dalla figura precedente o anche da quella seguente (in cui è riportato solo l'apparato di ricezione, con l'esclusione dei dispositivi per l'estrazione del timing, che in fase di dimensionamento non hanno alcuna importanza), la probabilità di errore $p(\epsilon)$ fa riferimento al punto terminale del sistema in trasmissione, ossia al punto in cui tiriamo fuori nuovamente i bit:



Si tratta, allora, di andare a ritroso, determinando la potenza P_T da trasmettere affinché $p(\epsilon)$ non superi il valore massimo tollerabile. Questo valore massimo tollerabile è dunque la prima specifica da cui partire nel dimensionamento di un sistema di questo tipo.

Rapporto S/N a valle del campionatore

In ricezione, le possibilità sono solo due:

- se è stato trasmesso il simbolo logico 0 e si è scelto di associare a tale simbolo una forma d'onda $s_0(t)$, il segnale ricevuto è

$$s_0(t) + n(t)$$

- se è stato trasmesso il simbolo logico 1 e si è scelto di associare a tale simbolo una forma d'onda $s_1(t)$, il segnale ricevuto è

$$s_1(t) + n(t)$$

Questo è quello che avviene a monte del filtro adattato, ossia quindi all'ingresso dell'apparato di ricezione.

Il segnale ricevuto, con sovrapposto il rumore, va in ingresso al filtro, il quale genera in uscita una forma d'onda che presenta il proprio valore massimo in corrispondenza di uno specifico istante. In questo istante, quindi, il campionatore misura il segnale (cioè misura il suo valore massimo) e lo trasferisce al sistema a soglia, che deve decidere per l'1 o per lo 0. Quindi, all'uscita dal campionatore, si ottiene un campione di segnale (ad esempio V_0 o V_1) con sommato un campione di rumore (ad esempio n_0 o n_1):

$$0 \text{ trasmesso} \rightarrow V_0 + n_0$$

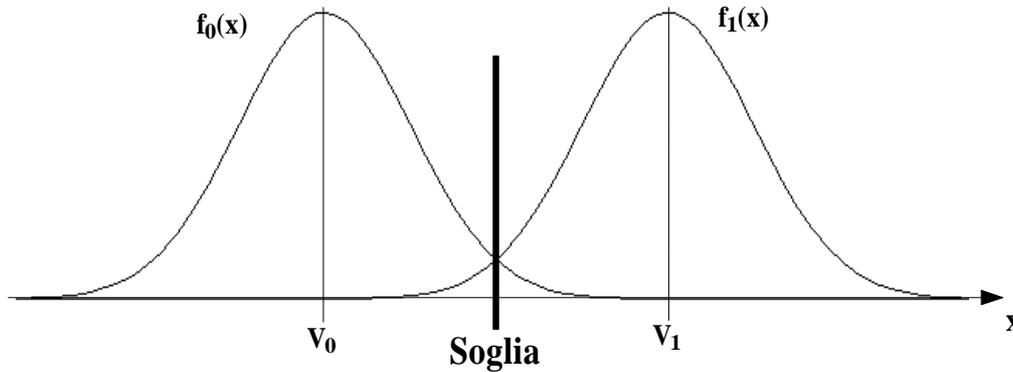
$$1 \text{ trasmesso} \rightarrow V_1 + n_1$$

A differenza di V_0 e V_1 , che sono quantità deterministiche, le quantità n_0 ed n_1 , essendo campioni di un processo casuale, sono variabili aleatorie. Non solo, ma si tratta di campioni dello stesso

¹³ E' ovvio che la responsabilità di errore è imputabile proprio al rumore sovrapposto al segnale: se non ci fosse alcun rumore, il sistema non avrebbe mai modo di sbagliare.

processo casuale, anche se misurati in istanti diversi: dato che il rumore è notoriamente un processo stazionario¹⁴, le caratteristiche statistiche di n_0 ed n_1 sono le stesse. Quali sono queste caratteristiche? Partiamo dal fatto che il rumore $n(t)$ sovrapposto a $s_R(t)$ è un rumore bianco gaussiano a valor medio nullo: all'uscita del filtro, il rumore non è più bianco, ma è comunque gaussiano e con valor medio nullo; di conseguenza, n_0 ed n_1 sono variabili aleatorie, di pari varianza (che indicheremo con σ_n^2), con densità di probabilità ancora gaussiana.

La situazione, peraltro già schematizzata e descritta in precedenza, è la seguente:



dove x è la tensione all'uscita dal campionatore: la funzione $f_0(x)$ è la densità di probabilità del segnale, in uscita dal campionatore, condizionata all'evento *0 trasmesso*; la funzione $f_1(x)$ è invece la densità di probabilità dello stesso segnale, condizionata però all'evento *1 trasmesso*. Tali densità di probabilità, per le considerazioni di prima, sono gaussiane con stessa deviazione standard σ_n ma diverso valor medio:

$$f_0(x) = \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\frac{(x-V_0)^2}{2\sigma_n^2}} \qquad f_1(x) = \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\frac{(x-V_1)^2}{2\sigma_n^2}}$$

Una volta misurata, nell'istante prefissato di lettura, la tensione x in uscita dal campionatore, si tratta di decidere se è stato trasmesso 1 oppure 0 e lo si fa mediante la soglia S : se la tensione è superiore alla soglia, si decide per l'1 trasmesso, mentre in caso contrario si decide per lo 0 trasmesso. Valutiamo la probabilità di errore nei due casi:

- supponiamo che sia stato trasmesso 0, per cui la curva cui fare riferimento è $f_0(x)$: in questo caso, il sistema sbaglia se il valore x del segnale nell'istante di lettura è superiore alla soglia S , per cui deduciamo che la probabilità di errore vale

$$P(\epsilon | 0) = P(x > S) = \int_S^{+\infty} f_0(x) dx = \int_S^{+\infty} \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\frac{(x-V_0)^2}{2\sigma_n^2}} dx$$

Abbiamo già osservato che $P(\epsilon|0)$ corrisponde all'area sottesa da $f_0(x)$ tra la soglia e $+\infty$;

- supponiamo adesso che sia stato trasmesso 1, per cui dobbiamo fare riferimento a $f_1(x)$: in questo caso, il sistema sbaglia se il valore x del segnale nell'istante di lettura è inferiore alla soglia S , per cui la probabilità di errore vale

¹⁴ avente cioè caratteristiche statistiche indipendenti dal tempo

$$P(\epsilon | 1) = P(x < S) = \int_{-\infty}^S f_1(x) dx = \int_{-\infty}^S \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\frac{(x-V_1)^2}{2\sigma_n^2}} dx$$

Ovviamente, $P(\epsilon | 1)$ corrisponde all'area sottesa da $f_1(x)$ tra $-\infty$ e la soglia.

Sempre riprendendo quanto già detto in precedenza, sappiamo che la probabilità di errore del sistema vale

$$p(\epsilon) = p(0T)P(\epsilon | 0) + p(1T)P(\epsilon | 1)$$

dove $p(1T)$ e $p(0T)$ sono, rispettivamente, la probabilità di aver trasmesso 1 e la probabilità di avere trasmesso 0. Per simboli equiprobabili, risulta $p(1T) = p(0T) = 0.5$, per cui abbiamo che

$$p(\epsilon) = 0.5 \cdot [P(\epsilon | 0) + P(\epsilon | 1)]$$

A questo punto, proprio perché i simboli 0 ed 1 sono equiprobabili, è chiaro che conviene porre la soglia esattamente al centro tra i valori V_0 e V_1 :

$$S = \frac{V_0 + V_1}{2}$$

Così facendo, anche le singole probabilità di errore $P(\epsilon | 0)$ e $P(\epsilon | 1)$ sono uguali. Ponendole entrambe uguali a $P(\epsilon | T)$, concludiamo dunque che

$$p(\epsilon) = P(\epsilon | T)$$

E' importante osservare che questo risultato non dipende affatto dalla equiprobabilità dei simboli: esso dipende dall'aver piazzato la soglia a metà, il che significa che $P(\epsilon | 0) = P(\epsilon | 1)$, e dal fatto che la densità di probabilità della tensione x abbia un andamento simmetrico rispetto al valore medio. Per esempio, supponendo $p(1T) \neq p(0T)$, si avrebbe

$$p(\epsilon) = p(0T)P(\epsilon | 0) + p(1T)P(\epsilon | 1) = [p(0T) + p(1T)]P(\epsilon | 1) = P(\epsilon | 1)$$

dove ovviamente potevamo considerare sia $P(\epsilon | 1)$ sia $P(\epsilon | 0)$, data l'uguaglianza.

L'equiprobabilità dei simboli fornisce indicazioni solo riguardo alla ottimalità o meno della scelta della soglia: una volta deciso di porre la soglia a metà, in modo che $P(\epsilon | 0) = P(\epsilon | 1)$, affinché questa scelta sia ottimale (nel senso di minimizzazione della probabilità di errore complessiva, allora i simboli devono essere equiprobabili. Se, invece, ci fosse un simbolo più probabile dell'altro, sarebbe conveniente sistemare la soglia in modo da minimizzare la probabilità di errore su quel simbolo, tollerando invece una probabilità di errore maggiore per l'altro.

Tornando adesso al dimensionamento, una volta fissata la posizione di S , si tratta di calcolare numericamente $P(\epsilon | 0)$ e $P(\epsilon | 1)$, ossia di calcolare gli integrali prima individuati. Essendo gaussiane le funzioni integrande, non è assolutamente possibile procedere per via analitica; allora, integrali di quel tipo sono stati ampiamente tabulati per diversi valori di varianza, di valor medio e di estremo di integrazione. In particolare, è disponibile, in forma appunto tabulata, il valore dell'integrale, tra w e $+\infty$, di una coda gaussiana di valor medio nullo e di varianza unitaria, al variare di w stesso:

$$Q(w) = \int_w^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Si può allora sfruttare questo integrale per calcolare $P(\epsilon|0)$ e $P(\epsilon|1)$. Infatti, per il calcolo di $P(\epsilon|0)$, basta effettuare, nell'espressione di $P(\epsilon|0)$ stessa, un cambio di variabile:

$$P(\epsilon|0) = \int_s^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(x-V_0)^2}{2\sigma_n^2}} dx \xrightarrow{y=\frac{x-V_0}{\sigma_n}} P(\epsilon|0) = \int_{\frac{s-V_0}{\sigma_n}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

L'integrale così ottenuto corrisponde proprio alla funzione $Q(w)$ vista prima, dove però l'estremo inferiore w di integrazione è questa volta $\frac{s-V_0}{\sigma_n}$: possiamo dunque porre

$$P(\epsilon|0) = Q\left(\frac{s-V_0}{\sigma_n}\right)$$

Potremmo procedere allo stesso modo per $P(\epsilon|1)$, ma abbiamo detto che questa quantità, avendo posto la soglia a metà, è esattamente pari a $P(\epsilon|0)$.

Quindi, la quantità che determina la probabilità di errore nel caso di trasmissione di 0 è

$$\gamma_0 = \frac{s-V_0}{\sigma_n}$$

Nel caso di trasmissione di 1, si tratta invece della quantità $\gamma_1 = \frac{V_1-s}{\sigma_n}$, ma è immediato verificare che i due γ sono identici (in modulo) se la soglia è posta a metà:

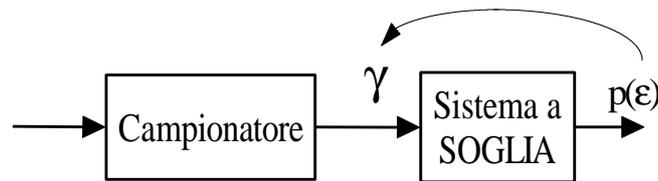
$$s = \frac{V_0 + V_1}{2} \longrightarrow \gamma = |\gamma_0| = |\gamma_1| = \frac{V_1 - V_0}{2\sigma_n}$$

Ci siamo riferiti ai moduli in quanto quello che conta è sempre la distanza tra V_1 e V_0 .

Con questi passaggi, siamo arrivati alla seguente conclusione: nell'ipotesi di simboli equiprobabili e nell'ipotesi di soglia del decisore posta esattamente a metà, la probabilità di errore $p(\epsilon)$ coincide con la probabilità di errore sul generico bit (quindi $P(\epsilon|0)$ oppure $P(\epsilon|1)$) e questa, a sua volta, coincide con la quantità $Q(\gamma)$:

$$p(\epsilon) = P(\epsilon|0) = Q(\gamma) = Q\left(\frac{V_1 - V_0}{2\sigma_n}\right)$$

Quindi, nota la legge con cui la funzione Q varia al variare di γ , abbiamo in pratica trovato una corrispondenza tra il vincolo $p(\epsilon)$ all'uscita del decisore ed il vincolo γ all'ingresso del decisore: un fissato valore di $p(\epsilon)$ corrisponde ad un fissato valore di γ .



Nell'ottica del procedimento a ritroso di cui abbiamo parlato all'inizio, siamo dunque passati dall'uscita del sistema all'uscita del campionario.

A questo punto, dovremmo passare a monte del campionario, trovando l'analogo della quantità γ . Prima di fare questo, però, possiamo anche scegliere il tipo di codifica di linea impiegata, al fine di stabilire quanto valgono V_0 e V_1 :

- una prima possibilità è la codifica ortogonale, in base alla quale trasmettiamo segnale nullo per lo 0 e forma d'onda $s_1(t)$ per l'1: in questo caso, l'uscita del campionario è $V_0=0$ in corrispondenza dello zero logico e $V_1=V_{\max}$ in corrispondenza dell'uno logico, il che significa che $S=V_{\max}/2$ e che $\gamma = \frac{V_{\max}}{2\sigma_n}$;
- la seconda possibilità è la codifica antipodale, in base alla quale trasmettiamo $s(t)$ per l'uno logico e $-s(t)$ per lo zero logico: in questo caso, l'uscita del campionario è $-V$ in corrispondenza dello zero logico e $+V$ in corrispondenza dell'uno logico, il che significa che $S=0$ e che $\gamma = \frac{V}{\sigma_n}$.

Mettiamoci, ad esempio, nella seconda ipotesi, per cui la codifica è antipodale e risulta

$$\gamma = \frac{V}{\sigma_n}$$

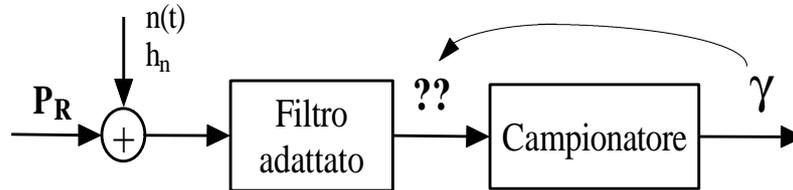
In questo caso, dunque, richiedere un prefissato valore di $p(\epsilon)$ a valle del decisore significa richiedere un prefissato valore del rapporto V/σ_n a valle del campionario: questo rapporto è una forma di rapporto segnale-rumore, in quanto confronta il valore campionato della forma d'onda in assenza di rumore e la deviazione standard del rumore stesso.

In un certo senso, quindi, siamo passati da un quantità numerica¹⁵ $p(\epsilon)$ ad una quantità tipicamente analogica quale è il rapporto S/N. Non poteva che essere così, in quanto abbiamo già osservato in precedenza che il sistema compreso tra l'uscita del generatore di forme d'onda (in trasmissione) e l'ingresso del decisore è puramente analogico.

¹⁵ nel senso di quantità riferita a simboli binari

Rapporto S/N a valle del filtro in ricezione

Il passo successivo è quello di legare la quantità $\gamma = \frac{V}{\sigma_n}$ ad una corrispondente quantità (ancora analogica) a monte del campionatore:



Dobbiamo in definitiva passare dalle misure istantanee compiute dal campionatore alle forme d'onda in uscita dal filtro di ricezione.

Il filtro fornisce in uscita forme d'onda il cui massimo, in istanti predeterminati, è campionato dal campionatore: di conseguenza, al valore V a monte del campionatore non può che corrispondere il valore di picco $V_{S,p}$ della tensione in assenza di rumore. Si tratta ora di vedere a cosa corrisponda la deviazione standard σ_n dei campioni di rumore.

Consideriamo la funzione di autocorrelazione del rumore filtrato $n_u(t)$, definita come l'antitrasformata della densità spettrale:

$$\mathfrak{R}_{n_u, n_u}(\tau) = E[n_u(t)n_u(t + \tau)] = \mathfrak{F}^{-1}[h_{nu}(f)] = \mathfrak{F}^{-1}[|H(f)|^2 h_n(f)]$$

Se indichiamo con $n_c(t)$ il rumore filtrato campionato, la sua funzione di autocorrelazione è

$$\mathfrak{R}_{n_c, n_c}(kT) = E[n_c(iT)n_c(iT + kT)]$$

Adesso, osserviamo che la funzione $\mathfrak{R}_{n_u, n_u}(\tau)$ dovrebbe, in teoria, dipendere sia da t sia da τ . Tuttavia, il fatto che il processo di rumore sia stazionario fa sì che essa dipenda solo da τ . Possiamo allora calcolare $\mathfrak{R}_{n_u, n_u}(\tau)$ nell'istante $\tau=kT$:

$$\mathfrak{R}_{n_u, n_u}(kT) = E[n_u(t)n_u(t + kT)]$$

Ponendo inoltre $t=iT$, con i generico, otteniamo

$$\mathfrak{R}_{n_u, n_u}(kT) = E[n_u(iT)n_u(iT + kT)]$$

Il valore numerico che il rumore assume in un certo istante di campionamento è lo stesso prima e dopo il campionamento, da cui deduciamo che

$$\mathfrak{R}_{n_u, n_u}(\tau) = \mathfrak{R}_{n_c, n_c}(kT)$$

Abbiamo cioè trovato che la funzione di autocorrelazione dei campioni di rumore all'uscita dal campionatore è numericamente uguale al valore della funzione di autocorrelazione del rumore prima del campionamento, presa per ritardi discreti pari a multipli di T ¹⁶.

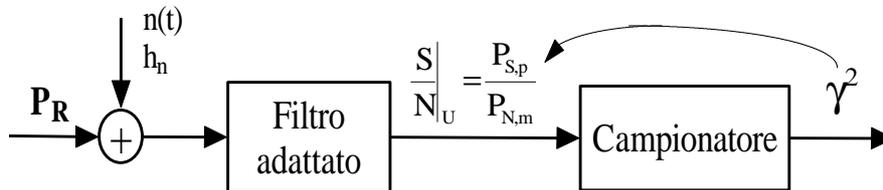
A questo punto, per semplice definizione, la varianza del rumore all'uscita dal campionatore vale $\sigma_n^2 = \mathfrak{R}_{nc,nc}(0)$ e quindi anche $\sigma_n^2 = \mathfrak{R}_{nu,nu}(0)$. Dato, però, che $\mathfrak{R}_{nu,nu}(0)$ è la varianza del rumore all'uscita dal filtro, deduciamo, in conclusione, che la varianza del rumore è identica prima e dopo il campionamento:

$$\sigma_n^2 = \sigma_{nu}^2$$

Non è ancora finita, in quanto possiamo sfruttare un'altra importante proprietà del processo di rumore e cioè la sua ergodicità: in base alla ergodicità, possiamo legare la varianza σ_{nu}^2 ad una quantità che è una media temporale del processo di rumore considerato, ossia la potenza media:

$$\sigma_n^2 = \sigma_{nu}^2 = P_{Nu,media}$$

Quindi, a conclusione di questo discorso, la quantità $\gamma = \frac{V}{\sigma_n}$ a valle del campionatore può essere legata al rapporto $\frac{V_{S,p}}{V_{N,eff}}$ tra il valore di picco del segnale e il valore efficace del rumore. In termini di rapporto segnale/rumore vero e proprio (cioè in termini di potenze), possiamo considerare il rapporto $\frac{P_{S,p}}{P_{N,m}}$ tra la potenza di picco di segnale (pari al quadrato del valore di picco¹⁷) e la potenza media di rumore (pari al quadrato del valore efficace):



Possiamo dunque porre

$$\frac{S}{N}_U = \frac{P_{S,p}}{P_{N,m}} = \gamma^2$$

Riepilogando ancora una volta, siamo risaliti da $p(\epsilon)$ a valle del decisore a γ a valle del campionatore a $\frac{P_{S,p}}{P_{N,m}}$ a valle del filtro. Questo, nell'ipotesi di codifica antipodale.

¹⁶ Tutto ciò non sarebbe invece valido se il processo di rumore non fosse stazionario

¹⁷ In tutti i nostri discorsi, noi diciamo che la potenza di picco è pari al quadrato del valore di picco: questo è vero solo se si ipotizza unitario il livello di impedenza. Quindi, tutte le volte che viene fuori questo discorso, si deve ritenere implicita l'ipotesi per cui $Z=1$.

Potenza ricevuta dal mezzo trasmissivo

A questo punto, dobbiamo passare dall'uscita all'ingresso del filtro di ricezione, in modo da determinare la potenza che è necessario ricevere dal mezzo trasmissivo. In altre parole, dobbiamo passare dal rapporto S/N all'uscita del filtro al rapporto S/N all'ingresso del filtro stesso. E' ovvio che dobbiamo considerare il funzionamento del filtro.

A questo proposito, abbiamo già visto che il filtro che trae maggiori informazioni dalla forma d'onda ricevuta $s_R(t)$ è il **filtro adattato**, il quale si deve adattare non solo alla forma d'onda ricevuta, ma anche alla densità spettrale del rumore sovrapposto. Nel caso da noi considerato, tale densità spettrale è costante (rumore bianco), per cui, in base a passaggi precedentemente effettuati, si trova che il rapporto S/N (potenza di picco su potenza media) all'uscita dal filtro adattato vale

$$\left. \frac{S}{N} \right|_U = \frac{E_R}{h_n / 2}$$

Questa formula è importante proprio perché lega il rapporto S/N in uscita dal filtro a quantità (E_R e h_n) relative all'ingresso del filtro stesso. Per risalire alla potenza ci basta dividere numeratore e denominatore per T:

$$\left. \frac{S}{N} \right|_U = \frac{\frac{E_R}{T}}{\frac{h_n}{2T}} = \frac{\frac{E_R}{T}}{h_n \frac{f_s}{2}}$$

Se E_R è l'energia corrispondente alla singola forma d'onda ricevuta, $\frac{E_R}{T}$ è la potenza media di tale forma d'onda, ossia la potenza media di segnale: infatti, ricordando che stiamo supponendo una codifica antipodale, in ogni periodo si presenta esattamente la stessa forma d'onda, positiva o negativa, per cui per il simbolo relativo ad ogni periodo arriva lo stesso contributo di energia; di conseguenza, effettuando la media, si ottiene effettivamente la potenza media.

La quantità $h_n \frac{f_s}{2}$, invece, può essere interpretata come la potenza media di rumore valutata in una banda convenzionale pari a metà della frequenza cifra f_s : si può infatti scrivere che

$$h_n \frac{f_s}{2} = \int_0^{f_s/2} h_n df$$

e questa è appunto la formula con cui si ricava la potenza media di rumore in una specifica banda.

Abbiamo dunque che

$$\left. \frac{S}{N} \right|_U = \frac{P_R}{h_n \frac{f_s}{2}} \quad (**)$$

A questo punto, abbiamo a disposizione due espressioni del rapporto S/N all'uscita del filtro: quella appena ricavata e quella $\left. \frac{S}{N} \right|_U = \frac{P_{S,p}}{P_{N,m}} = \gamma^2$ in funzione di γ .

Possiamo dunque porre

$$\frac{S}{N}|_U = \frac{P_R}{h_n \frac{f_S}{2}} = \gamma^2$$

Da questa relazione, noto il valore di γ , si può calcolare la potenza media P_R necessaria in ricezione. Nota questa, basta conoscere l'attenuazione prodotta dal mezzo trasmissivo per risalire alla potenza necessaria in trasmissione:

$$P_T[\text{dB}] = P_R[\text{dB}] + \alpha_{\text{tot}}[\text{dB}]$$

Come si vedrà bene nell'esempio numerico che seguirà, l'attenuazione prodotta dal mezzo trasmissivo dipende dalla massima frequenza del segnale da trasmettere sul mezzo stesso e quindi dalle forme d'onda scelte per la trasmissione. Per il momento, però, le forme d'onda in trasmissione non sono state ancora scelte, per cui la banda occupata non è ancora subentrata nei nostri discorsi.

Osservazione: caso della codifica ortogonale

Vediamo adesso se e come cambiano i discorsi appena seguiti nel caso in cui la codifica di linea adottata sia una codifica ortogonale.

Abbiamo già detto che, in questo caso, supponendo ancora di porre la soglia a metà e supponendo sempre equiprobabili i simboli binari, risulta

$$\gamma = \frac{V_{\text{max}}}{2\sigma_n}$$

Se allora ripetiamo i discorsi fatti prima a proposito del passaggio dall'uscita all'ingresso del campionario, la quantità cui fare riferimento è adesso $\frac{V_{S,p}/2}{V_{N,\text{eff}}}$, dove il 2 a denominatore deriva

appunto dal fatto che γ dipende non dal valore di picco, ma dalla sua metà. Se poi passiamo alle potenze, cioè eleviamo al quadrato, il 2 diventa un 4:

$$\frac{P_{S,p}}{4P_{N,m}} = \gamma^2$$

A questo punto, se vogliamo confrontare questo risultato con quello ottenuto per la codifica antipodale, dobbiamo necessariamente supporre che il filtro utilizzato sia lo stesso, in quanto solo questa ipotesi ci garantisce che il rumore in uscita dal filtro sia lo stesso.

Fatta dunque questa ipotesi, se ripetiamo gli stessi discorsi fatti in precedenza a proposito del filtro adattato, dall'ultima equazione scritta ricaviamo che

$$\frac{S}{N}|_U = \frac{E_R}{h_n/2} = 4\gamma^2$$

dove E_R rappresenta l'energia ricevuta, in un intervallo di tempo T , solo quando c'è la forma d'onda¹⁸, ossia quando è trasmesso un 1.

Dividendo ancora una volta numeratore e denominatore per T , abbiamo che

$$\frac{\frac{E_R}{T}}{2\frac{h_n}{T}} = \gamma^2$$

A numeratore, abbiamo la quantità $\frac{E_R}{T}$, che, nel caso della codifica antipodale, rappresentava la potenza media del segnale in ricezione; invece, questo non è più vero: infatti, dato che la forma d'onda è presente solo se è trasmesso un 1, T non è più il periodo di ripetizione della forma d'onda, ma semplicemente il periodo di cifra (ossia il periodo di ripetizione dei simboli). Nel caso in cui i simboli siano equiprobabili, $\frac{E_R}{T}$ rappresenta invece il doppio della potenza media del segnale ricevuto. Ponendo dunque $\frac{E_R}{T} = 2P_R$, abbiamo che

$$\gamma^2 = \frac{2P_R}{2\frac{h_n}{T}} = \frac{P_R}{h_n f_s}$$

Confrontando con l'espressione di $\left. \frac{S}{N} \right|_U$ ottenuta nel caso antipodale, osserviamo la mancanza del 2 a numeratore: ciò significa, come del resto avevamo trovato già in precedenza, che, *per ottenere, con la codifica ortogonale, la stessa probabilità di errore che si otterrebbe con codifica antipodale è necessario trasmettere una potenza media doppia (cioè 3dB in più di potenza media)*. La differenza deriva dal fatto che, per ottenere la stessa $p(\epsilon)$, il valore di picco della codifica ortogonale deve essere doppio di quello della codifica ortogonale.

¹⁸ in quanto ricordiamo che, con la codifica ortogonale, non si trasmette nulla in corrispondenza dello 0 logico, per cui si ha un'unica forma d'onda (rettangolare) solo quando va trasmesso un 1.

ESEMPIO NUMERICO: TRASMISSIONE NUMERICA DEL SEGNALE TV A COLORI

Siamo a questo punto in grado di svolgere un esercizio su un sistema di trasmissione numerico.

In primo luogo, dobbiamo decidere che tipo di segnale vogliamo trasmettere: scegliamo il **segnale televisivo**. In particolare, mentre in esercizi precedenti abbiamo considerato il segnale televisivo in bianco e nero, questa volta consideriamo quello **a colori**. La differenza con il segnale bianco/nero è nel fatto che dobbiamo trasmettere 3 distinti segnali: il segnale di luminanza (che indichiamo con Y), comune al segnale bianco/nero, e i due segnali di crominanza (U e V).

Il primo passo è quello di convertire il segnale analogico da trasmettere in forma numerica. Nel nostro caso, i segnali analogici da trasmettere sono 3: si procede, allora, prima passando in forma numerico i 3 segnali e poi facendone la *multiplazione a divisione di tempo*¹⁹ (**TDM**), in modo da ottenere un'unica sequenza di bit da trasmettere. La numerizzazione, come è noto, si svolge secondo una precisa sequenza di operazioni: prima si effettua un filtraggio passa-basso (necessario per prevenire fenomeni di aliasing), poi una operazione di campionamento (a frequenza almeno doppia della massima frequenza del segnale considerato) e poi una quantizzazione.

Cominciamo dal segnale di luminanza Y. Sappiamo che esso presenta una banda **B = 5 MHz**: di conseguenza, il *filtraggio anti-aliasing* viene fatto proprio con banda di 5 MHz. Il successivo *campionamento*, che deve essere fatto ad una frequenza di almeno 10 MHz, viene effettuato, in base agli standard internazionali, a **13.5 MHz** (cioè prelevando 1 campione ogni 74 nsec). Si procede, quindi, ad una quantizzazione uniforme con **8 bit** per campione (il che significa che vengono quantizzate $2^8=256$ ampiezze diverse). Questi numeri servono a valutare la velocità di trasmissione dei bit sul canale: Infatti, associando 8 bit a ciascun campione e avendo un campione ogni $1/13.5(\text{MHz})$ secondi, il numero di bit da trasmettere ogni secondo è

$$N = 8(\text{bit}) * 13.5 * 10^6 \left(\frac{1}{\text{sec}} \right) = 108 * 10^6 \left(\frac{\text{bit}}{\text{sec}} \right) = 108 \left(\frac{\text{Mbit}}{\text{sec}} \right)$$

La frequenza di campionamento di 13.5 MHz e gli 8 bit per la quantizzazione sono valori appartenenti allo **standard CCIR 4:4:2**. Questo prevede una ulteriore specifica, relativa alla risoluzione: viene infatti richiesto che il segnale di luminanza venga descritto con una risoluzione doppia rispetto ad entrambe le componenti di crominanza. Di conseguenza, per la numerizzazione di tali segnali di crominanza non avremo più 108 Mbit/sec, ma qualcosa in meno.

Eseguendo la numerizzazione dei segnali di crominanza così come per il segnale di luminanza, otteniamo, per ciascuno di essi, una sequenza di bit: per dimezzare la risoluzione, ci basta scartare un bit ogni due sia per la risoluzione orizzontale sia per la risoluzione verticale. Di conseguenza, ciascuno dei segnali di crominanza fornirà un tasso di bit al secondo pari ad un quarto di quello ottenuto per il segnale di luminanza:

¹⁹ Ricordiamo che esistono due diverse tecniche con cui trasmettere più segnali, sia in formato analogico sia in formato digitale, in un unico mezzo trasmissivo: la multiplazione a divisione di tempo (**TDM** - Time Division Multiplexing) e la multiplazione a divisione di frequenza (**FDM** - Frequency Division Multiplexing). Entrambe queste tecniche sono realizzabili sia in analogico sia in digitale, con differenti difficoltà: è molto più facile realizzare la tecnica TDM in digitale (basta inviare, in ogni unità di tempo, un campione per ogni segnale da trasmettere, sfruttando i *tempi morti* della conversione analogico→digitale) così come è molto più facile realizzare la tecnica FDM in analogico (basta modulare opportunamente i segnali, in modo da affiancarle i rispettivi spettri e lasciare lo *spazio* necessario a distinguerli).

$$\begin{aligned} \text{luminanza} &\longrightarrow 108 \left(\frac{\text{Mbit}}{\text{sec}} \right) \\ \text{crominanza U} &\longrightarrow \frac{108}{4} \left(\frac{\text{Mbit}}{\text{sec}} \right) = 27 \left(\frac{\text{Mbit}}{\text{sec}} \right) \\ \text{crominanza V} &\longrightarrow \frac{108}{4} \left(\frac{\text{Mbit}}{\text{sec}} \right) = 27 \left(\frac{\text{Mbit}}{\text{sec}} \right) \end{aligned}$$

A questo punto, avendo a disposizione tutti e tre i segnali da trasmettere in forma in numerica, va effettuata la moltiplicazione a divisione di tempo: anziché inviare 1 bit ogni unità di tempo, ne dovremo inviare 3. Ciò significa, che la frequenza di cifra (cioè appunto il numero di bit inviati ogni secondo) è

$$f_s = 108 \left(\frac{\text{Mbit}}{\text{sec}} \right) + 27 \left(\frac{\text{Mbit}}{\text{sec}} \right) + 27 \left(\frac{\text{Mbit}}{\text{sec}} \right) = 162 \left(\frac{\text{Mbit}}{\text{sec}} \right)$$

Abbiamo dunque ottenuto una frequenza di cifra **$f_s=162(\text{Mbit/sec})$** .

Fissata la frequenza di cifra, abbiamo in precedenza già osservato che la successiva specifica da conoscere è la *probabilità di errore*, ossia la probabilità $p(\epsilon)$ che l'utente finale riceva un bit diverso da quello effettivamente trasmesso. Possiamo valutare $p(\epsilon)$ fissando la qualità delle immagini da visualizzare sul nostro ipotetico televisore; in particolare, possiamo decidere qual è il numero massimo di bit sbagliati che possiamo tollerare per ogni quadro. Generalmente, si accetta **1 bit sbagliato per semiquadro**. Fissata questa indicazione, per giungere a $p(\epsilon)$ abbiamo bisogno di conoscere quanti quadri vediamo in 1 secondo e questo dipende dallo *standard televisivo* utilizzato: se, per esempio, scegliamo il **sistema PAL**, la frequenza di quadro è di 25 Hz, vale a dire 25 quadri al secondo²⁰. Noto il numero di quadri in 1 secondo, possiamo calcolare il numero di bit per ogni quadro:

$$\frac{162 \left(\frac{\text{Mbit}}{\text{sec}} \right)}{25 \left(\frac{\text{quadri}}{\text{sec}} \right)} = 6.5 \left(\frac{\text{Mbit}}{\text{quadro}} \right)$$

A questo punto abbiamo concluso, in quanto accettare un bit sbagliato per ogni quadro significa accettare una probabilità di errore

$$p(\epsilon) = \frac{1 \left(\frac{\text{bit sbagliato}}{\text{quadro}} \right)}{6.5 * 10^6 \left(\frac{\text{bit}}{\text{quadro}} \right)} = 1.538 * 10^{-7}$$

Possiamo dunque arrotondare, richiedendo una probabilità di errore **$p(\epsilon)=10^{-7}$** .

Abbiamo dunque le specifiche che il nostro sistema deve rispettare. Possiamo perciò passare al dimensionamento, seguendo gli stessi passi teorici indicati nei paragrafi precedenti. Tra l'altro, possiamo anche dimenticarci, a questo punto, del tipo di segnale che stiamo trasmettendo, in quanto le uniche informazioni che d'ora in poi utilizzeremo sono $p(\epsilon)$ ed f_s .

²⁰ Ricordiamo che, invece, nel sistema **NTSC** adottato negli Stati Uniti, la frequenza di quadro è di 30 Hz.

Per prima cosa, dobbiamo passare dalla $p(\varepsilon)$ a valle del decisore al rapporto segnale-rumore a valle del campionatore: ipotizzando, come sempre fatto nei discorsi precedenti, che il rumore sovrapposto sia quello termico e decidendo subito di porre la soglia a metà, sappiamo che la formula cui fare riferimento è

$$p(\varepsilon) = P(\varepsilon | 1) = P(\varepsilon | 0) = Q(\gamma) = \int_{\gamma}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

Il passaggio avviene dunque fondamentalmente da $p(\varepsilon)$ alla quantità γ , che ricordiamo essere, in generale, pari a $\frac{V_1 - V_0}{2\sigma_n}$.

Fissata $p(\varepsilon)$, dobbiamo conoscere γ e quindi dobbiamo conoscere come la funzione $Q(\gamma)$ varia con γ stesso. A tal proposito, si verifica che le variazioni di $Q(\gamma)$ con γ sono molto forti. In particolare, per valori di $Q(\gamma)$ (e quindi di $p(\varepsilon)$) compresi tra 10^{-3} e 10^{-9} , si può affermare, con buona approssimazione, che per ottenere un cambiamento di 1 ordine di grandezza (in più o in meno) di $Q(\gamma)$ è necessario variare γ di 1 dB (rispettivamente in meno o in più). Per esempio, volendo ridurre $p(\varepsilon)=Q(\gamma)$ di un fattore 10, è necessario aumentare γ di 1 dB.

Noi dobbiamo trovare quale valore di γ serve per ottenere $Q(\gamma)=p(\varepsilon)=10^{-7}$. Dalle apposite tabelle, si trova che, per ottenere $p(\varepsilon)=10^{-6}$, è necessario prendere $\gamma=13.5$ dB. Allora, in base a quanto appena detto, per ottenere $p(\varepsilon)=10^{-7}$, dobbiamo prendere **$\gamma=14.5$ dB**.

Su questi 14.5 dB è importante fare una osservazione: infatti, ricordando che γ è un rapporto di tensioni, è chiaro che il valore di γ in dB si ottiene dal valore di γ in unità naturali tramite la formula

$$(\gamma)_{\text{dB}} = 20 \log_{10} \gamma$$

(dove il termine moltiplicativo 20 deriva proprio dal fatto che si considera un rapporto di tensione, mentre si usa 10 per un rapporto tra potenze).

Questo ci servirà tra poco per un'altra osservazione.

Noto γ (che in pratica rappresenta il rapporto segnale/rumore a valle del campionatore), possiamo associare questo parametro al rapporto segnale/rumore $\left. \frac{S}{N} \right|_U$ a valle del filtro di ricezione. Qui, però, dobbiamo operare una ulteriore scelta, scegliendo tra una codifica ortogonale e una codifica antipodale:

$$\begin{aligned} \text{antipodale} &\longrightarrow \left. \frac{S}{N} \right|_U = \frac{P_{S,p}}{P_{N,m}} = \gamma^2 \\ \text{ortogonale} &\longrightarrow \left. \frac{S}{N} \right|_U = \frac{P_{S,p}}{P_{N,m}} = 4\gamma^2 \end{aligned}$$

dove $P_{S,p}$ rappresenta la potenza di picco del segnale all'uscita del filtro di ricezione e $P_{N,m}$ rappresenta la potenza media del rumore all'uscita dello stesso filtro.

Optiamo, ad esempio, per una **codifica antipodale**, per cui la definizione da considerare è la prima delle due appena elencate. Per passare all'ingresso del filtro di ricezione, dobbiamo scegliere il tipo di filtro, ossia la funzione di trasferimento. Optiamo per il solito **filtro adattato**, per il quale sappiamo che

$$\left. \frac{S}{N} \right|_U = \frac{E_R}{h_n / 2}$$

dove E_R è l'energia della forma d'onda ricevuta e h_n la densità spettrale di potenza del rumore che, all'uscita del mezzo trasmissivo, si sovrappone al segnale utile).

Se dividiamo numeratore e denominatore di quel rapporto per T , otteniamo

$$\frac{S}{N}\bigg|_U = \frac{\frac{E_R}{T}}{\frac{h_n}{2T}} = \frac{\frac{E_R}{T}}{h_n \frac{f_S}{2}} = \frac{P_R}{h_n \frac{f_S}{2}}$$

dove P_R è la potenza media (di segnale) ricevuta e $h_n \frac{f_S}{2}$ la potenza media di rumore valutata in una banda convenzionale pari a metà della frequenza di cifra f_S .

Possiamo dunque scrivere, con riferimento a quanto trovato prima, che

$$\frac{P_R}{h_n \frac{f_S}{2}} = \gamma^2 \longrightarrow \boxed{P_R = \gamma^2 h_n \frac{f_S}{2}}$$

Nota f_S e fissata h_n (cioè fissata la rumorosità delle apparecchiature riceventi), siamo dunque in grado di calcolare la minima²¹ potenza media necessaria in ricezione. Esprimendoci in unità logaritmiche, possiamo scrivere che

$$P_R [\text{dB}] = (\gamma^2)_{\text{dB}} + (h_n)_{\text{dB}} + \left(\frac{f_S}{2}\right)_{\text{dBHz}} = 10\log_{10} \gamma^2 + 10\log_{10} h_n + 10\log_{10} \frac{f_S}{2}$$

Ricordiamo adesso che $h_n = FkT_0$: fissando genericamente un **fattore di rumore** $F=10\text{dB}$ per le apparecchiature riceventi e una temperatura $T_0=293^\circ\text{K}$, risulta $(h_n)_{\text{dB}} = -174 + 10 = -164[\text{dB}]$.

Possiamo dunque concludere che

$$P_R [\text{dB}] = (14.5[\text{dB}]) + (-164[\text{dB}]) + \left(10\log_{10} \frac{162 \cdot 10^6}{2}\right)_{\text{dBHz}} = 14.5 - 164 + 79 \cong -70.5[\text{dB}]$$

Osserviamo che la quantità γ^2 , espressa in dB, vale ancora 14.5dB, ossia è pari a $\gamma[\text{dB}]$ calcolato prima. Il motivo, come si deduce dai passaggi effettuati, è che γ^2 è un rapporto tra tensioni al quadrato, ossia tra potenze, per cui il passaggio da unità naturali a unità logaritmiche è

$$(\gamma^2)_{\text{dB}} = 10\log_{10} \gamma^2 = 20\log_{10} \gamma = 14.5[\text{dB}]$$

Adesso, per risalire alla potenza da trasmettere, ci basta conoscere l'attenuazione introdotta dal mezzo trasmissivo: supponiamo, allora, di effettuare un collegamento lungo $L = 100 \text{ km}$, utilizzando un cavo²² con la solita attenuazione specifica $\alpha=2\text{dB/km @ 1 MHz}$.

²¹ Minima nel senso che, ovviamente, un eventuale suo aumento può portare nient'altro che un miglioramento delle prestazioni, ossia quindi una riduzione di $p(\epsilon)$.

Nota l'attenuazione specifica (cioè i dB/km) alla frequenza di 1 MHz, per ottenere l'attenuazione specifica complessiva α_T (e quindi l'attenuazione complessiva α_{tot} da sommare alla potenza P_R per ottenere la potenza P_T da trasmettere) dobbiamo conoscere la massima frequenza del segnale da trasmettere sul mezzo considerato²³. Questo perché, ipotizzando di effettuare una **equalizzazione passiva** del mezzo, l'attenuazione specifica complessiva, nota quella relativa ad una particolare frequenza di riferimento, si ottiene dalla formula

$$\alpha_T = \alpha \sqrt{\frac{f_{max}}{f_{rif}}} \quad (\text{unità logaritmiche})$$

La frequenza massima del segnale da trasmettere dipende dalle forme d'onda che intendiamo trasmettere. Avendo stabilito che il filtro deve essere adattato, sappiamo che tali forme d'onda dipendono da cosa vogliamo ottenere all'uscita del filtro stesso: come sappiamo, è importante ottenere, all'uscita del filtro, *forme d'onda ad intersimbolo nullo*. Supponiamo, allora, si effettuare un **progetto alla Nyquist**, imponendo cioè che l'uscita del filtro di ricezione siano forme d'onda appartenenti alla *famiglia di Nyquist*. Questa famiglia presenta diverse forme d'onda, tutte ad intersimbolo nullo, ciascuna delle quali caratterizzata da un proprio *roll off* δ (che quantifica lo smorzamento dello spettro dal valore alto al valore nullo). Scegliere, allora, una particolare forme d'onda significa sceglierne il δ .

Questo δ diventa importante, ai nostri fini, in quanto abbiamo visto che è legato alla banda $B_{Nyquist}$ dello spettro della corrispondente forma d'onda mediante la relazione

$$B_{Nyquist} = \frac{f_s}{2}(\delta + 1)$$

La minima banda si ottiene con $\delta=0$ (cui corrisponde uno spettro rettangolare e quindi una forma d'onda Seno Cardinale), mentre la massima banda si ottiene per $\delta=1$ (cui corrisponde uno spettro a forma di Cos^2 , detto *Coseno rialzato*). Scegliamo questa seconda possibilità, per cui la massima frequenza del segnale in uscita dal filtro di ricezione è

$$B_{Nyquist} = f_s$$

Da questa informazione è immediato risalire alla massima frequenza del segnale da inviare sul mezzo trasmissivo: infatti, il filtro adattato, essendo un operatore lineare, non può produrre in uscita frequenze diverse da quelle ricevute in ingresso (può tutt'al più attenuare quelle già presenti, ma non può certo *inventarne* di nuove), per cui anche il segnale in ingresso al filtro, cioè il segnale trasmesso, conterrà la frequenza f_s . D'altra parte, non ci interessano frequenze superiori ad f_s , in quanto stiamo usando una equalizzazione passiva fino a f_s e quindi sarebbe inutile trasmettere oltre tale frequenza .

²² Visto che parliamo di "cavo", stiamo considerando un mezzo trasmissivo passa-basso e quindi un **sistema di trasmissione numerico in banda base**: ciò significa che non abbiamo bisogno di alcuna modulazione (numerica) prima della trasmissione. Se, al posto del cavo, avessimo ipotizzato un mezzo passa-banda, allora avremmo dovuto considerare una preventiva modulazione, argomento del quale parleremo più avanti.

²³ Per un segnale passa-basso (come quelli che stiamo considerando adesso), la massima frequenza è per definizione la banda del segnale.

In definitiva, quindi, la massima frequenza del segnale trasmesso è numericamente pari alla frequenza di cifra, per cui poniamo $f_{\max}=162 \text{ MHz}$. Sostituendo questo valore nella formula per il calcolo dell'attenuazione specifica complessiva, deduciamo che quest'ultima vale

$$\alpha_T = \alpha \sqrt{\frac{f_{\max}}{f_{\text{rif}}}} = 2(\text{dB}) \cdot \sqrt{\frac{162(\text{MHz})}{1(\text{MHz})}} = 25.5 \left(\frac{\text{dB}}{\text{km}} \right)$$

A questo punto, avendo supposto un collegamento di 100 km, l'attenuazione complessiva è

$$\alpha_{\text{tot}} = \alpha_T \cdot 100 = 2550[\text{dB}]$$

Si tratta, chiaramente, di una attenuazione enorme, per compensare la quale sarebbe necessario trasmettere una potenza pari a

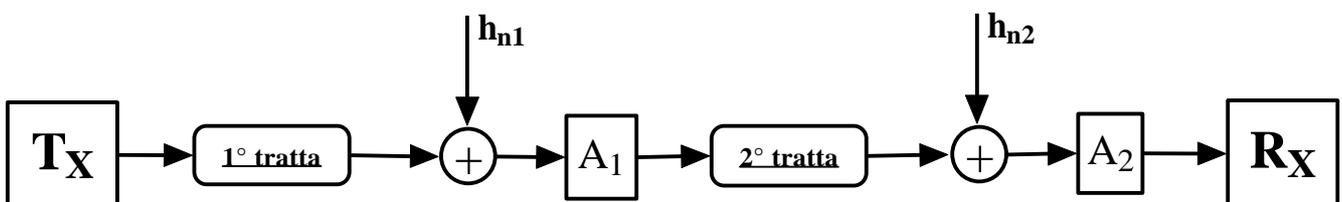
$$P_T[\text{dB}] = P_R[\text{dB}] + \alpha_{\text{tot}}[\text{dB}] = 2479.5[\text{dB}]$$

E' decisamente impensabile trasmettere una simile potenza. Non c'è altro rimedio, allora, che usare anche in questo caso un *sistema multitratta*.

Prima ancora, però, di parlare dei sistemi multitratta per la trasmissione numerica, facciamo una osservazione riguardo i calcoli appena effettuati: quando abbiamo risolto un esercizio analogo (cioè sempre trasmissione del segnale televisivo con le stesse specifiche sul mezzo trasmissivo), ma nel caso analogico, abbiamo trovato una attenuazione complessiva di 450 dB; il motivo per cui l'attenuazione, in quel caso, era minore di quella trovata poco fa è che la massima frequenza del segnale da trasmettere coincideva con la banda del segnale televisivo, ossia 5 MHz, mentre in questo caso abbiamo visto che la massima frequenza da trasmettere è 162 MHz (a causa sia della numerizzazione sia della presenza dei segnali di colore). Questo è dunque un sicuro svantaggio dei sistemi di trasmissione numerica: la numerizzazione del segnale analogico da trasmettere comporta un notevole allargamento della banda da utilizzare. A fronte di questo svantaggio, come vedremo tra un attimo, i sistemi numerici multitratta presentano dei pregi che invece i sistemi analogici multitratta non presentano.

SISTEMA NUMERICO MULTITRATTA SU CAVO

Il modo più semplice di realizzare un **sistema numerico multitratta** è analogo a quello visto nel caso della trasmissione analogica, nel senso che basta predisporre, dopo ogni *spezzone* di cavo, *apparecchiature di pura amplificazione*, che cioè amplifichino (in potenza) il segnale ricevuto e lo ritrasmettano così com'è:



(figura si riferisce evidentemente ad un sistema a 2 tratte).

In questo caso, il dimensionamento è del tutto analogo a quello seguito nel caso di sistema a singola tratta: l'obiettivo è quello di determinare il numero N di tratte necessario per ottenere, in ricezione, la potenza P_R desiderata e avendo fissato la P_T tollerabile.

La determinazione del numero N di tratte si fa in maniera identica a quella seguita nel caso analogico ed il motivo è che questa parte del sistema è assolutamente identica nel caso numerico e in quello analogico: supponendo che le tratte siano tutte uguali tra loro (incluso il fatto che la potenza trasmessa su ciascuna di esse è sempre pari a P_T), sappiamo che il rapporto rumore/segnale totale è la somma dei singoli rapporti rumore/segnale, da cui si ricava che la relazione da considerare è

$$10 \log_{10} N + 14.5[\text{dB}] \leq \left(P_T[\text{dBm}] - \frac{\alpha_{\text{tot}}}{N}[\text{dBm}] \right) - \left(h_n \frac{f_s}{2} \right)_{\text{dB}}$$

Il primo membro dell'equazione corrisponde al rapporto segnale/rumore da richiedere alla singola tratta; il secondo membro è il rapporto S/N ottenibile da ciascuna tratta. Quella relazione fornisce il valore di N necessario per ottenere, all'uscita da ogni singola tratta, il rapporto S/N di 14.5 dB trovato nell'esempio numerico di prima. Scegliendo una potenza in trasmissione pari a $P_T=100\text{mW}$ e risolvendo per tentativi, si trova **$N=35$** .

Notiamo subito come questo numero sia molto maggiore di quello trovato nel caso analogico: in quel caso, infatti, avendo sempre fissato $P_T=100\text{mW}$, avevamo trovato che erano necessarie solo 7 tratte.

Tratte rigenerative

Ci chiediamo se sia possibile migliorare il progetto, ossia, fondamentalmente, ridurre il numero delle tratte. Per sperare di ottenere questo miglioramento, l'unica possibilità è di utilizzare diversamente le singole tratte. Mentre prima abbiamo supposto di utilizzare, tra una tratta e l'altra, apparecchiature di pura amplificazione, possiamo pensare di usare delle **apparecchiature rigenerative**: questo significa che, all'uscita di ogni tratta, dobbiamo effettuare tutte le operazioni che permettono di tornare ai simboli binari, in modo che, una volta riottenuta la sequenza binaria teoricamente²⁴ trasmessa, essa possa essere nuovamente codificata e trasmessa. In altre parole, in ogni apparecchiatura intermedia si torna ai simboli binari prima della successiva trasmissione. Questa è una operazione che molto spesso è necessario fare, a prescindere dai problemi legati al dimensionamento: basti pensare ai casi in cui è necessario estrarre l'informazione in più punti della catena di trasmissione.

Vediamo allora se e come si modifica il progetto del sistema adottato apparecchiature rigenerative tra una tratta e l'altra.

In effetti, con una scelta di questo tipo, ogni tratta diventa sostanzialmente un sistema di trasmissione numerico a sé stante: di conseguenza, su ciascuna tratta di potranno verificare degli errori. C'è però una cosa fondamentale da osservare: supponiamo che, lungo una generica tratta, si verifichi un errore su un certo bit; quale probabilità c'è che anche la tratta successiva compia un errore sullo stesso bit? In generale, anzi, quale probabilità c'è che lo stesso bit venga sbagliato da più tratte? E' possibile fare un discorso prettamente matematico, ma è comunque intuitivo comprendere che è praticamente impossibile che su più tratte si sbagli lo stesso bit.

Questo fatto ha come conseguenza, che, mettendo in cascata tratte rigenerative, di tratta in tratta aumenta il numero medio di bit sbagliati: questo perché ogni tratta compie errori su bit diversi da quelli su cui sbagliano le altre tratte, per cui gli errori sulle varie tratte si vanno sommando gli uni agli altri. Allora, se la probabilità di errore sul generico bit è abbastanza piccola e se gli errori di

²⁴ "teoricamente" in quanto ci sono sempre da considerare gli eventuali errori

tratta in tratta sono statisticamente indipendenti tra loro²⁵, si può affermare che la probabilità di errore complessiva è pari alla somma delle probabilità di errore sulle singole tratte:

$$P_{\text{tot}}(\epsilon) = \sum_{i=1}^N P_i(\epsilon)$$

Questa formula è l'analogo della formula in base alla quale, nei sistemi analogici multitratta su cavo, il rapporto rumore/segnale complessivo è pari alla somma dei singoli rapporti segnale/rumore.

Ovviamente, la probabilità di errore complessiva $P_{\text{tot}}(\epsilon)$ coincide con la $p(\epsilon)$ considerata nei sistemi a singola tratta: allora, supponendo che le $P_i(\epsilon)$ siano tutte uguali tra loro, è chiaro che alla singola tratta è richiesta una probabilità di errore N volte più piccola di quella complessiva richiesta in ricezione:

$$P_{\text{tot}}(\epsilon) = N \cdot P_{\text{singola tratta}}(\epsilon) \longrightarrow P_{\text{singola tratta}}(\epsilon) = \frac{P_{\text{tot}}(\epsilon)}{N}$$

Di conseguenza, il dimensionamento di ogni tratta va effettuato così come nell'esempio di prima, considerando però due differenze: la prima, in base a quanto appena visto, è che la probabilità di errore richiesta è N volte più piccola di $p(\epsilon)$; la seconda è che la lunghezza (e quindi l'attenuazione) di ogni singola tratta è N volte più piccola della lunghezza complessiva L .

Da queste considerazioni, sembrerebbe allora che le cose siano peggiorate rispetto a prima, in quanto è diminuita l'attenuazione (che diventa α_{tot}/N come nel caso precedente), ma è anche diminuita la probabilità di errore richiesta. *Bisogna però considerare il modo con cui il rapporto S/N da ottenere all'uscita di ogni singola tratta varia al variare della probabilità di errore*: sappiamo, infatti, dallo studio della funzione $Q(\gamma)$, che una diminuzione della probabilità di errore di un fattore 10 equivale ad un aumento del rapporto S/N di 1 dB; se la diminuzione di $p(\epsilon)$ è di un fattore 100, l'aumento del rapporto S/N è dunque di 2dB; in generale, se la diminuzione di $p(\epsilon)$ è di un fattore N , l'aumento del rapporto S/N è dunque di $\log_{10} N$ dB. Questo comporta che la relazione da usare per il dimensionamento sia adesso

$$\log_{10} N + 14.5[\text{dB}] \leq \left(P_T[\text{dBm}] - \frac{\alpha_{\text{tot}}}{N}[\text{dBm}] \right) - \left(h_n \frac{f_s}{2} \right)_{\text{dB}}$$

dove il termine $\log_{10} N$, derivante appunto dalle considerazioni appena fatte, manca adesso del fattore moltiplicativo 10 che invece era presente prima, nel caso delle apparecchiature di pura amplificazione.

Risolvendo ancora una volta l'equazione per tentativi, si trova **N=29**, ossia una riduzione di ben 6 tratte rispetto al caso precedente.

Quindi, *il motivo per cui conviene notevolmente l'uso delle tratte rigenerative è che la probabilità di errore dipende in maniera molto critica dal rapporto segnale-rumore*: nel caso di tratte di pura amplificazione, di tratta in tratta si somma il rapporto tra rumore e segnale, per cui il rapporto S/N richiesto in ricezione va aumentato della quantità $10 \log_{10} N$; nel caso, invece, di tratte rigenerative, quello che si somma è la probabilità di errore, cui consegue che il rapporto S/N richiesto in ricezione vada aumentato della quantità $\log_{10} N$. Se ad ogni raddoppio di $p(\epsilon)$ corrispondesse un raddoppio anche del rapporto S/N, allora non cambierebbe niente rispetto alle tratte di pura amplificazione. Queste cose chiariscono

²⁵ il che equivale a dire che devono essere indipendenti i disturbi che, in ciascuna tratta, agiscono sul segnale provocando gli errori

l'osservazione, fatta all'inizio del capitolo, in cui si diceva che, mentre in un sistema analogico è necessario aumentare la potenza trasmessa di 10 dB per ottenere un incremento del rapporto S/N di 10 dB, in un sistema digitale questo non è invece necessario.

Detto anche in termini intuitivi, il vantaggio delle tratte rigenerative è che, alla fine di ogni tratta si ripulisce il segnale e si ritrasmette quindi un segnale esente da rumore (il rumore precedentemente sovrapposto si manifesta solo in termini di bit sbagliati). Questo si traduce in una minore probabilità di errore.

Autore: **SANDRO PETRIZZELLI**
e-mail: sandry@iol.it
sito personale: <http://users.iol.it/sandry>
succursale: <http://digilander.iol.it/sandry1>