

Alessandro Falaschi

Elementi di

Trasmissione dei Segnali e Sistemi di Telecomunicazione

Edizione .93

Roma, Dicembre 2001

a mia madre

© Alessandro Falaschi 2000-2001

- Editto presso <http://infocom.ing.uniroma1.it/alef/libro>
- Il presente testo è liberamente utilizzabile ad uso personale, e per questo è presente in rete.
- Se ne diffida viceversa la riproduzione, in qualunque forma, se non accordata dall'autore.
- L'uso integrale del testo, o di sue parti, come un ausilio alla didattica, come riferimento per la formazione del personale, o come materiale per la preparazione di corsi, necessita del preventivo consenso dall'autore.

Questo testo nasce come supporto ad un corso di telecomunicazioni per la laurea in Ingegneria Elettrica presso l'Università di Roma I, ed essendo l'unico corso di TLC per quella laurea, i suoi contenuti spaziano su molti temi, trattati altrove in diversi esami specifici. L'accento esplicito del titolo sia ai *Segnali* che ai *Sistemi*, oltre a rispecchiarne l'impostazione complessiva, ne caratterizza l'uso anche nel contesto delle nuove Lauree in due livelli. Sono infatti presenti elementi di *segnali, sistemi, probabilità, campionamento, quantizzazione, elaborazione, traffico, moltiplicazione, commutazione, trasmissione dati, reti, modulazione analogica e numerica, trasferimenti energetici, collegamenti radio, in cavo ed su fibra ottica, rumore*.

A fronte della molteplicità degli argomenti, l'esposizione adotta uno *stile chiaro, conciso ed esauriente*, ricorrendo ad una stesura su due livelli. Il primo è orientato ad una lettura più veloce, e attribuisce la giusta rilevanza a concetti, argomenti e tecniche; il secondo livello si sviluppa invece attraverso note ed appendici, costituendo l'approfondimento del primo, isolando gli aspetti più rigorosi, e fornendo esemplificazioni e chiarimenti. Ovunque, si tenta di mantenere vivo l'interesse del lettore grazie a *numerosi illustrazioni* a sostegno dei concetti, presentando i grafici delle relazioni analitiche, ed evidenziando al meglio le analogie ed i collegamenti tra gli argomenti. I sistemi di TLC stessi sono a volte presentati alla stregua di esempi applicativi della teoria esposta, guadagnando in concretezza e stimolando allo studio.

Il testo è rimasto per più di un anno disponibile per il download gratuito presso <http://infocom.uniroma1.it/alef/libro>, totalizzando nel periodo oltre 2000 scaricamenti. Nel frattempo si è provveduto, anche su indicazione dei lettori, alla correzione degli errori ed al completamento di alcune parti, ottenendo come risultato una qualità editoriale più che buona, e soprattutto un testo sostanzialmente esatto. Lo stesso indirizzo internet continuerà ad ospitare, oltre ad una *versione navigabile* del testo, anche le eventuali ulteriori correzioni, nonché le sezioni che si aggiungeranno nel tempo.

L'ampio spettro di argomenti trattati, associato ad una notazione tendenzialmente uniforme, rendono il testo idoneo a costituire un riferimento di base ideale sia per lo studente, sia per lo studioso che desideri comprendere i principi operativi degli apparati di telecomunicazione.

Alessandro Falaschi

è nato nel gennaio 1959. Nel 1983 si laurea con lode in Ingegneria Elettronica all'Università di Roma I, ove nel 1988 consegue il Dottorato di Ricerca in Ingegneria dell'Informazione e della Comunicazione. Dal 1992 al 95 è ricercatore presso il Dipartimento di Elettronica dell'Università di Perugia, dove svolge didattica per il corso di Reti di Telecomunicazione; ancora come ricercatore, torna all'Università di Roma I, presso il Dipartimento INFO-COM, dove è docente del corso di Comunicazioni Elettriche per la laurea in Ingegneria Elettrica. Si è occupato di elaborazione del segnale vocale e di trasmissioni numeriche, ed attualmente i suoi interessi sono orientati ai sistemi di Voice over IP ed alle architetture di audio-video conferenza di gruppo su rete internet.

Capitolo 1

Generalità

Effettuiamo subito una velocissima panoramica dei molteplici aspetti che intervengono nei sistemi di telecomunicazione, che oltre al semplice inoltro per via *elettrica* di un *messaggio informativo* da un luogo ad un altro, coinvolgono un discreto numero di apparati differenti e cooperanti, nel contesto di una organizzazione *in rete* dei dispositivi.

La trasmissione può riguardare un messaggio generato al tempo stesso della sua trasmissione, oppure esistente a priori. Il supporto fisico del messaggio, chiamato *segnale*, identifica due categorie molto generali: quella dei segnali *analogici*, e quella dei segnali *numerici*. Nel primo caso rientra ad esempio la voce umana, mentre esempi di segnali numerici sono i documenti conservati su di un computer.

1.1 Trasmissione dell'informazione

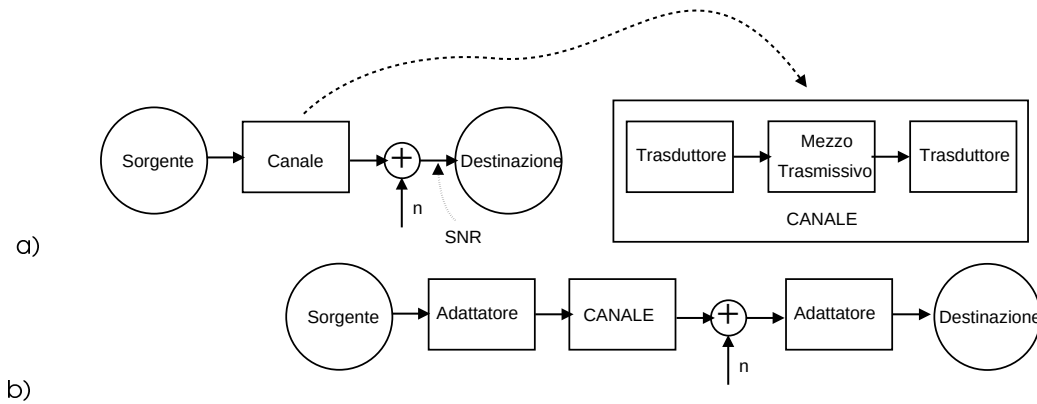
Sorgente, destinatario e canale L'origine del segnale da trasmettere è indicata (vedi Fig. 1.1) come *sorgente*, di tipo analogico o numerico per i due tipi di segnale. Ciò che giace tra sorgente e *destinatario* viene descritto da una entità astratta denominata *canale* di comunicazione, le cui caratteristiche condizionano i messaggi trasmessi.

Distorsioni e disturbi Il canale può ad esempio imporre una limitazione alla *banda di frequenze* del segnale in transito¹; cause fisiche ineliminabili producono inoltre, al lato ricevente, l'insorgere di un segnale di disturbo additivo, comunemente indicato con il termine di *rumore*, che causa la ricezione di un segnale diverso da quello stesso presente all'uscita del canale. Pertanto, ci si preoccupa di caratterizzare il canale in modo da scegliere i metodi di trasmissione più idonei a rendere minima l'alterazione sul messaggio trasmesso.

Rapporto segnale-rumore L'entità delle alterazioni subite dal messaggio viene spesso quantificata in termini di *rapporto segnale rumore* (SNR o SIGNAL-TO-NOISE RATIO), che rappresenta un indice di qualità del collegamento stesso, e che per ora definiamo genericamente come il rapporto tra l'entità del segnale utile ricevuto e quello del rumore ad esso sovrapposto, indicato con n nella figura 1.1.

Trasmissione La Fig. 1.1a) evidenzia come il canale, nella realtà fisica, è costituito da un *mezzo trasmissivo* su cui si propaga un segnale di natura elettrica, che

Figura 1.1: Elementi funzionali per la trasmissione dell'informazione



viene convertito in tale forma da appositi *trasduttori* di trasmissione e ricezione². Considerando per il momento i trasduttori come facenti parte del canale stesso, proseguiamo l'analisi concentrandoci sugli ulteriori aspetti del processo di comunicazione.

Adattatori La figura 1.1b) evidenzia l'esistenza (in trasmissione, in ricezione od a entrambe le estremità) di dispositivi *adattatori*, che hanno lo scopo di ridurre od eliminare le cause di deterioramento del messaggio introdotte dalla trasmissione: si può ad esempio ricorrere ad *equalizzatori* per correggere la risposta in frequenza di un canale, ad *amplificatori* per contrastare l'attenuazione subita dal segnale, ovvero a *codificatori di linea* per rendere le caratteristiche del segnale idonee ad essere trasmesse sul canale a disposizione.

Rete La trasmissione lungo un canale in uso esclusivo alla coppia sorgente-destinazione è piuttosto raro; di solito i collegamenti sono condivisi tra più comunicazioni, ognuna con differente origine e destinatario. Il problema della condivisione delle risorse trasmissive, ed il coordinamento di queste attività, produce la necessità di analizzare in modo esplicito le *reti di telecomunicazione*, che entrano a far parte integrante dei sistemi di trasmissione dell'informazione.

Gli aspetti delle telecomunicazioni brevemente accennati, sono immediatamente applicabili a segnali di natura *analogica*, in cui il segnale è definito per tutti gli istanti di tempo, ed assume valori qualsiasi. Nel caso invece in cui il segnale è definito solo per istanti di tempo discreti e valori discreti, si entra nell'ambito delle *trasmissioni numeriche*.

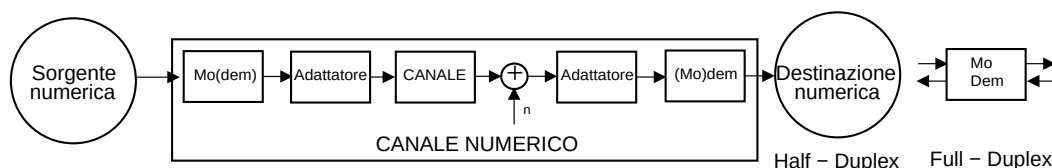
1.2 Trasmissioni numeriche

Modem Qualora si desideri trasmettere un segnale *numerico*, questo deve in generale essere convertito in un segnale analogico mediante l'utilizzo di dispositivi chiamati

¹Approfondiremo nel seguito il senso di questa locuzione; per ora è sufficiente interpretarla in termini generici, ovvero di fedeltà della riproduzione al segnale originario.

²Un classico esempio di trasduttore è quello dell'antenna, nel caso di trasmissione radio.

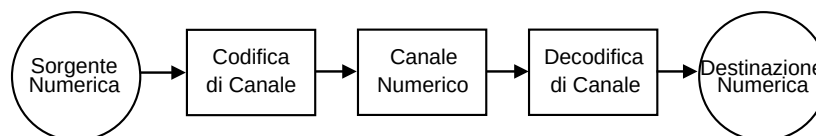
*Modem*³, come rappresentato dalla figura seguente, in cui è evidenziato come per una trasmissione unidirezionale⁴ occorra solo *metà* delle funzioni del modem per entrambi i lati del collegamento, mentre nel caso di collegamento *full duplex* (in cui entrambi gli estremi possono essere contemporaneamente sorgente e destinazione) il modem opera allo stesso tempo nelle due direzioni.



Canale Numerico La figura precedente suggerisce come sia possibile racchiudere tutto ciò che è compreso tra i due modem in un *unico* blocco, denominato *canale numerico*. Quest'ultimo è concettualizzato come una entità autonoma, e nel progetto di una sistema di comunicazione numerica è caratterizzato da un *fattore di qualità* individuato dalla ...

Probabilità di errore che è definita come la frequenza con cui *i simboli ricevuti differiscono* da quelli trasmessi, a causa di un errore verificatosi nel modem ricevente, e causato a sua volta dal rumore additivo presente in uscita dal canale analogico, e/o dalle alterazioni introdotte dal canale. Evidentemente, le *prestazioni* individuate dalla probabilità di errore sono strettamente legate a quelle individuate dal rapporto segnale-rumore per il canale analogico sottostante.

Codifica di Canale Nelle trasmissioni numeriche, si può introdurre una *ridondanza* nella sequenza trasmessa, inviando più simboli di quanti non ne produca la sorgente, e quindi di fatto *aumentando* il numero di simboli da trasmettere per unità di tempo; i simboli in più sono scelti in modo da essere in qualche modo *dipendenti* tra loro, e questa loro caratteristica rende possibile la riduzione della probabilità di errore di cui soffre il canale numerico. Infatti, grazie alla dipendenza (nota) tra i simboli trasmessi, il ricevitore è ora in grado di "accorgersi" che si è verificato un errore, in quanto la dipendenza prevista non è più rispettata; pertanto, il ricevitore può attuare delle contromisure. La ridondanza introdotta può essere così elevata da permettere la correzione di errori isolati⁵, oppure il ricevitore può semplicemente richiedere la *ritrasmissione* del simbolo errato. Le trasformazioni del segnale ora descritte prendono il nome di *codifica di canale*, e devono essere "rimosse" all'uscita dello stesso da un processo inverso di *decodifica*.

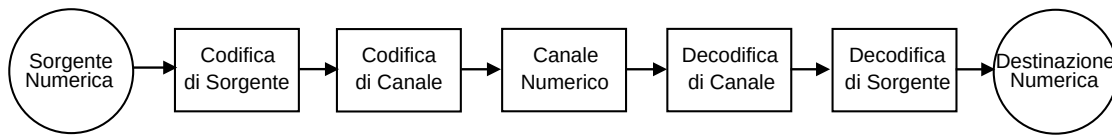


³La parola *Modem* è una contrazione delle due parole *modulatore-demodulatore*.

⁴Nelle trasmissioni unidirezionali, sorgente e destinazione non si scambiano i ruoli. La trasmissione stessa viene anche indicata con il termine di *half-duplex*.

⁵Si parla in questo caso di codifica FEC, ovvero di *Forward Error Correction*.

Codifica di Sorgente Possono essere introdotti due ulteriori blocchi, che operano una *codifica* (e relativa de-codifica) *di sorgente* sulla sequenza trasmessa, con uno scopo è per così dire “inverso” a quello della codifica di canale: infatti, la codifica di sorgente *rimuove le dipendenze* tra i simboli presenti nelle sequenze generate dalla sorgente, ottenendo di fatto una riduzione del numero di simboli da trasmettere per unità di tempo⁶. Un tipico esempio di codifica di sorgente è rappresentato dagli algoritmi di compressione esistenti per i file di computer (come i file *zippati*); in tal caso, il fattore di compressione ottenibile dipende dalla natura del file trattato, ed è tanto maggiore quanto più quest’ultimo presenta caratteristiche di ripetitività e quindi di predicibilità del suo contenuto. In altre parole, l’uscita di un codificatore di sorgente è una sequenza di simboli tendenzialmente *indipendenti* tra loro, nel senso che ogni forma di predicibilità di un simbolo a partire dai circostanti è stata rimossa.



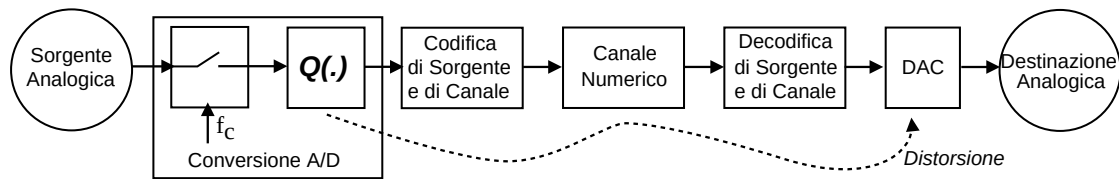
Campionamento e Quantizzazione Poniamoci ora il problema di utilizzare un *canale numerico* per effettuare una *trasmissione analogica*. Il vantaggio di tale “contorsione” è da ricercarsi nel migliore comportamento delle trasmissioni numeriche rispetto ai disturbi, nonché alla loro *generalità*⁷. Per ottenere il risultato desiderato, occorre applicare alla sorgente analogica un procedimento di *campionamento*, prelevandone i valori ad istanti discreti, e quindi di *quantizzazione*, rappresentando tali valori mediante un insieme finito di simboli. Il risultato è una sequenza numerica che può essere di nuovo convertita nel segnale originario, utilizzando un dispositivo di conversione *digitale-analogica* (DAC) dal lato del ricevitore. Esempi pratici di quest’ultimo processo sono ben noti, come ad esempio nel caso dei CD audio.

Rumore di quantizzazione La riduzione dei valori campionati nell’ambito di un insieme finito produce una *ulteriore distorsione*, che può essere pensata sommarsi in ricezione al segnale originario, producendo una nuova fonte di degrado. L’entità del rumore dovuto alla quantizzazione è inversamente legata alla *risoluzione del quantizzatore*, ovvero alla capacità di differenziare tra valori di ingresso molto vicini tra loro. In definitiva, la distorsione risulta tanto minore, quanto maggiore è il flusso informativo⁸ prodotto dal quantizzatore, espresso in bit/secondo.

⁶Pensiamo per similitudine ad un imballaggio, il cui contenuto è prima disposto in modo da occupare il minimo volume (codifica di sorgente), ed a cui viene poi aggiunto del materiale antiurto (codifica di canale).

⁷Nei collegamenti numerici, non occorre specializzare il collegamento allo specifico mezzo trasmissivo a disposizione, anzi quest’ultimo è totalmente “mascherato” dal fornitore del collegamento numerico stesso.

⁸Vedremo infatti nel seguito che un aumento della risoluzione del processo di quantizzazione corrisponde ad un aumento del numero di bit necessari a rappresentare ogni valore (o campione) di segnale.



Teoria velocità-distorsione Un teorico dell'informazione, CLAUDE SHANNON, enunciò negli anni '50 una serie di teoremi, che sono la base dell'analisi dei sistemi di comunicazione. Non è oggetto del presente testo approfondire tali aspetti; è comunque istruttivo accennare ai principali risultati, che possono essere riassunti nei seguenti tre enunciati:

- Un qualsiasi canale pone un limite al massimo flusso informativo che transita in esso. Il limite deriva dai vincoli che il canale impone sulla massima banda $\mathbf{B}$ del segnale in transito, sulla massima potenza di segnale $\mathbf{S}$ ricevuta, e sulla potenza di rumore $\mathbf{N}$ presente al ricevitore. Il massimo flusso di informazione in transito prende il nome di *capacità di canale* $\mathbf{C}$, e può essere espresso come $C = B \log_2 \left(1 + \frac{S}{N} \right)$ bit/sec. In questi termini, la massima velocità di trasmissione in un canale può dipendere da una limitazione sulla banda $\mathbf{B}$, o sulla potenza, od essere causata da un eccessivo rumore: solo alterando uno di questi fattori (scegliendo un diverso canale), è possibile aumentare la velocità di trasmissione.
- Una qualsiasi sorgente produce un flusso informativo in bit/secondo tanto più elevato quanto minore è la distorsione introdotta dal processo di quantizzazione.
- Considerando una coppia sorgente + canale, dato che il canale limita il massimo flusso informativo prodotto dalla sorgente, quest'ultima verrà necessariamente riprodotta con una distorsione tanto maggiore quanto minore è la capacità di canale. A meno di non impiegare più tempo per la trasmissione.... oppure di cambiare canale (ad esempio usandone uno con una banda maggiore).

1.3 Segnali analogici, certi ed aleatori

I segnali analogici, indicati con $s(t)$ (con le parentesi tonde), rappresentano l'*andamento nel tempo* di una grandezza elettrica. Come esempio possiamo citare il *segnale vocale*, in cui un'onda trasversale di pressione-velocità è convertita in una tensione da un microfono. Oppure citare un *segnale di immagine*, che è bidimensionale, e definito su di un piano anziché nel tempo, rappresentato da una grandezza $S(x, y)$ che ne individua la luminanza, e scandito per linee generando un segnale temporale. Un segnale può anche assumere *valori complessi*⁹, in questo caso il segnale assume contemporaneamente due diversi valori (parte reale e parte immaginaria, oppure modulo e fase).

E' importante distinguere tra i segnali cosiddetti *certi* e quelli *aleatori*. Un esempio di *segnale certo* può essere una cosinusoide di cui sia nota sia l'ampiezza che la fase, mentre un *segnale aleatorio* non è noto con esattezza prima che questo venga prodotto

⁹Come vedremo, un segnale a valori complessi è il risultato di una particolare rappresentazione, detta *inviluppo complesso*, utile nell'analisi dei segnali modulati.

(ad esempio il rumore di un ruscello, o le notizie presenti in un telegiornale). L'insieme di tutti i segnali aleatori appartenenti ad una medesima classe viene indicato nel suo complesso come *processo aleatorio*, ed un segnale particolare di questo insieme come una sua *realizzazione*.

1.3.1 Rappresentazione di segnali analogici

Lo studio delle proprietà dei segnali si articola prendendo in considerazione per gli stessi rappresentazioni alternative, scelte in modo da poter valutare più agevolmente le alterazioni subite dai segnali nel passaggio attraverso sistemi fisici. In particolare, sarà definito lo *sviluppo in serie di Fourier* per la rappresentazione dei segnali periodici, e quindi la *trasformata di Fourier* che descrive una classe più ampia di segnali.

L'analisi di Fourier consente di definire il concetto di *banda occupata* da un segnale, nonché di come la sua potenza e/o energia si *distribuisce in frequenza*; quest'ultimo andamento viene indicato con il termine di *Spettro di Densità di Potenza* (o di *Energia*).

1.3.2 Rappresentazione di processi aleatori

Anche nel caso in cui il segnale non è noto a priori, e dunque è impossibile calcolarne la trasformata di Fourier in forma chiusa, si può ugualmente giungere ad una rappresentazione che caratterizzi le realizzazioni del processo nei termini della distribuzione (statistica) in frequenza della potenza di segnale.

Ciò è possibile considerando la *funzione di autocorrelazione*, che esprime il grado di interdipendenza statistica tra i valori assunti in istanti diversi dalle realizzazioni del processo, e che costituisce un elemento unificante ai fini della *stima spettrale* dei segnali.

Osserveremo come processi molto correlati siano caratterizzati da una densità di potenza di tipo *colorato*, mentre processi scarsamente correlati saranno identificati da una densità di potenza di tipo *bianco*¹⁰.

1.3.3 Transitio dei segnali attraverso sistemi fisici

Saranno forniti i metodi di valutazione dei *peggioramenti* indotti su di un segnale che transiti in un mezzo trasmissivo, e di come riuscire a *ridurli* od *eliminarli*.

Risposta impulsiva e convoluzione Un concetto di base è l'integrale di *convoluzione*, per mezzo del quale si esprime in forma chiusa l'uscita di un circuito elettrico¹¹ in base alla conoscenza dell'ingresso, e di una particolare caratteristica della rete, la *risposta impulsiva*. Quest'ultima rappresenta l'uscita della rete quando in ingresso è presente una particolare funzione analitica, detta *impulso matematico*.

Risposta in frequenza Operando nel dominio della frequenza, osserveremo come la trasformata di Fourier della risposta impulsiva rappresenti la *risposta in frequenza* della rete, ovvero l'uscita alle diverse frequenze quando l'ingresso ha uno spettro bianco.

¹⁰I termini *colorato* e *bianco* hanno origine da una similitudine con l'energia luminosa, per cui se la luce bianca indica l'indiscriminata presenza di tutte le lunghezze d'onda, così uno spettro bianco indica la presenza in egual misura di tutte le frequenze; viceversa, come una luce colorata dipende dal prevalere di determinate frequenze nella radiazione elettromagnetica, così uno spettro colorato indica la prevalenza di alcune frequenze su altre.

Modulazione Nel caso in cui il segnale da trasmettere occupi una banda concentrata attorno ad una frequenza più o meno elevata (detta portante), come nel caso dei *segnali modulati*, si ricorre alla rappresentazione mediante le *componenti analogiche di bassa frequenza*. Il caso opposto, caratterizzato da una estensione frequenziale contigua alla frequenza zero, è detto di *banda base*. L'uso dei segnali modulati è obbligatorio, qualora questi debbano essere trasmessi su canali di tipo cosiddetto *passa-banda*.

La trasmissione dei segnali mediante un sistema di comunicazione coinvolge diversi altri aspetti, che sono brevemente introdotti al § 1.6, nel contesto della caratterizzazione dei canali di comunicazione.

1.4 Segnali numerici

Sono indicati con la notazione $s[k]$ (con le parentesi quadre), per evidenziare che il loro dominio è l'insieme dei numeri interi. Sono valide le stesse definizioni fornite al § 1.7.2 a riguardo dei segnali analogici, relativamente ai concetti di potenza, energia e periodicità, utilizzando qui delle sommatorie in luogo degli integrali.

Sequenze Un segnale viene chiamato numerico quando assume valori appartenenti ad un *insieme finito di simboli*; per questo motivo, la sua essenza è indicata anche come *sequenza simbolica*. Ad esempio, un testo scritto assume valori nell'ambito dei caratteri stampabili. Se si rappresenta ogni carattere con il suo numero ordinale, si ha allora una vera sequenza di numeri.

Segnali tempo-discreti Si può alternativamente rappresentare ogni carattere con un diverso valore di tensione, ottenendo un segnale analogico che è una rappresentazione a *più livelli* di tensione della sequenza originaria.

Frequenza di simbolo Il concetto di occupazione di banda, applicabile ai segnali analogici, è qui sostituito da quello di velocità di emissione, espressa in *simboli/secondo*, ed indicata come *frequenza di simbolo*. Una sequenza prodotta da una sorgente numerica si presta facilmente ad essere trasformata in un'altra, con un diverso alfabeto ed una differente frequenza di simbolo¹².

Frequenza binaria Qualora si desideri ottenere una *trasmissione binaria*, ossia rappresentabile come una sequenza di *zeri* ed *uni*, l'alfabeto di rappresentazione ha cardinalità pari a 2. In tal caso, ogni simbolo L -ario di ingresso produrrà gruppi di M caratteri binari (o bit), con M pari all'intero superiore di $\log_2 L$ bit/simbolo. Questa ultima grandezza M , pari al numero di bit/simbolo, moltiplicata per il numero di simboli a secondo f_s , permette di calcolare il flusso informativo in bit/secondo, che prende il nome di *frequenza binaria*: $f_b = M \cdot f_s$.

Campionamento Si è già accennato a come un segnale analogico possa essere rappresentato mediante i suoi valori campionati a frequenza di f_c *campioni/secondo* e

¹²Quando un circuito elettrico ha la funzione di trasportare un segnale tra una coppia di morsetti ad un'altra, il circuito prende il nome di *rete due porte* o *quadripolo*.

quantizzati con un numero M di *bit/campione*, consentendo l'uso di un canale numerico. In tal caso, la sorgente numerica equivalente sarà caratterizzata da una velocità di trasmissione di f_b bit/secondo, pari al prodotto $f_c \cdot M$.

Modulazione numerica Qualora la risposta in frequenza del canale imponga un processo di modulazione, esistono tecniche *specifiche* per i segnali numerici, che traggono vantaggio dalla natura discreta del messaggio da trasmettere.

Trasmissione a pacchetto Un segnale numerico può avere origini delle più disparate, e non necessariamente essere il risultato di un processo di quantizzazione. Ad esempio, può trattarsi di un file da trasmettere tra due computer; in tal caso, si può frammentare la sequenza numerica in messaggi più piccoli (chiamati *pacchetti di dati*), numerarli consecutivamente, ed inviarli singolarmente attraverso la rete di interconnessione, anche impiegando percorsi differenti per ogni sotto-messaggio: sarà compito del lato ricevente ri-assemblare i singoli pacchetti nell'ordine originario. Il caso descritto è un tipo particolare di *rete a commutazione di pacchetto*, di cui saranno esposti i principi di funzionamento e le metodologie di progetto di massima delle risorse, mirate all'ottenimento di *prestazioni* definite in termini di *ritardo medio* di trasmissione.

1.5 Teoria delle probabilità

Molti dei concetti utilizzati per trattare i processi aleatori, per definire la quantità di informazione di un messaggio, le prestazioni di un canale, il dimensionamento di reti di comunicazione, sono fondati sulla conoscenza della *teoria delle probabilità*, che verrà pertanto illustrata, almeno nei suoi concetti fondamentali, i quali saranno immediatamente applicati ai casi specifici che si verificano nei sistemi di telecomunicazione.

Teoria del traffico In particolare, è accennata l'applicazione della teoria delle probabilità al problema del dimensionamento di collegamenti che debbano trasportare più messaggi contemporaneamente, operando una moltiplicazione degli stessi su di un medesimo mezzo trasmissivo.

1.6 Sistemi di Telecomunicazione

Introduciamo brevemente quattro diversi *punti di vista* in cui è possibile inquadrare le problematiche di comunicazione: gli aspetti *fisici*, di *elaborazione*, di *sistema*, di *rete*, e di *trasporto*.

Aspetti fisici Un canale di comunicazione, *dal punto di vista fisico*, si identifica con il mezzo trasmissivo, per la descrizione del quale si adotta frequentemente un modello circuitale. Elenchiamo i mezzi comunemente adottati:

¹²Per fissare le idee, consideriamo i simboli di una sequenza numerica $s[k]$ ad L valori: questi possono essere presi a gruppi di K , producendo simboli a velocità K volte inferiore, ma con L^K valori distinti. Se si dispone di un alfabeto di uscita ad H valori, i gruppi di K simboli L -ari originari possono essere rappresentati con gruppi di M simboli H -ari purché $L^K \leq H^M$. Es.: per codificare in binario ($H = 2$) simboli con $L = 26$ livelli, occorrono almeno $M = 5$ bit/simbolo, ottenendo così $2^5 = 32 > L = 26$.

Collegamenti radio: il segnale si propaga nello spazio libero come onda elettromagnetica sferica, e viene irradiato mediante antenne, che ne focalizzano la potenza lungo direzioni privilegiate. La trasmissione è resa possibile grazie al processo di modulazione;

Collegamenti in cavo: da quelli tra computer e stampante, a quelli su doppino (telefonia), a quelli in cavo coassiale (televisione, ethernet). Possono essere di tipo half o full duplex a seconda che i due estremi della comunicazione siano unidirezionali o bidirezionali;

Collegamenti in fibra ottica: sono realizzati facendo viaggiare energia luminosa attraverso una guida d'onda di materiale dielettrico. La tecnica è idonea alla trasmissione dei soli segnali numerici, dato che la sorgente luminosa in trasmissione viene accesa e spenta velocissimamente in corrispondenza dei bit (zero od uno) del messaggio.

Modello circuitale Il collegamento ed i trasduttori ad esso relativi, sono spesso realizzati ricorrendo ad un circuito elettrico equivalente, in modo da poterne descrivere il comportamento mediante strumenti analitici noti.

Aspetti sistemistici Da un punto di vista *sistemistico*, il transito dei segnali attraverso sistemi fisici è analizzato in termini del peggioramento introdotto, che può essere catalogato nell'ambito di diverse categorie:

Distorsioni Si distinguono quelle cosiddette *lineari*, causate da una risposta in frequenza non ideale, dalle distorsioni *non lineari*, che causano invece una deformazione istantanea sulla forma d'onda in transito;

Non stazionarietà Questi fenomeni sono caratterizzati da una variazione nel tempo delle caratteristiche del canale, e ricorrono spesso nel caso di comunicazioni con mezzi mobili;

Attenuazione Un segnale in transito lungo un canale presenta in uscita una ampiezza inferiore a quella di ingresso. L'alterazione può aver luogo sia per cause fisiche intrinseche (lunghezza del collegamento, disadattamento di impedenze, tecnologia degli amplificatori), che in dipendenza di fatti contingenti (percorsi multipli, pioggia); in questo secondo caso, il fenomeno è trattato come l'esito di un processo aleatorio;

Portata Affinché possano essere soddisfatti i requisiti di qualità (ad esempio l'SNR) desiderati, risulta che la lunghezza del collegamento deve essere inferiore ad un massimo, in conseguenza dell'attenuazione del collegamento, della potenza trasmessa, e degli altri fatti contingenti;

Qualità del servizio Con questo termine sono indicate diverse grandezze, ognuna applicabile in un particolare contesto, e che rappresentano un indice di "bontà" del processo comunicativo. Tra queste grandezze possiamo citare il *rapporto segnale rumore* SNR e la *probabilità di errore* P_e , relative rispettivamente alle trasmissioni analogiche e numeriche; il *ritardo medio*, rilevante nel caso di trasmissioni a pacchetto; il *tempo di fuori servizio*, qualificante della affidabilità dei sottosistemi di comunicazione.

Rete Dal punto di vista della *rete di comunicazione*, la consegna del messaggio informativo alla destinazione deve tener conto di aspetti indicati come:

Commutazione La rete è costituita da un insieme di *nodi di commutazione*, interconnessi da collegamenti che vengono usati in modalità condivisa da molte comunicazioni contemporanee, e che sono *attraversati* dai messaggi in transito, che devono essere *smistati* verso la porta di uscita corretta;

Instradamento La determinazione del percorso dei messaggi nella rete, scelto tra i possibili percorsi che collegano la sorgente con la destinazione, prende il nome di *instradamento*;

Segnalazione Il coordinamento tra i nodi della rete avviene mediante lo scambio tra gli stessi di informazioni aggiuntive dette *di segnalazione*, che costituiscono un vero e proprio processo di comunicazione parallelo a quello prettamente informativo;

Protocollo Lo scambio dei messaggi di segnalazione tra le coppie di nodi della rete, od anche tra i nodi ed un organo di controllo centrale, avviene utilizzando particolari linguaggi, detti *protocolli* di segnalazione;

Elaborazione terminale Questa categoria comprende tutti gli aspetti legati alle trasformazioni operate sull'informazione ai due estremi del collegamento. L'importanza e la specificità di tali trasformazioni assume un rilievo sempre maggiore con l'evoluzione (in termini di miniaturizzazione e potenza di calcolo) dei dispositivi di elaborazione, in special modo per ciò che riguarda le trasmissioni numeriche.

Trasporto Dal punto di vista del *trasporto* dell'informazione, sono rilevanti gli aspetti di:

Multiplazione Si tratta di raggruppare tra loro le singole comunicazioni in transito per un tratto in comune; con il risultato di migliorare sensibilmente l'efficienza della rete. Infatti, mediante la multiplazione si può garantire un elevato tasso di utilizzo delle risorse, che non giacciono mai inutilizzate proprio grazie al ri-uso continuo e multiplo delle stesse per parecchie comunicazioni in transito. Tecniche di multiplazione comunemente adottate sono la *multiplazione ...*

- di *tempo*, in cui lo stesso collegamento è utilizzato per più comunicazioni contemporanee in base ad un uno schema di alternanza temporale;
- di *frequenza*, in cui le diverse comunicazioni occupano differenti regioni di frequenza, in uno stesso collegamento tra (ad esempio) due antenne;
- di *codice*, in cui diverse comunicazioni avvengono simultaneamente nella medesima banda di frequenza, adottando una particolare codifica che ne permette la separazione dal lato ricevente.

Controllo Riguarda la corretta consegna del messaggio al destinatario, e coinvolge principalmente la gestione degli errori di trasmissione di cui si è discusso a riguardo della codifica di canale, le problematiche di riassettablaggio delle

comunicazioni inoltrate in forma di pacchetti distinti, e la gestione della segnalazione per ciò che riguarda l'adattamento dei procolli di instradamento alle condizioni di carico della rete.

1.7 Segnali e sistemi

Vengono qui brevissimamente riassunte le definizioni ricorrenti nel descrivere gli elementi fondamentali nei cui termini sono descritti i sistemi di telecomunicazione.

In termini generali, un *sistema* è un gruppo di oggetti che interagiscono armoniosamente, e che sono combinati in modo da conseguire un obbiettivo desiderato. Un sistema può essere parte (sottosistema) di un sistema più grande, e si può definire una intera gerarchia di sistemi, ognuno con il proprio dominio.

Un *segnale* è un evento che determina l'inizio di qualche azione. Nel nostro caso, possiamo interessarci alla *risposta* di un sistema ad un dato segnale. A volte, un sistema è descritto unicamente in termini della sua risposta a determinati segnali.

1.7.1 Caratteristiche dei sistemi

Idealizziamo ora un sistema come una trasformazione $\mathcal{T}[\cdot]$, tale che ad ogni segnale di ingresso $x(t)$ corrisponda una uscita $y(t)$: $\mathcal{T}[x(t)] = y(t)$. In base a tale formalismo, riportiamo alcune caratteristiche dei sistemi, che ne descrivono il comportamento in termini più generali.

Linearità Un sistema è *lineare* quando l'uscita associata ad una combinazione lineare di ingressi, è la combinazione lineare delle uscite previste per ogni singolo ingresso: $\mathcal{T}[\sum_i a_i x_i(t)] = \sum_i a_i \mathcal{T}[x_i(t)]$. Al contrario, un legame ingresso-uscita *senza memoria*¹³ del tipo $y(t) = g(x(t))$, in cui $g(\cdot)$ è una generica funzione *non* lineare¹⁴ ... non è lineare !

Permanenza Un sistema è *permanente* (o stazionario) se l'uscita associata ad un ingresso traslato nel tempo, è la traslazione temporale dell'uscita che si avrebbe per lo stesso ingresso non traslato, ovvero: se $\mathcal{T}[x(t)] = y(t)$, allora $\mathcal{T}[x(t - \tau)] = y(t - \tau)$. Nel caso contrario, il sistema è detto tempo-variante.

Realizzabilità fisica E' detta anche *causalità*, perché determina l'impossibilità di osservare una uscita, prima di aver applicato un qualunque ingresso. Una definizione alternativa asserisce che i valori di uscita $y(t)$ ad un istante $t = t_0$, non possono dipendere da valori di ingresso $x(t)$ per $t > t_0$.

Stabilità è definita come la proprietà di fornire uscite limitate (in ampiezza) per ingressi limitati.

¹³Un operatore si dice *senza memoria* quando ogni valore dell'uscita dipende da un unico valore di ingresso.

¹⁴Una funzione è lineare quando è esprimibile in forma $y = ax + b$ (equazione di una retta).

1.7.2 Caratteristiche dei segnali

Da un punto di vista analitico, un segnale è una funzione del tempo, del tipo descritto in 1.3, e per il quale si possono operare le classificazioni:

Segnale di potenza Un segnale analogico può avere una estensione temporale limitata, oppure si può immaginare che si estenda da meno infinito a infinito. Nel secondo caso il segnale si dice di *potenza* se ne esiste (ed è diversa da zero) la media quadratica

$$0 < \mathcal{P}_s = \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} |s(t)|^2 dt < \infty$$

Un segnale di potenza è inoltre detto

Segnale periodico di periodo T , nel caso in cui si verifichi che

$$s(t) = s(t + T)$$

per qualsiasi valore di t , mentre si dice

Segnale di energia un segnale di durata limitata o illimitata, se esiste il valore

$$0 < \mathcal{E}_s = \int_{-\infty}^{\infty} |s(t)|^2 dt < \infty$$

Perché ciò avvenga, occorre che $s(t)$ tenda a zero (per t che tende ad ∞) più velocemente (od in modo uguale) ad $\frac{1}{\sqrt{t}}$ (e quindi $|s(t)|^2$ tenda a zero come $\frac{1}{t}$).

In particolare, se un segnale ha *durata limitata*, ovvero è nullo per t al di fuori di un intervallo $[t_1, t_2]$ (vedi Fig. a pagina seguente), allora è anche di energia. Infine, viene detto

Segnale impulsivo un segnale di energia, che tende a zero come (o più velocemente di) $\frac{1}{t}$:

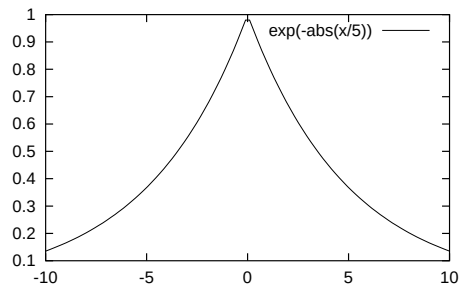
$$0 < \int_{-\infty}^{\infty} |s(t)| dt < \infty$$

E' il caso delle funzioni sommabili, per le quali $|s(t)|^2$ tende a zero come (o più di) $\frac{1}{t^2}$, e dunque di energia.

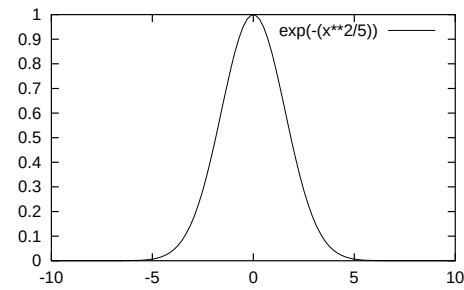
Riassumendo

- Un segnale *impulsivo* è di energia;
- Un segnale a *durata limitata* è impulsivo, e di energia;
- Un segnale *periodico* non è di energia, ma di potenza;

Esempi di segnali di energia

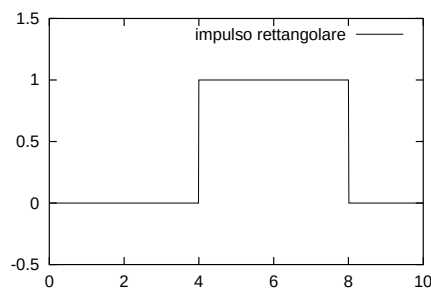


Impulso Esponenziale Bilatero

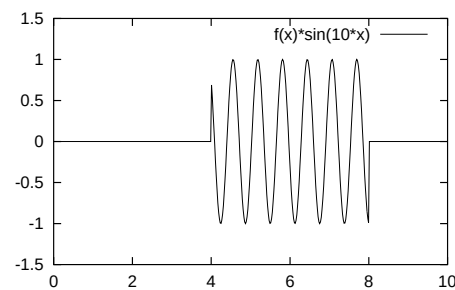


Impulso Gaussiano

Esempi di segnale a durata limitata



Impulso rettangolare tra 4 ed 8



Sinusoide troncata

Qualora il segnale sia associato a delle grandezze elettriche, allora i concetti di *Potenza* ed *Energia* hanno il correlato fisico illustrato di seguito.

1.7.3 Aspetti fisici delle grandezze energetiche

Potenza istantanea Se consideriamo una resistenza R , ed applichiamo ai suoi capi una tensione $v(t)$, in essa scorre una corrente $i(t) = \frac{v(t)}{R}$, e la *potenza ceduta* alla resistenza ad ogni istante t è pari a

$$p(t) = v(t) i(t)$$

che si misura in *Watt* (equivalente a Joule/secondo), e che rappresenta la *potenza istantanea* assorbita. Ricordando che $i(t) = \frac{v(t)}{R}$, si ottiene anche $p(t) = \frac{v^2(t)}{R} = i^2(t) R$.

Energia Se integriamo $p(t)$ su di un intervallo temporale T , si ottiene *l'energia complessiva* assorbita da R nell'intervallo T :

$$e_T(t) = \int_{t-T}^t p(\tau) d\tau \quad [\text{joule}]$$

Nello stesso intervallo T , la resistenza *assorbe* una potenza $p_T(t) = \overline{e_T(t)} = \frac{1}{T}e_T(t)$ [Watt], che costituisce una *media a breve termine* dell'energia assorbita nell'intervallo¹⁵.

Se un segnale $x(t)$ è *periodico* con periodo T (o $\frac{T}{n}$ con n intero), i valori di $\overline{e_T(t)} = p_T(t)$ coincidono con quelli calcolabili con T comunque grande. Se $R = 1\Omega$, tali valori coincidono inoltre con le definizioni di potenza ed energia del segnale:

$$\text{Energia: } \mathcal{E}_x = \int_{-\frac{T}{2}}^{\frac{T}{2}} |x|^2(t) dt = e_T\left(\frac{T}{2}\right) \quad [\text{Volt}^2 \cdot \text{sec}] \text{ oppure } [\text{Ampere}^2 \cdot \text{sec}]$$

$$\text{Potenza: } \mathcal{P}_x = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x|^2(t) dt = \frac{1}{T}e_T\left(\frac{T}{2}\right) = p_T\left(\frac{T}{2}\right) \quad [\text{Volt}^2] \text{ oppure } [\text{Ampere}^2]$$

Potenza dissipata Se la resistenza è diversa da 1Ω , le due quantità non coincidono più. Nelle misure fisiche in genere si ottiene la potenza *dissipata* sullo strumento di misura (o irradiata dall'antenna, o dagli altoparlanti) espressa in Watt. Per risalire alla potenza/energia *di segnale* delle grandezze elettriche presenti ai suoi capi (tensione o corrente) occorre dividere (o moltiplicare) la potenza in Watt per R . Ad esempio, una potenza assorbita $\mathcal{P}$ di 10 Watt su 8 Ohm fornisce una potenza di segnale $\mathcal{P} \cdot R = 80 (\text{Volt})^2$, ovvero di $\frac{\mathcal{P}}{R} = 1.25 (\text{Ampere})^2$.

Valore efficace Si indica allora come *valore efficace* quel livello di segnale continuo che produrrebbe lo stesso effetto energetico. Nell'esempio precedente, otteniamo: $V_{eff} = \sqrt{80} = 8.94 \text{ Volt}$; $I_{eff} = 1.118 \text{ Ampere}$. Infatti:

$$\mathcal{P}_T(\text{segnale}) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} V_{eff}^2 dt = \frac{T}{T} (8.94)^2 = 80 \text{ Volt}^2 \text{ che su } 8\Omega \text{ dissipa } \frac{V^2}{R} = 10 \text{ Watt.}$$

¹⁵Anticipando una notazione che verrà usata nel corso del testo, il pedice $_T$ indica l'estensione temporale a cui è riferita la grandezza che presenta il pedice, mentre la sopralineatura di una grandezza che dipende dal tempo, indica una media temporale della grandezza stessa.

Capitolo 2

Serie di Fourier

Sono qui impostate alcune relazioni trigonometriche fondamentali, per poi definire la rappresentazione di segnali periodici mediante lo sviluppo in serie di Fourier. Sono quindi introdotte le proprietà di simmetria e di approssimazione, e sviluppati i concetti di ortogonalità. E' infine enunciato il teorema di Parseval e definito lo spettro di potenza. Il capitolo termina con richiami di algebra vettoriale.

2.1 Prerequisiti trigonometrici

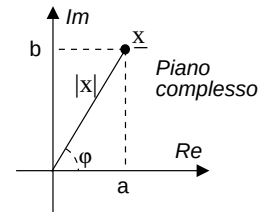
2.1.1 Numeri complessi

Un numero complesso $\underline{x}$ è costituito da una coppia di valori numerici a e b che ne rappresentano la parte reale e quella immaginaria:

$$\underline{x} = a + jb$$

E' spesso utile ricorrere ad una rappresentazione di $\underline{x}$ nel piano complesso, che mette in luce l'espressione alternativa¹ di $\underline{x}$ nei termini di modulo $|x|$ e fase φ :

$$\underline{x} = |x| e^{j\varphi}$$



Queste due quantità si ottengono dalle parti reale ed immaginaria, mediante le relazioni

$$|x| = \sqrt{a^2 + b^2} \quad \text{e} \quad \varphi = \arctan \frac{b}{a}$$

mentre le relazioni inverse risultano

$$a = |x| \cos \varphi \quad \text{e} \quad b = |x| \sin \varphi$$

Per ogni numero complesso $\underline{x}$, è definito il suo coniugato $\underline{x}^*$ come quel numero complesso con uguale parte reale, e parte immaginaria di segno opposto, ovvero uguale modulo, e fase cambiata di segno: $\underline{x}^* = a - jb = |x| e^{-j\varphi}$.

¹La rappresentazione in modulo e fase consente di calcolare il prodotto tra numeri complessi (es $\underline{x} = |x| e^{j\varphi}$ e $\underline{y} = |y| e^{j\theta}$) in modo semplice: $\underline{y} = \underline{x} \cdot \underline{y} = |x| |y| e^{j(\varphi+\theta)}$.

2.1.2 Formule di Eulero

L'esponenziale $e^{j\varphi}$ è un particolare numero complesso con modulo pari ad uno², e che quindi si scompone in parte reale ed immaginaria come

$$e^{\pm j\varphi} = \cos \varphi \pm j \sin \varphi$$

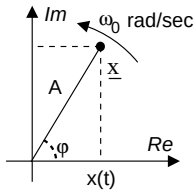
Da questa relazione sono derivabili le *formule di Eulero*, che esprimono le funzioni trigonometriche in termini di esponenziali complessi come

$$\cos \varphi = \frac{e^{j\varphi} + e^{-j\varphi}}{2} \quad e \quad \sin \varphi = \frac{e^{j\varphi} - e^{-j\varphi}}{2j}$$

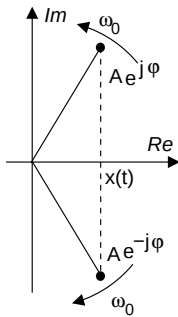
e che possono tornare utili nel semplificare i calcoli, trasformando i prodotti tra funzioni trigonometriche in somme di angoli³.

2.1.3 Fasori

Un segnale del tipo $x(t) = A \cos(2\pi f_0 t + \varphi)$ è completamente rappresentato dal numero complesso $\underline{x} = Ae^{j\varphi}$ detto *fasore*, la cui conoscenza permette di riottenere il segnale originario mediante la relazione $x(t) = \Re \{ \underline{x} \cdot e^{j2\pi f_0 t} \}$, che una volta sviluppata⁴ risulta infatti pari a



$$\begin{aligned} x(t) &= \Re \{ A \cdot e^{j(2\pi f_0 t + \varphi)} \} = A \cdot \Re \{ \cos(2\pi f_0 t + \varphi) + j \sin(2\pi f_0 t + \varphi) \} \\ &= A \cos(2\pi f_0 t + \varphi) \end{aligned}$$



Osserviamo che il risultato ottenuto può interpretarsi graficamente come l'aver impresso al fasore una rotazione di velocità angolare $\omega_0 = 2\pi f_0$ radianti/secondo in senso antiorario, ed aver proiettato il risultato sull'asse reale. In alternativa, possiamo esprimere il segnale di partenza anche come

$$x(t) = \frac{1}{2} \{ \underline{x} e^{j2\pi f_0 t} + \underline{x}^* e^{-j2\pi f_0 t} \}$$

Tale operazione coinvolge anche le *frequenze negative*, e corrisponde a tener conto anche di un secondo vettore rotante, che si muove ora in senso orario, che ha una parte immaginaria di segno sempre opposto al primo, e che è moltiplicato per il coniugato del fasore. Vedremo tra breve che l'ultima espressione fornita è esattamente quella della *serie di Fourier* per il caso in questione.

²L'espressione più generale e^γ con $\gamma = \alpha + j\varphi$ è ancora un numero complesso, di modulo e^α . Infatti $e^\gamma = e^{\alpha + j\varphi} = e^\alpha e^{j\varphi} = e^\alpha (\cos \varphi + j \sin \varphi)$.

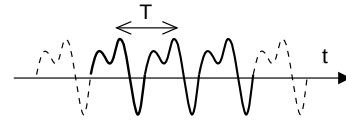
³L'affermazione nasce dalla relazione $e^\alpha e^\beta = e^{\alpha + \beta}$. Ad esempio quindi, il prodotto $\cos \alpha \cdot \sin \beta$ diviene

$$\begin{aligned} &= \frac{1}{4j} (e^{j\alpha} + e^{-j\alpha}) (e^{j\beta} - e^{-j\beta}) = \frac{1}{4j} [e^{j\alpha} e^{j\beta} - e^{j\alpha} e^{-j\beta} + e^{-j\alpha} e^{j\beta} - e^{-j\alpha} e^{-j\beta}] \\ &= \frac{1}{4j} [e^{j(\alpha + \beta)} - e^{-j(\alpha + \beta)} - e^{j(\alpha - \beta)} + e^{-j(\alpha - \beta)}] = \frac{1}{4j} [2j \sin(\alpha + \beta) - 2j \sin(\alpha - \beta)] \\ &= \frac{1}{2} [\sin(\alpha + \beta) - \sin(\alpha - \beta)] \end{aligned}$$

⁴Un modo alternativo di ottenere lo stesso risultato è quello di esprimere gli esponenziali complessi in termini trigonometrici, ottenendo $x(t) = \Re \{ |x| (\cos \varphi + j \sin \varphi) [\cos(2\pi f_0 t + \varphi) + j \sin(2\pi f_0 t + \varphi)] \}$, e sviluppare il calcolo facendo uso delle relazioni $\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha + \beta) + \cos(\alpha - \beta)]$ e $\sin \alpha \sin \beta = \frac{1}{2} [\cos(\alpha - \beta) - \cos(\alpha + \beta)]$, ma avremmo svolto più passaggi.

2.2 Serie di Fourier

Come anticipato a pag. 12, un segnale *periodico* $x(t)$ è un segnale di potenza, che assume ripetutamente gli stessi valori a distanza multipla di un intervallo temporale T denominato *periodo*, ovvero tale che $x(t) = x(t + T) \forall t$. L'inverso di T è detto *frequenza fondamentale* $F = \frac{1}{T}$ o *prima armonica* di $x(t)$, espressa in Hertz, dimensionalmente pari all'inverso di un tempo [sec^{-1}].



Per i segnali periodici esiste una forma di rappresentazione basata sulla conoscenza di una serie infinita di coefficienti complessi $\{X_n\}$ denominati *coefficienti di Fourier*, calcolabili a partire da un periodo del segnale come

$$X_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j2\pi n F t} dt \quad (2.1)$$

e che permettono la *ricostruzione* di $x(t)$, sotto forma di una combinazione lineare di infinite funzioni esponenziali complesse $e^{j2\pi n F t}$, mediante l'espressione nota come *serie di Fourier*:

$$x(t) = \sum_{n=-\infty}^{\infty} X_n e^{j2\pi n F t} \quad (2.2)$$

Osserviamo che:

- La conoscenza di $\{X_n\}$ *equivale* a quella di $x(t)$ e *viceversa*, esistendo il modo di passare dall'una all'altra rappresentazione;
- Le funzioni della base di rappresentazione $e^{j2\pi n F t}$ sono funzioni trigonometriche a frequenza multipla (*n-esima*) della fondamentale, dette anche armoniche *n-esime*⁵
- I termini $X_n e^{j2\pi n F t}$ sono chiamati *componenti armoniche* di $x(t)$ a frequenza $f = nF$;
- Il coefficiente $X_0 = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) dt$ rappresenta la componente continua (o *valor medio*) di $x(t)$;
- La serie di Fourier dà valori esatti in tutti i punti in cui $x(t)$ è continuo, mentre in corrispondenza di discontinuità di prima specie fornisce un valore pari alla media dei valori agli estremi, cosicché il valore dell'energia di un periodo è preservato;
- I coefficienti di Fourier X_n possono essere calcolati anche per un segnale di estensione finita T . Antitrasformando, il segnale diventa periodico!
- Se poniamo $nF = f$ (con f variabile continua), possiamo interpretare le componenti armoniche come i valori campionati di una funzione (complessa) delle frequenze: $X_n = \overline{X}(nF)$. Ad $\overline{X}(f)$ si dà il nome di *inviluppo dello spettro di ampiezza* di $x(t)$, che si ottiene estendendo la definizione dei coefficienti di Fourier: $\overline{X}(f) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j2\pi f t} dt$;

- I coefficienti X_n sono valori complessi. Al loro posto si possono usare, in alternativa:

$$\begin{cases} M_n = |X_n| & \text{Spettro di modulo} \\ \varphi_n = \arctan \frac{\Im\{X_n\}}{\Re\{X_n\}} & \text{Spettro di fase} \end{cases}$$

essendo

$$X_n = |X_n| e^{j\varphi_n} = \Re\{X_n\} + j\Im\{X_n\}$$

2.2.1 Segnali reali

2.2.1.1 Simmetria Coniugata

I coefficienti della serie di Fourier possono essere calcolati anche per segnali complessi; nel caso particolare di $x(t)$ reale i coefficienti di Fourier risultano godere della proprietà di *simmetria coniugata*, espressa come

$$X_{-n} = X_n^*$$

e che esprime la circostanza che i coefficienti con indice n negativo possiedono una parte reale uguale a quella dei coefficienti con (uguale) indice positivo, e parte immaginaria cambiata di segno⁶. Ciò comporta una proprietà analoga per il modulo e la fase di $\{X_n\}$, e dunque possiamo scrivere:

$$x(t) \text{ Reale} \Leftrightarrow \begin{cases} \Re\{X_{-n}\} = \Re\{X_n\} \\ \Im\{X_{-n}\} = -\Im\{X_n\} \end{cases} ; \begin{cases} |X_{-n}| = |X_n| \\ \arg\{X_{-n}\} = -\arg\{X_n\} \end{cases}$$

Tali relazioni evidenziano che

Se $x(t)$ è reale, i coefficienti X_n risultano avere modulo pari e fase dispari, ovvero parte reale pari e parte immaginaria dispari.

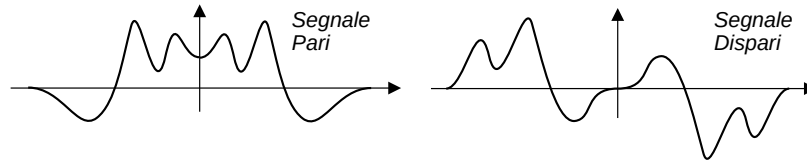
Un corollario di questo risultato è che⁷

Se $x(t)$ è reale pari, i coefficienti X_n sono reali (pari), mentre se $x(t)$ è reale dispari, gli X_n sono immaginari (dispari).

⁶Questa terminologia richiama alla mente nozioni di teoria musicale, in cui gli armonici di una nota sono appunto note di frequenza multipla della prima. In particolare la seconda armonica corrisponde ad un intervallo di ottava, e la 4^a a due ottave. E la terza armonica? Prendendo ad esempio un *La* a 440 Hz, la terza armonica si trova a $440 * 3 = 1320$ Hz. Sapendo che ogni semitono della scala temperata corrisponde ad un rapporto di frequenza pari a $2^{\frac{1}{12}}$ rispetto al semitono precedente, proviamo a determinare il numero di semitoni N_s tra la terza armonica ed il *La* (fondamentale). Risulta allora $2^{\frac{N_s}{12}} = \frac{1320}{440} = 3 \rightarrow \frac{N_s}{12} = \log_2 3 \simeq 1.5849 \rightarrow N_s = 19$ semitoni, ovvero un intervallo di quinta rispetto alla seconda armonica (il *La* dell'ottava successiva, che si trova ad 880 Hz).

⁶La dimostrazione di questa proprietà si basa sul fatto che, scomponendo l'esponenziale complesso che compare nella formula per il calcolo degli X_n come $e^{-j2\pi n F t} = \cos 2\pi n F t - j \sin 2\pi n F t$, ed essendo $x(t)$ reale, l'integrale stesso si suddivide in due, ognuno relativo al calcolo indipendente della parte reale e quella immaginaria: $X_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \cos 2\pi n F t dt - \frac{j}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \sin 2\pi n F t dt$. Essendo il coseno una funzione pari, il primo integrale fornisce gli stessi risultati per n cambiato di segno; il secondo integrale invece cambia segno con n , essendo il seno una funzione dispari.

⁷Con riferimento alla scomposizione del calcolo di X_n alla nota precedente, notiamo che se $x(t)$ è (reale) pari, allora $\Im\{X_n\} = 0$, in quanto $x(t) \sin 2\pi n F t dt$ è dispari, ed il suo integrale esteso ad un intervallo simmetrico rispetto all'origine è nullo. Se invece $x(t)$ è (reale) dispari, si ottiene $\Re\{X_n\} = 0$, per lo stesso motivo applicato al termine $x(t) \cos 2\pi n F t dt$.



2.2.1.2 Interpretazione degli X_n come fasori

Confrontando la formula di ricostruzione

$$x(t) = \sum_{n=-\infty}^{\infty} X_n e^{j2\pi n F t}$$

con quella

$$x(t) = \frac{1}{2} \left\{ \underline{x} e^{j2\pi f_0 t} + \underline{x}^* e^{-j2\pi f_0 t} \right\}$$

ricavata al § 2.1.3 per il caso di un coseno, e tenendo conto della proprietà di simmetria coniugata $X_{-n} = X_n^*$, si nota come un segnale reale possa essere pensato composto a partire da un insieme infinito di fasori $\underline{X}_n$ (pari al doppio dei coefficienti X_n), rotante ognuno con una velocità angolare $\omega_n = 2\pi n F$ multipla della frequenza fondamentale.

2.2.1.3 Serie Trigonometrica

Nel caso in cui gli X_n abbiano simmetria coniugata, la formula di ricostruzione può scriversi

$$x(t) = X_0 + \sum_{n=1}^{\infty} \left\{ X_n e^{j2\pi n F t} + X_{-n} e^{-j2\pi n F t} \right\} = M_0 + \sum_{n=1}^{\infty} M_n 2 \cos(2\pi n F t + \varphi_n)$$

ovvero in forma di serie di coseni; si noti che X_0 è necessariamente reale, in quanto la fase deve risultare una funzione dispari della frequenza.

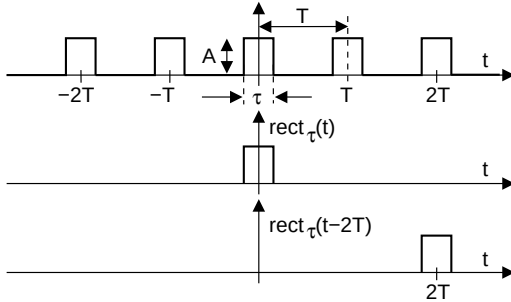
In modo simile, le proprietà relative alle parti reale ed immaginaria permettono di scrivere:

$$\begin{aligned} x(t) &= X_0 + \sum_{n=1}^{\infty} \left\{ (R_n + jI_n) e^{j2\pi n F t} + (R_n - jI_n) e^{-j2\pi n F t} \right\} \\ &= R_0 + \sum_{n=1}^{\infty} \{ 2R_n \cos(2\pi n F t) - 2I_n \sin(2\pi n F t) \} \end{aligned}$$

in cui

$$R_0 = M_0 = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) dt \quad \text{e} \quad \begin{cases} R_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \cos(2\pi n F t) dt \\ I_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \sin(2\pi n F t) dt \end{cases}$$

Pertanto, nel caso in cui $x(t)$ sia un segnale reale, la serie di Fourier si riduce ad uno sviluppo in termini di funzioni trigonometriche, ed in particolare ad una serie di soli coseni nel caso in cui $x(t)$ sia pari, oppure una serie di soli seni, nel caso in cui sia dispari.

Esempio: Serie di Fourier di un'onda rettangolare

Nella figura a lato è mostrata un'onda quadra con un *duty cycle*⁸ del 33%, con espressione analitica che può essere scritta come

$$x(t) = \sum_{n=-\infty}^{\infty} A \text{rect}_{\tau}(t - nT)$$

in cui l'espressione $\text{rect}_{\tau}(t)$ rappresenta un impulso rettangolare di base τ ed altezza unitaria, centrato nell'origine dei tempi.

L'argomento $(t - nT)$ indica una *traslazione* (o spostamento) dello stesso *in avanti* (ossia verso gli istanti *positivi*) di una quantità pari a nT , cosicché la sommatoria rappresenta appunto la replica dello stesso impulso rettangolare infinite volte in avanti ed all'indietro. Il calcolo dei coefficienti di Fourier per il segnale in questione non presenta particolari difficoltà, ed il risultato si presta ad alcune utili considerazioni. Applicando le formule note si ottiene

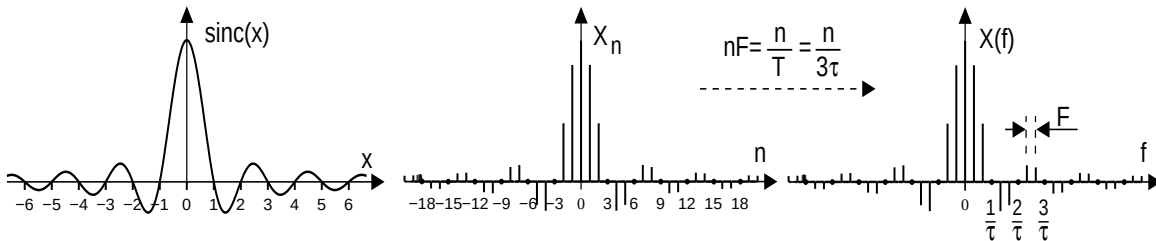
$$\begin{aligned} X_n &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j2\pi n F t} dt = \frac{1}{T} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} A e^{-j2\pi n F t} dt = \frac{A}{T} \frac{e^{-j2\pi n F t}}{-j2\pi n F} \Big|_{-\frac{\tau}{2}}^{\frac{\tau}{2}} = \\ &= \frac{A}{\pi n F T} \frac{e^{j2\pi n F \frac{\tau}{2}} - e^{-j2\pi n F \frac{\tau}{2}}}{2j} = A \frac{\tau}{T} \frac{\sin(\pi n F \tau)}{\pi n F \tau} = A \frac{\tau}{T} \text{sinc}(n F \tau) \end{aligned}$$

Nella seconda uguaglianza, gli estremi di integrazione sono stati ristretti all'intervallo di effettiva esistenza del segnale, mentre l'ultimo passaggio si giustifica ricordando le formule di Eulero.

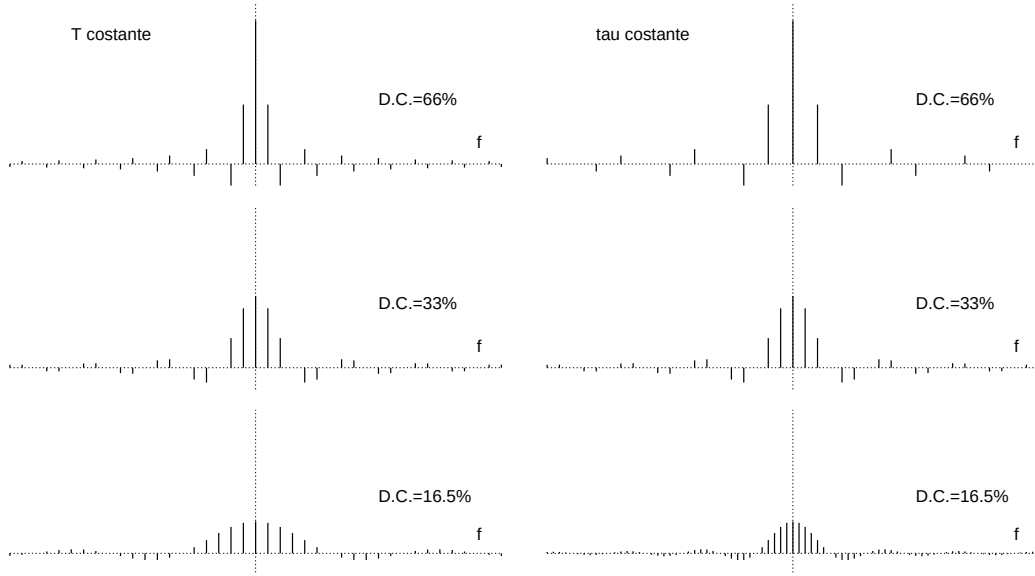
I valori X_n si ottengono quindi calcolando l'espressione $\frac{\sin(\pi n F \tau)}{\pi n F \tau}$ per n intero; nella figura in basso a sinistra è raffigurata per comodità la funzione

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

che passa da zero per x intero, e che ricorrerà spesso nel testo. Nella figura al centro sono mostrati i valori degli $X_n = A \frac{\tau}{T} \text{sinc}(n F \tau)$, in cui si è posto $A = 1$ e $\tau = \frac{T}{3}$ (corrispondente al duty cycle del 33%), e quindi i termini X_n sono nulli in corrispondenza dei valori di $n = 3, 6, 9, \dots$. La figura a destra infine, mostra ancora i coefficienti X_n , ma lungo una scala in Hertz, ottenuta considerando che $nF = \frac{n}{T}$ rappresenta la frequenza dell' n -esima armonica, e che la posizione $\tau = \frac{T}{3}$ adottata fornisce $nF = \frac{n}{3\tau}$. Osserviamo ora che, mentre la spaziatura tra le armoniche è pari ad $F = \frac{1}{T}$ e dipende esclusivamente dal *periodo* della forma d'onda, gli zeri della funzione $\text{sinc}(n F \tau)$ occorrono a frequenze multiple di $\frac{1}{\tau}$. Per meglio comprendere le implicazioni di tali osservazioni, valutiamo come si modificano i valori X_n al variare di τ e di T .



⁸Il DUTY CYCLE si traduce come *ciclo di impegno*, ed è definito come il rapporto percentuale per il quale il segnale è diverso da zero, ossia $\text{duty cycle} = \frac{\tau}{T} * 100 \%$.



Il lato sinistro della figura soprastante mostra cosa accade se τ raddoppia (sopra) o dimezza (sotto): le armoniche mantengono la stessa spaziatura $\frac{1}{T}$, ma l'involuppo $\text{sinc}(nF\tau)$ si *contrae* ed *espande* rispettivamente. Il lato destro della figura mostra invece il risultato ottenibile variando T con τ fisso, e scegliendo i valori di T in modo da ottenere lo stesso duty cycle $\frac{\tau}{T}$ del lato sinistro: questa volta rimane costante la velocità con cui gli X_n vanno a zero, mentre le armoniche si *diradano* (sopra) ed *infittiscono* (sotto) all'aumentare ed al diminuire di T rispettivamente. Infine, notiamo come al diminuire del duty cycle si assista in entrambi i casi ad una riduzione dell'ampiezza degli X_n , legata alla riduzione di potenza del segnale (vedi sezione 2.3).

2.2.2 Serie di Fourier a banda limitata

Consideriamo un'onda quadra con duty-cycle del 50%

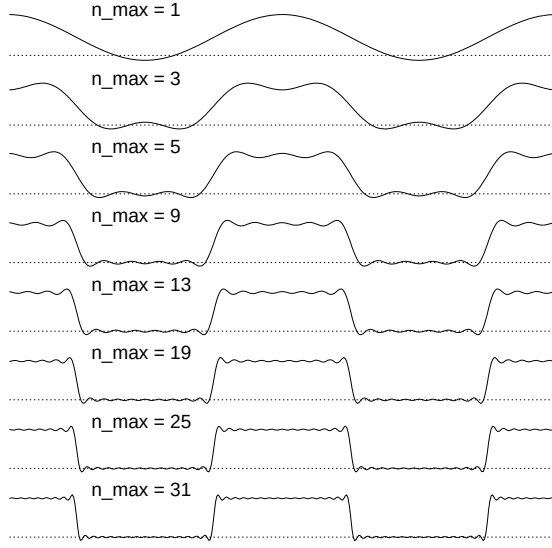
$$x(t) = \sum_{k=-\infty}^{\infty} \text{rect}_{\frac{T}{2}}(t - kT)$$

rappresentata mediante una serie troncata di Fourier in cui si considerano solo i coefficienti X_n con indice $-N \leq n \leq N$. Sappiamo che $X_n = \frac{\tau}{T} \frac{\sin(\pi n F \tau)}{\pi n F \tau}$ e, per $\tau = \frac{T}{2}$, si ottiene $X_n = \frac{1}{2} \frac{\sin(\frac{n\pi}{2})}{\frac{n\pi}{2}} = \frac{1}{2} \text{sinc}\left(\frac{n}{2}\right)$, che risulta diverso da zero solo con n *dispari*, e dunque:

$$X_0 = \frac{1}{2}; \quad X_n = \begin{cases} \frac{(-1)^{\frac{n-1}{2}}}{\pi n} & \text{con } n \text{ dispari} \\ 0 & \text{con } n \text{ pari} \end{cases}$$

Essendo inoltre $x(t)$ reale pari, sappiamo che può essere espresso come serie di coseni:

$$x(t) = X_0 + \sum_{n=1}^{\infty} 2X_n \cos(2\pi n F t)$$



Nella figura a fianco riportiamo il risultato ottenuto arrestando lo sviluppo in serie all'indice mostrato per ogni curva, e generando quindi il segnale

$$\hat{x}_N(t) = X_0 + \sum_{N=1}^{n_{Max}} 2X_n \cos(2\pi n F t)$$

Come osservabile, la ricostruzione è sempre più accurata, tranne che per le oscillazioni in prossimità della discontinuità, che prendono il nome di *Fenomeno di Gibbs*.

Il caso mostrato è emblematico della inaccuratezza che si commette considerando contributi frequenziali ridotti rispetto a quelli propri della forma d'onda⁹, a causa (ad esempio) di un *filtraggio* del segnale.

2.3 Teorema di Parseval

Stabilisce l'equivalenza di due rappresentazioni del segnale dal punto di vista energetico. La potenza, infatti, è calcolabile in modo simile in entrambi i domini del tempo e della frequenza, risultando

$$\mathcal{P}_x = \lim_{\Delta T \rightarrow \infty} \frac{1}{\Delta T} \int_{-\frac{\Delta T}{2}}^{\frac{\Delta T}{2}} |x(t)|^2 dt = \sum_{n=-\infty}^{\infty} |X_n|^2$$

Sviluppiamo i calcoli che danno luogo al risultato mostrato:

$$\begin{aligned} \mathcal{P}_x &= \lim_{\Delta T \rightarrow \infty} \frac{1}{\Delta T} \int_{-\frac{\Delta T}{2}}^{\frac{\Delta T}{2}} |x(t)|^2 dt = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) x^*(t) dt = \\ &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left[\sum_n X_n e^{j2\pi n F t} \right] \left[\sum_m X_m^* e^{-j2\pi m F t} \right] dt = \sum_n \sum_m X_n X_m^* \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{j2\pi(n-m)F t} dt = \\ &= \sum_{n=-\infty}^{\infty} X_n X_n^* = \sum_{n=-\infty}^{\infty} |X_n|^2 = \sum_{n=-\infty}^{\infty} M_n^2 = \sum_{n=-\infty}^{\infty} (R_n^2 + I_n^2) \end{aligned}$$

⁹Un risultato teorico, che qui citiamo solamente, mostra che l'errore quadratico di ricostruzione $\varepsilon = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} (x(t) - \hat{x}(t))^2 dt$ che è presente utilizzando solo le prime N armoniche è il minimo rispetto a quello ottenibile utilizzando un qualunque altro gruppo di N armoniche che non siano le prime.

Ortogonalità degli esponenziali complessi Nei precedenti calcoli si è fatto uso del risultato

$$\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{j2\pi(n-m)Ft} dt = \begin{cases} 0 & \text{con } n \neq m \\ 1 & \text{con } n = m \end{cases}$$

che deriva dalla circostanza che la funzione integranda (per $n \neq m$) è periodica con periodo uguale o sotto-multiplo di T , e quindi a valor medio nullo; per $n = m$ invece essa vale $e^0 = 1$, e dunque il risultato. Questo prende il nome di *Proprietà di Ortogonalità* degli esponenziali complessi, in base ai principi di algebra vettoriale forniti in appendice 2.4.1.

Spettro di Potenza per segnali periodici In appendice (pag. 25) si mostra come l'integrale

$$\|x(t)\|^2 = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) x(t)^* dt$$

oltre a misurare la potenza del segnale periodico $x(t)$, ne misuri la norma quadratica da un punto di vista algebrico.

Tornando ad eseminare il risultato $\mathcal{P}_x = \sum_{n=-\infty}^{\infty} |X_n|^2$ espresso dal teorema di Parseval, notiamo che $|X_n|^2$ è la potenza di una singola componente armonica di $x(t)$:

$$\mathcal{P}_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} [X_n e^{j2\pi n F t}] [X_n^* e^{-j2\pi n F t}] dt = \frac{|X_n|^2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} dt = |X_n|^2$$

e quindi osserviamo che

La potenza totale $\mathcal{P}_x$ di un segnale periodico $x(t)$ è pari alla somma delle potenze delle sue componenti armoniche $X_n e^{j2\pi n F t}$.

Si presti attenzione che il risultato è una diretta conseguenza dell'ortogonalità della base di rappresentazione. In generale, la potenza di una somma *non* è pari alla somma delle potenze¹⁰; l'uguaglianza ha luogo solo nel caso di in cui gli addendi siano ortogonali.

La successione $\{\mathcal{P}_n\} = \{\dots, |X_{-k}|^2, \dots, |X_0|^2, \dots, |X_k|^2, \dots\}$ rappresenta come la potenza totale si ripartisce tra le diverse armoniche a frequenza $f = nF$, e prende il nome di *Spettro di Potenza* del segnale $x(t)$.

Osserviamo che necessariamente i termini $\mathcal{P}_n = |X_n|^2$ risultano reali e positivi. Inoltre, se $x(t)$ è reale, risulta $|X_n|^2 = |X_{-n}|^2 = |X_{-n}|^2$, e quindi si ottiene $\mathcal{P}_n = \mathcal{P}_{-n}$; pertanto un segnale *reale* è caratterizzato da uno spettro di potenza *pari*.

Problema Si determini lo spettro di potenza di un'onda quadra.

Soluzione Essendo $X_n = \frac{1}{2} \text{sinc}(\frac{n}{2})$, si ottiene $\{\mathcal{P}_n\} = \{|X_n|^2\} = \frac{1}{4} \left\{ \text{sinc}^2\left(\frac{n}{2}\right) \right\}$.

¹⁰In generale risulta, con la notazione di prodotto scalare $(\bar{a}, \bar{b})$ tra vettori-segnali $\bar{a}$ e $\bar{b}$ introdotta al § 2.4.1: $(\bar{x} + \bar{y}, \bar{x} + \bar{y}) = (\bar{x}, \bar{x}) + (\bar{y}, \bar{y}) + (\bar{x}, \bar{y}) + (\bar{y}, \bar{x})$.

2.4 Appendici

2.4.1 Algebra Vettoriale

Spazio normato Un insieme di elementi viene detto *spazio lineare* (o *spazio vettoriale*), quando sono definite le operazioni di somma tra elementi e di moltiplicazione degli stessi per dei coefficienti, e queste operazioni danno come risultato ancora un elemento dell'insieme.

Lo *spazio prodotto interno* (o *spazio normato*) è quello spazio lineare, in cui è definito il *prodotto scalare* $(\bar{x}, \bar{y})$ tra generici vettori $\bar{x}$ ed $\bar{y}$ ¹¹. In tal caso, si può definire la *norma* $\|\bar{x}\|$ di un vettore $\bar{x}$ come

$$\|\bar{x}\| = \sqrt{(\bar{x}, \bar{x})}$$

Due vettori si dicono *ortogonali* se $(\bar{x}, \bar{y}) = 0$.

Un generico punto $\bar{x}$ dello spazio può esprimersi come combinazione lineare di vettori $\bar{u}_i$ di una base di rappresentazione, con coefficienti x_i :

$$\bar{x} = \sum_i x_i \bar{u}_i$$

Se per i vettori della base risulta $(\bar{u}_i, \bar{u}_j) = 0$ per tutti gli $i \neq j$, allora la base è detta *ortogonale*, ed i coefficienti x_i si determinano per proiezione di $\bar{x}$ lungo i vettori della base:

$$x_i = (\bar{x}, \bar{u}_i)$$

In tal caso, il prodotto scalare tra due vettori $\bar{x}$ ed $\bar{y}$ ha espressione¹²

$$(\bar{x}, \bar{y}) = \sum_i x_i y_i^* \|\bar{u}_i\|^2$$

Se $\|\bar{u}_i\|^2 = 1$, allora la base è detta *ortonormale*.

¹¹Il prodotto scalare è un operatore che associa ad una coppia di vettori uno scalare. Indicando con $(\bar{x}, \bar{y})$ il prodotto scalare tra $\bar{x}$ ed $\bar{y}$, tale operatore deve soddisfare alle seguenti tre proprietà:

- $(\bar{x}, \bar{y}) = (\bar{y}, \bar{x})$ - proprietà *commutativa*;
- $(a\bar{x} + b\bar{y}, \bar{z}) = a(\bar{x}, \bar{z}) + b(\bar{y}, \bar{z})$ - proprietà *distributiva*;
- $(\bar{x}, \bar{x}) \geq 0$ (con il segno uguale solo se $\bar{x} = 0$).

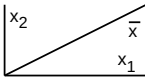
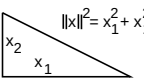
¹²E' facile verificare che il risultato ottenuto è direttamente applicabile allo spazio descritto dalla geometria euclidea, in cui gli u_i sono unitari ed orientati come gli assi cartesiani, ottenendo in definitiva

$$(\bar{x}, \bar{y}) = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Osserviamo inoltre come l'espressione che permette il calcolo della lunghezza di un vettore

$$\|\bar{x}\| = \sqrt{\sum_i (x_i)^2}$$

non sia nient'altro che la riproposizione del teorema di Pitagora, che (su due dimensioni) asserisce l'uguaglianza dell'area del quadrato costruito sull'ipotenusa, con la somma delle aree dei quadrati costruiti sui

cateti. Infatti  diventa  .

Spazio dei segnali periodici I concetti ora esposti sono immediatamente applicabili all'insieme dei segnali periodici di periodo T , idealizzati come elementi di uno spazio normato, per i quali viene definito un operatore di prodotto scalare tra due segnali $x(t)$ ed $y(t)$ come l'integrale

$$(x(t), y(t)) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) y^*(t) dt$$

a cui corrisponde una *norma quadratica* immediatamente riconoscibile come la *potenza* del segnale:

$$\|x(t)\|^2 = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt$$

E' dunque immediato constatare che il sottospazio costituito dall'insieme dei segnali $\{e^{j2\pi n F t}\}$, con $F = \frac{1}{T}$, costituisce una base ortonormale per i segnali periodici di periodo T ; in particolare si riconosce che l'espressione (2.1) di pag. 17 rappresenta la proiezione¹³ del segnale lungo i vettori della base, mentre la formula di ricostruzione (2.2) costituisce la rappresentazione del segnale nei termini delle sue componenti ortogonali.

Ri-definizione dei coefficienti di Fourier Moltiplicando ambo i membri della formula di ricostruzione (2.2) per $e^{-j2\pi m F t}$, ed eseguendo l'integrale tra due istanti t_1 e t_2 presi a distanza di un multiplo intero di periodi (ossia $t_2 - t_1 = kT$), si ottiene

$$\begin{aligned} \int_{t_1}^{t_2} x(t) e^{-j2\pi m F t} dt &= \int_{t_1}^{t_2} \left(\sum_{n=-\infty}^{\infty} X_n e^{j2\pi n F t} \right) e^{-j2\pi m F t} dt = \\ &= \sum_{n=-\infty}^{\infty} X_n \int_{t_1}^{t_2} e^{j2\pi(n-m)F t} dt = (t_2 - t_1) \cdot X_m \end{aligned}$$

in quanto per $n \neq m$ la funzione integranda ha valor medio nullo, dato che nell'intervallo (t_1, t_2) (dovunque collocato dell'asse dei tempi) presenta un numero intero di periodi. Pertanto, il calcolo dei coefficienti può ottenersi a partire da un qualunque intervallo esteso su un numero intero di periodi:

$$X_n = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t) e^{-j2\pi n F t} dt$$

Disuguaglianza di Schwartz Consiste nel risultato

$$\left| \int_{-\infty}^{\infty} x(t) y(t) dt \right|^2 \leq \int_{-\infty}^{\infty} |x(t)|^2 dt \cdot \int_{-\infty}^{\infty} |y(t)|^2 dt$$

che a volte può tornare utile nei calcoli che coinvolgono segnali di energia. La dimostrazione si basa sull'identificare l'insieme di tali segnali come uno spazio normato, dotato di un operatore di prodotto scalare tra $x(t)$ ed $y(t)$ definito come

$$(x(t), y(t)) = \int_{-\infty}^{\infty} x(t) y^*(t) dt \quad (2.3)$$

¹³ Infatti, il prodotto scalare si calcola come il prodotto dei moduli, moltiplicato per l'angolo compreso tra i due: $(\vec{x}, \vec{y}) = |\vec{x}| \cdot |\vec{y}| \cdot \cos \theta$. Se il secondo vettore ha lunghezza unitaria, si ottiene la proiezione del primo nella direzione del secondo.

Con tali posizioni, il risultato mostrato deriva da quello valido per un qualunque spazio vettoriale, che fa uso della disuguaglianza $|\cos \theta| \leq 1$, e che mostra che

$$(\bar{x}, \bar{y})^2 = (|x| \cdot |y| \cdot \cos \theta)^2 \leq |x|^2 \cdot |y|^2$$

Applicando quindi la definizione di prodotto scalare (2.3) ai due vettori-segnale $\bar{x} = x(t)$ e $\bar{y} = y(t)$, si ottiene il risultato espresso dalla disuguaglianza di Schwartz, in cui $\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} x(t) x^*(t) dt = (\bar{x}, \bar{x})$, e per la quale vale il segno di uguale se e solo se $x(t) = Ky^*(t)$, con K costante reale.

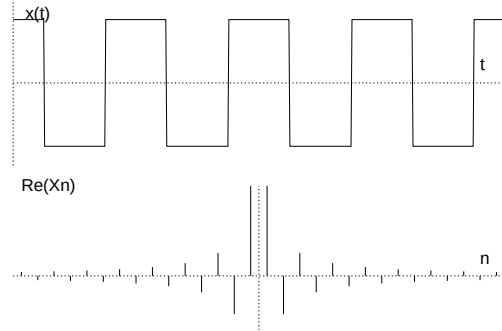
2.4.2 Esempi di Sviluppo in serie

Nello schema che segue, sono mostrate le ampiezze delle componenti armoniche X_n per alcuni segnali periodici di periodo T , di cui è fornita l'espressione nel tempo per $|t| < T/2$.

Onda quadra simmetrica

$$x(t) = \begin{cases} +1 & |t| < T/4 \\ -1 & T/4 \leq |t| < T/2 \end{cases}$$

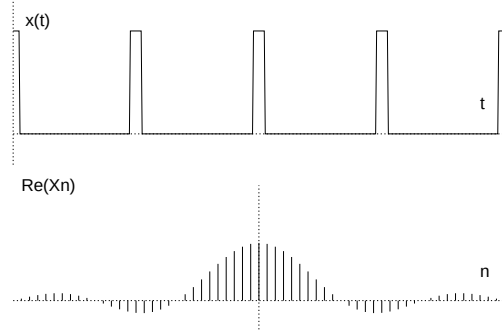
$$X_n = \begin{cases} \text{sinc}\left(\frac{n}{2}\right) & n \neq 0 \\ 0 & n = 0 \end{cases}$$



Treno di impulsi rettangolari

$$x(t) = \begin{cases} +1 & |t| < \tau/2 \\ 0 & \tau/2 \leq |t| < T/2 \end{cases}$$

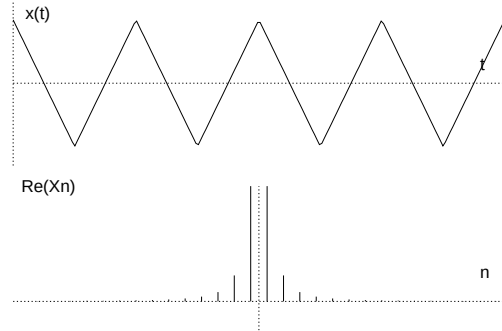
$$X_n = \frac{\tau}{T} \text{sinc}\left(\frac{n\tau}{T}\right)$$



Onda triangolare simmetrica

$$x(t) = 1 - 4\frac{|t|}{T} \quad |t| < T/2$$

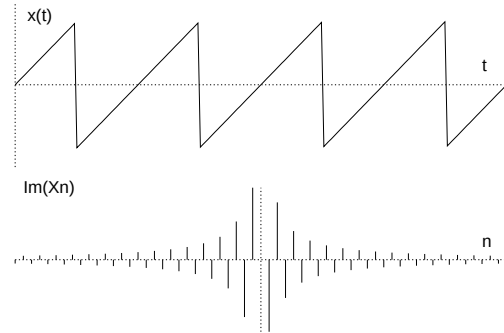
$$X_n = \begin{cases} \text{sinc}^2\left(\frac{n}{2}\right) & n \neq 0 \\ 0 & n = 0 \end{cases}$$



Dente di sega simmetrico

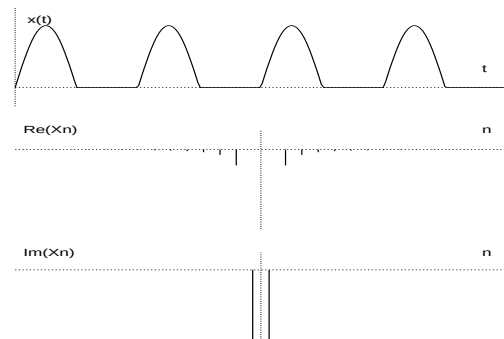
$$x(t) = 2\frac{t}{T} \quad |t| < T/2$$

$$X_n = \begin{cases} j \frac{(-1)^n}{n\pi} & n \neq 0 \\ 0 & n = 0 \end{cases}$$

**Rettificata a singola semionda**

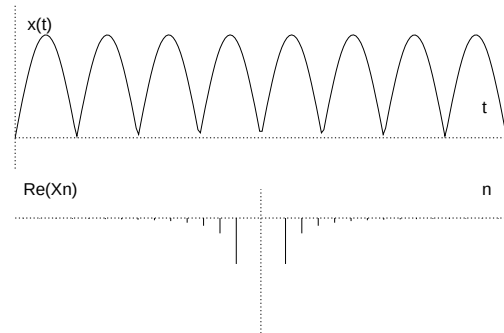
$$x(t) = \begin{cases} \sin \omega_0 t & 0 \leq t < T/2 \\ 0 & -T/2 \leq t < 0 \end{cases}$$

$$X_n = \begin{cases} \frac{1}{\pi(1-n^2)} & n \text{ pari} \\ -j\frac{1}{4} & n = \pm 1 \\ 0 & \text{altrimenti} \end{cases}$$

**Rettificata a onda intera**

$$x(t) = |\sin \omega_0 t|$$

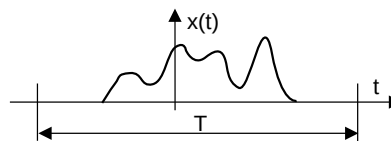
$$X_n = \begin{cases} \frac{2}{\pi(1-n^2)} & n \text{ pari} \\ 0 & \text{altrimenti} \end{cases}$$



Capitolo 3

Trasformata di Fourier

Abbiamo visto al capitolo 2 che lo sviluppo in serie di Fourier può essere applicato ad un segnale limitato nel tempo, e che l'uso della formula di ricostruzione rende periodico il segnale originario. Se però facciamo tendere ad infinito il periodo "fittizio" T su cui sono calcolati i coefficienti X_n , le armoniche della serie di Fourier tendono ad infittirsi, fino ad arrivare ad una distanza infinitesima; allo stesso tempo, la periodizzazione del segnale ricostruito tende via via a scomparire.



3.1 Definizione

La trasformata di Fourier serve a rappresentare quei segnali per i quali non sussiste una struttura periodica, ed è un operatore funzionale che, applicato ad un segnale definito nel dominio del tempo, ne individua un altro nel dominio della variabile *continua* frequenza (a differenza della serie *discreta* di Fourier, idonea al caso in cui siano presenti *solo armoniche* della fondamentale). L'operazione di trasformazione è spesso indicata con la simbologia $X(f) = \mathcal{F}\{x(t)\}$, ed il segnale trasformato si indica con la stessa variabile di quello nel tempo, resa maiuscola. La sua definizione formale dal punto di vista analitico è:

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

la cui esistenza è garantita per segnali $x(t)$ impulsivi (ovvero per i quali $\int_{-\infty}^{\infty} |x(t)| dt < \infty$, cioè assolutamente sommabili). Un segnale impulsivo è anche di energia, mentre non è sempre vero il viceversa. Spesso però, $X(f)$ esiste anche per segnali di energia; vedremo inoltre che può essere definita (grazie ad operazioni di passaggio al limite) anche per segnali di potenza periodici.

L'antitrasformata di Fourier $\mathcal{F}^{-1}\{\}$ è l'operatore analitico che svolge l'associazione inversa a $\mathcal{F}\{\}$, e che consente di ottenere, a partire da un segnale definito nel dominio della frequenza, quel segnale nel dominio del tempo la cui trasformata è il primo segnale. L'operazione di antitrasformazione è definita come

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df$$

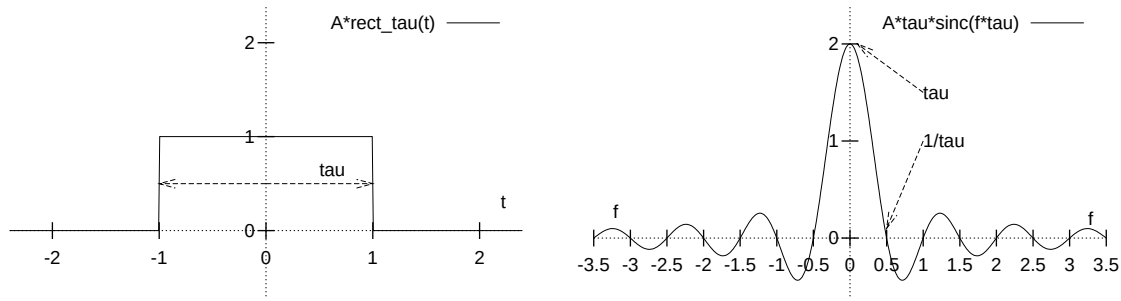


Figura 3.1: $\mathcal{F}$ -trasformata di un rettangolo di base $\tau = 2$ ed ampiezza $A = 1$

e vale ovunque $x(t)$ sia continuo, mentre nelle discontinuità di prima specie fornisce il valor medio di $x(t)$. Il risultato della trasformata $X(f) = M(f) \exp^{j\varphi(f)}$ è anche detto *spettro di ampiezza complessa*, mentre $M(f)$ ed $\varphi(f)$ sono detti spettri di *modulo* e *fase*.

La formula di ricostruzione, se messa a confronto con la serie di Fourier, può essere pensata come una somma integrale di *infinite* componenti $X(f) df e^{j2\pi ft}$ di ampiezza (complessa) infinitesima, evidenziando come ora siano presenti *tutte* le frequenze e non solo le armoniche. Una seconda analogia con la serie di Fourier deriva dal considerare un segnale $x(t)$ di durata limitata T , e calcolare $X(f) = \mathcal{F}\{x(t)\}$ per $f = \frac{n}{T} = nF$. In tal caso, è facile verificare che risulta $X(f = nF) = T \cdot X_n$ con X_n pari all' n -esimo coefficiente di Fourier calcolato per $x(t)$ su quello stesso periodo.

Prima di procedere con le proprietà e le caratteristiche di questa trasformata, svolgiamo un semplice esercizio.

Trasformata di un rettangolo Disponendo del segnale $x(t) = A \text{rect}_\tau(t)$, se ne vuol calcolare lo spettro di ampiezza $X(f)$. Svolgendo il calcoli si ottiene:

$$\begin{aligned}
 X(f) &= \int_{-\infty}^{\infty} A \text{rect}_\tau(t) e^{-j2\pi ft} dt = A \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} e^{-j2\pi ft} dt = \\
 &= A \left. \frac{e^{-j2\pi ft}}{-j2\pi f} \right|_{-\frac{\tau}{2}}^{\frac{\tau}{2}} = \frac{A}{\pi f} \frac{e^{j2\pi f \frac{\tau}{2}} - e^{-j2\pi f \frac{\tau}{2}}}{2j} = \\
 &= A\tau \frac{\sin(\pi f \tau)}{\pi f \tau} = A\tau \cdot \text{sinc}(f\tau)
 \end{aligned}$$

Questo risultato, graficato in fig 3.1, ricorda quello già incontrato a pag. 20 per la serie di Fourier dell'onda quadra. Il noto andamento $\frac{\sin x}{x}$ rappresenta ora la distribuzione *continua* in frequenza dello spettro di ampiezza, ed il primo zero della curva si trova presso $f = \frac{1}{\tau}$, in modo del tutto simile al treno di impulsi rettangolari di base τ . Notiamo esplicitamente inoltre che, *aumentando* la durata del *rect*, lo spettro si *restringe*, addensandosi nella regione delle frequenze più basse; mentre al contrario, qualora il *rect* sia più breve, $X(f)$ si estende a regioni di frequenza più elevata.

3.2 Densità di Energia

Similmente al caso dei segnali periodici, viene ora stabilita una relazione tra l'energia di un segnale, e la distribuzione della stessa nel dominio della frequenza. In base alle considerazioni geometriche esposte in § 2.4.1, definiamo come prodotto scalare tra i segnali di energia $x(t)$ e $y(t)$ (detto anche *energia incrociata*) il valore

$$\mathcal{E}_{xy} = (\overline{x}, \overline{y}) = \int_{-\infty}^{\infty} x(t) y^*(t) dt$$

che, nel caso in cui $x(t) = y(t)$, coincide con l'energia $\mathcal{E}_x$ di $x(t)$. Se entrambi $x(t)$ e $y(t)$ possiedono trasformata di Fourier possiamo scrivere:

$$\begin{aligned} \mathcal{E}_{xy} &= \int y^*(t) \left[\int X(f) e^{j2\pi ft} df \right] dt = \int X(f) \left[\int y^*(t) e^{j2\pi ft} dt \right] df \\ &= \int_{-\infty}^{\infty} X(f) Y^*(f) df \end{aligned}$$

Il risultato

$$\int_{-\infty}^{\infty} x(t) y^*(t) dt = \int_{-\infty}^{\infty} X(f) Y^*(f) df$$

costituisce il *teorema di Parseval* per segnali di energia, ed implica che le trasformate di segnali ortogonali, sono anch'esse ortogonali. Ponendo ora $x(t) = y(t)$, si ottiene:

$$\mathcal{E}_x = (\overline{x}, \overline{x}) = \|x\|^2 = \int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df$$

Esaminando quest'ultima espressione, possiamo indicare

$$\mathcal{E}_x(f) = |X(f)|^2$$

come lo *spettro di densità di energia* di $x(t)$. Infatti, l'integrale $\int_{f_1}^{f_2} |X(f)|^2 df$ rappresenta il contributo all'energia totale $\mathcal{E}_x$ di $x(t)$, limitatamente alla banda di frequenze comprese tra f_1 ed f_2 .

3.3 Prime proprietà della trasformata di Fourier

Simmetria coniugata Nel caso in cui $x(t)$ sia reale, risulta¹

$$X(f) = X^*(-f)$$

e quindi la parte reale di $X(f)$ è *pari*, e quella immaginaria *dispari*, ossia modulo $|X(f)|$ pari e fase $\arg\{X(f)\}$ dispari; si applica infine il corollario di pag. 18.

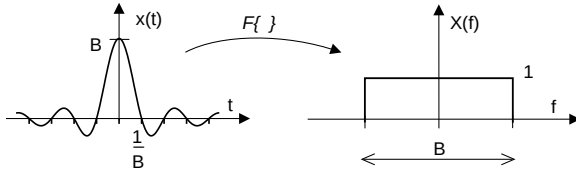
Dualità Trasformata ed antitrasformata differiscono solo per il segno. Ciò comporta che se sostituiamo alla variabile f del risultato $X(f)$ di una $\mathcal{F}$ -trasformata, la variabile t , si ottiene una funzione del tempo $X(t)$ che, se nuovamente trasformata, fornisce ... il segnale originario $x(t)$, calcolato nella variabile f , cambiata di segno: $x(-f)$. Il concetto esposto, verificabile analiticamente con facilità, si riassume come

¹Infatti $X^*(f) = \left[\int x(t) e^{-j2\pi ft} dt \right]^* = \int x^*(t) e^{j2\pi ft} dt = X(-f)$ dato che $x(t)$ è reale.

$$\begin{aligned} x(t) &\xrightarrow{\mathcal{F}\{\}} X(f) & t=f &\xrightarrow{\quad} X(t) \xrightarrow{\mathcal{F}\{\}} x(-f) \\ X(f) &\xrightarrow{\mathcal{F}^{-1}\{\}} x(t) & f=t &\xrightarrow{\quad} x(f) \xrightarrow{\mathcal{F}^{-1}\{\}} X(-t) \end{aligned}$$

e consente l'uso di risultati ottenuti “in un senso” (ad es. da tempo a frequenza) per derivare senza calcoli i risultati nell'altro (o viceversa).

Esempio: Trasformata di un *sinc*(*t*) Supponiamo di voler $\mathcal{F}$ -trasformare il segnale $x(t) = B \frac{\sin(\pi t B)}{\pi t B} = B \text{sinc}(tB)$: l'applicazione cieca dell'integrale che definisce la trasformata di Fourier appare un'impresa ardua...



Allora, ricordando che $\mathcal{F}\{\text{rect}_\tau(t)\} = \tau \text{sinc}(f\tau)$, scriviamo direttamente che

$$\mathcal{F}\{B \text{sinc}(tB)\} = \text{rect}_B(f)$$

Pertanto la trasformata di un $\frac{\sin x}{x}$ nel tempo, è un rettangolo in frequenza.

Linearità Discende molto semplicemente dalla proprietà distributiva dell'integrale che definisce la trasformata. Pertanto:

$$\text{se } z(t) = ax(t) + by(t) \quad \text{allora} \quad Z(f) = aX(f) + bY(f)$$

Valore medio e valore iniziale Subito verificabile una volta notato che la $\mathcal{F}$ -trasformata, calcolata per $f = 0$, si riduce all'integrale di $x(t)$, e quindi al suo *valor medio*. Pertanto:

$$m_x = \int_{-\infty}^{\infty} x(t) dt = X(f=0) \quad \text{e, per dualità:} \quad x_0 = x(t=0) = \int_{-\infty}^{\infty} X(f) df$$

dove l'ultima relazione esprime la proprietà del *valore iniziale*.

Come esempio di applicazione, troviamo subito che

$$\int_{-\infty}^{\infty} \text{sinc}(tB) dt = \frac{1}{B} \text{rect}_B(f=0) = \frac{1}{B}$$

Traslazione nel tempo Si tratta di una proprietà molto semplice, e che ricorre frequentemente nei calcoli sui segnali. Manifesta la relazione esistente tra la trasformata dei segnali e quella degli stessi traslati, e si esprime con il predicato:

$$\begin{aligned} \text{se } z(t) &= x(t-T) \\ \text{allora } Z(f) &= X(f) e^{-j2\pi fT} \end{aligned}$$

la cui dimostrazione è fornita sotto².

²La dimostrazione della proprietà di traslazione nel tempo si basa su di un semplice cambio di variabile: $Z(f) = \int x(t-T) e^{-j2\pi ft} dt = \int x(\theta) e^{-j2\pi f(T+\theta)} d\theta = e^{-j2\pi fT} \int x(\theta) e^{-j2\pi f\theta} d\theta = X(f) e^{-j2\pi fT}$

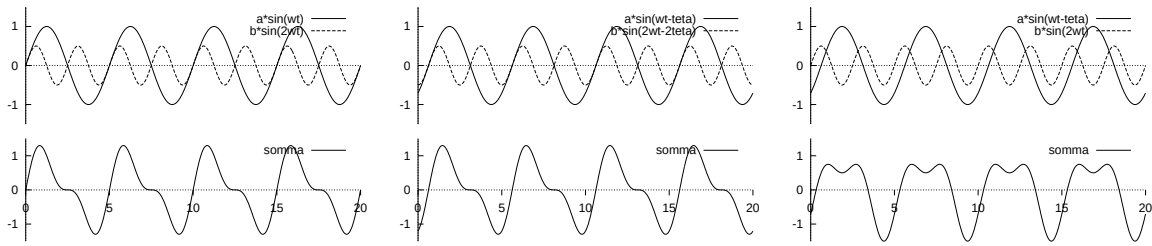
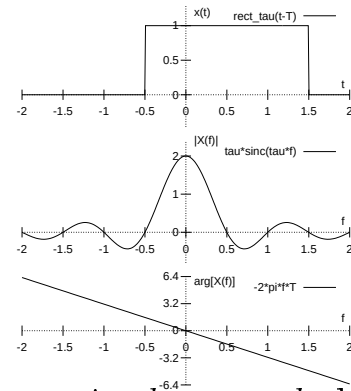


Figura 3.2: Confronto tra diversi spettri di fase

Esempio

La figura di lato esemplifica il risultato ottenuto nel caso in cui $x(t) = \text{rect}_\tau(t - T)$, mostrando come nello spettro di $x(t)$, di modulo $|X(f)| = \mathcal{F}\{\text{rect}_\tau\} = \tau \text{sinc}(f\tau)$, si aggiunga un contributo di fase lineare $\varphi(f) = -2\pi fT$. Nel caso in figura, si è posto $\tau = 2$ e $T = .5$, ottenendo in definitiva $Z(f) = X(f)e^{-j2\pi fT} = 2\text{sinc}(2f)e^{-j\pi f}$.

Poniamo ora l'attenzione sul fatto che l'espressione $x(t - T)$ indica un *ritardo* del segnale $x(t)$ di una quantità pari a T .



Conseguenze della linearità di fase La circostanza che un *ritardo temporale* del segnale $x(t)$ si traduca in una alterazione *lineare* della fase³ della sua trasformata $X(f)$, comporta una conseguenza notevole anche nel passaggio da f a t , ossia:

Se si desidera che un segnale mantenga inalterata la sua forma d'onda, pur subendo una alterazione della propria trasformata, l'unica possibilità è quella di modificare lo spettro di fase, con andamento lineare in frequenza.

Esempio Consideriamo un segnale periodico $x(t)$ costituito da due sole armoniche:

$$x(t) = a \sin(\omega t) + b \sin(2\omega t)$$

(avendo posto $2\pi F = \omega$), assieme alla sua versione ritardata

$$x(t - T) = a \sin(\omega(t - T)) + b \sin(2\omega(t - T)) = a \sin(\omega t - \omega T) + b \sin(2\omega t - 2\omega T)$$

Ponendo $\omega T = \theta$, otteniamo

$$x(t - T) = a \sin(\omega t - \theta) + b \sin(2\omega t - 2\theta)$$

e verifichiamo che la seconda armonica subisce un ritardo di fase esattamente doppio.

In fig 3.2 si è posto $a = 1$; $b = .5$; $\theta = \frac{\pi}{4}$ e $F = .2$, ed è mostrato sia il segnale somma originario, sia quello ottenuto considerando un contributo di fase lineare per le due armoniche. Verifichiamo che nel secondo caso, la forma d'onda è la stessa ottenibile per $T = 0$, in quanto le armoniche sono traslate del medesimo intervallo temporale. A destra invece, la fase della seconda armonica viene annullata, ottenendo dalla somma

³La circostanza evidenziata fa sì che, nel caso in cui $X(f)$ presenti un andamento *lineare* della fase, si usi dire che è presente un *ritardo di fase*.

un segnale $a \sin(2\pi Ft - \theta) + b \sin(2\pi Ft)$. Come è evidente, in questo caso il risultato assume una forma completamente diversa⁴.

Traslazione in frequenza (*Modulazione*) E' la proprietà duale della precedente, e stabilisce che

$$\text{se } Z(f) = X(f - f_0) \quad \text{allora } z(t) = x(t) e^{j2\pi f_0 t}$$

la cui dimostrazione è del tutto analoga a quanto già visto. Da un punto di vista mnemonico, distinguiamo la traslazione temporale da quella in frequenza per il fatto che, nel primo caso, i *segni* della traslazione e dell'esponenziale complesso sono *uguali*, e nel secondo, *opposti*.

Da un punto di vista pratico, può sorgere qualche perplessità per la comparsa di un segnale *complesso* nel tempo. Mostriamo però che anti-trasformando uno spettro ottenuto dalla somma di due traslazioni opposte, si ottiene un segnale reale:

$$\mathcal{F}^{-1}\{X(f - f_0) + X(f + f_0)\} = x(t) e^{j2\pi f_0 t} + x(t) e^{-j2\pi f_0 t} = 2x(t) \cos 2\pi f_0 t$$

Pertanto, lo sdoppiamento e la traslazione di $X(f)$ in $\pm f_0$ sono equivalenti ad un segnale cosinusoidale di frequenza f_0 , la cui ampiezza è modulata dal segnale $x(t) = \mathcal{F}^{-1}\{X(f)\}$. E' proprio per questo motivo, che la proprietà è detta anche di *modulazione* (vedi anche a pag. 40).

Coniugato Deriva direttamente⁵ dalla definizione di $\mathcal{F}$ -trasformata:

$$\mathcal{F}\{x^*(t)\} = X^*(-f); \quad \mathcal{F}^{-1}\{X^*(f)\} = x^*(-t)$$

Nel caso di segnali reali, ritroviamo la proprietà di simmetria coniugata $X(f) = X^*(-f)$.

Cambiamento di scala Quantifica gli effetti sullo spettro di una variazione nella velocità di scorrimento del tempo (e viceversa). Possiamo ad esempio pensare come, riavvolgendo velocemente un nastro magnetico, si ascolta un segnale di durata più breve, e dal timbro più *acuto*. Questo fenomeno viene espresso analiticamente come:

$$\mathcal{F}\{x(at)\} = \frac{1}{a} X\left(\frac{f}{a}\right)$$

in cui scegliendo $a > 1$ si ottiene accelerazione temporale ed allargamento dello spettro (ed il contrario con $a < 1$). La dimostrazione è riportata alla nota⁶.

3.4 Impulso matematico e trasformate notevoli

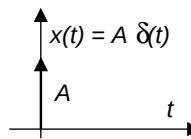
Prima di esporre altre proprietà della trasformata di Fourier, occorre definire ed analizzare le proprietà della "funzione" *impulso matematico*, indicato con $\delta(\cdot)$. Questo è

⁴Nel seguito illustreremo che una conseguenza del risultato discusso, è la sensibilità delle trasmissioni numeriche alle distorsioni di fase.

⁵Infatti $\mathcal{F}\{x^*(t)\} = \int_{-\infty}^{\infty} x^*(t) e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} [x(t) e^{j2\pi ft}]^* dt$.

⁶ $\int x(at) e^{-j2\pi ft} dt = \frac{1}{a} \int x(\beta) e^{-j2\pi \frac{f}{a} \beta} d(\beta) = \frac{1}{a} \int X(\beta) e^{-j2\pi \frac{f}{a} \beta} d\beta = \frac{1}{a} X\left(\frac{f}{a}\right)$

definito come un segnale $\delta(t)$ ⁷ che vale zero ovunque, tranne per $t = 0$ dove vale ∞ ; per contro, l'area di $\delta(t)$ è unitaria:



$$\delta(f) = \begin{cases} \infty & \text{con } f = 0 \\ 0 & \text{altrove} \end{cases} \quad \text{e} \quad \int_{-\infty}^{\infty} \delta(f) df = 1$$

Da un punto di vista analitico, $\delta(f)$ non è una funzione, ma una *distribuzione*, definita come il limite a cui tende una serie di funzioni. E' prassi rappresentare graficamente $A \cdot \delta(f)$ come una freccia (vedi figura) con scritto accanto il valore dell'area A.

3.4.1 Trasformata di una costante

La trasformata di una costante è un impulso di area pari a valore della costante.

Questa proprietà è valida per entrambi i domini (f e t) di partenza, fornendo

$$\mathcal{F}\{A\} = A \cdot \delta(f) \quad \text{e} \quad \mathcal{F}^{-1}\{A\} = A \cdot \delta(t)$$

In appendice 3.9.2 sono svolte riflessioni che illustrano come interpretare questo risultato. Qui osserviamo semplicemente che la costante A può essere vista come il limite, per $\tau \rightarrow \infty$, di un segnale rettangolare: $A = \lim_{\tau \rightarrow \infty} A \text{rect}_{\tau}(t)$, e quindi $\lim_{\tau \rightarrow \infty} A \tau \text{sinc}(f\tau) = \begin{cases} \infty & \text{con } f = 0 \\ 0 & \text{altrove} \end{cases}$. Ci troviamo pertanto nelle esatte circostanze che definiscono un impulso matematico, e resta da verificare che $\int_{-\infty}^{\infty} \tau \text{sinc}(f\tau) df = 1$: si può mostrare (pag. 32) che l'integrale vale uno per qualunque τ , e quindi possiamo scrivere $\mathcal{F}\{A\} = A \cdot \delta(f)$.

3.4.2 Trasformata per segnali periodici

Consideriamo un segnale periodico $x(t)$, del quale conosciamo lo sviluppo in serie

$$x(t) = \sum_{n=-\infty}^{\infty} X_n e^{j2\pi n F t}$$

Applicando la proprietà di linearità, il risultato per la trasformata di una costante, e ricordando la proprietà della traslazione in frequenza, troviamo⁸ che la $\mathcal{F}$ -trasformata di $x(t)$ vale:

$$X(f) = \sum_{n=-\infty}^{\infty} X_n \delta(f - nF)$$

Lo spettro di ampiezza di un segnale periodico è quindi costituito da *impulsi matematici*, situati in corrispondenza delle frequenze armoniche, e di area pari ai rispettivi coefficienti della serie di Fourier.

Un modo alternativo di calcolare la trasformata di segnali periodici è illustrato alla sezione 3.8.

⁷L'impulso matematico è noto anche con il nome di *Delta di Dirac*, e per questo è rappresentato dal simbolo δ .

⁸ $X(f) = \mathcal{F}\left\{\sum_{n=-\infty}^{\infty} X_n e^{j2\pi n F t}\right\} = \sum_{n=-\infty}^{\infty} X_n \mathcal{F}\{1 \cdot e^{j2\pi n F t}\} = \sum_{n=-\infty}^{\infty} X_n \cdot \delta(f - nF)$

Trasformata di un coseno Appliciamo il risultato trovato nel verso opposto, ossia per individuare le componenti armoniche, a partire dall'espressione della trasformata di Fourier. Nel caso di un coseno, che scriviamo

$$x(t) = A \cos(2\pi f_0 t + \varphi) = A \frac{e^{j(2\pi f_0 t + \varphi)} + e^{-j(2\pi f_0 t + \varphi)}}{2}$$

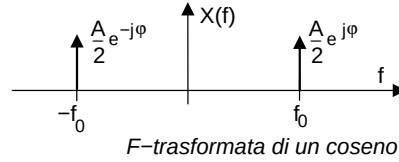
la $\mathcal{F}$ -trasformata risulta:

$$\begin{aligned} X(f) &= \mathcal{F} \left\{ \frac{A}{2} \left(e^{j2\pi f t} e^{j\varphi} + e^{-j2\pi f t} e^{-j\varphi} \right) \right\} \\ &= \frac{A}{2} \{ e^{j\varphi} \delta(f - f_0) + e^{-j\varphi} \delta(f + f_0) \} \end{aligned}$$

in cui riconosciamo

$$X_1 = \frac{A}{2} e^{j\varphi} \text{ e } X_{-1} = \frac{A}{2} e^{-j\varphi}$$

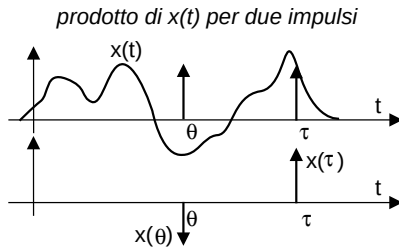
come mostrato in figura.



Potenza di un coseno Cogliamo l'occasione per calcolare la potenza di una sinusoide. Applicando il teorema di Parseval si ottiene:

$$\mathcal{P}_x = |X_1|^2 + |X_2|^2 = 2 \frac{A^2}{4} = \frac{A^2}{2}$$

3.4.3 Proprietà di setacciamento



Osserviamo innanzitutto che il *prodotto* di un segnale per un impulso unitario dà come risultato lo stesso impulso, con *area* pari al valore del segnale nell'istante in cui è centrato l'impulso:

$$x(t) \delta(t - \tau) = x(\tau) \delta(t - \tau).$$

Questa considerazione consente di scrivere il valore di $x(t)$ per un istante $t = \tau$, nella forma

$$x(\tau) = \int_{-\infty}^{\infty} x(t) \delta(t - \tau) dt$$

Quest'ultima proprietà è detta di *setacciamento* (SIEVING in inglese) in quanto consiste (metaforicamente !) nel passare al setaccio $x(t)$, che compare in entrambi i membri dell'espressione ottenuta, così come la farina compare su entrambi i lati (!) del setaccio stesso.

3.5 Risposta Impulsiva e Convoluzione

3.5.1 Risposta Impulsiva

Consideriamo un sistema fisico (elettrico, meccanico, pneumatico...) che venga sollecitato (in un punto considerato come ingresso) da un segnale impulsivo $\delta(t)$, e consideriamo l'andamento temporale di una grandezza (meccanica, pneumatica, elettrica...) che possiamo considerare una uscita. Tale risultato prende il nome di *risposta impulsiva* (ossia all'impulso) e viene indicata con $h(t)$. L'andamento di $h(t)$ rappresenta la grandezza

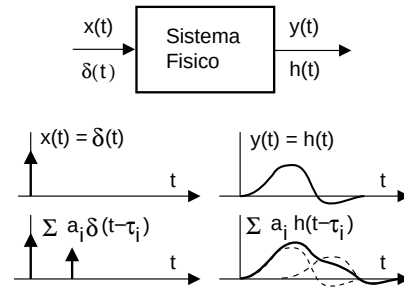
di uscita, osservata dopo che è passato un tempo pari a t da quando si è applicato in ingresso l'impulso $\delta(t)$.

Se il sistema è *lineare e permanente*⁹, applicando un ingresso costituito da *più impulsi*, ognuno con area differente a_i e centrato ad un diverso istante τ_i , ovvero

$$x(t) = \sum_{i=1}^N a_i \delta(t - \tau_i) \quad (3.1)$$

si ottiene una uscita pari a

$$y(t) = \sum_{i=1}^N a_i h(t - \tau_i) \quad (3.2)$$



Si rifletta sul significato della sommatoria, con l'aiuto della figura a lato: ad un dato istante t , il valore dell'uscita $y(t)$ risulta dalla somma di N termini, ognuno pari al valore della risposta impulsiva calcolata con argomento pari alla differenza tra istante di osservazione ed istante di applicazione dell' i -esimo impulso.

3.5.2 Integrale di convoluzione

Consideriamo ancora lo stesso sistema fisico, al cui ingresso sia posto un generico segnale $x(t)$ che, grazie alla proprietà di setacciamento, rappresentiamo come scomposto in infiniti termini, ossia come somma integrale di impulsi centrati in τ (variabile) ed area $x(\tau) d\tau$ (infinitesima):

$$x(t) = \int_{-\infty}^{\infty} x(\tau) d\tau \delta(t - \tau)$$

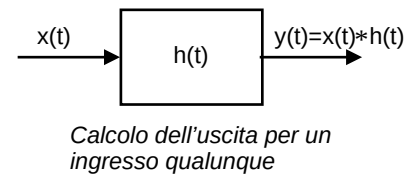
Questa espressione, formalmente simile alla (3.1), è equivalente alla proprietà di setacciamento, dato che $\delta(t)$ è una funzione pari. L'andamento della grandezza di uscita sarà il risultato della sovrapposizione di infinite risposte impulsive, ognuna relativa ad un diverso valore dell'ingresso:

$$y(t) = \int_{-\infty}^{\infty} x(\tau) h(t - \tau) d\tau$$

in cui $x(\tau) d\tau$ è l'area degli impulsi che costituiscono l'ingresso, e $h(t - \tau)$ è l'uscita all'istante t causata dall'impulso in ingresso centrato all'istante τ .

Il risultato ottenuto, formalmente simile a (3.2), prende il nome di *integrale di convoluzione*, viene indicato in forma simbolica da un asterisco (*), e gode della proprietà commutativa, ovvero

$$\begin{aligned} y(t) &= x(t) * h(t) = \\ &= \int_{-\infty}^{\infty} x(\tau) h(t - \tau) d\tau = \int_{-\infty}^{\infty} h(\tau) x(t - \tau) d\tau \end{aligned}$$

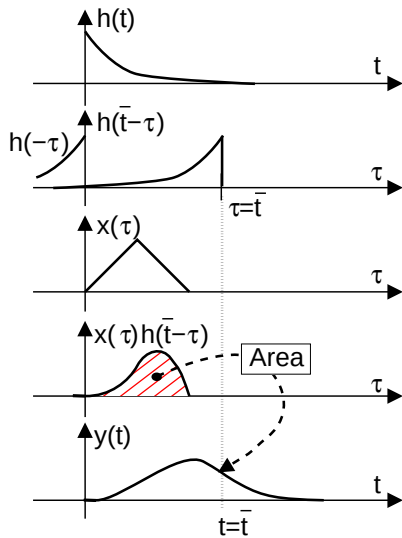


La notazione con l'asterisco fa sì che ci si possa riferire ad essa come al “prodotto di convoluzione”. Notiamo come $h(t)$ caratterizzi completamente il sistema fisico, in quanto permette di calcolarne l'uscita per un qualsiasi ingresso.

⁹Il significato di questa classificazione si trova al Capitolo 1, a pag. 11.

3.5.3 $h(t)$ come funzione memoria

Diamo ora un'interpretazione grafica della convoluzione: poniamo che $h(t)$ sia esponenziale decrescente ed $x(t)$ triangolare, come mostrato in figura.



Definiamo una costruzione grafica, che ci mostra la funzione integranda quando si deve calcolare l'uscita ad un generico istante $t = \bar{t}$. L'andamento di $h(\bar{t} - \tau)$ con τ variabile indipendente, si ottiene ribaltando $h(t)$ rispetto all'origine dei tempi e traslandola (nel passato, quindi a destra) di $\bar{t}$. Il risultato dell'integrale di convoluzione, quando $t = \bar{t}$, è pari a

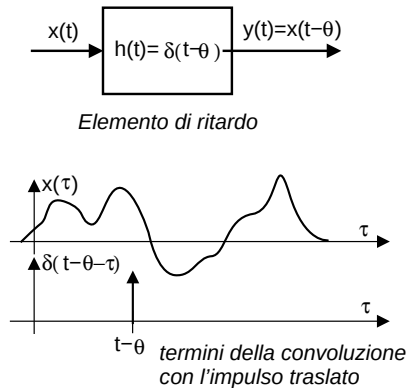
$$y(\bar{t}) = \int_{-\infty}^{\infty} x(\tau) h(\bar{t} - \tau) d\tau$$

ossia pari all'area del prodotto $x(\tau) h(\bar{t} - \tau)$, tratteggiata in figura; per altri valori di $\bar{t}$, il termine $h(\bar{t} - \tau)$ sarà traslato di una diversa quantità. Il calcolo dell'area di $x(\tau) h(\bar{t} - \tau)$ ha il significato di sommare le risposte causate da tutti i valori di ingresso, ogni risposta presa con il rispettivo ritardo $\bar{t} - \tau$ tra gli istanti (passati) $\tau \leq \bar{t}$ di applicazione dei valori di ingresso, e

l'istante $\bar{t}$ di osservazione. Pertanto, i valori di $h(t)$ rappresentano la memoria, da parte del sistema fisico, degli ingressi precedenti.

Il risultato della convoluzione, come si vede dalla costruzione grafica, ha una estensione (durata) maggiore dei due segnali di partenza.

3.5.4 Convoluzione con l'impulso traslato



Consideriamo un sistema fisico che operi un semplice ritardo θ sui segnali in ingresso: in tal caso risulterà $h(t) = \delta(t - \theta)$ ovvero, la risposta all'impulso è un impulso ritardato. Per calcolare l'uscita, che sappiamo essere pari a $y(t) = x(t - \theta)$, possiamo ricorrere all'integrale di convoluzione, ottenendo

$$\begin{aligned} y(t) &= x(t) * h(t) = x(t) * \delta(t - \theta) = \\ &= \int_{-\infty}^{\infty} x(\tau) \delta(t - \theta - \tau) d\tau = x(t - \theta) \end{aligned}$$

Questo risultato ci permette di enunciare un principio generale, che verrà utilizzato di frequente, e che recita:

La convoluzione tra un segnale $x(t)$ ed un impulso matematico $\delta(t - \theta)$ centrato ad un istante θ provoca la traslazione di $x(t)$ all'istante in cui è centrato l'impulso.

3.6 Moltiplicazione in Frequenza e nel Tempo

La descrizione di un sistema fisico per mezzo della sua risposta impulsiva è di fondamentale utilità soprattutto per merito della seguente proprietà della trasformata di Fourier:

La $\mathcal{F}$ -trasformata della convoluzione tra due segnali è pari al prodotto delle trasformate dei segnali:

$$\mathcal{F}\{x(t) * y(t)\} = X(f) Y(f)$$

La dimostrazione è riportata alla nota¹⁰. Sussiste inoltre anche la proprietà *duale*, ovvero ad un *prodotto nel tempo* corrisponde una *convoluzione in frequenza*, che si scrive

$$\mathcal{F}\{x(t) \cdot y(t)\} = X(f) * Y(f)$$

Trattiamo ora delle conseguenze e dei risvolti legati a queste due importanti proprietà, iniziando dalla prima.

Moltiplicazione in Frequenza L'applicazione più importante è legata al calcolo dell'uscita da un sistema fisico: questo può essere interamente condotto nel dominio della frequenza, calcolando

$$Y(f) = \mathcal{F}\{x(t) * h(t)\} = X(f) H(f)$$

e quindi ottenendo $y(t) = \mathcal{F}^{-1}\{Y(f)\}$. La trasformata della risposta impulsiva $H(f) = \mathcal{F}\{h(t)\}$ prende allora il nome di *risposta in frequenza*, in quanto rappresenta l'alterazione introdotta dal sistema per ogni frequenza presente in $X(f)$. Facciamo ora un paio di esempi per sperimentare questo modo di procedere.

Sistema passa tutto Poniamo di avere $H(f) = 1$, e che quindi risulti $h(t) = \delta(t)$. In questo caso le componenti di $X(f)$ alle diverse frequenze non subiscono nessuna alterazione, ottenendo $y(t) = \mathcal{F}^{-1}\{Y(f)\} = \mathcal{F}^{-1}\{X(f)\} = x(t)$, ed il sistema viene detto di tipo *passa tutto*. Per verifica, scriviamo l'integrale di convoluzione, che risulta $y(t) = \int_{-\infty}^{\infty} x(\tau) \delta(t - \tau) d\tau = x(t)$: ritroviamo quindi la proprietà di setacciamento.

Ritardo Se invece $H(f) = e^{-j2\pi f\theta}$, pari cioè ad un esponenziale complesso, il sistema equivale ad un elemento di ritardo, riproducendo in uscita l'ingresso presentato θ istanti prima. Infatti risulta: $y(t) = \mathcal{F}^{-1}\{Y(f)\} = \mathcal{F}^{-1}\{X(f) e^{-j2\pi f\theta}\} = x(t - \theta)$. D'altra parte, scrivendo l'integrale di convoluzione, e ricordando che $h(t) = \mathcal{F}^{-1}\{e^{-j2\pi f\theta}\} = \delta(t - \theta)$, avremmo ottenuto $y(t) = \int_{-\infty}^{\infty} x(\tau) \delta(t - \theta - \tau) d\tau = x(t - \theta)$, ritrovando la proprietà della convoluzione per un impulso traslato.

¹⁰
$$\begin{aligned} Z(f) &= \mathcal{F}\{x(t) * y(t)\} = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x(\tau) y(t - \tau) d\tau \right] e^{-j2\pi ft} dt = \\ &= \int_{-\infty}^{\infty} x(\tau) \left[\int_{-\infty}^{\infty} y(t - \tau) e^{-j2\pi ft} dt \right] d\tau = \int_{-\infty}^{\infty} x(\tau) Y(f) e^{-j2\pi f\tau} d\tau = \\ &= Y(f) \int_{-\infty}^{\infty} x(\tau) e^{-j2\pi f\tau} d\tau = Y(f) \cdot X(f) \end{aligned}$$

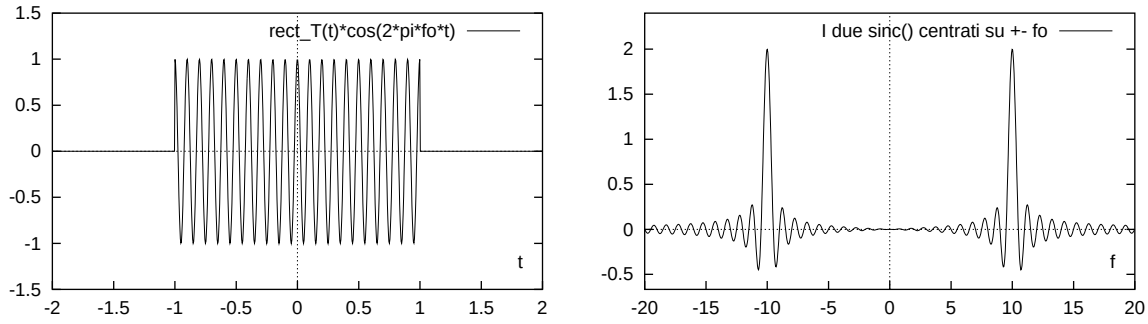


Figura 3.3: Trasformata di un coseno finestrato con $T = 2$, $f_0 = 10$

Moltiplicazione nel Tempo (*Modulazione*) La relazione

$$Z(f) = \mathcal{F}\{x(t)y(t)\} = X(f) * Y(f)$$

ci permette di investigare le conseguenze frequenziali del prodotto temporale di due segnali.

Esempio A titolo di esempio, analizziamo il caso in cui $z(t) = A \text{rect}_T(t) \cos 2\pi f_0 t$, la cui forma d'onda è graficata a sinistra della Fig. 3.3. Applicando i risultati noti e la proprietà di traslazione in frequenza, risulta:

$$\begin{aligned} Z(f) &= \frac{A}{2} \mathcal{F}\left\{\text{rect}_T(t) \left(e^{j2\pi f_0 t} + e^{-j2\pi f_0 t}\right)\right\} \\ &= \frac{AT}{2} [\text{sinc}[(f - f_0)T] + \text{sinc}[(f + f_0)T]] \end{aligned}$$

in cui $\mathcal{F}\{\text{rect}_T(t)\} = T \text{sinc}(fT)$ si è traslato in $\pm f_0$. Questo risultato, mostrato a destra in fig. 3.3¹¹, coincide con quello previsto: l'espressione di $Z(f)$ infatti è anche pari alla convoluzione tra $\mathcal{F}\{\text{rect}_T(t)\}$, ed i due impulsi traslati $\mathcal{F}\{\cos 2\pi f_0 t\} = \frac{1}{2}(\delta(f - f_0) - \delta(f + f_0))$.

L'esempio proposto ci permette di motivare il termine *Modulazione* associato a questa proprietà. L'ampiezza del coseno risulta infatti *modulata* dal rettangolo. La *modulazione di ampiezza* (AM) dei radio ricevitori casalinghi si riferisce esattamente a questo processo, svolto allo scopo di condividere tra più emittenti la banda prevista per le trasmissioni, assegnando a ciascuna di esse una diversa frequenza portante f_0 su cui trasmettere.

3.7 Derivazione ed Integrazione nel Tempo

Le ultime due proprietà riguardano un risultato di applicazione meno frequente, ma talvolta utile. Si ottiene infatti che le operazioni di derivata ed integrale di un segnale

¹¹Dalla figura si può anche arguire come, per T crescente, $Z(f)$ tenda sempre più ad assomigliare ad una coppia di impulsi, ossia al risultato noto per un coseno "infinito". Qualora si consideri invece solo un *breve intervallo* di un segnale, il suo spettro si modifica, a seguito della convoluzione con lo spettro della finestra di analisi.

L'operazione di estrazione di una porzione di segnale di durata limitata, a partire da un altro comunque esteso, è indicata come una operazione di *finestratura* (WINDOWING). In appendice (3.9.3) sono svolte considerazioni relative alla scelta di una finestra rettangolare o con *altro andamento*.

possono essere realizzate mediante il passaggio dello stesso attraverso un sistema fisico, dato che queste sono equivalenti a prodotti in frequenza, realizzabili quindi come convoluzione del segnale per la risposta impulsiva del sistema.

Derivazione nel Tempo E' equivalente a moltiplicare lo spettro per $j2\pi f$:

$$\mathcal{F}\left\{\frac{d}{dt}x(t)\right\} = j2\pi f \cdot X(f)$$

e più in generale $\mathcal{F}\left\{\frac{d^n}{dt^n}x(t)\right\} = (j2\pi f)^n \cdot X(f)$. Per segnali di energia, la dimostrazione è svolta nella nota¹². L'andamento dello spettro originario $X(f)$ risulta esaltato alle frequenze più elevate, in quanto il suo modulo è moltiplicato per $2\pi|f|$. La fase, invece, subisce un incremento di $\frac{\pi}{2}$ a tutte le frequenze (il numero immaginario puro $j2\pi f = 2\pi f e^{j\frac{\pi}{2}}$ ha fase $\frac{\pi}{2}$).

Esempio Calcolare $Y(f) = \mathcal{F}\{y(t)\}$ con $y(t) = \frac{d}{dt}x(t)$ e $x(t) = \cos 2\pi f_1 t + \cos 2\pi f_2 t$. Valutare quindi $y(t) = \mathcal{F}^{-1}\{Y(f)\}$ con $f_1 = 10$, $f_2 = 100$.

Si ottiene: $X(f) = \frac{1}{2}(\delta(f - f_1) + \delta(f + f_1) + \delta(f - f_2) + \delta(f + f_2))$. Dato che $f \cdot \delta(f - a) = a \cdot \delta(f - a)$, risulta:

$$Y(f) = j2\pi \frac{1}{2}[f_1(\delta(f - f_1) - \delta(f + f_1)) + f_2(\delta(f - f_2) - \delta(f + f_2))]$$

Considerando ora che $j2\pi \frac{1}{2} = -\frac{2\pi}{2j}$, si ottiene $y(t) = -2\pi f_1 \sin \omega_1 t - 2\pi f_2 \sin \omega_2 t$ e quindi, per $f_1 = 10$ e $f_2 = 100$, si ha

$$y(t) = -2\pi [10 \sin \omega_1 t + 100 \sin \omega_2 t]$$

Integrazione nel Tempo E' equivalente a dividere lo spettro per $j2\pi f$:

$$\mathcal{F}\left\{\int_{-\infty}^t x(\theta) d\theta\right\} = \frac{X(f)}{j2\pi f}$$

Tale risultato è diretta conseguenza del precedente, in virtù dei legami tra integrale e derivata. Infatti, $\int_{-\infty}^t x(\theta) d\theta$ è una funzione di t , che compare nel limite superiore di integrazione, e la sua derivata è proprio $x(t)$.

In questo caso, le basse frequenze del segnale originario sono esaltate seguendo un andamento $1/2\pi|f|$, mentre la fase subisce una variazione (un ritardo) costante pari a $-\frac{\pi}{2}$. Notiamo come questo risultato determini una singolarità per $f = 0$ in presenza di componenti continue per $x(t)$: in tal caso infatti il suo integrale tende a divergere, ed il risultato non è più di energia.

Esempio: Trasformata di un triangolo Consideriamo un segnale

$$x(t) = \text{rect}_T\left(t + \frac{T}{2}\right) - \text{rect}_T\left(t - \frac{T}{2}\right)$$

ed il suo integrale

$$z(t) = \int_{-\infty}^t x(\theta) d\theta = T \text{tri}_{2T}(t)$$

¹² $\mathcal{F}\left\{\frac{dx(t)}{dt}\right\} = \int_{-\infty}^{\infty} \frac{dx(t)}{dt} e^{-j2\pi ft} dt = x(t) e^{-j2\pi ft} \Big|_{-\infty}^{\infty} + j2\pi f \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt = j2\pi f X(f)$ in quanto $x(t)$ che compare nel primo termine dell'integrale per parti, essendo di energia, tende a zero per $t \rightarrow \infty$.

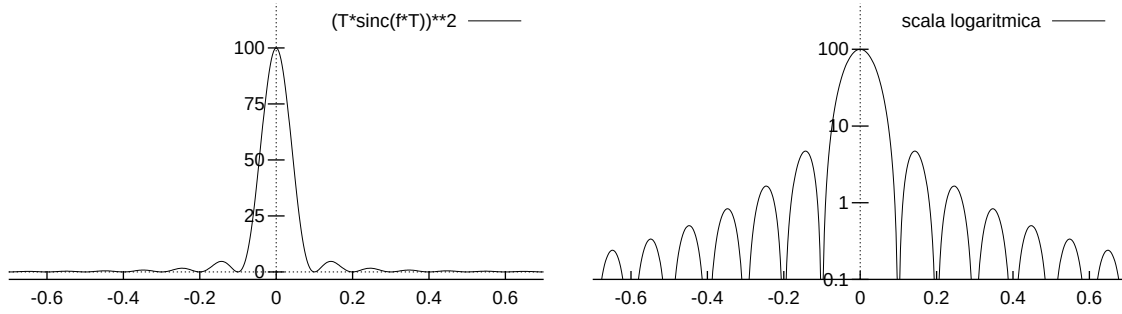
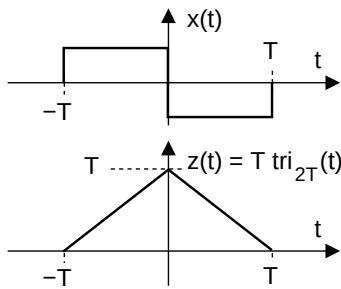


Figura 3.4: Andamento di $(T \operatorname{sinc}(fT))^2$ in scala lineare e logaritmica; $T = 10$.



entrambi rappresentati nella figura a lato: $z(t)$ è nullo fino a $t < -T$, cresce linearmente fino a $t = 0$, e quindi il contributo all'integrale dato dall'area del *rect* negativo torna ad annullarne il valore.

Per calcolare la $\mathcal{F}$ -trasformata di $z(t)$, calcoliamo prima quella di $x(t)$, e poi applichiamo la proprietà dell'integrazione. Applicando la proprietà di traslazione nel tempo, scriviamo

$$\begin{aligned} X(f) &= T \cdot \operatorname{sinc}(fT) \cdot e^{+j2\pi f \frac{T}{2}} - T \cdot \operatorname{sinc}(fT) \cdot e^{-j2\pi f \frac{T}{2}} = T \cdot \frac{\sin(\pi fT)}{\pi fT} \cdot 2j \sin \pi fT \\ &= j2T \frac{\sin^2(\pi fT)}{\pi fT} \end{aligned}$$

Dividendo quindi per $j2\pi f$ si ottiene

$$Z(f) = \frac{j2T \frac{\sin^2(\pi fT)}{\pi fT} \frac{T}{T}}{j2\pi f} = \left(T \frac{\sin(\pi fT)}{\pi fT} \right)^2 = (T \operatorname{sinc}(fT))^2$$

il cui andamento è mostrato in figura 3.4.

Lo stesso risultato può essere ottenuto per altra via, notando che

$$z(t) = T \cdot \operatorname{tri}_{2T}(t) = \operatorname{rect}_T(t) * \operatorname{rect}_T(t)$$

Come verifica, si ripercorra la costruzione grafica riportata alla sezione 3.6. E' quindi ora sufficiente applicare la proprietà del prodotto in frequenza, per ottenere:

$$Z(f) = [\mathcal{F}\{\operatorname{rect}_T(t)\}]^2 = [T \operatorname{sinc}(fT)]^2$$

Quest'ultima funzione $[T \operatorname{sinc}(fT)]^2$, indipendentemente da questo esercizio in cui si colloca, è anche pari alla densità di energia $\mathcal{E}_x(f)$ di un singolo segnale rettangolare $x(t) = \operatorname{rect}_T(t)$; infatti per il teorema di Parseval, si ha $\mathcal{E}_x(f) = X(f) X^*(f)$, ed $X(f) = \mathcal{F}\{\operatorname{rect}_T(t)\} = T \operatorname{sinc}(fT)$.

3.8 Trasformata di segnali periodici

Presentiamo ora un diverso modo di ottenere lo spettro di un segnale periodico, facendo uso di una particolare forma d'onda (ideale) nota come

Treno di impulsi che si esprime analiticamente come

$$\pi_T(t) = \sum_{m=-\infty}^{\infty} \delta(t - mT)$$

e che si rivelerà di utilizzo frequente nei contesti del campionamento e delle trasmissioni numeriche.

Segnale periodico Consideriamo un segnale periodico di periodo T espresso come

$$x(t) = \sum_{m=-\infty}^{\infty} g(t - mT)$$

di cui $g(t)$ costituisce un periodo: la riproduzione di infinite repliche di $g(t)$, spaziate di un periodo T l'una dall'altra, ricompongono il segnale periodico originario. Sfruttando la proprietà di convoluzione con l'impulso traslato, la stessa somma può essere scritta come

$$x(t) = \sum_{m=-\infty}^{\infty} g(t) * \delta(t - mT) = g(t) * \sum_{m=-\infty}^{\infty} \delta(t - mT) = g(t) * \pi_T(t)$$

dove nel secondo passaggio si è sfruttata la linearità della convoluzione. Ricordando ora la proprietà della moltiplicazione in frequenza, troviamo $X(f) = G(f) \cdot \mathcal{F}\{\pi_T(t)\}$; ci accingiamo allora a determinare $\mathcal{F}\{\pi_T(t)\}$, ossia la trasformata del treno di impulsi.

Trasformata del treno di impulsi

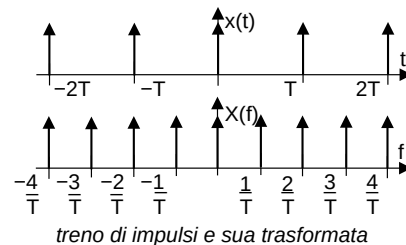
L'approccio che conviene seguire è di pensare a $\pi_T(t)$ come ad un segnale periodico, e svilupparlo in serie di Fourier. I coefficienti si calcolano allora come:

$$\begin{aligned} \Pi_n &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left[\sum_{m=-\infty}^{\infty} \delta(t - mT) \right] e^{-j2\pi n F t} dt \\ &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \delta(t) e^{-j2\pi n F t} dt = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} 1 \cdot \delta(t) dt = \frac{1}{T} \end{aligned}$$

in quanto, tra tutti gli impulsi della sommatoria, ne resta solo uno, quello centrato in zero, dato che gli altri sono tutti esterni ai limiti di integrazione; pertanto, tutti i coefficienti risultano avere lo stesso valore, pari ad $\frac{1}{T}$, e possiamo dunque scrivere

$$\mathcal{F}\{\pi_T(t)\} = \mathcal{F}\left\{ \sum_{n=-\infty}^{\infty} \Pi_n e^{j2\pi n F t} \right\} = \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right) = \frac{1}{T} \pi_{\frac{1}{T}}(f)$$

ottenendo il risultato cercato: $\mathcal{F}\{\pi_T(t)\} = \frac{1}{T} \pi_{\frac{1}{T}}(f)$. Pertanto, la trasformata di un treno di impulsi è *a sua volta* un treno di impulsi, di periodo inverso a quello originario.



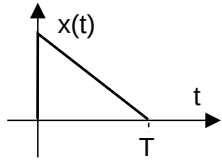
Trasformata di segnale periodico Siamo finalmente in grado di esprimere la trasformata di un segnale periodico come il prodotto tra la $\mathcal{F}$ -trasformata di un suo periodo ed un treno di impulsi in frequenza:

$$X(f) = G(f) \cdot \frac{1}{T} \pi_{\frac{1}{T}}(f)$$

Esempio Riprendendo in considerazione il caso dell'onda quadra affrontato in 2.2.1.3, non è difficile riconoscere come, ponendo $g(t) = \text{Arect}_\tau(t)$, e corrispondentemente $G(f) = A\tau \text{sinc}(f\tau)$, il prodotto di $G(f)$ per il treno di impulsi $\frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - nF)$ (con $F = \frac{1}{T}$) fornisce il risultato già incontrato: $X(f) = A\frac{\tau}{T} \sum_{n=-\infty}^{\infty} \text{sinc}(nF\tau) \delta(f - nF)$.

3.9 Appendici

3.9.1 Esercizio



Sia dato il segnale $x(t) = \begin{cases} 1 - \frac{t}{T} & \text{con } 0 \leq t \leq T \\ 0 & \text{altrimenti} \end{cases}$ mostrato in figura. Descrivere quanti più modi possibili di calcolarne lo spettro di densità di energia $\mathcal{E}_x(f)$.

1. Si calcola $X(f) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$ e quindi $\mathcal{E}_x(f) = |X(f)|^2$;
2. Anticipando un risultato del capitolo 7, è possibile calcolare $\mathcal{R}_x(f) = \int_{-\infty}^{\infty} x(t) x(t + \tau) dt$, e quindi $\mathcal{E}_x(f) = \mathcal{F}\{\mathcal{R}_x(f)\}$;
3. Notando che $x(t) = y(t) \cdot z(t)$ con $y(t) = \text{tri}_{2T}(t)$ e $z(t) = \text{rect}_T\left(t - \frac{T}{2}\right)$, possiamo scrivere $X(f) = Y(f) * Z(f)$, e quindi procedere come in 1);
4. Notiamo che la derivata¹³ di $x(t)$ vale $g(t) = \frac{d}{dt}x(t) = \delta(t) - \frac{1}{T}\text{rect}_T\left(t - \frac{T}{2}\right)$; questo ci permette di calcolare $G(f)$ come $G(f) = \mathcal{F}\{g(t)\} = 1 - \text{sinc}(fT)$. Otteniamo quindi $X(f) = \frac{G(f)}{j2\pi f}$, e procediamo come in 1).

3.9.2 Sulla trasformata di una costante

Svolgiamo alcune considerazioni sul risultato mostrato a pag. 35, illustrando come l'impulso $\delta(\cdot)$ permetta di rappresentare particolari situazioni. Consideriamo pertanto il segnale costante $x(t) = A$.

Trattiamo innanzitutto $x(t) = A$ come un segnale periodico di periodo T tendente ad ∞ ¹⁴, ed esprimiamolo nei termini dei coefficienti di Fourier: l'integrale $X_n =$

¹³La derivata di una discontinuità di prima specie è pari ad un impulso matematico, di area uguale all'altezza della discontinuità. Infatti l'integrale dell'impulso $\int_{-\infty}^t \delta x(\theta) d\theta$ è proprio un gradino. Questa considerazione consente di risolvere in modo semplice le trasformate di segnali in cui è presente una discontinuità.

¹⁴In effetti, un segnale costante è un segnale periodico, con periodo T qualsiasi.

$\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} A e^{-j2\pi n F t} dt$ per $T \rightarrow \infty$ fornisce zero per tutti gli n tranne che per $n = 0$, e quindi si ottiene $X_n = A$ se $n = 0$, mentre $X_n = 0$ se $n \neq 0$.

In alternativa, pensiamo la costante come il limite a cui tende un'onda quadra con duty-cycle $\frac{\tau}{T}$ al tendere di τ a T : lo spettro di ampiezza è stato calcolato al Capitolo 2, pag. 20, presenta righe alle armoniche $f = \frac{n}{T}$, mentre gli X_n con andamento $\text{sinc}(nF\tau)$ si azzerano alle frequenze $f = \frac{n}{\tau}$. Se $\tau \rightarrow T$, gli zeri annullano tutte le armoniche tranne X_0 , il cui valore $A\frac{\tau}{T}$ tende ora ad A .

Qualora invece si desideri calcolare la *trasformata* di Fourier anziché la *serie*, applicando la definizione $X(f) = \int_{-\infty}^{\infty} A e^{-j2\pi f t} dt$ si ottiene $X(f) = 0$ ovunque, tranne che in $f = 0$ dove $X(0) = \infty$.

Convienne allora ricorrere ad un'operazione di passaggio al limite, e pensare il segnale $x(t) = A$ come il risultato dell'allargamento progressivo di un $\text{rect}_\tau(t)$, cioè come $x(t) = \lim_{\tau \rightarrow \infty} A \text{rect}_\tau(t)$. La figura a lato mostra come, considerando valori τ_i via via più grandi, si ottenga una trasformata $X_{\tau_i}(f) = A\tau_i \text{sinc}(f\tau_i)$ sempre più *alta e stretta*.

Notiamo ora che l'energia di $x_\tau(t) = A \text{rect}_\tau(t)$ vale $\mathcal{E}_{x_\tau} = \int_{-\infty}^{\infty} |x_\tau(t)|^2 dt = A^2\tau$; per il teorema di Parseval, l'energia coincide nei domini di tempo e frequenza, e quindi risulta

$$\mathcal{E}_{x_\tau} = \int_{-\infty}^{\infty} |X_\tau(f)|^2 df = A^2\tau$$

Al tendere di τ ad ∞ , l'energia diviene infinita, mentre la potenza vale

$$\mathcal{P}_x = \lim_{\tau \rightarrow \infty} \frac{\mathcal{E}_{x_\tau}}{\tau} = \lim_{\tau \rightarrow \infty} \int_{-\infty}^{\infty} \frac{|X_\tau(f)|^2}{\tau} df = A^2$$

L'espressione $\lim_{\tau \rightarrow \infty} \frac{|X_\tau(f)|^2}{\tau}$ rappresenta dunque lo *spettro di densità di potenza* $\mathcal{P}_x(f)$ della costante A , che finalmente scriviamo come $\mathcal{P}_x(f) = A^2\delta(f)$, in cui $\delta(f)$ è la funzione *impulso matematico* introdotta in 3.4. In tal modo infatti, è facile verificare che risulta

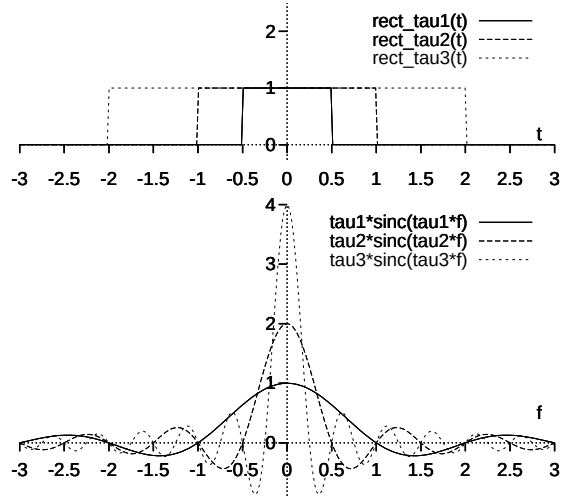
$$\mathcal{P}_x = \int_{-\infty}^{\infty} A^2\delta(f) df = A^2, \text{ e } \mathcal{P}_x(f) = \begin{cases} \infty & \text{con } f = 0 \\ 0 & \text{altrove} \end{cases}. \text{ In altre parole, il formalismo}$$

dell'impulso matematico rende possibile trattare questo caso, dove la potenza (finita) è tutta concentrata in un unico punto ($f = 0$) dando luogo ad una densità infinita.

3.9.3 Finestratura e stima spettrale

Prima di chiudere il capitolo, applichiamo la teoria svolta per speculare sul problema dell'analisi spettrale effettuata a partire da un solo segmento temporale del segnale.

Il calcolo dello spettro di $y(t) = x(t)w(t)$ fornisce, come noto, la trasformata $Y(f) = X(f) * W(f)$. Quindi, il vero spettro $X(f)$ di $x(t)$ non può essere conosciuto, se non



tramite l'effetto della convoluzione con quello $W(f)$ della funzione finestra $w(t)$. Già a pagina 40 si è fatto notare come, se $x(t) = A \cos 2\pi f_0 t$ e $w(t) = \text{rect}_T(t)$, si ottiene che

$$W(f) = T \text{sinc}(fT)$$

ed il risultato (vedi Fig. 3.3) tende a due linee spettrali all'aumentare di T .

Valutiamo ora gli effetti derivanti dall'uso di una funzione finestra diversa da quella rettangolare. Sa ad esempio si sceglie di adottare una finestra triangolare di eguale base, il risultato mostrato pag. 41 permette di ottenere

$$W(f) = \mathcal{F}\{w(t) = \text{tri}_T(t)\} = T \left[\text{sinc}\left(\frac{fT}{2}\right) \right]^2$$

La finestra triangolare esibisce un andamento nel tempo più dolce rispetto al rect , e ciò si riflette in una maggiore concentrazione della sua trasformata alle frequenze più basse. Infatti, $W(f)$ ha ora un lobo principale di estensione doppia (il primo zero si trova ad $f = \frac{2}{T}$ anziché ad $\frac{1}{T}$ come per il rect), mentre le "code" laterali decrescono più rapidamente, andando a zero come $\frac{1}{f^2}$ (vedi fig. 3.4 (lato sinistro) a pag. 42).

Le conseguenze del diverso andamento in frequenza delle due funzioni finestra possono essere percepite qualora esse vengano applicate ad un segnale composto (ad esempio) da due sinusoidi a frequenza f_1 ed $f_2 > f_1$: per la linearità della trasformata, il risultato sarà la replica di $W(f)$ centrata alle due frequenze presenti. Allora, l'andamento delle "code" di $W(f)$ potrà avere conseguenze diverse per uno stesso scarto di frequenza $f_2 - f_1$: nel caso di una $w(t)$ rettangolare, si assisterà ad una sovrapposizione maggiore delle code delle repliche di $W(f)$, rispetto alla $w(t)$ triangolare, tanto da far preferire la seconda, nel caso in cui si desideri distinguere la presenza di entrambe le frequenze (potenzialmente di ampiezze differenti), anche se con scarto di frequenza $f_2 - f_1$ ridotto. Per scarti inferiori a $\frac{4}{T}$, può invece convenire adottare una finestra rettangolare, con il lobo principale di minore estensione.

Considerazioni analoghe a quelle ora svolte, possono essere intraprese per diverse scelte di funzione finestra, in dipendenza dal particolare obiettivo della stima spettrale.

3.9.4 Trasformata discreta di Fourier

L'analisi in frequenza di un segnale può essere condotta ricorrendo a programmi di elaborazione numerica, a partire dai campioni $x_m = x(mT_c)$ dello stesso, prelevati¹⁵ ad intervalli fissi T_c . Disponendo di una sequenza di N valori x_m , $m = 0, 1, \dots, N-1$, si definisce DISCRETE FOURIER TRANSFORM (**DFT**) la nuova sequenza

$$X_n = \frac{1}{N} \sum_{m=0}^{N-1} x_m e^{-j2\pi \frac{m}{N} n} \quad (3.3)$$

univocamente definita per $n = 0, 1, \dots, N-1$, e che costituisce una approssimazione¹⁶ della sequenza di campioni della trasformata $X(f) = \mathcal{F}\{x(t)\}$, calcolata per $f = \frac{n}{NT_c}$, e divisa per NT_c :

$$X_n \simeq \frac{1}{NT_c} X\left(f = \frac{n}{NT_c}\right)$$

¹⁵Nel prossimo capitolo è illustrato come scegliere T_c .

¹⁶Una prima fonte di approssimazione deriva dall'operazione di finestrazione legata all'uso di un numero finito di campioni, operando quindi su $y(t) = x(t)w(t_c)$ anziché su $x(t)$. Per analizzare le altre fonti di

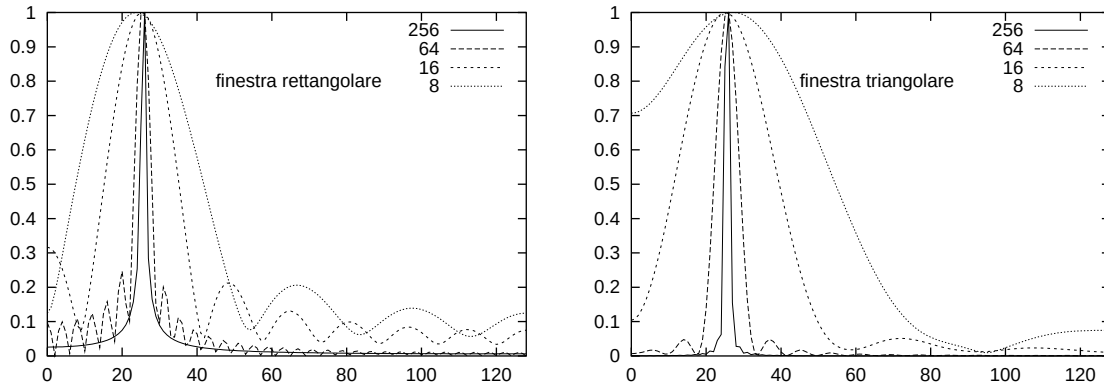


Figura 3.5: DFT di una sinusoide a 10 Hz, $f_c = 100$ Hz, con finestre di diversa lunghezza

Notiamo subito che la (3.3) è valida per qualsiasi n , ed ha un andamento periodico con periodo N , a cui corrisponde una frequenza $f = \frac{1}{T_c}$, in accordo con la separazione tra le repliche spettrali prevista dal teorema del campionamento; per questo motivo, qualora il segnale originario $x(t)$ contenga componenti a frequenze maggiori di $\frac{1}{2T_c}$, gli X_n con indici prossimi ad $\frac{N}{2}$ presentano errore di aliasing. A titolo di esempio, in fig. 3.5 sono riportati i valori $|X_n|$ per una DFT a 256 punti¹⁷, adottando due diverse finestre di analisi, al variare della loro estensione, espressa come numero di campioni.

Il passaggio dai campioni x_m alla sequenza X_n è invertibile, ricorrendo alla INVERSE DISCRETE FOURIER TRASFORM (**IDFT**)

$$x_m = \sum_{n=0}^{N-1} X_n e^{j2\pi \frac{n}{N} m}$$

che per m esterno a $[0, N-1]$ continua a valere, ed assume valori periodici, coerentemente a quanto accade per lo sviluppo in serie di Fourier. Infatti, i valori X_n sono equivalenti

approssimazione, iniziamo a scrivere l'espressione di $X(f) = \mathcal{F}\{y(t)\}$ per $f = \frac{n}{NT_c}$:

$$X\left(\frac{n}{NT_c}\right) = \int_0^{(N-1)T_c} x(t) e^{-j2\pi \frac{n}{NT_c} t} dt \simeq \sum_{m=0}^{N-1} x_m \cdot \int_0^{(N-1)T_c} \text{sinc}\left(\frac{t-mT_c}{T_c}\right) e^{-j2\pi \frac{n}{NT_c} t} dt$$

in cui la seconda eguaglianza utilizza l'interpolazione $x(t) = \sum_{m=-\infty}^{\infty} x_m \cdot \text{sinc}\left(\frac{t-mT_c}{T_c}\right)$ fornita dal teorema del campionamento (vedi prossimo capitolo), ed introduce una seconda fonte di approssimazione legata all'intervallo *finito* di variazione per m . L'ultimo integrale è a sua volta una approssimazione (a causa degli estremi di integrazione limitati) della trasformata di sinc $\left(\frac{t-mT_c}{T_c}\right)$, pari a $T_c \text{rect}_{\frac{1}{T_c}}(f) e^{-j2\pi f m T_c}$, che quando valutata per $f = \frac{n}{NT_c}$, fornisce il risultato

$$X\left(\frac{n}{NT_c}\right) = T_c \sum_{m=0}^{N-1} x_m e^{-j2\pi \frac{m}{N} n}$$

per valori $|n| \leq \frac{N}{2}$, a causa della estensione limitata (in frequenza) di $\text{rect}_{\frac{1}{T_c}}(f)$. E' però facile verificare che $X\left(\frac{n}{NT_c}\right)$ è periodica in n con periodo N , cosicché i valori assunti per $n = \frac{N}{2} + 1, \frac{N}{2} + 2$ sono uguali a quelli per $n = -\frac{N}{2} + 1, -\frac{N}{2} + 2$.

¹⁷Nel caso in cui i campioni di segnale siano in numero minore di quelli della DFT, i campioni mancanti sono posti a zero. Nelle figure sono mostrati solo i primi 128 valori, essendo i rimanenti speculari.

ad $\frac{1}{N}$ -esimo dei rispettivi coefficienti di Fourier, calcolati sulla finestra temporale di $x(t)$, come se questa fosse un periodo di un segnale periodico.

3.9.5 Trasformata di un gradino

Definiamo la funzione gradino come $g(t) = \begin{cases} 1 & \text{per } t > 0 \\ \frac{1}{2} & \text{per } t = 0 \\ 0 & \text{per } t < 0 \end{cases}$ che, fornendo $\int_{-\infty}^{\infty} |g(t)| dt =$

∞ , non dovrebbe avere una trasformata. Come già visto per la costante, il calcolo può essere condotto a termine come limite a cui tende la trasformata di una diversa funzione, il cui limite tende al gradino. Scegliamo quindi $g_\alpha(t) = e^{-\alpha t}$ per $t > 0$, per la quale risulta $\lim_{\alpha \rightarrow 0} g_\alpha(t) = g(t)$, e troviamo

$$G_\alpha(f) = \int_0^\infty e^{-\alpha t} e^{-j2\pi f t} dt = \frac{e^{-(\alpha + j2\pi f)t}}{-(\alpha + j2\pi f)} \Big|_0^\infty = \frac{1}{\alpha + j2\pi f} = \frac{\alpha - j2\pi f}{\alpha^2 + (2\pi f)^2}$$

Si può mostrare che $\lim_{\alpha \rightarrow 0} \Re\{G_\alpha(f)\} = \lim_{\alpha \rightarrow 0} \frac{\alpha}{\alpha^2 + (2\pi f)^2} = \frac{1}{2}\delta(t)$, mentre risulta $\lim_{\alpha \rightarrow 0} \Im\{G_\alpha(f)\} = \lim_{\alpha \rightarrow 0} \frac{-j2\pi f}{\alpha^2 + (2\pi f)^2} = \frac{1}{j2\pi f}$, ottenendo¹⁸ in definitiva

$$G(f) = \mathcal{F}\{g(t)\} = \frac{1}{2} \left(\delta(t) - \frac{j}{\pi f} \right)$$

3.9.6 Sintesi delle proprietà della trasformata di Fourier

PROPRIETÀ	$z(t)$	$Z(f) = \mathcal{F}\{z(t)\}$
Linearità	$ax(t) + by(t)$	$aX(f) + bY(f)$
Coniugato	$x^*(t)$	$X^*(-f)$
Cambiamento di scala	$x(at)$	$\frac{1}{a}X\left(\frac{f}{a}\right)$
Ritardo	$x(t - T)$	$X(f) e^{-j2\pi f T}$
Traslazione in frequenza	$x(t) e^{j2\pi f_0 t}$	$X(f - f_0)$
Modulazione di ampiezza	$x(t) \cos 2\pi f t$	$\frac{1}{2}X(f - f_0) + \frac{1}{2}X(f + f_0)$
Prodotto in frequenza	$\int_{-\infty}^{\infty} x(\tau) y(t - \tau) d\tau$	$X(f) Y(f)$
Prodotto nel tempo	$x(t) y(t)$	$\int_{-\infty}^{\infty} X(\sigma) Y(f - \sigma) d\sigma$
Dualità	$X(t)$	$x(-f)$
Simmetria coniugata	$x(t)$ reale	$X(f) = X^*(-f)$
Derivazione	$\frac{d}{dt}x(t)$	$j2\pi f \cdot X(f)$
Integrazione	$\int_{-\infty}^t x(\theta) d\theta$	$\frac{X(f)}{j2\pi f}$

¹⁸In realtà per $f = 0$ si ottiene che $\Im\{G_\alpha(f)\} = 0$, e lo stesso vale per $G(f)$, ossia $\Im\{G(f = 0)\} = 0$.

Capitolo 4

Campionamento e Trasmissione dei Dati

In questo capitolo è esposta la teoria che consente di rappresentare un segnale per mezzo dei suoi *campioni* temporali, permettendone la trasmissione in forma numerica. I campioni infatti sono una sequenza di *numeri*¹. Inoltre, si discute di come trasmettere segnali numerici (che siano ottenuti o meno per campionamento) su canali analogici, e sono introdotti i concetti di *codifica di canale* e di *linea*. Infine, sono illustrate le tecniche di moltiplicazione temporale dei segnali dati, ed accennate alcune problematiche attinenti le reti telefoniche commutate.

4.1 Teorema del Campionamento

Un segnale con spettro nullo a frequenze maggiori di W , è univocamente definito a partire dai valori che assume agli istanti $t = \frac{n}{2W}$, con n intero.

La frequenza $2W$ è chiamata *frequenza di Nyquist*. In virtù del teorema, l'andamento di un segnale $x(t)$ limitato in banda tra $-W$ e W può essere ricostruito in base ai suoi campioni, presi a frequenza doppia della sua banda a frequenze positive, per mezzo della formula:

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2W}\right) \cdot \text{sinc}\left(2W\left(t - \frac{n}{2W}\right)\right) \quad (4.1)$$

¹DIGITS in inglese, da cui il termine *digitale* come sinonimo di *numerico*.

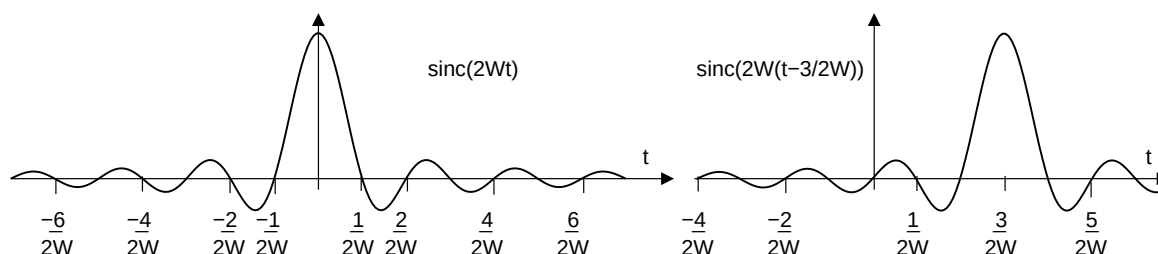


Figura 4.1: La funzione $\text{sinc}(2Wt)$ centrata in 0 e traslata in $\frac{3}{2W}$

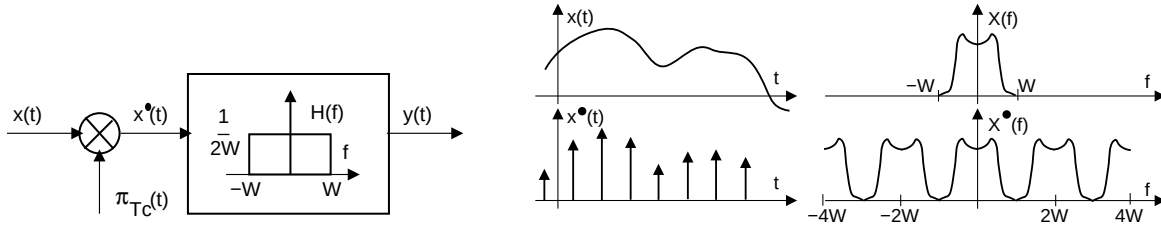


Figura 4.2: Circuito di campionamento e restituzione, e spettri dei segnali

in cui la funzione $\text{sinc}(2Wt) = \frac{\sin 2\pi Wt}{2\pi Wt}$ è mostrata in Fig. 4.1, assieme ad una sua replica traslata.

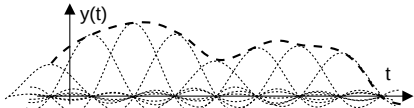
Per dimostrare il risultato, studiamo il circuito riportato in Fig. 4.2, che mostra uno schema simbolico che (come vedremo) realizza le stesse operazioni della formula di ricostruzione, in cui si opera un campionamento con periodo $T_c = \frac{1}{2W}$. Calcoliamo innanzitutto lo spettro di ampiezza $X^*(f)$ del segnale che esce dal moltiplicatore, che ha subito un'alterazione notevole rispetto a quello di $X(f)$ in ingresso. Infatti, il segnale $x^*(t)$ ha uno spettro di ampiezza

$$\begin{aligned} X^*(f) &= \mathcal{F}\{x(t) \cdot \pi_{T_c}(t)\} = X(f) * \frac{1}{T_c} \Pi_{\frac{1}{T_c}}(f) = X(f) * \frac{1}{T_c} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_c}\right) = \\ &= 2W \cdot \sum_{n=-\infty}^{\infty} X(f) * \delta(f - n2W) = 2W \cdot \sum_{n=-\infty}^{\infty} X(f - n2W) \end{aligned}$$

dove il penultimo passaggio scambia l'integrale di una somma con una somma di integrali, e l'ultimo passaggio tiene conto della proprietà di convoluzione con un impulso. In definitiva si è mostrato che $X^*(f)$ è costituito dalle repliche di $X(f)$ centrate a multipli della frequenza di campionamento. Pertanto, il filtro passa-basso $H(f)$ (chiamato anche con il nome di *filtro di restituzione*) lascia passare solo una delle repliche spettrali, e dunque è evidente come $Y(f) = \frac{1}{2W} 2W \cdot X(f) = X(f)$.

Per quanto riguarda la formula di ricostruzione che fa uso dei campioni $x(\frac{n}{2W})$ e delle funzioni $\text{sinc}(2Wt)$, anch'essa deriva dallo schema illustrato. Infatti $y(t)$ è il risultato della convoluzione tra $x^*(t)$ e $h(t) = \mathcal{F}^{-1}\{H(f)\} = \mathcal{F}^{-1}\{\frac{1}{2W} \text{rect}_{2W}(f)\} = \text{sinc}(2Wt)$ e dunque ogni impulso di cui è composto $x^*(t)$, quando convoluto con $h(t)$, ne trasla la forma d'onda al proprio istante. In formule:

$$\begin{aligned} y(t) &= \left[x(t) \cdot \sum_n \delta(t - nT_c) \right] * h(t) = \sum_n x(nT_c) \delta(t - nT_c) * \text{sinc}(2Wt) \\ &= \sum_n x(nT_c) \text{sinc}(2W(t - nT_c)) \end{aligned}$$



Questo risultato mostra come il teorema del campionamento definisca essenzialmente una formula di interpolazione: i valori del segnale ricostruito hanno l'esatto valore dei campioni di segnale negli istanti di campionamento, mentre negli istanti intermedi il valore si forma dalla somma di tutte le "code" dei sinc adiacenti. Il processo di costruzione grafica ora descritto è riportato nella figura precedente.

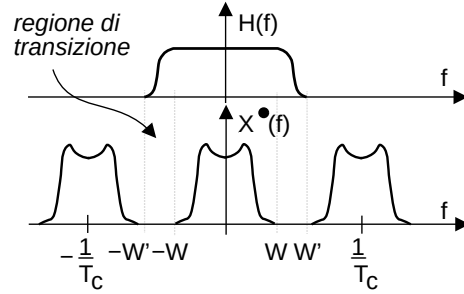
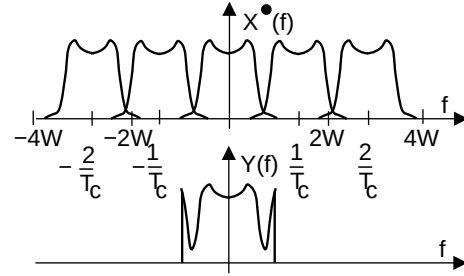
4.1.1 Aliasing

Questo termine ha origine dalla parola inglese² *alias* (copia, clone) e sta ad indicare il fenomeno che si produce nell'applicare il teorema del campionamento quando i requisiti non sono soddisfatti, e cioè quando la frequenza di campionamento è inferiore alla frequenza di Nyquist, ossia $\frac{1}{T_c} < 2W$ (ovvero $T_c > \frac{1}{2W}$). In questo caso le repliche spettrali che compongono $X^*(f)$ sono più ravvicinate, e si sovrappongono (l'aliasing è indicato anche come *fold-over*, *ripiegamento*).

Quando questo avviene, il filtro passa-basso di restituzione non è più in grado di estrarre la replica centrata in $f = 0$ (vedi figura a lato), e dunque alla sua uscita è presente $y(t) \neq x(t)$, che si differenzia da $x(t)$ in particolar modo per i contenuti energetici nella regione delle frequenze più elevate. In un segnale audio, ad esempio, ci si accorge che c'è aliasing quando è udibile una distorsione (rumore) congiuntamente ai passaggi con maggior contenuto di alte frequenze.

Il fenomeno dell'aliasing può insorgere, oltre che nel caso in cui si commetta il banale errore di adottare $T_c > \frac{1}{2W}$, anche a causa di una imperfetta limitazione in banda di $x(t)$ (che viene in genere filtrato proprio per accertarsi che sia $X(f) \approx 0$ con $|f| > W$).

Altri problemi possono essere causati dal filtro di restituzione $H(f)$, che difficilmente si riesce a realizzare ideale. Questo può presentare infatti una regione di transizione tra banda passante e banda soppressa di larghezza non nulla (vedi figura). In questo caso occorre sovracampionare con periodo $T_c = \frac{1}{2W'} < \frac{1}{2W}$, in modo che le repliche spettrali siano più distanziate tra loro, e quindi il filtro di ricostruzione possa isolare la replica centrale.



4.1.2 Energia di un segnale campionato

Si può dimostrare che le funzioni *sinc* costituiscono una base di rappresentazione ortogonale, in quanto

$$\int_{-\infty}^{\infty} \text{sinc}(2W(t - kT_c)) \text{sinc}(2W(t - hT_c)) dt = \begin{cases} 0 & \text{se } h \neq k \\ \frac{1}{2W} & \text{se } h = k \end{cases}$$

Pertanto, il valore dell'energia di un segnale limitato in banda è calcolabile a partire dai suoi campioni, e vale:

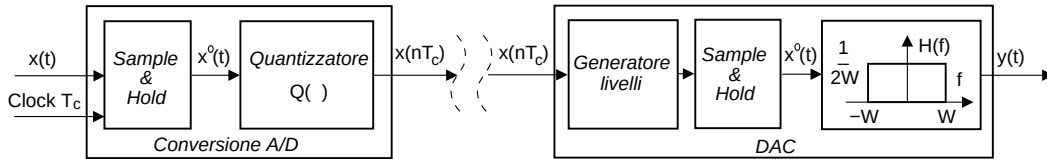
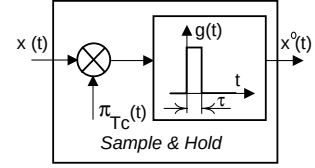
$$\begin{aligned} \mathcal{E}_x &= \int_{-\infty}^{\infty} x(t) x^*(t) dt = \sum_k \sum_h x_k x_h^* \int_{-\infty}^{\infty} \text{sinc}(2W(t - kT_c)) \text{sinc}(2W(t - hT_c)) dt \\ &= \sum_k \sum_h x_k x_h^* \frac{1}{2W} \delta(h, k) = \frac{1}{2W} \sum_k |x_k|^2 \end{aligned}$$

²In realtà *alias* è di origine latina !!!

4.1.3 Uso pratico

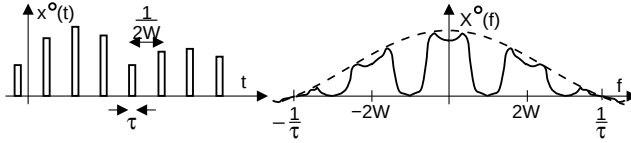
Lo schema proposto in Fig. 4.2 aveva il solo scopo di visualizzare gli aspetti teorico-matematici del teorema del campionamento.

Per coglierne i lati pratici, consideriamo innanzitutto che *non viene* generato il segnale $X^\bullet(f)$, in quanto il campionamento di $x(t)$ è realizzato per mezzo di un circuito *Sample and Hold* (S&H, ovvero campiona e mantiene) il quale produce una uscita costante (per un tempo τ), pari al valore assunto dall'ingresso negli istanti di campionamento (detti di *clock* = orologio); l'uscita del S&H viene quindi *quantizzata*, ovvero misurata e convertita in un valore numerico dal dispositivo $Q(\cdot)$.



Campionamento e restituzione nel mondo reale

I valori così ottenuti possono essere memorizzati, oppure trasmessi. Per ricostruire il segnale originario si adotta un DAC (*Digital to Analog Converter*, ossia Convertitore Digitale-Analogico) che può essere realizzato dai tre componenti mostrati in figura, ossia un dispositivo che per ogni diverso valore numerico genera un segnale di ampiezza pari ad uno dei livelli di quantizzazione, un S&H ed un filtro passa-basso di restituzione³.



Osserviamo ora come il S&H “emuli” il segnale $x^\bullet(t)$, realizzando al suo posto il segnale $x^\circ(t)$, mediante un treno di impulsi rettangolari modulati in ampiezza, in accordo allo schema

di principio disegnato al suo interno. Pertanto, il filtro di ricostruzione *non* è alimentato da $x^\bullet(t)$, ma dal segnale $x^\circ(t) = \sum_n x(nT_c) \cdot \text{rect}_\tau(t - nT_c)$. Per determinare quale sia in questo caso l'uscita del filtro di restituzione $H(f)$, valutiamone l'ingresso:

$$\begin{aligned} x^\circ(t) &= \sum_n x(nT_c) \cdot \text{rect}_\tau(t) * \delta(t - nT_c) = \\ &= \text{rect}_\tau(t) * \sum_n x(nT_c) \cdot \delta(t - nT_c) = \text{rect}_\tau(t) * x^\bullet(t) \end{aligned}$$

e dunque $X^\circ(f) = X^\bullet(f) \cdot \tau \text{sinc}(f\tau)$.

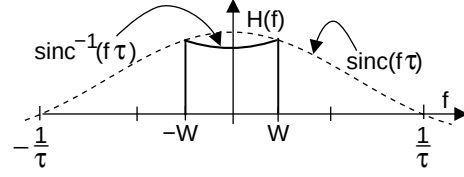
Osserviamo quindi che usare rettangoli di base $\tau < T_c$ al posto degli impulsi, equivale a moltiplicare $X^\bullet(f)$ per un involuppo di tipo $\frac{\sin x}{x}$ che, seppur con $\tau \ll T_c$ non causa grossi inconvenienti (gli zeri posti ad $\frac{1}{\tau}$ si allontanano dall'origine e $\frac{\sin x}{x}$ vicino ad $x = 0$ è quasi costante), per τ prossimo a T_c produce una alterazione dell'ampiezza della replica in banda base.

³In effetti, il DAC necessita di un segnale di temporizzazione, sincronizzato con T_c . Questo segnale può essere trasmesso separatamente, o essere ri-generato localmente a partire dalla stima della velocità alla quale sono ricevuti i valori $x(nT_c)$.

In tal caso (τ è noto) il filtro di ricostruzione può essere realizzato in modo da avere un andamento *inverso* a quello del $\frac{\sin x}{x}$, e tale che $H(f) \cdot \tau \text{sinc}(f\tau) = \text{costante}$. Infatti, questo accorgimento prende il nome di $\frac{\sin x}{x}$ *correction*.

All'appendice 4.4.1 è illustrato un metodo di MULTIPLAZIONE di più segnali campionati in una unica trasmissione.

La sezione seguente, pur esibendo alcuni aspetti analitici legati a quelli introdotti per il campionamento, ne è del tutto indipendente.



4.2 Trasmissioni numeriche

Ci occupiamo ora del problema di trasmettere su di un canale analogico una informazione che “nasce” discreta, a prescindere se derivi da un segnale campionato o meno.

4.2.1 Trasmissione numerica su canale analogico di Banda Base

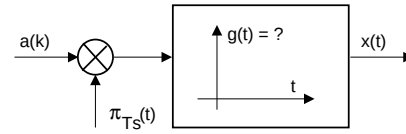
4.2.1.1 Segnale dati

Consideriamo una sorgente discreta, che produca simboli a_k ad una frequenza f_s , con cui costruiamo un segnale $x(t)$ idoneo ad essere trasmesso su un canale analogico

$$x(t) = \sum_k a_k \cdot g(t - kT_s) \quad (4.2)$$

come illustrato nella figura a lato⁴, del tutto simile al S&H introdotto al § 4.1.3, tranne che ora $g(t)$ è generico.

Svolgiamo ora una analisi allo scopo di determinare un criterio per scegliere $g(t)$, evidenziando subito i tre aspetti che sono alla base della sua corretta individuazione.



Costruzione del SEGNALE DATI

4.2.1.2 Requisiti per l'impulso di trasmissione

Limitazione di banda Osserviamo innanzitutto che (sotto opportune condizioni) la densità di potenza di (4.2) risulta pari a $\mathcal{P}_x(f) = \sigma_A^2 \frac{|G(f)|^2}{T}$ (vedi pag. 165), e quindi la sua trasmissione inalterata è possibile solo su di un canale che presenti una risposta in frequenza di tipo *passa-tutto* (ossia con modulo costante e fase lineare), od almeno simile, nella banda di frequenze occupata da $G(f)$.

Nel caso in cui, ad es., si adotti $g(t) = \text{rect}_\tau(t)$, allora $G(f) = \tau \text{sinc}(f\tau)$, con $\tau \leq T_s$; pertanto, il primo passaggio per zero è a frequenza $\frac{1}{\tau} \geq \frac{1}{T_s}$, ed il suo spettro può essere trascurato solo dopo qualche multiplo di tale valore.

Se invece il canale sopprime dal segnale dati (4.2) contenuti frequenziali non trascurabili, allora gli impulsi $g(t)$ si deformano e non sono più rettangolari; in particolare,

⁴Nel caso in cui a_k assume valori discreti in un alfabeto ad L livelli, codificati con $M = \lceil \log_2 L \rceil$ cifre binarie (*bit*) (il simbolo $\lceil \cdot \rceil$ rappresenta l'intero superiore), la trasmissione convoglia una *frequenza binaria* di valore pari a $f_b \left[\frac{\text{bit}}{\text{secondo}} \right] = M \left[\frac{\text{bit}}{\text{simbolo}} \right] \cdot f_s \left[\frac{\text{simboli}}{\text{secondo}} \right]$.

possono estendersi per una durata maggiore di T_s , causando problemi di *interferenza tra simboli*⁵ (ISI).

Limitazione nel tempo Il problema della limitazione di banda potrebbe essere risolto adottando, in linea di principio, un impulso elementare di tipo $g(t) = \text{sinc}\left(\frac{t}{T_s}\right)$, che ha trasformata $G(f) = T_s \cdot \text{rect}_{\frac{1}{T_s}}(f)$ strettamente limitata in banda, con frequenza massima $W = \frac{1}{2T_s}$, e che non subisce alterazioni purché il canale abbia un comportamento “passa tutto” in tale ristretto intervallo di frequenze. Notiamo che questa $g(t)$ passa da zero per $t = nT_s$, e pertanto non provoca interferenza tra simboli collocati agli istanti nT_s , come verificabile notando che in tal caso l'espressione (4.2) risulta del tutto simile alla (4.1); ora però *non siamo interessati* ai valori di $x_{PAM}(t)$ tra due istanti, mentre desideriamo unicamente recuperare i singoli valori originali, che troviamo *in modo esatto* agli istanti $t = nT_s$.

Lo svantaggio di adottare una forma d'onda $g(t)$ limitata in frequenza, è che il suo andamento è illimitato nel tempo, e dunque $g(t)$ può essere realizzata solo in modo approssimato⁶.

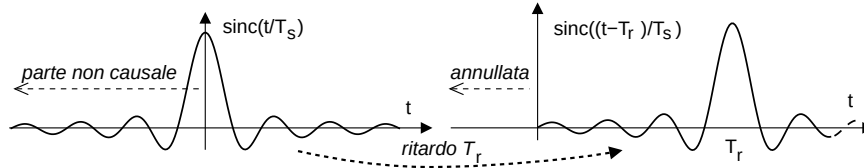
Limitazione di precisione Abbiamo appena mostrato come, adottando una $g(t) = \text{sinc}\left(\frac{t}{T_s}\right)$, si evita l'interferenza tra simboli, purché i campioni vengano prelevati esattamente agli istanti nT_s ⁽⁷⁾. Al di fuori di tali istanti infatti, il valore del segnale dipende dal valore delle code degli impulsi $g(t)$ centrati sugli altri simboli.

L'orologio (*clock*) del ricevitore, però, non ha una precisione infinita, e gli istanti di campionamento saranno affetti da errori di fase. Pertanto, è interessante ricercare una soluzione per $g(t)$ che dia luogo ad errori quanto più ridotti possibile, in presenza di errori di precisione nella determinazione degli istanti di campionamento.

Riepilogando: vorremmo soddisfare contemporaneamente le esigenze:

⁵Indicando con $g(t)$ la forma dell'impulso, e con $h(t)$ la risposta impulsiva del canale, all'uscita dello stesso il segnale dati è composto (vedi pag. 55) da impulsi con andamento pari a $g'(t) = g(t) * h(t)$, con una durata che può divenire maggiore di T_s , essendo questa pari alla somma delle durate di $g(t)$ e $h(t)$. Se ciò avviene, l'impulso $a_n g(t - nT_s)$ si estende a valori di $t > (n+1)T_s$, e quindi $x_{PAM}((n+1)T_s) = a_{n+1}g(0) + a_n g(T_s)$, introducendo un errore pari a $a_n g(T_s)$, detto appunto *interferenza tra simboli*.

⁶Un sistema fisico che debba realizzare una risposta impulsiva $g(t) = \text{sinc}\left(\frac{t}{T_s}\right)$, non può presentare $g(t) \neq 0$ per $t < 0$: questo equivarrebbe infatti ad un sistema in grado di produrre una uscita *prima ancora* che sia applicato un segnale al suo ingresso.



Se $g(t)$ ha estensione temporale illimitata, occorre ricorrere ad una versione ritardata e limitata $g'(t) = \begin{cases} g(t - T_R) & \text{con } t \geq 0 \\ 0 & \text{altrimenti} \end{cases}$. Se $T_R \gg T_s$, l'entità dell'approssimazione è accettabile, ed equivale ad un semplice ritardo pari a T_R ; d'altro canto, quanto maggiore è la durata della risposta impulsiva, tanto più difficile (ossia costosa) risulta la realizzazione del filtro relativo.

⁷Al contrario, se $g(t) = \text{rect}_{T_s}(t)$, il campionamento può avvenire ovunque nell'ambito del periodo di simbolo, ma si torna al caso di elevata occupazione di banda.

1. Occupare una banda contenuta
2. Ricorrere ad un filtro poco complesso
3. Ridurre la sensibilità agli errori di campionamento

Per i punti 2 e 3, è sufficiente adottare $g(t)$ di tipo rettangolare, generando un segnale dati del tipo $x(t) = \sum_k a_k \cdot \text{rect}_\tau(t - kT_s)$, che ha lo svantaggio di occupare un banda infinita⁸ e quindi la sua trasmissione “intatta” è possibile solo per canali ideali⁹.

Prima di esporre una soluzione di compromesso a tutti e 3 i problemi, consideriamo di inglobare in $g(t)$ sia l'effetto del filtro di trasmissione che di quello del canale $H(f)$; indichiamo pertanto il filtro di trasmissione come $g'(t)$. Con riferimento alla figura a lato, in ⁽¹⁰⁾ viene svolta la convoluzione tra $x(t)$ ed $h(t)$, che evidenzia come il segnale ricevuto $y(t)$ abbia espressione

$$y(t) = \sum_k a_k \cdot g'(t - kT_s) \quad \text{in cui} \quad g'(t) = g(t) * h(t)$$

Pertanto, i risultati che troveremo (validi per $g'(t)$) individueranno in realtà un formatore di impulsi con $G(f) = \frac{G'(f)}{H(f)}$.

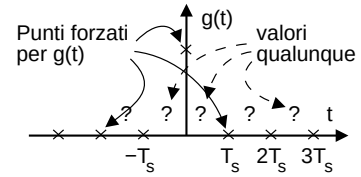
4.2.1.3 Condizioni di Nyquist

Torniamo a riferirci alla (4.2) per osservare che, affinché $x(nT_s)$ dipenda da uno solo degli $\{a_k\}$, deve risultare

$$g[(n - k)T_s] = \begin{cases} 1 & \text{se } n = k \\ 0 & \text{se } n \neq k \end{cases} \quad (4.3)$$

e cioè $g(t)$ deve passare da zero in tutti gli istanti multipli di T_s , tranne che per $t = 0$ dove deve valere 1; infatti, in tal caso si ottiene:

$$x(nT_s) = \sum_k a_k \cdot g((n - k)T_s) = a_n$$



⁸Che questo sia il caso, può essere verificato per alcuni segnali che abbiamo studiato o studieremo:

1. Segnale campionato. In questo caso $a_k = s(kT_c)$ sono i campioni di segnale, ed abbiamo visto che $x^\circ(t)$ ha spettro periodico in frequenza, con un involuppo di ampiezza dato da $\text{sinc}(\frac{f}{\tau})$;
2. Segnale periodico. Ponendo $a_k = \pm 1$ si genera un'onda quadra, il cui spettro è a righe con lo stesso involuppo citato prima;
3. Segnale dati. Se a_k sono variabili aleatorie statisticamente indipendenti (si dimostrerà che) $X(f)$ è di tipo continuo, con involuppo ancora pari a $\text{sinc}(\frac{f}{\tau})$.

⁹Nella tabella che segue è riportata l'occupazione di banda necessaria a contenere 10 lobi del $\text{sinc}(\frac{f}{\tau}) = \frac{1}{\tau} \mathcal{F}\{\text{rect}_\tau(t)\}$, per τ pari al periodo di simbolo T_s , in modo da dare un'idea delle specifiche necessarie al canale: osserviamo allora che il rettangolo può andare bene a basse velocità di trasmissione, infatti già per 10 Msimboli/sec, velocità di una LAN (*Local Area Network*, ossia una rete “locale” tra computer in uno stesso edificio), occorrono 100 MHz di banda.

f_s	apparato	T_s	$1/T_s$	$10/T_s$
$2.4 \cdot 10^3$	Modem (anni '80)	$4.2 \cdot 10^{-3}$	$2.4 \cdot 10^3$	24 KHz
$28.8 \cdot 10^3$	Modem (anni '90)	$3.5 \cdot 10^{-5}$	$28.8 \cdot 10^3$	288 KHz
$10 \cdot 10^6$	LAN	10^{-7}	10^7	100 MHz

$$\begin{aligned}
 {}^{10} y(t) &= \left[\sum_k a_k \cdot g(t - kT_s) \right] * h(t) = \left[g(t) * \sum_k a_k \cdot \delta(t - kT_s) \right] * h(t) = \\
 &= g(t) * h(t) * \sum_k a_k \cdot \delta(t - kT_s) = g'(t) * \sum_k a_k \cdot \delta(t - kT_s) = \\
 &= \sum_k a_k \cdot g'(t - kT_s)
 \end{aligned}$$

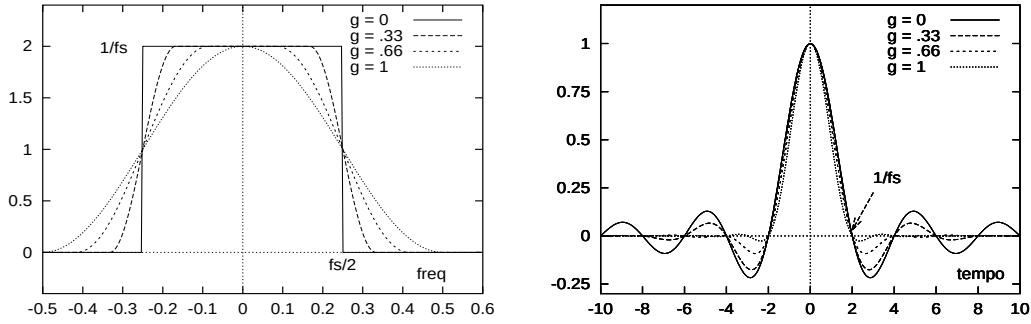


Figura 4.3: Caratteristica a coseno rialzato e impulso di Nyquist per $f_s = 0.5$, al variare di γ

Le condizioni (4.3) prendono il nome di *condizioni di Nyquist per l'assenza di interferenza intersimbolo* (ISI, *Inter Symbol Interference*) nel dominio del tempo. Se una forma d'onda $g(t)$ soddisfa tali condizioni, allora viene detta *impulso di Nyquist*⁽¹¹⁾.

Dalle condizioni di Nyquist *nel tempo* se ne derivano altre *in frequenza*, mediante i seguenti passaggi. Moltiplicando $g(t)$ per un treno di impulsi $\pi_{T_s}(t) = \sum_k \delta(t - kT_s)$, si ottiene

$$g(t) \cdot \pi_{T_s}(t) = \delta(t)$$

dato che $g(nT_s) = 0$ e $g(0) = 1$. Trasformando si ottiene:

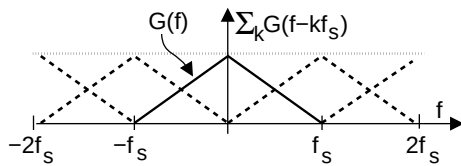
$$1 = G(f) * \frac{1}{T_s} \cdot \Pi_{\frac{1}{T_s}}(f) = G(f) * \frac{1}{T_s} \cdot \sum_k \delta\left(f - k\frac{1}{T_s}\right)$$

Indicando con $f_s = \frac{1}{T_s}$ la frequenza di simbolo, ed eseguendo le convoluzioni, risulta infine

$$\sum_k G(f - kf_s) = T_s$$

che rappresenta la condizione *in frequenza* per l'assenza di interferenza intersimbolo.

Il risultato ottenuto si interpreta considerando che una qualunque $G(f)$ va bene purché, se sommata con le sue repliche traslate di multipli di f_s , dia luogo ad una costante.



In questo caso si dice che $G(f)$ è una *caratteristica di Nyquist*. Notiamo che, seppure $G(f)$ possa essere qualsiasi, anche non limitata in banda, il nostro interesse è appunto per le $G(f)$ limitate in banda, come quella triangolare dell'esempio a lato.

Coseno Rialzato Una famiglia parametrica di caratteristiche di Nyquist limitate in banda, è quella cosiddetta a *coseno rialzato*, che è composta da 2 archi di coseno raccordati da una retta (vedi Fig. 4.3). La banda occupata ha espressione

$$B = \frac{f_s}{2} (1 + \gamma)$$

¹¹Ad esempio, l'impulso rettangolare è di Nyquist, in quanto $\text{rect}_{T_s}(t) = \begin{cases} 1 & \text{se } t < \frac{T_s}{2} \\ 0 & \text{se } t = kT_s \end{cases}$.

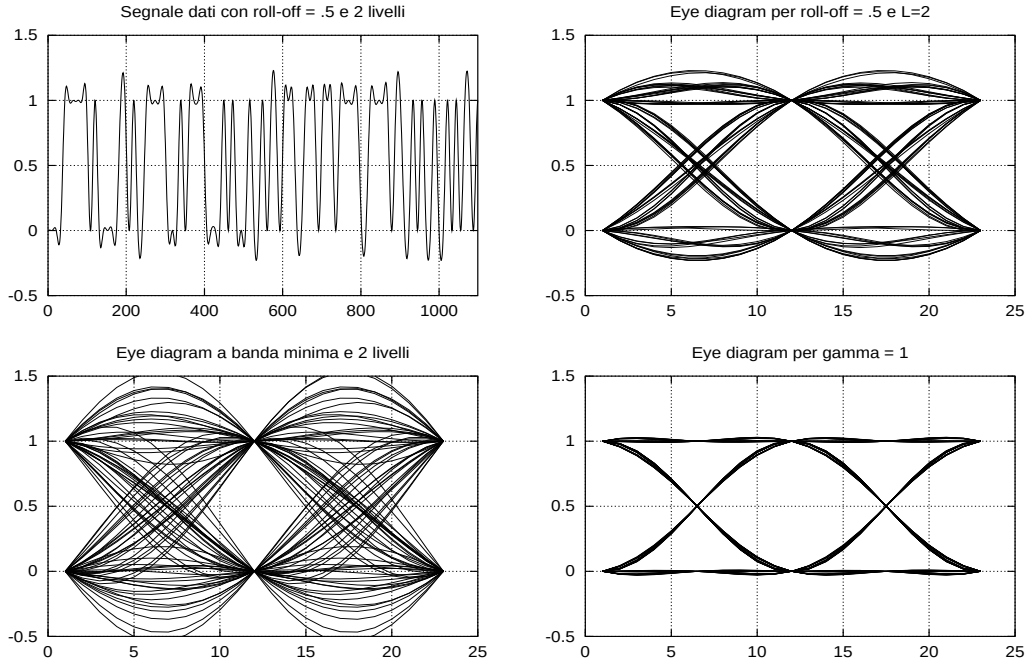


Figura 4.4: Segnale dati ed eye diagram per diversi valori di roll-off

in cui γ è chiamato *coefficiente di roll-off*¹², compreso tra 0 e 1, e rappresenta un indice di dispersione del ramo di coseno. La banda di $G(f)$ varia quindi da un minimo $B = f_s/2$ (per $\gamma = 0$) ad un massimo di $B = f_s$ (per $\gamma = 1$, nel qual caso $G(f)$ è proprio un periodo di coseno).

Il caso di *banda minima* si ottiene per $\gamma = 0$, ottenendo $G(f) = \frac{1}{f_s} \text{rect}_{f_s}(f)$, corrispondente ad una $g(t) = \text{sinc}(f_s t)$, come già discusso a pag. 54 al § sulla limitazione nel tempo. Occupare una banda inferiore a quella minima non è possibile, perchè in tal caso non sarebbero verificate le condizioni di Nyquist in frequenza, in quanto nella $\sum_k G(f - kf_s)$ resterebbero dei “buchi”.

Abbiamo già osservato alla nota (6) a pagina 54 come la realizzazione di $G(f)$ a banda minima sia difficile, e produca eccessiva sensibilità agli errori di campionamento. La situazione però migliora decisamente usando $\gamma > 0$, via via più grande. Con $\gamma \neq 0$ la $g(t)$ ha espressione

$$g(t) = \text{sinc}(tf_s) \cdot \frac{\cos \gamma \pi t f_s}{1 - (2\gamma t f_s)^2} \quad (4.4)$$

presentando una forma d'onda simile al $\frac{\sin(x)}{x}$, ma che va a zero molto più rapidamente, come valutabile visivamente osservando la Fig. 4.3. Pertanto, con $\gamma \rightarrow 1$ ogni singola onda $g(t - kT_s)$ estenderà il suo influsso ad un numero di impulsi limitrofi molto ridotto rispetto al caso $\gamma = 0$ in quanto le oscillazioni sono molto più smorzate, e dunque il termine di errore di ampiezza in presenza di un errore di istante di campionamento è ridotto, in quanto dipende da un minore numero di impulsi limitrofi.

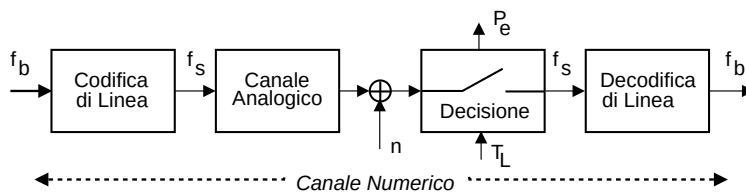
La fig. 4.4 mostra l'andamento del segnale dati con $g(t)$ fornito dalla (4.4), calcolata per $\gamma = 0.5$, e per a_k a due valori, pari a 0 e 1. Notiamo che, al di fuori degli istanti

¹²Il termine ROLL-OFF può essere tradotto come “rotola fuori”.

caratteristici, il segnale può assumere valori arbitrari, anche esterni alla dinamica degli a_k . E' inoltre mostrato il cosiddetto *diagramma ad occhio* (EYE DIAGRAM), così chiamato dalla sua forma, e che si ottiene visualizzando il segnale dati mediante un oscilloscopio con la base dei tempi sincronizzata al periodo di simbolo. Tale rappresentazione permette di meglio valutare la precisione di temporizzazione che è necessaria ad evitare ISI, e che è pari alla apertura orizzontale dell'occhio. Gli ultimi due diagrammi permettono il confronto con i casi rispettivamente a banda minima e per $\gamma = 1$, evidenziando l'influenza del roll-off.

4.2.2 Codifica di linea

Al primo capitolo (§ 1.2) abbiamo elencato le funzioni dei dispositivi previsti nel caso di trasmissioni numeriche, ed il nome di questa sezione si riferisce proprio ad una di tali funzioni. Ci riferiamo ora infatti agli accorgimenti necessari a trasmettere un segnale numerico su di un canale analogico, ed in particolare al caso in cui i requisiti *non siano* così stringenti da richiedere l'uso di una caratteristica di Nyquist. In tal caso, si adotta il nome di *codifica di linea*¹³, mentre nel primo capitolo la stessa funzione è stata assegnata al *modem*, termine che è più appropriato al caso di trasmissioni modulate (non in banda base), a cui ci dedicheremo al Capitolo 11.



In figura è riportato uno schema che mostra il segnale numerico prodotto dal codificatore di linea, caratterizzato da una *velocità di simbolo* f_s diversa da quella bi-

narica f_b propria del messaggio¹⁴; inoltre, si mostra come nel ricevitore avvenga un *processo di decisione* in merito a quale simbolo sia stato ricevuto, decisione potenzialmente errata, di cui vedremo in seguito come calcolare la probabilità di errore P_e .

4.2.2.1 Codici di Linea a banda infinita

Nel caso in cui si operi a bassa velocità, si può considerare il canale a banda infinita, e dunque adottare una forma d'onda rettangolare per il segnale di dati. A pag. 165 si mostra come determinare lo spettro del segnale dati (4.2) a simboli a_k indipendenti, che per $g(t) = \text{rect}_\tau(t)$ ha andamento di tipo $\text{sinc}(f\tau)$. In questo caso si ricorre ad una tra diverse soluzioni, che prendono il nome di *codici di linea*, e che sono effettivamente usati per trasmettere informazioni di natura numerica.

Elenchiamo qui alcuni codici di linea, assieme alle loro proprietà e caratteristiche, aiutandoci con gli esempi riportati in figura.

a) Neutrale o Unipolare NRZ

La sigla NRZ (*No-Return-to-Zero*) significa appunto che $g(t)$ “non torna a zero”, mentre *unipolare* è riferito alla presenza di soli valori positivi. Questo codice ha spettro di tipo $\text{sinc}(fT_s)$, con il primo zero a $f = 1/T_s$, e presenta una componente continua. Rima-

¹³Questi dispositivi sono a volte anche indicati come *line driver*.

¹⁴ T_s è il periodo di simbolo ed il suo inverso $f_s = 1/T_s$ è detto *frequenza di simbolo* (o *baud-rate*), detta anche *frequenza di segnalazione*, e si misura in simboli/secondo (detti appunto *baud*).

ne costante per dati costanti e ciò complica la sincronizzazione del clock del ricevitore¹⁵. Inoltre, lo zero ad $f = 1/T_s$ aggrava la situazione.

b) Unipolare RZ

La sigla RZ (*Return-to-Zero*) indica che l'impulso ha durata inferiore a T_s . Possiede una minore energia di a), ed occupa una banda doppia. Lo spettro ora è diverso da zero per $f = 1/T_s$, ma il segnale nel tempo si mantiene comunque costante per lunghe sequenze di zeri.

c) Polare NRZ

Ha media nulla (se 0 ed 1 sono equiprobabili) ma mantiene gli altri difetti del caso a).

d) Polare RZ

Ha media nulla (come sopra), potenza ad $f = 1/T_s$, non è mai costante.

e) Diphase NRZ (Manchester)

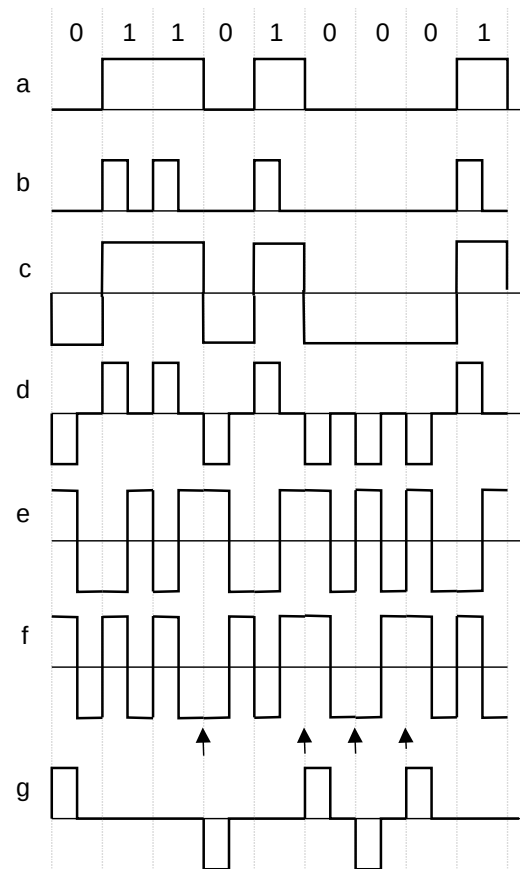
Codifica i dati nelle transizioni e per questo non è mai a media non nulla; rimane una ambiguità se si scambiano i fili. Come anche il seguente, occupa una banda di estensione intermedia ai casi precedenti.

f) Differenziale

Tra simboli contigui si inverte polarità se arriva un uno, mentre non si inverte se arriva uno zero (con le frecce). Banda come per e).

g) Alternate Mark Inversion (AMI)

Codifica gli 1 con polarità alternate, garantendo così l'assenza di valore medio, ma si perde la temporizzazione se sono presenti troppi zeri; per questo motivo, esistono altri codici (derivati) che inseriscono degli uni "in più".

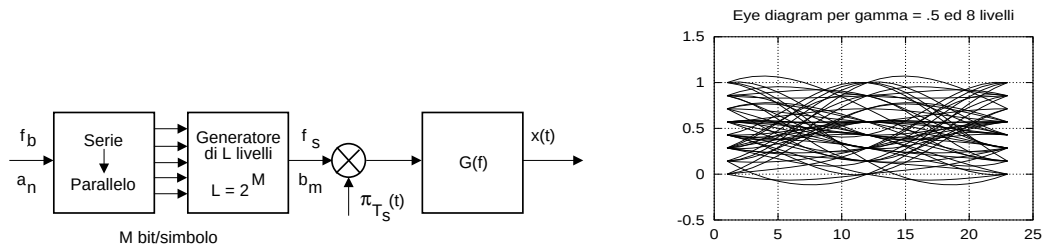


4.2.2.2 Trasmissione multilivello

Una volta stabilito il tipo di impulso da adottare (rettangolare o con caratteristica a coseno rialzato), può accadere che la banda a disposizione per la trasmissione sia insufficiente. Una soluzione di semplice attuazione è quella di ricorrere ad una trasmissione non più "binaria", con 2 soli livelli di segnale, ma con L possibili diversi livelli.

A tale scopo, occorre raggruppare M bits del messaggio a_n (che arrivano a velocità f_b bits/sec) in una unica parola binaria. Scegliendo $M = \log_2 L$, possiamo rappresentare ognuna delle $L = 2^M$ diverse configurazioni di bit con una di L diverse ampiezze del segnale, producendo così un segnale b_m , a velocità $f_s = f_b/M$, i cui simboli ad L livelli sono separati da un intervallo temporale pari a $T_s = MT_b$.

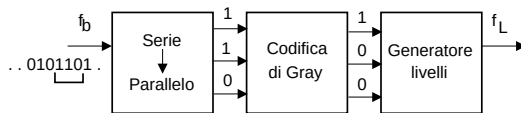
¹⁵ Il circuito che sincronizza l'orologio del campionario dello stadio ricevente con il periodo di simbolo del segnale ricevuto, effettua un confronto tra il *tempo* dell'orologio locale ed un *ritmo* presente nel segnale in arrivo, ed il segnale di errore alimenta un circuito di controreazione che mantiene il clock locale *al passo* con quello dei dati in arrivo. Un caso particolare di questo stesso principio è analizzato in 9.2.1.3 e 9.3.1.1 a proposito del PLL.



In questo modo l'occupazione di banda si riduce di *un M-esimo*, e dunque può essere ridotta semplicemente aumentando il numero di bit raggruppati in una singola parola. Se la potenza del segnale non viene aumentata, però, si assisterà ad un peggioramento della probabilità di errore del ricevitore, in quanto a parità di potenza di segnale, i livelli risultano ora ravvicinati, come mostrato dal diagramma ad occhio di figura per $L = 8$. L'argomento verrà ripreso al § 5.5.

4.2.2.3 Codice di Gray

Esponiamo ora un accorgimento che consente di ridurre l'effetto degli errori commessi dal decisore in presenza di trasmissione multilivello. Per fissare le idee, supponiamo che i valori dei livelli prodotti dal generatore serie-parallelo, ovvero i livelli sono "numerati in binario", e livelli contigui rappresentano configurazioni di bit "in sequenza naturale", come mostrato in figura. Allora, se trasmettiamo ad esempio il livello associato a 100, ed il decisore commette l'errore¹⁶ di ritenere di aver ricevuto il livello contiguo, che rappresenta la sequenza 011, abbiamo tutti e tre i bit sbagliati!



Ingresso binario	Uscita (livello)
100	111
101	110
111	101
110	100
010	011
011	010
001	001
000	000

Il procedimento illustrato, in presenza di un errore sul simbolo, produce un solo bit

ferimento alla tabella, osserviamo che (ad esempio) per trasmettere la sequenza 110 di ingresso, si usa il livello numero 100, ossia il quarto (partendo da zero), lo stesso dell'esempio precedente; se il decisore sbaglia e ritiene di aver ricevuto il terzo livello (011, stesso errore precedente), a questo il decodificatore di Gray associa la sequenza 010, che infatti differisce dall'originale per un solo bit (il primo).

Il procedimento illustrato, in presenza di un errore sul simbolo, produce un solo bit

¹⁶L'errore è causato dal rumore che, sommandosi al segnale ricevuto (vedi Fig a pagina 58), può falsarne l'ampiezza in modo che il decisore si trovi ad operare con valori diversi da quelli trasmessi.

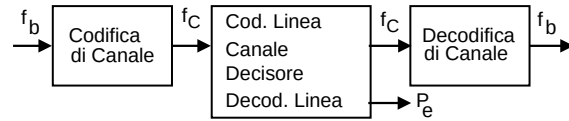
Il *codice di Gray* consiste in una tabella di conversione, che sostituisce ai bit uscenti dal convertitore serie-parallelo una diversa configurazione di bit. Possiamo immaginare l'operazione come quella di un accesso a memoria, in cui la parola originaria costituisce l'indirizzo, per mezzo del quale si individua la parola codificata da trasmettere al suo posto.

La conversione è biunivoca (a partire dal codice si risale alla parola originaria), e le parole del codice di Gray hanno la proprietà di rappresentare i livelli di segnale contigui come configurazioni di bit che differiscono solo in una cifra binaria (ossia in un bit). Con ri-

errato. Ciò comporta che la probabilità di osservare un bit errato è pari a $P_e^b = P_e^s/M$, con M pari al numero di bit/simbolo. Infatti, $P_e^b = \frac{N.Bit\ errati}{N.Bit\ totali} = \frac{N.Simboli\ errati}{M \cdot N.Simboli} = P_e^s \frac{1}{M}$.

4.2.3 Codifica di Canale

Con questo termine si è indicato, nel primo capitolo, il procedimento con il quale vengono “protette” le informazioni da trasmettere, aumentando il numero di bit inviati (passando da f_b a $f_C > f_b$ ⁽¹⁷⁾ bps,



vedi figura), in modo da rendere ridondante il messaggio e facilitare la gestione degli eventuali errori di trasmissione. In appendice § 4.4.3 è descritto un metodo particolare di codifica, mentre di seguito sono forniti principi e metodi di carattere generale.

4.2.3.1 Codici a blocchi

Senza ancora entrare nei dettagli, definiamo come codifica a blocchi quella operata aggiungendo, per ogni m bit in ingresso, altri k bit aggiuntivi, scelti in funzione dei primi m . Ovviamente in tal caso sarà $f_C = \frac{m+k}{m} f_b$.



Controllo di parità E' una codifica a blocchi che consiste nell'adottare $k = 1$, scegliendo questo bit in modo che nel gruppo di $m + 1$ bit ci sia un numero *pari* di uni¹⁸, applicando quindi una regola di parità *pari* (EVEN). Il caso opposto, ossia l'aggiunta di un bit in modo da rendere *dispari* il numero di uni, prende nome di parità ODD.

In entrambi i casi¹⁹, quando il ricevitore raggruppa i bit pervenuti, esegue un *controllo* detto appunto *di parità*, semplicemente contando il numero di uni, ed accorgendosi così se nella parola si sia verificato un errore (uno zero divenuto uno o viceversa). In tal caso, il ricevitore invierà all'altro estremo del collegamento una richiesta di ritrasmissione del gruppo di bit errati.

Il controllo di parità come ora descritto è tipicamente usato per i collegamenti seriali a bassa velocità, in particolare tra computer e terminale²⁰, stampante o *mouse*. Si parla in tal caso di *trasmissione asincrona*, o START-STOP, che viene descritta in appendice 4.4.2.

4.2.3.2 Codici a correzione

Si è detto che un codificatore di canale *a blocchi* può aggiungere k bit (e non uno soltanto!) ogni m . Questo può essere realizzato, ad esempio, se ognuno dei k bit rappresenta la parità calcolata su un sottogruppo degli m bit di ingresso. Dato che questi bit in più non

¹⁷Non si confonda questa f_C “frequenza di canale” con quella di *campionamento* di inizio capitolo... :-)

¹⁸Ad esempio, alla sequenza 001001verrà aggiunto uno 0, mentre a 010101 si aggiungerà ancora un 1, perché altrimenti gli uni complessivi sarebbero stati 3, che è dispari.

¹⁹Il ricevitore deve comunque essere al corrente del fatto se la parità sia ODD o EVEN !

²⁰Almeno così era, ai tempi in cui ancora esistevano questi apparati !!! :-)

sono qualsiasi, ma dipendono dai primi m (²¹), non tutte le 2^{m+k} possibili configurazioni possono essere emesse: tutte quelle possibili costituiscono il *codebook*, e prendono il nome di *codeword* (parole di codice). In presenza di errori, il decodificatore se ne accorge per il fatto di aver ricevuto una *non-codeword*²²; se le codeword sono sufficientemente “differenti” l’una dall’altra, allora si può anche tentare di *correggere* l’errore, anziché chiedere la ritrasmissione della parola.

Codice a ripetizione Un esempio molto semplice di codice a blocchi con proprietà correttive è il codice a ripetizione $k : 1$, che per ogni bit in arrivo ne produce k identici in uscita, in modo che (se gli errori sono indipendenti) il decisore possa correggere l’errore in base ad una “votazione a maggioranza” (*majority voting*)²³. In presenza di errori a gruppi, occorre ricorrere allo scrambler, descritto nel seguito.

Distanza di Hamming. La capacità di un codice, di correggere uno o più errori, oltre che rivelarne la presenza, è legata (come abbiamo detto) alla “differenza” tra le parole del codebook. Una misura di questa differenza è la *distanza di Hamming* (d_H), che rappresenta il minimo numero di bit diversi tra due parole di codice (ad esempio, nel codice a ripetizione $3 : 1$ si hanno 2 parole di codice, 000 ed 111, e risulta $d_H = 3$). Si dimostra che un codice di canale è in grado di correggere $\frac{d_H-1}{2}$ errori per parola e di rivelare $d_H - 1$ errori per parola.

FEC e ARQ Qualora un codice di canale sia utilizzato in modalità *correttiva*, si usa descrivere la codifica come una FORWARD ERROR CORRECTION, ossia una correzione di errore “in avanti”. Se invece il codice è usato solo per rivelare gli errori, allora sarà necessario prevedere l’uso di protocolli detti ARQ (AUTOMATIC REQUEST) che si occupano di chiedere la ritrasmissione della parola errata.

4.2.4 Scrambling

Nel caso in cui la distribuzione temporale degli errori non sia uniforme²⁴, ma possano invece presentarsi ravvicinati e nella stessa parola, i codici di canale falliscono nel loro scopo. La soluzione in questo caso è il cosiddetto *scrambler*²⁵ o *interleaver* che, a patto di

²¹Ad esempio, ognuno dei k bit aggiunti (detti di *controllo*) può essere calcolato applicando la regola della parità ad un sottoinsieme degli m bit di sorgente, con i sottoinsiemi eventualmente sovrapposti. Un codice del genere prende il nome di codice di *Hamming*.

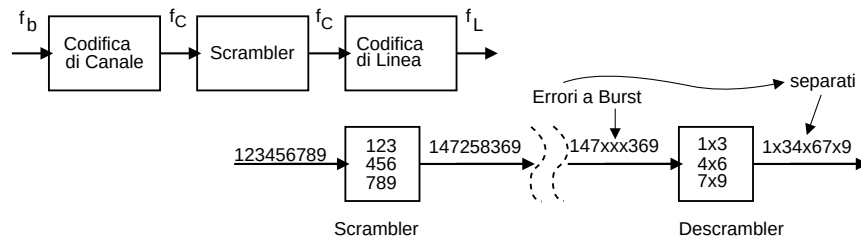
²²I k bit di controllo (dipendenti dagli altri m , tramessi inalterati) sono a volte indicati come CRC (che sta per CYCLIC REDUNDANCY CHECK, *Controllo Ciclico di Ridondanza*), e costituiscono appunto una quantità che viene ricalcolata al ricevitore in base ai primi m della parola, e confrontata con il CRC ricevuto per rivelare eventuali errori.

²³Poniamo, a titolo di esempio, che $k = 3$, e di dover trasmettere 0110. La sequenza diventa 000 111 111 000 e quindi, a causa di errori, ricevo 000 101 110 100. Votando a maggioranza, ricostruisco la sequenza corretta 0 1 1 0.

²⁴Se gli errori avvengono in maniera indipendente, la probabilità di un bit errato su N è pari a NP_e , cioè N volte la probabilità di un bit errato, mentre la probabilità di due bit errati è circa pari a $N(N-1)P_e^2$, e quindi molto minore se $P_e \ll 1$.

Più errori consecutivi o comunque prossimi vengono indicati in inglese come “errori a burst” che si traduce “*scoppio*”; mentre in italiano lo stesso evento è indicato come errori a *pacchetto*.

²⁵Letteralmente: arrampicatore, ma anche “arruffatore”, vedi *scrambled eggs*, le uova strapazzate dell’*english breakfast*.



accettare un ritardo, modifica l'ordine dei dati codificati, in modo che gli errori avvenuti su bit vicini si riflettano in errori su bit... lontani e quindi appartenenti a parole differenti. Ovviamente, occorre prevedere un processo inverso (*descrambling*) all'altro capo del collegamento. E' appena il caso di notare che lo scrambler (similmente al codice di Gray) *non altera* il numero dei bit trasmessi.

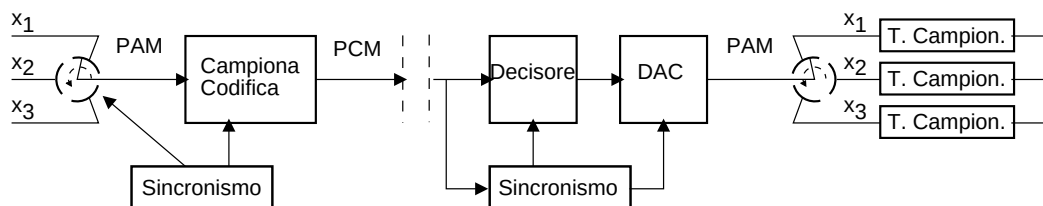
4.3 Reti di trasmissione

Tutti i casi di trasmissione finora illustrati fanno riferimento a collegamenti *punto-punto*, in cui una unica sorgente di informazione intende comunicare con un unico destinatario. Nella realtà, è assai più frequente il caso in cui i soggetti coinvolti nella comunicazione affidino la stessa ad una *rete* di collegamento, consegnando il messaggio al *nodo* di commutazione a cui hanno accesso. Il messaggio quindi, una volta determinato un percorso di *attraversamento* che coinvolga i nodi della rete più opportuni, giunge al destinatario, grazie anche alla presenza di informazioni aggiuntive, dette *di segnalazione*.

A completamento del presente capitolo, sono quindi illustrati alcuni argomenti propri delle *reti di telefonia*, dette anche a *commutazione di circuito*, allo scopo di fornire una panoramica dei principali aspetti delle stesse. Altri aspetti legati alle *reti di trasmissione dati* saranno illustrati al Capitolo 6, nel contesto della teoria del traffico, assieme alle reti a *commutazione di pacchetto*.

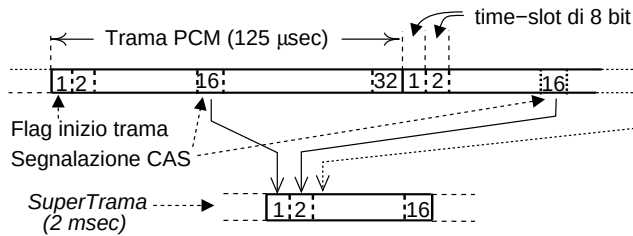
4.3.1 Multiplazione di tempo numerica

Le moderne reti telefoniche sono completamente numeriche²⁶, e trasportano i segnali vocali sotto forma di campioni, rappresentati ognuno da un gruppo di M bit. Uno stesso collegamento è utilizzato per trasportare contemporaneamente N segnali vocali, alternando nel tempo i rispettivi campioni, realizzando così una *multiplazione temporale* in forma numerica.



²⁶Quelle "non numeriche" sono ormai abbandonate...

Trama PCM In figura sono rappresentati tre *tributari*, campionati a turno alla stessa frequenza di 8 KHz, quantizzati ad 8 bit per campione, e trasmessi (8 bit alla volta²⁷) a turno (vedi § 4.4.1) su di un unico collegamento. Il segnale binario risultante prende il nome di *PCM* (PULSE CODE MODULATION²⁸). In figura è evidenziato inoltre un blocco di sincronismo necessario a ricostruire la corretta sequenza ricevuta, in modo da redistribuire correttamente i campioni ai filtri di restituzione.



La struttura temporale ripetitiva che ospita i campioni dei singoli tributari prende il nome di *trama* (FRAME²⁹), ed è composta 32 intervalli detti *time-slot*. Trenta di questi, ospitano a turno i bit di un campione proveniente da un numero massimo di 30 tributari³⁰, mentre i rimanenti due

convogliano le informazioni di segnalazione³¹, che indicano lo stato dei singoli collegamenti (il 16° intervallo) e forniscono il sincronismo relativo all'inizio della trama stessa (il primo). La velocità binaria complessiva risulta quindi di 32 intervalli * 8 bit/intervallo * 8000 campioni/secondo = 2048000 bit per secondo; per questo motivo, all'insieme ci si riferisce come alla *trama PCM a 2 Mbit*. La durata della trama deve essere invece la stessa del periodo di campionamento, ossia $1/8000 = 125 \mu\text{sec}$.

Il primo time-slot della trama contiene una configurazione di bit sempre uguale, chiamata FLAG (*bandiera*), che ha lo scopo di indicare ai circuiti di sincronismo l'inizio della trama stessa. I dati di segnalazione contenuti nel 16° intervallo devono essere diluiti su più trame, per poter rappresentare tutti i 30 tributari³². Si è stabilito che occorra prelevare il 16° intervallo di 16 trame successive, per ricostruire una struttura detta *supertrama* (di $16 \cdot 8 = 128$ bit) che rappresenta le informazioni di tutti i tributari (disponendo così di 4 bit/canale/supertrama), e che si ripete ogni $16 \cdot 125 = 2000 \mu\text{sec}$.

In effetti nel 16° time-slot della trama sono presenti a turno, oltre ai bit di segnalazione relativi allo stato dei tributari, anche bit necessari alla sincronizzazione della supertrama (ossia un *flag*), mentre le informazioni di segnalazione sono ripetute più volte nel-

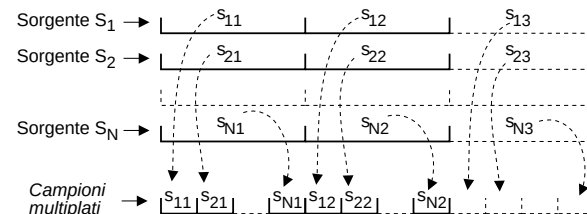
²⁷La tecnica di moltiplicare un *blocco di bit* (in questo caso 8) alla volta prende il nome di *word interleaving*, distinto dal *bit interleaving*, in cui l'alternanza è a livello di bit.

²⁸Il segnale PCM ispira il suo nome dal PAM (vedi § 4.4.1) in quanto ora, anziché trasmettere le *ampiezze* degli impulsi, si inviano i *codici* binari dei livelli di quantizzazione.

²⁹FRAME significa più propriamente *telaio*, e in questo caso ha il senso di individuare una struttura, da "riempire" con il messaggio informativo.

³⁰

In figura è mostrato un esempio, in cui i campioni s_{ij} di N sorgenti S_i si alternano a formare una trama. Dato che nell'intervallo temporale tra due campioni, nella trama devono essere collocati tutti gli M bit/campione delle N sorgenti, la frequenza binaria (in bit/secondo) complessiva sarà pari a $f_b = f_c$ (campioni/secondo/sorgente) $\cdot N$ (sorgenti) $\cdot M$ (bit/campione).



³¹Vedi anche la sezione 4.3.4.

³²Gli 8 bit del 16° intervallo sono infatti insufficienti a codificare lo stato dei 30 tributari di una trama.

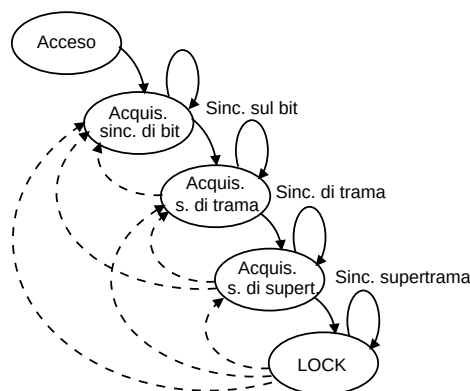
la stessa supertrama, per proteggersi da eventuali errori di ricezione, che danneggiando l'informazione sullo stato dei canali potrebbero causare la “caduta della linea”.

CAS e CCS La particolarità di inserire le informazioni di segnalazione assieme a quelle informative prende il nome di *Segnalazione Associata al Canale* (CAS, CHANNEL ASSOCIATED SIGNALING), ed ha origine dalla sostituzione dei precedenti collegamenti analogici con quelli numerici, conservandone alcune particolarità strutturali.

Attualmente, la rete telefonica evolve invece verso una *Segnalazione a Canale Comune* (CCS, COMMON CHANNEL SIGNALING), che non prevede time-slot riservati per il controllo, ma distribuisce queste informazioni tramite una rete parallela a quella di comunicazione tra utenti. In questo caso, esiste anche un organo di controllo centrale della rete che riceve, elabora ed invia tutte le informazioni di segnalazione, rendendo così possibile la disponibilità di servizi come il trasferimento di chiamata, la conversazione a tre, l'avviso di chiamata....

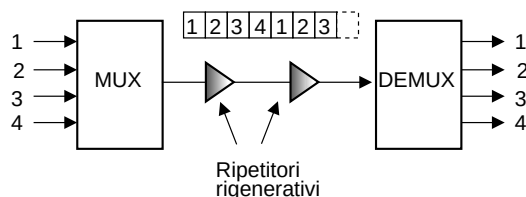
Sincronizzazione di centrale

In figura sono mostrati i diversi stati attraverso cui deve evolvere il dispositivo di sincronizzazione prima di entrare nello stato di LOCK (*aggancia*) ed iniziare a poter leggere e smistare i contenuti dei diversi time-slot. Occorre, nell'ordine, acquisire prima il sincronismo sul bit, quindi quello sulla trama, ed infine quello relativo alla supertrama. In ogni stato poi, esiste la possibilità (fortunatamente remota) di perdere il sincronismo ed “indietreggiare” (linee tratteggiate) nel diagramma di stato, perdendo le comunicazioni in corso.



4.3.2 Multiplazione Asincrona e Rete Plesiocrona³³

Nella trama PCM, tutti i 30 canali sono campionati congiuntamente, e più flussi a 2 Mbit possono a loro volta essere “messi assieme” in modalità *bit interleaved* (prendendo un bit alla volta da ogni tributario) da appositi dispositivi *multiplatori* (o MULTIPLEXER, o MUX). Il collegamento può prevedere più dispositivi detti *ripetitori rigenerativi*, che oltre ad amplificare il segnale, lo “puliscono” dal rumore accumulato, decodificando i dati in ingresso per poi generare ex-novo il segnale numerico.



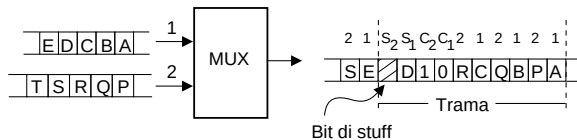
N. Canali	Vel. (Mbps)	Teorica
30	2.048	
120	8.448	8.192
480	34.368	32.768
1920	139.264	131.072
7680	565.148	524.288

³³L'argomento di questo paragrafo non va confuso con la *trasmissione* asincrona (quella start-stop mostrata al § 4.4.2), che descrive una modalità di *inviare* informazioni numeriche; qui invece si tratta di *multiplare*, ossia come *mettere assieme* più comunicazioni.

Il problema con questo modo di procedere è che i singoli tributari possono ragionevolmente avere origine da centrali differenti, ognuno con un proprio orologio indipendente, e quindi le velocità possono essere lievemente differenti l'una dall'altra³⁴, pur essendo molto simili. In questo caso si dice che la rete opera in modo *plesiocrono*, ossia *quasi* isocrono (ma non del tutto).

In tabella riportiamo la gerarchia CCITT³⁵, secondo la quale ad esempio 4 flussi da 2 Mbps sono multiplati in uno da 8 Mb/sec: notiamo che sebbene siano teoricamente sufficienti 8192 Mb/sec, in realtà il Multiplexer ne produce di più (8448). Questo avviene proprio per permettere la trasmissione di segnali non necessariamente sincroni, mediante la tecnica del *Bit Stuffing*³⁶.

Bit Stuffing Consideriamo 2 tributari i cui bit vengono inseriti alternativamente in una trama da 4 bit/canale; il secondo risulta lievemente più lento.



I primi 3 + 3 bit vengono trasmessi comunque, mentre il 4° può essere trasmesso o meno, a seconda se i tributari lo abbiano pronto. Per ottenere questo risultato, i bit C_1 e C_2 (di controllo) valgono 0

oppure 1 a seconda se l'intervallo seguente (S_1 e S_2) contenga un dato valido oppure sia solo un bit di stuff, cioè vuoto, in quanto il tributario corrispondente è più lento rispetto alla velocità nominale. Ecco perché le velocità delle gerarchie superiori sono abbondanti: per ospitare i bit di controllo, necessari a gestire tributari non sincronizzati. Il metodo illustrato permette in ricezione di effettuare il *destuffing*, e riottenere i flussi originari. Nella realtà le informazioni di controllo sono molto ridondate, perché se scambiassimo un bit di stuff per uno buono (o viceversa), distruggeremmo anche la struttura di trama del tributario che ha subito l'errore.

4.3.3 Sincronizzazione di Rete

Se tutti i nodi della rete operassero alla stessa velocità, non sussisterebbero problemi nella multiplazione di più tributari. Nel caso in cui la sincronizzazione tra nodi sia completamente affidata ad un orologio di centrale di elevata precisione, si verifica il caso di funzionamento *plesiocrono*, che è quello prescritto per le centrali che interconnettono le reti di due diverse nazioni, o di due diversi operatori di telecomunicazioni. Ma questa non è l'unica soluzione.

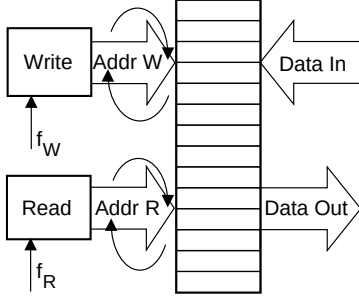
Una alternativa è la sincronizzazione *mutua* tra centrali, in cui ognuna di queste emette dati in uscita ad una frequenza pari alla media delle frequenze dei dati in ingresso. A parte fenomeni transitori durante i quali la rete è soggetta ad oscillazioni di velocità, relativi all'inserimento od alla disattivazione di centrali "topologicamente importanti", il metodo funziona ragionevolmente bene. Una seconda soluzione è una sincronizzazione di rete di tipo *gerarchico* in cui le centrali ricevono informazioni di sincronismo da soggetti "più importanti", come per configurazioni *Master-Slave* in cui il Master è una centrale ad elevata precisione, od un riferimento in comune come ad esempio un satellite in orbita terrestre.

³⁴Un oscillatore con precisione di una parte su milione, produce un ciclo in più o in meno ogni 10^6 ; ad una velocità di 2 Mbps, ciò equivale a un paio di bit in più od in meno ogni secondo.

³⁵Comité Consultatif International pour la Telephonie et Telegraphie. Questo organismo non esiste più, ed ora l'ente di standardizzazione ha nome ITU-T.

³⁶Da: TO STUFF = riempire.

Elastic Store Si tratta di un accorgimento³⁷ idoneo ad “assorbire” fluttuazioni della velocità di trasmissione, come ad esempio nel caso della sincronizzazione mutua.



E' realizzato mediante un banco di memoria (di dimensione pari ad una trama), riempito (ciclicamente) con le parole (word) del flusso binario in ingresso, alla velocità f_W di quest'ultimo, utilizzando il puntatore WRITE che si incrementa³⁸ appunto a velocità f_W , e che torna all'inizio della memoria una volta raggiunto il fondo. Un secondo puntatore READ viene utilizzato per leggere la memoria, alla velocità f_R richiesta, e prelevare i dati da inviare in uscita: se f_R e f_W sono differenti, READ e WRITE prima o poi si sovrappongono, causando la perdita o la ripetizione di una intera trama, e nulla più³⁹.

4.3.4 Messaggi di segnalazione

Forniamo ora un istruttivo esempio di ciò che accade *prima, durante e dopo* essere riusciti a parlare per telefono con qualcuno. Quando si solleva la cornetta, ovvero si *sgancia*⁴⁰, viene chiuso un interruttore interno al telefono, facendo scorrere una corrente nei 2 fili che giungono fino a casa, e la centrale *si accorge* della nostra intenzione di telefonare. Di conseguenza, vengono *impegnate* (SEIZED) delle risorse della centrale, ed è prenotato un time-slot uscente, alla disponibilità del quale corrisponde un segnale *tono di centrale*, che ci viene inviato per invitarci a comporre il numero. A seguito della ricezione del numero, la centrale *di origine* coinvolge il resto della rete, impegnando risorse della stessa, ed individuando quali nodi attraversare per giungere a destinazione (fase di *istradamento*, in inglese ROUTING). Una volta contattata la centrale di destinazione, questa provvede a far squillare il telefono chiamato, ed inviare indietro un segnale di *RingBack* che produce presso il chiamante un *tono di libero* (oppure un segnale di *Busy*, nel caso che il chiamato sia già impegnato in altra conversazione).

Chiamante	C. di Origine	Rete	C. di Destin.	Chiamato
Sgancia	→ Impegno	→ Indicazione		
	← Tono di centr.	← Risposta OK		
Compone il numero	→ Richiesta di connessione	→ Istradamento e impegno risorse	Indicazione di chiamata →	Suona
Suona libero	← Conferma	← Ringback	← Risp. Libero	
Parla	← Connette	← Connette	← Connette ←	Risponde Parla
		← Mantenimento →		
Attacca	→ Svincolo →	Rilascio risorse →	Occupato →	Attacca

³⁷Letteralmente: magazzino elastico.

³⁸Possiamo pensare di aver realizzato il puntatore mediante un contatore binario che si incrementa con f_W (f_R). WRITE fornisce l'indirizzo in cui scrivere i dati in ingresso, mentre READ quello per i dati in uscita.

³⁹Infatti il sincronismo di trama viene preservato; inoltre l'evento di sovrapposizione dei puntatori può essere rilevato, e segnalato ai dispositivi di demultiplazione, in modo che tengano conto dell'errore che si e' verificato.

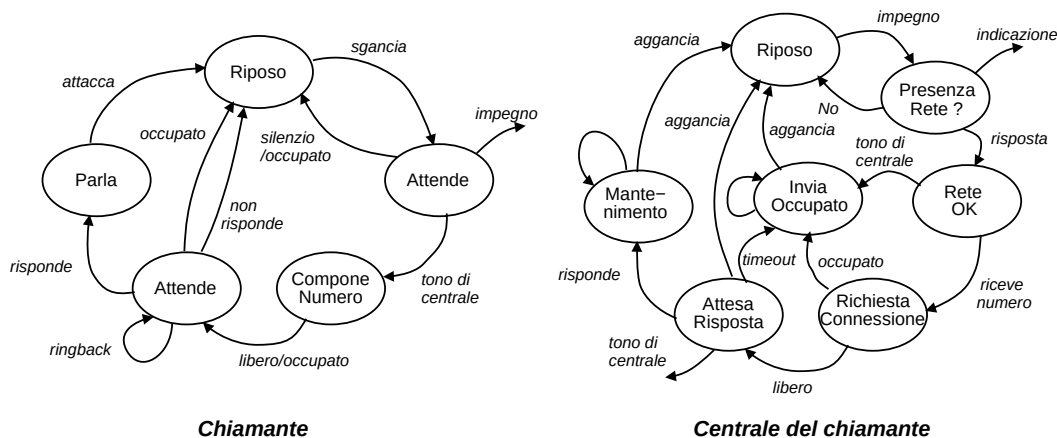
⁴⁰in inglese si dice andare OFF-HOOK, con riferimento storico all'uncino su cui riporre la cornetta, presente fin dai primi modelli di telefono “a manovella”.

Quando il chiamato risponde, la sua centrale se ne accorge e lo connette al collegamento instaurato, che entra in uno stato di mantenimento, e che permane finché una delle due parti non attacca. A questo punto, la centrale locale invia alla rete un segnale di *svincolo*, che produce il rilascio delle risorse impegnate ed un tono di occupato all'altro terminale.

Grazie a questo “raccontino” abbiamo evidenziato:

- La rilevanza delle informazioni di *segnalazione* nelle comunicazioni tramite una rete;
- Il carattere *segmentale* della segnalazione, che avviene a livello delle singole entità che si affacciano (ad esempio, le due centrali periferiche svolgono in tandem un colloquio di tipo $\left\{ \begin{array}{l} \text{Richiesta} \rightarrow \text{Indicazione} \\ \text{Conferma} \leftarrow \text{Risposta} \end{array} \right\}$ con i rispettivi “interlocutori”: l'apparecchio di un utente da un lato e la rete dall'altro;
- la necessità di disporre di strategie per determinare il percorso che la chiamata deve seguire all'interno della rete (*instradamento*), intimamente legato a come la rete è strutturata (*topologia*), ed a come sono assegnati i numeri telefonici agli utenti (*indirizzamento*);
- la necessità di disporre, presso i nodi della rete che devono essere attraversati dalla comunicazione, di opportune funzionalità di *commutazione*, allo scopo di realizzare la serie di collegamenti previsti in fase di istradamento;

Lo scambio di messaggi di segnalazione prende il nome di *protocollo*, in virtù dell'ordine schematico che ne definisce le varie fasi. La sequenza dei messaggi del protocollo può essere rappresentata sia nella forma mostrata nella tabella, sia in forma di un diverso diagramma di transizione⁴¹ per ogni “lato” coinvolto nella comunicazione, come è esemplificato nella figura che segue. E' da notare che i due automi che descrivono le varie fasi della segnalazione evolvono in modo congiunto, essendo i segnali di *uscita* dell'uno *ingressi* dell'altro.

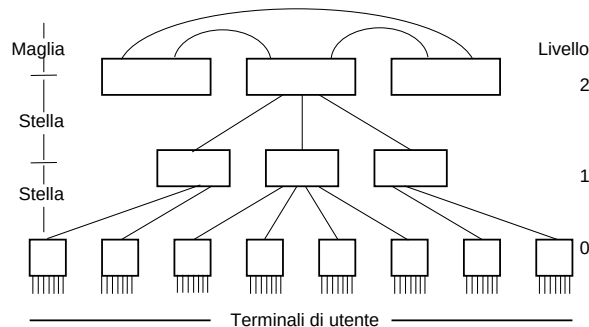


⁴¹Si usa il formalismo di scrivere sugli archi i segnali di ingresso/uscita.

4.3.5 Topologia di rete

Nella figura seguente è riportata una possibile topologia di rete, a scopo puramente esemplificativo, in quanto questo è un elemento su cui i progettisti godono di massima libertà. Il nostro esempio si riferisce ad una rete a 2 livelli, in cui i terminali di utente che fanno capo ad una stessa centrale locale di livello zero, accedono alla rete per mezzo di un collegamento che moltiplica le comunicazioni provenienti dalla stessa centrale.

Come vedremo nel prossimo capitolo, il numero di collegamenti contemporaneamente possibili è inferiore al numero di terminali di utente, ed è dimensionato applicando i metodi della teoria del traffico al caso in questione. Le centrali locali sono connesse *a stella* a quelle di livello 1, che sono quindi attraversate dagli instradamenti relativi alle comunicazioni tra utenti geograficamente vicini, ma connessi a centrali locali diverse. Se, al contrario, due utenti sono connessi alla stessa centrale di livello zero, l'instradamento non esce dalla centrale locale.



Le centrali di livello 1 sono a loro volta connesse a stella a quelle di livello 2, che gestiscono il traffico a livello nazionale; alcune di queste poi, consentono di instradare anche i collegamenti verso altre reti (ad es. di altre nazioni od operatori). Le centrali di 2° livello sono connesse tra loro a *maglia* completa (MESH in inglese) in modo da consentire instradamenti alternativi anche nel caso in cui un collegamento tra centrali vada fuori servizio.

4.3.6 Instradamento

Anche per questo argomento, forniamo solo una definizione molto sommaria di tre possibili strategie:

END to END o *Right-through* (da estremo ad estremo o dritto per dritto): la scelta del percorso è effettuata dalla centrale di origine, ad esempio in base al prefisso od all'inizio del numero, utilizzando delle *tabelle di routing* statiche. E' la modalità dell'inizio della telefonia, in cui i commutatori erano elettromeccanici, ed i collegamenti verso altre centrali erano *cablati*. Ha l'enorme svantaggio che i cambiamenti alla topologia della rete si devono riflettere in cambiamenti di tutte le tabelle (o dei morsetti !);

Link-by-link o *Own-Exchange* (tratta per tratta o scelta dalla centrale): ogni centrale decide in autonomia dove instradare (in uscita) le connessioni entranti, in base a sue tabelle dinamiche, ovvero informazioni che giungono dalla rete stessa. Si adatta alle modifiche della topologia ma non è affidabile al 100 %, potendo ad esempio produrre dei *loop* (leggi circoli viziosi);

Tramite CCS (*Common Channel Signaling*, segnalazione a canale comune): le decisioni sull'instradamento sono demandate ad una rete di segnalazione parallela ed indipendente da quella del traffico smaltito, e che collega tutte le centrali ad un

unico organo di controllo (il *canale comune*), il quale determina l'instradamento in base alla sua conoscenza dello stato del traffico nella rete, e comunica contemporaneamente a tutte le centrali coinvolte nell'instradamento, come configurare i propri organi di commutazione per realizzare il collegamento richiesto.

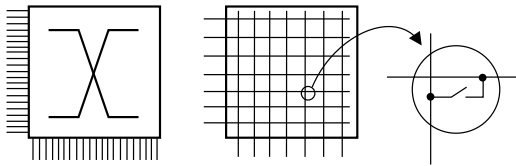
4.3.7 Commutazione

Illustreremo ora alcune architetture di dispositivi che consentono la cosiddetta *commutazione di circuito*, che rappresenta la creazione di un collegamento *stabile* tra due porte del commutatore, con un *impegno permanente* di risorse fisiche per tutta la durata del collegamento. Un'altra modalità di commutazione, quella di *pacchetto*, sarà illustrata al Capitolo 6.

4.3.7.1 Reti a divisione di spazio

Sono chiamati così gli organi di commutazione che realizzano un collegamento fisico (elettrico) tra uno degli N ingressi ed una delle M uscite. Nel caso in cui $N > M$, la rete è un *concentratore*⁴², mentre se $N < M$ la rete è un *espansore*; se $N = M$ la rete è quadrata e *non bloccante*.

Il commutatore è rappresentato da un blocco con una "X" (in inglese *cross*, od incrocio), e può essere pensata come una matrice binaria in cui ogni elemento (1 o 0) rappresenta lo stato (chiuso od aperto) di un interruttore (realizzato ad esempio mediante un transistor) che collega una linea di ingresso ad una di uscita.



Realizzare in questo modo una rete non bloccante prevede l'uso di un numero di interruttori pari ad $N \cdot M$, dei quali solo $\min(N, M)$ sono utilizzati, anche nelle condizioni di massimo carico. Inoltre, nessun interruttore può essere "eliminato" senza precludere irrimediabilmente

la possibilità di collegare qualunque ingresso a qualunque uscita. Allo scopo di utilizzare un numero ridotto di interruttori, sia per costruire reti non bloccanti oppure bloccanti con bassa probabilità di blocco, si ricorre alle...

4.3.7.2 Reti multistadio

di cui in figura è riportato un esempio a 3 stadi, in cui gli N ingressi sono ripartiti su r_1 reti più piccole con n ingressi, e le M uscite su r_3 reti con m uscite. Nel mezzo ci sono r_2 reti con r_1 ingressi ed r_3 uscite. Si può dimostrare che la rete complessiva è

⁴²come ad esempio un centralino (PBX, PRIVATE BRANCH EXCHANGE) con 8 derivati (interni) e 2 linee esterne: se due interni parlano con l'esterno, un terzo interno che vuole anche lui uscire trova occupato. Si dice allora che si è verificata una condizione di *blocco*.

⁴³E' una condizione *sufficiente* a scongiurare il blocco anche nella condizione *peggiore*. Tale circostanza si verifica quando:

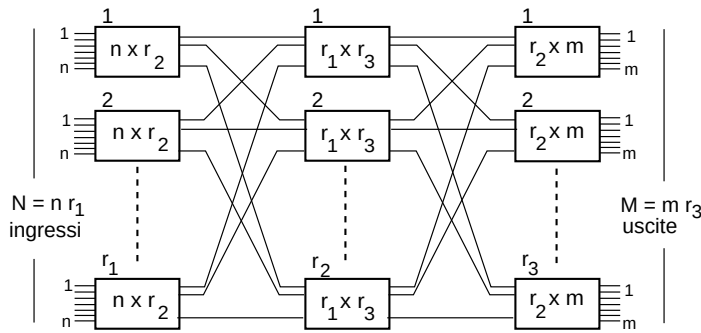
- ▷ una matrice del primo stadio (i) ha $n - 1$ terminazioni occupate
- ▷ una matrice del terzo stadio (j) ha $m - 1$ terminazioni occupate e
- ▷ tali terminazioni non sono connesse tra loro, anzi le connessioni associate impegnano ognuna una diversa matrice intermedia e

- ▷ si richiede la connessione tra le ultime due terminazioni libere di i e j

⇒ in totale si impegnano allora $m - 1 + n - 1 + 1 = m + n - 1$ matrici intermedie.

non bloccante se il numero di matrici dello stadio intermedio è almeno $r_2 \geq n + m - 1$ (condizione di CLOS⁴³). Una connessione da sinistra a destra ha ora la possibilità di scegliere attraverso quale matrice intermedia passare.

Nel caso di reti quadrate ($N = M$), ponendo $n = m = \sqrt{\frac{N}{2}}$, si ottiene un numero complessivo di incroci pari a $4 \left(\sqrt{2} N^{\frac{3}{2}} - N \right)$, che risulta inferiore ad N^2 (e dunque vantaggioso rispetto ad un commutatore monostadio) a partire da $N \geq 24$.

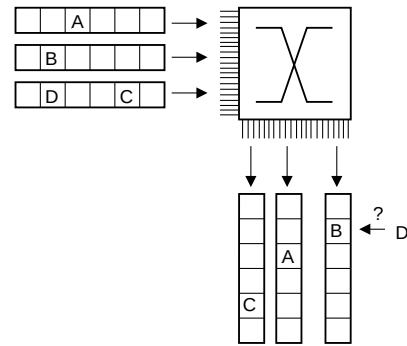


missioni numeriche a divisione di tempo.

Ovviamente, la problematica relativa alle matrici di commutazione è molto articolata, coinvolgendo topologie più complesse, filosofie di instradamento, e tecniche per la stima delle probabilità di blocco. Tralasciamo ulteriori approfondimenti, per illustrare invece come realizzare dispositivi di commutazione per tra-

4.3.7.3 Commutazione numerica a divisione di tempo

Consideriamo il caso in cui si debbano commutare le comunicazioni associate ai singoli *time-slot* presenti in diversi flussi⁴⁴ numerici organizzati in trame. Avendo a disposizione solamente una matrice di commutazione spaziale, quest'ultima può essere riprogrammata alla stessa frequenza dei *time-slot*, consentendo alle comunicazioni entranti di dirigersi verso i flussi uscenti in direzione delle rispettive destinazioni finali. La matrice spaziale, però, non può alterare l'ordine temporale dei dati in ingresso; pertanto, non può (ad esempio) inviare le conversazioni *B* e *D* sulla stessa linea uscente, in quanto si verifica un conflitto temporale. E' quindi evidente la necessità di introdurre uno stadio di *commutazione temporale*.

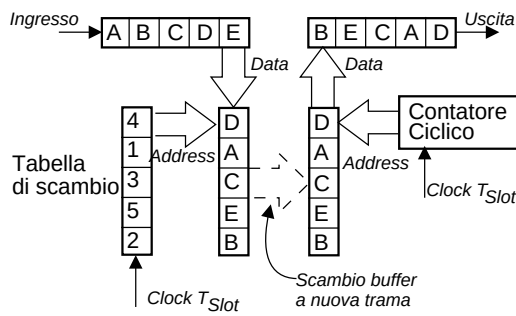


Time Slot Interchanger

Questo dispositivo è indicato come TSI (*time slot interchanger*) ed ha la funzione di produrre in uscita una sequenza di dati identica a quella in ingresso, tranne per averne cambiato l'ordine temporale.



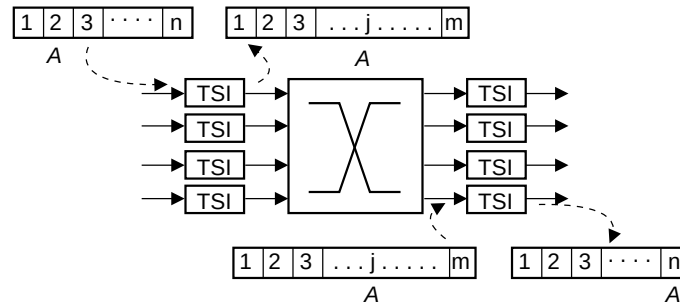
⁴⁴Le comunicazioni presenti in uno stesso flusso, ovvero appartenenti alla stessa trama, condividono la stessa origine/destinazione.



In figura è mostrato un possibile schema di funzionamento: una trama entrante viene scritta, agli indirizzi ottenuti leggendo sequenzialmente la tabella di scambio, in un buffer di memoria (Es.: entra *E* e lo scrivo al 4° posto, poi entra *D* e va al 1° posto, etc.). Prima dell'inizio di una nuova trama, il buffer è scambiato con un altro uguale, e quello appena scritto viene letto con ordine sequenziale (partendo dall'alto), per creare la nuova trama in uscita.

Ovviamente, è possibile anche la realizzazione opposta, con scrittura sequenziale e lettura "casuale".

Commutazione bidimensionale Così come un commutatore spaziale non è sufficiente, anche un TSI "da solo" è di scarsa utilità, non potendo instradare le comunicazioni su vie diverse. Combinando assieme le due funzioni, si giunge a realizzare commutatori sia di tempo che di spazio, come la struttura a 3 stadi in figura, chiamata "TST" perché alterna uno stadio temporale, uno spaziale ed uno temporale.



Notiamo subito che, in questo schema, il numero di intervalli temporali in uscita dai TSI di ingresso è $m > n$ ⁴⁵: ciò determina, per lo stadio spaziale, una frequenza di commutazione più elevata della frequenza dei time-slot in ingresso. Una generica conversazione "A" che occupa il 2° slot del primo flusso può raggiungere (ad esempio) l'ultimo slot dell'ultimo flusso, occupando uno qualsiasi (*j*) degli *m* slot utilizzati dal commutatore spaziale. Aumentando il valore di *m*, si riduce la probabilità di blocco; in particolare, questa è nulla se $m = 2n - 1$ ⁴⁶.

Analizziamo i vantaggi conseguiti dalla commutazione numerica con un semplice esempio. Poniamo di voler commutare con lo schema illustrato 4 flussi PCM (con $n = 30$): i $4 * 30 = 120$ canali presenti sono commutati utilizzando solo $4 * 4 = 16$ interruttori, contro i $120 * 120 = 14.400$ interruttori necessari ad una matrice spaziale monostadio che svolga la commutazione dei 120 canali analogici !

⁴⁵Ovviamente, $m - n$ intervalli sono lasciati vuoti, in ordine *sperso* tra gli *m*.

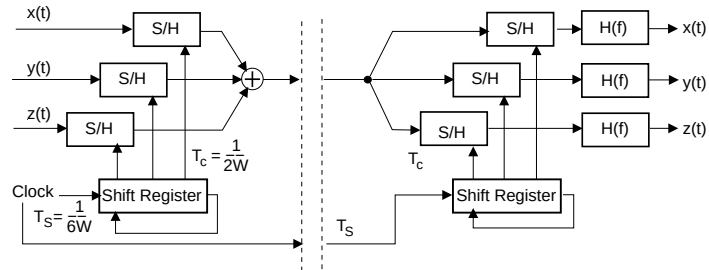
⁴⁶Si confronti questo risultato con la condizione di Clos, fornita al § 4.3.7.2.

4.4 Appendici

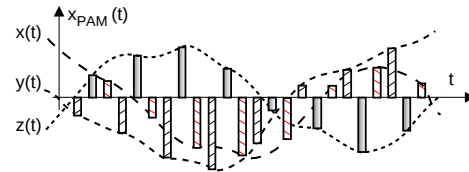
4.4.1 Multiplazione a divisione di tempo mediante modulazione di ampiezza degli impulsi

Al tempo in cui la realizzazione del componente di quantizzazione presentava discrete difficoltà circuitali, si pensò⁴⁷ di sfruttare il teorema del campionamento per inviare su di un unico collegamento più comunicazioni multipliate a divisione di tempo (TDM = *Time Division Multiplex*).

E' sufficiente infatti sommare alla funzione $x^\circ(t)$ introdotta al § 4.1.3 altri segnali simili, ad esempio $y^\circ(t)$, $z^\circ(t)$ come mostrato in Figura, ognuno campionato a frequenza $2W$, ma sfasato rispetto agli altri, realizzando la cosiddetta onda PAM (*Pulse Amplitude Modulation*, modulazione ad ampiezza di impulsi), con impulsi separati da un intervallo $T_s = \frac{1}{2NW} = \frac{T_c}{N}$, con N pari al numero di segnali multiplati. Il pedice s indica che si tratta di un *periodo di simbolo*.



Il segnale $x_{PAM}(t)$ composto dalle 3 sorgenti dell'esempio della figura in alto è mostrato a lato, e può essere nuovamente campionato estraendo $x(nT_c)$, $y(nT_c)$, $z(nT_c)$, mentre i segnali $x(t)$, $y(t)$ e $z(t)$ sono riprodotti facendone passare gli impulsi campionati a frequenza $2W$ in un filtro di ricostruzione con banda W .



4.4.2 Trasmissione Asincrona

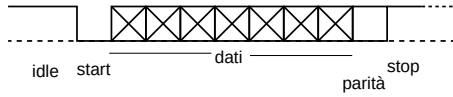
Come anticipato in 4.2.3.1, il controllo di parità “a bit unico” è spesso usato nel caso di collegamenti seriali a bassa velocità, che assume le caratteristiche particolari che qui illustriamo.

Nel caso di un terminale collegato ad un computer centrale, ad esempio, la trasmissione avviene in modo “sporadico”, ossia quando l'operatore “digita” sui tasti del terminale; tale modalità di trasmissione è indicata come *asincrona*, per distinguerla da quella (*isocrona*) che ha luogo quando sorgente e destinatario si scambiano informazioni in modo continuo e senza pause. Svolgiamo quindi un piccolo *excursus* sul tema, già che ci siamo.

Nella trasmissione asincrona, la linea di comunicazione permane abitualmente in uno stato di *libero* (IDLE), contraddistinta (ad esempio) da uno stato di tensione positi-

⁴⁷La “pensata” non ebbe molte applicazioni, se non in ambito della commutazione interna ad esempio ad un centralino, a causa della sensibilità del metodo agli errori di temporizzazione, ed alle caratteristiche del mezzo trasmissivo su cui inviare il segnale PAM.

va. Quando il $\mathbf{T}\mathbf{x}$ ⁴⁸ vuole trasmettere, porta la linea a 0 per la durata di 1 simbolo, che prende il nome di bit di *start*: la transizione *in discesa* viene rilevata da $\mathbf{R}\mathbf{x}$, che si pre-dispone a contare un numero fisso di simboli (7 in figura) basandosi su di un suo orologio indipendente. Segue poi un bit di *parità* (se prevista) ed uno di *stop*, presente quest'ultimo per assicurare una durata minima dello stato di IDLE, prima della trasmissione del carattere successivo.



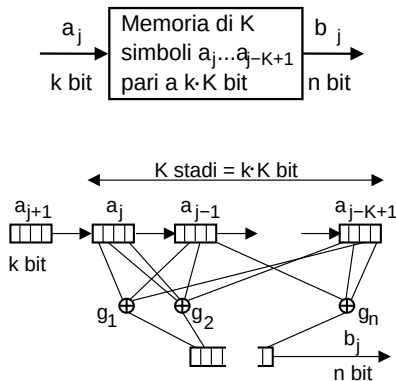
Il vantaggio di una simile modalità operativa, è che $\mathbf{R}\mathbf{x}$ non ha bisogno (come invece accade nella trasmissione isocrona) di ricostruire con esattezza la temporizzazione del segnale entrante: si

basa piuttosto su di un proprio orologio locale, di precisione anche scadente, che viene "risvegliato" in corrispondenza del bit di start.

Tale semplicità operativa produce una inefficienza, in quanto oltre ai dati ed al bit di parità, si introduce anche lo *start* e lo *stop*, utili solo ai fini della sincronizzazione ma privi di contenuto informativo.

4.4.3 Codici convoluzionali

Si tratta di una classe di codificatori di canale che, a differenza dei codici a blocchi, produce una sequenza binaria i cui valori dipendono da gruppi di bit di ingresso *temporalmente sovrapposti*, in analogia all'integrale di convoluzione che fornisce valori che dipendono da *intervalli* dell'ingresso, pesato dai valori della risposta impulsiva.



Un generico codice convoluzionale è indicato con la notazione $CC(n, k, K)$, che lo descrive capace di generare gruppi di n bit di uscita (sequenza $\{b\}$) in base alla conoscenza di K simboli di ingresso (sequenza $\{a\}$), ognuno composto da k bit. La memoria dei K ingressi è usualmente rappresentata mediante un registro a scorrimento che ospita gli ultimi $K \cdot k$ bit di ingresso, dove per ogni nuovo simbolo a_j che entra da sinistra, i precedenti scorrono a destra, ed il più "vecchio" viene dimenticato. Ognuno degli n bit di uscita $b_j(i)$, $i = 1, 2, \dots, n$ è ottenuto a partire dai $k \cdot K$ bit di memoria, eseguendo una somma modulo 2⁽⁴⁹⁾ tra alcuni di essi⁵⁰. Il rappor-

to $R = \frac{k}{n}$ è chiamato *coding rate* e rappresenta *quanti bit di informazione* sono presenti per ogni bit di uscita dal codificatore.

4.4.3.1 Diagramma di transizione

Una volta definito il meccanismo di calcolo di b_j a partire da a_j , il codificatore può essere descritto mediante un *diagramma di transizione*, costruito da $2^{(K-1)k}$ stati S , che rappresentano gli ultimi $K - 1$ simboli di ingresso, ed in cui ad ogni stato competono 2^k

⁴⁸A volte, trasmettitore e ricevitore sono indicati con l'abbreviazione $\mathbf{T}\mathbf{x}$ ed $\mathbf{R}\mathbf{x}$, rispettivamente.

⁴⁹La somma modulo due tra una coppia di cifre binarie è indicata con $\oplus$, e coincide con l'operatore logico di OR ESCLUSIVO. Pertanto risulta $0 \oplus 0 = 1 \oplus 1 = 0$ e $1 \oplus 0 = 0 \oplus 1 = 1$.

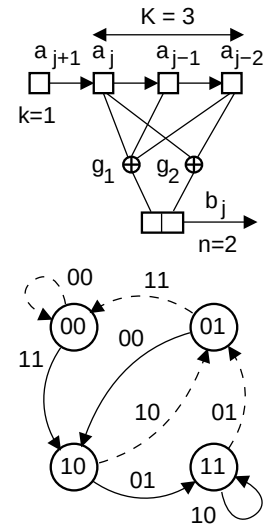
⁵⁰Gli n modi di scegliere quali dei $k \cdot K$ bit sommare, per ottenere ognuno degli n bit di uscita, sono determinati mediante n vettori generatori g_i , $i = 1, 2, \dots, n$, di lunghezza $k \cdot K$, contenenti una serie di cifre binarie zero od un uno, a seconda che l'*i*-esimo sommatore modulo due sia connesso (o meno) al corrispondente bit di memoria.

transizioni (una per ogni possibile nuovo a_j di ingresso). Ad ogni transizione, è infine associato in modo univoco il gruppo di n bit $b_j(a_j, S_j)$ da produrre in uscita⁵¹.

Esempio: CC(2,1,3)

Per fissare le idee, definiamo un $CC(n, k, K) = CC(2, 1, 3)$ con $g_1 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ e $g_2 = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$, con coding rate $\frac{1}{2}$, mostrato in figura, per il quale costruiamo il diagramma di transizione mostrato sotto, dove le transizioni tra stati sono tratteggiate o piene in corrispondenza di valori a_j pari a zero o ad uno, e sono etichettate con la coppia di bit in uscita b_j , calcolabile mediante la seguente tabella della verità:

a_j	S_j	$b_j(1)$	$b_j(2)$
0	00	0	0
0	01	1	1
0	10	1	0
0	11	0	1
1	00	1	1
1	01	0	0
1	10	0	1
1	11	1	0



Come si può notare, ogni stato ha *solo due* transizioni, e quindi solo *due valori* di uscita (la metà dei 4 possibili con due bit); inoltre, questi valori differiscono in *entrambi i bit*⁵².

Qualora si ponga in ingresso la sequenza $\{a\} = \{\dots 010100\}$ ⁽⁵³⁾, è possibile osservare che la sequenza di stati risulta $\{S\} = \{\dots, 01, 10, 01, 10, 00\}$, mentre quella di uscita è $\{b\} = \{\dots, 10, 00, 10, 11\}$, semplicemente seguendo il diagramma di transizione, a partire dallo stato iniziale 00. D'altra parte, è possibile anche il procedimento opposto: conoscendo $\{b\}$, si può risalire ad $\{a\}$, percorrendo di nuovo le transizioni etichettate con i simboli b_j . In definitiva, osserviamo come ad ogni coppia $(\{a\}, \{b\})$ sia biunivocamente associata una sequenza di stati $\{S\}$.

4.4.3.2 Diagramma a traliccio

Per meglio visualizzare le possibili sequenze di stati, costruiamo il *diagramma a traliccio* (TRELLIS) del codificatore, mostrato a sinistra di fig. 4.5 riportando sulle colonne i possibili stati attraversati ai diversi istanti j , collegando i *nodi* del traliccio con transizioni piene o tratteggiate nei casi in cui siano relative ad ingressi pari ad 1 o 0, e riportando sulle transizioni stesse i valori di uscita.

Con riferimento alla sequenza *codificata* $\{b\}$ riportata in basso, la successione di stati $\{S\}$ è rappresentata dalle linee più spesse.

⁵¹Lo stesso valore di b_i potrebbe essere prodotto da più di una delle $2^{k \cdot K}$ diverse memorie del codificatore.

⁵²La dipendenza di b_j da (a_j, S_j) è legata alla scelta dei generatori g_i . Nel caso in cui un valore $b_j(i)$ sia sempre uguale ad uno dei k bit di a_j , il codice è detto *sistematico*. La scelta dei g_i può essere effettuata via computer, per individuare il gruppo che determina le migliori prestazioni.

⁵³Per comodità di rappresentazione, il bit più a destra nella sequenza $\{a\}$ è il primo in ordine di tempo ad entrare nel codificatore.

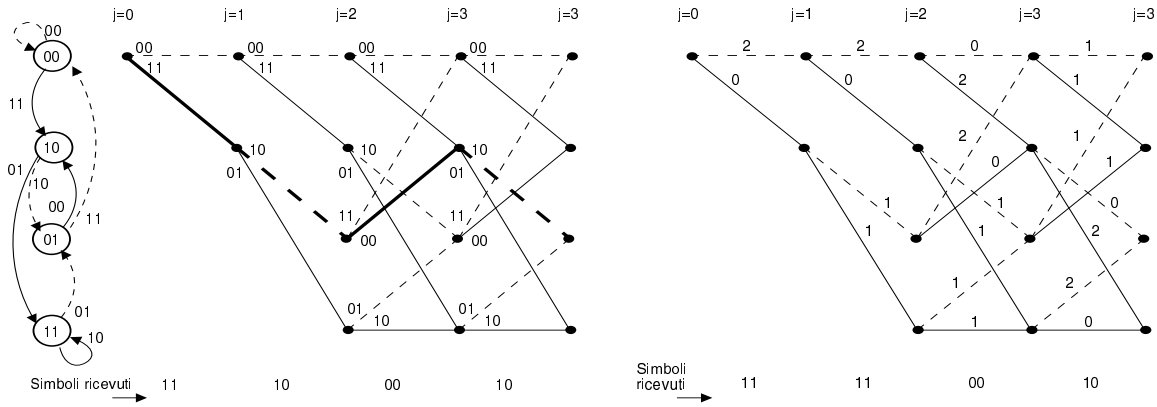
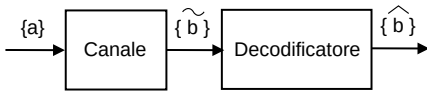


Figura 4.5: Diagramma a traliccio e costi d_H per la sequenza ricevuta

4.4.3.3 Decodifica di Viterbi



Consideriamo ora il caso in cui la sequenza codificata $\{b\}$ venga trasmessa su di un canale, e che sia ricevuta *con errori*. In generale, non sarà più possibile rintracciare una sequenza di stati tale da produrre *esattamente* la sequenza ricevuta $\{\tilde{b}\}$, ed il problema diviene quello di individuare la sequenza di stati $\{S\}$ tale da produrre una $\{\hat{b}\}$ la più vicina a $\{\tilde{b}\}$. Allo scopo di misurare questa differenza, utilizziamo la *distanza di Hamming* $d_H(\tilde{b}, \hat{b})$ pari al numero di bit diversi tra $\{\tilde{b}\}$ e le possibili $\{\hat{b}\}$ ⁵⁴, e etichettando gli archi del traliccio con $d_H(\tilde{b}_j, \hat{b}_j)$ tra il simbolo da emettere $\hat{b}_j$ e quello osservato in ricezione $\tilde{b}_j$ (vedi lato destro di fig. 4.5). In tal modo, per ogni particolare sequenza di stati $\{S\}$ è possibile determinare un *costo* pari alla somma delle $d_H(\tilde{b}_j, \hat{b}_j)$ relative alle transizioni attraversate dal traliccio⁵⁵. Pertanto, la sequenza $\{\hat{b}\}$ più vicina a $\{\tilde{b}\}$, ossia $d_H(\tilde{b}, \hat{b}) = \min_{\{S\}} \{\sum_j d_H(\tilde{b}_j, \hat{b}_j)\}$, può essere individuata come quella associata alla sequenza di stati $\{\hat{S}\}$ di *minimo costo*⁵⁶.

Dato che da ogni stato si dipartono 2^k archi, ad ogni istante il numero di percorsi alternativi aumenta di un fattore 2^k , crescendo molto velocemente all'aumentare di j . L'enumerazione *completa* dei percorsi può essere però evitata, notando che quando due percorsi con costi diversi si incontrano in uno stesso nodo, quello di costo *maggiore* sicuramente *non* è la parte iniziale del percorso di minimo costo, e quindi può essere eliminato.

⁵⁴Questo caso viene indicato con il termine *hard-decision decoding* in quanto il ricevitore *ha già* operato una decisione (quantizzazione) rispetto a $\tilde{b}$. Al contrario, se i valori ricevuti sono passati *come sono* al decodificatore di Viterbi, questo può correttamente valutare le probabilità $p(\tilde{b}/\hat{b})$ ed operare in modalità *soft decoding*, conseguendo prestazioni migliori.

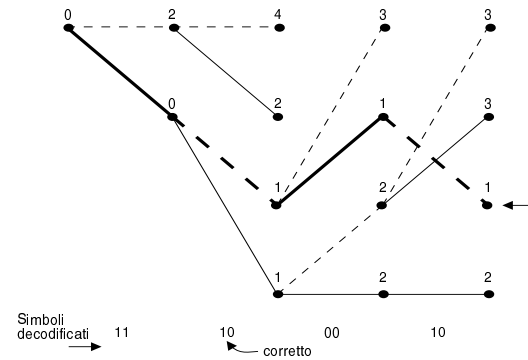
⁵⁵Ad esempio, con riferimento alla fig. 4.5, la $\{S\} = \{00, 10, 11, 01, 10\}$ ha un *costo* pari a 3.

⁵⁶Qualora la distanza tra $\tilde{b}_j$ ed un possibile $\hat{b}_j$ sia espressa come probabilità condizionata $p(\tilde{b}_j/\hat{b}_j)$, il processo di decodifica è detto di *massima verosimiglianza*.

Questa filosofia si applica al caso in questione con riferimento alla figura che segue, che mostra come il calcolo dei *costi parziali* avvenga *per colonne* da sinistra a destra, scrivendo sopra ad ogni nodo il costo del *miglior* percorso che lo raggiunge.

Ad ogni colonna sono scartati i percorsi che si incontrano con uno migliore, cosicchè il numero di percorsi *sopravvissuti* è sempre pari al numero di stati 2^k . All'estremità destra della figura, una freccia indica la minima $d_H(\hat{b}, \tilde{b})$, associata al percorso a tratto spesso, e che permette di individuare la $\{\tilde{b}\}$, che come si vede è *quella esatta*.

Tralasciamo ora di approfondire la teoria che consente l'analisi dettagliata dell'algoritmo, e ci limitiamo alle seguenti



Riflessioni

- l'esempio fornito si è mostrato in grado di correggere un errore pur impiegando un coding rate pari ad $\frac{1}{2}$, superiore a quello ($\frac{1}{3}$) del codice a ripetizione;
- la d_H del miglior percorso corrisponde al numero di bit errati (nel caso in cui siano stati corretti) nella $\hat{b}$ ricevuta;
- si verifica errore (cioè $\{\hat{b}\} \neq \{b\}$) se $d_H(\hat{b}, \tilde{b})$ è *minore* di $d_H(b, \tilde{b})$;
- le capacità di correzione del codice migliorano aumentando la d_H tra le possibili sequenze $\{b\}$ ⁵⁷;
- la d_H tra diverse $\{b\}$ aumenta con $\frac{K}{k}$, in quanto la matrice di transizione tra stati diviene più sparsa, ed i valori di $\{b\}$ sono più interdipendenti.

⁵⁷La minima distanza tra le sequenze codificate è indicata come d_{min} , e può essere trovata come la d_H tra una $\{b^0\}$ tutta nulla ($\{b^0\} = \{...00000000\}$) e quella con il minor numero di uni, che si diparte e ritorna (nel traliccio) dallo/allo stato 00.

Capitolo 5

Probabilità, Processi ed Errori

In questo capitolo sono trattati una serie di argomenti che hanno in comune l'utilizzo dei risultati della teoria delle probabilità, brevemente riassunti nelle prime sezioni. Dopo aver caratterizzato le variabili aleatorie mediante i momenti, e definite le proprietà di ergodicità dei processi, viene valutata l'entità del rumore di quantizzazione, e calcolata la probabilità di errore per una trasmissione numerica. Un'ampia appendice è dedicata ad approfondire alcuni aspetti particolari della teoria, come elementi di decisione statistica, l'entropia e la codifica di sorgente, la quantizzazione logaritmica, e le distribuzioni di Raileigh e Rice.

5.1 Teoria delle probabilità

Tratta delle caratteristiche *regolari* di fenomeni *irregolari* o *casuali*. Una prima definizione di probabilità è quella fornita dalla teoria frequentistica, per cui se, ripetendo N volte un esperimento, n_A volte si verifica la circostanza A , a quest'ultima si assegna una frequenza relativa n_A/N , da cui si deriva la probabilità di A come

$$Pr_A = \lim_{N \rightarrow \infty} \frac{n_A}{N}$$

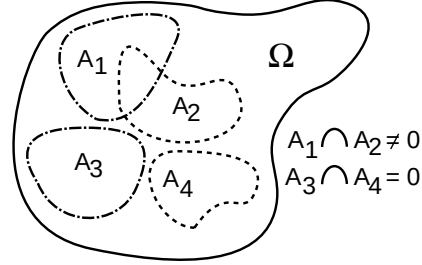
In termini più astratti, si guarda all'insieme di tutte le circostanze possibili come a degli insiemi algebrici, i cui elementi (o punti) sono appunto le diverse circostanze. I punti possono essere raggruppati in sottoinsiemi (eventualmente vuoti o di un solo punto) per i quali valgono le proprietà di unione, intersezione, complemento, inclusione...

I fenomeni fisici sono posti in relazione con i punti degli insiemi suddetti mediante il concetto di *spazio campione* Ω , che è l'unione di tutti i possibili risultati di un *fenomeno aleatorio*. Sottoinsiemi dello spazio campione sono detti *eventi*. L'intero spazio è l'*evento certo*, mentre l'insieme vuoto corrisponde all'evento impossibile ϕ (od evento *nullo*). Una *unione* $\cup$ di eventi, corrisponde all'evento che si verifica ogni qualvolta se ne verifichi *un suo componente*, mentre *l'intersezione* $\cap$ è verificata se *tutti* i componenti lo sono. Esempio: il lancio di un dado genera uno spazio con 6 punti (eventi) disgiunti. Uno spazio campione può avere un numero di punti finito, infinito numerabile, o infinito.

5.1.1 Assiomi delle probabilità

Costituiscono le basi su cui sono costruiti i teoremi seguenti, ed affermano che:

- $0 \leq P(A) \leq 1$: la probabilità di un evento è compresa tra 0 ed 1;
- $Pr(\Omega) = 1$: la probabilità dell'evento certo è 1;
- Se $Pr(A_i \cap A_j) = \phi$ allora $Pr(\bigcup A_i) = \sum Pr(A_i)$: la probabilità dell'unione di eventi mutuamente indipendenti è la somma delle singole probabilità.



5.1.2 Teoremi di base

- $Pr(\phi) = 0$: la probabilità dell'evento impossibile è nulla.
- $Pr(A \cap B) + Pr(A \cap \overline{B}) = Pr(A)$ e $Pr(B) + Pr(\overline{B}) = 1$: un evento ed il suo complemento riempiono lo spazio (detto anche teorema delle *probabilità totali*¹)
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$: la probabilità dell'evento intersezione si conta una volta sola. Ad esempio: lanciando un dado, la probabilità $Pr(\text{pari} \cup > 2)$ di ottenere un numero pari, oppure più grande di due, è la *somma* delle probabilità dei singoli eventi $Pr(\text{pari}) = \frac{3}{6}$ e $Pr(> 2) = \frac{4}{6}$, meno quella che si verificano assieme $Pr(\text{pari} \cap > 2) = \frac{2}{6}$. Pertanto: $Pr(\text{pari} \cup > 2) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$.
- Se $B \subseteq A$ allora $Pr(B) \leq Pr(A)$: quando l'evento B è contenuto in A il verificarsi del primo implica il secondo.

5.1.3 Probabilità condizionali

Può avvenire che il verificarsi di un evento *influenzi* il verificarsi o meno di un altro: Si dice allora che lo condiziona, ovvero che l'evento influenzato è *condizionato*. La probabilità che avvenga A , noto che B si sia verificato, si scrive $Pr(A/B)$, e si legge probabilità (condizionata) di A dato B , che è definita come

$$Pr(A/B) = \frac{Pr(A, B)}{Pr(B)}$$

in cui $Pr(A, B) = Pr(A \cap B)$ è la probabilità *congiunta* che A e B si verificano entrambi, ed a patto che $Pr(B) \neq 0$ (altrimenti anche $Pr(A/B)$ è zero!).

¹Utile per scrivere la probabilità di un evento come “1 meno” quella dell'evento complementare.

Esercizio: Valutare la probabilità condizionata $Pr(A/B)$ che lanciando un dado si otenga un numero pari (evento $A = (\text{pari})$), condizionatamente all'evento B che il numero sia > 2 . Soluzione alla nota².

A partire dalla precedente definizione, si ottiene quella della probabilità *congiunta*: $Pr(A, B) = Pr(A/B) Pr(B)$; inoltre, gli eventi condizionante e condizionato possono invertire i rispettivi ruoli, permettendo di scrivere anche: $Pr(A, B) = Pr(B/A) Pr(A)$. Eguagliando le due espressioni, si ottiene:

$$Pr(A/B) = \frac{Pr(B/A) Pr(A)}{Pr(B)} \quad \text{ed anche} \quad Pr(B/A) = \frac{Pr(A/B) Pr(B)}{Pr(A)}$$

Come ultima definizione, rammentiamo che le probabilità $Pr(A)$ e $Pr(B)$ sono indicate come *marginali*.

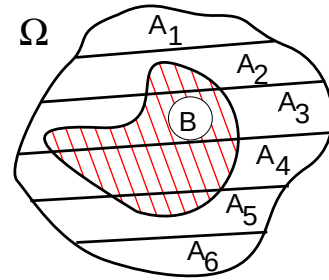
5.1.4 Teorema di Bayes

A volte, non tutti i possibili eventi sono direttamente osservabili: in tal caso la probabilità marginale $Pr(A)$ è indicata come probabilità *a priori*. Qualora l'evento A sia in qualche modo legato ad un secondo evento B , che invece possiamo osservare, la probabilità condizionata $Pr(A/B)$ prende il nome di probabilità *a posteriori* perché, a differenza di quella a priori, rappresenta un valore di probabilità valutata *dopo* la conoscenza di B .

In generale, però, si conosce solamente $Pr(A)$ e $Pr(B/A)$ (queste ultime sono dette probabilità condizionate *in avanti*), e per calcolare $Pr(A/B)$ occorre conoscere anche $Pr(B)$. Quest'ultima quantità si determina *saturando* la probabilità congiunta $Pr(A, B)$ rispetto a tutti gli eventi marginali A_i possibili:

$$Pr(B) = \sum_i Pr(B, A_i) = \sum_i Pr(B/A_i) Pr(A_i)$$

a patto che risulti $Pr(A_i, A_j) = 0$ e $\bigcup A_i = \Omega$, e cioè che l'insieme degli $\{A_i\}$ costituisca una partizione dello spazio degli eventi Ω . Tale circostanza è mostrata in figura.



L'ultima relazione ci permette di enunciare il *teorema di Bayes*, che mostra come ottenere le probabilità *a posteriori* a partire da quelle *a priori* e da quelle condizionate *in avanti*:

$$Pr(A_i/B_j) = \frac{Pr(B_j/A_i) Pr(A_i)}{\sum_k Pr(B_j/A_k) Pr(A_k)}$$

Nella appendice 5.6.1 è mostrata l'applicazione del risultato ad un problema di decisione statistica tipico delle telecomunicazioni, relativo alla ricezione binaria. In 5.1.6 è invece proposto un esempio più diretto.

²Il risultato è pari alla probabilità $Pr(A, B) = Pr(\text{pari}, > 2)$ che i due eventi si verifichino contemporaneamente, divisa per la probabilità $Pr(B) = Pr(> 2)$ che il numero sia > 2

Si rifletta sulla circostanza che la probabilità del pari $Pr(A) = \frac{1}{2}$, quella $Pr(B) = \frac{4}{6}$, o quella congiunta di entrambi $Pr(A, B) = \frac{2}{6}$, sono tutte riferite ad un qualunque lancio di dado, mentre $Pr(\text{pari} / > 2)$ è relativa ad un numero ridotto di lanci, ossia solo quelli che determinano un risultato > 2 . Pertanto, essendo $Pr(B) \leq 1$, si ottiene $Pr(A/B) \geq Pr(A)$; infatti per l'esempio del dado si ottiene $Pr(\text{pari} / > 2) = Pr(\text{pari}, > 2) / Pr(> 2) = \frac{2/6}{4/6} = \frac{1}{2}$, che è maggiore di $Pr(\text{pari}, > 2) = \frac{1}{3}$ (i valori di probabilità sono ottenuti come rapporto tra il numero di casi favorevoli e quello dei casi possibili).

Si ottiene invece $Pr(A/B) = Pr(A)$ solo se $Pr(B) = 1$, ossia se B corrisponde all'unione di tutti gli eventi possibili.

5.1.5 Indipendenza statistica

Si verifica quando

$$Pr(A/B) = Pr(A)$$

in quanto il verificarsi di B non influenza A . Come conseguenza, per due eventi statisticamente indipendenti avviene che

$$Pr(A, B) = Pr(A) Pr(B)$$

5.1.6 Esempi riassuntivi

- Un sistema di comunicazione radio è affetto da attenuazioni supplementari causate da pioggia. Indicando con FS l'evento che il sistema vada fuori servizio, e conoscendo le probabilità condizionate $Pr(FS/piove) = .5$, $Pr(FS/non\ piove) = .05$ e la probabilità marginale $Pr(piove) = .03$, determinare:

1. La probabilità di fuori servizio $Pr(FS)$, indipendentemente dal verificarsi o meno dell'evento piovoso;
2. La probabilità che stia piovendo, sapendo che il sistema è fuori servizio.

Risultato ⁽³⁾.

- Quale è la probabilità che, lanciando 3 volte un dado, esca 3 volte 1 ? Risultato ⁽⁴⁾.
- Un'urna contiene 2 biglie bianche e 3 nere. Qual è la probabilità che su 2 estrazioni consecutive, escano le 2 biglie bianche ? Risultato ⁽⁵⁾.
- Qual è la probabilità che 2 carte, estratte a caso da un mazzo da bridge da 52, siano K e Q ? Risultato ⁽⁶⁾.

³La probabilità marginale di fuori servizio si calcola applicando il teorema delle probabilità totali:

$$\begin{aligned} Pr(FS) &= Pr(FS/piove) \cdot Pr(piove) + Pr(FS/non\ piove) \cdot Pr(non\ piove) = \\ &= .5 \cdot .03 + .05 \cdot .97 = .0635 = 6.35\% \end{aligned}$$

in quanto $Pr(non\ piove) = 1 - Pr(piove) = .97$. Applicando il teorema di Bayes si trova quindi:

$$Pr(piove/FS) = \frac{Pr(FS/piove) \cdot Pr(piove)}{Pr(FS)} = \frac{.5 \cdot .03}{.0635} = .236 = 23.6\%$$

Si noti come la probabilità *a priori* che piovva (3 %) venga rimpiazzata dal suo valore *a posteriori* (23,6 %) grazie alla nuova informazione di cui disponiamo (collegamento fuori servizio). Per una definizione precisa delle probabilità *a priori* ed *a posteriori* si veda l'appendice 5.6.1.

⁴E' pari al prodotto delle probabilità marginali, essendo i lanci statisticamente indipendenti, visto che il dado è "senza memoria". Pertanto il risultato è $(\frac{1}{6})^3 = \frac{1}{216} \simeq 4.6296 \cdot 10^{-3}$.

⁵Anche qui l'urna è senza memoria; però dopo la prima estrazione le biglie restano in 4! Pertanto ora il prodotto delle probabilità marginali risulta $\frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$.

⁶ $Pr(K, Q) = Pr(K\ prima, Q\ seconda) + Pr(Q\ prima, K\ seconda) = Pr(K\ prima) \cdot Pr(Q\ seconda/K\ prima) + Pr(Q\ prima) \cdot Pr(K\ seconda/Q\ prima) = 2 \left(\frac{4}{52} \cdot \frac{4}{51} \right) = \frac{8}{663} \simeq 1.2 \cdot 10^{-2}$.

5.2 Variabili aleatorie

Finora si è parlato di *eventi* in modo astratto, mentre spesso ci si trova ad associare ad ogni punto dello spazio campione un valore numerico: lo spazio campione Ω diventa allora l'*insieme dei numeri* e prende il nome di *variabile aleatoria*. La realizzazione di un evento corrisponde ora all'assegnazione di un valore (tra i possibili) alla variabile aleatoria; tale valore "prescelto" prende dunque il nome di *realizzazione* della v.a. Distinguiamo inoltre tra variabili aleatorie *discrete* e *continue*, a seconda se la grandezza che descrivono abbia valori numerabili o continui⁷. La caratterizzazione della variabile aleatoria, in termini probabilistici, si ottiene indicando come la "massa di probabilità" si distribuisce sull'insieme di valori che la variabile aleatoria può assumere, per mezzo delle 2 funzioni di variabile aleatoria seguenti.

5.2.1 Funzioni di densità e di distribuzione di probabilità

Così come un oggetto non omogeneo è più o meno denso in regioni differenti del suo volume complessivo, così la *densità di probabilità* mostra su quali valori della variabile aleatoria si concentra la probabilità. Così, ad esempio, la densità della v.a. discreta associata al lancio di un dado può essere scritta: $p_D(x) = \sum_{n=1}^6 \frac{1}{6} \delta(x - n)$ il cui significato discutiamo subito, con l'aiuto dei due grafici seguenti.

D e x indicano rispettivamente la v.a. (il numero che uscirà) ed una sua realizzazione (una delle 6 facce). I 6 impulsi centrati in $x = n$ rappresentano una concentrazione di probabilità nei sei possibili valori; l'area di tali impulsi è proprio pari alla probabilità di ognuno dei sei risultati. E' facile verificare che

$$\int_{-\infty}^{\infty} p_D(x) dx = 1$$

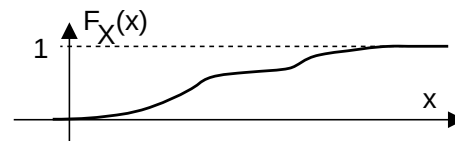
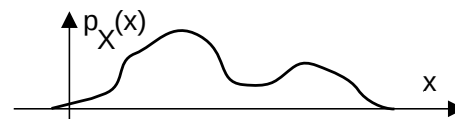
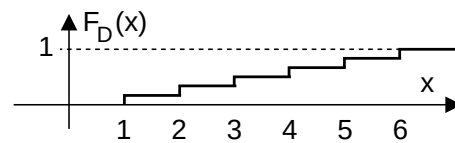
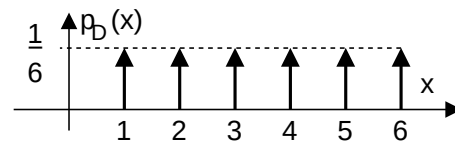
e che risulta

$$\int_a^b p_D(x) dx = Pr\{a < D \leq b\}$$

ovvero pari alla probabilità che la v.a. D assuma un valore tra a e b . In particolare, non potendosi verificare una probabilità negativa, si ha $p_D(x) \geq 0$ con $\forall x$.

Una funzione di v.a. strettamente collegata alla densità è la funzione *distribuzione di probabilità*⁸, definita come

$$F_X(x) = \int_{-\infty}^x p_X(\xi) d\xi = Pr\{X \leq x\}$$



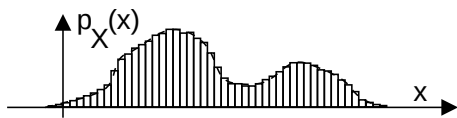
⁷Un esempio classico di v.a. discreta è quello del lancio di un dado, un altro sono i numeri del lotto. Una v.a. continua può essere ad esempio un valore di pressione atmosferica in un luogo, oppure l'attenuazione di una trasmissione radio dovuta a fenomeni atmosferici.

⁸In realtà, l'ordine storico è quello di definire prima $F_X(x)$ come la probabilità che X sia non superiore ad un valore x , ovvero $F_X(x) = Pr\{X \leq x\}$, e quindi $p_X(x) = \frac{dF_X(x)}{dx}$. Il motivo di tale "priorità" risiede nel fatto che $F_X(x)$ presenta minori "difficoltà analitiche" di definizione (ad esempio presenta solo discontinuità di prima specie, anche con v.a. discrete).

che risulta una funzione non decrescente di x , limitata ad un massimo valore di 1, ed il cui andamento mostriamo sotto alla $p_D(x)$, nel caso dell'esempio del lancio del dado⁹.

Le definizioni date mantengono validità nel caso di v.a. continua, originando le curve mostrate nei due grafici a lato. Ora è ancora più evidente la circostanza che $p_X(x)$ è una *densità*, e diviene una probabilità solo quando moltiplicata per un intervallo di x ⁽¹⁰⁾.

Istogramma Citiamo infine la stretta relazione che intercorre tra la densità di probabilità e l'*istogramma*.



Quest'ultimo può essere realizzato se si dispone di una serie di realizzazioni della v.a., e si ottiene suddividendo il campo di variabilità della grandezza X in sotto-intervalli, e disegnando rettangoli verticali, ognuno di altezza pari al numero di volte

che (nell'ambito del campione statistico a disposizione) X assume un valore in quell'intervallo. Dividendo l'altezza di ogni rettangolo per il numero di osservazioni N , si ottiene una approssimazione di $p_X(x)$, via via più precisa con $N \rightarrow \infty$, e con una conseguente riduzione dell'estensione degli intervalli.

5.2.2 Medie, momenti e momenti centrati

Indichiamo con $g(x)$ una funzione di una variabile aleatoria¹¹.

Si definisce *valore atteso* (o *media*, *media di insieme*, *media statistica*) di $g(x)$ rispetto alla variabile aleatoria X la quantità:

$$E_X \{g(x)\} = \int_{-\infty}^{\infty} g(x) p_X(x) dx$$

che corrisponde ad una media (integrale) pesata¹² dei diversi valori $g(x)$, ognuno con peso pari alla probabilità $p_X(x) dx$; la notazione $E_X \{.\}$ indica quindi¹³ tale operazione di media integrale, assieme alla v.a. (x) rispetto a cui eseguirla.

Nel caso in cui $g(x) = x^n$, il valore atteso prende il nome di *momento di ordine n* , che corrisponde quindi al valore atteso della n -esima potenza della v.a., e che si indica come

$$m_X^{(n)} = E \{x^n\} = \int_{-\infty}^{\infty} x^n p_X(x) dx$$

Verifichiamo subito che $m_X^{(0)} = 1$. Il momento di primo ordine

$$m_X = m_X^{(1)} = \int_{-\infty}^{\infty} x p_X(x) dx$$

⁹Si ricorda che la derivata di un gradino è un impulso di area pari al dislivello. Infatti, sapevamo già che l'integrale di un impulso è una costante - ammesso che l'impulso cada dentro gli estremi di integrazione!

¹⁰Infatti la probabilità che X cada tra x_0 e $x_0 + \Delta x$ vale $\int_{x_0}^{x_0 + \Delta x} p_X(x) dx \simeq p_X(x_0) \Delta x$.

¹¹Un esempio di funzione di v.a. potrebbe essere il valore della vincita associata ai 13 in schedina, che dipende dalla v.a. rappresentata dai risultati delle partite, una volta noto il montepremi e le giocate. Infatti, per ogni possibile vettore di risultati, si determina un diverso numero di giocate vincenti, e quindi un diverso modo di suddividere il montepremi.

¹²Notiamo che se al posto delle probabilità $p_X(x) dx$ utilizziamo i valori di un istogramma $Pr(x_i) = \frac{N(x_i < x \leq x_i + \Delta x)}{N}$, l'integrale si trasforma in una sommatoria, il cui sviluppo evidenzia l'equivalenza con una media pesata: $\sum x_i Pr(x_i) = \frac{x_1 N_1 + x_2 N_2 + \dots + x_n N_n}{N}$.

¹³In effetti, la E simboleggia la parola EXPECTATION, che è il termine inglese usato per indicare il valore atteso.

prende il nome di *media*¹⁴ della v.a. X (a volte denominata *centroide*), mentre con $n = 2$ si ha la *media quadratica* $m_X^{(2)} = \int_{-\infty}^{\infty} x^2 p_X(x) dx$.

Nel caso di variabili aleatorie discrete, i momenti sono definiti come $m_X^{(n)} = \sum_i x_i^n p_i$, in cui $p_i = \text{Pr}\{x = x_i\}$, pesando quindi le possibili realizzazioni x_i con le rispettive probabilità.

Nel caso in cui $g(x) = (x - m_X)^n$, il relativo valore atteso è chiamato *momento centrato* di ordine n , ed indicato come

$$\mu_X^{(n)} = E\{(x - m_X)^n\} = \int_{-\infty}^{\infty} (x - m_X)^n p_X(x) dx$$

E' immediato constatare che $\mu_X^{(0)} = 1$ e che $\mu_X^{(1)} = 0$. Il momento centrato del 2° ordine prende il nome di *varianza*, e si indica

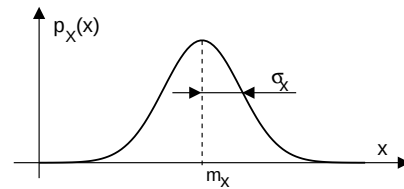
$$\sigma_X^2 = \mu_X^{(2)} = E\{(x - m_X)^2\} = \int_{-\infty}^{\infty} (x - m_X)^2 p_X(x) dx$$

Una relazione notevole che lega i primi due momenti (centrati e non) è ⁽¹⁵⁾:

$$\sigma_X^2 = m_X^{(2)} - (m_X)^2$$

La radice quadrata della varianza, σ_X , prende il nome di *deviazione standard*. Mentre la media m_X indica dove si colloca il “centro statistico” della densità di probabilità, σ_X indica quanto le singole determinazioni della v.a. siano disperse attorno ad m_X .

In appendice 5.6.2 è presentata una interessante applicazione dell'operatore di valore atteso alla misura della *quantità media di informazione* presente in un messaggio, assieme ad una tecnica di *codifica di sorgente* idonea a ridurre la dimensione del messaggio al minimo, conservandone per intero il contenuto informativo.



5.2.3 Variabile aleatoria a distribuzione uniforme

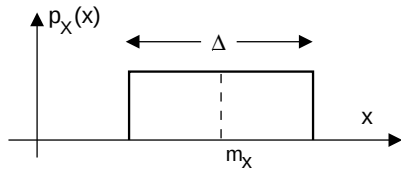
Applichiamo le definizioni dei momenti ad un caso pratico: la variabile aleatoria uniforme è caratterizzata da uno stesso valore di probabilità per tutta la gamma di realizzazioni possibili, limitate queste ultime ad un unico intervallo non disgiunto; pertanto, la densità di probabilità è esprimibile mediante una funzione rettangolare:

$$p_X(x) = \frac{1}{\Delta} \text{rect}_{\Delta}(x - m_X)$$

in cui Δ rappresenta l'estensione dell'intervallo di esistenza della variabile aleatoria.

¹⁴Supponiamo che X rappresenti l'altezza degli individui; l'altezza *media* sarà allora calcolabile proprio come momento del primo ordine.

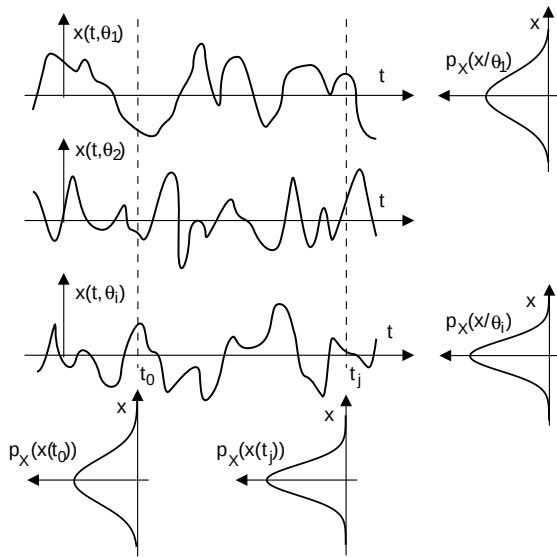
¹⁵Infatti risulta: $\sigma_X^2 = E\{(x - m_X)^2\} = E\{x^2 + m_X^2 - 2xm_X\} = E\{x^2\} + m_X^2 - 2m_X E\{x\} = m_X^{(2)} + (m_X)^2 - 2(m_X)^2 = m_X^{(2)} - (m_X)^2$. Si è preferito usare la notazione $E\{x\}$, più compatta rispetto all'indicazione degli integrali coinvolti; i passaggi svolti si giustificano ricordando la proprietà distributiva degli integrali (appunto), ed osservando che il valore atteso di una costante è la costante stessa.



E' facile verificare che il parametro m_X , che indica l'ascissa a cui è centrato il rettangolo, corrisponde esattamente al momento di primo ordine di X . Il calcolo della varianza¹⁶ invece fornisce: $\sigma_X^2 = \frac{\Delta^2}{12}$.

5.3 Processi Stazionari ed Ergodici

Dopo aver descritto come caratterizzare statisticamente singoli valori (denominati variabili aleatorie), occupiamoci del caso in cui si voglia descrivere da un punto di vista probabilistico un intero segnale, la cui reale identità non sia nota a priori¹⁷. Un segnale siffatto viene detto *membro* (o realizzazione) di un *processo aleatorio*, e può essere indicato come $x(t, \theta)$, che corrisponde ad una descrizione formale che prevede una coppia di insiemi: il primo di questi è l'insieme $\mathcal{T}$ degli istanti temporali (tipicamente un intervallo) su cui sono definiti i membri del processo; il secondo è relativo ad una variabile aleatoria Θ , i cui valori θ identificano ognuno una particolare realizzazione del processo.



Pertanto, una singola realizzazione $\theta = \theta_i$, per così dire, *indicizza* il processo, le cui istanze effettive $x(t, \theta_i)$, con $t \in \mathcal{T}$, sono note solo dopo la conoscenza di $\theta_i \in \Theta$ (¹⁸). Il processo aleatorio è quindi definito come l'insieme dei segnali $\{x(t, \theta)\}$, con $t \in \mathcal{T}$ e $\theta \in \Theta$.

Se viceversa fissiamo un particolare istante temporale t_j , il valore $x(t_j, \theta)$ è una variabile aleatoria, la cui realizzazione dipende da quella di $\theta \in \Theta$; pertanto, è definita la densità $p_X(x(t_j))$ (indipendente da θ), che possiamo disegnare in corrispondenza dell'istante t_j in cui è prelevato il campione¹⁹; a tale riguardo, si faccia riferimento alla figura a fianco, che mostra le densità di probabilità definite a partire dai membri di un processo.

Figura 5.1: Un processo *non ergodico*

5.3.1 Media di insieme

E' definita come il *valore atteso* di una *potenza n-esima* dei valori del segnale, ossia un suo momento, eseguito rispetto alla variabilità dovuta a Θ , ed è pertanto calcolata come

¹⁶Anzichè calcolare σ_X^2 per la $p_X(x)$ data, calcoliamo $m_X^{(2)}$ per una v.a. uniforme con $m_X = 0$: in tal caso infatti $m_X^{(2)} = \sigma_X^2$. Si ha: $\sigma_X^2 = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} x^2 \frac{1}{\Delta} dx = \frac{x^3}{3\Delta} \Big|_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} = \frac{1}{3\Delta} \left(\frac{\Delta^3}{8} + \frac{\Delta^3}{8} \right) = \frac{1}{3\Delta} 2 \frac{\Delta^3}{8} = \frac{\Delta^2}{12}$.

¹⁷Chiaramente, la maggioranza dei segnali trasmessi da apparati di TLC sono di questo tipo.

¹⁸Per fissare le idee, conduciamo parallelamente al testo un esempio "reale" in cui il processo aleatorio è costituito da.... la selezione musicale svolta da un dj. L'insieme $\mathcal{T}$ sarà allora costituito dall'orario di apertura delle discoteche (dalle 22 all'alba ?), mentre in θ faremo ricadere tutte le caratteristiche di variabilità (umore del dj, i dischi che ha in valigia, la discoteca in cui ci troviamo, il giorno della settimana...).

¹⁹Nell'esempio, $x(t_0, \theta)$ è il valore di pressione sonora rilevabile ad un determinato istante (es. le 23.30) al variare di θ (qualunque dj, discoteca, giorno...).

$$m_X^{(n)}(t_j) = E_\Theta \{x^n(t_j, \theta)\} = \int_{-\infty}^{\infty} x^n(t_j, \theta) p_\Theta(\theta) d\theta = \int_{-\infty}^{\infty} x^n p_X(x(t_j)) dx$$

in cui l'ultima eguaglianza evidenzia come una media di insieme dipenda dalla distribuzione $p_X(x(t_j))$ di $x(t_j, \theta)$ per tutti i $\theta \in \Theta$, mostrata in basso in fig. 5.1. Notiamo che in linea di principio, la media di insieme dipende dall'istante t_j in cui è prelevato un valore²⁰.

5.3.2 Medie temporali

In alternativa, possiamo fissare una particolare realizzazione θ_i di Θ , e quindi identificare un singolo membro $x(t, \theta_i)$, che è ora un segnale certo²¹; per lo stesso, possono quindi essere calcolate le *medie temporali*, indicate con una linea sopra alla quantità di cui si calcola la media $\overline{(\cdot)}$:

$$m_X^{(n)}(\theta_i) = \overline{x^n(t, \theta_i)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^n(t, \theta_i) dt$$

In particolare, ritroviamo il *valore medio* $m_X(\theta_i) = \overline{x(t, \theta_i)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t, \theta_i) dt$ e la potenza²² (*media quadratica*) $m_X^{(2)}(\theta_i) = \overline{x^2(t, \theta_i)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t, \theta_i) dt$. Notiamo che una generica media temporale:

- non dipende dal tempo;
- è una variabile aleatoria (dipende infatti dalla realizzazione di Θ).

5.3.3 Medie temporali calcolate come medie di insieme

L'estrazione da $x(t, \theta_i)$ di un valore ad un istante casuale $t = t_j \in \mathcal{T}$, definisce una ulteriore variabile aleatoria, descritta dalla densità di probabilità (condizionata) $p_X(x/\theta_i)$, che disegniamo a fianco dei singoli membri mostrati in fig. 5.1. Qualora la $p_X(x/\theta_i)$ sia nota, le medie temporali di ordine n possono essere calcolate (per quel membro) come i momenti:

$$m_X^{(n)}(\theta_i) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^n(t, \theta_i) dt = \int_{-\infty}^{\infty} x^n p_X(x/\theta_i) dx = E_{X/\Theta=\theta_i} \{x^n\}$$

Ciò equivale infatti ad effettuare una media ponderata, in cui ogni possibile valore di x è pesato per la sua probabilità $p_X(x/\theta_i) dx$.

²⁰Ad esempio, se in tutte le serate il volume aumenta progressivamente nel tempo, la $p_X(x(t_j))$ si allargherà per t_j crescenti.

²¹ $x(t, \theta_i)$ rappresenta, nel nostro esempio, l'intera selezione musicale (detta *serata*) proposta da un ben preciso dj, in un preciso locale, un giorno ben preciso.

²² $m_X^{(2)}(\theta_i)$ in questo caso rappresenta la potenza media con cui è suonata la musica nella particolare serata θ_i .

5.3.4 Processi stazionari

Qualora $p_X(x(t_j))$ non dipenda da t_j , ma risulti $p_X(x(t_j)) = p_X^T(x)$ per qualsiasi $t_j \in \mathcal{T}$, il processo $\{x(t, \theta)\}$ è detto *stazionario*²³ *in senso stretto*. In tal caso tutte le medie di insieme non dipendono più dal tempo, ossia $m_X^{(n)}(t) = m_X^{(n)}$ per $\forall t \in \mathcal{T}$, e le curve in basso in fig. 5.1 sono tutte uguali.

Se invece solamente le prime due medie di insieme $m_X(t)$ e $m_X^{(2)}(t)$ non dipendono da t , il processo $\{x(t, \theta)\}$ è detto *stazionario in media* ed *in media quadratica*, od anche *stazionario in senso lato*²⁴.

Supponiamo ora di suddividere il membro $x(t, \theta_i)$ in più intervalli temporali, e di calcolare per ciascuno di essi le medie temporali, limitatamente al relativo intervallo. Nel caso in cui queste risultino uguali tra loro, e di conseguenza uguali alla media temporale $m_X^{(n)}(\theta_i)$, il membro è (individualmente) stazionario²⁵. Ovviamente, se tutti i membri sono individualmente stazionari, lo è anche il processo a cui appartengono.

5.3.5 Processi stazionari ed ergodici

Questa importante sottoclasse di processi stazionari identifica la circostanza che *ogni membro del processo è statisticamente rappresentativo di tutti gli altri*. Ciò si verifica quando la densità di probabilità (a destra in fig. 5.1) dei valori estratti da un singolo membro $p_X(x/\theta_i)$ è sempre la stessa, indipendentemente dal particolare θ_i , ottenendo in definitiva $p_X(x/\theta_i) = p_X^\Theta(x)$ indipendentemente dalla realizzazione e, per la stazionarietà, anche $p_X(x/\theta_i) = p_X^T(x)$, e dunque $p_X^\Theta(x) = p_X^T(x) = p_X(x)$. In questo caso le medie temporali $m_X^{(n)}(\theta_i)$, calcolabili come momenti sulla singola realizzazione come illustrato in 5.3.3, sono identiche per tutti i membri²⁶ θ_i , ed identiche anche alle medie di insieme $m_X^{(n)}(t_j)$ calcolate per un qualunque istante. Enunciamo pertanto la definizione:

Un processo stazionario è ergodico se la media temporale calcolata su di una qualunque realizzazione del processo, coincide con la media di insieme relativa ad una variabile aleatoria estratta ad un istante qualsiasi (per la stazionarietà) da una realizzazione qualsiasi (per l'ergodicità).

Esempio: la potenza di segnale Mostriamo come il calcolo della potenza di un membro di un processo ergodico sia equivalente a quello del momento di 2° ordine del processo:

$$\mathcal{P}_X(\theta) = \overline{x^2(\theta)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(\tau, \theta) d\tau = \int_{-\infty}^{\infty} x^2 p_X(x/\theta) dx =$$

²³La “serata in discoteca” stazionaria si verifica pertanto se non mutano nel tempo il genere di musica, il volume dell'amplificazione... o meglio se eventuali variazioni in alcune particolari discoteche-realizzazioni sono compensate da variazioni opposte in altrettanti membri del processo.

²⁴In questo caso la $p_X(x(t))$ non è nota, oppure non è stazionaria, ma le maggiori applicazioni della proprietà di stazionarietà dipendono solo da $m_X(t)$ e $m_X^{(2)}(t)$, che possono essere misurati (o per meglio dire *stimati*), e risultare stazionari anche se $p_X(x(t))$ non lo è.

²⁵Questo accade se la selezione musicale di una particolare serata si mantiene costante (es. solo raggamuffin) oppure variata ma in modo omogeneo (es. senza tre “lenti” di fila).

²⁶Volendo pertanto giungere alla definizione di una serata *ergodica* in discoteca, dovremmo eliminare quei casi che, anche se individualmente stazionari, sono decisamente “fuori standard” (tutto metal, solo liscio...).

$$= \int_{-\infty}^{\infty} x^2 p_X(x) dx = m_X^{(2)} = E\{x^2\} = \mathcal{P}_X$$

Questo risultato mostra come sia possibile calcolare la potenza di una realizzazione di un processo, senza conoscerne la forma d'onda.

In particolare osserviamo che, essendo $\mathcal{P}_X = m_X^{(2)} = \sigma_x^2 + (m_x)^2$, per i segnali a media nulla ($m_x = 0$) si ottiene $\mathcal{P}_X = \sigma_x^2$ ed il valore efficace $\sqrt{\mathcal{P}_X}$ coincide con la deviazione standard σ_x . La radice della potenza è inoltre spesso indicata come *valore RMS* (ROOT MEAN SQUARE), definito come $x_{RMS} = \sqrt{x^2(t)}$, ovvero la radice della media quadratica (nel tempo). Se il segnale è a media nulla, x_{RMS} coincide con il valore efficace; se $x(t)$ è membro di un processo ergodico a media nulla, x_{RMS} coincide con la deviazione standard.

5.3.6 Riassumendo

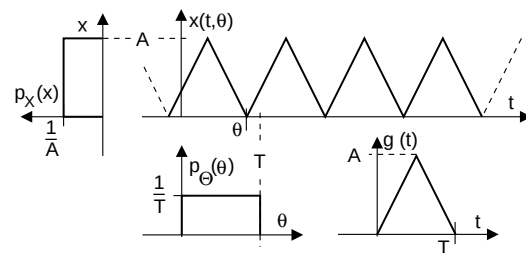
- Se un processo è ergodico, è anche stazionario, ma non il viceversa. Esempio: se $x(t, \theta) = C_\theta$ pari ad una costante (aleatoria), allora è senz'altro stazionario, ma $p_X(x/\theta) = \delta(x - C_\theta)$, e quindi non ergodico.
- Se un processo è ergodico è possibile:
 - Calcolare le medie di insieme in forma di medie temporali a partire da una singola realizzazione *oppure*
 - Ottenere le medie temporali di una qualunque realizzazione a partire dalle medie di insieme, disponendo della statistica $p_X(x)$.
- Se l'eguaglianza tra medie di insieme e temporali sussiste solo fino ad un determinato ordine e non oltre, il processo *non* è ergodico *in senso stretto*. Per ciò che concerne le Telecomunicazioni, è spesso sufficiente la proprietà di ergodicità in senso lato, ovvero limitata al 2° ordine, che garantisce $\overline{x(t)} = E\{x\} = m_x$; $\overline{x^2(t)} = E\{x^2\} = m_x^{(2)}$.

5.3.7 Processo ad aleatorietà parametrica

Questo è il nome dato a processi $\{x(t, \theta)\}$ per quali il parametro θ compare in modo esplicito nella espressione analitica dei segnali membri.

Come esempio, il segnale periodico $x(t, \theta) = \sum_{n=-\infty}^{\infty} A \cdot \text{tri}_T(t - \theta - nT)$, rappresentato in figura, ha come parametro un ritardo θ , che è una variabile aleatoria che ne rende imprecisata la fase iniziale.

Se θ è una v.a. uniformemente distribuita tra 0 e T (ovvero $p_\Theta(\theta) = \frac{1}{T} \text{rect}_T(\theta - \frac{T}{2})$), allora il processo è stazionario, ergodico, e la sua densità di probabilità risulta $p_X(x) = \frac{1}{A} \text{rect}_A(x - \frac{A}{2})$. Il valor medio $m_X = E\{x\}$ è pari alla media temporale $\frac{A}{2}$, la varianza è quella della d.d.p.²⁷ uniforme $\sigma_X^2 = \frac{A^2}{12}$ e la potenza vale $\mathcal{P}_X = \sigma_X^2 + m_X^2 = \frac{A^2}{12} + \frac{A^2}{4} = \frac{4A^2}{12} = \frac{A^2}{3}$.

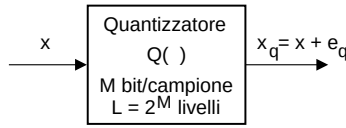


²⁷d.d.p. è l'abbreviazione comunemente usata per Densità di Probabilità.

Se la $p_\Theta(\theta)$ fosse stata diversa, il processo avrebbe potuto perdere ergodicità. Se ad esempio $p_\Theta(\theta) = \frac{4}{T} \text{rect}_{\frac{T}{4}}(\theta)$, si sarebbe persa la stazionarietà: infatti prendendo ad esempio $-\frac{T}{8} < t < \frac{T}{8}$, tutte le realizzazioni avrebbero valori minori del valor medio $\frac{A}{2}$.

Un secondo esempio di processo ad aleatorietà parametrica è il *processo armonico*, i cui membri hanno espressione $x(t, \theta) = A \cos(2\pi f_0 t + \theta)$, dove θ è una v.a. uniforme con d.d.p. $p_\Theta(\theta) = \frac{1}{2\pi} \text{rect}_{2\pi}(\theta)$. In tal caso il processo è stazionario ed ergodico, e si ottiene che una valore estratto a caso da un membro qualsiasi è una v.a. con d.d.p. $p_X(x) = \frac{1}{\pi \sqrt{A^2 - x^2}}$.

5.4 SNR di Quantizzazione



Trattiamo ora della questione, lasciata in sospeso, di come scegliere il numero M di bit con cui rappresentare i campioni di un segnale, ovvero della *risoluzione* con cui realizzare il dispositivo già indicato in 4.1.3 come *quantizzatore*. Tale processo consiste nel rappresentare i valori x in ingresso mediante un insieme finito di $L = 2^M$ valori quantizzati

$$x_q = x + \epsilon_q$$

introducendo dunque un errore ϵ_q .

L'obiettivo è quello di scegliere L in modo da mantenere il rapporto segnale rumore di quantizzazione SNR_q migliore di un valore desiderato. Dato che l' SNR_q è pari al rapporto tra le potenze del segnale $\mathcal{P}_x$ e del rumore $\mathcal{P}_\epsilon$, procediamo nel determinare queste ultime due, con l'aiuto del grafico mostrato in Fig. 5.2.

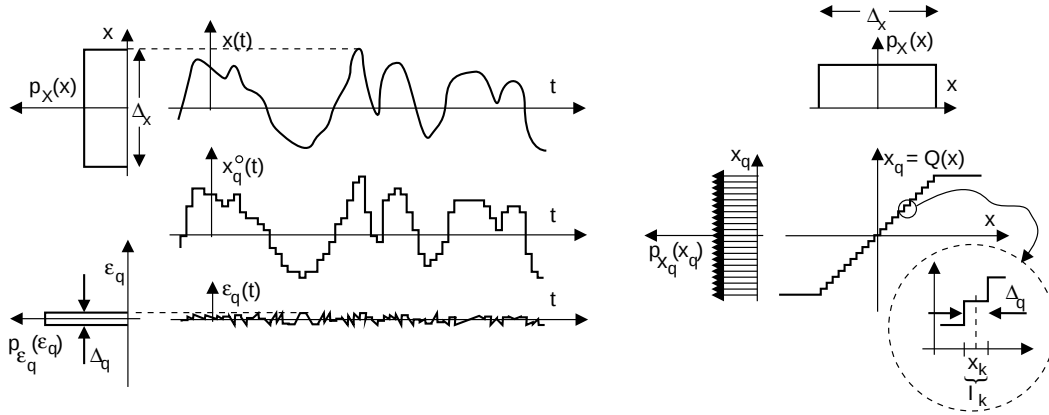


Figura 5.2: Processo di quantizzazione per segnali a distribuzione di ampiezza uniforme

Adottiamo l'ipotesi semplificativa che i valori in ingresso al campionatore abbiano origine da un processo ergodico a media nulla, e siano rappresentati da una v.a. con densità di probabilità uniforme $p_X(x) = \frac{1}{\Delta_x} \text{rect}_{\Delta_x}(x)$; pertanto la potenza dei campioni, pari alla varianza della v.a., risulterà

$$\mathcal{P}_x = \sigma_x^2 = \frac{\Delta_x^2}{12}$$

Lo stesso intervallo di valori di ingresso $\pm \frac{\Delta_x}{2}$ è suddiviso dal quantizzatore in L intervalli I_k di eguale ampiezza $\Delta_q = \frac{\Delta_x}{(L-1)}$, centrati sui valori $x_k = k\Delta_q - \frac{\Delta_x}{2}$, con $k = 0, 1, 2, \dots, L-1$. Tutti i valori di ingresso x , che cadono all'interno di I_k , ovvero tali che $x_k - \frac{\Delta_q}{2} \leq x < x_k + \frac{\Delta_q}{2}$, sono codificati con l'intero k , rappresentato in binario da una parola di $M = \log_2 L$ bit.

Il componente che, a partire dai valori quantizzati, ricostruisce il segnale $x_q^\circ(t)$ da inviare al filtro di restituzione, tipicamente associa ad ogni intero k il valore centrale $x_q = x_k$ dell'intervallo di quantizzazione, commettendo così un errore $\epsilon_q = x_q - x$, di entità limitata entro l'intervallo $\pm \frac{\Delta_q}{2}$. Di nuovo, si suppone che anche ϵ_q sia una v.a. uniformemente distribuita tra $\pm \frac{\Delta_q}{2}$, ed indipendente²⁸ da x_k ; pertanto, la potenza della componente di errore è anche qui pari alla varianza, e cioè

$$\mathcal{P}_\epsilon = \sigma_\epsilon^2 = \frac{\Delta_q^2}{12} = \frac{1}{12} \left(\frac{\Delta_x}{L-1} \right)^2$$

Siamo finalmente in grado di valutare l' SNR di quantizzazione:

$$SNR_q = \frac{\mathcal{P}_x}{\mathcal{P}_\epsilon} = \frac{\Delta_x^2}{12} 12 \left(\frac{L-1}{\Delta_x} \right)^2 = (L-1)^2 \cong L^2$$

in cui l'ultima approssimazione ha validità nel caso evidente in cui $L \gg 1$. Il risultato mostra che l' SNR_q cresce in modo quadratico con l'aumentare dei livelli, ovvero se L raddoppia SNR_q quadruplica. Ricorrendo alla notazione in decibel²⁹ per l' SNR , otteniamo il risultato $SNR_q(L)|_{dB} = 10 \log_{10} L^2 = 20 \log_{10} L$ e, ricordando che $L = 2^M$, si ottiene

$$SNR_q(M)|_{dB} = M \cdot 10 \log_{10} 2 \simeq 6 \cdot M \quad \text{dB}$$

dato che $\log_{10} 2 \simeq 0.3$. In modo simile, valutiamo il miglioramento in dB ottenibile aumentando di uno il numero di bit per ogni campione, ovvero raddoppiando il numero di livelli: $SNR_q(2L)|_{dB} = 20 \log_{10} 2L = 20 \log_{10} L + 20 \log_{10} 2 \cong SNR_q(L)|_{dB} + 6 \text{ dB}$. Pertanto ogni bit in più provoca un miglioramento di 6 dB per l' SNR_q .

Consideriamo ora cosa accade se il segnale in ingresso x ha una dinamica *minore* di quanto previsto: in tal caso σ_x^2 si riduce, mentre $\sigma_\epsilon^2 = \frac{1}{12} \left(\frac{\Delta_x}{L-1} \right)^2$ non cambia, e dunque SNR_q peggiora come se avessimo ridotto i livelli. In appendice § 5.6.4 è illustrata la tecnica usata *nella pratica* per mantenere un SNR_q elevato anche con bassi livelli di segnale.

5.5 Probabilità di Errore nelle Trasmissioni Numeriche di Banda base

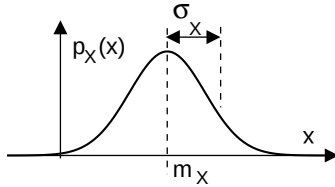
Con l'aiuto degli strumenti probabilistici studiati, analizziamo il funzionamento del dispositivo di ricezione di una trasmissione numerica, dal punto di vista della probabilità di errore. Come già illustrato, ogni ricevitore ha a che fare, oltre che con il segnale effettivamente ricevuto $r(t)$, anche con un segnale $n(t)$ sommato al primo, la cui forma

²⁸Questa ed altre ipotesi adottate sono palesemente non vere, ma permettono di giungere ad un risultato abbastanza semplice, e che può essere molto utile nel dimensionamento *di massima* degli apparati.

²⁹Una breve introduzione alle misure in decibel è presentato in appendice 5.6.3.

d'onda effettiva è una realizzazione del processo di *rumore additivo* (appunto). Il processo di rumore è stazionario ed ergodico, e la densità di probabilità del primo ordine che lo descrive, è la “famosa” *gaussiana*, che introduciamo subito.

5.5.1 Gaussiana

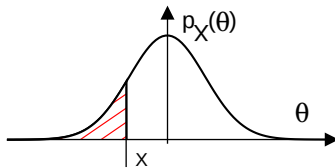


Una variabile aleatoria gaussiana x è descritta da una densità di probabilità di espressione

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp \left\{ -\frac{(x - m_x)^2}{2\sigma_x^2} \right\}$$

il cui andamento è mostrato in figura, dove è posto in evidenza come m_x e σ_x (media e deviazione standard) siano in relazione la prima con la centratura orizzontale, e la seconda con la dispersione della curva attorno alla media. Oltre che da un punto di vista grafico, i primi due momenti della v.a. descrivono completamente la densità anche dal punto di vista analitico; pertanto, la stima di questi (ad esempio a partire da un buon numero di realizzazioni³⁰) è sufficiente per descrivere completamente il fenomeno aleatorio.

La v.a. gaussiana descrive bene una moltitudine di fenomeni naturali, ed è dimostrabile analiticamente che la sua densità è caratteristica di grandezze generate dalla somma di un numero molto elevato di cause aleatorie, tutte con la medesima d.d.p.³¹ (*teorema del limite centrale*).

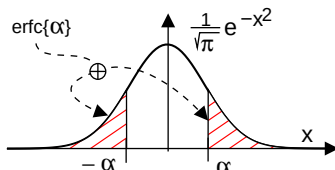


La funzione di distribuzione di questa v.a. non è calcolabile in forma chiusa; fortunatamente però, stante la necessità di conoscere la probabilità di eventi gaussiani, sono disponibili tabelle e grafici, che riportano il valore numerico dell'integrale che definisce tali valori di probabilità.

Per renderci conto della situazione, proviamo a calcolare la probabilità che X non superi un certo valore x , pari per definizione alla funzione di distribuzione, e rappresentata dall'area tratteggiata in figura (per semplicità ci riferiamo ad una gaussiana a media nulla):

$F_X(x) = \Pr\{X \leq x\} = \int_{-\infty}^x p_X(\theta) d\theta = 1 - \int_x^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{\theta^2}{2\sigma_x^2}} d\theta$. Effettuiamo ora un cambio di variabile, ponendo $\frac{\theta^2}{2\sigma_x^2} = \eta^2$, per cui in corrispondenza di $\theta = x$ si ha $\eta = \frac{x}{\sqrt{2}\sigma_x}$, e risulta $d\theta = \sqrt{2}\sigma_x d\eta$. Con queste posizioni, possiamo riscrivere

$$F_X(x) = 1 - \int_{\frac{x}{\sqrt{2}\sigma_x}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\eta^2} \sqrt{2}\sigma_x d\eta = 1 - \int_{\frac{x}{\sqrt{2}\sigma_x}}^{\infty} \frac{1}{\sqrt{\pi}} e^{-\eta^2} d\eta = 1 - \frac{1}{2} \operatorname{erfc} \left\{ \frac{x}{\sqrt{2}\sigma_x} \right\}$$



Cosa è successo? Semplicemente, abbiamo espresso l'integrale (irrisolvibile in forma chiusa) nei termini della “funzione” $\operatorname{erfc}\{\cdot\}$, che rappresenta la probabilità che una v.a. gaussiana a media nulla e varianza $\frac{1}{2}$ superi (in valore assoluto) il valore dato come argomento, come mostrato in figura, e pari a

³⁰Disponendo di un insieme $\{x_n\}$ di N realizzazioni di una variabile aleatoria X , possiamo effettuare le seguenti stime: $\hat{m}_x = \frac{1}{N} \sum_{n=1}^N x_n$ e $\hat{m}_x^{(2)} = \frac{1}{N} \sum_{n=1}^N x_n^2$, il cui valore tende asintoticamente a quello delle rispettive medie di insieme, come N (la dimensione del campione statistico) tende a ∞ .

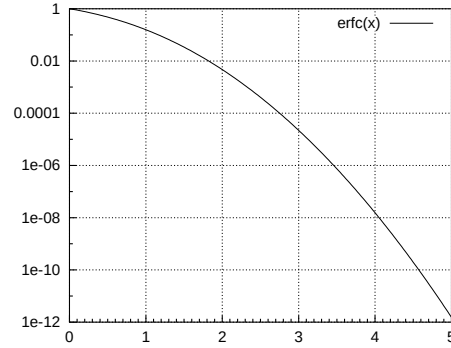
³¹Il suo scopritore, K.F. Gauss, denominò la v.a. e la sua ddp come *Normale*, indicando con questo il fatto che il suo uso potesse essere “quotidiano”.

$$\operatorname{erfc}\{\alpha\} = 2 \frac{1}{\sqrt{\pi}} \int_{\alpha}^{\infty} e^{-x^2} dx$$

I valori di erfc in funzione del suo argomento sono reperibili sia in forma di tabelle numeriche, che in forma di diagrammi quotati³².

In linea generale quindi, volendo calcolare la probabilità che una v.a. gaussiana X , con media m_x e varianza σ_x^2 , superi in ampiezza un determinato valore α , l'unica strada percorribile è quella di utilizzare la funzione erfc , avendo cura di porre come argomento il valore di α debitamente scalato, per ricondursi ad una gaussiana a media nulla e varianza $\frac{1}{2}$:

$$\Pr\{X > \alpha\} = \frac{1}{2} \operatorname{erfc}\left\{\frac{\alpha - m_X}{\sqrt{2}\sigma_x}\right\}$$

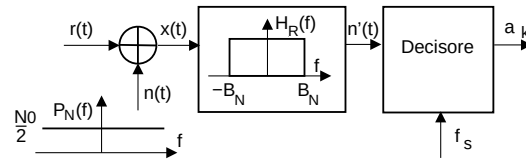


5.5.2 Probabilità di errore

Come anticipato, il decisore presente al lato ricevente di una trasmissione numerica opera su un segnale $x(t) = r(t) + n(t)$, ovvero il segnale ricevuto $r(t)$ è “corrotto” da un disturbo additivo $n(t)$ realizzazione di un processo stazionario ergodico gaussiano. Nel caso in cui siano presenti più cause di disturbo, anche localizzate in punti diversi del collegamento, si fa in modo di ricondurle tutte ad un'unica fonte di rumore (equivalente) in ingresso al decisore.

Più precisamente, il disturbo $n(t)$ è caratterizzato da un valor medio nullo, da uno spettro di densità di potenza *bianco* (ossia costante)

$$\mathcal{P}_N(f) = \frac{N_0}{2}$$



e da una varianza σ_N^2 pari alla potenza³³ $\mathcal{P}_N$ di una sua realizzazione qualsiasi (per l'ergodicità). Allo scopo di limitare $\mathcal{P}_N$ alla minima possibile, in ingresso al ricevitore è posto un filtro passa-basso ideale con risposta in frequenza $H_R(f)$ limitata in una banda $\pm B_N$ (detta *banda di rumore*, vedi Cap. 10), tale da lasciar passare il segnale $r(t)$ per intero, e limitare al tempo stesso la $\mathcal{P}_N(f)$ alla minore estensione possibile: pertanto, la potenza del rumore in uscita da $H_R(f)$ risulta (come vedremo al Capitolo 7) pari a

$$\mathcal{P}_{N'} = \int_{-\infty}^{\infty} \mathcal{P}_N(f) |H_R(f)|^2 df = \int_{-B_N}^{B_N} \frac{N_0}{2} df = N_0 B_N$$

Svolgiamo l'analisi indicando il segnale ricevuto nella forma

$$r(t) = \sum_k a[k] \cdot g(t - kT_s)$$

³²Il termine erfc sta per *funzione di errore complementare*, e trae origine dai risultati della misura di grandezze fisiche, in cui l'errore di misura, dipendente da cause molteplici, si assume appunto gaussiano.

³³In realtà $n(t)$ non ha $\mathcal{P}_N(f)$ costante per qualsiasi valore di f fino ad infinito, ma occupa una banda indeterminata ma limitata. Altrimenti, come vedremo al capitolo 7, avrebbe una potenza infinita.

in cui $g(t)$ è una caratteristica di Nyquist, in modo da evitare l'insorgenza di ISI³⁴; ed ogni simbolo $a[k]$ è un elemento di una sequenza aleatoria a valori discreti, pari questi ultimi agli L valori di ampiezza a_i , $i = 1, 2, \dots, L-1$, distribuiti entro un intervallo con dinamica pari a $a_L - a_1 = \Delta$. Agli istanti multipli del periodo di simbolo $t = kT_s = k/f_s$, il decisore esamina il valore del segnale $x(t) = r(t) + n(t)$, ed anzichè ritrovare i valori a_i trasmessi, osserva la realizzazione di una variabile aleatoria gaussiana, con media pari al valore a_i , e varianza $\sigma_N^2 = N_0 B_N$. Chiaramente, il ricevitore non conosce quale valore sia stato trasmesso in quell'istante, ed effettua una decisione di massima verosimiglianza confrontando tra loro le densità di probabilità condizionate alle diverse ipotesi a_i

$$P_{X/a_i}(x) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp \left\{ -\frac{(x - a_i)^2}{2\sigma_N^2} \right\} \quad (5.1)$$

e scegliendo per l' $\hat{a}_i$ tale che $P_{X/\hat{a}_i}(x)$ è la più grande.

Questo procedimento, nel caso in cui i valori a_i siano equispaziati, equivale (vedi figura seguente) a definire $L-1$ soglie di decisione θ_i , $i = 1, 2, \dots, L-1$, poste a metà tra i valori a_i ed a_{i+1} , e decidere per il valore a_i se il segnale ricevuto $x(kT_s)$ cade all'interno dell'intervallo compreso tra θ_{i-1} e θ_i (³⁵). La probabilità di errore che vogliamo determinare costituisce quindi la probabilità che $x(kT_s)$ oltrepassi una soglia di decisione, ovvero che un campione di rumore sia (in modulo) più grande di $\alpha = |\theta_i - a_i| = \frac{\Delta}{2(L-1)}$.

L'evento ora definito è un errore, condizionato alla trasmissione di a_i , la cui probabilità di verificarsi prende il nome di probabilità di errore *condizionata* e vale

$$P_{e/a_i} = 2 \int_{\theta_i}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_N} \exp \left\{ -\frac{(x - a_i)^2}{2\sigma_N^2} \right\} dx = P_\alpha$$

che chiameremo P_α . Lo stesso valore P_α è valido per tutti gli indici i compresi tra 2 ed $L-1$, mentre per a_1 ed a_L la probabilità di errore è dimezzata perché l'errore si verifica solo su di una soglia di decisione: $P_{e/a_1} = P_{e/a_L} = \frac{1}{2}P_\alpha$. Applicando il cambiamento di variabile illustrato nella sezione precedente, troviamo che $P_\alpha = \text{erfc} \left\{ \frac{\theta_i - a_i}{\sqrt{2}\sigma_N} \right\}$; sostituendo a $\theta_i - a_i$ la sua ampiezza espressa in termini della dinamica di segnale Δ , troviamo

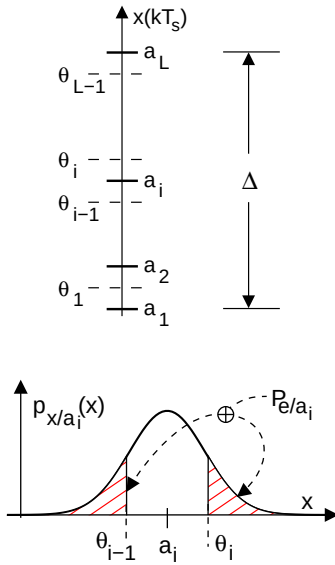
$$P_\alpha = \text{erfc} \left\{ \frac{\Delta}{2\sqrt{2}\sigma_N(L-1)} \right\} \quad (5.2)$$

Per arrivare all'espressione della probabilità di errore incondizionata³⁶, occorre eseguire una operazione di valore atteso rispetto a tutti gli indici i , con $i = 1, 2, \dots, L$, ovvero pesare le diverse probabilità di errore condizionate per le rispettive probabilità degli eventi condizionanti. Nel caso in cui i valori a_i siano *equiprobabili*, con probabilità $Pr(a_i) = \frac{1}{L}$,

³⁴Si tratta dell'*Inter Symbol Interference*, introdotta al Capitolo precedente.

³⁵Chiaramente, tutti i valori inferiori a θ_1 provocano la decisione a favore di a_1 , e quelli superiori di θ_{L-1} indicano la probabile trasmissione di a_L .

³⁶Che non dipende cioè da quale simbolo sia stato trasmesso.



si ottiene:

$$P_e = E_{a_i} \{P_{e/a_i}\} = \sum_{i=1}^L Pr(a_i) P_{e/a_i} = \frac{1}{L} \left[(L-2) P_\alpha + 2 \frac{1}{2} P_\alpha \right] = \left(1 - \frac{1}{L} \right) P_\alpha$$

in cui si è tenuto conto della diversa probabilità condizionata per i livelli interni e per i due agli estremi. Il risultato ottenuto, benchè già idoneo a valutare la P_e , può essere ulteriormente elaborato per ottenere espressioni più adatte ai progetti di dimensionamento.

5.5.3 Dipendenza di P_e da SNR

Allo scopo di rappresentare il risultato appena trovato in termini più fisici, esplicitiamo innanzitutto la potenza $\mathcal{P}_R$ del segnale utile $r(t)$. Si può dimostrare che sotto le ipotesi in cui:

- 1) si adotti un impulso di Nyquist a coseno rialzato con roll-off γ ;
- 2) i simboli $a[k]$ siano statisticamente indipendenti ed a valori a_i equiprobabili;
- 3) tali valori siano distribuiti uniformemente su L livelli, con dinamica $a_L - a_1 = \Delta$;

al § 7.9.3 si ottiene³⁷

$$\mathcal{P}_R = \frac{\Delta^2}{12} \frac{L+1}{L-1} \left(1 - \frac{\gamma}{4} \right)$$

Se in particolare risulta $L = 2$ e $\gamma = 0$ (come nel caso di trasmissione binaria a banda minima) allora si ha $\mathcal{P}_R = \frac{\Delta^2}{4}$.

Indichiamo ora con $y^2 = \frac{\Delta^2}{8\sigma_N^2(L-1)^2}$ il quadrato dell'argomento dell' $erfc$ in (5.2), ed esprimiamo Δ in funzione di $\mathcal{P}_R$. Sostituendo Δ in y^2 , e ricordando che $\sigma_N^2 = \mathcal{P}_N$, si ottiene $y^2 = \frac{3}{2} \frac{\mathcal{P}_R}{\mathcal{P}_N} \frac{1}{(L^2-1)(1-\frac{\gamma}{4})}$ e dunque

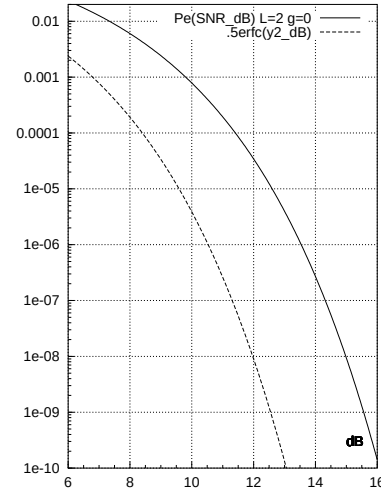
$$\begin{aligned} P_e &= \left(1 - \frac{1}{L} \right) \operatorname{erfc} \left\{ \sqrt{\frac{3}{2} SNR \frac{1}{(L^2-1)(1-\frac{\gamma}{4})}} \right\} \\ &= \left(1 - \frac{1}{L} \right) \operatorname{erfc} \{y\} \end{aligned}$$

Il termine sotto radice, che abbiamo indicato come y^2 , viene spesso espresso in dB:

$$y_{dB}^2 = 10 \log_{10} y^2 = SNR_{dB} + 1.76 - 10 \log_{10} (L^2 - 1) - 10 \log_{10} \left(1 - \frac{\gamma}{4} \right)$$

che costituisce una importante formula di progetto³⁸, in quanto permette l'uso diretto del rapporto segnale rumore, già noto come indice della "bontà" del collegamento; infatti all'aumentare di SNR , aumenta anche y^2 , e di conseguenza si riduce P_e .

La comodità di utilizzare direttamente grandezze in dB è tale da rendere disponibili curve di $erfc$ per le quali l'argomento è espresso in unità logaritmiche, come riportato nella figura precedente, in cui è graficato il valore di P_e in funzione sia di y_{dB}^2 che di



³⁷Anche se il risultato sarà dimostrato al § 7.9.3, merita comunque un commento: osserviamo che $\mathcal{P}_R$ diminuisce all'aumentare di γ (si stringe infatti l'impulso nel tempo); inoltre $\mathcal{P}_R$ diminuisce al crescere di L , in quanto nel caso di più di 2 livelli, la forma d'onda assume valori molto vari all'interno della dinamica di segnale, mentre con $L = 2$ ha valori molto più "estremi".

³⁸Il misterioso termine 1.76 non è altro che $10 \log_{10} \frac{3}{2}$.

SNR_{dB} (nel caso $L = 2$ e $\gamma = 0$). Possiamo notare che valori dell'ordine di $P_e \simeq 10^{-5}$ si ottengono già per SNR_{dB} inferiore a 13 dB, decisamente scarsi per una trasmissione analogica.

Volendo utilizzare il grafico riportato per valori di L e γ diversi da 2 e 0, nel caso in cui SNR_{dB} non variasse, occorre calcolare la penalizzazione aggiuntiva al valore di y_{dB}^2 , equivalente ad una riduzione di SNR di pari entità. In corrispondenza del nuovo valore in ascissa per y_{dB}^2 , si trova un nuovo (peggior) valore di P_e a partire da quello fornito dalla curva, moltiplicandolo per il fattore $2 \left(1 - \frac{1}{L}\right)$, che tiene conto del diverso numero di livelli.

5.5.4 Dipendenza di P_e da E_b/N_0

L' SNR che appare nell'argomento dell'*erfc* in realtà *non* è una variabile indipendente del collegamento, in quanto la potenza di rumore $\mathcal{P}_N = N_0 B_N$, da cui SNR dipende, a sua volta è funzione dalla banda del filtro di ricezione, che come abbiamo visto è posta pari alla massima frequenza presente in $r(t)$, e (nel consueto caso di $g(t)$ a coseno rialzato) pari a

$$B_N = \frac{f_s}{2} (1 + \gamma) = \frac{f_b}{2 \log_2 L} (1 + \gamma)$$

dependendo quindi anch'essa da L e γ , oltre che dalla f_b . Pertanto, al variare di L e γ , varia anche SNR , complicando il ragionamento esposto al termine della precedente sezione.

Allo scopo di mantenere separati tra loro i diversi elementi che determinano la probabilità di errore, e di porre nella giusta evidenza i parametri di sistema f_b ed N_0 , esprimiamo le potenze $\mathcal{P}_N$ e $\mathcal{P}_R$ in funzione di $T_b = 1/f_b$:

$$\mathcal{P}_N = N_0 B_N = \frac{N_0 (1 + \gamma)}{T_b 2 \log_2 L} \quad \text{e} \quad \mathcal{P}_R = \frac{E_b}{T_b} \quad (5.3)$$

in modo da ottenere

$$SNR = \frac{\mathcal{P}_R}{\mathcal{P}_N} = \frac{E_b T_b 2 \log_2 L}{T_b N_0 (1 + \gamma)} = \frac{E_b 2 \log_2 L}{N_0 (1 + \gamma)}$$

La grandezza E_b rappresenta l'*energia per bit*³⁹ e la sua definizione

$$E_b = \mathcal{P}_R T_b = \frac{\mathcal{P}_R}{f_b}$$

mostra come essa riassume in sé i parametri di sistema *potenza di segnale* e *velocità binaria*, mentre invece non dipende dai *parametri di trasmissione* L e γ . Anche N_0 , inoltre, costituisce un parametro di sistema, rappresentando una grandezza su cui non è possibile intervenire.

Adottando le posizioni esposte, il quadrato dell'argomento dell'*erfc* che dipende da SNR può ora essere espresso come:

$$y^2 = \frac{3}{2} SNR \frac{1}{(L^2 - 1) \left(1 - \frac{\gamma}{4}\right)} = \frac{E_b}{N_0} \frac{3 \log_2 L}{(L^2 - 1) (1 + \gamma) \left(1 - \frac{\gamma}{4}\right)}$$

³⁹ si rifletta sulla circostanza che la potenza è una energia per unità di tempo.

Come la precedente, anche questa relazione può ben essere utilizzata per valutare il valore dell'argomento dell' $erfc$ ed usare quindi le curve per determinare la probabilità di errore. In particolare, notiamo che per $L = 2$ e $\gamma = 0$ si ottiene: $P_e = \frac{1}{2}erfc\left\{\sqrt{\frac{E_b}{N_0}}\right\}$, che assume il ruolo di valore di riferimento: infatti, il termine sotto radice contraddistingue le condizioni operative della trasmissione (racchiudendo potenza ricevuta, frequenza binaria e livello di rumore). Le scelte progettuali (roll-off e numero di livelli) diverse da $L = 2$ e $\gamma = 0$ determinano immancabilmente un peggioramento della P_e , ma vengono intraprese per soddisfare esigenze di risparmio di banda (aumentando L)⁴⁰, e per ridurre i termini di interferenza intersimbolica (aumentando γ). In particolare, la compensazione (con f_b fissa) del peggioramento di P_e dovuto ad uno (o ad entrambi) questi fattori, può avvenire solamente grazie ad un aumento di E_b , ovvero un aumento della potenza trasmessa: tale circostanza è nota come *compromesso banda-potenza*⁴¹.

Due domande riassuntive:

1. Perché P_e peggiora se aumento i livelli ? Risposta (⁴²).
2. Perché P_e peggiora se aumento γ ? Risposta (⁴³).

5.5.5 Diagramma ad occhio

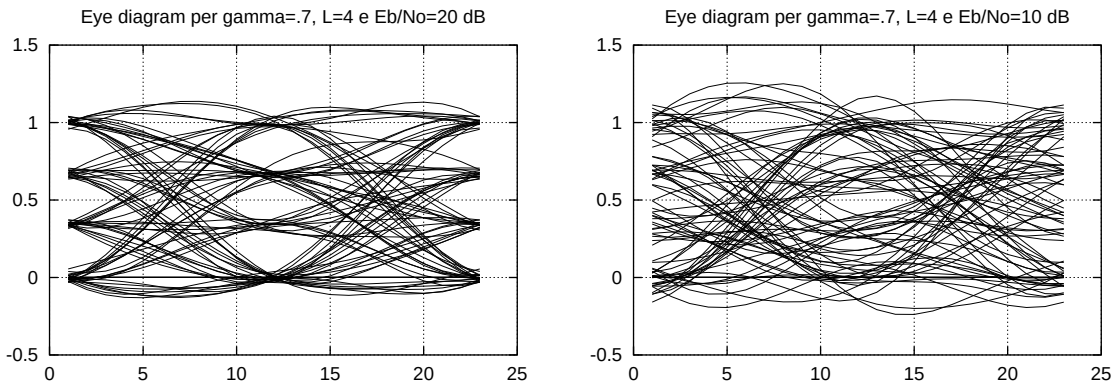


Figura 5.3: Diagramma ad occhio per segnale rumoroso con E_b/N_0 pari a 20 e 10 dB, $\gamma = .7$, $L = 4$

⁴⁰Aumentando L , y^2 diminuisce, in quanto $(L^2 - 1)$ cresce più velocemente di $\log_2 L$.

⁴¹Nel caso in cui vi siano limitazioni di potenza, come ad esempio nelle comunicazioni satellitari, conviene occupare la maggior banda possibile, mantenendo $L = 2$, in modo da risparmiare potenza.

Il teorico dell'informazione Claude Shannon ha trovato il risultato che è possibile trasmettere senza errori (ricorrendo a codifiche di canale sempre più sofisticate) purchè la velocità di trasmissione f_b non ecceda la *Capacità di Canale*, definita come $C = W \log_2 \left(1 + \frac{P_R}{N_0 W}\right)$, in cui W è la banda del canale, ed $N_0 W$ la potenza del rumore. Un secondo canale, con W maggiore, dispone di una Capacità maggiore, in quanto $\log_2(\cdot)$ cresce più lentamente di W ; pertanto, per mantenere la stessa capacità, è sufficiente trasmettere con una potenza di segnale P_R ridotta.

⁴²Perché a parità di P_R gli intervalli di decisione sono più ravvicinati, e le "code" della gaussiana sottendono un'area maggiore.

⁴³Perché occorre aumentare la banda del filtro di ricezione e dunque far entrare più rumore. D'altra parte questo peggioramento è compensato dalla riduzione dell'ISI.

Si tratta dello stesso tipo di grafico già descritto a pag. 57, e che ora ci aiuta a valutare in modo visivo la qualità di una trasmissione numerica. In fig. 5.3 sono riportati i grafici per un segnale dati a 4 livelli, in presenza di due diverse potenze di rumore: notiamo che al peggiorare del rapporto $\frac{E_b}{N_0}$, la zona priva di traiettorie (*l'occhio*) riduce la sua estensione verticale (*tende a chiudersi*). Pertanto, disponendo di un segnale numerico di qualità sconosciuta, questa può essere valutata in modo approssimato, qualora si disponga di un oscilloscopio, esaminando il grado di apertura dell'*occhio*.

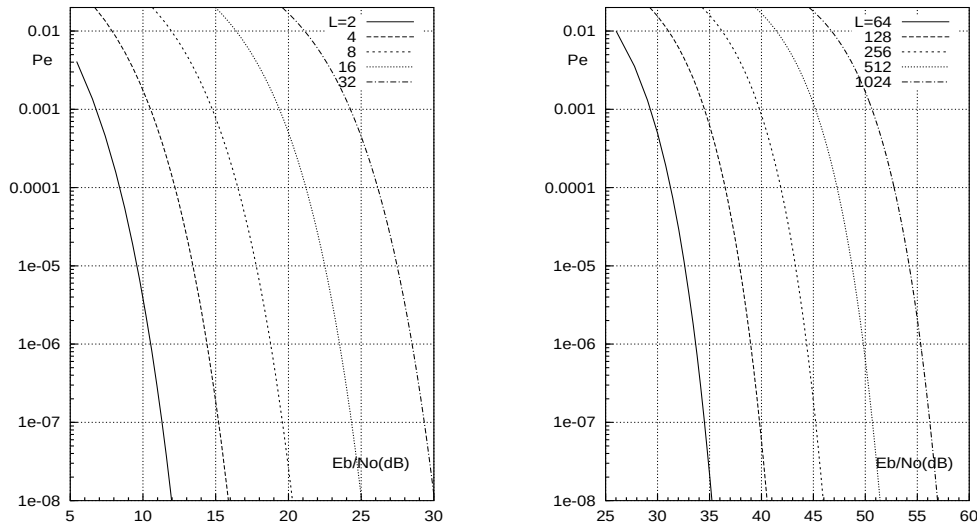
5.5.6 Uso del codice di Gray

Le probabilità di errore ricavate caratterizzano l'evento di decidere per la ricezione di a_i quando invece è stato trasmesso a_{i-1} o a_{i+1} ⁽⁴⁴⁾. Abbiamo già discusso in precedenza come esista un modo particolare (il codice di Gray) di assegnare ai simboli a_i delle codifiche binarie, in modo da associare ai simboli relativi a valori (livelli) contigui, parole differenti per un solo bit.

In presenza della codifica di Gray, pertanto, ogni simbolo errato contiene un solo bit errato, e quindi l'evento di errore *sul bit* si verifica quando il simbolo a cui appartiene è errato, e il bit è quello errato, ovvero: $Pr\{bit\ errato\} = Pr\{simbolo\ errato\} \cdot Pr\{bit\ errato / simbolo\ errato\} = P_e \cdot \frac{1}{\log_2 L}$. Ad esempio, con $L = 256$ livelli, la P_e sul bit si riduce di $\log_2 L = 8$ volte. L'espressione della P_e per bit nel caso si adotti una codifica di Gray, diviene quindi:

$$P_e^{bit} = \frac{1}{\log_2 L} \left(1 - \frac{1}{L}\right) \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0} \frac{3 \log_2 L}{(L^2 - 1)(1 + \gamma)(1 - \frac{\gamma}{4})}} \right\} \quad (5.4)$$

Le curve seguenti mostrano il valore di P_e^{bit} così determinato, in funzione di $\frac{E_b}{N_0}$ espresso in dB, per diversi valori di L . Valori di $\gamma \neq 0$ determinano un peggioramento⁴⁵ di $10 \log_{10} (1 + \gamma)(1 - \frac{\gamma}{4})$, che deve essere compensato da un eguale incremento in dB di $\frac{E_b}{N_0}$, per ottenere la stessa P_e .



Probabilità di errore sul bit per trasmissione multilivello di banda base con codifica di Gray

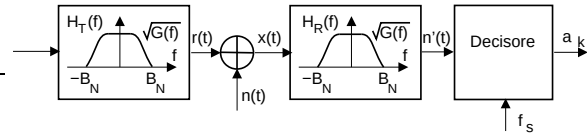
⁴⁴La probabilità di un errore legato al salto di due o più livelli θ è così piccola da potersi trascurare.

⁴⁵Di non grande entità: per $\gamma = 1$ il peggioramento risulta di 1.761 dB.

5.5.7 Ricevitore ottimo

Come sarà illustrato in fondo al cap. 7, per massimizzare l' SNR in ricezione, occorre impiegare un filtro adattato alla forma dell'impulso trasmesso $g(t)$, per il quale risulti $H(f) = G^*(f)$. Nello schema adottato per la figura a pag. 93, il filtro di ricezione possiede invece il solo scopo di limitare la banda del rumore, ed è sempre un passa-basso ideale, indipendente da $g(t)$. Allora, se si adotta una $G(f)$ di Nyquist *non* a banda minima, i campioni di rumore che, sovrapposti a quelli di segnale, danno luogo alle v.a. $x(kT_L)$ con d.d.p (5.1), non sono più indipendenti⁴⁶. Per questo motivo, sarebbe in linea di principio possibile realizzare un dispositivo *predittore lineare* che, in base alla conoscenza dei precedenti valori di rumore⁴⁷, calcola una stima del valore corrente che, sottratta al valore effettivamente ricevuto, consente di ridurre la varianza della grandezza di osservazione, consentendo una riduzione della probabilità di errore.

Allo scopo di realizzare comunque le condizioni di Nyquist in ricezione, e contemporaneamente rendere incorrelati i campioni di rumore, la caratteristica di Nyquist viene decomposta in parti uguali tra



trasmettitore e ricevitore, dando luogo allo schema di figura, in cui $H_T(f) = H_R(f) = \sqrt{G(f)}$. In tal modo, al decisore giunge esattamente lo stesso segnale di prima, mentre la densità di potenza del rumore non è più costante, ma ora vale $\mathcal{P}_N(f) = \frac{N_0}{2} |H_R(f)|^2 = \frac{N_0}{2} G(f)$. Pertanto, i campioni di rumore presi a distanza T_s sono ora incorrelati, in quanto $\mathcal{R}_N(\tau) = \mathcal{F}^{-1}\{\mathcal{P}_N(f)\}$ è ora un impulso di Nyquist, che passa da zero per $\tau = kT_s$. Notiamo che, essendo $G(f)$ reale pari, la scomposizione $H_T(f) = H_R(f) = \sqrt{G(f)}$ realizza la condizione $H_R(f) = H_T^*(f)$ che definisce un filtro adattato.

Per determinare le nuove prestazioni, notiamo che mentre la banda passante di $H_R(f)$ (e dunque del rumore) si è mantenuta pari a $B_N = \frac{f_s}{2}(1 + \gamma)$, la potenza del rumore ora vale⁴⁸ $\mathcal{P}_N = \int_{-\infty}^{\infty} \frac{N_0}{2} G(f) df = N_0 f_s = \frac{N_0}{T_b 2 \log_2 L}$, riducendosi di un fattore $(1 + \gamma)$. Questo valore, sostituito nella (5.3), causa una riduzione di $(1 + \gamma)$ nell' SNR , e la stessa quantità è rimossa di conseguenza anche nella (5.4), portando a

$$P_e^{bit} = \frac{1}{\log_2 L} \left(1 - \frac{1}{L}\right) \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0} \frac{3 \log_2 L}{(L^2 - 1) \left(1 - \frac{\gamma}{4}\right)}} \right\}$$

il valore della probabilità di errore sul bit adottando il ricevitore ottimo ed il codice di Gray.

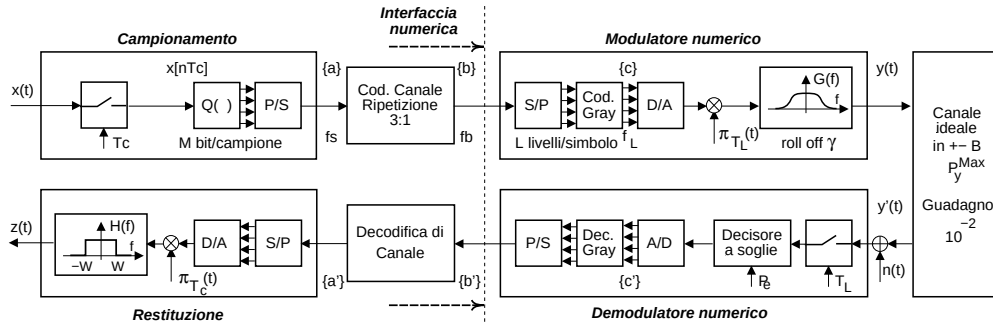
⁴⁶Infatti, il segnale $n(t)$ uscente da $H(f) = \operatorname{rect}_{2B}(f)$ possiede autocorrelazione $\mathcal{R}_N(\tau) = \mathcal{F}^{-1}\{|H(f)|^2\} = 2B \operatorname{sinc}(2B\tau)$ (vedi cap. 7), che passa da zero per $\tau = \frac{1}{2B}$. Se si utilizza una $G(f)$ a coseno rialzato con $\gamma > 0$, occorre una banda di rumore $B_N = \frac{f_s}{2}(1 + \gamma)$, a cui corrisponde l'incorrelazione tra campioni di rumore prelevati a distanza multipla di $\frac{1}{2B} = \frac{1}{f_s(1+\gamma)}$. Invece, il rumore è campionato con frequenza pari a quella di simbolo f_s , e dunque con campioni a distanza $T_s = \frac{1}{f_s}$. Pertanto, i campioni di rumore sono correlati, con autocorrelazione pari a $\mathcal{R}_N(T_s) = 2B \operatorname{sinc}(1 + \gamma)$.

⁴⁷Non entriamo qui nei dettagli dei metodi di predizione lineare, ma notiamo semplicemente che la *correlazione* tra grandezze aleatorie ne caratterizza la *dipendenza*, e dunque la conoscenza dei valori passati consente di ridurre l'incertezza relativa ai nuovi valori. I valori dei precedenti campioni di rumore, sono calcolabili a partire da quelli dei simboli *decisi* senza commettere errore, sottratti al valore del segnale ricevuto in corrispondenza dei precedenti simboli.

⁴⁸Il risultato si ottiene visivamente, calcolando l'area di $G(f)$ a coseno rialzato, e considerando le sue proprietà di simmetria attorno ad f_s .

5.5.8 Esercizio

Un sistema di trasmissione basato sul campionamento e sulla trasmissione numerica è rappresentato in figura.



Il canale riportato all'estremità destra è considerato ideale entro una banda $\pm B = \pm 31.5$ KHz, purchè la potenza al suo ingresso non superi il valore $\mathcal{P}_y^{Max} = 1$ Volt²; in tal caso la potenza in uscita risulta $\mathcal{P}_{y'} = 0.01 \cdot \mathcal{P}_y$. Al segnale ricevuto è sovrapposto un rumore additivo gaussiano bianco stazionario ergodico a media nulla, con spettro di densità di potenza $\mathcal{P}_N(f) = \frac{N_0}{2} = 4.61 \cdot 10^{-14}$ Volt²/Hz, e limitato nella banda $\pm B$.

- 1) Se $G(f)$ è a coseno rialzato con $\gamma = .5$, determinare la massima frequenza di simbolo $f_L = \frac{1}{T_L}$.
- 2) Ponendo che $\mathcal{P}_y = \mathcal{P}_y^{Max}$, determinare l' SNR in ingresso al demodulatore numerico.
- 3) Desiderando una $P_e = P_e^c$ per la sequenza $\{c'\}$ pari a $P_e = 10^{-4}$, determinare il massimo numero di livelli/simbolo L .
- 4) Indicare la frequenza binaria f_b per la sequenza $\{b'\}$.
- 5) Valutare P_e^b per la sequenza $\{b'\}$ e mostrare che il numero di errori per unità di tempo in $\{b'\}$ è lo stesso che in $\{c'\}$.
- 6) Mostrare che, adottando una codifica di canale a ripetizione 3 : 1, la probabilità di errore P_e^a per la sequenza $\{a'\}$ risulta pari a circa $P_e^a \simeq 3 (P_e^b)^2$.
- 7) Indicare la frequenza binaria f_a per le sequenze $\{a\}$ ed $\{a'\}$.

Supponiamo ora che $P_e^a = 0$, e si desideri un $SNR = \mathcal{P}_x / \mathcal{P}_{z-x} = 10000$. Nel caso in cui $x(t)$ sia un processo con densità di probabilità $p(x)$ uniforme, ed indicando con W la banda di $x(t)$;

- 8) Determinare il minimo numero di bit/campione M .
- 9) Determinare la massima banda W .
- 10) Se la banda è ridotta a $W' = \frac{1}{2}W$, determinare il nuovo valore di SNR ottenibile.

Soluzione

- 1) La banda B occupata dal segnale y vale $B = \frac{f_L}{2} (1 + \gamma)$, e quindi deve risultare $f_L = \frac{2B}{1+\gamma} = \frac{2 \cdot 31.5 \cdot 10^3}{1.5} = 42 \cdot 10^3 = 42.000$ baud (baud = simboli/secondo).
- 2) Per definizione $SNR = \frac{\mathcal{P}_{y'}}{\mathcal{P}_N}$ in cui $\mathcal{P}_{y'} = \mathcal{P}_y^{Max} \cdot 0.01$ e $\mathcal{P}_N = \frac{N_0}{2} 2B = N_0 B$ e quindi $SNR = \frac{0.01 \cdot \mathcal{P}_y^{Max}}{N_0 B} = \frac{0.01}{2 \cdot 4.61 \cdot 10^{-14} \cdot 31.5 \cdot 10^3} = 3.44 \cdot 10^6$.

- 3) Sappiamo che, in presenza di rumore gaussiano, risulta $P_e = \left(1 - \frac{1}{L}\right) \operatorname{erfc}\{y\}$ e quindi, desiderando un valore di P_e , occorre che risulti $\frac{1}{2} \operatorname{erfc}\{y\} = \frac{1}{2} P_e \left(1 - \frac{1}{L}\right)^{-1} \simeq \frac{1}{2} P_e$ in quanto $\frac{1}{L} \ll 1$ se $L \gg 1$ (cosa che verificheremo successivamente). Dalla curva di pag. 96 si trova che ad un valore di $\frac{1}{2} \operatorname{erfc}\{y\} = 5 \cdot 10^{-4}$ corrisponde un $y_{dB}^2 \simeq 7.5$. Si applica ora la formula di progetto che mette in relazione y con i parametri del sistema:

$$y_{dB}^2 = 10 \log_{10} y^2 = SNR_{dB} + 1.76 - 10 \log_{10} (L^2 - 1) - 10 \log_{10} \left(1 - \frac{\gamma}{4}\right)$$

da cui a questo punto è facile ottenere

$$\begin{aligned} 10 \log_{10} (L^2 - 1) &= -y_{dB}^2 + SNR_{dB} + 1.76 - 10 \log_{10} \left(1 - \frac{\gamma}{4}\right) \\ &= -7.5 + 10 \log_{10} 3.44 \cdot 10^6 + 1.76 - 10 \log_{10} 0.875 \\ &= -7.5 + 65.36 + 1.76 + 0.58 = 60.2 \end{aligned}$$

e quindi $L^2 - 1 = 10^{6.02}$, ossia $L = \sqrt{1 + 10^{6.02}} = 1023$ che viene arrotondato a 1024 livelli, che è una potenza di due.

- 4) Dato che ad ogni simbolo di $\{c\}$ ad L livelli, con frequenza di emissione pari a f_L , corrisponde ad un gruppo di $N_b = \log_2 L = 10$ bit della sequenza $\{b\}$, la frequenza f_b è di 10 volte f_L , e quindi $f_b = 10 \cdot f_L = 10 \cdot 42 \cdot 10^3 = 420$ Kbps.
- 5) Grazie all'adozione del codice di Gray, in caso di errore tra livelli contigui per i simboli di $\{c'\}$, nella sequenza $\{b'\}$ solo uno (tra N_b) dei bit associati ad un simbolo è errato; il bit errato è uno qualsiasi del gruppo di N_b , e pertanto la probabilità che un bit specifico sia errato (quando è errato il simbolo di $\{c'\}$) è $\frac{1}{N_b}$. Pertanto $P_e^b = P_e^{b/c} P_e^c = \frac{1}{N_b} P_e^c$, in cui $P_e^{b/c}$ è la probabilità condizionata che un generico bit di $\{b'\}$ sia sbagliato quando è sbagliato il simbolo di $\{c'\}$ da cui ha origine.

- Il numero di bit (della sequenza $\{b'\}$) errati per unità di tempo è dato da $P_e^b \cdot f_b$; sostituendo: $P_e^b \cdot f_b = \frac{P_e^c}{N_b} \cdot f_b = P_e^c \cdot \frac{f_b}{N_b} = P_e^c \cdot f_L$, ovvero è numericamente pari ai simboli errati (nella sequenza $\{c'\}$) per unità di tempo;
- risulta dunque infine:
 - $P_e^b = \frac{P_e^c}{N_b} = \frac{10^{-4}}{10} = 10^{-5}$;
 - $P_e^b \cdot f_b = P_e^c \cdot f_L = 10^{-5} \cdot 420 \cdot 10^3 = 10^{-4} \cdot 42 \cdot 10^3 = 4.2 \frac{\text{errori}}{\text{secondo}}$

- 6) Ogni bit di $\{a'\}$ è sbagliato solo se sono sbagliati 2 o più bit in un gruppo di 3; come mostrato al capitolo seguente, la probabilità di 2 bit errati su 3 è calcolabile dalla distribuzione di Bernoulli, e vale $\binom{3}{2} p_e^2 (1 - p_e) = 3 p_e^2 (1 - p_e)$, a cui va sommata la probabilità di 3 bit errati, pari a p_e^3 . Pertanto $p_e^a = 3 p_e^2 (1 - p_e) + p_e^3 = 3 p_e^2 - 3 p_e^3 + p_e^3 \simeq 3 p_e^2$ in cui ovviamente $p_e = P_e^b$, e l'approssimazione è legittima in quanto se $p_e = 10^{-5}$ allora $p_e^2 = 10^{-10}$ e $p_e^3 = 10^{-15}$, trascurabili rispetto a p_e . Lo stesso risultato si ottiene osservando che 2 bit errati su 3 hanno probabilità $p_e^2 (1 - p_e)$, e questi possono essere scelti in tre modi diversi (1° e 2° , 1° e 3° , 2° e 3°). In definitiva, risulta $P_e^a \simeq 3 (P_e^b)^2 = 3 \cdot 10^{-10}$.

- 7) Dato che ad ogni 3 bit di $\{b'\}$ corrisponde un solo bit di $\{a'\}$, si ottiene $f_a = \frac{f_b}{3} = \frac{420 \cdot 10^3}{3} = 140$ Kbps, a cui corrisponde $P_e^a \cdot f_a = 3 \cdot 10^{-10} \cdot 140 \cdot 10^3 = 4.2 \cdot 10^{-5} \frac{\text{errori}}{\text{secondo}}$.

- 8) Sappiamo che per un processo uniforme l' SNR di quantizzazione risulta approssimativamente $SNR_q = (L - 1)^2$, in cui L è il numero di livelli del quantizzatore, a cui corrisponde l'utilizzo di $M = \log_2 L$ bit/campione. Risulta pertanto $L = 1 + \sqrt{SNR_q} = 1 + \sqrt{10^4} = 101$ livelli. Per ottenere un numero intero di bit/campione ed un SNR_q migliore od uguale a quello desiderato, determiniamo l'intero superiore: $M = \lceil \log_2 L \rceil = 7$ bit/campione (equivalente a 128 livelli).

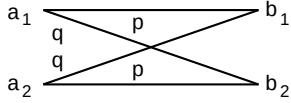
- 9) Come sappiamo, la frequenza di campionamento $f_c = \frac{1}{T_c}$ non può essere inferiore a $2W$; inoltre, la frequenza binaria f_a risulta pari al prodotto dei bit/campione per i campioni a secondo: $f_a = f_c \cdot M$; pertanto $f_c = \frac{f_a}{M} = \frac{140 \cdot 10^3}{7} = 20$ KHz e dunque la W massima risulta $W_{max} = \frac{f_c}{2} = 10$ KHz.

- 10) Nel caso in cui $W' = \frac{1}{2}W$, allora si può dimezzare anche la frequenza di campionamento $f'_c = \frac{f_c}{2} = 10$ KHz, e pertanto utilizzare un $M' = 2M$ per ottenere la stessa f_a . Pertanto il nuovo SNR_q risulta $SNR'_q = (L' - 1)^2 = (2^{2M} - 1)^2 = (2^{14} - 1)^2 \simeq 2.68 \cdot 10^8$, ovvero SNR'_q (dB) = 84.3 dB.

5.6 Appendici

5.6.1 Canale simmetrico binario

In figura è rappresentato uno schema, che rappresenta un canale numerico, al cui ingresso si presenta uno tra due simboli a_1 e a_2 , mentre alla sua uscita si osservano i simboli b_1 e b_2 . Il canale è caratterizzato da due probabilità condizionate *in avanti*: una probabilità di errore $P_e = Pr(b_2/a_1) = Pr(b_1/a_2)$ che indichiamo con q , ed una probabilità (complementare) di non-errore $1 - P_e = Pr(b_1/a_1) = Pr(b_2/a_2)$ indicata con $p = 1 - q$.



Rapporto di verosimiglianza Qualora si osservi in uscita uno dei due valori (ad es. b_1), si possono confrontare le probabilità a posteriori per le due possibili *ipotesi* che in ingresso fosse presente a_1 od a_2 , valutando il rapporto

$$RV_{ML} = \frac{Pr(b_1/a_1)}{Pr(b_1/a_2)} = \frac{p}{q} \quad (5.5)$$

e quindi decidere per l'ipotesi *più verosimile* (es. si decide per a_1 se $p > q$, ovvero se $P_e < \frac{1}{2}$).

Se disponiamo della conoscenza delle probabilità a priori $Pr(a_1)$ e $Pr(a_2)$, ed i due simboli a_1 ed a_2 non sono equiprobabili⁴⁹, possiamo costruire il *rapporto di verosimiglianza* utilizzando le probabilità a posteriori $Pr(a_1/b_1)$ e $Pr(a_2/b_1)$, calcolabili applicando il teorema di Bayes. Mantenendo il caso di aver ricevuto il simbolo b_1 , il *rapporto di verosimiglianza* si scrive ora come

$$RV_B = \frac{Pr(a_1/b_1)}{Pr(a_2/b_1)} = \frac{Pr(b_1/a_1) Pr(a_1)}{Pr(b_1/a_2) Pr(a_2)} \quad (5.6)$$

RV può assumere valore $>$, $<$ od $=$ ad 1, a seconda di quale delle due probabilità a posteriori sia più grande, portando la decisione a favore di a_1 , a_2 , o l'indifferenza - rispettivamente.

Verifica di Ipotesi di ML e Bayesiana La metodologia ora descritta prende il nome di *verifica di ipotesi* statistica e si basa appunto sul confronto di quanto la grandezza osservata sia *verosimile*, compatibilmente con le *ipotesi* possibili. Se RV utilizza solamente le probabilità in avanti (5.5), la decisione si dice di *massima verosimiglianza* (indicata con *ML* o MAXIMUM LIKELIHOOD), mentre se si impiegano le probabilità a posteriori (5.6), si sta effettuando una *decisione bayesiana*.

⁴⁹In caso contrario, la 5.6 è equivalente alla 5.5. Pertanto, nei casi in cui non si conosca la *statistica di sorgente*, è sufficiente attuare una decisione di massima verosimiglianza.

Riflessioni Il meccanismo con cui, nella decisione Bayesiana, le probabilità in avanti si combinano con quelle a priori, può essere analizzato mediante alcune osservazioni: innanzi tutto, a_1 potrebbe essere così “raro” che, in presenza di una moderata probabilità di errore, il ricevitore potrebbe preferire di decidere sempre a_2 , attribuendo l’eventuale ricezione di b_1 dovuta più verosimilmente ad un errore del canale, piuttosto che all’effettiva estrazione di a_1 . In assenza di canale inoltre, l’unico rapporto di verosimiglianza possibile sarebbe stato quello tra le probabilità a priori $Pr(a_1)$ e $Pr(a_2)$; la ricezione di un simbolo b_i dunque ha portato un miglior livello di informazione, alterando il RV , in misura tanto maggiore quanto più bassa è la probabilità di errore.

Applicazione Verifichiamo i ragionamenti appena svolti riscrivendo per esteso una probabilità a posteriori:

$$Pr(a_1/b_1) = \frac{Pr(b_1/a_1) Pr(a_1)}{Pr(b_1/a_1) Pr(a_1) + Pr(b_1/a_2) Pr(a_2)} = \frac{p \cdot Pr(a_1)}{p \cdot Pr(a_1) + q \cdot Pr(a_2)}$$

Se $p = q$, il canale è inservibile e non aggiunge informazione: infatti si ottiene $Pr(a_1/b_1) = Pr(a_1)$ in quanto $Pr(a_1) + Pr(a_2) = 1$. D’altra parte, se $q < p$, risulta

$$Pr(a_1/b_1) = \frac{Pr(a_1)}{Pr(a_1) + \frac{q}{p}Pr(a_2)} > Pr(a_1)$$

aumentando quindi la probabilità di a_1 rispetto a quanto previsto dalla probabilità a priori; se la probabilità di errore tende a zero ($q \rightarrow 0$), osserviamo che $Pr(a_1/b_1) \rightarrow 1$.

5.6.2 Entropia e Codifica di Sorgente

Prendiamo in considerazione una sorgente discreta, che emetta simboli x_k appartenenti ad un alfabeto di cardinalità L (ossia con $k = \{1, 2, \dots, L\}$), con probabilità $p_k = Pr(k)$. Definiamo l’informazione associata all’osservazione del simbolo x_k come⁵⁰

$$I_k = I(x_k) = \log_2 \frac{1}{p_k} = -\log_2 p_k$$

La scelta fatta, conduce ai seguenti risultati:

Prob. p_k	Informazione $-\log_2 p_k$	Commento
1	0	L’evento certo non fornisce informazione
0	∞	L’evento impossibile dà informazione infinita
$\frac{1}{2}$	1	In caso di scelta binaria (es. testa o croce) occorre una cifra binaria (<i>bit = binary digit</i>) per indicare il risultato

L’informazione legata ad un evento è anche rappresentativa dell’incertezza a suo riguardo, prima che questo si verifichi.

Passiamo ora a definire *Entropia* (indicata con H) di una sorgente discreta S , il valore atteso dell’informazione apportata dalla conoscenza dei simboli da essa generati:

$$H_S = E\{I_k\} = \sum_{k=1}^L p_k I_k = \sum_{k=1}^L p_k \log_2 \frac{1}{p_k}$$

⁵⁰Per calcolare il logaritmo in base 2, si ricordi che $\log_2 \alpha = \frac{\log_{10} \alpha}{\log_{10} 2} \simeq 3.32 \log_{10} \alpha$.

che costituisce una interessante applicazione dell'operatore di *valore atteso* da poco introdotto. H rappresenta l'*informazione media* di un generico simbolo emesso dalla sorgente, ottenuta pesando l'informazione di ogni possibile simbolo con la probabilità del simbolo stesso. Osserviamo ora che:

- Se i simboli sono equiprobabili ($p_k = \frac{1}{L}$ con $\forall k$), la sorgente è *massimamente informativa*, e la sua entropia è la massima possibile per un alfabeto ad L simboli, e cioè $H_{S_{Max}} = \frac{1}{L} \sum_{k=1}^L \log_2 L = \log_2 L$.
- Se i simboli non sono equiprobabili, allora $H_S < \log_2 L$.
- Se la sorgente emette sempre e solo lo stesso simbolo, allora $H_S = 0$. Questa circostanza, assieme alla precedente, consente di scrivere che $0 \leq H_S \leq \log_2 L$.

Un caso particolare è quello delle sorgenti binarie, che producono uno tra due simboli $\{x_0, x_1\}$ con probabilità rispettivamente $\{Pr(x_0) = p, Pr(x_1) = q = 1 - p\}$, che inserite nella formula dell'Entropia, forniscono l'espressione

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

il cui andamento è mostrato nella figura seguente, in funzione di p .

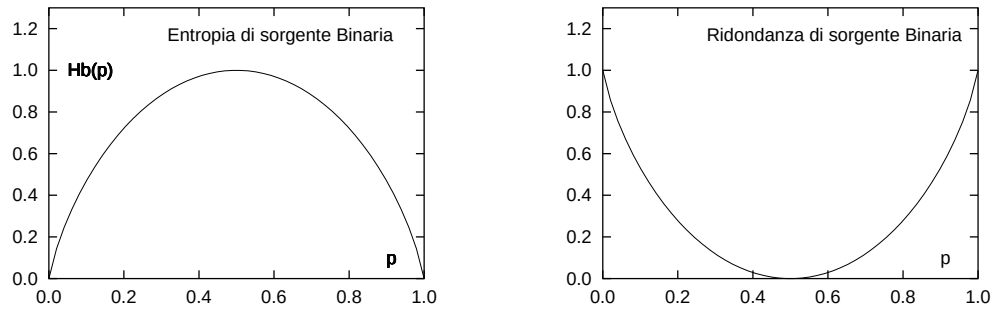


Figura 5.4: Entropia di sorgente binaria, e ridondanza associata

I due simboli $\{x_0, x_1\}$ possono essere rappresentati dalle 2 cifre binarie $\{0, 1\}$, che in questo caso chiamiamo *binit*, per non confonderli con la misura dell'informazione (il bit), ed allora se $p \neq .5$, usiamo 1 binit/simbolo per rappresentare un messaggio con solo $H_b < 1$ bit/simbolo di informazione, dunque la nostra codifica ha una ridondanza pari a $1 - H_b(p)$ (graficata)⁵¹.

Lo stesso concetto, si applica ad una sorgente ad L livelli, con Entropia $H_S = \log_2 L$, i cui simboli sono codificati con $\lceil \log_2 L \rceil$ binit/simbolo⁵², con ridondanza $\lceil \log_2 L \rceil - H_S$. Ci chiediamo allora: è possibile trasmettere tanti binit quanti ne servono, e non di più? La risposta è sí, ricorrendo ad una CODIFICA DI SORGENTE, di cui appresso forniamo un esempio.

⁵¹Si presti attenzione sulla differenza: la ridondanza della codifica di sorgente, indica i binit/simbolo sprecati, mentre la ridondanza della codifica di canale, indica il rapporto tra binit di protezione e quelli di informazione.

⁵²La notazione $\lceil \alpha \rceil$ indica l'intero superiore ad α : ad esempio, se $\alpha = 3.7538$, si ha $\lceil \alpha \rceil = 4$.

Codifica di Huffman e codifica entropica a lunghezza variabile Consiste nell'usare CODEWORDS (*parole di codice*) di lunghezza variabile, ed usare le più lunghe per i simboli meno probabili. Consideriamo ad esempio una sorgente con alfabeto di cardinalità $L = 4$, ai cui simboli compete la probabilità riportata nella tabella che segue.

In questo caso l'Entropia vale

$$\begin{aligned} \sum_k p_k \log_2 \frac{1}{p_k} &= \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 + \frac{2}{8} \log_2 8 \\ &= \frac{1}{2} + \frac{1}{2} + \frac{2}{8} \cdot 3 = 1.75 \text{ bit/simbolo} \end{aligned}$$

Simbolo	Prob.	Codeword
x_1	.5	1
x_2	.25	01
x_3	.125	001
x_4	.125	000

mentre il numero medio di binit/simbolo usati è pari a

$$\bar{N} = E \{ \text{Lunghezza} \} = \sum_k \text{Lungh.}(k) p_k = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{2}{8} = 1.75 \text{ binit/simbolo}$$

Osserviamo che:

- le codewords a lunghezza variabile non si confondono tra loro in quanto soddisfano la *regola del prefisso*, ovvero nessuna codeword è uguale all'inizio di una codeword più lunga: se così non fosse, non riusciremmo ad isolarle nel flusso numerico.
- I due risultati (entropia e lunghezza media) coincidono, in virtù del fatto che i valori di probabilità sono potenze non negative di 2.

Nel caso generale (valori di probabilità di emissione qualsiasi), si introduce un ritardo di codifica, in modo da raggruppare più simboli di sorgente, ed effettuare così una cosiddetta codifica a blocchi: per fissare le idee, consideriamo una sorgente binaria, della quale si raggruppino coppie di bit simulando così una sorgente a quattro valori.

Con i risultati riportati in tabella a fianco, si può verificare che si ottiene una Entropia di sorgente

$$H_s = .8 \log_2 \frac{1}{.8} + .2 \log_2 \frac{1}{.2} = 0.72$$

bit/simbolo, mentre adottando una codifica con 1 binit per ogni simbolo generato, si ottiene ovviamente una lunghezza media di 1 binit/simbolo. Al contrario, la lunghezza media ottenibile raggruppando coppie di simboli risulta pari a

Simbolo	Prob.	Codeword
x_1	.8	1
x_2	.2	0
$x_1 x_1$	.64	1
$x_1 x_2$	.16	01
$x_2 x_1$	.16	001
$x_2 x_2$	.04	000

$$\bar{N} = 1 \cdot .64 + 2 \cdot .16 + 3 \cdot .16 + 3 \cdot .04 = 1.5792$$

binit ogni 2 simboli, ossia pari a 0.7896 bit/simbolo.

In generale, prendendo blocchi via via più lunghi, è possibile ridurre la velocità di codifica R (in binit/simbolo) rendendola sempre più simile all'Entropia, ovvero

$$\min [R] = H_S + \varepsilon \quad \text{con } \varepsilon \rightarrow 0$$

se la lunghezza del blocco tende ad infinito (*Teorema di Shannon* sulla codifica di sorgente).

5.6.3 Misura delle grandezze in Decibel

Molto spesso, quando si deve indicare un *rapporto* tra le potenze, si ricorre ad una *scala logaritmica*, definita come: $R(dB) = 10 \log_{10} R$, in cui R è un rapporto tra potenze ed $R(dB)$ il suo valore in scala logaritmica. La relazione inversa è $R = 10^{\frac{R(dB)}{10}}$. Dato che $\log_{10} 2 = 0.30102\dots$, un raddoppio di R produce un aumento di circa 3 dB.

R	0	10^{-3}	1	2	5	10	10^n
R(dB)	$-\infty$	-30	0	~ 3	~ 7	10	$n \cdot 10$

Una misura in dB è la rappresentazione logaritmica di un rapporto, e dunque costituisce una grandezza *relativa*.

Il dB può essere una unità di misura *assoluta*, come ad esempio il dB riferito al Watt, che si scrive dBW. Conoscendo la potenza del rumore $\mathcal{P}_n$ in dBW e l'SNR desiderato, si ottiene per la potenza di segnale $\mathcal{P}_s(dBW) = \mathcal{P}_n(dBW) + SNR(dB)$: infatti,

$$SNR(dB) = 10 \log_{10} \frac{\mathcal{P}_s}{\mathcal{P}_n} = 10 \log_{10} \mathcal{P}_s - 10 \log_{10} \mathcal{P}_n = \mathcal{P}_s(dBW) - \mathcal{P}_n(dBW)$$

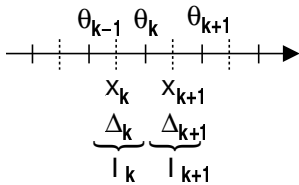
Si noti che $\mathcal{P}_s$ e $\mathcal{P}_n$ sono entrambi espressi in Watt.

Esiste inoltre il dBm, che è una misura logaritmica di un potenza, espressa in milli-Watt; risulta: $\mathcal{P}(dBm) = 10 \log_{10} [\mathcal{P}(mW)] = 10 \log_{10} [10^3 \cdot \mathcal{P}(W)] = \log_{10} 10^3 + \mathcal{P}(dBW) = 30 + \mathcal{P}(dBW)$.

5.6.4 Quantizzazione logaritmica

Abbiamo osservato al termine del § 5.4 che, nel caso in cui il segnale x da quantizzare presenti valori *ridotti* rispetto alla sua dinamica presunta Δ_x , si assiste ad un *peggioramento* di SNR_q , in quanto ciò equivale a disporre di un numero di livelli $L = 2^M$ ridotto. Aggiungiamo ora che se il processo x *non ha* distribuzione di ampiezza uniforme come ipotizzato, il risultato $SNR_q \simeq L^2$ non è più valido.

Si può mostrare che, per una generica $p_X(x)$, il quantizzatore *ottimo* (che rende minimo SNR_q) *non* suddivide la dinamica Δ_x in $L - 1$ intervalli I_k di *uguale* estensione $\Delta_q = \frac{\Delta_x}{L-1}$, ma utilizza intervalli di estensione $\Delta_q(k)$ *ridotta* nelle regioni per le quali $p_X(x)$ è più grande.



Questi valori di x sono infatti i più frequenti, e vengono quantizzati con un errore di potenza $\sigma_q^2(k)$ ridotta rispetto al caso di Δ_q costante. La potenza dell'errore $\mathcal{P}_\epsilon$ si ottiene come valore atteso

$$\mathcal{P}_\epsilon = E \left\{ \sigma_q^2(k) \right\} = \sum_{k=1}^L p_k \sigma_q^2(k)$$

⁵³ Il metodo è iterativo, ed inizia suddividendo l'intervallo Δ_x in modo uniforme. Per ogni iterazione:

- si determinano i valori quantizzati x_k come $x_k = E \{ x \in I_k \} = \int_{I_k} x \cdot p_X(x/k) dx = \frac{\int_{I_k} x p_X(x) dx}{p_k}$. In tal modo, i valori x_k si spostano (internamente a I_k) verso la regione in cui $p_X(x)$ è maggiore;
- si ri-calcolano i confini di decisione θ_k come $\theta_k = \frac{x_k + x_{k+1}}{2}$, seguendo lo spostamento degli x_k .

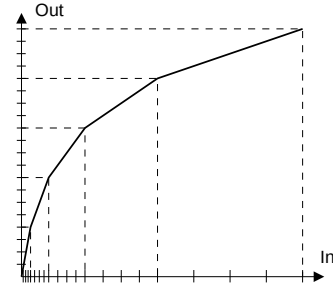
Le iterazioni si arrestano quando non si riscontrano cambiamenti apprezzabili.

avendo indicato $p_k = \int_{I_k} p_X(x) dx$ la probabilità che $x \in I_k$. Il modo ottimo di disporre i confini (θ_{k-1}, θ_k) che delimitano I_k , in modo da rendere minima $\mathcal{P}_\epsilon$, è noto come algoritmo di LLOYD-MAX⁵³.

Nella quantizzazione del segnale vocale, anche se è arbitrario identificare con esattezza una $p_X(x)$, si verifica strumentalmente che quest'ultima è addensata nelle regioni con valori più piccoli. Per questo motivo, la legge di quantizzazione che si è adottata per ottenere gli 8 bit a campione utilizzati nel PCM segue un andamento logaritmico⁵⁴, dimezzando progressivamente la pendenza della caratteristica di ingresso-uscita del quantizzatore.

La figura a fianco mostra un esempio di tale realizzazione (per i soli valori positivi), evidenziando come il risultato possa essere approssimato individuando (a partire dall'origine) regioni di ingresso di ampiezza che man mano raddoppia, e suddividendo la regione in un uguale numero di intervalli equispaziati.

La caratteristica non lineare è realizzabile per via completamente numerica: si attua un campionamento con legge lineare, con un numero di livelli molto maggiore del necessario (es con $4096 = 2^{12}$ livelli); i bit più significativi del campione individuano la regione di ingresso, ed i bit rimanenti sono *shiftati* a destra, per mantenere costante il numero di intervalli per regione. I risultati di tale *mapping* possono infine essere inseriti in una ROM come coppie di *ingresso-uscita*, ovvero come coppie *uscita-ingresso* per il dispositivo che effettua il procedimento inverso.



5.6.5 Funzione caratteristica

E' definita come l'antitrasformata di Fourier di una densità di probabilità, ovvero (equivalentemente) come il valore atteso di e^{jwz} :

$$\Phi_z(w) = \mathcal{F}^{-1}\{p_Z(z)\} = E_Z\{e^{jwz}\} = \int p_Z(z) e^{jwz} dz$$

Osserviamo che, se applicata alla somma di v.a. indipendenti, si ha:

$$\Phi_z(w) = E_Z\{e^{jw(x+y)}\} = E_Z\{e^{jwx}e^{jwy}\} = E_X\{e^{jwx}\} E_Y\{e^{jwy}\} = \Phi_x(w) \Phi_y(w)$$

ovvero la funzione caratteristica di una somma di v.a. indipendenti è pari al prodotto delle funzioni caratteristiche.

Effettuando ora l'operazione inversa (trasformata di Fourier della funzione caratteristica della somma) e ricordando che un prodotto in frequenza è una convoluzione nel tempo (e viceversa) si ottiene il risultato $p_Z(z) = \mathcal{F}\{\Phi_z(w)\} = \mathcal{F}\{\Phi_x(w) \Phi_y(w)\} = p_X(x) * p_Y(y)$ che ci permette di enunciare:

La densità di probabilità della somma di v.a. indipendenti è pari alla convoluzione tra le rispettive densità di probabilità marginali.

La funzione caratteristica ha altri usi... ma non approfondiamo oltre.

⁵⁴L'andamento esatto della curva segue uno di due standard, denominati legge μ (per USA e Giappone) e legge A (per gli altri), livemente diverse nella definizione, ma sostanzialmente equivalenti.

5.6.6 Trasformazioni di v.a. e cambio di variabili

Quando più v.a. si combinano con leggi diverse dalla somma, il risultato del § precedente non è più sufficiente a fornire una espressione per la d.d.p. risultante. Illustriamo allora il procedimento analitico generale, necessario ad ottenere una espressione per la d.d.p. di generiche funzioni di v.a.

5.6.6.1 Caso unidimensionale

Consideriamo una prima v.a. X , ed una seconda Y da essa derivata per mezzo della relazione $y = f(x)$, che si applica alle determinazioni di x di X . La caratterizzazione statistica di Y in termini della sua d.d.p. $p_Y(y)$ può ottenersi a partire da quella di X , nei termini della funzione di distribuzione di Y , come:

$$F_Y(y) = Pr\{Y \leq y\} = Pr\{X \leq f^{-1}(y)\} \quad (5.7)$$

e calcolando poi $p_Y(y) = \frac{dF_Y(y)}{dy}$.

L'espressione (5.7) può dare luogo a risultati più o meno usabili a seconda della natura della trasformazione $f(x)$: in fig. 5.5a) troviamo un esempio in cui i valori $y \leq \tilde{y}$ hanno origine da due diversi intervalli di X ; in corrispondenza di questi, l'area sottesa dalla $p_X(x)$ individua la probabilità cercata.

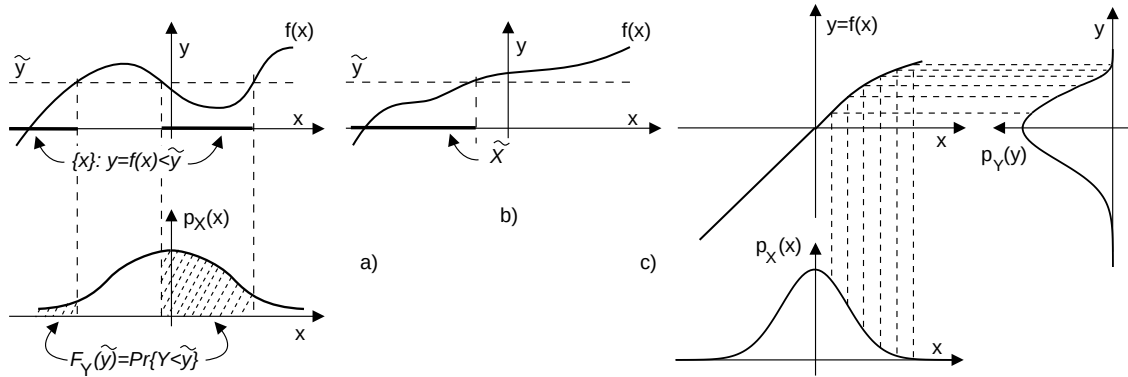


Figura 5.5: Trasformazioni tra variabili aleatorie

Nel caso in cui $f(x)$ sia monotona crescente fig. 5.5 b), per ogni valore di $\tilde{y}$ esiste un solo intervallo di $\tilde{X} \subset X$ tale che $y = f(x \in \tilde{X}) \leq \tilde{y}$, e la (5.7) può essere riscritta come

$$F_Y(y) = Pr\{X \leq f^{-1}(y)\} = F_X(x = f^{-1}(y))$$

che, derivata, permette di giungere alla espressione di calcolo della $p_Y(y)$:

$$p_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dx} \frac{df^{-1}(y)}{dy} = p_X(f^{-1}(y)) \frac{df^{-1}(y)}{dy} \quad (5.8)$$

La (5.8) può essere interpretata a parole, osservando che la nuova v.a. $y = f(x)$ possiede una d.d.p. pari a quella di x , calcolata con argomento pari alla funzione inversa

$x = f^{-1}(y)$, moltiplicata per la derivata di $f^{-1}(y)$. La d.d.p. risultante da una trasformazione di v.a. si presta inoltre ad un processo di costruzione grafica, come esemplificato nella fig. 5.5 c).

Esempio

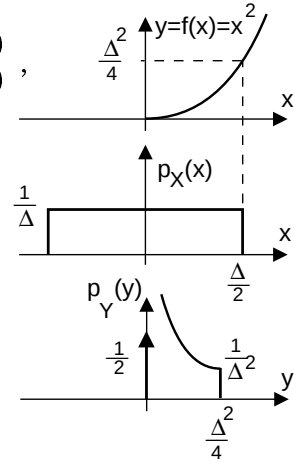
Determinare $p_Y(y)$, qualora risulti $y = f(x) = \begin{cases} 0 & \text{con } x \leq 0 \\ x^2 & \text{con } x > 0 \end{cases}$,

nel caso in cui $p_X(x) = \frac{1}{\Delta} \text{rect}_{\Delta}(x)$.

Osserviamo innanzitutto che tutte le determinazioni $x \leq 0$ danno luogo ad un unico valore $y = 0$; pertanto si ottiene $p_Y(0) = \frac{1}{2}\delta(y)$. Per $0 < y \leq \frac{\Delta^2}{4}$ (corrispondente ad $0 < x \leq \frac{\Delta}{2}$) si applica la teoria svolta, ottenendo $F_Y(y) = Pr\{x \leq \sqrt{y}\} = F_X(\sqrt{y})$, e dunque

$$p_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dx} \bigg|_{x=\sqrt{y}} \frac{d(x=\sqrt{y})}{dy} = \frac{1}{\Delta} \frac{1}{2\sqrt{y}}$$

in cui l'ultima eguaglianza tiene conto che $\frac{dF_X(x)}{dx} = p_X(x)$, che vale $\frac{1}{\Delta}$ per tutti gli x nell'intervallo in considerazione. L'ultima curva mostra la d.d.p. risultante per questo esempio.



5.6.6.2 Caso multidimensionale

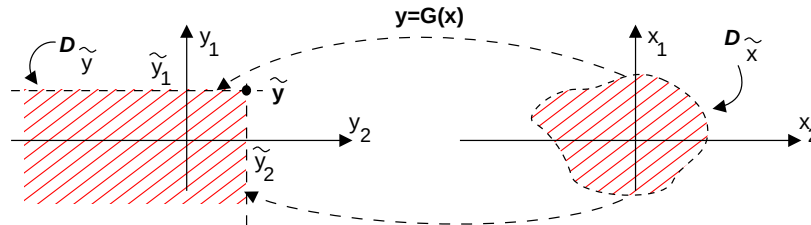
Descriviamo questo caso per mezzo del vettore di v.a. $\mathbf{X} = (x_1, x_2, \dots, x_n)$, a cui è associata una d.d.p. congiunta $p_{\mathbf{X}}(x_1, x_2, \dots, x_n)$, e di un secondo vettore aleatorio $\mathbf{Y}$ dipendente dal primo mediante la trasformazione $\mathbf{Y} = \mathbf{G}(\mathbf{X})$, ovvero

$$\begin{cases} y_1 = g_1(x_1, x_2, \dots, x_n) \\ y_2 = g_2(x_1, x_2, \dots, x_n) \\ \vdots \\ y_n = g_n(x_1, x_2, \dots, x_n) \end{cases}$$

Se esiste la relazione inversa $\mathbf{X} = \mathbf{G}^{-1}(\mathbf{Y})$, composta dall'insieme di funzioni $x_i = g_i^{-1}(y_1, y_2, \dots, y_n)$ per $i = 1, 2, \dots, n$, allora per la d.d.p. di $\mathbf{Y}$ sussiste⁵⁵ un risultato

⁵⁵La dimostrazione segue le medesime linee guida del caso precedente, ed è impostata sulla base della considerazione che la funzione di distribuzione di $\mathbf{Y}$, calcolata in un generico punto $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$, rappresenta la probabilità che $\mathbf{Y}$ appartenga alla regione (dominio) delimitata dal punto $\tilde{\mathbf{y}}$, indicata con $\mathcal{D}_{\tilde{\mathbf{y}}}^{\sim}$:

$$F_{\mathbf{Y}}(\tilde{\mathbf{y}}) = Pr\{\mathbf{Y} \leq \tilde{\mathbf{y}}\} = Pr\{\mathbf{Y} \in \mathcal{D}_{\tilde{\mathbf{y}}}^{\sim}\}$$



formalmente molto simile a quello valido nel caso monodimensionale, e cioè

$$p_{\mathbf{Y}}(y_1, y_2, \dots, y_n) = p_{\mathbf{X}}(\mathbf{X} = \mathbf{G}^{-1}(\mathbf{Y})) \cdot |\mathbf{J}(\mathbf{X}/\mathbf{Y})|$$

in cui $p_{\mathbf{X}}(\mathbf{x} = \mathbf{G}^{-1}(\mathbf{y}))$ è la d.d.p. di $\mathbf{X}$ calcolata con argomento dipendente da $\mathbf{Y}$, e $|\mathbf{J}(\mathbf{X}/\mathbf{Y})|$ è il *giacobiano* della trasformazione inversa $\mathbf{G}^{-1}$, ossia il determinante della matrice costituita da tutte le sue derivate parziali:

$$\mathbf{J}(\mathbf{X}/\mathbf{Y}) = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \dots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \dots & \dots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

v.a. di Raileigh Come applicazione della teoria sopra illustrata, consideriamo la relazione che lega il modulo ρ e la fase φ di un numero complesso $z = x + jy$, alle sue parti reale ed immaginaria, assieme alle corrispondenti funzioni inverse:

$$\begin{cases} \rho = \sqrt{x^2 + y^2} \\ \varphi = \arctan \frac{y}{x} \end{cases} \quad \begin{cases} x = \rho \cos \varphi \\ y = \rho \sin \varphi \end{cases} \quad (5.9)$$

Nel caso in cui x ed y siano due v.a. gaussiane indipendenti, a media nulla e varianza σ^2 , la d.d.p. congiunta di (x, y) si ottiene come prodotto delle d.d.p. marginali, e vale

$$p_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

La $p_{P,\Phi}(\rho, \varphi)$ si ottiene come descritto, valutando⁵⁶ i due termini $p_{X,Y}(x(\rho, \varphi), y(\rho, \varphi))$ e $|\mathbf{J}(x, y/\rho, \varphi)|$, ed ottenendo quindi

$$p_{P,\Phi}(\rho, \varphi) = \frac{\rho}{2\pi\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right)$$

Alla stessa regione $\mathcal{D}_{\tilde{\mathbf{Y}}}$, ne corrisponde una diversa $\mathcal{D}_{\tilde{\mathbf{X}}}$ nello spazio $\mathbf{X}$, tale che per ogni valore $\mathbf{x}^\circ \in \mathcal{D}_{\tilde{\mathbf{X}}}$ risulti $\mathbf{y}^\circ = \mathbf{G}(\mathbf{x}^\circ) \in \mathcal{D}_{\tilde{\mathbf{Y}}}$. Con queste posizioni, la $F_{\mathbf{Y}}(\tilde{\mathbf{y}}) = Pr\{\mathbf{Y} \in \mathcal{D}_{\tilde{\mathbf{Y}}}\}$ si calcola a partire dalla d.d.p. $p_{\mathbf{X}}(\mathbf{x})$, integrata sul dominio $\mathcal{D}_{\tilde{\mathbf{X}}}$:

$$F_{\mathbf{Y}}(\tilde{\mathbf{y}}) = Pr\{\mathbf{X} \in \mathcal{D}_{\tilde{\mathbf{X}}}\} = \int_{\mathcal{D}_{\tilde{\mathbf{X}}}} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

Infine, osservando che

$$p_{\mathbf{Y}}(y_1, y_2, \dots, y_n) = \frac{\partial^n F_{\mathbf{Y}}(y_1, y_2, \dots, y_n)}{\partial y_1 \partial y_2 \dots \partial y_n}$$

si ottiene il risultato mostrato.

⁵⁶Il calcolo dei due termini si esegue come

$$p_{X,Y}(x(\rho, \varphi), y(\rho, \varphi)) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\rho^2(\cos^2 \varphi + \sin^2 \varphi)}{2\sigma^2}\right) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right)$$

$$|\mathbf{J}(x, y/\rho, \varphi)| = \left| \begin{bmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \varphi} \end{bmatrix} \right| = \left| \begin{bmatrix} \cos \varphi & -\rho \sin \varphi \\ \sin \varphi & \rho \cos \varphi \end{bmatrix} \right| = \rho(\cos^2 \varphi + \sin^2 \varphi) = \rho$$

Le d.d.p. marginali $p_P(\rho)$ e $p_\Phi(\varphi)$ si ottengono saturando⁵⁷ la d.d.p. congiunta rispetto all'altra variabile, ricavando

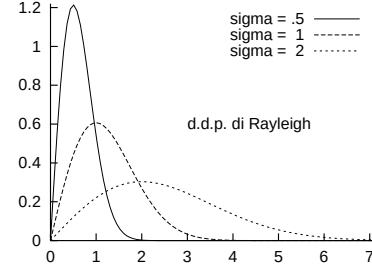
$$p_P(\rho) = \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \quad \text{con } \rho \geq 0; \quad p_\Phi(\varphi) = \frac{1}{2\pi} \quad \text{con } -\pi < \varphi \leq \pi \quad (5.10)$$

L'espressione di $p_P(\rho)$ in (5.10) prende nome di d.d.p. di RAYLEIGH, graficata a sotto.

Il valor medio e la varianza della v.a. ρ valgono rispettivamente $m_P = \sigma\sqrt{\frac{\pi}{2}}$ e $\sigma_P^2 = \sigma^2(2 - \frac{\pi}{2})$, ed è possibile mostrare che per essa vale la proprietà

$$Pr\{\rho > \gamma\} = \int_\gamma^\infty p_P(\rho) d\rho = \exp\left(-\frac{\gamma^2}{2\sigma^2}\right) \quad (5.11)$$

Quest'ultimo valore può rappresentare la probabilità di *mancare* un bersaglio per una distanza superiore a γ , nell'ipotesi che gli errori di puntamento orizzontale e verticale siano entrambi gaussiani, indipendenti, a media nulla ed uguale varianza.

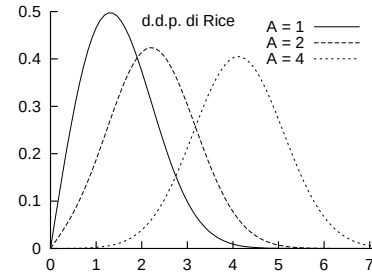


v.a. di Rice Consideriamo nuovamente la trasformazione (5.9), in cui si considera, al posto di x , la v.a. x' , somma di una v.a. gaussiana x e di una costante A . In tal caso, la d.d.p. $p_P(\rho)$ è detta di RICE, ed ha⁵⁸ espressione

$$p_P(\rho) = \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2 + A^2}{2\sigma^2}\right) I_0\left(\frac{\rho A}{\sigma^2}\right) \quad \text{con } \rho \geq 0 \text{ e } I_0(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{z \cos \varphi} d\varphi \quad (5.12)$$

dove $I_0(z)$ è la funzione *modificata* di Bessel del primo tipo ed ordine zero, la cui espressione non ne permette il calcolo in forma chiusa, ma che può essere approssimata come $I_0(z) \sim e^{\frac{z^2}{4}}$ per $z \ll 1$, e come $I_0(z) \sim \frac{e^z}{\sqrt{2\pi z}}$ per $z \gg 1$.

Di lato mostriamo l'andamento di $p_P(\rho)$ con $\sigma = 1$ e tre diversi valori di A , e osserviamo che queste curve vanno rapportate a quelle del secondo grafico per la d.d.p. di Rayleigh, ottenuto per lo stesso valore di σ . Notiamo infine che per $A = 0$ si torna al caso di Rayleigh, mentre per valori crescenti di A , l'andamento della d.d.p. di Rice approssima sempre più quello di una gaussiana.



⁵⁷Svolgiamo il calcolo solo per la prima relazione:

$$p_P(\rho) = \int_{-\pi}^{\pi} p_{P,\Phi}(\rho, \varphi) d\varphi = \frac{\rho}{2\pi\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \cdot \int_{-\pi}^{\pi} d\varphi = \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right)$$

⁵⁸Osserviamo innanzitutto che d.d.p. congiunta di partenza si scrive ora come

$$p_{X',Y}(x', y) \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x' - A)^2 + y^2}{2\sigma^2}\right)$$

in quanto x' è una v.a. gaussiana con media A e varianza σ^2 . Sostituendo quindi nell'esponente $x' = \rho \cos \varphi$ e $y = \rho \sin \varphi$, si ottiene

$$(x' - A)^2 + y^2 = \rho^2 \cos^2 \varphi + A^2 - 2\rho A \cos \varphi + \rho^2 \sin^2 \varphi = \rho^2 + A^2 - 2\rho A \cos \varphi$$

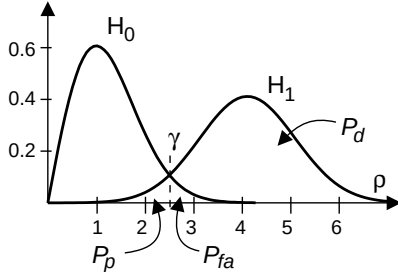
Osservando ora che il giacobiano della trasformazione ha un valore pari a ρ anche in questo caso, otteniamo

$$p_{P,\Phi}(\rho, \varphi) = p_{X',Y}(x'(\rho, \varphi), y(\rho, \varphi)) |J(x', y/\rho, \varphi)| = \frac{\rho}{2\pi\sigma^2} \exp\left(-\frac{\rho^2 + A^2}{2\sigma^2}\right) \exp\left(\frac{2\rho A \cos \varphi}{2\sigma^2}\right)$$

La saturazione di questa d.d.p. congiunta, operata eseguendo $p_P(\rho) = \int_{-\pi}^{\pi} p_{P,\Phi}(\rho, \varphi) d\varphi$, determina il risultato mostrato.

5.6.7 Detezione di sinusoidi nel rumore

Applichiamo i principi della decisione statistica ad un caso “classico”, mostrando come la teoria ora svolta si applichi allo schema di demodulazione incoerente mostrato al § 9.2.1.2. Qualora in ingresso al demodulatore sia presente un rumore di potenza σ^2 , pari a quella delle singole componenti analogiche, ed a questo sia sovrapposta una sinusoidale di ampiezza A , per il modulo dell’involuppo complesso $\rho = |\underline{x}(t)|$ valgono esattamente le considerazioni svolte per la v.a. di Rice.



La ricezione del segnale dà infatti luogo (vedi § 10.1.2) a due processi di rumore $x_c(t)$ ed $x_s(t)$ gaussiani ed ergodici, che in assenza di sinusoidi hanno media nulla e varianza σ^2 , e quindi la d.d.p. di ρ assume un andamento di Rayleigh, come mostrato in figura per il caso indicato con H_0 , e relativo a $\sigma = 1$. La curva H_1 mostra invece la d.d.p. di ρ nel caso di presenza di segnale, con andamento di Rice, per $A = 4$ e $\sigma = 1$.

Il problema nasce qualora il tono sinusoidale non sia presente con continuità, e si desideri operare una decisione relativa all’ipotesi di una sua presenza (indicata con H_1) od assenza (H_0). Nel caso in cui A e σ siano note, il problema è ben posto, e si riduce a determinare una soglia γ con cui confrontare ρ , e decidere per H_0 od H_1 nei casi in cui $\rho < \gamma$ e $\rho > \gamma$ rispettivamente. In tal caso, vengono definiti tre possibili eventi, assieme alle rispettive probabilità, che dipendono dal valore assegnato a γ :

- Probabilità di detezione $P_d = \int_{\gamma}^{\infty} H_1(\rho) d\rho$
- Probabilità di falso allarme $P_{fa} = \int_{\gamma}^{\infty} H_0(\rho) d\rho$
- Probabilità di perdita $P_p = \int_0^{\gamma} H_1(\rho) d\rho$

in cui gli ultimi due valori sono riferiti ad eventi di errore.

La nomenclatura adottata è quella tipica dei radar; in tale contesto, può aver senso tentare di spostare γ in modo da favorire l’uno o l’altro evento in base a considerazioni strategiche⁵⁹. Nel caso in cui i due errori siano equivalenti, e se le probabilità a priori di H_0 ed H_1 sono uguali, la probabilità di errore

$$P_e = Pr(H_0) Pr(e/H_0) + Pr(H_1) Pr(e/H_1) = \frac{1}{2}P_{fa} + \frac{1}{2}P_p \quad (5.13)$$

risulta minimizzata qualora si adotti una decisione di *massima verosimiglianza*, ponendo γ nel punto in cui le due curve si intersecano (come in figura), in modo da preferire l’ipotesi i la cui probabilità a posteriori $Pr(H_i/\rho)$ è più grande.

Una valutazione asintotica delle prestazioni può essere svolta notando che all’aumentare di $\frac{A}{\sigma}$, il valore di γ si avvicina (da destra) ad $\frac{A}{2}$; ponendo quindi $\gamma = \frac{A}{2}$ e sostituendo

⁵⁹In ambito militare, può aver senso ridurre la probabilità di perdita, a spese di un aumento di quella di falso allarme, tranne nel caso in cui quest’ultimo non provochi conseguenze del tutto irreversibili, e “sbagliate” in caso di errore. Ragionamenti analoghi possono essere svolti in campo medico, in cui si dovrebbe preferire un falso allarme, piuttosto che trascurare l’importanza di un sintomo.

le espressioni (5.10) e (5.12) delle d.d.p. a posteriori in quella (5.13) della P_e , otteniamo

$$P_e = \frac{1}{2} \int_{\frac{A}{2}}^{\infty} \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) d\rho + \frac{1}{2} \int_0^{\frac{A}{2}} \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2 + A^2}{2\sigma^2}\right) I_0\left(\frac{\rho A}{\sigma^2}\right) d\rho \quad (5.14)$$

Per ciò che riguarda il primo termine, applicando il risultato (5.11) si trova che il valore $\int_{\frac{A}{2}}^{\infty} \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) d\rho = \exp\left(-\frac{A^2}{8\sigma^2}\right)$. Per il secondo termine, osserviamo che il suo valore è ben più piccolo del primo (si veda la figura tracciata per $A = 4$, o le considerazioni riportate in nota⁶⁰), e quindi può essere trascurato, fornendo in definitiva

$$P_e = \frac{1}{2} \exp\left(-\frac{A^2}{8\sigma^2}\right)$$

per $\frac{A}{\sigma} \gg 1$. Notiamo infine che $\frac{A^2}{2}$ rappresenta la potenza della sinusoide, e che σ^2 è la potenza del rumore. Pertanto, il risultato trovato ha una immediata interpretazione in termini di $SNR = \frac{A^2/2}{\sigma^2}$:

$$P_e = \frac{1}{2} \exp\left(-\frac{SNR}{4}\right)$$

⁶⁰Come già osservato, la funzione modificata di Bessel può essere approssimata come $I_0\left(\frac{\rho A}{\sigma^2}\right) \sim \frac{\exp\left(\frac{\rho A}{\sigma^2}\right)}{\sqrt{2\pi \frac{\rho A}{\sigma^2}}}$ per $\frac{\rho A}{\sigma^2} \gg 1$, e quindi il secondo termine di (5.14) diviene

$$\frac{1}{2} \int_0^{\frac{A}{2}} \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \exp\left(-\frac{A^2}{2\sigma^2}\right) \exp\left(\frac{\rho A}{\sigma^2}\right) \frac{\sigma}{\sqrt{2\pi\rho A}} d\rho = \frac{1}{2} \int_0^{\frac{A}{2}} \sqrt{\frac{\rho}{2\pi\sigma^2 A}} \exp\left(\frac{(\rho - A)^2}{2\sigma^2}\right) d\rho$$

Notiamo ora che per $\frac{A}{\sigma} \gg 1$ l'integrando è trascurabile tranne che per valori di ρ vicini ad A , cosicché il limite inferiore può essere esteso a $-\infty$, e il termine $\sqrt{\frac{\rho}{2\pi\sigma^2 A}}$ può essere sostituito con $\frac{1}{\sqrt{2\pi}\sigma}$, ottenendo quindi la maggiorazione $\frac{1}{2} \int_{-\infty}^{\frac{A}{2}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{\rho^2 - A^2}{2\sigma^2}\right) d\rho$, che è l'espressione dell'integrale di una gaussiana con media A^2 , e che può essere espresso come $\frac{1}{2} \cdot \frac{1}{2} \operatorname{erfc}\left(\frac{A}{2\sqrt{2}\sigma}\right)$.

Se $z \gg 1$, $\frac{1}{2} \operatorname{erfc}(z)$ può essere approssimata come $\frac{1}{2z\sqrt{\pi}} \exp(-z^2)$, ottenendo

$$\frac{1}{2} \int_{-\infty}^{\frac{A}{2}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{\rho^2 - A^2}{2\sigma^2}\right) d\rho \sim \frac{\sigma}{A\sqrt{2\pi}} \exp\left(-\frac{A^2}{8\sigma^2}\right)$$

che, essendo per ipotesi $\frac{A}{\sigma} \gg 1$, risulta trascurabile rispetto a $\frac{1}{2} \exp\left(-\frac{A^2}{8\sigma^2}\right)$.

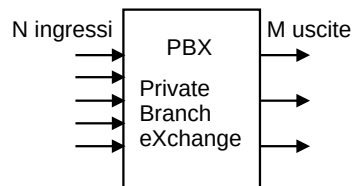
Capitolo 6

Traffico, Code e Reti a Pacchetto

Questo capitolo raggruppa argomenti omogenei, che possono essere trattati avendo già definito i concetti di probabilità necessari allo sviluppo della teoria qui esposta. Sono presentati i concetti e le relazioni che descrivono la probabilità di blocco nei sistemi di servizio orientati alla perdita, nei casi di popolazione finita ed infinita; sono quindi introdotti i sistemi orientati al ritardo, e dimostrati i risultati relativi al caso di coda infinita e servente unico. Infine, sono fornite definizioni e concetti relativi alle reti per trasmissione dati, con particolare riguardo alle reti a pacchetto, nei termini dei diversi modi di funzionamento e delle gerarchie di protocolli adottate.

6.1 Distribuzione Binomiale per popolazione finita

Iniziamo con il chiederci quante linee uscenti M siano necessarie ad un centralino con N interni, per garantire una probabilità di trovare tutte le linee occupate, inferiore ad un valore massimo, chiamato *grado di servizio*¹. Per trovare il risultato, calcoliamo prima la probabilità che tutte le linee uscenti siano occupate, assumendo noti N ed M .



Affrontiamo il problema in termini ancor più generali, chiedendoci quale sia la probabilità $p_B(k)$ che un numero k di persone (su N) sia contemporaneamente al telefono. Assumiamo che ognuno degli N interni abbia una probabilità p di telefonare (in media ci saranno Np telefoni occupati) e che le telefonate siano indipendenti. La probabilità che un ben preciso gruppo di k individui telefoni (ed $N - k$ no), è pari a $p^k q^{N-k}$ (in cui $q = 1 - p$).

Dato che il numero di differenti modi di scegliere k oggetti tra N è pari a $\binom{N}{k} = \frac{N!}{k!(N-k)!}$, allora la probabilità di avere k (qualsiasi) persone al telefono è pari a

$$p_B(k) = \binom{N}{k} p^k q^{N-k}$$

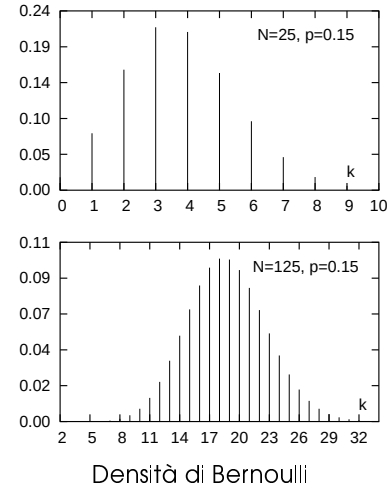
¹Il termine grado di servizio esprime un concetto di *qualità*, ed è usato in contesti diversi per indicare differenti grandezze associate appunto alla qualità dei servizi di telecomunicazione. Nel caso presente, qualità è una piccola probabilità di occupato.

Dato che $\sum_{k=1}^N p_B(k) = 1$, la funzione $p_B(k)$ è una densità di probabilità di v.a. discreta, nota con il nome di variabile aleatoria di *Bernoulli*².

Al variare di k , si ottengono tutte le probabilità cercate, rappresentate nella figura al lato nel caso in cui $p = 0.15$ e $N = 25$; osserviamo che (ovviamente!) non si possono avere più di N utenti al telefono.

Per conoscere il numero di linee occorrenti a garantire una probabilità di *congestione* (o di blocco) P_B inferiore ad un massimo, si sommano (partendo *da destra*) i valori di probabilità $p_B(k)$, finché non si supera la probabilità prefissata: allora M sarà pari all'ultimo indice k . Infatti in tal modo la probabilità che ci siano più di M interni a voler telefonare è pari a

$$Pr(k > M) = \sum_{k=M+1}^N p_B(k) = \sum_{k=M+1}^N \binom{N}{k} p^k q^{N-k} < P_B$$



La distribuzione Binomiale è detta anche *delle prove ripetute* poiché può essere usata per calcolare la probabilità di un certo numero di eventi favorevoli, a seguito della ripetizione dello stesso fenomeno aleatorio³.

Il valore medio della distribuzione Binomiale è $m_B = Np$, e la varianza $\sigma_B^2 = Npq$. Tornando al caso del centralino, il numero medio di linee occupate è Np : tale quantità rappresenta il *traffico offerto*, che si misura in ERLANG. (ad esempio, 3 Erlang = in media ci sono 3 linee occupate).

Il rapporto $\frac{\sigma_B^2}{m_B} = \frac{Npq}{Np} = q < 1$ è un indice di come la variabile aleatoria si distribuisca attorno alla media. Il caso di Bernoulli in cui $\frac{\sigma_B^2}{m_B} < 1$ è rappresentativo di un traffico *dolce*, che deriva dall'ipotesi di popolazione finita: infatti, all'aumentare delle linee occupate, diminuisce la probabilità di una nuova chiamata, in quanto diminuiscono le persone *non* al telefono.

²La $p_B(k)$ è detta anche Binomiale, in quanto i fattori $\binom{N}{k}$ sono quelli della potenza di un binomio $(p+q)^N$, calcolabili anche facendo uso del triangolo di Pascal, mostrato per riferimento di seguito.

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 1 & & 1 & & \\ & 1 & & 2 & & 1 & \\ & 1 & 3 & & 3 & 1 & \\ 1 & 1 & 4 & 6 & 4 & 1 & \\ 1 & 1 & 5 & 10 & 10 & 5 & 1 \\ 1 & 1 & 6 & 15 & 20 & 15 & 6 & 1 \end{array}$$

³Esempio: si voglia calcolare la probabilità di osservare 3 volte testa, su 10 lanci di una moneta. Questa risulta pari a $p_B(3) = \binom{10}{3} p^3 q^7 = 120 \cdot .5^3 \cdot .5^7 = 0.117$, ovvero una probabilità dell'11,7 %.

La stessa distribuzione Binomiale, può anche essere usata per calcolare la probabilità di errore complessiva in una trasmissione numerica realizzata mediante un collegamento costituito da più tratte, collegate da ripetitori rigenerativi. In una trasmissione binaria, si ha errore se un numero dispari di tratte causa un errore per lo stesso bit, e cioè $P_e = \sum_{k=1, k \text{ dispari}}^N \binom{N}{k} p^k q^{N-k}$, in cui p è la prob. di errore per una tratta; inoltre risulta che se $p \ll 1$ e $Np \ll 1$, allora $P_e \simeq Np$.

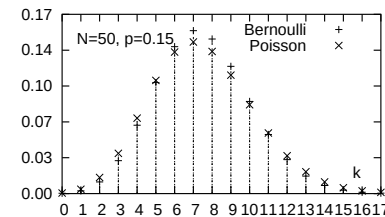
6.2 Distribuzione di Poisson ed Esponenziale negativa

Al crescere del numero N di utenti, l'utilizzo della distribuzione Binomiale può essere scomodo, per via dei fattoriali, e si preferisce trattare il numero di conversazioni attive k come una variabile aleatoria di POISSON, la cui densità ha espressione

$$p_P(k) = e^{-\alpha} \frac{\alpha^k}{k!}$$

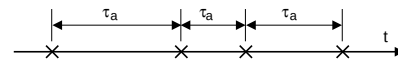
ed è caratterizzata da valor medio e varianza $m_P = \sigma_P^2 = \alpha$.

La Poissoniana costituisce una buona approssimazione della ddp di Bernoulli, adottando per la prima lo stesso valor medio della seconda $m_P = m_B = Np$, come mostrato in figura. In generale, questa densità è impiegata per descrivere la probabilità che si verifichino un numero di eventi *indipendenti e completamente casuali* di cui è noto solo il numero medio α ⁴.



D'altra parte, al tendere di N ad ∞ , il modello Bernoulliano adottato per questo fenomeno aleatorio perde di validità. Infatti, se $N \rightarrow \infty$, il numero di nuove chiamate *non diminuisce* all'aumentare del numero dei collegamenti in corso. A questo scopo, si preferiscono descrivere gli eventi "indipendenti e completamente casuali" in base ad una *frequenza media di interarrivo* λ che rappresenta la velocità⁵ con cui si presentano nuove chiamate⁶.

L'inverso di λ rappresenta un tempo, ed esattamente $\bar{\tau}_a = 1/\lambda$ è il *valor medio* della variabile aleatoria τ_a costituita dall'intervallo di tempo tra l'arrivo di due chiamate. Con queste definizioni, è possibile riferire la v.a. di Poisson ad un intervallo temporale di osservazione t , durante il quale si presentano un numero medio α di chiamate⁷ pari a $\alpha = \lambda t$. Pertanto, possiamo scrivere la v.a. Poissoniana come



$$p_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

che indica la probabilità che in un tempo t si verifichino k eventi (indipendenti e completamente casuali) la cui frequenza media è λ ⁸).

⁴Con questa approssimazione, pertanto, la probabilità che (ad esempio) si stiano svolgendo *meno* di 4 conversazioni contemporanee è pari a $p_P(0) + p_P(1) + p_P(2) + p_P(3) = e^{-\alpha} \left(1 + \alpha + \frac{\alpha^2}{2} + \frac{\alpha^3}{6}\right)$.

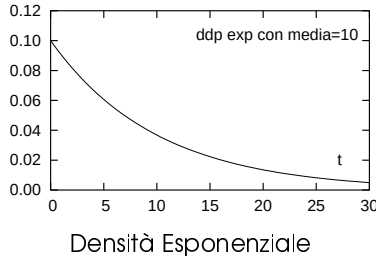
⁵ λ viene espresso in *richieste per unità di tempo*.

⁶La trattazione può facilmente applicarsi a svariate circostanze: dalla frequenza con cui si presentano richieste di collegamento ad una rete di comunicazioni, alla frequenza con cui transitano automobili sotto un cavalcavia, alla frequenza con cui particelle subatomiche transitano in un determinato volume...

⁷Esempio: se da un cavalcavia osserviamo (mediamente) $\lambda = 3$ auto/minuto, nell'arco di $t = 2$ minuti, transiteranno (in media) $3 \cdot 2 = 6$ autovetture.

⁸Sapendo che l'autobus (completamente casuale!) che stiamo aspettando ha una frequenza di passaggio (media) di 8 minuti, calcolare: **A**) la probabilità di non vederne nessuno per 15 minuti e **B**) la probabilità che ne passino 2 in 10 minuti.

Soluzione. Si ha $\lambda = 1/8$ passaggi/minuto e quindi: **A**) $p_0(15) = e^{-\frac{15}{8}} = 0.15$ pari al 15%; **B**) $p_2(10) = e^{-\frac{10}{8}} \frac{(\frac{10}{8})^2}{2} = 0.224$ pari al 22.4%



L'uso della poissoniana è intimamente legato (ma non viene qui dimostrato) al considerare il tempo di interarrivo tra chiamate come una variabile aleatoria completamente casuale⁹, descritta da una densità di probabilità *esponenziale negativa*, espressa analiticamente come

$$p_E(t) = \lambda e^{-\lambda t}$$

per $t \geq 0$ e mostrata in figura; tale v.a. è caratterizzata dai momenti $m_E = \frac{1}{\lambda}$ e $\sigma_E^2 = \frac{1}{\lambda^2}$. La probabilità che un tempo di attesa (a distribuzione esponenziale) superi un determinato valore t_0 è allora subito calcolato come

$$Pr(t > t_0) = \int_{t_0}^{\infty} \lambda e^{-\lambda t} dt = e^{-\lambda t} \Big|_{t_0}^{\infty} = e^{-\lambda t_0}$$

6.3 Sistema di servizio orientato alla perdita

Un *sistema di servizio* è una entità in grado di accogliere *richieste di servizio* (che sono gli eventi che definiscono il *processo di ingresso*) fino al raggiungimento della propria capacità limite, determinata dal numero M di *serventi* di cui il sistema dispone¹⁰. Una volta occupati tutti i serventi, e finché non se ne liberi qualcuno, le successive richieste possono essere poste in coda (in un sistema *orientato al ritardo*) oppure rifiutate, come avviene per i sistemi *orientati alla perdita* che analizziamo ora. Scopo della presente sezione sarà pertanto quello di determinare il numero di serventi necessario a garantire una *probabilità di rifiuto* della richiesta di servizio pari ad un valore che descrive il *grado di servizio* che si intende fornire.

Mentre il processo di ingresso è descritto in termini della frequenza *media* di arrivo λ , il tempo medio di occupazione dei serventi (indicato come *processo di servizio*) è descritto nei termini del *tempo medio di servizio* τ_S , ovvero dal suo inverso $\mu = 1/\tau_S$, pari alla *frequenza media* di servizio. Nella trattazione seguente si fa l'ipotesi che entrambi i processi (di ingresso e di servizio) siano descrivibili in termini di v.a. a distribuzione

⁹La ddp esponenziale è spesso adottata come un modello approssimato ma di facile applicazione per rappresentare un tempo di attesa, ed applicato ad esempio alla durata di una conversazione telefonica, oppure all'intervallo tra due malfunzionamenti di un apparato.

Da un punto di vista formale, considerare eventi che non hanno memoria di quando siano accaduti l'ultima volta, ed indicati come *eventi completamente casuali*, equivale a scrivere che

$$Pr(t > t_0 + \theta / t > t_0) = Pr(t > \theta)$$

esplicitando come la probabilità di attendere altri θ istanti, avendone già attesi t_0 , non dipende da t_0 . Per verificare che la ddp esponenziale rappresenti proprio questo caso, svolgiamo i passaggi:

$$Pr(t > t_0 + \theta / t > t_0) = \frac{Pr(t > t_0 + \theta; t > t_0)}{Pr(t > t_0)} = \frac{Pr(t > t_0 + \theta)}{Pr(t > t_0)} = \frac{e^{-\lambda(t_0 + \theta)}}{e^{-\lambda t_0}} = e^{-\lambda \theta} = Pr(t > \theta)$$

¹⁰Gli esempi dalla vita reale sono molteplici, dal casello autostradale presso cui arrivano auto richiedenti il servizio del casellante (M =numero di caselli aperti), al distributore automatico di bevande (servente unico), all'aereo che per atterrare richiede l'uso della pista (servente unico)... nel contesto delle telecomunicazioni, il modello si applica ogni qualvolta vi siano un numero limitato di risorse a disposizione, come ad esempio (ma non solo!) il numero di linee telefoniche uscenti da un organo di commutazione, od il numero di *time-slot* presente in una trama PCM, od il numero di operatori di un *call-centre*...

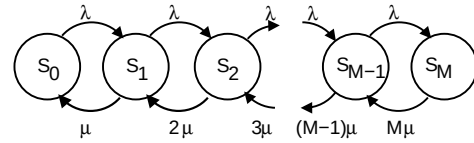
esponenziale¹¹, ovvero che le durate degli eventi “nuova richiesta” e “servente occupato” siano *completamente casuali*¹².

Il rapporto $A_o = \frac{\lambda}{\mu}$ è indicato come *intensità media* del traffico *offerto*¹³ e descrive quanti serventi (in media) *sarebbero* occupati ad espletare le richieste arrivate e non ancora servite, nel caso in cui M fosse infinito. L'aggettivo *offerto* indica il fatto che, essendo invece M finito, alcune richieste non sono accolte, ed A_o risulta diverso dal traffico A_s che può essere effettivamente *smaltito*. L'unità di misura dell'intensità di traffico è l'ERLANG, il cui valore indica appunto il numero medio di serventi occupati. *Ad esempio*: se ad un centralino giungono una media di $\lambda = 3$ chiamate al minuto, e la durata media di una conversazione è $1/\mu = 3$ minuti, allora $A_o = 3 \cdot 3 = 9$ Erlang, corrispondenti al potenziale impegno di una *media* di 9 centralinisti (e nove linee telefoniche).

6.3.1 Probabilità di rifiuto

La teoria che porta a determinare la probabilità che una nuova richiesta di servizio non possa essere accolta per esaurimento dei serventi si basa sull'analisi di un cosiddetto *processo di nascita e morte*, che descrive da un punto di vista statistico l'evoluzione di una popolazione descritta in termini di una frequenza di nascita (nuova conversazione) e di morte (termine di conversazione). In questi termini, un numero di serventi occupati in un sistema di servizio corrisponde al numero di individui della popolazione, e la sua descrizione statistica è condotta con l'ausilio del grafo di figura, in cui il generico stato S_i rappresenta la circostanza che i serventi siano occupati, circostanza a cui compete una probabilità $p_i = Pr(S_i)$.

Gli stati del grafo sono collegati da archi etichettati con la frequenza λ delle transizioni tra gli stati, ovvero dal ritmo con cui si passa da S_i a S_{i+1} a causa di una nuova richiesta, indipendente (per ipotesi) dal numero di serventi già occupati, e dal ritmo $i \cdot \mu$ con cui si torna da S_{i+1} ad S_i , a causa del termine del servizio espletato da uno tra gli $i + 1$ serventi occupati, e proporzionale quindi a questo numero.



Il modello illustrato permette, in base a considerazioni di equilibrio statistico, di giungere¹⁴ ad un risultato in forma chiusa per la probabilità che il sistema si trovi in

un suo generico stato: $p_i = Pr(S_i) = \frac{\frac{A_o^i}{i!}}{\sum_{k=0}^M \frac{A_o^k}{k!}}$. Notiamo ora che p_M è la probabilità che

tutti i serventi siano occupati, pari dunque alla probabilità che una nuova richiesta di servizio sia rifiutata. Chiamiamo allora questo valore *Probabilità di Blocco, di Rifiuto* o

¹¹L'ipotesi permette di valutare la probabilità che l'intervallo temporale tra due eventi di ingresso sia superiore a θ , come $e^{-\lambda\theta}$ (ad esempio, la prob. che tra due richieste di connessione presentate ad una centrale telefonica passi un tempo almeno pari a θ); allo stesso modo, la probabilità che il servizio abbia una durata maggiore di θ è pari a $e^{-\mu\theta}$ (ad esempio, la prob. che una telefonata duri più di θ).

¹²Le ipotesi poste fanno sì che i risultati a cui giungeremo siano conservativi, ovvero il numero di serventi risulterà maggiore od uguale a quello realmente necessario; l'altro caso limite è quello in cui il tempo di servizio non varia, ma è costante (ed in questo caso, non trattato, la stessa intensità di traffico può essere gestita con un numero molto ridotto di serventi); nella realtà ci si troverà in situazioni intermedie.

¹³Si noti che il pedice o è una “o” e non uno “0”, ed identifica appunto l'aggettivo *offerto*.

¹⁴Non ci addentriamo in questa sede nei calcoli relativi, non particolarmente complessi ma comunque abbondanti. La metodologia generale sarà invece illustrata in occasione della trattazione dei sistemi di servizio orientati al ritardo.

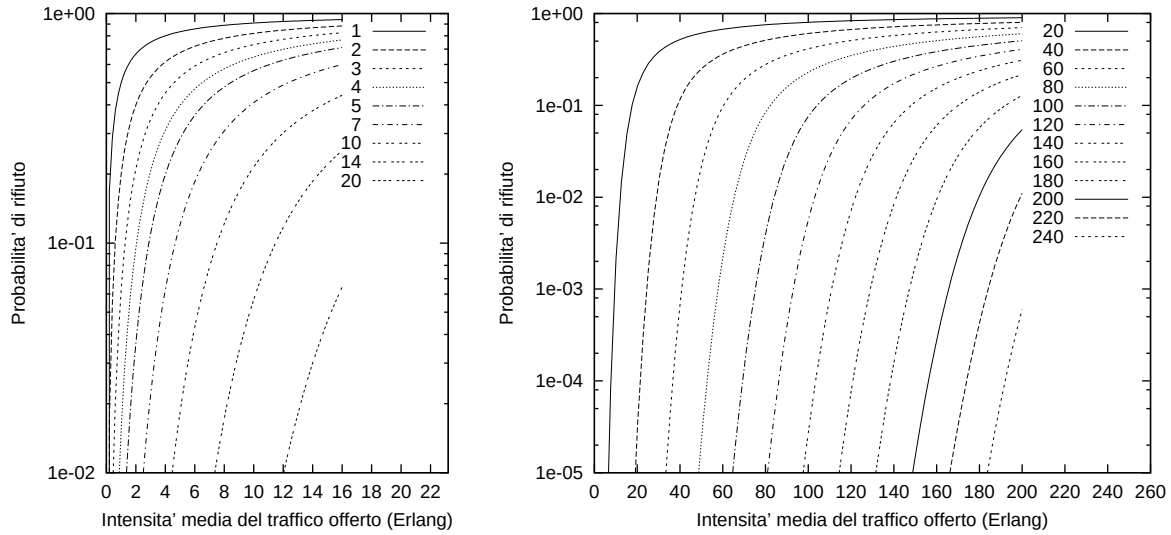


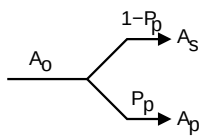
Figura 6.1: Andamento della probabilità di blocco P_B in un sistema orientato alla perdita, al variare di A_o , per il numero di serventi indicato sulle curve

di *Perdita*, la cui espressione prende il nome di Formula B di Erlang, del primo tipo, di ordine M ed argomento A_o :

$$P_B = p(S_M) = \frac{\frac{A_o^M}{M!}}{\sum_{k=0}^M \frac{A_o^k}{k!}} = E_{1,M}(A_o)$$

L'andamento di P_B in funzione di M e di A_o è graficato in Fig. 6.1, e mostra (ad esempio) come per una intensità di traffico offerto pari a 40 Erlang, siano necessari più di 50 serventi per mantenere una P_B minore dell'1%, che salgono a più di 60 per una $P_B = 10^{-3}$.

6.3.2 Efficienza di giunzione



In presenza di una intensità di traffico offerto A_o , solo il $(1 - P_p) \cdot 100\%$ delle richieste è smaltito (pari ad una intensità $A_s = A_o(1 - P_p)$), mentre viene perso un traffico di intensità $A_p = A_o P_p$.

Si può definire un coefficiente di utilizzazione

$$\rho = \frac{A_s}{M} = \frac{A_o}{M} (1 - P_p)$$

che rappresenta la percentuale di impegno dei serventi, e che è graficato nella Fig. 6.2 al variare di A_o , per una P_B assegnata e pari a $2 \cdot 10^{-3}$. Come si vede, una volta fissato il grado di servizio, all'aumentare del numero di serventi il traffico smaltito cresce più in fretta di quanto non crescano i serventi, cosicché (a parità di P_p) l'efficienza aumenta con l'intensità di traffico offerto, e per questo i collegamenti (*giunzioni*) in grado di smaltire un numero più elevato di connessioni garantiscono una maggiore economicità di esercizio.

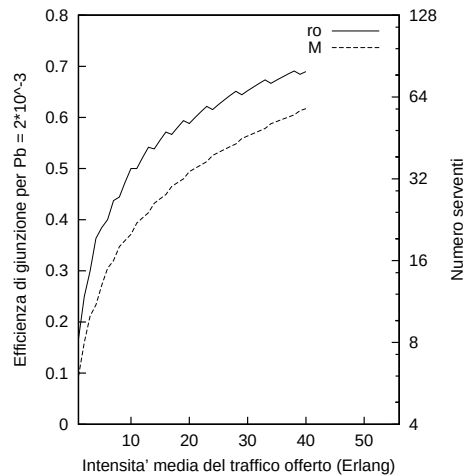


Figura 6.2: Andamento dell'efficienza ρ e del numero di serveri per una probabilità di blocco P_B prefissata

6.3.3 Validità del modello

Le considerazioni esposte si riferiscono ad una ipotesi di traffico completamente casuale con tempi di interarrivo e di servizio esponenziali, e processo di traffico incidente di Poisson. In queste ipotesi, il rapporto $\frac{\sigma_P^2}{m_P} = 1$ tra varianza e media delle distribuzioni di Poisson, è rappresentativo appunto di un traffico *completamente casuale*.

Del tutto diversa può risultare l'analisi, nel caso di una giunzione usata solo nel caso di trabocco del traffico da una giunzione piena. In questo caso λ non è più costante, anzi aumenta con l'aumentare delle connessioni già avvenute, tipico di "traffico a valanga"¹⁵.

Esempio Un numero molto elevato di sorgenti analogiche condivide uno stesso mezzo trasmissivo, caratterizzato da una capacità complessiva di 25.6 Mbps. Le sorgenti sono campionate a frequenza $f_c = 21.33$ KHz e con una risoluzione di 12 bit/campione; ogni sorgente trasmette ad istanti casuali per un tempo casuale e gli intervalli di interarrivo e di servizio sono entrambi v.a. a distribuzione esponenziale negativa di valor medio rispettivamente $\lambda = 20$ richieste/minuto e $\frac{1}{\mu} = 4.25$ minuti.

- 1) Determinare la f_b di una sorgente nelle fasi di attività;
- 2) determinare il numero massimo di sorgenti contemporaneamente attive;
- 3) determinare il grado di servizio (Probabilità di rifiuto) ottenibile con il mezzo trasmissivo indicato;
- 4) indicare la capacità da aggiungere al collegamento per garantire un grado di servizio cento volte migliore.

Risposte

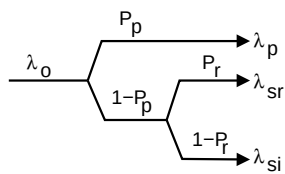
- 1) $f_b = \frac{\text{bit}}{\text{campione}} \cdot \frac{\text{campioni}}{\text{secondo}} = 12 \cdot 21.33 \cdot 10^3 = 256$ Kbps;
- 2) Il numero massimo di sorgenti contemporaneamente attive coincide con il numero di serveri M del collegamento, e quindi $M = \frac{25.6 \cdot 10^6}{256 \cdot 10^3} = 100$ serveri;

¹⁵Un esempio di tale tipo di traffico potrebbe essere... l'uscita da uno stadio (o da un cinema, una metropolitana,...) in cui il flusso di individui non è casuale, ma aumenta fino a saturare le vie di uscita.

- 3) L'intensità media di traffico offerto risulta pari a $A_o = \frac{\lambda}{\mu} = \frac{20}{1/4.25} = 85$ Erlang, e pertanto dalle curve di Fig. 6.1 si trova una probabilità di rifiuto pari a circa 10^{-2} ;
- 4) Si richiede quindi una probabilità di rifiuto 100 volte inferiore, e cioè pari a 10^{-4} . Dalle curve di Fig. 6.1 si osserva che ciò richiede (a parità di A_o) almeno 120 (circa) serventi, 20 in più, pari ad una capacità aggiuntiva di $20 \cdot 256 \cdot 10^3 = 5.120$ Mbps.

6.4 Sistemi di servizio orientati al ritardo

Mentre i sistemi orientati alla perdita rappresentano il modo di operare delle reti di telecomunicazione a *commutazione di circuito*, in cui ogni connessione impegna in modo permanente alcune delle risorse di rete, che una volta esaurite producono un *rifiuto* della richiesta di connessione, i sistemi *orientati al ritardo* sono rappresentativi di reti a *commutazione di pacchetto*, in cui i messaggi sono suddivisi in unità elementari (detti pacchetti, appunto) la cui ricezione non deve più avvenire in tempo reale, e che condividono le stesse risorse fisiche (degli organi di commutazione e di trasmissione) con i pacchetti di altre comunicazioni. Pertanto, l'invio di un pacchetto può essere *ritardato* se la rete è in grado di gestire delle *code di attesa*, in cui accumulare le richieste di servizio che eccedono il numero di serventi a disposizione, e da cui prelevare (con ritardo) i pacchetti stessi non appena si rendano disponibili le risorse necessarie.



In questo caso, il grafico che mostra come si ripartiscono i flussi di richieste si modifica come in figura, dove è evidenziato come la frequenza di richieste λ_o si suddivide tra la frequenza delle richieste perse λ_p , di quelle servite con ritardo λ_{sr} e di quelle servite immediatamente λ_{si} , in funzione della probabilità di perdita P_p e di ritardo P_r . In termini di queste quantità,

valgono le relazioni: $\lambda_p = P_p \lambda_o$; $\lambda_{sr} = P_r (1 - P_p) \lambda_o$; $\lambda_{si} = (1 - P_r) (1 - P_p) \lambda_o$. Indicando con $\tau_S = \frac{1}{\mu}$ il tempo medio di servizio di ogni richiesta, (che non comprende quindi il tempo di accodamento), si definisce, come già noto, una intensità di traffico offerto $A_o = \frac{\lambda_o}{\mu} = \lambda_o \tau_S$, che deve risultare $A_o = A_p + A_{sr} + A_{si}$, e quindi $A_p = \frac{\lambda_p}{\mu}$; $A_{sr} = \frac{\lambda_{sr}}{\mu}$; $A_{si} = \frac{\lambda_{si}}{\mu}$.

Considerando il caso in cui la coda abbia una lunghezza finita e pari ad L , osserviamo che anche le L richieste successive all'impegno di tutti gli M serventi sono accolte (e poste in coda), come se i serventi fossero divenuti $M + L$. In realtà l'analisi fornisce risultati differenti, in quanto le richieste accodate devono essere ancora servite, e quindi il calcolo della P_p non è una diretta estensione dei risultati ottenuti per i sistemi orientati alla perdita. E' comunque abbastanza semplice verificare¹⁶ che ora la P_p risulta inferiore alla P_B del caso senza coda, e pertanto l'intensità di traffico smaltito $A_s = A_{sr} + A_{si} = (1 - P_p) A_o$ aumenta, a parità di offerta.

6.4.1 Risultato di Little

Si tratta di un risultato molto generale, valido per qualsiasi distribuzione dei tempi di interarrivo e di servizio, la cui applicazione può tornare utile nell'analisi, e che recita:

¹⁶Se P_B è la probabilità di blocco derivante dalla disponibilità di M serventi, una frequenza di richieste pari a $P_B \cdot \lambda_o$ non può essere servita immediatamente; adottando una coda, la frequenza delle richieste non servite immediatamente $P_B \cdot \lambda_o$ è uguale a $\lambda_o (P_p + P_r (1 - P_p))$, ed eguagliando le due espressioni si ottiene $P_p = \frac{P_B - P_r}{1 - P_r}$, che è sempre minore di P_B .

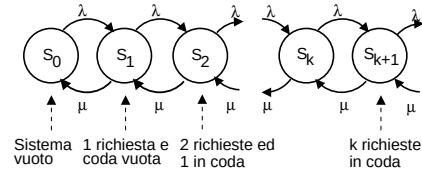
Il numero medio $\bar{N}$ di utenti contemporaneamente presenti in un sistema di servizio è pari al prodotto tra frequenza media di smaltimento delle richieste λ_s ed il tempo medio di permanenza τ_p dell'utente nel sistema

e quindi in definitiva $\bar{N} = \lambda_s \cdot \tau_p$. Nell'applicazione al caso di servizi orientati alla perdita, si ha $\tau_p = \tau_S$, mentre nei servizi a coda risulta $\tau_p = \tau_c + \tau_S$ in cui τ_c rappresenta il tempo medio di coda.

6.4.2 Sistemi a coda infinita ed a servente unico

Prima di fornire risultati più generali, svolgiamo l'analisi per questo caso particolare, in cui la frequenza di richieste perse λ_p è nulla, dato che una coda di lunghezza infinita le accoglie tutte.

Il sistema è descritto da un punto di vista statistico mediante il diagramma di nascita e morte riportato a fianco, in cui ogni stato S_k rappresenta k richieste nel sistema, di cui una sta ricevendo servizio e $k - 1$ sono accodate.



Per procedere nell'analisi, si applica un principio di equilibrio statistico che asserisce che, esaurito il transitorio relativo alla “accensione” del sistema (che ovviamente è inizialmente vuoto), la frequenza media di transizioni tra S_k e S_{k+1} deve eguagliare quella da S_{k+1} ad S_k . Indicando con $p_k = Pr(S_k)$ la probabilità che il sistema contenga k richieste, l'equilibrio statistico si traduce nell'insieme di equazioni

$$\lambda_o p_k = \mu p_{k+1} \quad \text{con} \quad k = 0, 1, 2, \dots, \infty \quad (6.1)$$

Infatti, $\lambda_o p_k$ è pari alla frequenza media con cui il numero di richieste accolte passa da k a $k + 1$; il fatto che il servente sia unico comporta che la frequenza di servizio sia sempre $\mu = \frac{1}{\tau_S}$ indipendentemente dal numero di richieste accodate, e dunque μp_{k+1} è proprio la frequenza media con cui il sistema passa da $k + 1$ a k richieste accolte.

La relazione (6.1) è di natura ricorsiva, e può esprimersi come

$$p_k = \left(\frac{\lambda_o}{\mu} \right)^k p_0 = A_o^k p_0$$

Per determinare il valore $p_0 = Pr(S_0)$, uguale alla probabilità che il sistema sia vuoto, ricordiamo¹⁷ che deve risultare

$$1 = \sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} p_0 A_o^k = p_0 \frac{1}{1 - A_o}$$

da cui otteniamo $p_0 = 1 - A_o$ e dunque

$$p_k = (1 - A_o) A_o^k$$

che corrisponde ad una densità di probabilità esponenziale discreta.

¹⁷Nella derivazione del risultato si fa uso della relazione $\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1-\alpha}$, nota con il nome di *serie geometrica*, e valida se $\alpha < 1$, come infatti risulta nel nostro caso, in quanto necessariamente deve risultare $A_o = \frac{\lambda_o}{\mu} < 1$; se il servente è unico infatti, una frequenza di arrivo maggiore di quella di servizio preclude ogni speranza di funzionamento, dato che evidentemente il sistema non ha modo di smaltire tutte le richieste.

Siamo ora in grado di determinare alcune grandezze di interesse:

Probabilità di ritardo P_r : risulta pari alla probabilità che il sistema non sia vuoto, e cioè che ci sia già almeno una richiesta accolta, ed è pari a¹⁸

$$P_r = 1 - p_0 = 1 - (1 - A_0) = A_0$$

Ricordiamo di aver già definito l'efficienza come il rapporto $\rho = \frac{A_s}{M}$ tra il traffico smaltito ed il numero dei serventi; nel nostro caso $M = 1$ e $A_s = A_o$: dunque $\rho = A_o$. Pertanto, il risultato $P_r = A_o = \rho$ indica come, al tendere ad 1 dell'efficienza, la probabilità di ritardo tenda anch'essa ad 1.

Lunghezza media di coda indicata con $\bar{L}$: risulta essere semplicemente il valore atteso del numero di richieste presenti nel sistema, ovvero¹⁹

$$\bar{L} = E\{k\} = \sum_{k=0}^{\infty} k p_k = (1 - A_o) \sum_{k=0}^{\infty} k A_o^k = \frac{A_o}{1 - A_o}$$

da cui risulta che per $A_o \rightarrow 1$ la coda tende ad una lunghezza infinita.

Tempo medio di permanenza indicato con τ_p , e scomponibile nella somma $\tau_p = \tau_s + \tau_c$ tra il tempo medio di servizio ed il tempo medio di coda. Possiamo applicare qui il risultato di Little $\bar{N} = \lambda_s \cdot \tau_p$, che esprime la relazione tra numero medio $\bar{N}$ di richieste presenti, frequenza di smaltimento (qui pari a quella di offerta²⁰), e tempo medio di permanenza; infatti accade che $\bar{N} = \bar{L}$, ed utilizzando il risultato $\bar{L} = \frac{A_o}{1 - A_o}$ si ottiene

$$\tau_p = \frac{\bar{N}}{\lambda_s} = \frac{\bar{L}}{\lambda_o} = \frac{A_o}{1 - A_o} \frac{1}{\lambda_o} = \frac{\lambda_o}{\mu} \frac{1}{\lambda_o} \frac{1}{1 - \lambda_o/\mu} = \frac{1}{\mu - \lambda_o}$$

da cui si osserva che, se la frequenza di offerta tende al valore della frequenza di servizio, il tempo medio di permanenza tende ad ∞ .

Tempo medio di coda si calcola come

$$\tau_c = \tau_p - \tau_s = \frac{1}{\mu - \lambda_o} - \frac{1}{\mu} = \frac{\mu - \mu + \lambda_o}{\mu(\mu - \lambda_o)} = \frac{A_o}{\mu(1 - A_o)} = \frac{1}{\mu} \frac{\rho}{1 - \rho} = \tau_s \frac{\rho}{1 - \rho}$$

Questo risultato mostra che il tempo medio di coda è legato al tempo medio di servizio e all'efficienza di giunzione, confermando ancora i risultati per $\rho \rightarrow \infty$.

¹⁸Ricordiamo che p_0 è la probabilità che il sistema sia vuoto, e dunque $1 - p_0$ quella che *non* sia vuoto.

¹⁹si fa uso della relazione $\sum_{k=0}^{\infty} k \alpha^k = \alpha \sum_{k=0}^{\infty} k \alpha^{k-1} = \alpha \frac{\partial}{\partial \alpha} \sum_{k=0}^{\infty} \alpha^k = \alpha \frac{\partial}{\partial \alpha} \frac{1}{1 - \alpha} = \frac{\alpha}{(1 - \alpha)^2}$

²⁰Non può essere $\lambda_s > \lambda_o$, perchè si servirebbero più richieste di quante se ne presentano. Se fosse invece $\lambda_s < \lambda_o$, la coda crescerebbe inesorabilmente e sarebbe quindi inutile.

6.4.3 Sistemi a coda finita e con più serventi

Riportiamo solo i risultati, validi se entrambi i processi di ingresso e di servizio sono esponenziali con frequenza media λ_o e μ , la coda è lunga L , i serventi sono M e le sorgenti infinite.

Probabilità di k richieste nel sistema

$$p_k(A_o) = \begin{cases} \frac{A_o^k}{k! \alpha(A_o)} & 0 \leq k \leq M \\ \frac{A_o^k}{M^{k-M} M! \alpha(A_o)} & M \leq k \leq M + L \end{cases}$$

in cui $\alpha(A_o) = \frac{1}{p_0(A_o)} = \sum_{k=0}^{M+L} \frac{A_o^k}{k!}$ e $A_o = \frac{\lambda_o}{\mu}$. Si noti come per $0 \leq k \leq M$ ed $L = 0$ si ottenga lo stesso risultato già esposto per i sistemi orientati alla perdita, mentre per $M = 1$ ed $L = \infty$ ci si riconduca al caso precedentemente analizzato.

Probabilità di ritardo $P_r = \sum_{k=M}^{M+L} p_k(A_o) = p_M(A_o) \frac{1-\rho^{L+1}}{1-\rho}$ in cui $\rho = \frac{A_o}{M}$

Probabilità di perdita $P_p = p_{M+L}(A_o) = \frac{A_o^{M+L}}{M^L M! \alpha(A_o)}$

Tempo medio di coda $\tau_c = \tau_S \frac{P_r - L \cdot P_{M+L}(A_o)}{M - A_o}$

La Figura 6.3 che segue, descrive la probabilità di perdita per un sistema a servente singolo (a sinistra) e con 10 serventi (a destra), in funzione dell'intensità di traffico offerto e della lunghezza di coda. Nel caso di trasmissione di pacchetti di lunghezza fissa, per i quali il tempo di servizio è fisso e *non* a distribuzione esponenziale²¹, i risultati ottenuti costituiscono una *stima conservativa* delle prestazioni del sistema (che potranno cioè essere migliori). L'analisi delle curve permette di valutare con esattezza il vantaggio dell'uso di una coda (a spese del tempo di ritardo). Infatti, aumentando il numero di posizioni di coda si mantiene una probabilità di blocco accettabile anche con un elevato traffico.

Ad esempio, per $P_b = 1\%$ ed $M = 1$, osserviamo che con una coda di $L = 20$ posizioni si gestisce un traffico di $A_o = 0.83$ Erlang, contro gli $A_o = 0.11$ Erlang del caso senza coda. Ciò corrisponde ad un aumento dell'efficienza di $\frac{0.83}{0.11} = 7.54$ volte. D'altra parte ora il tempo medio di coda (calcolato in modo conservativo applicando la relazione per coda infinita) è $\tau_c = \tau_S \frac{\rho}{1-\rho} = \frac{0.83}{1-0.83} \tau_S = 4.9 \tau_S$, ed è quindi aumentato (rispetto a τ_S) di quasi 5 volte.

Esempio Un nodo di una rete per dati effettua la moltiplicazione di pacchetti di dimensione media di 8 Kbyte²² su collegamenti con velocità binaria $f_b = 100$ Mbps²³

- 1) Determinare il tempo medio di servizio di ogni singolo pacchetto;
- 2) determinare il tempo medio di interarrivo τ_a tra pacchetti corrispondente ad un traffico di ingresso di 1200 pacchetti/secondo, e l'associata intensità A_o ;

²¹In una trasmissione a pacchetto, operata a frequenza binaria f_b e con pacchetti di lunghezza media $\bar{L}_p$ bit, il tempo *medio* di servizio per un singolo pacchetto è pari a quello medio necessario alla sua trasmissione, e cioè $\tau_S = \bar{L}_p / f_b$.

²²1 byte = 8 bit, 1 K = $2^{10} = 1024$. Il "K" in questione è "un K informatico". Nel caso invece in cui ci si riferisca ad una velocità di trasmissione, il prefisso K torna a valere $10^3 = 1000$.

²³In virtù di quanto esposto alla nota precedente, in questo caso $1M = 10^6 = 1000000$.

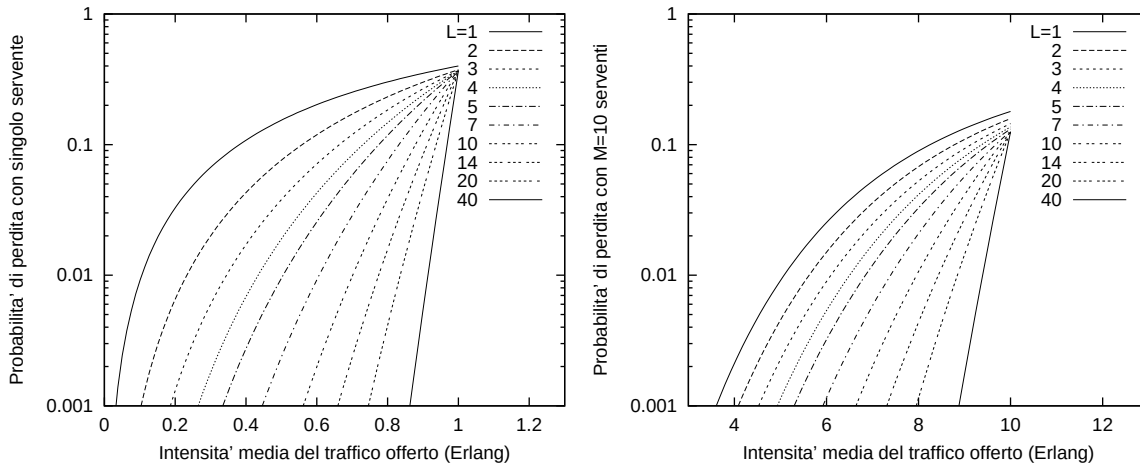


Figura 6.3: Probabilità di perdita per un sistema a coda finita L ed $M = 1, 10$

- 3) assumendo che la dimensione dei pacchetti sia una v.a. con densità esponenziale negativa, così come il tempo di interarrivo tra pacchetti, e che la memoria del moltiplicatore sia così grande da approssimare le condizioni di coda infinita, determinare il ritardo medio di un pacchetto, ossia il tempo medio trascorso tra quando un pacchetto si presenta in ingresso al nodo e quando ne esce;
- 4) calcolare la quantità di memoria necessaria ad ospitare i dati che si accumulano in un intervallo temporale pari al ritardo medio, considerando pacchetti di lunghezza fissa e pari alla media.

Risposte

- 1) Il tempo medio di servizio di un pacchetto è pari al tempo occorrente per trasmetterlo: $\tau_s = \frac{1}{\mu} = \text{durata di un bit} \cdot \frac{\text{bit}}{\text{pacchetto}} = \frac{1}{10^8} \left[\frac{\text{secondi}}{\text{bit}} \right] \cdot 8 \cdot 1024 \left[\frac{\text{byte}}{\text{pacchetto}} \right] \cdot 8 \left[\frac{\text{bit}}{\text{byte}} \right] \simeq 655 \mu\text{sec};$
- 2) $\tau_a = \frac{1}{\lambda} = \frac{1}{1200} = 833 \mu\text{sec}; A_o = \frac{\lambda}{\mu} = 1200 \cdot 655 \cdot 10^{-6} = 0.786 \text{ Erlang};$
- 3) Le condizioni poste corrispondono a quelle di traffico poissoniano e sistema a singolo servente e coda infinita, per il quale la teoria fornisce per il tempo di permanenza il risultato $\tau_p = \frac{1}{\mu - \lambda_o} = \frac{1}{\frac{10^8}{655} - \frac{10^6}{833}} = \frac{1}{326} \simeq 3 \text{ msec};$
- 4) La memoria necessaria è pari al prodotto tra il tempo medio di permanenza ed il numero di bit che si accumulano in quel periodo, ovvero $3 \cdot 10^{-3} [\text{sec}] \cdot 1200 \left[\frac{\text{pacch}}{\text{sec}} \right] \cdot 8 \cdot 1024 \left[\frac{\text{byte}}{\text{pacch}} \right] \simeq 29.5 \text{ Kbyte}.$

6.5 Reti per trasmissione dati

In questa sezione illustriamo le particolarità legate alle *trasmissioni dati*, e come queste possano essere vantaggiosamente sfruttate per conseguire la *maggior efficienza* che i sistemi di servizio a coda presentano rispetto a quelli orientati alla perdita. Le particolari *modalità* e *funzioni* legate alle trasmissioni dati saranno classificate secondo uno schema che ne consente il confronto in termini di prestazioni e vincoli sulla realizzazione della rete. Infine, verranno formalizzate le esigenze legate alla soluzione dei problemi di trasmissione dati, introducendo i concetti legati alle *architetture protocollari*, assieme ad alcuni esempi reali.

Le trasmissioni dati si prestano bene a comunicazioni in cui siano possibili ritardi temporali variabili, attuando una filosofia di tipo *ad immagazzinamento a rilancio* (STORE AND FORWARD) basata sul suddividere il messaggio in unità informative elementari denominate *pacchetti*, che possono essere inoltrati sulla rete di comunicazione, assieme a quelli prodotti da altre trasmissioni. L'applicazione della stessa metodologia a trasmissioni (ad esempio) vocali non è per nulla semplice, in quanto la presenza di un ritardo variabile per la trasmissione dei pacchetti comporta problemi non trascurabili, a meno di non attuare speciali meccanismi di priorità e prenotazione della banda, tuttora oggetto di ricerca.

6.5.1 Il pacchetto dati

Discutiamo brevemente, in termini generali, i possibili contenuti di un pacchetto dati; il suo formato effettivo dipenderà dal particolare protocollo di trasmissione adottato.

La prima osservazione da fare, è che la suddivisione del messaggio in pacchetti comporta un aumento delle informazioni da trasmettere, in quanto ognuno di questi dovrà contenere informazioni aggiuntive per consentire un suo corretto recapito e la sua ricombinazione con gli altri pacchetti dello stesso messaggio. Occorre inoltre affrontare gli ulteriori problemi, tipici di una comunicazione di dati, ai fini del recupero degli errori di trasmissione, nonché relativi alla corretta gestione delle risorse di rete.

In termini generali, un pacchetto è quindi composto da una *intestazione* (HEADER), dalla parte di messaggio che trasporta (*dati*), e da un campo *codice di parità* (CRC) necessario a rivelare l'occorrenza di errori di trasmissione²⁴.



L'*header* a sua volta può essere suddiviso in campi, in cui trovano posto (tra le altre cose) gli *indirizzi* del destinatario e della sorgente, un *codice di controllo* che causa in chi lo riceve l'esecuzione di una procedura specifica, un *numero di sequenza* che identifica il pacchetto all'interno del messaggio originale, ed un campo che indica la *lunghezza* del pacchetto. Nonostante la presenza delle informazioni aggiuntive²⁵, la trasmissione a pacchetto consegue una efficienza maggiore di quella a circuito, in quanto è attuata mediante sistemi a coda.

Può sembrare vantaggioso mantenere elevata la dimensione dei pacchetti, per ridurre la rilevanza delle informazioni aggiuntive, ma si verificano controindicazioni. Infatti, suddividere messaggi lunghi in pacchetti più piccoli garantisce l'inoltro di (altre) comunicazioni più brevi durante la trasmissione di messaggi lunghi, che altrimenti bloccherebbero i sistemi di coda se realizzate con un unico "pacchettone". Inoltre, all'aumentare della lunghezza di un pacchetto, aumenta proporzionalmente la probabilità di uno (o più) bit errati, e dunque l'uso di dimensioni ridotte riduce le necessità di ritrasmissione.

²⁴La sigla CRC significa *Cyclic Redundancy Check* (controllo ciclico di ridondanza) ed indica una parola binaria i cui bit sono calcolati in base ad operazioni algebriche operate sui bit di cui il resto del messaggio è composto. Dal lato ricevente sono eseguite le stesse operazioni, ed il risultato confrontato con quello presente nel CRC, in modo da controllare la presenza di errori di trasmissione.

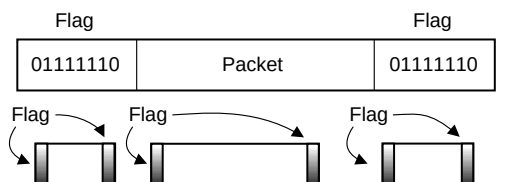
²⁵L'entità delle informazioni aggiuntive rispetto a quelle del messaggio può variare molto per i diversi protocolli, da pochi bit a pacchetto fino ad un 10-20% dell'intero pacchetto (per lunghezze ridotte di quest'ultimo).

6.5.2 Modo di trasferimento delle informazioni

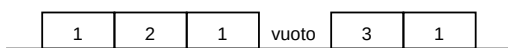
È definito in base alla specificazione di 3 caratteristiche che lo contraddistinguono: lo *schema di multiplazione*, il *principio di commutazione* e l'*architettura protocollare*.

6.5.2.1 Schema di multiplazione

È stato già illustrato uno schema a divisione tempo che prevede l'uso di una *trama*, in cui trovano posto diverse comunicazioni vocali (PCM telefonico), e che necessita di un funzionamento sincronizzato dei nodi di rete. La trasmissione a pacchetti invece non prevede l'uso esclusivo di risorse da parte delle singole comunicazioni, e *non fa uso* di una struttura di trama e pertanto occorrono soluzioni particolari per permettere la *delimitazione* dei pacchetti.



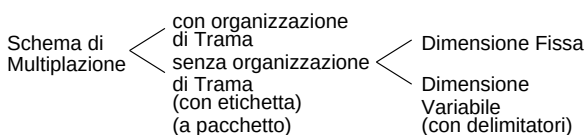
Per evitare che i dati “propri” del pacchetto possano simulare un flag, in trasmissione viene inserito un bit 0 dopo 5 uni di fila, che (se presente) viene rimosso al ricevitore. Se dopo 5 uni c'è ancora un 1, allora è un flag.



in presenza di trama, tranne che... la trama non c'è, e dunque l'ordine dei pacchetti è qualsiasi, ma viene meno l'esigenza dei flag di delimitazione.

Ad esempio, i protocolli HDLC ed X.25 presentano pacchetti di dimensioni variabili, e fanno uso di un byte di *flag* costituito dalla sequenza 01111110 in testa ed in coda, per separare tra loro i pacchetti di comunicazioni differenti.

Nel caso in cui il pacchetto invece abbia una *dimensione fissa*²⁶, ci si trova ad operare in una situazione simile a quella



In entrambi i casi (lunghezza di pacchetto fissa o variabile) i nodi della rete non necessitano di operare in sincronismo tra loro; lo schema di multiplazione è quindi detto *a divisione di tempo senza orga-*

nizzazione di trama, asincrono, con etichetta. Il termine etichetta indica che ogni pacchetto deve recare con sé le informazioni idonee a ricombinarlo assieme agli altri dello stesso messaggio.

6.5.2.2 Principio di commutazione

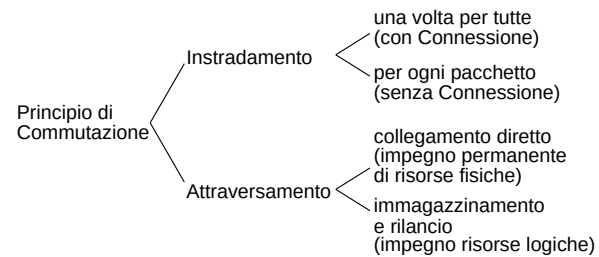
È definito in base a come sono realizzate le due funzioni di *instradamento* (come individuare un percorso nella rete) e *attraversamento* (come permettere l'inoltro del messaggio).

Se l'*instradamento* (ROUTING) viene determinato una volta per tutte all'inizio del collegamento, il modo di trasferimento viene detto *con connessione*. Se al contrario l'*instradamento* avviene in modo indipendente per ogni pacchetto, il collegamento è detto *senza connessione* ed ogni pacchetto di uno stesso messaggio può seguire percorsi differenti.

²⁶Un modo di trasferimento con pacchetti di dimensione fissa è l'ATM (ASYNCHRONOUS TRANSFER MODE) che viene descritto sommariamente in appendice.

L'*attraversamento* di un nodo di rete consiste invece nel *demoltiplicare* le informazioni in ingresso e moltiplicarle di nuovo su uscite diverse: ciò può avvenire mediante un *collegamento diretto* o per *immagazzinamento e rilancio*.

Sulla base di queste considerazioni, definiamo:

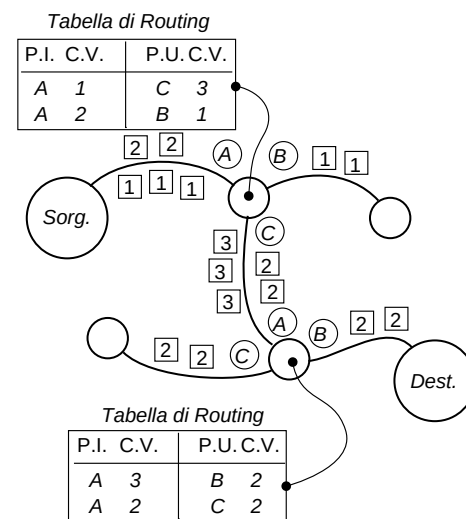


Commutazione di circuito: l'instradamento avviene una volta per tutte prima della comunicazione, e l'attraversamento impegna in *modo permanente* le *risorse fisiche* dei nodi della rete; è il caso della telefonia, sia POTS che PCM²⁷.

Commutazione di pacchetto a Circuito Virtuale: L'instradamento avviene *una volta per tutte* all'inizio della trasmissione, che prevede una prima fase di richiesta di connessione, durante la quale si determina appunto l'instradamento tra i nodi della rete. I pacchetti di uno stesso messaggio seguono quindi tutti uno stesso percorso, e l'attraversamento si basa sull'impegno di *risorse logiche*²⁸ ed avviene per *immagazzinamento e rilancio*. La trasmissione ha luogo dopo aver contrassegnato ogni pacchetto con un *identificativo di connessione (IC)* che identifica un *canale virtuale*²⁹ tra coppie di nodi di rete, e che ne individua l'appartenenza ad uno dei collegamenti in transito.

L'intestazione del pacchetto può essere ridotta, al limite, a contenere il solo *IC* del canale virtuale. L'attraversamento avviene consultando apposite tabelle (dette di *routing*), generate nella fase di instradamento iniziale, in cui è indicato il ramo di uscita per tutti i pacchetti appartenenti ad uno stesso messaggio.

Facciamo un esempio: una sorgente, a seguito della fase di instradamento, invia i pacchetti con identificativo $IC = 1$ al primo nodo individuato dal routing. Consultando la propria tabella, il nodo trova che il canale virtuale *1* sulla *porta di ingresso* (P.I.) *A* si connette al c.v. *3* sulla *porta di uscita* (P.U.) *C*. Ora i pacchetti escono da *C* con $IC = 3$ ed una volta giunti al nodo seguente sulla P.I. *A*, escono dalla P.U. *B* con $IC = 2$ e giungono finalmente a destinazione.



Notiamo che su di un collegamento *tra due nodi*, i numeri dei canali virtuali identificano in modo univoco il collegamento a cui appartengono i pacchetti, mentre uno stesso

²⁷La sigla POTS è l'acronimo di *Plain Old Telephone Service* (buon vecchio servizio telefonico), e si riferisce all'era della telefonia "analogica", in cui si creava un vero e proprio circuito elettrico. Le risorse fisiche impegnate in questo caso sono gli organi di centrale ed i collegamenti tra centrali, assegnati per tutta la durata della comunicazione in esclusiva alle due parti in colloquio. Nel caso del PCM, le risorse allocate cambiano natura (ad esempio consistono anche nell'intervallo temporale assegnato al canale all'interno della trama) ma ciononostante ci si continua a riferire ugualmente alla cosa come ad una rete a *commutazione di circuito*. Infine, ricordiamo che la rete telefonica pubblica è anche indicata come PSTN (*Public Switched Telephone Network*).

²⁸Le risorse impegnate sono dette *logiche* in quanto corrispondono ad entità concettuali (i *canali virtuali* descritti nel seguito).

²⁹Il termine *Canale Virtuale* simboleggia il fatto che, nonostante i pacchetti di più comunicazioni viaggino "rimiscolati" su di uno stesso mezzo, questi possono essere distinti in base alla comunicazione a cui appartengono, grazie ai differenti *IC* (numeri) con cui sono etichettati; pertanto, è come se i pacchetti di una stessa comunicazione seguissero un proprio *canale virtuale* indipendente dagli altri.

numero di canale virtuale può essere riutilizzato su porte differenti³⁰. La concatenazione dei canali virtuali attraversati viene infine indicata con il termine *Circuito Virtuale* per similitudine con il caso di commutazione di circuito, con la differenza che ora il percorso individuato è definito solo in termini di tabelle e di etichette, e non di risorse fisiche (tranne che per la memoria della tabella).

Al termine della comunicazione, sul circuito virtuale viene inviato un apposito pacchetto di controllo, che provoca la rimozione del routing dalle tabelle.

Congestione e controllo di flusso Durante la fase di instradamento, il percorso nella rete è determinato in base alle condizioni di traffico del momento, ed eventualmente la connessione può essere rifiutata in caso in cui la memoria di coda nei nodi coinvolti sia quasi esaurita, evento indicato con il termine di *congestione*.

D'altra parte, se alcune sorgenti origine dei Canali Virtuali già assegnati e che si incrociano in uno stesso nodo intermedio, iniziano ad emettere pacchetti a frequenza più elevata del previsto, il nodo intermedio si congestiona (ossia esaurisce la memoria di transito) ed inizia a *perdere pacchetti*, penalizzando tutti i Canali Virtuali che attraversano il nodo.

Per questo motivo, sono indispensabili strategie di *controllo di flusso* che permettano ai nodi di regolare l'emissione delle sorgenti. Il controllo di flusso è attuato anch'esso mediante pacchetti (di controllo), privi del campo di dati, ma contenenti un codice identificativo del comando che rappresentano. Ad esempio, un nodo non invia nuovi pacchetti di un circuito virtuale finché non riceve un *pacchetto di riscontro* relativo ai pacchetti precedenti. D'altra parte, nel caso di una rete congestionata, la perdita di pacchetti causa il mancato invio dei riscontri relativi, e dunque i nodi a monte cessano l'invio di nuovi pacchetti³¹. Dopo un certo periodo di tempo (TIMEOUT) il collegamento è giudicato interrotto e viene generato un pacchetto di *Reset* da inviare sul canale virtuale, e che causa, nei nodi attraversati, il rilascio delle risorse logiche (tabelle) relative al Canale Virtuale.

Discutiamo ora invece di un ulteriore possibile principio di commutazione:

Commutazione di pacchetto a Datagramma Anche in questo caso, l'*attraversamento* dei nodi avviene per *immagazzinamento e rilancio*, mentre la funzione di *instradamento* è svolta in modo distribuito tra i nodi di rete *per ogni pacchetto*, i quali (chiamati ora *datagrammi*) devono necessariamente contenere l'indirizzo completo della destinazione. Infatti, in questo caso manca del tutto la fase iniziale del collegamento, in cui prenotare l'impegno delle risorse (fisiche o logiche) che saranno utilizzate³². Semplicemente, non è previsto alcun impegno a priori, ed ogni pacchetto costituisce un collegamento individuale che impegna i nodi di rete solo per la durata del proprio passaggio. L'instradamento avviene mediante tabelle presenti nei nodi, sia di tipo statico che dinamico (nel qual caso tengono conto delle condizioni di carico e di coda dei nodi

³⁰I numeri di c.v. sono negoziati tra ciascuna coppia di nodi durante la fase di instradamento, e scelti tra quelli non utilizzati da altre comunicazioni già in corso. Alcuni numeri di c.v. inoltre possono essere riservati, ed utilizzati per propagare messaggi di segnalazione inerenti il controllo di rete.

³¹In realtà vengono prima fatti dei tentativi di inviare nuovamente i pacchetti "vecchi". Questi ultimi infatti sono conservati da chi li invia (che può anche essere un nodo intermedio), finché non sono riscontrati dal ricevente. Quest'ultimo fatto può causare ulteriore congestione, in quanto restano impegnate risorse di memoria "a monte" della congestione che così *si propaga*.

³²Per questo motivo, il collegamento è detto *senza connessione*.

limitrofi) che indicano le possibili porte di uscita per raggiungere la destinazione scritta sul pacchetto. Quest'ultimo quindi viene fatto uscire *senza nessuna alterazione* dalla porta di uscita.

Uno dei maggiori vantaggi dei datagrammi rispetto ai circuiti virtuali è una migliore resistenza ai guasti e malfunzionamenti: in questo caso infatti, a parte una eventuale necessità di ritrasmettere i pacchetti persi, il collegamento prosegue attraverso percorsi alternativi; inoltre l'elevato numero di percorsi alternativi, può permettere (in condizioni di carico leggero) di soddisfare brevi richieste di trasmissione a velocità elevate. Allo stesso tempo, in presenza di messaggi molto brevi, l'invio di un singolo datagramma è più che sufficiente, mentre nel caso a circuito virtuale le fasi di instaurazione ed abbattimento sarebbero state un lavoro in più da svolgere (tanto che ad es. l'X.25, che è nato a c.v., prevede anche il funzionamento a datagramma).

Consegna ordinata e congestione Uno dei maggiori problemi legati all'uso di datagrammi è che l'ordine di arrivo dei pacchetti può essere diverso da quello di partenza, potendo questi seguire percorsi differenti. Per questo motivo, nei datagrammi è presente un *numero di sequenza* che si incrementa ad ogni pacchetto trasmesso, ed alla destinazione sono predisposti dei *buffer*³³ di memoria nei quali ricostruire l'ordine esatto dei pacchetti.

Nel caso di un pacchetto mancante, il ricevente non sa se questo è semplicemente ritardato oppure è andato perso, rendendo problematico il controllo di flusso. In questo caso si produce un impegno anomalo dei buffer di ingresso, che non possono essere rilasciati perchè incompleti, e ciò può causare il rifiuto dell'accettazione di nuovi pacchetti, provocando un impegno anomalo anche per i buffer di uscita di altri nodi, causando congestione³⁴.

Prima di effettuare un trasferimento a datagramma, è opportuno (a parte il caso di messaggi composti da un singolo datagramma) verificare la disponibilità del destinatario finale, e preavvisarlo di riservare una adeguata quantità di memoria. Ad esempio, in Internet avviene proprio questo (vedi pag. 139).

Proseguiamo la descrizione delle reti per dati con l'ultima caratteristica di un modo di trasferimento:

6.5.2.3 Architettura protocollare

Definisce la stratificazione delle funzioni di comunicazione, sia per gli apparati terminali che per i nodi di transito, e di come queste interagiscono reciprocamente sia tra nodi diversi, che nell'ambito di uno stesso nodo. Alcune di queste sono già state introdotte, e le citiamo per prime, seguite da quelle più rilevanti illustrate di seguito:

- il *controllo di flusso*, che impedisce la saturazione dei buffer;
- la *consegna ordinata*, per riassemblare messaggi frammentati su più datagrammi;

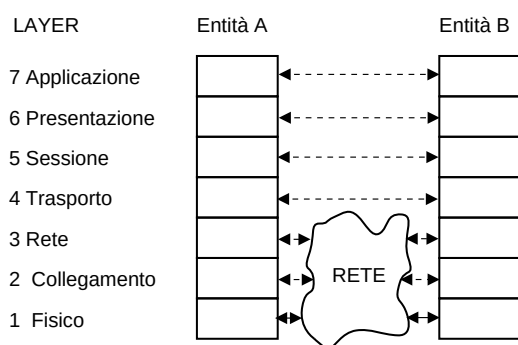
³³Il termine *buffer* ha traduzione letterale "respingente, paracolpi, cuscinetto" ed è a volte espresso in italiano dalla locuzione *memoria tampone*.

³⁴La soluzione a questa "spirale negativa" si basa ancora sull'uso di un allarme a tempo (timeout), scaduto il quale si giudica interrotto il collegamento, e sono liberati i buffer.

- la *segmentazione e riassemblaggio*, che definisce le regole per frammentare un messaggio in pacchetti e ricomporli, ad esempio in corrispondenza dei “confini” tra sottoreti con differente lunghezza di pacchetto;
- il *controllo di connessione*, che provvede ad instaurare la connessione, eseguire l’instradamento, impegnare le risorse, supervisionare il controllo di flusso, abbattere la connessione al suo termine;
- il *controllo di errore*, che provvede a riscontrare le unità informative, a rilevare gli errori di trasmissione, a gestire le richieste di trasmissione;
- l’*incapsulamento*, che aggiunge ai pacchetti di dati da trasmettere le informazioni di protocollo come l’header, gli indirizzi, il controllo di parità...

Stratificazione ISO-OSI Allo scopo di fornire un aiuto alla schematizzazione delle interazioni tra le funzioni illustrate, l’*International Standard Organization* (ISO) ha formalizzato un modello concettuale per sistemi di comunicazione denominato *Open System Interconnection* (OSI)³⁵, che individua una relazione gerarchica tra i protocolli. In particolare sono definiti sette *strati* o *livelli* (LAYERS) ognuno dei quali raggruppa un insieme di funzioni affini. Gli strati più elevati (4-7) sono indicati anche come *strati di utente*, in quanto legati a funzioni relative ai soli apparati terminali; gli *strati di transito* invece (1-3) riguardano funzioni che devono essere presenti anche nei nodi intermedi.

La relazione gerarchica individuata stabilisce tra due strati contigui un rapporto di tipo *utente-servizio*; ovvero lo svolgimento delle funzioni di strato superiore necessita dei servizi offerti dallo strato inferiore.



A titolo di esempio, si pensi all’invio di un documento mediante un corriere espresso: ci si affida allora ad uno strato di trasporto che offre all’utente (strato di *sessione*) un servizio (appuntamento) di *trasporto* che ha il compito di “far apparire” il documento presso il destinatario. La sede locale del corriere si affida quindi alla propria divisione interna che gestisce la *rete* dei corrispondenti, la quale si affida a sua volta ai corrispondenti stessi, che hanno il compito di assistere alla consegna ed all’arrivo (*col-*

legamento) del documento. Il trasferimento *fisico* dello stesso può quindi avvenire mediante un ultimo strato funzionale (treno, nave, aereo, auto...) che provvede esclusivamente al recapito in base alle informazioni ricevute dallo strato di collegamento.

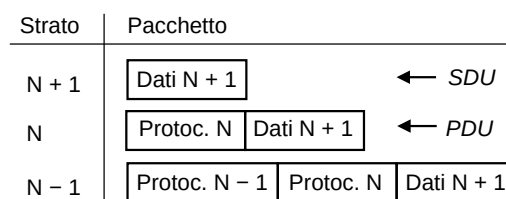
Per terminare l’esempio, facciamo notare come in ogni livello avvengano *due* tipi di colloqui (regolati da altrettanti protocolli): uno è *orizzontale*, detto anche *tra pari* (PEER-TO-PEER), come è ad esempio il contenuto del documento che spediamo, od i rapporti tra corrispondenti locali (che nel caso di un sistema di comunicazione corrisponde allo strato di collegamento, relativo ai procolli tra singole coppie di nodi di rete); il secondo tipo di colloquio avviene invece in forma *verticale*, o *tra utente e servizio*, in quanto per realizzare le funzioni di uno strato *utente* ci si affida ad un *servizio* di comunicazione

³⁵In virtù dell’intreccio di sigle, il modello di riferimento prende il nome (palindromo) di modello *ISO-OSI*.

offerto dallo strato inferiore (che a sua volta può avvalersi dei servizi degli altri strati ancora inferiori)³⁶.

Incapsulamento A riguardo della modalità con cui un protocollo tra pari di strato N affidi i suoi dati ad un servizio di strato $N - 1$, questo avviene (nella commutazione di pacchetto) mediante la funzione di *incapsulamento*, di cui viene data una interpretazione grafica alla figura seguente.

I dati che lo strato $N + 1$ vuol trasmettere al suo pari, indicati anche come *Service Data Unit* (SDU), sono prefissi dalle informazioni di protocollo necessarie alla gestione del collegamento tra entità allo strato N . Questa nuova unità informativa prende il nome di *Protocol Data Unit* (PDU) per lo strato N , e viene passata in forma di SDU al servizio di collegamento offerto dallo strato $N - 1$, che ripete l'operazione di incapsulamento con le proprie informazioni di protocollo, generando una nuova PDU (di strato $N - 1$). Pertanto, lo strato fisico provvederà a trasmettere pacchetti contenenti tutte le informazioni di protocollo degli strati superiori.



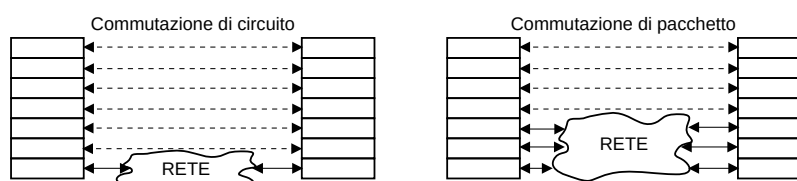
Indipendenza dei servizi tra pari rispetto al servizio di collegamento Quando uno strato affida il collegamento con un suo pari allo strato inferiore, quest'ultimo maschera al superiore la modalità con cui viene realizzato il trasferimento.

In particolare, se ci riferiamo all'interfaccia tra strato di trasporto e strato di rete, quest'ultimo può realizzare con il suo pari collegamenti con o senza connessione, offrendo allo stesso tempo (ma in modo indipendente) allo strato di trasporto un servizio con o senza connessione, dando luogo alle seguenti 4 possibilità:

Servizio di rete	SERVIZIO DI TRASPORTO	
	CIRCUITO VIRTUALE	DATAGRAMMA
Circuito Virtuale	SNA, X.25	Insolito
Datagramma	Arpanet, TCP/IP	Decnet

SNA (SYSTEM NETWORK ARCHITECTURE) è una architettura proprietaria IBM, in cui il trasferimento avviene in modo ordinato, richiedendo al livello di trasporto un circuito virtuale, che è realizzato da una serie di canali virtuali tra i nodi di rete. La stessa architettura è adottata anche dall'X.25, che costituisce l'insieme di protocolli che descrivono il funzionamento di reti pubbliche a commutazione di pacchetto, presenti in tutto il mondo: quella italiana prende il nome di ITAPAC.

³⁶Il modo di trasferimento è completamente definito dopo che sia stato specificato in quale strato siano svolte le funzioni di commutazione e multiplazione. In una rete a commutazione di circuito, queste sono realizzate dallo strato fisico che, esaurita la fase di instradamento ed impegno di risorse fisiche, collega in modo trasparente sorgente e destinazione. Nella commutazione di pacchetto, invece, le funzioni di multiplazione e commutazione coinvolgono (per tutti i pacchetti del messaggio) tutti i nodi di rete interessati; si dice pertanto che i protocolli di collegamento e di rete devono essere *terminati* (nel senso di gestiti) da tutti i nodi di rete.



Arpanet è l'architettura di Internet, in cui sebbene lo strato di rete operi con principio di commutazione a datagramma, mediante il protocollo *IP* (INTERNET PROTOCOL), lo strato di trasporto (*TCP*, TRANSFER CONTROL PROGRAM) offre a quelli superiori un servizio con connessione, attuato mediante circuiti virtuali, in modo da garantire il corretto sequenziamento delle unità informative, ed offrire canali di comunicazione formalmente simili ai files presenti localmente su disco. Il mascheramento del servizio di rete interna a datagramma in un servizio con connessione avviene a carico dello strato *TCP* di trasporto presente nei nodi terminali, che appunto affronta il riassettaggio ordinato dei datagrammi ricevuti dallo strato di rete.

Decnet è (o meglio era) l'architettura Digital, in cui il controllo di errore, la sequenzializzazione, ed il controllo di flusso sono realizzati dal livello di trasporto.

Soluzione insolita non è praticata perché equivale a fornire alla rete pacchetti disordinati, farli consegnare nello stesso identico disordine a destinazione, dove poi sono riassettrati. Può avere un senso se la comunicazione è sporadica, ma sempre per la stessa destinazione, nel qual caso somiglia ad un circuito virtuale permanente.

6.6 Appendici

In questa sezione sono raccolte le descrizioni di sistemi telematici che costituiscono l'immediata applicazione dei principi esposti nel capitolo. Il numero di esempi trattati crescerà con l'evolvere delle edizioni del testo.

Iniziamo subito con il dire che il modello a strati ISO-OSI è una astrazione concettuale utile per individuare raggruppamenti di funzioni, e serve ottimamente come modello per stimolare l'interoperabilità di apparati di diversi costruttori. D'altra parte, alcune delle realizzazioni esistenti (come ad esempio INTERNET) si sono sviluppate precedentemente alla definizione del modello, mentre altre (come ATM) seguono filosofie che solo successivamente sono state incorporate nel modello di riferimento. Pertanto, utilizzeremo le classificazioni ISO-OSI come parametro di riferimento, mediante il quale analizzare le funzioni delle reti reali.

6.6.1 La rete Internet

6.6.1.1 Storia

Nel 1964 L. Kleinrock (UCLA) propone un modello di rete non gerarchica e con parti ridondanti, che attuasse una modalità di trasferimento senza connessione e senza garanzie di qualità del servizio, rimandando queste ultime ai livelli superiori dell'architettura protocollare. Tale tipologia di servizio è oggi indicata con il termine *best effort*³⁷. Nel '69 sono operativi cinque nodi nelle università americane, e nel '72 avviene la prima dimostrazione pubblica di ARPANET, basata su NCP. Nel '73 Kahn e Cerf iniziano a definire TCP, da cui viene successivamente separato l'IP per la convenienza di non dover necessariamente aprire sempre una connessione. Fino all'80, il DoD³⁸ sovvenziona le università per implementare in ambiente UNIX i protocolli, che nel frattempo si

³⁷Migliore sforzo, ossia la rete dà il massimo, senza però garantire nulla.

³⁸Department of Defense.

vanno arricchendo di servizi, mentre la trasmissione Ethernet (del 1973) è adottata per realizzare LAN.

Nel 1983 il DoD decreta che tutti i calcolatori connessi a ARPANET adottino i protocolli TCP/IP, e separa la rete in due parti: una civile (ARPANET) ed una militare (MILNET). Negli anni seguenti i finanziamenti della National Science Foundation permettono lo sviluppo di una rete di trasporto a lunga distanza e di reti regionali, che interconnettono LAN di altre università e di enti di ricerca alla rete ARPANET, alla quale si collegano poi anche le comunità scientifiche non americane.

Nel 1990 ARPANET cessa le sue attività, e Barners-Lee (CERN) definisce il WWW, mentre nel '93 Andreessen (NCSA) sviluppa il primo *browser* WWW. Dal 1995 l'NSF non finanzia più la rete di interconnessione, ed il traffico inizia ad essere trasportato da operatori privati.

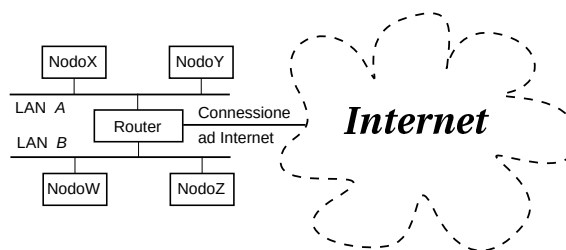
6.6.1.2 Le caratteristiche

La parola *Internet* in realtà è composta da due parole, INTER e NET, in quanto le caratteristiche della rete Internet sono quelle di fondere in una unica architettura una infinità di singole reti locali, potenzialmente disomogenee, e permettere la comunicazione tra i computer delle diverse sottoreti.

Ogni nodo della rete è connesso ad una rete locale (LAN³⁹), la quale a sua volta è interconnessa ad Internet mediante dei nodi detti *router*⁴⁰ che sono collegati ad una o più LAN e ad Internet, e svolgono la funzione di instradare le comunicazioni verso l'esterno. L'instradamento ha luogo in base ad un *indirizzo IP*⁴¹, che individua i singoli nodi in modo univoco su scala mondiale.

Come anticipato, lo strato di rete (o strato IP) realizza un modo di trasferimento a datagramma e non fornisce garanzie sulla qualità di servizio (QoS, QUALITY OF SERVICE) in termini di ritardi, errori e pacchetti persi. La situazione è mitigata dalla strato di trasporto (TCP, TRANSMISSION CONTROL PROTOCOL) che offre ai processi applicativi un servizio a circuito virtuale.

I protocolli di Internet sono realizzati in software e sono pubblici; gli utenti stessi e molte sottoreti private contribuiscono significativamente al trasporto, all'indirizzamento, alla commutazione ed alla notifica delle informazioni. Queste sono alcune ragioni fondamentali per cui Internet *non è di nessuno* ed è un patrimonio dell'umanità.



³⁹LOCAL AREA NETWORK, ossia *rete locale*. Con questo termine si indica un collegamento che non si estende oltre (approssimativamente) un edificio.

⁴⁰La funzione di conversione di protocollo tra reti disomogenee è detta di *gateway*, mentre l'interconnessione tra reti locali è svolta da dispositivi *bridge* oppure da *ripetitori* se le reti sono omogenee. Con il termine *router* si indica più propriamente il caso in cui il nodo svolge funzioni di instradamento, che tipicamente avviene nello *strato di rete*. Nel caso in cui invece si operi un instradamento a livello dello *strato di collegamento*, ossia nell'ambito di sezioni diverse (collegate da bridge o ripetitori) di una stessa LAN, il dispositivo viene detto *switch*. Infine, un *firewall* opera a livello di trasporto, e permette di impostare *regole di controllo* per restringere l'accesso alla rete interna in base all'indirizzo di *sorgente*, al tipo di *protocollo*, e/o a determinati *servizi*.

⁴¹IP = Internet Protocol.

6.6.1.3 Gli indirizzi

Strato	Indirizzo
Applicazione	<i>prot://nodo.dominio.tld</i>
Trasporto	<i>socket TCP <=> prot</i>
Rete	<i>Indirizzo IP x.y.w.z</i>
Collegamento	<i>a:b:c:d:e:f (Ethernet)</i>

Iniziamo l'argomento discutendo subito la stratificazione degli indirizzi coinvolti in una comunicazione via Internet. Ogni livello funzionale infatti utilizza le proprie convenzioni di indirizzamento, come illustrato nella tabella a fianco. Se a prima vista questa abbon-

danza di indirizzi può apparire esagerata, è proprio in questo modo che si realizza l'interoperabilità tra ambienti di rete differenti.

IP ed Ethernet I computer connessi ad Internet (detti *nodi*) sono le sorgenti e le destinazioni dell'informazione, e sono individuati da un *indirizzo IP*, che consiste in un gruppo di 4 byte⁴² e che si scrive *x.y.w.z* con ognuna delle 4 variabili pari ad un numero tra 0 e 255.

I nodi sono connessi alla rete mediante una interfaccia a volte indicata come MAC (MEDIA ACCESS CONTROL). Prendendo come esempio⁴³ i nodi connessi ad una LAN Ethernet, l'interfaccia di rete è individuata a sua volta da un *indirizzo Ethernet* composto da 6 byte. Quest'ultimo è unico in tutto il mondo, ed impresso dal costruttore nella scheda di interfaccia. L'indirizzo Ethernet viene però utilizzato solo nell'ambito della LAN di cui il nodo fa parte, ossia dopo che i pacchetti sono stati instradati dai router, per mezzo dell'indirizzo IP, verso la LAN.

Sottoreti Ogni nodo conosce, oltre al proprio indirizzo IP, anche una *maschera di sottorete* composta da una serie di uni seguita da zeri, in numero complessivo di 32 bit, tanti quanti ne sono presenti nell'indirizzo IP. Il termine *maschera* è dovuto all'operazione di AND binario (vedi tabella) operata tra la maschera e gli indirizzi IP, per determinare se questi appartengano alla propria stessa LAN oppure risiedano altrove.

Indirizzo IP	Maschera Sottorete	Indirizzo sottorete
151.100.8.33	255.255.255.0	151.100.8.0

Nel caso in cui la sottorete di un nodo Y verso cui il nodo X deve inviare un pacchetto è la stessa su cui è connesso X, allora questi può individuare l'indirizzo Ethernet del destinatario⁴⁴ ed inviargli il pacchetto direttamente. In caso contrario, X invierà il pacchetto al proprio *gateway* verso Internet.

Intranet Alcuni gruppi di indirizzi IP (come quelli 192.168.x.y oppure 10.x.y.z) non vengono instradati dai router, e possono essere riutilizzati nelle *reti private* di tutto

⁴²Con 4 byte si indirizzano (in linea di principio) $2^{32} = 4.29 \cdot 10^9$ diversi nodi (più di 4 miliardi). E' tuttora in sviluppo il cosiddetto IPv6, che estenderà l'indirizzo IP a 16 byte, portando la capacità a $3.4 \cdot 10^{38}$. L'IPv6 prevede inoltre particolari soluzioni di suddivisione dell'indirizzo, allo scopo di coadiuvare le operazioni di *routing*.

⁴³Evidentemente esistono molte diverse possibilità di collegamento ad Internet, come via telefono (tramite provider), collegamento satellitare, Frame Relay, linea dedicata, ISDN... ma si preferisce svolgere un unico esempio per non appesantire eccessivamente l'esposizione. La consapevolezza delle molteplici alternative consente ad ogni modo di comprendere la necessità di separare gli strati di trasporto e di rete dall'effettiva modalità di trasmissione.

⁴⁴Mostriamo in seguito che questo avviene mediante il protocollo ARP.

il mondo per realizzare le cosiddette *reti intranet* operanti con gli stessi protocolli ed applicativi che funzionano via Internet.

DNS L'utente di una applicazione internet in realtà non è a conoscenza degli indirizzi IP dei diversi nodi, ma li identifica per mezzo di nomi simbolici del tipo *nodo.dominio.tld*, detti anche *indirizzi Internet*. Il processo di risoluzione che individua l'indirizzo IP associato al nome avviene interrogando un particolare nodo, il DOMAIN NAME SERVER (*servente dei nomi di dominio*). La struttura dei nomi, scandita dai punti, individua una gerarchia di autorità per i diversi campi. Il campo *tld* è chiamato *dominio di primo livello* (TOP LEVEL DOMAIN⁴⁵), mentre il campo *dominio* in genere è stato registrato da qualche organizzazione che lo giudica rappresentativo della propria offerta informativa. Il campo *nodo* rappresenta invece una ben determinata macchina, il cui indirizzo Internet completo è *nodo.dominio.tld*, e che non necessariamente è collegato alla stessa LAN a cui sono connessi gli altri nodi con indirizzo che termina per *dominio.tld*.

Quando un *nodoX* generico deve comunicare con *nodoY.dominio.tld*, interroga il proprio DNS⁴⁶ per conoscerne l'IP. Nella rete sono presenti molti DNS, alcuni dei quali detengono informazioni *autorevoli*⁴⁷ riguardo ai nodi di uno o più domini, altri (i DNS *radice*, o ROOT) detengono le informazioni relative a quali DNS siano autorevoli per i domini di primo livello, ed altri fanno da tramite tra i primi due ed i *client* che richiedono una risoluzione di indirizzo. Se il DNS di *nodoX* non è autorevole per *nodoY*, allora⁴⁸ provvede ad inoltrare la richiesta, interrogando prima un DNS radice per individuare chi è autorevole per *.tld*, quindi interroga questo per trovare chi è autorevole per *.dominio.tld*, e quindi usa la risposta ottenuta per dirigere la richiesta di risoluzione originaria. Se la cosa può sembrare troppo macchinosa per funzionare bene, è perchè la stessa sequenza di operazioni *non deve* essere effettuata sempre: il DNS utilizzato da *nodoX* riceve infatti, assieme all'IP di *nodoY*, anche una informazione detta TIME TO LIVE (TTL o *tempo di vita*) che descrive la scadenza della coppia *nome-IP* ottenuta. Genericamente il TTL è di qualche giorno, e fino alla sua scadenza il DNS *ricorda*⁴⁹ la corrispondenza, in modo da fornire la propria copia in corrispondenza delle richieste future, e ridurre sensibilmente il traffico legato alla risoluzione degli indirizzi Internet. L'insieme delle risoluzioni apprese è denominata *cache del DNS*⁵⁰.

⁴⁵I top level domain possono essere pari ad un identificativo geografico (.it, .se, .au...) od una delle sigle .com, .org, .net, .mil, .edu, che sono quelle utilizzate quando internet era solo americana.

⁴⁶Il "proprio" DNS viene configurato per l'host in modo fisso, oppure in modo dinamico da alcuni Service Provider raggiungibili per via telefonica, e convenientemente corrisponde ad un nodo situato "vicino" al nodo che lo interroga.

⁴⁷Chi registra il dominio deve disporre necessariamente di un DNS in cui inserire le informazioni sulle corrispondenze tra i nomi dei nodi del proprio dominio ed i loro corrispondenti indirizzi IP. In tal caso quel DNS si dice autorevole per il dominio ed è responsabile di diffondere tali informazioni al resto della rete.

⁴⁸In realtà esiste anche una diversa modalità operativa, che consiste nel delegare la ricerca ad un diverso DNS (detto *forwarder*), il quale attua lui i passi descritti appresso, e provvede per proprio conto alla risoluzione, il cui esito è poi comunicato al primo DNS e da questi ad *hostX*. Il vantaggio di tale procedura risiede nella maggiore ricchezza della *cache* (descritta appresso) di un DNS utilizzato intensivamente.

⁴⁹Il DNS ricorda anche le altre corrispondenze ottenute, come il DNS autorevole per *.tld* e per *.dominio.tld*; nel caso infine in cui si sia utilizzato un forwarder, sarà quest'ultimo a mantenere memoria delle corrispondenze per i DNS intermedi.

⁵⁰CACHE è un termine generico, che letteralmente si traduce *nascondiglio dei viveri*, e che viene adottato ogni volta si debba indicare una memoria che contiene copie di riserva, o di scorta...

Indirizzi TCP Si è detto che ogni nodo è individuato in Internet mediante il proprio indirizzo IP, ma questo non è sufficiente ad indicare con quale particolare programma (che può essere uno specifico *server* come nel caso del DNS) si vuole entrare in comunicazione. I programmi che sono pronti a ricevere connessioni si pongono *in ascolto* su ben determinate *porte* (o *socket*⁵¹), identificate da numeri⁵², e che sono referenziati in modo simbolico (es. *http://*, *ftp://*) dagli applicativi di utente che si rivolgono allo strato di trasporto (il TCP) per stabilire un collegamento con un server presente su di un nodo remoto.

Alcuni servizi rispondono ad indirizzi *ben noti*, fissi per tutti i nodi, in quanto il chiamante deve sapere a priori a quale porta connettersi. Il nodo contattato invece, apre con il chiamante una connessione di ritorno, su di un numero di porta diverso, che è stato comunicato dal chiamante al momento della richiesta di connessione, e per il quale il chiamante non ha già aperto altre connessioni differenti.

6.6.1.4 TCP

Discutiamo ora del TCP⁵³, che offre ai processi applicativi un servizio di trasporto a circuito virtuale, *attaccato* ad una porta⁵⁴ di un nodo remoto individuato dall'indirizzo IP. Il suo compito è quello di ricevere dai processi applicativi dei dati, suddividerli in pacchetti, ed inviarli al suo pari che svolge il processo inverso.

Il pacchetto TCP

1		8		16		24	
Porta Sorgente				Porta Destinazione			
Numero di Sequenza (Tx)							
Numero di Riscontro (Rx)							
Offset		Riser.	Contr.	Finestra			
Checksum				Puntatore Urgente			
Opzioni				Riempimento			
Dati							
Dati							
...							

La struttura di un pacchetto TCP è mostrata in figura, e comprende una intestazione composta da 6 gruppi (o più) di 4 byte per un minimo di 192 bit, a cui segue un numero variabile di gruppi di 4 byte di Dati, provenienti dagli strati applicativi superiori.

Troviamo subito i numeri delle porte a cui si riferisce la connessione, mentre gli indirizzi IP sono aggiunti dallo strato di rete. I numeri di *Sequenza* e di *Riscontro* servono rispettivamente a numerare i

bytes dei pacchetti uscenti, ed a notificare l'altro lato del collegamento del numero di sequenza del prossimo byte che si aspetta di ricevere⁵⁵, riscontrando implicitamente come

⁵¹*Socket* è un termine che corrisponde alla... presa per l'energia elettrica casalinga, ed in questo contesto ha il significato di una *presa* a cui si "attacca" il processo che richiede la comunicazione.

⁵²Spesso gli indirizzi che identificano i punti di contatto di servizi specifici vengono indicati come SERVICE ACCESS POINT (SAP), anche per situazioni differenti dal caso specifico delle porte del TCP.

⁵³TCP = *Transport Control Protocol*.

⁵⁴Il numero di porta costituisce in pratica l'*identificativo di connessione* del circuito virtuale. Nel caso in cui un server debba comunicare con più client, dopo avere accettato la connessione giunta su di una *porta ben nota*, apre con i client diversi canali di ritorno, differenziati dall'uso di porte di risposta differenti.

La lista completa dei servizi standardizzati e degli indirizzi ben noti (*socket*) presso i quali i serventi sono in attesa di richieste di connessione, è presente in tutte le distribuzioni Linux nel file */etc/services*.

⁵⁵Il numero di sequenza si incrementa ad ogni pacchetto di una quantità pari alla sua dimensione in bytes, ed ha lo scopo di permettere le operazioni di controllo di flusso. Il valore iniziale del numero di

correttamente arrivati i pacchetti con numero di sequenza più basso.

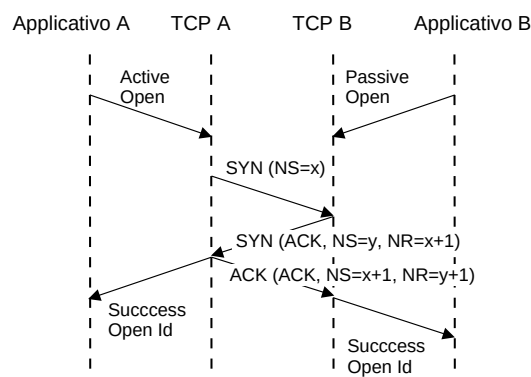
Offset (4 bit) codifica il numero di parole da 4 byte dell'intestazione, mentre nei 6 bit *Riservati* non è mai stato scritto nulla. I 6 bit del campo *Controllo* hanno ognuno un nome ed un significato preciso, qualora posti ad uno. Il primo (URG) indica che il campo urgent pointer contiene un valore significativo; ACK indica che si sta usando il Numero di Riscontro; PSH indica un pacchetto urgente che non può rispettare la coda in ricezione; RST segnala un malfunzionamento e impone il reset della connessione; SYN è pari ad uno solo per il primo pacchetto inviato per richiedere di creare una connessione; FIN indica che la sorgente ha esaurito i dati da trasmettere.

I 16 bit di *Finestra* rappresentano il numero di byte che, a partire dal numero espresso dal *Numero di Riscontro*, il trasmittente è in grado di ricevere, ed il suo utilizzo sarà meglio illustrato tra breve nel contesto del controllo di flusso. Il *Checksum* serve al ricevente per verificare se si sia verificato un errore, il *Puntatore Urgente* contiene il numero di sequenza dell'ultimo byte di una sequenza di dati urgenti, e le *Opzioni* (di lunghezza variabile) sono presenti solo raramente, ed utilizzate a fini di controllo, ad esempio per variare la dimensione della finestra. Infine, il *Riempimento* conclude l'ultima parola da 32 bit.

Uno stesso pacchetto TCP può svolgere solo funzioni di segnalazione, o di sola trasmissione dati, od entrambe.

Apertura e chiusura della connessione Il TCP offre un servizio di circuito virtuale, e quindi prima di inviare dati, deve effettuare un colloquio iniziale con il nodo remoto di destinazione. In particolare, il colloquio ha lo scopo di accertare la disponibilità del destinatario ad accettare la connessione, e permette alle due parti di scambiarsi i rispettivi numeri di sequenza descritti alla nota 55.

Dato che anche il “chiamante” deve riscontrare il numero che il chiamato ha fornito, occorrono tre pacchetti per terminare il dialogo, che prende il nome di THREE WAY HANDSHAKE⁵⁶. Il diagramma a lato mostra l'evoluzione temporale del colloquio tra processo applicativo che intende comunicare (A), ed un secondo che si pone in ascolto (B), mostrando come al primo SYN che pone $NS_A = x$, ne segua un altro che pone $NS_B = y$, seguito a sua volta dall'ACK di chi ha iniziato⁵⁷. La chiusura può avvenire per diverse cause: o perchè è termi-



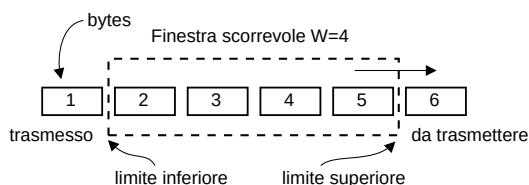
sequenza e di riscontro è diverso per ogni connessione, e generato in modo pseudo-casuale da entrambe le parti in base ai propri orologi interni, allo scopo di minimizzare i problemi dovuti all'inaffidabilità dello strato di rete (IP) che può perdere o ritardare i datagrammi, nel qual caso il TCP trasmittente ri-invia i pacchetti precedenti dopo un time-out. Questo comportamento può determinare l'arrivo al lato ricevente di un pacchetto duplicato, e consegnato addirittura dopo che la connessione tra i due nodi è stata chiusa e riaperta. In tal caso però la nuova connessione adotta un diverso numero di sequenza iniziale, cosicchè il pacchetto duplicato e ritardato risulta fuori sequenza, e non viene accettato.

⁵⁶HANDSHAKE = *stretta di mano*.

⁵⁷Per ciò che riguarda i valori dei numeri di riscontro NR , questi sono incrementati di 1, perchè la *finestra* (descritta nel seguito) inizia dai bytes del prossimo pacchetto, a cui competeranno appunto valori di NS incrementati di uno.

nato il messaggio, segnalato dal bit FIN, o per situazioni anomale, che il TCP indica con il bit RST.

Protocollo a finestra Allo scopo di realizzare un controllo di flusso, il TCP prevede l'uso dell'*NR* inviato dal ricevente per dosare il ritmo con cui trasmettere i propri pacchetti.



La lunghezza di *Finestra* comunicata con il SYN del ricevente, determina la quantità di memoria riservata per i buffer dedicati alla connessione, che viene gestita come una memoria a scorrimento o *finestra scorrevole* (SLIDING WINDOW). Questa memoria è presente per ge-

stire i casi di pacchetti ritardati o fuori sequenza, e contiene i bytes già trasmessi. Il trasmittente (vedi figura) non fa avanzare il limite inferiore finché non riceve un riscontro con *NR* maggiore di tale limite. In questo modo non occorre attendere il riscontro di tutti i bytes, o di tutti i pacchetti (che devono comunque essere di dimensione inferiore alla finestra), ma ci può avvantaggiare trasmettendo l'intero contenuto della finestra.

Una finestra del tutto analoga è utilizzata dal ricevente, allo scopo di ricomporre l'ordine originario dei pacchetti consegnati disordinatamente dallo strato IP di rete. Non appena il ricevente completa un segmento contiguo al limite inferiore, sposta quest'ultimo in avanti di tanti bytes quanti ne è riuscito a leggere in modo contiguo, ed invia un riscontro con *NR* pari al più basso numero di byte che ancora non è pervenuto⁵⁸.

Nel caso in cui sia settato il bit URG⁵⁹, si stanno inviando dati urgenti fuori sequenza, e che non devono rispettare il protocollo a finestra, come ad esempio per recapitare un segnale di interrupt relativo ad una sessione Telnet per terminare una applicazione remota.

Controllo di errore Trascorso un certo tempo (detto *timeout*) nell'attesa di un riscontro, il trasmittente ritiene che alcuni pacchetti sono andati persi, e li re-invia⁶⁰. Il valore del *timeout* viene calcolato dinamicamente dal TCP in base alle sue misure di *round-trip delay*⁶¹, ossia del tempo che intercorre in media tra invio di un pacchetto e ricezione del suo riscontro. In questo modo il TCP si adatta alle condizione di carico della rete ed evita di ri-spediti pacchetti troppo presto o di effettuare attese inutili. In particolare, nel caso di rete congestionata aumenta la frequenza dei pacchetti persi, e valori di *timeout* troppo ridotti potrebbero peggiorare la situazione.

⁵⁸Il riscontro può viaggiare su di un pacchetto già in "partenza" con un carico utile di dati e destinato al nodo a cui si deve inviare il riscontro. In tal caso quest'ultimo prende il nome di PIGGYBACK (*rimorchio*), o *riscontro rimorchiato*.

⁵⁹In tal caso, il campo *Puntatore Urgente* contiene il numero di sequenza del byte che delimita superiormente i dati che devono essere consegnati urgentemente.

⁶⁰Il mancato invio del riscontro può anche essere causato dal verificarsi di un *checksum* errato dal lato ricevente, nel qual caso quest'ultimo semplicemente evita di inviare il riscontro, confidando nella ritrasmissione per timeout.

⁶¹Con licenza poetica: *il ritardo del girotondo*, che qui raffigura un percorso di andata e ritorno senza soste.

Controllo di flusso Il meccanismo a finestra scorrevole determina, istante per istante, il numero massimo di bytes che possono essere trasmessi verso il destinatario, e pertanto consente al nodo meno veloce di adeguare la velocità di trasmissione alle proprie capacità. La dimensione della finestra può essere variata (su iniziativa del ricevente) nel corso della connessione, in accordo al valore presente nel campo *Finestra* dell'intestazione TCP. Ad esempio, una connessione può iniziare con una dimensione di finestra ridotta, e poi aumentarla nel caso in cui non si verifichino errori, la rete sopporti il traffico, ed i nodi abbiano memoria disponibile.

Controllo di congestione Il TCP può usare la sua misura di *round-trip delay* come un indicatore di congestione della rete, e lo scadere di un *timeout* come un segnale del peggioramento della congestione. In tal caso quindi, può essere ridotta la dimensione della finestra di trasmissione, caricando così di meno la rete.

UDP Lo *User Datagram Protocol* è ancora un protocollo di trasporto, che opera senza connessione, e sostituisce il TCP per inviare pacchetti isolati, o serie di pacchetti la cui ritrasmissione (se perduti) sarebbe inutile. Ad esempio, è utilizzato nella trasmissione di dati in tempo reale, oppure per protocolli di interrogazione e controllo come il DNS.

6.6.1.5 IP

L'*Internet Protocol* costituisce l'ossatura della rete internet, realizzandone i servizi di rete ed interfacciando le diverse sottoreti a cui sono connessi i nodi. Le sue principali funzioni sono pertanto l'indirizzamento, l'instradamento e la variazione della dimensione⁶² dei pacchetti prodotti dal TCP o da altri protocolli degli strati superiori. Ogni pacchetto è inviato come un messaggio indipendente, in modalità datagramma; la consegna dei datagrammi non è garantita⁶³, e questi possono essere persi, duplicati o consegnati fuori sequenza.

L'IP riceve dallo strato superiore (il TCP od un altro protocollo) un flusso di byte suddivisi in pacchetti, a cui si aggiunge l'indirizzo IP di destinazione; tale flusso è utilizzato per riempire un proprio buffer di dimensione opportuna, che quando pieno (od al termine del pacchetto ricevuto *dall'alto*) è *incapsulato* aggiungendo una intestazione (*l'header*) che codifica la segnalazione dello strato di rete realizzato dal protocollo IP.

L'intestazione IP contiene le informazioni mostrate alla figura successiva.

Il campo *VER* indica quale versione si sta utilizzando, e permette sperimentazioni e miglioramenti senza interrompere il servizio. *HLEN* e *TLEN* indicano rispettivamente la lunghezza dell'header e di tutto il pacchetto, mentre *TOS* codifica un *Type of Service*

⁶²L'IP può trovarsi a dover inoltrare i pacchetti su sottoreti che operano con dimensioni di pacchetto inferiori. Per questo, deve essere in grado di frammentare il pacchetto in più datagrammi, e di ricomporli nell'unità informativa originaria all'altro estremo del collegamento.

⁶³Si suppone infatti che le sottoreti a cui sono connessi i nodi non garantiscano affidabilità. Ciò consente di poter usare sottoreti le più generiche (incluse quelle affidabili, ovviamente).

per differenziare ad esempio la QoS⁶⁴ richiesta. L'*identificazione* riporta lo stesso valore per tutti i frammenti di uno stesso datagramma, mentre l'*Offset di frammento* indica la posizione del frammento nel datagramma (con frammenti di dimensione multipla di 8 byte).

1	5	9	17	20	32
VER	HLEN	TOS	TLEN		
Identificazione			Flags	Frag. Offset	
TTL	Protocollo		Checksum		
IP Address Sorgente					
IP Address Destinazione					
Opzioni			Riempimento		

Solo 2 dei tre bit di *Flag* sono usati, DF (*Don't Fragment*) per richiedere alla rete di non frammentare il datagramma, e MF (*More Fragments*) per indicare che seguiranno altri frammenti. Il *TTL* (Time To Live) determina la massima permanenza del pacchetto nella rete⁶⁵, il *protocollo* indica a chi consegnare il datagramma all'arrivo (ad es. TCP o UDP), e

Checksum serve per verificare l'assenza di errori nell'header⁶⁶.

Gli *Indirizzi IP* di sorgente e destinazione hanno l'evidente funzione di recapitare correttamente il messaggio, mentre il campo *Opzioni* ha una lunghezza variabile, può essere omesso, e consente ad esempio di richiedere il tracciamento della serie di router attraversati.

Indirizzamento e Routing A pagina 136 si è anticipata la relazione che lega la parte iniziale dell'indirizzo IP ad una determinata sottorete, in modo da partizionare i 2³² indirizzi su di una gerarchia a due livelli e delegare la consegna all'host finale ad uno o più router responsabili di servire la sottorete⁶⁷. In realtà la gerarchia presenta una ulteriore suddivisione, dettata sia da esigenze amministrative che funzionali.

I bit più significativi dell'indirizzo IP identificano 5 diversi gruppi (o *classi*) di indirizzi, descritti dalla seguente tabella:

Inizio IP addr	Classe	bit rete/nodo	N. reti	N. nodi per rete
0	A	7/24	128	16 777 216
10	B	14/16	16 384	65 536
110	C	21/8	2 097 152	256
1110	D	28 bit di indirizzo multicast per 268 435 456 canali		
11110	E	27 bit per usi futuri e ricerca		

Quando una organizzazione decide di essere presente in internet, richiede l'assegna-

⁶⁴La Qualità del Servizio richiesta per il particolare datagramma può esprimere necessità particolari, come ad esempio il ritardo massimo di consegna. La possibilità di esprimere questa esigenza a livello IP fa parte dello standard, ma per lunghi anni non se ne è fatto uso. L'avvento delle comunicazioni multimediali ha risvegliato l'interesse per il campo TOS.

⁶⁵Lo scopo del TTL è di evitare che si verifichino fenomeni di loop infinito, nei quali un pacchetto "rimbalza" tra due nodi per problemi di configurazione. Per questo, TTL è inizializzato al massimo numero di nodi che il pacchetto può attraversare, e viene decrementato da ogni nodo che lo riceve (e ritrasmette). Quando TTL arriva a zero, il pacchetto è scartato.

⁶⁶In presenza di un frammento ricevuto con errori nell'header, viene scartato tutto il datagramma di cui il frammento fa parte, delegando allo strato superiore le procedure per l'eventuale recupero dell'errore.

⁶⁷Possiamo portare come analogia un indirizzo civico, a cui il postino consegna la corrispondenza, che viene poi smistata ai singoli condomini dal portiere dello stabile. Il servizio postale, così come la rete Internet, non ha interesse di sapere come sono suddivise le sottoreti delle diverse organizzazioni, ed i router instradano i pacchetti IP in base alla parte "rete" dell'indirizzo, delegando ai router della rete di destinazione il completamento dell'instradamento.

zione di un lotto di indirizzi IP ad apposti organismi, i quali attribuiscono all'organizzazione un gruppo di indirizzi di classe A, B o C in base al numero di nodi che l'organizzazione prevede di mettere in rete. Una rete in classe B ad esempio è individuata da 14 bit (ossia, assieme ai bit di classe, dai primi due bytes dell'indirizzo IP), e quindi esistono $2^{14} = 16384$ diverse reti in classe B, ognuna con una capacità di $2^{16} = 65536$ diversi nodi. Chi è intestatario di un gruppo di indirizzi, provvede ad assegnarli ai singoli nodi della propria sottorete.

Subnetting e Supernetting Osserviamo ora che la maschera di sottorete presentata a pag. 136 *non* coincide con il gruppo di bit che identifica la classe e la rete: infatti, l'insieme di indirizzi 151.100.x.y corrisponde ad una rete in classe B, mentre la maschera di sottorete 255.255.255.0 individua una sottorete in classe C. Praticamente, la rete in classe B è stata ulteriormente suddivisa (*subnettata*) in 256 sottoreti di classe C, permettendo di realizzare un instradamento gerarchico su due livelli nell'ambito dell'organizzazione intestaria della rete in classe B⁶⁸. L'operazione inversa (detta *supernetting*), ossia quella di aggregare più reti di dimensione ridotta in una di dimensione maggiore, ha senso all'interno del router che instrada il traffico verso l'organizzazione intestataria delle sottoreti, in quanto permette di ridurre la dimensione delle tabelle di routing, che contengono così un solo elemento relativo alla super-rete, anziché un elemento per ogni singola sottorete.

Classless Interdomain Routing - CDIR Nella prima metà degli anni '90 apparve evidente che il partizionamento degli indirizzi nelle tre classi A, B e C non era rispondente alle richieste dell'utenza; accadeva infatti che le reti in classe C erano troppo "piccole", mentre quelle in classe B rischiavano di esaurirsi a breve, pur essendo sfruttate molto poco⁶⁹. Per questo motivo, è stata rimossa la suddivisione rigida nelle tre classi, e si è applicato sistematicamente il principio del supernetting. In pratica, si è ridefinita la maschera di sottorete, come una sequenza di *uni* allineata a sinistra, permettendo così di definire reti di dimensione pari a una potenza di due qualsiasi. Come risultato, ora una sottorete è identificata da una coppia indirizzo/maschera del tipo (ad es.) 172.192.0.0/12, che rappresenta tutti 2^{20} indirizzi che vanno da 172.192.0.0 a 172.207.255.255, che hanno i 12 bit più elevati uguali a 101011001100: questa sequenza prende il nome di *prefisso* della rete. In definitiva quindi, la maschera è espressa come il numero di bit più significativi in comune a tutti i nodi della sottorete.

Longest Match Un router decide su che porta instradare un pacchetto IP in base al confronto tra l'indirizzo di destinazione e tutti i prefissi presenti nella tabella di routing, associati ciascuno alla "migliore" porta di uscita verso la sottorete definita dal prefisso. Nel caso in cui si verifichi più di una uguaglianza, si sceglie l'instradamento caratterizzato dal *maggior numero* di bit coincidenti, ossia relativo al prefisso *più lungo*. Infatti, in tal modo viene preferita la direzione *più specifica* verso la destinazione finale. In as-

⁶⁸In questo caso, l'Università di Roma "La Sapienza" è intestataria della rete 151.100.

⁶⁹Ad esempio, organizzazioni con poco più di un migliaio di nodi erano costrette a richiedere una intera classe B con capacità di 65536 nodi.

senza di uguaglianze invece, il pacchetto è inoltrato in base ad una *default route*, che tipicamente rimanda la decisione ad un router “gerarchicamente più elevato”⁷⁰.

Sistemi Autonomi e Border Gateway Vi sono router collegati direttamente con le LAN, e configurati per instradare correttamente i pacchetti diretti a destinazioni locali. Vi sono poi router collegati solo ad altri router, che *apprendono* gli instradamenti verso le reti locali mediante appositi *protocolli di routing* che consentono ai router di primo tipo di *pubblicizzare* (ADVERTISE) le reti raggiungibili direttamente, ed ai router del secondo tipo di fare altrettanto nei confronti dei loro pari.

L'insieme di sottoreti (e router, nodi e DNS) gestite da una stessa organizzazione prende il nome di *Autonomous System* (AS), e nel suo ambito sono attivi protocolli di routing detti *Interior Gateway Protocols* (IGP), che distribuiscono le informazioni di raggiungibilità interna. Alcuni router di uno stesso AS svolgono il ruolo di *Border Gateway* (BG), e comunicano con i BG di altri AS mediante appositi *Exterior Gateway Protocols* (EGP), pubblicizzando all'esterno le proprie sottoreti, apprendendo dagli altri BG la raggiungibilità delle sottoreti esterne, e distribuendo tali informazioni ai router interni. Un compito particolare dell'EGP, è quello di attuare qualche politica nei confronti del *traffico di transito* tra due AS diversi dall'AS di cui il BG fa parte: in tal caso, il protocollo prende il nome di *Border Gateway Protocol* (BGP).

L'applicazione del CDIR comporta, per ogni scambio di informazioni di routing, la necessità di aggregare o disaggregare i prefissi di sottorete, in modo da mantenere al minimo la dimensione delle tabelle di instradamento.

Multicast Tornando all'esame della tabella di pag. 142, in cui la classe E costituisce evidentemente una “riserva” di indirizzi per poter effettuare sperimentazioni, la classe D individua invece dei canali *multicast*⁷¹. Quando un nodo decide di aderire ad un canale multicast, invia un messaggio⁷² in tal senso al proprio router più vicino, che a sua volta si occupa di informare gli altri router. Questi ultimi provvederanno quindi, qualora osservino transitare un pacchetto avente come destinazione un canale multicast, ad instradarlo verso l'host aderente. In presenza di più nodi nella stessa sottorete in ascolto dello stesso canale, solo una copia dei pacchetti attraverserà il router: il traffico multicast⁷³ evita infatti di aprire una connessione dedicata per ogni destinatario, ma si suddivide via via nella rete solo quando i destinatari sono raggiungibili da vie diverse.

6.6.1.6 Ethernet

Ci occupiamo qui di un caso particolare di realizzazione dei primi due livelli del modello ISO-OSI. Come anticipato a pag. 136, molti nodi di Internet sono univocamente individuati da un indirizzo (Ethernet) di 6 byte che, sebbene sia unico al mondo, viene usato

⁷⁰Sebbene la topologia di Internet possa essere qualunque, nella pratica esistono dei *carrier* internazionali che svolgono la funzione di *backbone* (spina dorsale) della rete, interconnettendo tra loro i continenti e le nazioni.

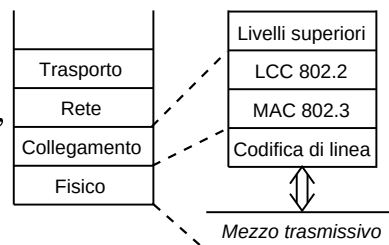
⁷¹Il termine *multicast* è ispirato alle trasmissioni *broadcast* effettuate dalle emittenti radio televisive.

⁷²Mediante il protocollo IGMP (*Internet Group Management Protocol*) che opera sopra lo strato IP, ma (a differenza del TCP) fa uso di datagrammi non riscontrati, similmente all'UDP ed all'ICMP.

⁷³Data l'impossibilità a stabilire un controllo di flusso con tutti i destinatari, il traffico multicast viaggia all'interno di pacchetti UDP.

solamente nell'ambito della LAN a cui il nodo è connesso, in quanto la distribuzione mondiale degli indirizzi Ethernet è casuale⁷⁴: se infatti questi fossero usati come indirizzi a livello di rete, le tabelle di instradamento dovrebbero essere a conoscenza di *tutti* i nodi esistenti⁷⁵. Puntualizziamo inoltre che un nodo di Internet può essere connesso alla rete anche per via telefonica, o con svariati altri metodi; ci limitiamo qui a descrivere il caso delle LAN Ethernet, peraltro particolarmente diffuso.

Ethernet individua un particolare tipo di pacchetto dati, adottato inizialmente dalla Xerox, adatto ad incapsulare dati provenienti da protocolli diversi. Successivamente, il formato è stato standardizzato dall'IEEE, e per ciò che ci interessa le specifiche sono quelle identificate dalle sigle 802.2 (LOGICAL LINK CONTROL, LCC) e 803.3 (CARRIER SENSE MULTIPLE ACCESS - COLLISION DETECT, CSMA/CD). La figura mostra il legame tra queste due sigle e gli strati del modello; lo strato MAC in cui si realizza il CSMA/CD individua il MEDIA ACCESS CONTROL.



Il mezzo trasmissivo è un cavo, coassiale o coppia simmetrica, sul quale sono collegati tutti nodi della LAN, che si *contendono* il mezzo trasmissivo, in quanto vi può trasmettere solo un nodo per volta. Inoltre, tutti i nodi sono in ascolto sullo stesso mezzo per ricevere i pacchetti a loro destinati, riconoscibili per la presenza del proprio indirizzo Ethernet nel campo destinazione. Un pacchetto Ethernet può inoltre riportare un indirizzo di destinazione particolare, detto di *Broadcast*, che obbliga *tutti* i nodi presenti alla ricezione del pacchetto.

Address Resolution Protocol - ARP Quando un pacchetto IP giunge ad un router, e l'indirizzo IP indica che il destinatario è connesso ad una delle LAN direttamente raggiungibili dal router⁷⁶, questo invia su quella LAN un pacchetto *broadcast*, su cui viaggia una richiesta ARP (ADDRESS RESOLUTION PROTOCOL), allo scopo di individuare l'indirizzo Ethernet del nodo a cui è assegnato l'indirizzo IP di destinazione del pacchetto arrivato al router. Se tale nodo è presente ed operativo, riconosce che la richiesta è diretta a lui, ed invia un pacchetto di risposta comunicando il proprio indirizzo Ethernet, che viene memorizzato dal router in una apposita tabella⁷⁷.

Operazioni simili sono svolte da ognuno dei nodi della LAN, ogni volta che debbano inviare un pacchetto ad un altro nodo direttamente connesso alla stessa rete locale. Se

⁷⁴E rappresenta quindi ciò che viene detto uno *spazio di indirizzi piatto* (FLAT ADDRESS SPACE).

⁷⁵Al contrario, il partizionamento dell'indirizzo IP in rete+nodo permette di utilizzare tabelle di routing di dimensioni gestibili.

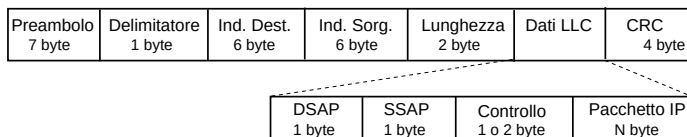
⁷⁶Ad ogni porta del router è associata una coppia sottorete/maschera (vedi pag. 136) che descrive l'insieme degli indirizzi direttamente connessi alla porta. La verifica di raggiungibilità (o *adiacenza*) è attuata mettendo in AND l'IP di destinazione con le maschere, e confrontando il risultato con quello DELL'AND tra le maschere e gli indirizzi delle sottoreti collegate.

⁷⁷Dato che i nodi possono essere spostati, possono cambiare scheda di rete e possono cambiare indirizzo IP assegnatogli, la corrispondenza IP-Ethernet è tutt'altro che duratura, ed ogni riga della tabella ARP indica anche quando si sia appresa la corrispondenza, in modo da poter stabilire una scadenza, ed effettuare nuovamente la richiesta per verificare se sono intervenuti cambiamenti topologici.

Se il nodo ha cambiato IP, ma non il nome, sarà il TTL del DNS (mantenuto aggiornato per il dominio del nodo) a provocare il rinnovo della richiesta dell'indirizzo.

al contrario l'IP di destinazione non fa parte della stessa LAN, il pacchetto è inviato al *default gateway*.

Formato di pacchetto Il pacchetto Ethernet è generato dall'LLC e dal MAC, ognuno dei quali incapsula il pacchetto IP con le proprie informazioni di protocollo.



Nella figura è mostrato il risultato finale delle operazioni. In testa troviamo 7 byte di *preambolo*, necessario a permettere la sincronizzazione dell'orologio del

ricevente con quello in trasmissione; dato che la sincronizzazione richiede un tempo non noto a priori, un byte di *flag* segnala l'inizio del pacchetto. Troviamo quindi gli *indirizzi Ethernet* di sorgente e destinazione, due byte che indicano la *lunghezza* della restante parte del pacchetto, e quindi l'incapsulamento dei dati prodotti dall'LLC. In fondo, sono presenti 4 byte che realizzano il *controllo di errore*.

L'LLC inserisce invece, in testa al pacchetto IP, due indirizzi *SAP* (SERVICE ACCESS POINT) di sorgente e destinazione, che rappresentano il codice del programma da cui proviene ed a cui è indirizzato il pacchetto; è qui che si specifica quale protocollo di strato superiore è incapsulato (ad es., IP od ARP). Nel campo di *controllo* possono essere anche ospitati numeri di sequenza per i casi che lo possano richiedere, ed infine troviamo il pacchetto IP originario.

Collisione Come anticipato, il mezzo trasmissivo è in comune con tutti i nodi, e dunque si è studiata una particolare soluzione il cui nome CSMA/CD indica che l'*Accesso Multiplo* avviene in due fasi: prima di trasmettere, si ascolta se non vi sia già qualcuno che trasmette (CARRIER SENSE), e durante la trasmissione, si verifica che nessun altro stia trasmettendo contemporaneamente (COLLISION DETECT). Pertanto, ogni nodo che debba trasmettere si pone prima in ascolto, e se osserva che già vi sono trasmissioni in corso, attende un tempo casuale e riprova. Quando trova il mezzo "libero", inizia a trasmettere, ma contemporaneamente verifica che nessun altro inizi a sua volta la trasmissione: questo fatto può accadere, in virtù del tempo di propagazione⁷⁸ non nullo, e determina un periodo (detto di *contesa*, e che dipende dalla massima lunghezza del cavo) entro il quale un nodo può erroneamente credere che nessun altro stia trasmettendo.

Qualora sia rilevata una contesa, i due nodi smettono di trasmettere, e riprovano solo dopo una attesa di durata casuale.

Trasmissione Il segnale relativo al pacchetto Ethernet viene trasmesso adottando una codifica di linea di tipo Manchester differenziale. La configurazione con tutti i nodi collegati su di uno stesso cavo è detta *a bus*, e sono state coniate apposite sigle per identificare il tipo di connessione, come ad esempio 10BASE5 e 10BASE2, relative al

⁷⁸ Su di un cavo coassiale *tick* da 50 Ω , la velocità di propagazione risulta di $231 \cdot 10^6$ metri/secondo. Su di una lunghezza di 500 metri, occorrono 2.16 μsec perchè un segnale si propaghi da un estremo all'altro. Dato che è permesso di congiungere fino a 5 segmenti di rete per mezzo di ripetitori, e che anch'essi introducono un ritardo, si è stabilito che la minima lunghezza di un pacchetto Ethernet debba essere di 64 byte, che alla velocità di trasmissione di 10 Mbit/sec corrisponde ad una durata di 54.4 μsec , garantendo così che se si è verificata una collisione, le due parti in causa possano accorgersene.

collegamento di banda base a 10 Mbps, su cavo *tick* e *thin*⁷⁹, con estensione massima 500 e 200 metri⁸⁰.

6.6.1.7 Fast e Gigabit Ethernet

Mentre si proponeva ATM come una soluzione idonea per quasi tutti gli ambiti, la tecnologia Ethernet ha incrementato la velocità trasmissiva di un fattore pari a mille, e si propone sempre più come soluzione generalizzata.

Fast Ethernet Nel 1995 è stato definito lo standard IEEE 802.3u detto *Fast Ethernet*, che eleva la velocità di trasmissione a 100 Mbps ed impiega due diversi cavi UTP⁸¹ per le due direzioni di trasmissione, rendendo la comunicazione *full-duplex*. In quest'ambito son definiti i sistemi 10BASET e 100BASET, relativi all'uso del cavo UTP anzichè di un coassiale, e prevedono una topologia *a stella* per la LAN, realizzata utilizzando una unità centrale (detta HUB=*mozzo di ruota*) da cui si dipartono tanti cavi, ognuno che collega un unico nodo. Nel caso di un HUB economico, questo svolge solo le funzioni di ripetitore (ritrasmette tutto su tutte le sue porte) e dunque le collisioni possono ancora verificarsi.

LAN Switch D'altra parte, i dispositivi detti BRIDGE o LAN SWITCH *apprendono* dai pacchetti in transito gli indirizzi ethernet dei nodi collegati alle porte, ed evitano di ritrasmettere i pacchetti sulle porte dove *non si trova* il destinatario. Dato che gran parte del traffico è inviato verso il *gateway* della LAN, lo SWITCH apprende in fretta su che porta questo si trovi, cosicchè tutti i pacchetti destinati all'esterno non sono ritrasmessi sugli altri rami della LAN, ed il traffico tra i nodi connessi allo SWITCH non si propaga al resto della LAN.

La lunghezza massima dei collegamenti è ora ridotta a 100 metri, per il motivo che un pacchetto di dimensione minima di 64 byte trasmesso a 100 Mbps, impiega un tempo che è $\frac{1}{10}$ di quello relativo alla velocità di 10 Mbps, e quindi per consentire la detezione di collisione, si è dovuta ridurre di pari misura la massima distanza tra nodi trasmettenti.

Gigabit Ethernet Nel giugno 1998 viene standardizzato l'IEEE 802.3z, che porta ad 1 Gbps la velocità di trasmissione delle trame Ethernet, rimpiazzando lo strato di codifica di linea dell'802.3 con i due strati inferiori dell'ANSI X3T11 *Fiber Channel*. In questo modo, si mantiene la compatibilità con gli strati LLC e MAC di Ethernet, mentre la trasmissione avviene su fibra ottica o su cavo in accordo alla tabella seguente.

media	distanza	mezzo	sorgente
1000BASE-SX	300 m	f.o. multimodo ϕ 62.5 μm	laser 850 nm
	550 m	f.o. multimodo ϕ 50 μm	laser 850 nm
1000BASE-LX	550 m	f.o. multimodo ϕ 50 o 62.5 μm	laser 1300 nm
	3000 m	f.o. monomodo ϕ 9 μm	laser 1300 nm
1000BASE-CX	25 m	cavo STP (<i>shielded twisted pair</i>)	
1000BASE-T	25-100 m	4 coppie di cavo UTP	

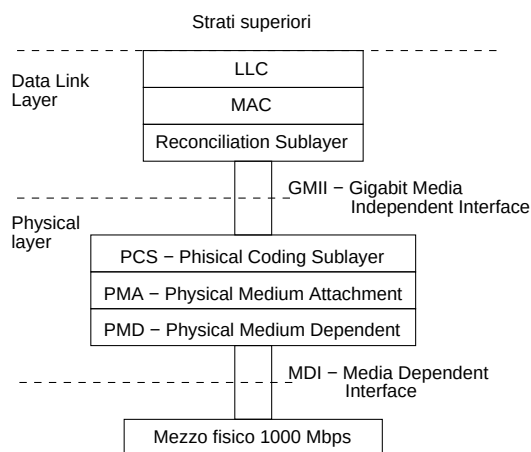
⁷⁹TICK = *duro* (grosso), THIN = *sottile*. Ci si riferisce al diametro del cavo.

⁸⁰Le sigle indicano infatti la velocità, se in banda base o meno, e la lunghezza della tratta.

⁸¹UNSHIELDED TWISTED PAIR (UTP), ossia la coppia ritorta non schermata.

Packet bursting Dato che ora la velocità di trasmissione è 10 volte quella del fast Ethernet, la compatibilità con il MAC CSMA/CD richiederebbe di ridurre la massima lunghezza del collegamento a 10 metri. Al contrario, è stata aumentata la durata minima di una trama portandola a 512 byte, in modo da aumentare la durata della trasmissione e garantire la detezione di collisione. In effetti, il MAC ethernet continua a produrre pacchetti di durata minima 64 byte, e questi sono riempiti (PADDED) fino a 512 byte con una *carrier extension* di simboli speciali. Questa operazione è particolarmente inefficiente se i pacchetti da 64 byte sono frequenti; in tal caso si attua allora il *packet bursting* che, esauriti i 512 byte minimi realizzati come indicato, accoda gli ulteriori pacchetti nello stesso burst trasmissivo, fino ad una lunghezza di 1500 byte.

Architettura La figura seguente mostra la pila protocollare per Gigabit Ethernet.



La GMII permette di usare lo strato MAC con qualunque strato fisico, ed opera sia in full-duplex che in half-duplex, alle velocità di 10, 100 e 1000 Mbps, mediante due percorsi dati (TX e RX) da 8 bit, più due segnali di strato per indicare presenza di portante e detezione di collisione, che sono mappati dal RS nelle primitive riconosciute dallo strato MAC preesistente.

Lo strato fisico è suddiviso in tre sottolivelli. Il PCS fornisce una interfaccia uniforme al RS per tutti i media. Provvede alla conversione 8B/10B tipica del *Fiber Channel*, che rappresenta gruppi di 8 bit mediante *code group*

da 10 bit, alcuni dei quali rappresentano i simboli, ed altri sono codici di controllo, come quelli usati per la *carrier extension*. Il PCS genera inoltre le indicazioni sulla portante e sulla collisione, e gestisce la auto-negoziazione sulla velocità di trasmissione e sulla bidirezionalità del media.

Il PMA provvede alla conversione parallelo-serie e viceversa, mentre il PMD definisce l'MDI, ossia la segnalazione di strato fisico necessaria ai diversi media, così come il tipo di connettore.

Ripetitore full-duplex e controllo di flusso Qualora tutte le porte di un ripetitore siano di tipo full-duplex, non può più verificarsi contesa di accesso al mezzo, e la contesa avviene all'interno del ripetitore, che (non essendo un SWITCH) copia tutte le trame in ingresso (debitamente bufferizzate in apposite code) in tutte le code associate alle porte di uscita. Pertanto, la lunghezza massima dei collegamenti non è più dettata dalla necessità di rilevare collisioni, ma dalle caratteristiche del mezzo trasmissivo. D'altra parte, possono verificarsi situazioni di *flooding* delle code di ingresso; il comitato IEEE 802.3x ha quindi definito un meccanismo di controllo di flusso, che mette in grado i ripetitori (e gli switch) di richiedere ai nodi connessi la sospensione temporanea della trasmissione.

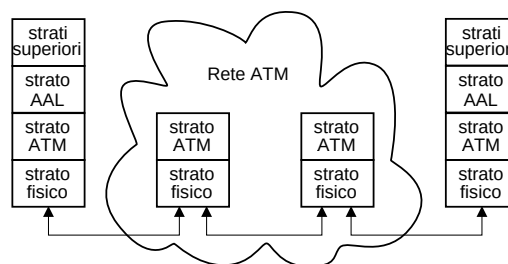
10 Gigabit Ethernet Nel 2002 viene definito lo standard IEEE 802.3ae, che stabilisce le modalità operative di un collegamento Ethernet operante solo in full duplex su fibra ottica. Lo standard prevede di interoperare con la trasmissione SONET/SDH.

6.6.2 Rete ATM

La sigla ATM sta per *Asynchronous Transfer Mode*, ed identifica una particolare rete progettata per trasportare indifferentemente traffico di diversa natura, sia di tipo dati che real-time⁸², che per questo motivo è indicata anche come B-ISDN⁸³. Il suo funzionamento si basa sul principio della *commutazione di cella* (CELL SWITCHING), dove per cella si intende un pacchetto di lunghezza fissa di 53 byte. I primi 5 byte delle celle contengono un identificativo di connessione, ed il loro instradamento avviene mediante dei circuiti virtuali. La commutazione delle celle tra i nodi di rete ha luogo in maniera particolarmente efficiente, e questa è una delle caratteristiche più rilevanti dell'ATM.

Architettura La rete ATM viene indicata anche come una *Overlay Network*, in quanto operativamente si *sovrappone* ai livelli inferiori di una rete esterna.

Dal canto suo, ATM è strutturata sui tre strati funzionali di adattamento (AAL), di commutazione ATM, e fisico. Mentre i nodi ai bordi della rete devono realizzare tutti e tre gli strati, i nodi interni svolgono solo le funzioni attuate da quelli inferiori. La tabella 6.1 riporta le principali funzioni svolte dai tre strati, e pone in evidenza come in uno stesso strato siano identificabili diverse sotto-funzioni.



Strato fisico Il mezzo primario di trasmissione (con cui è in contatto il sotto-strato PM) per ATM è la fibra ottica, in accordo alla struttura di trama dell'SDH/SONET, per la quale sono state standardizzate le velocità di 1.5 e 2 Mbps (DS1/E1), 155 Mbps (OC3) e 622 Mbps (OC12c). La velocità di 155 Mbps è disponibile anche su FIBRE CHANNEL, e su cavo ritorto, mentre la velocità di 100 Mbps è disponibile su FDDI. Infine, sono previste anche velocità di interconnessione di 139, 52, 45, 34 e 25 Mbps.

In funzione del mezzo trasmissivo, può variare la *struttura di trama*⁸⁴ in cui vanno inserite le celle. Il quinto byte della intestazione di cella, contiene l'*Header Error Code* (HEC) calcolato sui 4 byte precedenti, che viene usato in ricezione per rivelare due errori

⁸²Per traffico real-time si intende sia quello telefonico, sia più in generale quello di natura multimediale.

⁸³Siamo alla fine degli anni '80, e la definizione *Integrated Service Data Network* (ISDN) si riferisce ad una rete in grado di permettere, oltre al normale trasporto dei dati, anche servizi di rete. La rete ISDN era però limitata ad una velocità massima (presso l'utente) di 2 Mbps, e per questo venne chiamata *narrow-band ISDN* (N-ISDN). A questa, avrebbe fatto seguito la *broad-band ISDN* (B-ISDN) che ha poi dato luogo alla definizione dell'ATM.

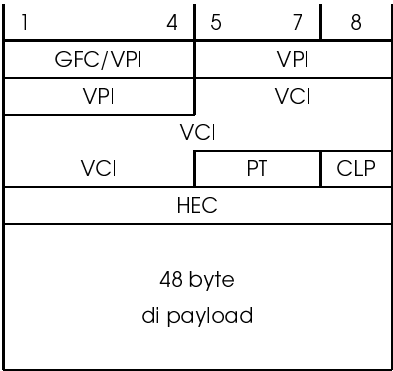
⁸⁴Sono definite due tipi di *interfaccia utente-network* (UNI): quella SDH/SONET, in cui le celle sono inserite nel *payload* della trama SDH, e quella CELL-BASED, che prevede un flusso continuo di celle. Mentre nel primo caso il bit rate lordo comprende l'*overhead* di trama, nel secondo comprende la presenza di celle di tipo *Operation and Maintenance* (OAM).

strato	sotto-strato	funzioni
ATM Adaptation Layer (AAL)	• Convergenza (CS) • Segmentazione e Riasssemblaggio (SAR)	Definisce il servizio offerto agli strati superiori Suddivide i dati in modo compatibile con la dimensione di cella, e li ricostruisce in ricezione
ATM layer		Multiplazione e demultiplazione delle celle Traslazione delle etichette VPI/VCI Generazione/estrazione dell'HEADER della cella Gestione del controllo di flusso GFC
Physical Layer (PL)	• Convergenza di trasmissione (TC) • Mezzo Fisico (PM)	Delimitazione delle celle Inserimento celle IDLE per adattamento velocità Generazione e verifica dell'HEC (controllo di errore) Generazione della trama di trasmissione Temporizzazione e sincronizzazione Gestione del mezzo

Tabella 6.1: Stratificazione delle funzioni in una rete ATM

e correggerne uno⁸⁵. Nel caso in cui la sorgente produca dati a velocità inferiore a quella del collegamento, sono inserite celle aggiuntive di tipo IDLE, rimosse al ricevitore⁸⁶. Infine, la *delimitazione delle celle* è attuata in ricezione in base alla correlazione tra i primi quattro byte dell'header, ed il campo HEC dello stesso.

Strato ATM Mentre lo strato fisico si occupa di trasmettere e ricevere celle, lo strato ATM si occupa di elaborarle. Nei nodi di frontiera, le celle sono multiplate e demultiplate, mentre dentro la rete, sono commutate tra gli ingressi e le uscite.



Formato della cella ATM

Nei primi quattro byte dell'header di cella trova posto *l'etichetta* necessaria a realizzare il trasferimento a circuito virtuale; questa etichetta è suddivisa in due campi, il *Virtual Path Identifier* (VPI) ed il *Virtual Channel Identifier* (VCI)⁸⁷. Il motivo della suddivisione risiede nella possibilità di raggruppare logicamente diversi circuiti virtuali che condividono lo stesso percorso nella rete. Nei collegamenti di cui è composto il percorso comune, viene usato uno stesso vpi per tutte le celle, mentre le diverse connessioni su quel percorso sono

identificate mediante diversi vci. L'instradamento congiunto delle celle con uguale vpi è

⁸⁵Nel primo caso la cella viene scartata, mentre nel secondo inoltrata correttamente. La presenza di più di due errori, provoca un errato inoltro della cella.

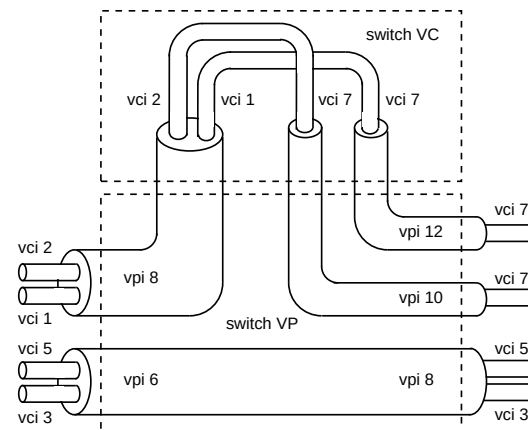
⁸⁶Le celle IDLE sono riconoscibili in base ad una particolare configurazione dei primi 4 byte dell'header, così come avviene per le celle OAM, nonché per altri tipi particolari di cella, che trasportano la segnalazione degli strati superiori.

⁸⁷Mentre per VCI sono riservati 16 bit, per VPI si usano 12 bit all'interno della rete, e 8 bit ai suoi bordi, riservando 4 bit indicati come *Generic Flow Control* (GFC) per regolare il flusso delle sorgenti.

effettuato nei nodi (vp switch), che si occupano solo⁸⁸ di scambiare il VPI delle celle, e di porle sulla porta di uscita corretta, come indicato dalle tabelle di instradamento.

La sequenza dei nodi attraversati dall'instradamento è indicata come *Virtual Path Connection* (VPC), è composta da zero o più VP switch, ed è delimitata tra due nodi (VC o VP/VC switch⁸⁹) che elaborano anche i VCI. La sequenza dei VC switch che elaborano i VCI, e che si estende tra due nodi che terminano lo strato di adattamento, è indicata invece con il termine *Virtual Channel Connection* (VCC)⁹⁰ e comprende uno o più VPC, coincidendo spesso⁹¹ con il percorso tra ingresso ed uscita⁹² della rete ATM.

La creazione delle tabelle di instradamento può essere di tipo manuale, dando luogo ad una *Permanent Virtual Connection* (PVC), oppure può essere il risultato di una richiesta estemporanea, dando luogo ad una *Switched Virtual Connection* (SVC)⁹³; L'OGGETTO DELLA RICHIESTA PUÒ ESSERE UNA VCC od una VPC, ed in questo secondo caso la VPC verrà usata per tutte le VCC future tra i due nodi.



Classi di traffico e Qualità del Servizio (QoS) Nella fase di *setup*, sono attuate delle verifiche dette *Connession Admission Control* (CAC) per assicurarsi che la nuova connessione non degradi le prestazioni di quelle già in corso, nel qual caso la chiamata è rifiutata. La sua accettazione determina invece la stipula tra utente e rete di un *Traffic Contract* a cui la sorgente si deve attenere. Nel corso della trasmissione, i nodi ATM verificano che le caratteristiche del traffico in transito nelle VCC siano conformi al rispettivo contratto, svolgendo un *Usage Parameter Control* (UPC) detto anche *policing*⁹⁴. Prima di proseguire, forniamo però alcune definizioni.

Come anticipato, ATM si è sviluppata per trasportare diversi tipi di traffico, classificabili come segue, nei termini dei parametri indicati di seguito:

⁸⁸Questa semplificazione del lavoro di instradamento, quando confrontata con quello relativo ad una rete IP, è all'origine della vocazione *fast switching* della rete ATM. Per di più, permette la realizzazione *hardware* dei circuiti di commutazione. D'altra parte, mentre per IP l'instradamento avviene al momento della trasmissione, in ATM avviene durante la *set-up* della connessione, quando le tabelle di instradamento sono inizializzate.

⁸⁹Nel caso in cui venga invece scambiato solo il VCI, si ottiene uno switch VC puro.

⁹⁰La rete ATM assicura la consegna delle celle di una stessa VCC nello stesso ordine con cui sono state trasmesse, mentre non assicura l'ordinamento per le celle di una stessa VPC.

⁹¹Può accadere infatti di incontrare uno switch VC puro, in cui è scambiato solo il VCI, ed al quale fanno capo due diverse VCC.

⁹²I nodi di ingresso ed uscita sono indicati come *ingress* ed *egress* nella terminologia ATM.

⁹³Nella richiesta di una SVC, l'utente invia i messaggi di *setup* su di una particolare (*well known*) coppia VPI/VCI=0/5. In generale, le prime 32 VCI di ogni VPI sono riservate per propositi di controllo.

ntenuti dei messaggi di segnalazione aderiscono alle specifiche Q.2931, che fanno parte di *User Network Interface* (UNI) 3.1, e sono un adattamento di Q.931 per N-ISDN. Le specifiche UNI 4.0 prevedono la negoziazione della QoS, e la capacità di richiedere una SVC per una VPC.

⁹⁴Letteralmente: POLIZIOTTAMENTO. Il controllo può anche essere effettuato su di una intera VPC.

- *Constant Bit Rate* (CBR) identifica il traffico real-time come la voce⁹⁵ ed il video non codificato;
- *Variable Bit Rate* (VBR) può essere di tipo real time (es. video MPEG) oppure no, ed allora può tollerare variazioni di ritardo (CDV) ma non l'eccessiva perdita di dati (CLR);
- *Available Bit Rate* (ABR) tenta di sfruttare al meglio la banda disponibile. Il contratto prevede la fornitura di un MCR da parte della rete, e le sorgenti sono in grado di rispondere ad una indicazione di congestione, riducendo di conseguenza l'attività;
- *Unspecified Bit Rate* (UBR) condivide la banda rimanente con ABR, ma non gli è riconosciuto un MBR, né è previsto nessun controllo di congestione. Le celle in eccesso sono scartate. Idonea per trasmissioni insensibili a ritardi elevati, e che dispongono di meccanismi di controllo di flusso indipendenti⁹⁶.

Le classi di traffico sono descrivibili mediante i parametri

- *Peak Cell Rate* (PCR) applicabile a tutte le classi, ma è l'unico parametro per CBR;
- *Sustainable Cell Rate* (SCR) assieme ai tre seguenti, descrive le caratteristiche di VBR: velocità comprese tra SCR e PCR sono non-conformi, se di durata maggiore di MBS;
- *Minimum Cell Rate* (MCR) caratterizza la garanzia di banda offerta alla classe ABR;
- *Maximum Burst Size* (MBS) descrive la durata dei picchi di traffico per sorgenti VBR.

Il contratto di traffico, mentre impegna la sorgente a rispettare i parametri di traffico dichiarati, vincola la rete alla realizzazione di una *Quality of Service* (QoS), rappresentata dalle grandezze (tra le altre)

- *Cell Transfer Delay* (CTD) assieme alla seguente, è molto importante per la classe CBR;
- *Cell Delay Variation* (CDV) rappresenta la variabilità nella consegna delle celle, dannosa per le applicazioni real-time. La presenza di una CDV elevata può inoltre provocare fenomeni di momentanea congestione all'interno della rete, e può essere ridotta adottando degli *shaper*⁹⁷, che riducono la variabilità di ritardo a spese un aumento di CTD;
- *Cell Loss Ratio* (CLR) rappresenta il tasso di scarto di celle del collegamento.

⁹⁵La classe CBR si presta bene a trasportare traffico telefonico PCM. In questo caso, può trasportare solo gli intervalli temporali realmente occupati.

⁹⁶La classe UBR è particolarmente adatta al trasporto di traffico IP, in quanto questo è un protocollo senza connessione, e gli strati superiori (ad es. il TCP) sono in grado di gestire correttamente un servizio di collegamento con perdita di dati.

⁹⁷Un *sagomatore* è composto in prima approssimazione da un buffer di memoria, il cui ritmo di svuotamento *non è mai* superiore ad un valore costante.

Nel caso in cui il policing rilevi che una connessione viola le condizioni contrattuali⁹⁸, può intraprendere svariate azioni, e se può non scartare immediatamente una cella, provvede comunque a segnalarne l'anomalia, ponendo pari ad uno il bit *Cell Loss Priority* (CLP) dell'header. Ciò fa sì che la cella divenga scartabile⁹⁹ in caso di congestione in altri nodi. Un ulteriore campo dell'header, il *Payload Type* (PT), può infine ospitare una *segnalazione in avanti*, che manifesta il fatto che la cella in questione ha subito congestione.

Indirizzamento I nodi di una rete ATM sono identificati da un indirizzo di 20 byte, di diverso significato nei casi di reti pubbliche o private, come indicato dal primo byte (AFI).

Nel primo caso, detto *formato NSAP*¹⁰⁰, il DCC o l'ICD sono assegnati dall'ISO, e l'indirizzo del nodo è disposto nei 10 byte indicati come *High-Order Domain*

Rete Privata				
AFI	ICD/DCC	HO-DSP	ESI	SEL
Rete Pubblica				
AFI	E.164	HO-DSP	ESI	SEL

Specific Part (HO-DSP). I sei byte dell'*End System Identifier* (ESI) sono forniti dal dispositivo connesso ai bordi della rete, e coincidono con il suo indirizzo *Ethernet*: in tal modo la rete comunica un prefisso che identifica il nodo di ingresso, ed il dispositivo lo associa al proprio ESI per forgiare il proprio indirizzo completo. Infine, il byte SEL può essere usato per moltiplicare più entità presso il terminale, ed è ignorato dalla rete.

Nel caso di rete pubblica, il campo HO-DSP è ristretto a 4 byte, e gli 8 byte di E.164 contengono un indirizzo appartenente alla numerazione telefonica mondiale.

Strato di adattamento Come mostrato in tab. 6.1, l'AAL è suddiviso in due componenti, *Segmenting and Reassembly* (SAR) e *Convergence Sublayer* (CS); le funzioni di quest'ultimo sono ulteriormente ripartite tra una *Common Part* (CPCS) ed un *Service Specific cs* (SSCS).

Il compito di AAL è quello di generare i 48 byte del payload per le celle ATM, a partire dalle SDU ricevute, e di ricomporre queste ultime in ricezione, a partire dal risultato della loro demoltiplicazione operata (in base alle etichette VPI/VCI) dallo strato ATM ricevente. Mentre il SAR si interfaccia con lo strato ATM, il CS interagisce con i protocolli superiori, e le esatte operazioni svolte dipendono dalla natura del traffico trasportato: la tabella che segue mostra quattro diverse situazioni.

⁹⁸Ad esempio, una CBR supera il proprio PCR, od una VBR oltrepassa il PCR per più tempo di MBS, oppure il traffico generato da una UBR non può essere instradato per l'esaurimento della banda.

⁹⁹Alcune classi di traffico pongono CLP=1 già in partenza, sia per una capacità indipendente di risolvere situazioni di perdita di dati, sia per la diversa natura dei dati che possono inviare, come ad esempio una codifica di segnale in cui alcuni dati possono essere interpolati, mentre altri no. Al contrario, alcune sorgenti confidano molto nel rispetto del proprio CLP=0, come ad esempio nel caso in cui queste inviino pacchetti di dati ben più grandi delle celle ATM, e che sono di conseguenza frammentati in molte unità, ed in presenza di una sola cella mancante, devono ritrasmettere l'intero pacchetto. In quest'ultimo caso, sono state elaborate strategie di *scarto precoce* (EARLY DISCARD) di tutte le celle di un pacchetto, per il quale si è già verificato lo scarto di una cella componente.

¹⁰⁰Il formato NSAP si ispira al *Network Service Access Point* dell'OSI, e se ne differenzia per aver fuso i campi *Routing Domain* e *Area* in un solo campo HO-DSP, per il quale si è adottata una gerarchia di instradamento basata su di un prefisso mobile, in modo simile al CDIR dell'IP.

A	B	C	D
servizio isocrono		ritardo variabile consentito	
bit rate costante	bit rate variabile		
con connessione			senza connessione
AAL 1	AAL 2	AAL 3/4 o 5	AAL 3/4 o 5

La classe A è un classico caso CBR, ed in tal caso si adotta un AAL di tipo 1, in cui lo strato CS è assente, ed il SAR utilizza il primo dei 48 byte di cella per inserire informazioni di controllo sull'ordine di consegna, e di ausilio al recupero della temporizzazione di sorgente presso la destinazione.

La classe B (AAL 2) individua sorgenti multimediali a pacchetto, mentre per la C (AAL 3/4 o 5) siamo più tipicamente in presenza di una connessione dati a circuito virtuale. In questa categoria rientra il trasporto di collegamenti X.25 e *frame relay*, sia di tipo ABR che UBR. Lo stesso tipo di AAL (3/4 o 5) è infine usato anche per la classe D, in cui rientra pienamente il trasporto di traffico IP su ATM.

Quando il CS di AAL 3/4 riceve una SDU (di dimensione massima $2^{16} - 1$) dagli strati superiori, la allinea ad un multiplo di 32 byte, e vi aggiunge 32 byte in testa ed in coda con informazioni di lunghezza e di controllo di integrità. La CS-PDU risultante è passata al SAR, che la suddivide in blocchi di 44 byte, a cui ne aggiunge 2 in testa e due in coda¹⁰¹, e completa così la serie di 48 byte da passare allo strato ATM. Al contrario, il SAR dell'AAL 5 suddivide la CS-PDU in blocchi da 48 byte e non aggiunge informazioni¹⁰², demandando il riconoscimento dell'ultima cella di una stessa CS-PDU ad un bit del campo PT presente nell'header di cella ATM. D'altra parte, la lunghezza della CS-PDU dell'AAL 5 è multipla di 48 byte, aggiungendone un numero appropriato, oltre ai 64 byte di intestazione (ora posta in coda), in cui ora sono presenti anche 8 bit di informazione da utente ad utente.

Ip su ATM classico Allo stesso tempo in cui si diffonde l'uso di ATM tra gli operatori di TLC, il TCP/IP emerge come lo standard comune per l'interconnessione tra elaboratori. Sebbene il TCP/IP si appoggi ad *Ethernet* in area locale, per i collegamenti a lunga distanza¹⁰³ l'ATM presenta indubbi vantaggi come la disponibilità di banda su richiesta, la coesistenza con il traffico di tipo diverso, l'elevata efficienza della commutazione, e la possibilità di raggiungere diverse destinazioni. Una prima soluzione, subito scartata, fu quella nota come *peer model*, in cui i nodi ATM possiedono un indirizzo IP, ed usano i protocolli di routing IP. ATM risulta così *appaiata* alla rete IP, ma ciò complica la realizzazione dei nodi ATM, ed il metodo non si generalizza per protocolli diversi da IP.

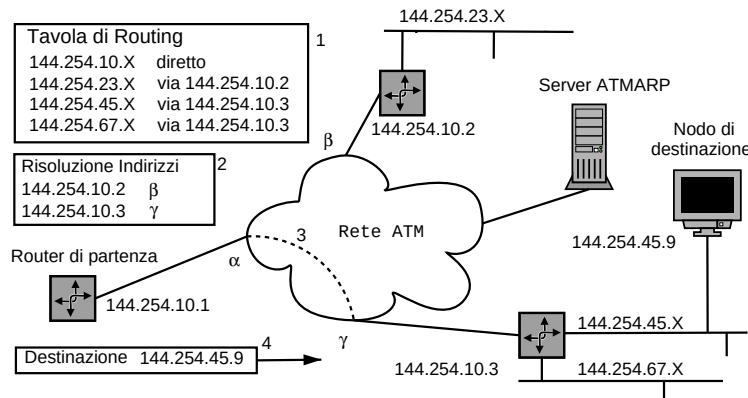
L'alternativa seguita, detta *overlay model*, vede ATM come uno stato di collegamento su cui opera l'IP, che si comporta come se si trovasse su di una LAN. In particolare, solo i nodi di frontiera tra IP ed ATM prendono un doppio indirizzo, ed individuano una *Logical Subnet* (LIS) definita da uno stesso prefisso IP ed una stessa maschera di sottorete. Con riferimento alla figura, quando il router di partenza vuole contattare il nodo di destinazione, trova (1) prima l'IP del router di destinazione, e quindi invia una richiesta

¹⁰¹Questi ultimi 4 byte contengono l'indicazione (2 bit) se si tratti della prima, ultima od intermedia cella di una stessa CS-PDU, la lunghezza dei dati validi se è l'ultima (6 bit), un numero di sequenza (4 bit), un controllo di errore (10 bit), ed una etichetta (10 bit) che rende possibile interallacciare temporalmente le celle di diverse CS-PDU.

¹⁰²In questo modo si risparmiano 4 byte ogni 48. Ora però è indispensabile che le celle arrivino in sequenza, e non è più possibile alternare diverse CS-PDU.

¹⁰³Quando la distanza tra i nodi oltrepassa dimensioni di un edificio, si parla di *Campus Network* o di *Wide Area Network* (WAN), ed a volte è usato il termine *Metropolitan Area Network* (MAN) per estensione cittadine. Per estensioni ancora maggiori si parla di *reti in area geografica*.

ARP al server ATMARP presente nella LIS¹⁰⁴, che risponde comunicando l'indirizzo γ , il quale è così risolto (2). A questo punto si può instaurare una VCC con B mediante la segnalazione ATM (3), ed effettuare la comunicazione (4). Questa soluzione è nota come *VC multiplexing*, ed i dati sono incapsulati direttamente nella CPCS-PDU di AAL5. In ricezione, l'etichetta VPI/VCI è usata per consegnare il pacchetto al protocollo di strato superiore che ha realizzato la connessione ATM. D'altra parte, questa elaborazione deve avvenire a *diretto contatto* con AAL5, e ciò preclude la possibilità di interlavoro con nodi esterni alla rete ATM.



Nel caso in cui sia antieconomico creare un gran numero di VC, o se si dispone unicamente di un PVC¹⁰⁵, il pacchetto IP viene incapsulato in un header LLC IEEE 802.2 prima di essere consegnato all'AAL5. In tal modo, il router ricevente esamina l'header LLC del pacchetto ricevuto dal nodo ATM di egress, per consegnare il pacchetto al protocollo appropriato, realizzando così un *trasporto multiprotocollo* su ATM.

LANE, NHRP e MPOA Discutiamo qui brevemente ulteriori possibilità di utilizzo di ATM come trasporto IP, ma a cui verosimilmente sarà preferito l'MPLS.

Mentre l'approccio classico aggiunge un substrato tra IP ed AAL5, per così dire *esterno* alla rete ATM, l'approccio LANE (LAN *Emulation*) ne aggiunge uno *esterno* alla rete IP, che *crede* di avere a che fare con una LAN ethernet. In questo caso anziché una LIS, si definisce una *Emulated LAN* (ELAN), il cui esatto funzionamento prevede diversi passaggi¹⁰⁶.

Sia nel caso classico che in quello LANE, se due router IP sono su due LAN (LIS o ELAN) differenti (con prefissi differenti) la comunicazione tra i due deve necessariamente

¹⁰⁴ Tutti i nodi della LIS hanno configurato manualmente l'indirizzo ATM del server ATMARP.

¹⁰⁵ Un VC permanente collega solamente una coppia di nodi, ed in tal caso è possibile anche fare a meno del server ATMARP, in quanto un PVC è configurato manualmente. Nei fatti, questo è l'uso più diffuso del trasporto IP over ATM, ed è tipicamente utilizzato per collegare sedi distanti di una stesso sistema autonomo, eliminando la necessità di sviluppare in proprio un impianto di TLC tra le sedi.

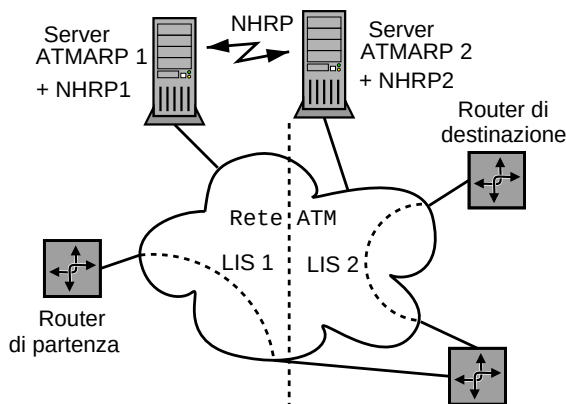
¹⁰⁶ La emulazione di una LAN da parte della rete ATM è possibile dopo aver definito per ogni ELAN un LAN *Emulation Server* (LES) a cui ogni LAN *Emulation Client* (LEC) si rivolge per conoscere l'indirizzo ATM di un altro LEC, a partire da suo indirizzo MAC (la traduzione da IP a MAC è già avvenuta tramite ARP a livello IP). In una ELAN deve inoltre essere presente un dispositivo *Broadcast and Unknown Server* (BUS) che diffonde a tutti i LEC i pacchetti broadcast Ethernet (come ad es. le richieste ARP), e che viene usato dai LEC che devono inviare un broadcast. Infine, occorre un LAN *Emulation Configuration Server* (LECS) che conosce, per ogni ELAN della rete ATM, l'elenco dei LEC, del LES e del BUS.

All'accensione di un LEC, questo contatta il LECS (conoscendone l'indirizzo ATM, oppure su di una VCC ben nota, o tramite segnalazione ATM) per apprendere gli indirizzi ATM del proprio LES e del BUS. Quindi, registra presso il LES la corrispondenza tra i propri indirizzi MAC ed ATM. Quando un LEC desidera inviare dati ad un altro LEC, dopo averne risolto l'indirizzo ATM interrogando il LES, incapsula le trame IP con un header LLC IEEE 802.2 proprio come nel caso classico.

attraversare un terzo router IP, anche se esiste un collegamento diretto tra i primi due, tutto interno alla rete ATM. La situazione è illustrata in figura, per il caso classico.

Come possiamo notare, i router di partenza e di destinazione potrebbero dialogare direttamente tramite la rete ATM, diminuendo il carico di traffico della stessa, e risparmiando al router intermedio il compito di riassemblare e disassemblare i pacchetti IP in transito, oltre a riclassificarli ai fini del routing. Se i server ATMARP delle due LIS possono scambiarsi le proprie informazioni, il router di partenza può arrivare a conoscere l'indirizzo ATM di quello di destinazione, e creare un collegamento diretto.

Lo scambio delle corrispondenze $\langle ind. IP; ind. ATM \rangle$ avviene per mezzo del *Next Hop Resolution Protocol* (NHRP) tra entità indicate come *NHRP Server* (NHS), che possono appartenere ognuno a più LIS, e che instaurano tra di loro un meccanismo di *passa-parola*¹⁰⁷, per rispondere alle interrogazioni che ricevono. L'applicazione di un meccanismo in parte simile, porta nel caso delle ELAN alla definizione del *Multi Protocol over ATM* (MPOA)¹⁰⁸.



MPLS Il *Multi Protocol Label Switching* (MPLS) è un metodo di realizzare una trasmissione a circuito virtuale su reti IP, la cui architettura è descritta nella RFC 3031 dell'IETF, e che verrà esposto meglio in una prossima edizione. Qui illustriamo i legami che MPLS presenta con ATM.

Lo sviluppo di MPLS ha origine dalle iniziative industriali tese a realizzare router internet economici di prestazioni elevate, e capaci di gestire la banda in modo appropriato. Lo IETF ha ricevuto il compito di armonizzare in una architettura standardizzata i diversi approcci, basati sul principio di inoltrare i pacchetti in base ad una etichetta (LABEL) impostata dal primo router della rete, proprio come avviene in ATM. Dato che erano già disponibili i dispositivi hardware per realizzare i nodi di switching ATM, i primi prototipi hanno semplicemente utilizzato tali switch sotto il diretto controllo di un router IP, collegato ad altri simili tramite la rete ATM. L'MPLS è tuttavia più generale, sia verso l'alto (è *multiprotocollo* in quanto si applica oltre che ad IP, a qualunque altro strato di

¹⁰⁷I NHS risiedono su dispositivi che sono anche router IP, e che quindi mantengono aggiornate le tabelle di instradamento che indicano il prossimo salto (*next hop*) verso destinazioni IP. Le richieste di risoluzione ATMARP per un certo indirizzo IP sono instradate mediante queste stesse tabelle, giungendo di salto in salto fino al router-NHS appartenente alla stessa LIS dell'IP di destinazione, che conosce la risposta. Quest'ultima ripercorre all'indietro il percorso fatto dalla richiesta, fino alla sorgente. I router attraversati dal *passa parola, ricordano* (per un pò) le risposte trasportate, riducendo il traffico NHRP.

¹⁰⁸Il metodo si basa su di un meccanismo indicato come *flow detection*, attuato dal router IP-ATM prossimo alla sorgente, che è in grado di accorgersi di traffico non sporadico diretto verso una medesima destinazione. Questo router impersona allora un *MPOA Client* (MPC), ed interroga un *MPOA server* (MPS) per conoscere l'indirizzo ATM della destinazione, in modo da creare un collegamento diretto. Ogni MPS serve una o più ELAN, e gli MPS comunicano tra loro mediante il NHRP.

L'MPOA realizza la separazione tra il calcolo dell'instradamento e l'inoltro dei dati. A differenza di un router tradizionale, che svolge entrambi i compiti, l'MPC svolge solo l'inoltro verso l'indirizzo ATM di destinazione, mentre quest'ultimo è fornito dall'MPS, che si comporta quindi come un *route server*.

rete) che verso il basso (funziona indifferentemente dall'implementazione dello strato di collegamento, sia ATM, *ethernet* od altro).

La *label* apposta dal primo MPLS Router (LSR) dipende dalla destinazione IP del pacchetto; diverse destinazioni possono coincidere con una sola *Forwarding Equivalence Class* (FEC)¹⁰⁹, identificata da una singola *label*. Tutti i pacchetti di una stessa FEC sono inoltrati verso il medesimo *next hop*, indicato dalla tabella di routing, indicizzata dalla *label*¹¹⁰. Nella stessa tabella, si trova anche la nuova *label* da assegnare al pacchetto, prima di consegnarlo all'LSR seguente. In tutti i LSR successivi, il pacchetto non è ri-classificato, ma solo inoltrato verso il *next hop* con una nuova *label* come ordinato dalla tabella di routing. Pertanto, è il primo LSR a decidere tutto il tragitto, ed i pacchetti di una stessa FEC seguono tutti lo stesso *Label Switched Path* (LSP). In tal modo, gli switch possono essere più semplici, si possono stabilire instradamenti diversi per una stessa destinazione¹¹¹ in base al punto di ingresso, così come le FEC posso essere rese dipendenti non solo dalla destinazione, ma anche da altri parametri, come la classe di servizio richiesta.

L'associazione tra *label* e FEC (ossia il *next hop* per i pacchetti con quella *label*) è stabilita dal LSR di *destinazione*¹¹², e cioè un LSR indica agli LSR dai quali *si aspetta di ricevere* traffico, quale *label* usare in corrispondenza delle FEC per le quali conosce l'instradamento. Dato che la conoscenza di un istradamento è anche il prerequisito sulla cui base sono annunciate le informazioni di routing *hop-by-hop* in internet, il *Label Distribution Protocol* (LDP) può essere vantaggiosamente associato ai protocolli di distribuzione delle informazioni di routing già esistenti (es. BGP). Le associazioni tra FEC e *label* si propagano dunque fino ai nodi di ingresso, realizzando un reticolo di "alberi" di LSP, costituiti dagli LSP definiti da una stessa FEC, e che convergono verso uno stesso *egress* a partire da diversi *ingress*. Nel nodo in cui più LSP si riuniscono, è possibile effettuare il *label merging* assegnando la stessa *label* ai pacchetti uscenti, riducendo così la dimensione delle tabelle di routing.

L'etichetta *label* su cui si basa l'MPLS può genericamente consistere in un incapsulamento della PDU dello strato di rete, prima che questa sia passata allo strato di collegamento. Quando i LSR sono realizzati mediante switch ATM, la *label* è efficacemente realizzata usando la coppia VPI/VCI, realizzando i LSP come delle VCC. In questo caso però, sorgono problemi nel caso in cui si debba effettuare il *merge* di più LSP relative ad una stessa FEC, che passano da uno stesso ATM-LSR. Infatti, se un nodo adottasse in uscita una stessa *label-VCC* per differenti VCC entranti, le celle in cui sono segmentati i pacchetti IP, ed ora con uguale *label-VCC*, si alternerebbero, rendendo impossibile il riasssemblaggio dei pacchetti. Per questo motivo, MPLS può operare anche con LSR che non permettono il *merging*, e che possono quindi essere utilizzati assieme ad altri che ne sono capaci; in tal caso, L'LSR non-merging *non* è notificato automaticamente delle associazioni FEC-*label*, ma gli viene comunicata una (diversa) *label* ogni volta che ne chiede

¹⁰⁹Nel routing ip tradizionale, una FEC coincide con l'instradamento individuato dal *longest match*.

¹¹⁰Nel routing IP convenzionale, per ogni router, la tabella di routing deve essere esaminata per intero per ogni pacchetto, alla ricerca del *longest match* tra le regole presenti.

¹¹¹Il routing IP tradizionale opera su di una base *hop-by-hop*, e per questo non può tenere conto della provenienza. Quando due pacchetti per una medesima destinazione passano da uno stesso router, proseguono per lo stesso percorso.

¹¹²Infatti, è la *label* del pacchetto *ricevuto* che determina il *next hop*, e quindi è quest'ultimo a definire la semantica della *label* presso i propri vicini.

una (da associare ad una FEC), usando così più *label* del necessario. Una alternativa, è quella di codificare la FEC mediante il solo VPI, ed usare il VCI per indicare il nodo di partenza. In questo modo, il *merging* è per così dire *automatico*, senza problemi di alternanza temporale delle celle di diversi pacchetti IP, ed il metodo può essere applicato se è possibile coordinare l'assegnazione dei VCI tra sorgenti diverse, e se il numero delle *label* non oltrepassa la capacità di indirizzamento.

L'esposizione svolta è volutamente semplificata, e trascura per comodità alcune importanti caratteristiche di MPLS.

Capitolo 7

Densità Spettrale e Filtraggio

Descriviamo qui l'effetto del passaggio di un segnale, o di un processo, attraverso un sistema fisico, concentrando l'attenzione sulle modifiche subite dallo spettro di densità di potenza, o di energia. Mentre per i segnali periodici e di energia siamo già in grado di determinare lo spettro di ampiezza dell'uscita come $Y(f) = H(f) X(f)$, e da questo ottenere (ad es.) lo spettro di densità di potenza come $\mathcal{P}_y(f) = |Y(f)|^2$, nel caso in cui $x(t)$ rappresenti una generica realizzazione di un processo, prima di giungere ad un risultato analogo occorre quantomeno definire lo spettro di potenza del generico ingresso $\mathcal{P}_x(f)$; è esattamente questo lo scopo della prima parte del capitolo, in cui i concetti statistici già definiti vengono estesi per giungere alla definizione della *funzione di autocorrelazione*, la cui trasformata di Fourier è pari appunto alla densità cercata (teorema di WIENER).

Il resto del capitolo procede applicando ad esempi pratici la teoria fin qui sviluppata, presentando alcuni casi tipici di determinazione di uno spettro di densità di potenza, descrivendo possibili architetture dei filtri, fino ad elencare in modo sistematico le alterazioni delle diverse grandezze descrittive dei segnali, in conseguenza del loro passaggio attraverso unità elementari di elaborazione.

Il capitolo termina con una appendice in cui sono riferiti ulteriori risultati, e dimostrate alcune relazioni usate fin qui nel testo.

7.1 Correlazione e Covarianza

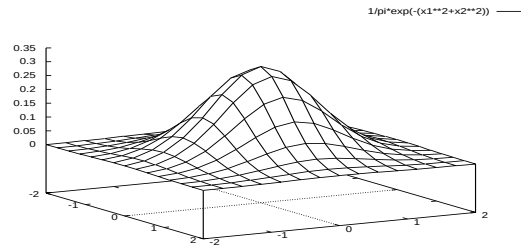
Abbiamo già osservato come la caratterizzazione statistica del primo ordine $p_X(x)$ di un processo $\{x(t, \theta)\}$ stazionario ergodico, consenta il calcolo di valor medio m_x e varianza σ_X^2 , nonché della potenza $\mathcal{P}_X = E_X\{x^2\} = \sigma_X^2 - (m_x)^2$, valida per una qualunque realizzazione θ del processo. Definiamo ora una statistica *del secondo ordine* che permetterà di determinare anche lo spettro di densità di potenza delle realizzazioni del processo.

La statistica di secondo ordine si basa sulla considerazione di 2 istanti t_1 e t_2 , in corrispondenza dei quali estraiamo, da una realizzazione $\bar{\theta}$ di un processo $x(t, \theta)$, due variabili aleatorie $x_1 = x(t_1)$, $x_2 = x(t_2)$. Al variare della realizzazione campionata, tutte le coppie di valori estratti sono altrettante determinazioni di una variabile aleatoria bidimensionale, descritta da una densità



di probabilità congiunta $p_{X_1 X_2}(x_1 x_2; t_1 t_2)$, che in linea di principio dipende anche dagli istanti t_1 e t_2 .

La figura alla pagina precedente mostra l'operazione di campionamento di una realizzazione, mentre quella a fianco esemplifica come la densità congiunta sottenda un volume unitario, e descriva con il suo andamento le regioni del piano $x_1 x_2$ in cui cadono un maggior numero di coppie (ovvero dove la probabilità è più densa).



7.1.1 Correlazione

Questa grandezza dipende dalla definizione di un valore atteso, formalmente analogo a quanto già visto per il caso unidimensionale, che prende il nome di *momento misto di ordine* (i, j) e che risulta:

$$m_{X_1 X_2}^{(i,j)} = E_{X_1 X_2} \{x_1^i x_2^j\} = \int \int x_1^i x_2^j p_{X_1 X_2}(x_1 x_2; t_1 t_2) dx_1 dx_2$$

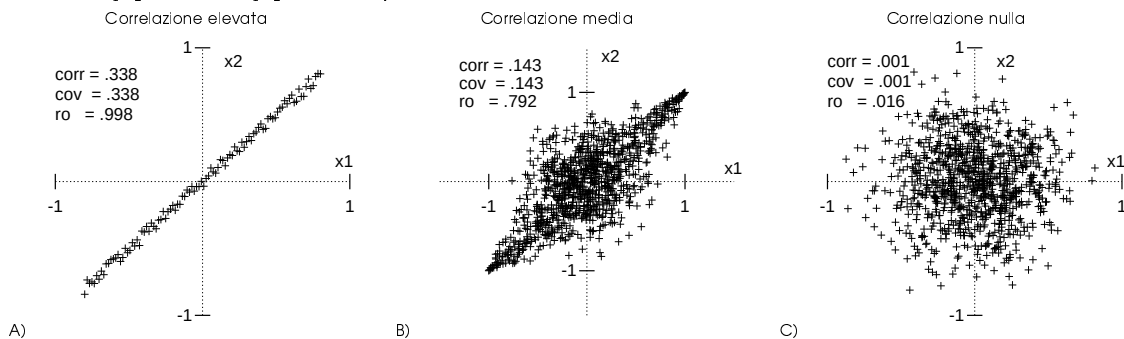
e che nel caso in cui $i = j = 1$ prende il nome di *correlazione* e si indica come

$$\mathcal{R}_X(x(t_1), x(t_2)) = m_{X_1 X_2}^{(1,1)}$$

Prima di proseguire, soffermiamoci un istante per meglio comprendere il significato di $m_{X_1 X_2}^{(1,1)}$. La correlazione fra due variabili aleatorie indica il legame che esiste tra le due¹, nel senso di quanto l'una ha un valore simile all'altra, ed ha un valore assoluto tanto più elevato quanto più i valori di x_1 e x_2 sono legati in modo deterministico²: in effetti però, seguendo gli esempi riportati nella nota, non sempre questo accade, e pertanto è opportuno basarsi sull'uso di momenti centrati come descritto al punto successivo.

¹Intese qui nel senso più generale, per esempio come temperatura e pressione in un punto ben preciso di un cilindro di un motore a scoppio, oppure come pressione e velocità in un circuito idraulico, o pneumatico... astruendo cioè dal caso particolare di due v.a. estratte da una stessa forma d'onda.

²Può tornare utile pensare $m_{X_1 X_2}^{(1,1)}$ come una media pesata dei possibili valori del prodotto $x_1 x_2$; i termini di eguale ampiezza e segno opposto possono elidersi se equiprobabili. Negli esempi che seguono, riportiamo dei *diagrammi di scattering* per sei diversi casi di distribuzione delle coppie di valori x_1 e x_2 , assieme ai valori di $\mathcal{R}_{x_1 x_2}$ (Corr), $\sigma_{x_1 x_2}$ (COV), e ρ (ro) (vedi § (7.9.1) per quest'ultimo parametro)



7.1.2 Covarianza e Indipendenza Statistica

Nel caso in cui le due v.a. siano statisticamente indipendenti, e cioè si possa scrivere $p_{X_1 X_2}(x_1, x_2; t_1, t_2) = p(x_1)p(x_2)$ ⁽³⁾, l'integrale che definisce la correlazione si fattorizza e pertanto

$$\begin{aligned}\mathcal{R}_X(x_1, x_2) &= \int \int x_1 x_2 p(x_1) p(x_2) dx_1 dx_2 = \int x_1 p(x_1) dx_1 \cdot \int x_2 p(x_2) dx_2 = \\ &= E\{x_1\} E\{x_2\} = m_{X_1} m_{X_2}\end{aligned}$$

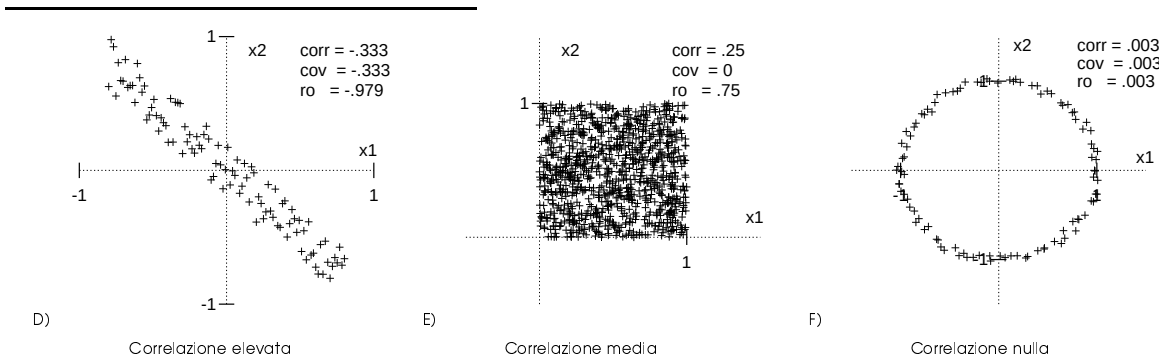
Se a questo punto definiamo *covarianza* il momento misto centrato

$$\sigma(x_1, x_2) = E\{(x_1 - m_{X_1})(x_2 - m_{X_2})\} = E\{x_1 x_2\} - m_{X_1} m_{X_2} = \mathcal{R}_X(x_1, x_2) - m_{X_1} m_{X_2}$$

otteniamo⁴ il risultato che

Se due variabili aleatorie x_1 ed x_2 sono statisticamente indipendenti, queste si dicono INCORRELATE, in quanto la covarianza $\sigma(x_1, x_2)$ è nulla⁵.

La proprietà esposta ha valore in una sola direzione, in quanto se due v.a. esibiscono $\sigma(x_1, x_2) = 0$ non è detto che siano statisticamente indipendenti⁶. L'unico caso in cui ciò si verifica, è quello relativo alle variabili aleatorie estratte da un processo gaussiano.



In A) e D) le coppie di valori sono praticamente legate da una legge deterministica di tipo lineare; in B) ed E) c'è più variabilità ma si nota ancora una certa dipendenza tra le due. Nei casi C) ed F), osserviamo due v.a. per le quali risulta $\mathcal{R}_X(x_1, x_2) = 0$, in quanto le coppie di valori si distribuiscono in egual misura in tutte le direzioni. Osserviamo infine come C) ed E) siano casi in cui x_1 e x_2 risultano *statisticamente indipendenti*, dato che $p_{X_1 X_2}(x_1, x_2)$ è fattorizzabile come $p_{X_1}(x_1)p_{X_2}(x_2)$.

³Omettiamo per brevità di indicare la variabile aleatoria a pedice della densità di probabilità.

⁴L'uguaglianza appena mostrata si ottiene ricordando che un valore atteso è in realtà un integrale, e che la proprietà distributiva di quest'ultimo permette di scrivere

$$E\{(x_1 - m_{X_1})(x_2 - m_{X_2})\} = E\{x_1 x_2\} - E\{x_1 m_{X_2}\} - E\{m_{X_1} x_2\} + E\{m_{X_1} m_{X_2}\}$$

che fornisce il risultato esposto.

⁵Notiamo immediatamente che il termine più corretto sarebbe "incovarianzate"; l'uso (ormai storico e consolidato) dell'espressione *incorrelate* deriva probabilmente dal considerare usualmente grandezze a media nulla, per le quali le due espressioni coincidono.

⁶Vedi ad esempio il caso F) della nota (2), in cui le variabili aleatorie risultano incorrelate ma non sono per nulla indipendenti, in quanto l'una dipende strettissimamente dall'altra, dato che le coppie di valori si dispongono su di un cerchio.

7.1.3 Statistiche dei Processi

Nel caso in cui il processo da cui si estraggono x_1 ed x_2 sia *stazionario*, si ottiene che

$$\mathcal{R}_X(x(t_1), x(t_2)) = \mathcal{R}_X(x(t_1), x(t_2 = t_1 + \tau)) = \mathcal{R}_X(\tau)$$

e cioè la correlazione *dipende solo dall'intervallo* $\tau = t_2 - t_1$. Infatti se un processo è stazionario, le proprietà statistiche non dipendono da traslazioni temporali.

Nel caso in cui il processo sia anche *ergodico*, allora le medie di insieme hanno lo stesso valore delle corrispondenti medie temporali, e dunque la correlazione (media di insieme) può essere calcolata in base alla sua media temporale equivalente, a partire da una qualunque realizzazione del processo

$$\mathcal{R}_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) x(t + \tau) dt$$

Ovviamente è vero anche l'inverso, e cioè: il valor medio del prodotto tra due campioni, estratti (a caso) a distanza τ a partire da una specifica realizzazione⁷, ha un valore che è calcolabile come media di insieme a partire dalla conoscenza della densità di probabilità congiunta $p_{X_1 X_2}(x_1, x_2; t_1, t_1 + \tau)$.

7.1.4 Autocorrelazione

La media temporale $\mathcal{R}_x(\tau)$ appena introdotta prende il nome di *integrale di autocorrelazione*, ed è definito anche per segnali di energia, come

$$\mathcal{R}_x(\tau) = \int_{-\infty}^{\infty} x^*(t) x(t + \tau) dt$$

in cui l'operatore di coniugato generalizza l'operazione anche al caso di segnali complessi. Simile, ma diversa, è la definizione di integrale di *intercorrelazione*, che esprime lo stesso calcolo, ma relativo a due segnali $x(t)$ ed $y(t)$ processi diversi:

$$\mathcal{R}_{xy}(\tau) = \int_{-\infty}^{\infty} x^*(t) y(t + \tau) dt \quad (7.1)$$

L'*integrale* di autocorrelazione è anche detto *funzione* di autocorrelazione, in quanto il suo argomento è un tempo (l'intervallo tra due campioni) e dunque $\mathcal{R}_x(\tau)$ può essere visto come un segnale (funzione di τ anziché di t). Nello studio abbiamo già incontrato un integrale (di convoluzione) il cui risultato è una funzione del tempo; la somiglianza tra i due è più profonda di una semplice analogia, in quanto si può scrivere

$$\mathcal{R}_x(\tau) = \int_{-\infty}^{\infty} x^*(t) x(t + \tau) dt = x^*(-t) * x(t)$$

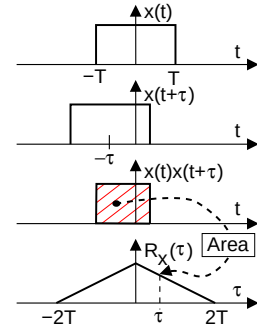
in cui $*$ è il consueto simbolo di convoluzione⁸.

⁷Si rifletta sulla descrizione ora data a parole dell'operazione di media temporale scritta sopra.

⁸Il risultato ottenuto si basa sul cambio di variabile $\theta = t + \tau$ che permette di scrivere

$$\mathcal{R}_x(\tau) = \int_{-\infty}^{\infty} x^*(\theta - \tau) x(\theta) d\theta = \int_{-\infty}^{\infty} x^*(-(\tau - \theta)) x(\theta) d\theta = x^*(-t) * x(t)$$

In base a quest'ultima osservazione otteniamo che la costruzione grafica, che fornisce il risultato dell'integrale di autocorrelazione, è del tutto simile a quella già illustrata per la convoluzione, con la differenza che ora non si effettuano ribaltamenti di asse. La figura a lato ne illustra l'applicazione ad un caso noto, per il quale $x(t) = x^*(-t)$, e che fornisce quindi lo stesso risultato di $x(t) * x(t)$.



7.1.4.1 Proprietà dell'autocorrelazione

Elenchiamo ora alcune *caratteristiche* della funzione di autocorrelazione:

Traslazioni temporali. Se consideriamo i segnali $x(t)$ e $y(t) = x(t + \theta)$, le rispettive autocorrelazioni $\mathcal{R}_x(\tau)$ ed $\mathcal{R}_y(\tau)$ sono identiche⁹. Questo risultato mostra come l'autocorrelazione non tenga conto dell'informazione legata alla fase dei segnali: infatti $x(t)$ e $y(t)$ hanno la stessa densità spettrale, a meno di un contributo di fase lineare, ed hanno uguale autocorrelazione.

Durata Limitata. La funzione di autocorrelazione di un segnale di durata limitata è anch'essa a durata limitata, e di estensione doppia rispetto alla durata del segnale originario.

Segnali Periodici. L'autocorrelazione di un segnale periodico di periodo T è anch'essa periodica, con lo stesso periodo. Infatti per $\tau = nT$ il secondo fattore integrando è traslato di un numero intero di periodi.

Illustriamo ora invece due proprietà *fondamentali*:

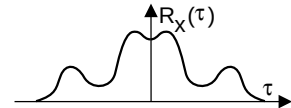
Massimo nell'origine. La $\mathcal{R}_x(\tau)$ calcolata in $\tau = 0$ fornisce il valore massimo di $\mathcal{R}_x(\tau)$ per qualunque altro valore di τ . In particolare, $\mathcal{R}_x(\tau = 0)$ è uguale alla potenza del segnale $x(t)$, od all'energia se $x(t)$ è di energia.

$$0 \leq \mathcal{R}_x(0) = \begin{cases} \int |x(t)|^2 dt = \mathcal{E}_x > |\mathcal{R}_x(\tau \neq 0)| & \text{se } x(t) \text{ di energia} \\ \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt = \mathcal{P}_x \geq |\mathcal{R}_x(\tau \neq 0)| & \text{se } x(t) \text{ di potenza} \end{cases}$$

in cui l'ultimo segno $\geq$ risulta = per $\tau = nT$ se $x(t)$ è periodico.

Simmetria coniugata:

è possibile verificare che risulta $\mathcal{R}_x(-\tau) = \mathcal{R}_x^*(\tau)$, da cui osserviamo subito che $\mathcal{F}\{\mathcal{R}_x(\tau)\}$ è reale. Nel caso in cui $x(t)$ è reale, si ottiene $\mathcal{R}_x(-\tau) = \mathcal{R}_x(\tau)$, ovvero l'autocorrelazione di un segnale reale è *reale pari*.



⁹Infatti otteniamo:

$$\mathcal{R}_y(\tau) = \int_{-\infty}^{\infty} y(t)y(t+\tau)dt = \int_{-\infty}^{\infty} x(t+\theta)x(t+\theta+\tau)dt = \int_{-\infty}^{\infty} x(\alpha)x(\alpha+\tau)d\alpha = \mathcal{R}_x(\tau)$$

7.2 Densità Spettrale

Come anticipato ad inizio capitolo, mostriamo il metodo con cui determinare lo spettro di densità di potenza nel caso di processi. La cosa decisamente gradevole è che questo stesso strumento è valido anche per gli altri tipi di segnale.

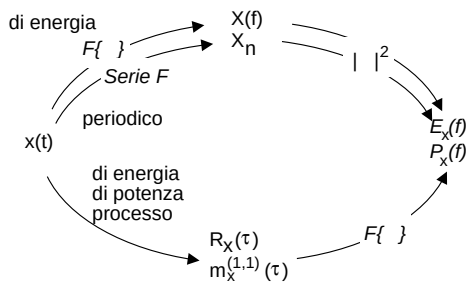
7.2.1 Teorema di Wiener

Lo spettro di densità di potenza $\mathcal{P}_x(f)$ (o di energia $\mathcal{E}_x(f)$) di $x(t)$ è uguale alla trasformata di Fourier della sua funzione di autocorrelazione $\mathcal{F}\{\mathcal{R}_x(\tau)\}$.

La dimostrazione del teorema è straordinariamente semplice, e per segnali di energia si scrive

$$\begin{aligned}\mathcal{R}_x(\tau) &= \int_{-\infty}^{\infty} x^*(t) x(t+\tau) dt = \int_{-\infty}^{\infty} X^*(f) X(f) e^{j2\pi f\tau} df = \\ &= \mathcal{F}^{-1}\{X^*(f) X(f)\} = \mathcal{F}^{-1}\{\mathcal{E}_x(f)\}\end{aligned}$$

in cui abbiamo prima applicato il teorema di Parseval, poi la proprietà di traslazione nel tempo, e quindi espresso $X^*(f) X(f)$ come $\mathcal{E}_x(f)$.



Per segnali di potenza la dimostrazione è del tutto simile, e valida anche per realizzazioni di processi. In particolare, se il processo è ergodico, la media di insieme $m_{X_1 X_2}^{(1,1)}$ risulta uguale alla media temporale calcolata per ogni realizzazione del processo, e pertanto lo spettro di densità di potenza di un processo si ottiene trasformando la funzione di autocorrelazione calcolata come media di insieme, oppure trasformando quella calcolata come media temporale per una delle sue realizzazioni.

Pertanto, grazie al Teorema di Wiener, è possibile ottenere $\mathcal{P}_x(f)$ anche per processi, oppure fare “la prova del nove” per segnali di energia o periodici.

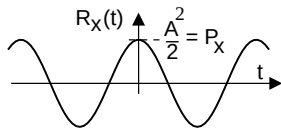
7.2.2 Esempi

Processo armonico. E' definito in base ad una sua generica realizzazione

$$x(t, \theta) = A \cos(2\pi f_0 t + \theta)$$

che, se il parametro θ è una variabile aleatoria uniformemente distribuita tra $-\pi$ e π (ossia $p_\theta(\theta) = \frac{1}{2\pi} \text{rect}_{2\pi}(\theta)$), descrive un processo ergodico.

Sappiamo che una sua realizzazione (ad esempio quella con $\theta = 0$) ha una densità di potenza $\mathcal{P}_x(f) = \frac{A^2}{4} [\delta(f - f_0) + \delta(f + f_0)]$. Possiamo quindi ottenere l'autocorrelazione senza dover svolgere l'integrale:



$$\mathcal{R}_x(t) = \mathcal{F}^{-1}\{\mathcal{P}_x(f)\} = \frac{A^2}{4} [e^{j2\pi f_0 t} + e^{-j2\pi f_0 t}] = \frac{A^2}{2} \cos(2\pi f_0 t)$$

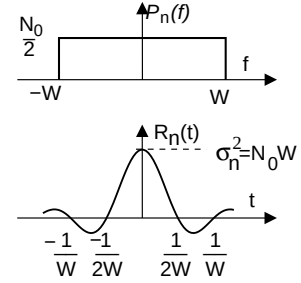
Il risultato trovato conferma che l'autocorrelazione di un segnale periodico è periodica; riflettiamo dunque sulla circostanza che anche un seno, od un coseno con qualunque altra fase, avrebbe avuto la stessa $\mathcal{R}_x(t)$. Ciò è d'altra parte evidente, avendo tutti questi segnali uguali densità $\mathcal{P}_x(f)$.

Processo gaussiano bianco limitato in banda. Il processo $n(t)$ è chiamato *bianco* perché costante in frequenza, e descritto da una densità di potenza pari a

$$\mathcal{P}_n(f) = \frac{N_0}{2} \text{rect}_{2W}(f)$$

in cui W è l'occupazione di banda a frequenze positive. In tali ipotesi otteniamo

$$\mathcal{R}_n(t) = \mathcal{F}^{-1}\{\mathcal{P}_n(f)\} = N_0 W \text{sinc}(2Wt)$$



In particolare, campioni presi a distanza $T_c = 1/2W$ sono incorrelati (ed essendo il processo gaussiano anche statisticamente indipendenti); questo risultato giustifica, almeno da un punto di vista teorico, una ipotesi che viene spesso fatta: quella di trovare sovrapposti ai campioni di segnale, dei campioni di rumore *statisticamente indipendenti*.

All'aumentare di W , $\mathcal{R}_n(t)$ tende a zero più rapidamente, cosicché il rumore si mantiene correlato per un tempo sempre minore, ovvero due campioni estratti ad una stessa distanza t hanno una correlazione sempre minore. Un risultato simile vale anche più in generale, in quanto l'autocorrelazione $\mathcal{R}_x(t)$ di un qualsiasi processo (tranne nel caso periodico, riconducibile ad una combinazione di processi armonici) tende a 0 con $t \rightarrow \infty$, ovvero da un certo t in poi la correlazione è trascurabile.

Onda PAM. Abbiamo già descritto un generico segnale numerico come una somma di repliche di una funzione $g(t)$, con ampiezze a_n rappresentative dei valori da trasmettere:

$$x(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT + \theta)$$

La presenza della variabile aleatoria θ a distribuzione uniforme tra $\pm \frac{T}{2}$ (per cui $p_\theta(\theta) = \frac{1}{T} \text{rect}_T(\theta)$), rende $x(t)$ un processo ergodico.

Si mostrerà in appendice 7.9.2 che, nelle ipotesi in cui le ampiezze a_n siano determinazioni di variabili aleatorie *indipendenti ed identicamente distribuite*, a media nulla e varianza σ_A^2 , l'autocorrelazione di $x(t)$ vale $\mathcal{R}_x(\tau) = \sigma_A^2 \frac{\mathcal{R}_G(\tau)}{T}$ in cui $\mathcal{R}_G(\tau)$ è l'autocorrelazione di $g(t)$, e dunque

$$\mathcal{P}_x(f) = \sigma_A^2 \frac{\mathcal{E}_G(f)}{T}$$

Osserviamo innanzitutto che è per questa via che si sono caratterizzate le densità di potenza proprie dei codici di linea. Limitandoci a voler interpretare il risultato, notiamo che $\mathcal{E}_G(f)$ è la densità di energia di una singola replica di $g(t)$. La sua ripetizione, con periodo T , fornisce una densità di potenza *media* $\frac{\mathcal{E}_G(f)}{T}$. Se ogni replica di $g(t)$ è moltiplicata per una v.a. indipendente a media nulla e varianza (potenza) σ_A^2 , la densità di potenza $\mathcal{P}_x(f)$ aumenta di egual misura. Un'ultima avvertenza riguarda il fatto che, se gli a_n non sono indipendenti, il risultato è più complesso (vedi appendice 7.9.2).

7.3 Stima spettrale

Il teorema di Wiener ci aiuta qualora di desideri conoscere la densità di potenza per un processo, di cui siamo in grado di stimare o postulare un $m_X^{(1,1)}(\tau) = \mathcal{R}_X(\tau)$. Spesso però si ha a che fare con segnali di cui, pur ricorrendo le ipotesi di appartenenza ad un processo ergodico, si ignorano le statistiche di insieme. Un altro caso tipico è quello di un segnale che, pur se rappresentativo di molti altri, non presenta caratteristiche spettrali costanti nel tempo, e sono proprio le variazioni di queste ultime ad interessare¹⁰. In questi casi, tutto ciò che si può fare è di tentare una stima dello spettro di potenza del segnale, a partire da un suo segmento temporale. Esistono al riguardo tecniche differenti, di cui citiamo solo quella che segue.

7.3.1 Periodogramma

Definiamo un intervallo T in cui isoliamo un segnale $x_T(t) = x(t, \theta) \text{rect}_T(t)$ da una realizzazione $x_T(t, \theta)$. Questo segmento di segnale è di energia, con $\mathcal{E}_{x_T}(f) = |X_T(f)|^2$, e al quale si può attribuire una densità di potenza

$$\mathcal{P}_{x_T}(f) = \frac{|X_T(f)|^2}{T}$$

che viene detta *periodogramma*¹¹. Al tendere di T ad ∞ , il risultato trovato tende alla densità di potenza $\mathcal{P}_x(f) = \lim_{T \rightarrow \infty} \frac{|X_T(f, \theta)|^2}{T}$ della realizzazione $x_T(t, \theta)$ e, se questa appartiene ad un processo ergodico, a quella di un qualunque altro membro.

Nel caso più verosimile in cui T non tende ad infinito, tentiamo di valutare l'errore che si commette ad usare $\mathcal{P}_{x_T}(f)$ come una stima $\hat{\mathcal{P}}_x(f)$ della vera densità $\mathcal{P}_x(f)$ del processo. Per una determinata frequenza f_0 , il valore $\frac{|X_T(f_0)|^2}{T}$ è una variabile aleatoria (dipende infatti da θ), che vorremmo avesse un valore atteso $m_T = E_\theta \{\mathcal{P}_{x_T}(f_0)\}$ pari alla densità del processo ($m_T = \mathcal{P}_x(f_0)$) ed una varianza $\sigma_T^2 = E_\theta \{(\mathcal{P}_{x_T}(f_0) - m_T)^2\}$ che diminuisce la crescere di T ¹². Per verificare se tali proprietà siano soddisfatte, valutiamo il valore atteso del periodogramma, a partire dalle relazioni fornite dal teorema di Wiener, applicato ad $X_T(f)$, e cioè $|X_T(f)|^2 = \mathcal{E}_{x_T}(f) = \mathcal{F}\{\mathcal{R}_{x_T}(\tau)\}$:

$$\begin{aligned} E_\theta \{\mathcal{P}_{x_T}(f)\} &= E_\theta \left\{ \mathcal{F} \left\{ \frac{1}{T} \int_{-\infty}^{\infty} x(t) \text{rect}_T(t) x(t+\tau) \text{rect}_T(t+\tau) dt \right\} \right\} = \\ &= \mathcal{F} \left\{ \frac{1}{T} \int_{-\infty}^{\infty} E_\theta \{x(t) x(t+\tau)\} \text{rect}_T(t) \text{rect}_T(t+\tau) dt \right\} = \\ &= \mathcal{F} \left\{ \mathcal{R}_x(\tau) \frac{1}{T} \int_{-\infty}^{\infty} \text{rect}_T(t) \text{rect}_T(t+\tau) dt \right\} = \mathcal{F} \{\mathcal{R}_x(\tau) \cdot \text{tri}_{2T}(\tau)\} = \\ &= \mathcal{P}_x(f) * T (\text{sinc}(fT))^2 \end{aligned}$$

¹⁰Un esempio può essere un segnale sonoro, ad esempio una voce recitante, per il quale vogliamo studiare le caratteristiche spettrali dei diversi suoni della lingua (i fonemi), per confrontarle con quelle di un'altro individuo, o per ridurre la quantità di dati necessaria a trasmettere il segnale in forma numerica, o per realizzare un dispositivo di riconoscimento vocale.

¹¹Il termine periodogramma ha origine dall'uso che fu inizialmente fatto del risultato, per scoprire tracce di periodicità all'interno di segnali qualsiasi.

¹²Quando il valore atteso di uno stimatore tende al valore vero, si dice che lo stimatore è *non polarizzato* (UNBIASED); se aumentando la dimensione del campione la varianza della stima tende a zero, lo stimatore è detto *consistente*.

che ci ripropone il risultato ottenuto al §3.9.3 relativo al processo di *finestratura*, e che mostra come lo stimatore è *polarizzato*¹³. In base allo scopo con cui è condotta la stima spettrale, può essere opportuno adottare al posto della finestra rettangolare un diverso andamento $w_T(t)$, in modo da calcolare $x_T(t, \theta) = x(t, \theta) \cdot w_T(t)$.

Il calcolo del periodogramma viene svolto mediante una *Trasformata Discreta di Fourier* (DCT), o meglio ancora con una FFT (vedi § 3.9.4). In tal caso, l'aumento di T corrisponde all'aumento del numero di campioni utilizzati, e all'aumento della risoluzione in frequenza ottenibile. In base a queste considerazioni, si può mostrare che la varianza dello stimatore non decresce con T .

7.4 Filtraggio di segnali e processi

Ora che siamo in grado di descrivere da un punto di vista spettrale tutti i segnali, indaghiamo sui valori relativi alla grandezza $y(t) = x(t) * h(t)$ uscente da un filtro con risposta impulsiva $h(t)$. In particolare, ci chiediamo quanto valgano $\mathcal{P}_y$ e $\mathcal{P}_y(f)$, oppure $\mathcal{E}_y$ ed $\mathcal{E}_y(f)$ se $x(t)$ è di energia.

7.4.1 Segnali di energia

Sappiamo che per il teorema di Parseval risulta $\mathcal{E}_y(f) = Y(f) Y^*(f)$; dato che $Y(f) = X(f) H(f)$, allora

$$\mathcal{E}_y(f) = X(f) H(f) X^*(f) H^*(f) = |X(f)|^2 |H(f)|^2 = \mathcal{E}_x(f) |H(f)|^2$$

A questo punto, eseguendo l'antitrasformata di Fourier di ambo i membri, si ottiene:

$$\mathcal{R}_y(\tau) = \mathcal{F}^{-1} \{ \mathcal{E}_y(f) \} = \mathcal{F}^{-1} \{ \mathcal{E}_x(f) |H(f)|^2 \} = \mathcal{R}_x(\tau) * \mathcal{R}_h(\tau)$$

Il risultato ottenuto è posto in evidenza perché è valido anche per i due casi successivi di segnale periodico e di processo, e mostra come l'autocorrelazione dell'uscita di un filtro è pari alla convoluzione tra l'autocorrelazione dell'ingresso e quella della risposta impulsiva.

A corollario di quanto esposto, sussistono i seguenti integrali¹⁴, equivalenti ai fini del calcolo dell'energia totale:

$$\begin{aligned} \mathcal{E}_y &= \int_{-\infty}^{\infty} \mathcal{E}_y(f) df = \int_{-\infty}^{\infty} \mathcal{E}_x(f) |H(f)|^2 df = \int_{-\infty}^{\infty} \mathcal{R}_x(\tau) \mathcal{R}_h(\tau) d\tau = \\ &= \int_{-\infty}^{\infty} \mathcal{R}_x(\tau) \mathcal{R}_h^*(\tau) d\tau = \mathcal{R}_y(0) \end{aligned}$$

Pertanto è possibile utilizzare tutte queste come relazioni di equivalenza, quando si ha necessità di determinare una grandezza a partire da altre note.

¹³Ci consola verificare che, per $T \rightarrow \infty$, la polarizzazione tende a scomparire.

¹⁴La terza uguaglianza sussiste in virtù del teorema di Parseval, mentre la quarta è valida se $\mathcal{R}_h(\tau)$ è reale, ossia $h(t)$ è idealmente realizzabile, come illustrato nel seguito.

7.4.2 Segnali periodici

Se in ingresso ad un filtro è presente un segnale periodico, il segnale in uscita è anch'esso periodico¹⁵, e per esso è valido lo sviluppo in serie di Fourier

$$y(t) = \sum_n Y_n e^{j2\pi n F t}$$

in cui i coefficienti Y_n possono esprimersi in termini dei coefficienti di Fourier dell'ingresso X_n e dei valori della risposta in frequenza (vedi sezione 7.5) come $Y_n = X_n H(nF)$, ovvero in modulo e fase come

$$|Y_n| = |X_n| |H(nF)|; \quad \arg(Y_n) = \arg(X_n) + \arg(H(nF))$$

Pertanto scriviamo che $\mathcal{P}_y(f) = \sum_n |Y_n|^2 \delta(f - nF)$; dato che $|Y_n|^2 = |H(nF)|^2 |X_n|^2$, si ha

$$\mathcal{P}_y(f) = \sum_n |H(nF)|^2 |X_n|^2 \delta(f - nF) = |H(f)|^2 \mathcal{P}_x(f)$$

Di nuovo, antitrasformando si ottiene $\mathcal{R}_y(\tau) = \mathcal{R}_x(\tau) * \mathcal{R}_H(\tau)$.

7.4.3 Processi ergodici

Anche in questo caso, si verifica (in appendice 7.9.4) che $m_Y^{(1,1)}(\tau) = m_X^{(1,1)}(\tau) * m_H^{(1,1)}(\tau)$, e dunque

$$\mathcal{P}_y(f) = |H(f)|^2 \mathcal{P}_x(f)$$

Passiamo quindi a calcolare le altre grandezze rappresentative:

Media:

$$\begin{aligned} m_Y &= E\{y(t)\} = E\{x(t) * h(t)\} = E\{x(t)\} * h(t) = m_X \int_{-\infty}^{\infty} h(\tau) d\tau \\ &= m_X H(0) \end{aligned}$$

è pari cioè a quella dell'ingresso, moltiplicata per il guadagno in continua del filtro.

Potenza: in linea generale, è sempre vero che $\mathcal{P}_y = \sigma_y^2 + (m_y)^2$; inoltre, valgono le relazioni

$$\begin{aligned} \mathcal{P}_y &= \mathcal{R}_y(0) = \int_{-\infty}^{\infty} \mathcal{P}_y(f) df = \int_{-\infty}^{\infty} \mathcal{P}_x(f) |H(f)|^2 df = \\ &= \int_{-\infty}^{\infty} \mathcal{R}_x(\tau) \mathcal{R}_H(\tau) d\tau \end{aligned}$$

Se ad esempio $x(t)$ è un processo bianco con $\mathcal{P}_x(f) = \frac{N_0}{2}$ e $\mathcal{R}_x(\tau) = \frac{N_0}{2} \delta(\tau)$, ed inoltre $x(t)$ è a media nulla, allora risulta

$$\sigma_y^2 = \mathcal{P}_y = \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2 df = \frac{N_0}{2} \mathcal{R}_H(0)$$

mentre $\mathcal{P}_y(f) = \frac{N_0}{2} |H(f)|^2$: pertanto, il processo in uscita dal filtro *non è più bianco*, ed in questo caso il processo si dice *colorato*. A questo fenomeno corrisponde anche una modifica della funzione di autocorrelazione: questa infatti non è più un impulso, ma vale $\mathcal{R}_y(\tau) = \frac{N_0}{2} \mathcal{R}_H(\tau)$; mentre prima quindi (per il processo bianco)

¹⁵Questa affermazione è vera, purché il filtro *non* presenti fenomeni di non linearità: in tal caso infatti, vale la sovrapposizione degli effetti, e non possono prodursi armoniche diverse da quelle in ingresso.

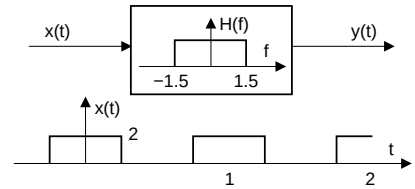
due suoi valori estratti in istanti diversi erano comunque incorrelati, la colorazione introdotta dal filtro ha causato l'insorgenza di una dipendenza statistica tra i valori estratti ad intervalli τ ¹⁶.

Densità di probabilità. A riguardo della $p_Y(y)$ non si può dire nulla di generale, tranne che essa dipende dalla $p_X(x)$ di ingresso e dalle operazioni compiute dal filtro; la sua espressione esatta va però determinata caso per caso. L'unico caso in cui la teoria fornisce una regola certa, è relativo ancora una volta al caso di processi gaussiani: posti in ingresso ad un filtro, producono in uscita un processo anch'esso gaussiano. Questo risultato è una diretta conseguenza della definizione stessa di processo gaussiano, come risultato della somma di infinite cause identicamente distribuite: dato che l'integrale di convoluzione effettivamente esegue una somma tra infiniti valori di ingresso, in linea di principio è lecito affermare che nel transito in un filtro la densità di probabilità dell'uscita si "gaussianizza".

7.4.4 Esercizio

Sia dato il filtro in figura, con $H(f) = \text{rect}_3(f)$, al cui ingresso viene posto il segnale $x(t) = 2 \sum_{n=-\infty}^{\infty} \text{rect}_{\frac{1}{2}}(t - n)$. Calcolare:

- 1) la potenza in ingresso $\mathcal{P}_x$,
- 2) la potenza in uscita $\mathcal{P}_y$,
- 3) l'espressione di $y(t)$.



Risposte

- 1) Calcoliamo la media temporale: $\mathcal{P}_x = \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt = \frac{1}{T} \int_{-1/4}^{1/4} 2^2(t) dt = \frac{4}{2} = 2$;
- 2) Sappiamo che $\mathcal{P}_y = \int_{-\infty}^{\infty} \mathcal{P}_y(f) df$, in cui $\mathcal{P}_y(f) = |Y(f)|^2$, ed a sua volta $Y(f) = X(f)H(f)$. Calcoliamo perciò innanzitutto

$$X(f) = F\{2\text{rect}_{\tau}(t) * \sum_{n=-\infty}^{\infty} \delta(t - nT)\} = 2\tau \cdot \text{sinc}(f\tau) \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T});$$

essendo $\tau = \frac{1}{2}$ e $T = 1$, risulta

$$X(f) = \text{sinc}\left(\frac{f}{2}\right) \sum_{n=-\infty}^{\infty} \delta(f - n) = \sum_{n=-\infty}^{\infty} X_n \delta(f - n),$$

con $X_n = \text{sinc}\left(\frac{n}{2}\right)$. Dunque, dato che gli unici impulsi che cadono entro la risposta in frequenza $H(f)$ sono quelli per $f = -1, 0$ e 1 , si ha:

$$Y(f) = X(f)H(f) = \sum_{n=-1}^1 X_n H(n) \delta(f - n)$$

il cui modulo quadro fornisce $\mathcal{P}_y(f) = \sum_{n=-1}^1 |X_n|^2 |H(n)|^2 \delta(f - n)$, e pertanto si ottiene

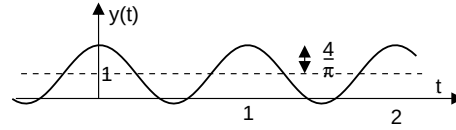
$$\mathcal{P}_y = \int_{-\infty}^{\infty} \mathcal{P}_y(f) df = \left(\frac{\sin -\frac{\pi}{2}}{-\frac{\pi}{2}}\right)^2 + 1 + \left(\frac{\sin \frac{\pi}{2}}{\frac{\pi}{2}}\right)^2 = 1 + 2\left(\frac{2}{\pi}\right)^2 = 1.811$$

¹⁶Il motivo di questo risultato può essere meglio compreso ricordando che l'integrale di convoluzione calcola i singoli valori in uscita da un filtro, come dipendenti da tutti gli ingressi passati, ognuno pesato con il valore della risposta impulsiva relativo al ritardo tra ingresso passato ed uscita presente. Pertanto, anche se i singoli valori in ingresso sono incorrelati, quelli di uscita (distanti meno della durata della risposta impulsiva) condividono una porzione di storia comune, e quindi i loro valori non sono più indipendenti.

$$3) y(t) = \sum_{n=-1}^1 X_n H(n) e^{j2\pi n t} =$$

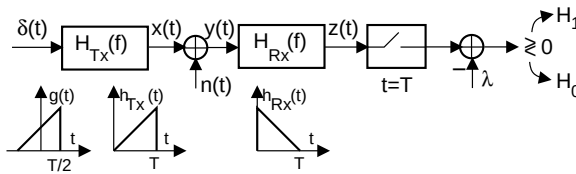
$$= 1 + \frac{2}{\pi} (e^{j2\pi t} + e^{-j2\pi t}) = 1 + \frac{4}{\pi} \cos 2\pi t.$$

Notiamo come il filtro abbia lasciato passare solamente la componente continua e la prima armonica.



7.4.5 Filtro Adattato

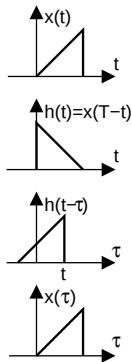
Supponiamo di trasmettere un segnale $x(t)$, generato (ad esempio) facendo transitare un impulso $\delta(t)$ in un filtro con risposta impulsiva $h_{Tx}(t) = g(t - \frac{T}{2})$, e di ricevere lo stesso segnale in presenza di rumore gaussiano a media nulla $n(t)$, con densità spettrale $\mathcal{P}_N(f) = \frac{N_0}{2}$.



Il ricevitore decide per la presenza (H_1) o l'assenza (H_0) del segnale (vedi figura), confrontando l'uscita di un filtro di ricezione $h_{Rx}(t)$, ad un istante $t = T$, con una soglia λ . Si dimostra che la probabilità di errore può essere resa minima, così come

il rapporto SNR all'istante di decisione, se si sceglie

$$H_{Rx}(f) = H_{Tx}^*(f) = G^*(f) e^{-j2\pi f \frac{T}{2}} \quad \text{ovvero} \quad h_{Rx}(t) = h_{Tx}^*(-t) = g^*\left(\frac{T}{2} - t\right)$$



Con tali scelte infatti, nel caso H_0 in cui $x(t)$ è assente, risulta $y(t) = n(t)$, e la grandezza di decisione $z(T)$ è una v.a. definita come

$$z(T) = \int_{-\infty}^{\infty} h_{Rx}(\tau) y(T - \tau) d\tau = \int_0^T x^*(T - \tau) n(T - \tau) d\tau = \mathcal{R}_{XN}(0)$$

e dunque presenta valore atteso $m_{z(T)}$ nullo¹⁷ e varianza¹⁸ $\sigma_{z(T)}^2 = \frac{N_0}{2} \mathcal{E}_G$. Se invece (H_1) il segnale è presente, si ottiene

$$z(T) = \int_0^T x^*(T - \tau) [x(T - \tau) + n(T - \tau)] d\tau = \mathcal{R}_X(0) + \mathcal{R}_{XN}(0)$$

dando luogo ad un valore $m_{z(T)} = \mathcal{E}_G$, mentre per $\sigma_{z(T)}^2$ vale lo stesso risultato precedente¹⁹.

¹⁷Risulta $m_{z(T)} = E\{\mathcal{R}_{XN}(0)\} = E\left\{\int_0^T x^*(t) n(t) dt\right\}$, che è (in senso statistico) pari a zero se $E\{n(t)\} = 0$.

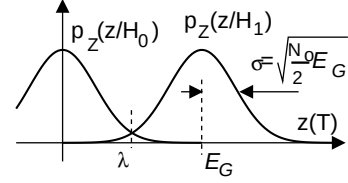
¹⁸Risulta $\sigma_{z(T)}^2 = E\{z^2(T)\} = \mathcal{R}_Z(0)$, e sappiamo che $\mathcal{R}_Z(0) = \mathcal{R}_N(\tau) * \mathcal{R}_{H_{Rx}}(\tau)|_{\tau=0}$; ora risulta $\mathcal{R}_Z(\tau) = \frac{N_0}{2} \delta(\tau) * \mathcal{R}_{H_{Rx}}(\tau) = \frac{N_0}{2} \mathcal{R}_{H_{Rx}}(\tau)$ e quindi

$$\sigma_{z(T)}^2 = \frac{N_0}{2} \mathcal{R}_{H_{Rx}}(0) = \frac{N_0}{2} \int_{-\infty}^{\infty} h_{Rx}^*(t) h_{Rx}(t) dt = \frac{N_0}{2} \mathcal{E}_G$$

in quanto $h_{Rx}(t)$ ha la stessa energia di $g(t)$.

¹⁹Infatti, ora risulta $m_{z(T)} = E\{\mathcal{R}_X(0) + \mathcal{R}_{XN}(0)\}$, in cui il secondo termine è nullo come già osservato, mentre il primo non è aleatorio e vale $\mathcal{R}_X(0) = \int_0^T x^*(t) x(t) dt = \mathcal{E}_G$, in quanto il segnale $x(t)$ ha la stessa energia di $g(t)$. Per ciò che riguarda $\sigma_{z(T)}^2$, osserviamo che essendo $h_{Rx}(t)$ lineare, l'uscita si ottiene come sovrapposizione degli effetti delle due cause $x(t)$ ed $n(t)$, e la componente aleatoria dell'uscita è dovuta al solo $n(t)$; pertanto, la sua varianza è la stessa calcolata per il caso di segnale assente.

La figura a fianco mostra la d.d.p. di $z(T)$ nelle ipotesi H_0 ed H_1 , che risulta gaussiana se $n(t)$ lo è. Notiamo che $m_{z(T)/H_1} = \mathcal{E}_G$ non dipende dalla particolare $g(t)$ adottata, né dalla sua durata T , ma solo dalla sua energia, ed è per questo che il filtro di ricezione è detto *adattato*.



Il rapporto $\frac{(m_{z(T)/H_1})^2}{\sigma_{z(T)}^2}$ rappresenta²⁰ l' SNR all'istante di decisione, ed il suo valore $SNR = \frac{(\mathcal{E}_G)^2}{\frac{N_0}{2}\mathcal{E}_G} = \frac{2\mathcal{E}_G}{N_0}$ costituisce il massimo²¹ che si può ottenere tra tutti filtri di ricezione $h_{Rx}(t)$ di energia pari a $\mathcal{E}_G$.

A commento del risultato, notiamo innanzi tutto che questo è valido in presenza di rumore bianco, mentre se questo è colorato, l' SNR diminuisce, ed il filtro ottimo va determinato in altro modo. Osserviamo poi che, se non fosse presente rumore, l'andamento dell'uscita del filtro adattato sarebbe proprio pari alla funzione di autocorrelazione di $g(t)$, che viene campionata in corrispondenza del suo massimo. Infine, notiamo che la $H_{Rx}(f)$ non presenta modulo costante e fase lineare, dato che lo scopo qui non è quello di preservare la forma d'onda in transito, ma di massimizzare l' SNR all'istante di decisione.

Segnalazione antipodale Desiderando voler distinguere tra due possibili messaggi (ad es, x_1 ed x_2), e di voler rendere minima la probabilità di errore, la scelta *ottima* consiste nell'adottare $x_2(t) = -x_1(t)$, e di impiegare al ricevitore un filtro adattato ad $x_1(t)$. In tal modo, all'istante di campionamento si avrà un valore positivo o negativo, a seconda se sia presente x_1 od x_2 .

Segnalazione ortogonale Dovendo trasmettere N diversi messaggi ($x_1, x_2, \dots, x_N$), associamo ad ognuno di essi una forma d'onda $x_i(t)$ tale che $\int x_i(t) x_j(t) dt = 0$ con $i \neq j$, ovvero in modo che i segnali $x_i(t)$ siano *ortogonali*. In tal caso il ricevitore ottimo è costituito da un banco di filtri adattati, ognuno ad una diversa $x_i(t)$, e la ricezione di una delle forme d'onda $x_i(t)$ non produce nessuna uscita sui filtri del banco per $j \neq i$. La decisione su chi sia stato trasmesso, viene presa valutando quale dei filtri presenti il valore massimo in corrispondenza dell'istante di campionamento, realizzando un ricevitore a correlazione (vedi § 11.1.3).

²⁰Il significato fisico del rapporto indicato può essere meglio visualizzato considerandone le radici quadrate, ossia $\frac{m_{z(T)/H_1}}{\sigma_{z(T)}}$, che costituisce il rapporto tra l'uscita per $t = T$ in presenza di solo segnale, e la deviazione standard di tale valore introdotta dal rumore. In altre parole, è indicativo della separazione tra le gaussiane riportate in figura. Pertanto, maggiore è questo rapporto, e minore sarà la probabilità di errore.

²¹Consideriamo infatti il caso in cui si abbia una $H_{Rx}(f) = H(f)$ generica. In presenza di solo segnale, si ottiene

$$|z(T)|^2 = |h(t) * x(t)|_{t=T}|^2 = \left| \int_{-\infty}^{\infty} h(\tau) x(T - \tau) d\tau \right|^2 = \left| \int_{-\infty}^{\infty} H(f) X(f) e^{j2\pi fT} df \right|^2$$

in cui nell'ultimo passaggio si è invocato il teorema di Parseval. Riportiamo ora la disuguaglianza di Schwartz (vedi § 2.4.1), che afferma la relazione $\left| \int_{-\infty}^{\infty} a(\theta) b(\theta) d\theta \right|^2 \leq \int_{-\infty}^{\infty} |a(\theta)|^2 d\theta \cdot \int_{-\infty}^{\infty} |b(\theta)|^2 d\theta$, con l'uguaglianza solo se $a(\theta) = k \cdot b^*(\theta)$. Se ora associamo $H(f)$ ad $a(\theta)$ e $X(f) e^{j2\pi fT}$ a $b^*(\theta)$, otteniamo che

$$(m_{z(T)/H_1})^2 \leq \int_{-\infty}^{\infty} |H(f)|^2 df \cdot \int_{-\infty}^{\infty} |X(f)|^2 df$$

con l'uguaglianza solo se $H(f) = kX^*(f)$, ossia se $H(f)$ è adattata a $X(f)$. Scegliendo $k = 1$, i due integrali a prodotto hanno lo stesso valore, pari a $\mathcal{E}_G$.

7.5 Caratteristiche dei sistemi fisici

Abbiamo già osservato come il legame $\mathcal{T}[x(t)] = y(t)$ definito in 1.7.1 tra ingresso $x(t)$ ed uscita $y(t)$ di un sistema fisico sia definito, in base alla conoscenza della risposta impulsiva, dall'integrale di convoluzione $y(t) = x(t) * h(t)$, che descrive il risultato di una operazione di filtraggio. Alla luce di questo risultato, torniamo ad analizzare le proprietà enunciate in 1.7.1, assieme a due altre.

Linearità Un sistema è lineare se, in presenza di una combinazione lineare di ingressi, l'uscita è la combinazione lineare delle uscite, ossia

$$\mathcal{T}\left[\sum_i a_i x_i(t)\right] = \sum_i a_i \mathcal{T}[x_i(t)]$$

Il legame ingresso-uscita descritto dall'integrale di convoluzione è di tipo lineare, in virtù della distributività dell'integrale. Cogliamo l'occasione per notare che un sistema descritto da una risposta impulsiva $h(t)$ con estensione temporale non nulla, è *con memoria*, in quanto i singoli valori di uscita dipendono da tutti i valori di ingresso "raccolti" dalla risposta impulsiva.

Permanenza Un sistema è permanente se la risposta ad un ingresso ritardato è ritardata anch'essa. Nel caso in cui la risposta impulsiva non cambi nel tempo, il sistema è permanente.

Realizzabilità ideale Un sistema è idealmente realizzabile se $h(t)$ è reale.

Realizzabilità fisica La proprietà di osservare valori di uscita che dipendono solo dagli ingressi passati, è automaticamente verificata se $h(t) = 0$ con $t < 0$. Abbiamo già osservato (nota 6 a pag. 54) come sistemi non realizzabili fisicamente possano essere approssimati da sistemi realizzabili accettando un ritardo dell'uscita.

Stabilità Si può mostrare che la proprietà di stabilità equivale alla condizione $\int |h(t)| dt < \infty$, ovvero che $h(t)$ sia un segnale impulsivo. Notiamo che questa circostanza garantisce l'esistenza della sua trasformata $H(f)$.

Risposta in frequenza Se un sistema, oltre che stabile, è anche idealmente realizzabile, allora $H(f) = H^*(-f)$, e dunque è sufficiente conoscere la parte a frequenze positive indicata con $H^+(f)$, dato che l'altra metà è ottenibile mediante una operazione di coniugazione. Questo fatto permette di misurare

$$H(f) = M(f) e^{j\varphi(f)}$$

(ossia modulo e fase di $H(f)$), che prende il nome di *risposta in frequenza*, utilizzando come ingresso una funzione sinusoidale con ampiezza A e fase θ note: $x(t) = A \cos(2\pi f_0 t + \theta)$. Il segnale in uscita sarà ancora una cosinusoide²² e avrà ampiezza $A \cdot M(f_0)$ e fase $\varphi(f_0) + \theta$; pertanto potremo ricavare

$$M(f_0) = \frac{\max\{y(t)\}}{\max\{x(t)\}}, \quad \text{e} \quad \varphi(f_0) = \arg\{y(t)\} - \arg\{x(t)\}$$

²²Svolgiamo i calcoli nel dominio della frequenza:

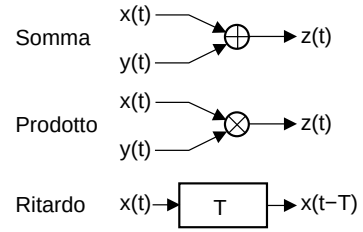
$$X(f) = \frac{A}{2} (e^{j\theta} \delta(f - f_0) + e^{-j\theta} \delta(f + f_0));$$

Ripetendo il procedimento per diverse f_0 , possiamo “campionare” $H(f)$.

7.6 Unità di elaborazione

Mentre un filtro viene definito, in base alla precedente classificazione, un operatore lineare con memoria, può essere utile considerare altri operatori elementari, che funzionano come “mattoni” di operazioni più complesse. Gli operatori elementari, rappresentati in figura, sono la somma, il prodotto, ed il ritardo.

Forniamo dapprima i risultati della combinazione di processi e segnali certi mediante gli operatori introdotti, e quindi mostriamo alcune semplici realizzazioni di *filtri digitali*, definiti nei termini degli operatori elementari.



7.6.1 Prodotto

Nel caso in cui un fattore sia un processo, e l'altro un segnale certo, il risultato (in generale) è un processo *non stazionario*. Infatti ora le medie d'insieme dipendono, istante per istante, dal valore che il segnale certo assume in quell'istante (tranne il caso in cui sia una costante)²³.

Se uno dei due fattori (ad es $y(t) = \tilde{y}$) è una costante, $z(t)$ è un processo della stessa natura di $x(t)$, con media $m_z = m_x \cdot \tilde{y}$, potenza $\mathcal{P}_z = \mathcal{P}_x \cdot \tilde{y}^2$, e autocorrelazione $\mathcal{R}_z(\tau) = \mathcal{R}_x(\tau) \cdot \tilde{y}^2$.

Se i fattori $x(t)$ ed $y(t)$ sono processi statisticamente indipendenti, stazionari e congiuntamente²⁴ ergodici, si ottiene:

Valor medio

$$\begin{aligned} m_z &= E\{z(t)\} = E\{x(t)y(t)\} = E\{x(t)\}E\{y(t)\} = \\ &= m_x \cdot m_y \end{aligned}$$

Potenza totale

$$\begin{aligned} \mathcal{P}_z &= E\{z^2(t)\} = E\{x^2(t)y^2(t)\} = E\{x^2(t)\}E\{y^2(t)\} \\ &= \mathcal{P}_x \cdot \mathcal{P}_y \end{aligned}$$

Varianza

$$Y(f) = X(f)H(f) = \frac{A}{2}M(f_0)(e^{j\theta}e^{j\varphi(f_0)}\delta(f-f_0) + e^{-j\theta}e^{-j\varphi(f_0)}\delta(f+f_0))$$

e antitrasformando si ottiene

$$y(t) = A \cdot M(f_0) \cos(2\pi f_0 t + \varphi(f_0) + \theta)$$

²³Se il segnale certo è periodico, il risultato della moltiplicazione per un processo stazionario dà luogo ad un processo detto *ciclostazionario*, in quanto le statistiche variano nel tempo, ma assumono valori identici con periodicità uguale a quella del segnale certo.

²⁴La proprietà di ergodicità congiunta corrisponde a verificare le condizioni ergodiche anche per coppie di valori estratti da realizzazioni di due differenti processi.

$$\begin{aligned}\sigma_z^2 &= E\{(z(t) - m_z)^2\} = \mathcal{P}_z - (m_z)^2 = \\ &= \mathcal{P}_x \cdot \mathcal{P}_y - (m_x \cdot m_y)^2\end{aligned}$$

Funzione di autocorrelazione

$$\begin{aligned}\mathcal{R}_z(\tau) &= E\{z(t)z(t+\tau)\} = E\{x(t)y(t)x(t+\tau)y(t+\tau)\} = \\ &= E\{x(t)x(t+\tau)\}E\{y(t)y(t+\tau)\} = \mathcal{R}_x(\tau) \cdot \mathcal{R}_y(\tau)\end{aligned}$$

In particolare, notiamo che l'incorrelazione di uno dei due processi, per un certo valore di τ , provoca l'incorrelazione del prodotto.

Spettro di densità di potenza

$$\begin{aligned}\mathcal{P}_z(f) &= \mathcal{F}\{\mathcal{R}_z(\tau)\} = \mathcal{F}\{\mathcal{R}_x(\tau) \cdot \mathcal{R}_y(\tau)\} = \\ &= \mathcal{P}_x(f) * \mathcal{P}_y(f)\end{aligned}$$

ossia è pari alla convoluzione tra le densità spettrali dei fattori. Notiamo quindi che la densità di potenza del prodotto presenta una occupazione di banda maggiore di quella dei singoli fattori.

Densità di probabilità Si calcola con le regole per il cambiamento di variabile, illustrate al § 5.6.6. Nel caso in cui i due processi siano statisticamente indipendenti, il risultato è

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(\theta) p_Y\left(\frac{z}{\theta}\right) \frac{d\theta}{\theta}$$

In Appendice (pag. 180), troviamo l'applicazione di questi risultati al calcolo della densità di potenza dell'onda PAM.

7.6.2 Somma

Anche in questo caso, se un termine è un processo e l'altro un segnale certo, la somma è (in generale) un segnale non stazionario. Se il segnale certo è costante, si torna al caso stazionario²⁵. Procediamo ora nel calcolo delle solite grandezze, con l'ipotesi che $x(t)$ ed $y(t)$ siano processi *statisticamente indipendenti*.

Valore medio

$$\begin{aligned}m_z &= E\{x(t) + y(t)\} = E\{x(t)\} + E\{y(t)\} \\ &= m_x + m_y\end{aligned}$$

Potenza totale

$$\begin{aligned}\mathcal{P}_z &= E\{(x(t) + y(t))^2\} = E\{x^2(t)\} + E\{y^2(t)\} + 2E\{x(t) \cdot y(t)\} \\ &= \mathcal{P}_x + \mathcal{P}_y + 2m_x m_y\end{aligned}$$

Pertanto, se almeno uno dei due processi è a media nulla, le potenze dei due processi si sommano.

²⁵Come per il prodotto, se il segnale certo è periodico, la somma si dice *ciclostazionaria* perché la dipendenza temporale non è assoluta, ma periodica.

Varianza

$$\begin{aligned}\sigma_z^2 &= E\{(z(t) - m_z)^2\} = \mathcal{P}_z - (m_z)^2 = \mathcal{P}_x + \mathcal{P}_y + 2m_x m_y - (m_x + m_y)^2 = \\ &= \mathcal{P}_x - (m_x)^2 + \mathcal{P}_y - (m_y)^2 = \sigma_x^2 + \sigma_y^2\end{aligned}$$

Autocorrelazione

$$\begin{aligned}\mathcal{R}_z(\tau) &= E\{z(t)z(t+\tau)\} = E\{x(t)x(t+\tau)\} + \\ &E\{y(t)y(t+\tau)\} + E\{x(t)y(t+\tau)\} + E\{x(t+\tau)y(t)\} = \\ &= \mathcal{R}_x(\tau) + \mathcal{R}_y(\tau) + 2m_x m_y\end{aligned}$$

Osserviamo come per $\tau = 0$ si ritrovi il valore della potenza totale.

Spettro di densità di potenza

$$\begin{aligned}\mathcal{P}_z(f) &= \mathcal{F}\{\mathcal{R}_z(\tau)\} \\ &= \mathcal{P}_x(f) + \mathcal{P}_y(f) + 2m_x m_y \delta(f)\end{aligned}$$

Densità di probabilità applicando le regole del cambiamento di variabile, nel caso di $x(t)$ ed $y(t)$ indipendenti, si ottiene

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(\theta) p_Y(z - \theta) d\theta = p_X(x) * p_Y(y)$$

La relazione esprime l'importante risultato che la densità di probabilità della somma di variabili aleatorie è la *convoluzione* tra le densità dei termini della somma. Esempio alla nota²⁶.

7.7 Filtri Digitali

I filtri digitali sono una particolare classe di filtri analogici, per i quali l'integrale di convoluzione si riduce in una sommatoria²⁷. La figura sottostante mostra una particolare architettura di filtro digitale denominata *filtro trasversale*, per il quale è facile verificare che la risposta impulsiva ha espressione

$$h(t) = \sum_{n=0}^N c_n \delta(t - nT)$$

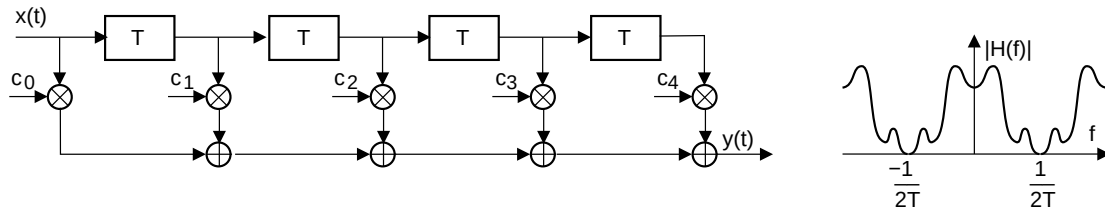
in cui N è l'ordine del filtro²⁸. E' altrettanto facile verificare che la funzione di trasferimento ha espressione

$$H(f) = \mathcal{F}\{h(t)\} = \sum_{n=0}^N c_n e^{-j2\pi f n T}$$

²⁶Se x ed y sono ad es. due v.a. a distribuzione uniforme tra $\pm\Delta$, la loro somma ha densità di probabilità triangolare con base 2Δ . Pertanto, nel lancio di 2 dadi il risultato più probabile è 7. Infatti può essere ottenuto come 6+1, 5+2, 4+3, 3+4, 2+5, 1+6, ovvero in 6 modi diversi, ognuno con probabilità $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$ e dunque $Pr\{7\} = 6 \cdot \frac{1}{36} = \frac{1}{6}$.

²⁷In questo senso, i filtri digitali presentano strette analogie con i *filtri numerici* (che non trattiamo), che operano sulla base di dati campionati, svolgendo le operazioni elementari direttamente via software (o mediante apposito hardware dedicato).

²⁸I coefficienti c_n vengono indicati nei testi inglesi come *taps* (rubinetti) in quanto possono essere pensati "spillare" frazioni del segnale. Per effetto di un processo di trasposizione linguistica, gli stessi coefficienti in italiano vengono a volte indicati discorsivamente come *tappi* (!).



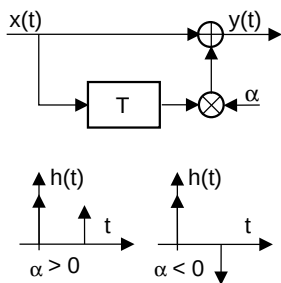
Una tale architettura può essere rappresentativa di un effettivo fenomeno naturale, come ad esempio la presenza di “echi” nel segnale ricevuto. Come si vede, $H(f)$ risulta periodica (in frequenza) con periodo $F = \frac{1}{T}$. Pertanto, l'architettura può essere adottata per *sintetizzare* una risposta in frequenza desiderata in una certa gamma di frequenze, e “qualsiasi” altrove. In particolare, se il segnale di ingresso $x(t)$ è limitato in banda tra $\pm \frac{1}{2T}$, l'azione filtrante ha luogo appunto nella sola banda del segnale.

Per *sintetizzare* il filtro a partire dall'andamento desiderato di $H(f)$ nella banda di interesse, si calcolano i coefficienti c_k mediante la formula

$$c_k = T \int_{-1/2T}^{1/2T} H(f) e^{j2\pi f k T} df$$

che è proprio (a parte un segno) l'espressione dei coefficienti per lo sviluppo in serie di Fourier di un periodo di segnale! Ovviamente, se $H(f)$ è qualsiasi, occorrerebbero un numero infinito di coefficienti c_k ; usandone un numero inferiore (finito) si produce una approssimazione della $H(f)$ desiderata²⁹.

7.7.1 Filtro trasversale del 1° ordine



E' descritto dalla architettura in figura, a cui corrisponde una risposta impulsiva

$$h(t) = \delta(t) + \alpha \delta(t - T)$$

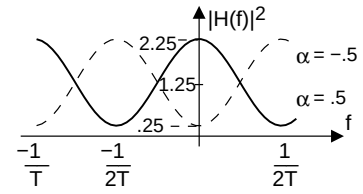
la cui trasformata è $H(f) = 1 + \alpha e^{-j2\pi f T}$ ⁽³⁰⁾, e quindi

$$\begin{aligned} |H(f)|^2 &= (1 + \alpha \cos 2\pi f T)^2 + (\alpha \sin 2\pi f T)^2 = \\ &= 1 + 2\alpha \cos 2\pi f T + \alpha^2 (\cos^2 2\pi f T + \sin^2 2\pi f T) = \\ &= 1 + \alpha^2 + 2\alpha \cos 2\pi f T \end{aligned}$$

²⁹Chiaramente il troncamento della serie di coefficienti c_k avverrà in modo simmetrico rispetto a c_0 , prendendo cioè sia gli indici positivi che quelli negativi. Viceversa, nello schema di filtro trasversale si usano solo coefficienti con indici ≥ 0 . Nel caso in cui l' $H(f)$ da cui partiamo sia reale (e pari), allora i c_k sono una serie reale pari, il che garantisce un filtro idealmente realizzabile, ma la cui $h(t) = \sum_{k=-N/2}^{N/2} c_k \delta(t - kT)$ necessita di una traslazione temporale per risultare anche fisicamente realizzabile. Se invece $H(f)$ ha un andamento qualunque, non si può dire nulla a riguardo di eventuali simmetrie per i coefficienti c_k .

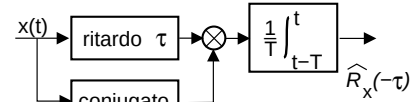
³⁰In questo caso $H(f)$, pur risultando a simmetria coniugata ($H(f) = H^*(-f)$), è complessa. Pertanto, i coefficienti c_k ottenibili dalla relazione $c_k = T \int_{-1/2T}^{1/2T} H(f) e^{j2\pi f k T} df$ sono reali, ma non necessariamente pari. Svolgendo i calcoli, si ha: $c_k = T \int_{-1/2T}^{1/2T} (1 + \alpha e^{j2\pi f T}) e^{j2\pi f k T} df = T \int_{-1/2T}^{1/2T} e^{j2\pi f k T} df + \alpha T \int_{-1/2T}^{1/2T} e^{j2\pi f (k-1) T} df$. Il primo integrale è nullo per $k \neq 0$, mentre il secondo per $k \neq 1$, in quanto le funzioni integrande hanno media nulla sull'intervallo $1/T$; pertanto $c_0 = 1$ e $c_1 = \alpha$, esattamente come è definita la risposta impulsiva.

In figura, è riportato l'andamento di $|H(f)|^2$ per due valori di $\alpha = \pm 0.5$, ed osserviamo che nell'intervallo di frequenze $|f| < \frac{1}{2T}$ può assumere un comportamento passa-alto oppure passa-basso³¹. Notiamo inoltre che ponendo $\alpha = -1$ si ottiene un differenziatore, in grado di rimuovere dall'ingresso segnali periodici di periodo T .



7.7.2 Stima della autocorrelazione di un processo ergodico

L'architettura riportata in figura *non* è un filtro (c'è un moltiplicatore anziché un sommatore) ma viene illustrata ora per "similitudine". Si tratta di uno schema idoneo a misurare una stima $\hat{R}_x(\tau)$ della funzione di autocorrelazione temporale di una realizzazione $x(t)$ di un processo. Variando il ritardo τ si ottiene $\hat{R}_x(\tau)$ per diversi valori dell'argomento. Dunque sarà poi possibile calcolare $\hat{P}_x(f) = \mathcal{F}\{\hat{R}_x(\tau)\}$, che rappresenta una stima dello spettro di densità di potenza di una qualunque realizzazione del processo, in virtù della proprietà di ergodicità.



7.7.3 Filtro digitale a risposta impulsiva infinita del 1° ordine

Lo schema riportato a fianco è un filtro di tipo *ricor-sivo* in cui l'uscita dipende dalle *precedenti uscite*. La sua risposta impulsiva ha durata infinita ed è pari a

$$h(t) = \sum_{n=0}^{\infty} \alpha^n \delta(t - nT)$$

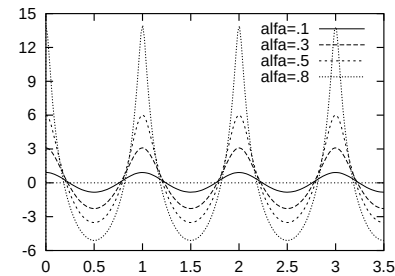
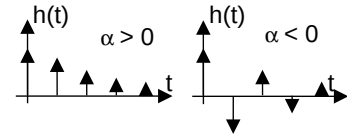
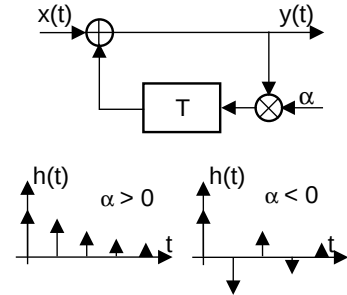
La sua funzione di trasferimento risulta pari a $H(f) = \sum_{n=0}^{\infty} \alpha^n e^{-j2\pi f nT}$ e, ricordando la formula dello sviluppo in serie geometrica $\sum_{n=0}^{\infty} \beta^n = \frac{1}{1-\beta}$, si può scrivere

$$H(f) = \sum_{n=0}^{\infty} (\alpha e^{-j2\pi f T})^n = \frac{1}{1 - \alpha e^{-j2\pi f T}}$$

Notiamo subito che il filtro è stabile purché $|\alpha| < 1$, altrimenti si ha una uscita anche senza ingresso, ovvero uscita infinita con ingresso limitato. Per ciò che riguarda $|H(f)|^2$, otteniamo

$$\begin{aligned} |H(f)|^2 &= \frac{1}{(1 - \alpha \cos 2\pi f T)^2 + (\alpha \sin 2\pi f T)^2} = \\ &= \frac{1}{1 + \alpha^2 - 2\alpha \cos 2\pi f T} \end{aligned}$$

Il grafico soprastante rappresenta $10 \log_{10} |H(f)|^2$, con $T = 1$ e diversi valori di α . Osserviamo infine che il caso $\alpha = 1$ corrisponde ad avere un integratore perfetto che, ad esempio, produce una rampa in uscita, se in ingresso c'è un gradino.



³¹In altre parole, l'andamento ondulatorio di $|H(f)|^2$ rende il filtro idoneo a diversi utilizzi, in funzione dell'andamento in frequenza del segnale di ingresso.

7.8 Filtri Analogici

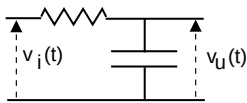
Sono ottenuti mediante componenti elettrici a costanti concentrate come condensatori, induttori e resistori.

Applicando la trasformata di *Laplace* alle equazioni differenziali che descrivono la relazione ingresso-uscita, si ottiene una funzione di trasferimento razionale del tipo

$$H(s) = \frac{\sum_{i=0}^N a_i s^i}{\sum_{j=0}^M b_j s^j}$$

(in cui $N \leq M$), definita su di un piano complesso $s = \sigma + j2\pi f$. Ponendo $s = j2\pi f$ si ottiene la funzione di trasferimento in f : $H(f) = H(s = j2\pi f)$. Questo procedimento è valido solo se il filtro è stabile, che nel dominio di Laplace equivale a richiedere che tutti i poli di $H(s)$ siano a sinistra dell'asse immaginario.

7.8.1 Filtro analogico ad un polo



Nella figura a fianco è riportato un filtro RC di tipo passa basso, per il quale la relazione tra $v_u(t)$ e $v_i(t)$ è descritta da una risposta impulsiva con espressione

$$h(t) = \frac{1}{RC} e^{-\frac{t}{RC}}$$

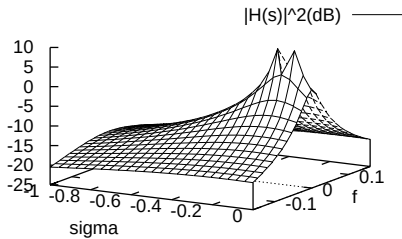
L'analisi del circuito mostra che la funzione di trasferimento risulta

$$H(f) = \mathcal{F}\{h(t)\} = \frac{1/j\omega C}{R + 1/j\omega C} = \frac{1}{1 + j2\pi f RC}$$

ovvero, nel dominio di Laplace

$$H(s) = \frac{1}{1 + sRC}$$

Pertanto, $H(s)$ presenta un polo in $s = -\frac{1}{RC}$ che fa sì che $H(s)|_{s=-\frac{1}{RC}} = \infty$.

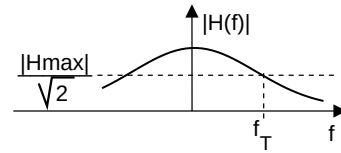


La figura a lato mostra l'andamento di $|H(s)|^2$, disegnata in decibel, e con $RC = 8$. Come evidente, $|H(s)|^2$ può essere pensata come una sorta di "cono vulcanico" attorno al polo, le cui falde, quando intersecate dal piano verticale infisso sull'asse $j2\pi f$, individuano la funzione di trasferimento in frequenza $H(f) = H(s = j2\pi f)$. Come si vede dalla figura, $H(f)$ risulta di tipo passa basso, con fianchi tanto più ripidi quanto più il polo è vicino all'origine.

7.8.2 Frequenza di taglio

Definiamo *frequenza di taglio* di un filtro la frequenza f_T per la quale

$$|H(f_T)| = \frac{|H_{Max}|}{\sqrt{2}}$$



Nel caso del filtro RC, si ha $|H_{Max}| = H(0) = 1$ e dunque scriviamo

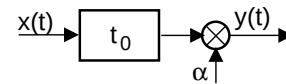
$$|H(f)| = \frac{1}{\sqrt{1 + (2\pi f RC)^2}} = \frac{1}{\sqrt{1 + \left(\frac{f}{f_T}\right)^2}}$$

in cui $f_T = \frac{1}{2\pi RC}$, pari quindi alla frequenza di taglio (infatti $|H(f_T)| = \frac{1}{\sqrt{1+1}} = \frac{1}{\sqrt{2}}$).

Notiamo anche che $|H(f_T)|^2 = \frac{1}{2}$ e dunque $|H(f_T)|^2|_{dB} = -3 \text{ dB}$; per questo la frequenza di taglio è indicata anche come frequenza a 3 dB.

7.8.3 Assenza di distorsioni lineari

Quali proprietà devono essere verificate da un filtro affinché l'uscita non differisca dall'ingresso per più di un fattore di scala ed un ritardo? Si richiede cioè che $y(t) = \alpha x(t - t_0)$, ovvero $Y(f) = \alpha X(f) e^{-j2\pi f t_0}$ e perciò



$$H(f) = \frac{Y(f)}{X(f)} = \alpha e^{-j2\pi f t_0}$$

dunque: $h(t) = \alpha \delta(t - t_0)$. Pertanto le richieste nel tempo si riflettono su di una risposta in frequenza con *modulo costante* e *fase lineare* (quantomeno nella banda del segnale).

7.9 Appendici

7.9.1 Coefficiente di Correlazione

I diagrammi di esempio presentati alla nota (2) a pag. 160, basano la valutazione di quanto una coppia di v.a. x ed y siano correlate, sul calcolo del coefficiente di correlazione ρ_{xy} , che ha valori compresi tra $+1$ e -1 , ed è definito come

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

In tal modo, si opera una normalizzazione del valore della covarianza σ_{xy} , rispetto alle deviazioni standard σ_x e σ_y delle due v.a., rendendo così il valore di ρ indipendente dalla estensione dei valori assunti da x ed y .

Il coefficiente ρ si presta ad una interessante interpretazione geometrica, una volta identificate (vedi § 2.4.1) la deviazione standard σ_x come la norma $\|\bar{x}\|$ di x , e la covarian-

za σ_{xy} come il prodotto scalare $(\bar{x}, \bar{y})$ tra x ed y ³². In tale contesto, possiamo indicare due v.a. come *ortogonali* se risulta $\sigma_{xy} = 0$, mentre un valore $\sigma_{xy} = \pm 1$ indica che una delle due v.a. è costantemente uguale all'altra, a meno di un coefficiente costante. Notiamo che l'ortogonalità $\sigma_{xy} = 0$ esprime unicamente l'assenza di legami di tipo *lineare* tra x ed y , come esemplificato dal caso F) della nota (2) a pag. 160.

Per ultimo citiamo l'estensione formale del risultato noto come *diseguaglianza di Schwartz* (pag. 25), una volta che al coefficiente di correlazione ρ_{xy} sia stato associato il concetto di coseno tra x ed y : tale posizione deriva dall'essere $-1 < \rho_{xy} < 1$, e permette di asserire che $|\sigma_{xy}| \leq \sigma_x \sigma_y$.

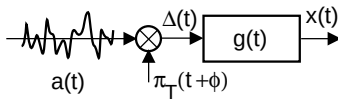
7.9.2 Onda PAM

Accingiamoci a verificare che il processo

$$x(t) = \sum a_n g(t - nT + \theta)$$

con θ v.a. aleatoria uniformemente distribuita tra $\pm \frac{T}{2}$, e con gli a_n v.a. statisticamente indipendenti, a media nulla ed identicamente distribuite, sia descritto da uno spettro di densità di potenza pari a

$$\mathcal{P}_x(f) = \sigma_A^2 \frac{|G(f)|^2}{T}$$



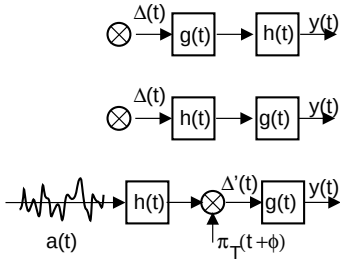
A tale scopo, adottiamo lo schema funzionale in figura. Un processo bianco $a(t)$ è moltiplicato per un processo impulsivo di periodo T , dando luogo ad un processo impulsivo $\Delta(t)$ con

$$\mathcal{R}_\Delta(\tau) = \mathcal{R}_\pi(\tau) \cdot \mathcal{R}_a(\tau) = \mathcal{R}_\pi(\tau) \cdot \sigma_A^2 \delta(\tau) = \sigma_A^2 \mathcal{R}_\pi(0) \delta(\tau)$$

Osserviamo ora che essendo $\pi_T(t)$ periodico, anche $\mathcal{R}_\pi(\tau)$ lo sarà. Ma

$$\mathcal{R}_\pi(0) = \frac{1}{T} \int_{-T/2}^{T/2} \delta(t) dt = \frac{1}{T}$$

e dunque $\mathcal{R}_\pi(\tau) = \frac{1}{T} \sum \delta(\tau - nT)$. Infatti, se $\mathcal{R}_\pi(\tau)$ è periodica di periodo T , deve assumere gli stessi valori, mentre per $\tau \neq nT$ è nulla per qualunque realizzazione. A parte ciò, vale $\mathcal{R}_\pi(0) = \frac{1}{T}$, e dunque



$$\mathcal{R}_\Delta(\tau) = \sigma_A^2 \frac{\delta(\tau)}{T}$$

Il segnale $x(t)$ pertanto ha $\mathcal{P}_x(f) = \mathcal{P}_\Delta(f) |G(f)|^2 = \sigma_A^2 \frac{|G(f)|^2}{T}$, come anticipato.

Ci chiediamo ora cosa accade, se il segnale passa per un secondo filtro $h(t)$ (vedi figura a lato): $y(t) = h(t) * x(t)$ e quindi, trasformando:

$$\mathcal{P}_y(f) = \mathcal{P}_x(f) |H(f)|^2 = \sigma_A^2 \frac{|G(f)|^2 |H(f)|^2}{T}$$

³²L'analogia non è poi troppo peregrina, considerando che se x è estratta da un processo ergodico a media nulla, la sua varianza σ_x^2 coincide con la potenza del segnale da cui è estratta, mentre se x ed y sono estratte da segnali congiuntamente ergodici, la covarianza σ_{xy} coincide con la funzione di intercorrelazione definita in (7.1) per segnali di energia.

Qualora si invertano tra loro h e g , la proprietà commutativa della convoluzione ci assicura di ottenere lo stesso risultato; al contrario, scambiando di posto $h(t)$ ed il moltiplicatore, le aree degli impulsi non sono più v.a. indipendenti, ed il risultato cambia. Infatti:

$$\mathcal{R}_{\Delta'}(\tau) = \mathcal{R}_{\pi}(\tau) \cdot \mathcal{R}_H(\tau) * \mathcal{R}_a(\tau) = \mathcal{R}_{\pi}(\tau) \cdot \sigma_A^2 \mathcal{R}_H(\tau)$$

e dunque ora $\Delta'(t) = \sum b_n \delta(t - nT + \theta)$ con i b_n non più indipendenti, ma $E\{b_n b_m\} = \sigma_A^2 \mathcal{R}_H((n - m)T)$; si ottiene dunque

$$\mathcal{P}_y(f) = |G(f)|^2 \mathcal{P}_{\Delta'}(f)$$

Dato però che $\mathcal{P}_{\Delta'}(f) = \sigma_A^2 \mathcal{P}_{\pi}(f) * |H(f)|^2 = \sigma_A^2 \frac{1}{T^2} (\sum_n \delta(f - \frac{n}{T})) * |H(f)|^2 = \frac{\sigma_A^2}{T^2} * \sum_n |H(f - \frac{n}{T})|^2$, risulta³³

$$\mathcal{P}_y(f) = \frac{\sigma_A^2}{T^2} |G(f)|^2 \sum_n \left| H\left(f - \frac{n}{T}\right) \right|^2$$

Dunque attenzione, se i simboli non sono indipendenti !

7.9.3 Potenza di un segnale dati

Al § 5.5.3 si è affermato che la potenza di un segnale

$$s(t) = \sum_n a_n g(t - nT)$$

in cui $g(t)$ è una caratteristica di Nyquist a coseno rialzato con roll-off γ , e gli a_i sono una sequenza di v.a. discrete, statisticamente indipendenti, ed uniformemente distribuite su L livelli in una dinamica $-\frac{\Delta}{2} \leq a_i \leq \frac{\Delta}{2}$, ha valore

$$\mathcal{P}_s = \frac{\Delta^2}{12} \frac{L+1}{L-1} \left(1 - \frac{\gamma}{4}\right)$$

Nella precedente appendice, si è mostrato che per lo stesso segnale risulta $\mathcal{P}_s(f) = \sigma_A^2 \frac{|G(f)|^2}{T}$, e dunque

$$\mathcal{P}_s = \int \mathcal{P}_s(f) df = \int \sigma_A^2 \frac{|G(f)|^2}{T} df$$

Svolgendo i relativi calcoli, si può mostrare che $\int |G(f)|^2 df = T(1 - \frac{\gamma}{4})$, e quindi $\mathcal{P}_s = \sigma_A^2 (1 - \frac{\gamma}{4})$; resta pertanto da calcolare σ_A^2 :

$$\sigma_A^2 = (a_i \text{ a media nulla}) = E_X \{a_i^2\} = \sum_{i=0}^{L-1} p_{a_i} \cdot a_i^2 = (a_i \text{ equiprobabili}) =$$

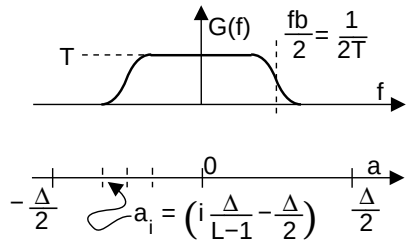
$$= \frac{1}{L} \sum_{i=0}^{L-1} \left(i \frac{\Delta}{L-1} - \frac{\Delta}{2}\right)^2 = \frac{\Delta^2}{L} \sum_{i=0}^{L-1} \left(\frac{i^2}{(L-1)^2} + \frac{1}{4} - \frac{i}{L-1}\right) =$$

$$= \frac{\Delta^2}{L} \left(\frac{L}{4} - \frac{1}{L-1} \sum_{i=0}^{L-1} i + \frac{1}{(L-1)^2} \sum_{i=0}^{L-1} (i)^2\right) =$$

$$(\text{facendo uso delle relazioni } \sum_{n=1}^N n = \frac{N(N+1)}{2} \text{ e } \sum_{n=1}^N n^2 = \frac{N(N+1)(2N+1)}{6})$$

$$= \frac{\Delta^2}{L} \left(\frac{L}{4} - \frac{1}{L-1} \frac{L(L-1)}{2} + \frac{1}{(L-1)^2} \frac{(L-1)L(2(L-1)+1)}{6}\right) = \Delta^2 \left(\frac{1}{4} - \frac{1}{2} + \frac{2L-2+1}{6(L-1)}\right) =$$

³³Nell'ultima serie di passaggi si è valutato $\mathcal{P}_{\pi}(f)$ come $\frac{1}{T^2} (\sum_n \delta(f - \frac{n}{T}))$, in quanto $\Pi(f) = \mathcal{F}\{\pi(t)\} = \mathcal{F}\{\sum_n \delta(t - nT)\} = \frac{1}{T} \sum_n \delta(f - \frac{n}{T})$.



$$= \Delta^2 \frac{6L-6-12L+12+8L-8+4}{24(L-1)} = \Delta^2 \frac{2L+2}{24(L-1)} = \frac{\Delta^2}{12} \frac{L+1}{L-1}.$$

7.9.4 Autocorrelazione dell'uscita di un filtro

Al §7.4.3 si è affermato che, quando un processo attraversa un filtro, il processo di uscita è caratterizzato da $\mathcal{R}_y(\tau) = \mathcal{R}_x(\tau) * \mathcal{R}_H(\tau)$. Mostriamo che è vero.

$$\begin{aligned} \mathcal{R}_y(t, t+\tau) &= E\{y(t)y(t+\tau)\} = E\left\{ \int h(\alpha)x(t-\alpha)d\alpha \int h(\beta)x(t+\tau-\beta)d\beta \right\} = \\ &= \int \int h(\alpha)h(\beta)E\{x(t-\alpha)x(t+\tau-\beta)\}d\alpha d\beta = (\text{se } x(t) \text{ è stazionario}) \\ &= \int h(\alpha) \int h(\beta)\mathcal{R}_x(\tau+\alpha-\beta)d\beta d\alpha = \int h(\alpha)\mathcal{R}_{xy}(\tau+\alpha)d\alpha = \mathcal{R}_{xy}(\tau) * h(-\tau) \end{aligned}$$

in cui $\mathcal{R}_{xy}(\tau) = \mathcal{R}_x(\tau) * h(\tau)$ è l'intercorrelazione tra $x(t)$ ed $y(t)$.

Scritto in altra forma: $\mathcal{R}_y(\tau) = \mathcal{R}_x(\tau) * h(\tau) * h(-\tau)$, e dunque antitrasformando si ottiene $\mathcal{P}_y(f) = \mathcal{P}_x(f) \cdot H(f) \cdot H^*(f) = \mathcal{P}_x(f) \cdot |H(f)|^2 = \mathcal{F}\{\mathcal{R}_x(\tau) * \mathcal{R}_H(\tau)\}$.

7.9.5 Grafici di esempio

Appresso sono riportati i grafici di forma d'onda, dell'autocorrelazione, delle densità spettrale e densità di probabilità, per alcuni segnali tipici.

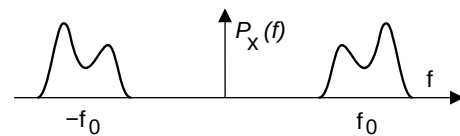
Segnale	Forma d'onda	Autocorrelazione	Densità spettrale	Densità di probabilità
Sinusoide				
Onda quadra				
Impulso rettangolare				
Triangolare				
Dente di sega				
Rumore gaussiano bianco				
Rumore gaussiano limitato in banda				

Capitolo 8

Segnali modulati

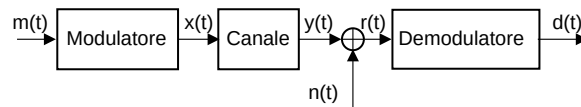
8.1 Generalità

I segnali *modulati*, di cui occuperemo ora, sono anche indicati come segnali in *banda traslata*, per la particolarità di esibire una occupazione spettrale limitata ad una banda ristretta attorno ad una frequenza f_0 chiamata *portante*.



Questi segnali sono prodotti mediante un *processo di modulazione*, che consiste nell'operare una trasformazione su di un segnale $m(t)$ *modulante* in modo da ottenerne un secondo $x(t)$ *modulato*, il cui contenuto spettrale è ora idoneo alla trasmissione mediante il canale a disposizione.

Il segnale *ricevuto* $r(t)$ (a cui è sovrapposto un processo di rumore $n(t)$) deve quindi essere *de-modulato* (operazione che, se applicata ad $x(t)$, restituirebbe $m(t)$) per ottenere $d(t)$, che rappresenta il segnale trasmesso, più eventuali distorsioni $\varepsilon(t)$. Per evidenziare la situazione, scriviamo il segnale demodulato come $d(t) = m(t) + \varepsilon(t)$, in cui $\varepsilon(t) = Dem\{n(t)\} + Dem\{y(t) - x(t)\}$: evidentemente, le distorsioni hanno origine sia dal risultato della demodulazione del rumore in ingresso al demodulatore, sia dagli effetti che la demodulazione ha sugli errori introdotti, dal canale, sul segnale in transito.



Il processo di modulazione è spesso associato ad una trasmissione radio, ma può rendersi necessario e/o utile anche per trasmissioni in cavo. In generale, individuiamo almeno tre situazioni in cui è necessario l'impiego di segnali modulati:

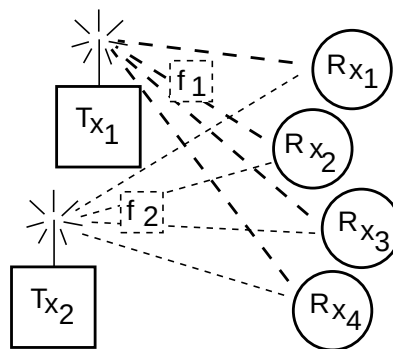
- 1) il canale non permette la trasmissione di frequenze contigue all'origine
- 2) il canale presenta un comportamento ideale (modulo costante e fase lineare) solo in determinati intervalli di frequenza
- 3) il canale presenta disturbi additivi solo in determinate regioni di frequenza.

8.1.1 Multiplazione a divisione di frequenza - FDM

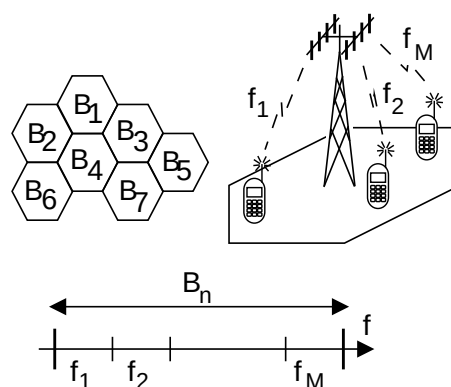
Come intuibile dal nome, si tratta di una tecnica di trasmissione in cui più comunicazioni contemporanee avvengono condividendo lo stesso mezzo fisico, ed impegnando ognuna una diversa regione di frequenze, per il semplice motivo che se utilizzassero tutte le

medesima banda, costituirebbero termini di interferenza reciproca. Molto spesso tutti i segnali multiplati sono di natura simile, ed ognuno è il risultato di una modulazione operata con una diversa frequenza portante. Portiamo ad esempio tre casi:

Collegamenti Punto-Multipunto. E' una topologia che si verifica ad esempio nel caso di trasmissioni televisive o radiofoniche (dette trasmissioni *Broadcast*), in cui ogni emittente (in figura indicata come T_x) trasmette a tutti i ricevitori (R_x) sintonizzati sulla propria portante (i famosi "canali" della TV), mentre altre emittenti utilizzano contemporaneamente lo stesso mezzo trasmissivo, occupando canali centrati ad altre frequenze.



Accesso Multiplo. E' la problematica tipica delle comunicazioni mobili, e quindi dei "telefonini". In tal caso, il territorio è suddiviso in *celle*, per ognuna delle quali è definita una regione di radiofrequenze (B_n) dedicata alla comunicazione tra i terminali ed un unico ripetitore. All'interno della cella, la banda a disposizione è suddivisa tra più canali, ognuno associato ad una diversa portante (f_i), che vengono usate a turno dai terminali che desiderano comunicare¹.



Sotto certi aspetti, questo caso è in qualche modo antitetico rispetto al precedente, e potrebbe essere indicato come collegamento *multipunto-punto*. In realtà la situazione è un pò più complessa, e gli aspetti qualificanti da un punto di vista sistemistico sono i protocolli di rete, necessari per consentire le fasi di richiesta di accesso, la localizzazione dei radiomobili, e la corretta gestione dell'*handover* (il cambio di cella)².

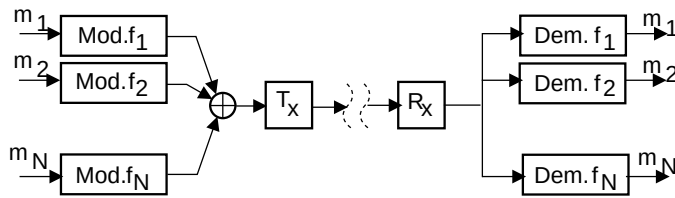
¹Senza entrare nei dettagli, specifichiamo semplicemente che celle limitrofe adottano regioni di frequenza differenti, onde evitare interferenze tra celle; inoltre, nell'ambito di uno stesso canale, è realizzata una struttura di trama, in modo da permettere l'utilizzo dello stesso canale da parte di più terminali contemporaneamente, multiplati a divisione di tempo.

²Un minimo di approfondimento però ci sta bene... Aggiungiamo quindi che la scelta del canale su cui comunicare avviene in base alle condizioni di ricezione del singolo radiomobile che, per effetto di cammini multipli del segnale ricevuto, può ricevere meglio certe portanti che non altre.

La trasmissione che ha luogo su di una portante, inoltre, può aver origine da più terminali, che si ripartiscono la medesima banda a divisione di tempo, in accordo ad una suddivisione di trama dell'asse dei tempi. Pertanto, dopo che un terminale si è aggiudicato una portante ed un intervallo temporale, la trasmissione (attuata mediante una modulazione numerica) ha luogo solo per brevi periodi, in corrispondenza del time-slot di propria pertinenza.

Dato che i singoli terminali si trovano a distanze diverse dal ripetitore di cella, diversi sono i tempi di propagazione del segnale di sincronismo di trama e di time-slot, e dunque l'intervallo temporale che viene "riempito" da ogni terminale giunge al ripetitore con un ritardo variabile. Per questo motivo, i time-slot della trama sono separati da piccoli periodi di inattività, chiamati *intervalli di guardia*, che garantiscono l'assenza di sovrapposizioni temporali delle trasmissioni originate dai diversi terminali.

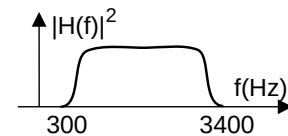
Collegamenti Punto-Punto. E' una forma di multiplazione più simile a quella già illustrata nel caso della TDM, in cui il collegamento avviene tra due località ben determinate. Un insieme di N segnali m_i , $i = 1, 2, \dots, N$, transita su di uno stesso mezzo trasmissivo, occupando ognuno una differente banda, centrata su di una diversa portante f_i , $i = 1, 2, \dots, N$, e può essere individualmente demodulato e separato in ricezione.



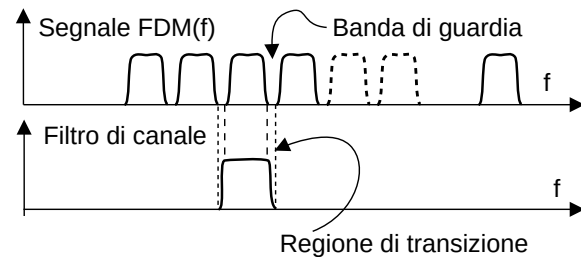
La trasmissione può avvenire sia mediante un collegamento in cavo, che mediante una trasmissione radio; in questa seconda evenienza, il collegamento è spesso indicato come *ponte radio*.

8.1.2 Canale telefonico

Le caratteristiche del collegamento offerto dalla comune linea telefonica tengono conto di molteplici aspetti. Uno di questi, forse il principale³, è la limitazione della banda del canale, per cui la trasmissione è garantita solo in un intervallo di frequenze comprese tra i 300 ed i 3400 Hz, mentre la banda nominale (ovvero l'occupazione di banda in una trasmissione FDM) risulta essere di 4000 Hz⁴. Discutiamo brevemente le origini storiche di tali limitazioni. L'assenza della regione -300÷300 Hz è legata alla presenza, all'interno del telefono, di un componente (detto *ibrido*⁵) che di fatto impedisce la trasmissione di frequenze molto basse.



Per lungo tempo, il traffico telefonico è stato multiplato su collegamenti FDM punto-punto, con i singoli canali modulati AM-BLU (vedi §9.1.2), che richiede la rimozione delle componenti frequenziali più basse. Inoltre, la necessità di separare tra loro i canali multiplati FDM mediante i *filtri di canale* che, per essere economicamente realizzabili, devono presentare una regione di transizione di estensione apprezzabile, ha determinato l'esigenza di prevedere tra due canali contigui un intervallo di frequenze detto *banda di guardia* (pari a 900 Hz) che determina la limitazione a 3400 Hz per la massima frequenza di segnale, in modo da ottenere $300 + (4000 - 3400) = 900$ Hz.



³Un altro fattore particolarmente rilevante è la *limitazione della potenza* che è possibile immettere su di un singolo collegamento telefonico che, associato al precedente, identifica il canale telefonico come limitato sia in banda che in potenza, e dunque con capacità $C = W \log_2 \left(1 + \frac{P_s}{N_0 W}\right)$ dipendente solo dal livello del rumore. La limitazione in potenza è storicamente motivata da problemi di *diafonia* (interferenza tra comunicazioni) dovuti a fenomeni di induzione elettromagnetica. Attualmente, è legata alla dinamica limitata del segnale che viene campionato e trasmesso in forma numerica.

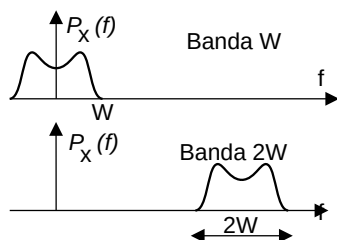
⁴Questo valore massimo nominale determina che la frequenza di campionamento del PCM telefonico è pari a $2 \cdot 4000 = 8000$ campioni al secondo. Utilizzando 8 bit/campione, si ottiene la velocità binaria $f_b = 64000$ campioni/secondo. Velocità inferiori si possono conseguire adottando metodi di codifica di sorgente per il segnale vocale.

⁵L'ibrido telefonico è un trasformatore con quattro porte, che realizza la separazione tra le due vie di comunicazione che viaggiano sullo stesso cavo. Nel caso di una linea ISDN, invece, il telefono stesso effettua la conversione numerica, ed i campioni di voce viaggiano nei due sensi (tra utente e centrale) secondo uno schema a divisione di tempo.

In caso contrario infatti, all'uscita del filtro di canale si troverebbe anche parte del segnale di un canale contiguo producendo una interferenza tra messaggi diversi.

La limitazione in banda di un canale telefonico tra 300 e 3400 Hz è dunque uno dei motivi per i quali la connessione telefonica di un computer ad un fornitore di connettività numerica (ad es. un provider Internet) richiede l'uso di un dispositivo (il modem) che effettui una forma di modulazione sul segnale da trasmettere.

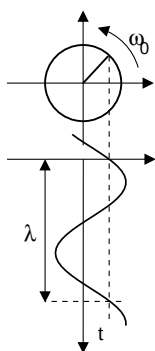
8.1.3 Banda di segnale



La banda occupata da un segnale è la regione di frequenze al di fuori della quale non vi sono componenti energetiche; la sua misura in Hz è indicata come larghezza di banda. Per segnali reali l'occupazione di banda è espressa in termini del solo contenuto a frequenze positive; dato che in tal caso lo spettro di potenza è una funzione pari di f , la banda totale è doppia. Tale definizione è pertanto non ambigua, ed in accordo alla comune accezione di frequenza

(positiva); pertanto, viene spesso indicata come *banda a frequenze positive*⁶.

8.1.4 Antenne e lunghezza d'onda



La trasmissione di un segnale via onda radio necessita di una antenna di dimensioni comparabili con quelle della lunghezza d'onda. Quest'ultima quantità (indicata con λ) è pari allo spazio percorso dall'onda in un tempo pari ad un periodo: dato che *spazio* = *velocità* · *tempo*, e considerando che le onde elettromagnetiche si propagano alla velocità della luce ($c = 3 \cdot 10^8$ m/s), si ha $\lambda = c \cdot T = \frac{c}{f}$.

Nel caso di segnali modulati, il valore di f è quello della portante, in quanto in genere il segnale modulato occupa una banda ristretta attorno alla portante. Trasmissioni con frequenze più elevate necessitano di antenne di dimensioni ridotte; se per assurdo trasmettessimo con portante di 300 Hz, occorrerebbe una antenna di dimensioni $\lambda = \frac{c}{f} = \frac{3 \cdot 10^8}{300} = 10^6$ m = 1000 Km !⁷

8.1.5 Trasmissione a banda laterale unica

Consideriamo un segnale $a(t)$ reale e limitato in banda, con $A(f) = A^*(-f)$ (grafico [A]), a simmetria coniugata. Come già illustrato, la conoscenza di $A^+(f) = A(f) \text{rect}_W\left(f - \frac{W}{2}\right)$ è sufficiente a definire $a(t)$ in modo completo, in virtù della simmetria coniugata. Se definiamo $x(t) = a(t) \cos \omega_0 t$, anch'esso reale, otteniamo che $X(f)$ [B], oltre ad essere a simmetria coniugata rispetto all'origine, ha simmetria coniugata anche rispetto ad f_0 : $X^+(f + f_0) = \{X^+(-f + f_0)\}^*$ [C].

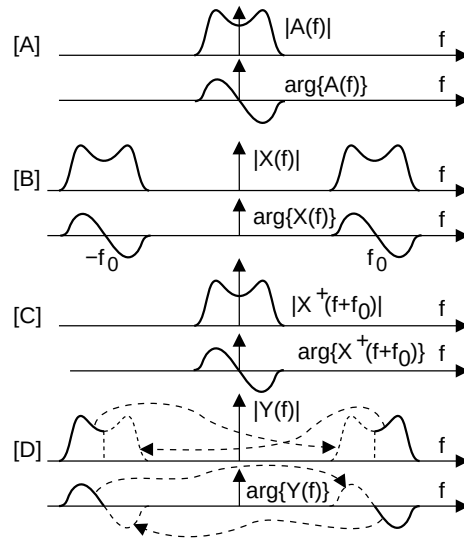
⁶Se un segnale è strettamente limitato in banda, deve avere durata infinita, e viceversa. E' pratica comune, invece, parlare di limitazione in banda anche per segnali di durata finita. Nel fare questo, si considera un $X(f)$ pari a zero per le frequenze f tali che $|X(f)| < \varepsilon$, ovvero considerare anziché $X(f)$ a banda illimitata, una sua finestra in frequenza $X_W(f) = X(f)W(f)$ a banda limitata, la cui antitrasformata $x_W(t)$ è diversa da $x(t)$ (sappiamo infatti che si ha $x_W(t) = x(t) * w(t)$), ma ne costituisce una approssimazione.

⁷Antenne più corte hanno una efficienza ridotta, ma sono ancora buone. Altrimenti la radio AM (540 - 1600 KHz) avrebbe bisogno di $\frac{3 \cdot 10^8}{1000 \cdot 10^3} = 300$ metri !

Questo risultato mostra come sia teoricamente possibile (con una fotocopiatrice ed un paio di forbici!) ottenere il segnale $X(f)$ a partire da un $Y(f)$ [D], con $Y(f)$ ottenuta da $X(f)$ eliminandone metà banda. La ricostruzione di $X(f)$ avviene infatti (freccie tratteggiate) spostando le copie duplicate di $Y^+(f)$ e $Y^-(f)$ come indicato dalle frecce.

Una volta verificata l'esattezza del procedimento illustrato, che ci consente di ricevere per intero $X(f)$ trasmettendone solo metà (cioè $Y(f)$), osserviamo che anche $Y(f)$ è a simmetria coniugata rispetto a zero (ossia $Y(f) = Y^*(-f)$), e quindi la sua antitrasformata $y(t)$ è reale, e dunque può essere realmente trasmesso.

A parte il “dettaglio” di come ricostruire “veramente” $X(f)$ a partire da $Y(f)$, ci chiediamo: esiste una formula per ottenere $y(t)$ in modo *diretto* a partire da $a(t)$? La risposta è positiva; per provarla occorre però affrontare alcune pagine di teoria, che illustrano un metodo di rappresentazione (nel dominio del tempo) per segnali modulati.



8.2 Rappresentazione dei segnali modulati

Questa sezione è dedicata alla teoria che permette di scrivere un qualunque segnale $x(t)$ nella forma

$$x(t) = x_c(t) \cos 2\pi f_0 t - x_s(t) \sin 2\pi f_0 t$$

che assume una particolare rilevanza nel caso in cui $x(t)$ sia un segnale modulato attorno ad f_0 , perchè allora $x_c(t)$ e $x_s(t)$ sono segnali limitati in banda con banda contigua all'origine, e le alterazioni prodotte sul segnale modulato, compresa l'estrazione del messaggio modulante $m(t)$, possono essere descritte mediante operazioni condotte su $x_c(t)$ ed $x_s(t)$.

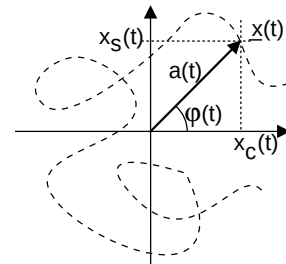
8.2.1 Involuppo complesso e modulazione di ampiezza e/o angolare

Introduciamo l'argomento ricordando come un segnale $x(t) = a \cos(\omega_0 t + \varphi)$ possa essere rappresentato per mezzo del fasore $\underline{x} = a e^{j\varphi}$, mediante la relazione $x(t) = \Re\{\underline{x} e^{j\omega_0 t}\}$. Estendiamo ora il concetto, definendo l'*inviluppo complesso* $\underline{x}(t)$ come un fasore per il quale il modulo a e la fase φ siano funzioni del tempo

$$\underline{x}(t) = a(t) e^{j\varphi(t)}$$

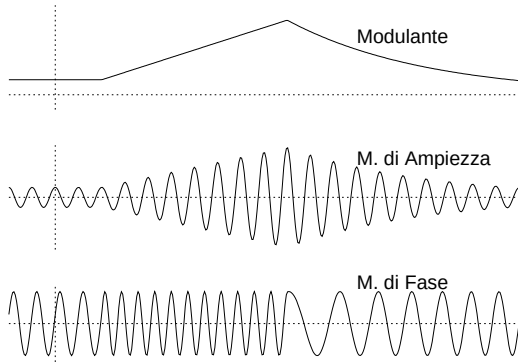
rappresentato nella figura a fianco assieme ad una sua potenziale traiettoria temporale. Ad $\underline{x}(t)$ possiamo associare un segnale reale

$$x(t) = \Re\{\underline{x}(t) e^{j\omega_0 t}\} = a(t) \cos(\omega_0 t + \varphi(t))$$



corrispondente ad imprimere al piano dell'involuppo complesso una rotazione antioraria a velocità angolare ω_0 .

L'involuppo complesso è un potente strumento che permette di descrivere il processo di modulazione in modo semplice ed omogeneo. Ad esempio, il caso (già noto) di traslazione in frequenza del segnale $a(t)$ mediante moltiplicazione per un coseno, corrisponde ad un involuppo complesso $\underline{x}(t) = a(t)$ a fase nulla: ad esso si dà il nome di *modulazione di ampiezza*⁸ per (l'evidente) ragione che l'ampiezza del coseno varia in funzione del segnale $a(t)$; la frequenza $f_0 = \frac{\omega_0}{2\pi}$ prende il nome di *frequenza portante*. Se al contrario consideriamo un involuppo complesso con modulo costante $\underline{x}(t) = e^{j\varphi(t)}$, l'andamento della fase $\varphi(t)$ imprime alla portante non modulata un diverso tipo di modulazione, detto *modulazione di fase*⁹ o *angolare* in quanto il segnale modulante ($\varphi(t)$ in questo caso) altera l'argomento del coseno.



Prima di proseguire, riflettiamo sul risultato mostrato in figura, in cui si considera un segnale modulante $m(t)$ prima costante, poi a rampa lineare, e quindi decrescente. Ponendo $\underline{x}(t) = m(t)$ si ottiene una portante modulata in ampiezza, mentre con $\underline{x}(t) = e^{jm(t)}$ la portante modulata angularmente $x(t) = \cos(2\pi f_0 t + m(t))$ presenta una ampiezza costante, ed una frequenza che nell'intervallo in cui $m(t)$ aumenta linearmente cambia in un valore più elevato, per poi diminuire. In pra-

tica, se $m(t) = \alpha t$, allora l'argomento del coseno diviene $2\pi f_0 t + \alpha t = 2\pi(f_0 + \alpha)t$.

Per meglio descrivere il caso di modulazione angolare, definiamo una *fase istantanea*

$$\psi(t) = 2\pi f_0 t + \varphi(t)$$

ed una *frequenza istantanea*

$$f_i(t) = \frac{1}{2\pi} \frac{d}{dt} \psi(t) = f_0 + \frac{1}{2\pi} \frac{d}{dt} \varphi(t)$$

In questi termini, la modulazione angolare viene distinta in *modulazione di fase* propriamente detta quando

$$\varphi(t) = k_\varphi m(t)$$

mentre viene detta *modulazione di frequenza* quando

$$\varphi(t) = 2\pi k_f \int_{-\infty}^t m(\tau) d\tau$$

in quanto in questo caso è la frequenza istantanea a dipendere direttamente dal segnale modulante: $f_i(t) = f_0 + k_f m(t)$.

8.2.2 Componenti analogiche di bassa frequenza

L'involuppo complesso $\underline{x}(t) = a(t) e^{j\varphi(t)}$ può ovviamente essere espresso tramite le sue parti reale ed immaginaria: $\underline{x}(t) = a(t) \cos \varphi(t) + ja(t) \sin \varphi(t) = x_c(t) + jx_s(t)$. I se-

⁸Indicata anche come AM (*amplitude modulation*).

⁹Indicata anche come PM (*phase modulation*).

gnali $x_c(t) = a(t) \cos \varphi(t)$ ed $x_s(t) = a(t) \sin \varphi(t)$ prendono il nome di componenti analogiche di bassa frequenza di $x(t)$ per un motivo presto chiaro. Intanto sviluppiamo¹⁰ l'espressione del corrispondente segnale modulato:

$$\begin{aligned} x(t) &= \Re \{ \underline{x}(t) e^{j\omega_0 t} \} = a(t) \cos(\omega_0 t + \varphi(t)) = a(t) [\cos \omega_0 t \cos \varphi(t) - \sin \omega_0 t \sin \varphi(t)] \\ &= x_c(t) \cos 2\pi f_0 t - x_s(t) \sin 2\pi f_0 t \end{aligned}$$

Il risultato ottenuto è valido per un qualunque segnale; ma riveste una importanza particolare nel caso in cui $x(t)$ sia di tipo limitato in banda con banda $2W$ centrata attorno ad f_0 , con $W < f_0$: in tal caso, sia $x_c(t)$ che $x_s(t)$ risultano *limitate in banda tra $\pm W$ e contigue all'origine*.

Che sia vero anche il viceversa, può essere verificato in modo intuitivo, partendo da $x_c(t)$ e $x_s(t)$ limitate in banda, e moltiplicandole per coseno e seno. Quest'ultima osservazione ci mostra una via per *sintetizzare* un segnale modulato in ampiezza, od angularmente, od entrambe le cose, mediante il semplice schema circuitale disegnato a fianco, che si basa sulla conoscenza delle componenti analogiche di bassa frequenza, che a loro volta sono ottenibili a partire da $a(t)$ e $\varphi(t)$. Restano comunque (per ora) aperti i seguenti problemi:

- Noto $x(t)$, come ottenere $x_c(t)$ ed $x_s(t)$?
- Noto lo spettro di densità di potenza $\mathcal{P}_x(f)$, che dire di $\mathcal{P}_{\underline{x}}(f)$?

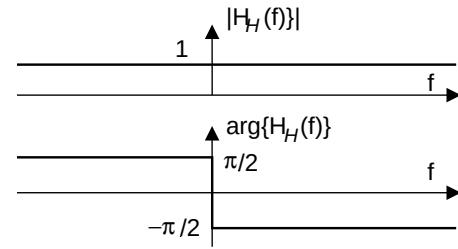
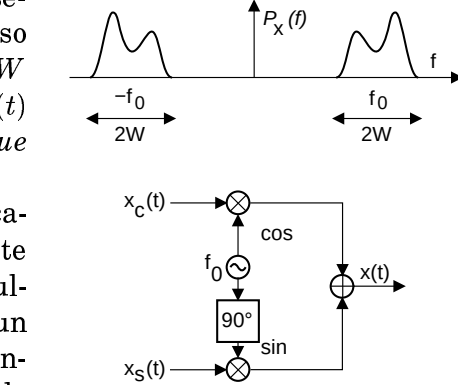
Per procedere oltre occorre introdurre:

8.2.3 Filtro di Hilbert

Il *filtro di Hilbert* è definito da una risposta in frequenza

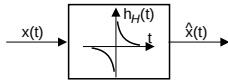
$$H_{\mathcal{H}}(f) = -j \cdot \text{sgn}(f)$$

ed in figura è mostrato l'andamento costante del modulo $|H_{\mathcal{H}}(f)| = 1$, e quello della fase $\angle H_{\mathcal{H}}(f)$ che passa da $\frac{\pi}{2}$ per $f < 0$ a $-\frac{\pi}{2}$ per $f > 0$. Il risultato del passaggio di un segnale $x(t)$ attraverso¹¹



¹⁰Si faccia uso della relazione $\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$.

¹¹L'antitrasformata di Fourier di $H_{\mathcal{H}}(f)$ fornisce l'espressione della risposta impulsiva del filtro di Hilbert $h_{\mathcal{H}}(t) = \mathcal{F}^{-1}\{H_{\mathcal{H}}(f)\} = \frac{1}{\pi t}$, permettendo di scrivere la trasformata di Hilbert nella forma di un integrale di convoluzione: $\hat{x}(t) = \mathcal{H}\{x(t)\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau = x(t) * \frac{1}{\pi t}$.



La realizzazione di un filtro di Hilbert secondo la sua definizione può risultare difficoltosa, a causa della brusca transizione della fase in corrispondenza di $f = 0$. In realtà, il filtro di Hilbert si usa principalmente per segnali modulati, che non presentano componenti spettrali a frequenze prossime allo zero. Pertanto, lo stesso scopo può essere svolto da un diverso filtro, con andamento più dolce della fase, e che presenti gli stessi valori nominali del filtro di Hilbert solamente per le frequenze comprese nella banda di segnale.

il filtro di Hilbert è un un secondo segnale, identificato con la *trasformata di Hilbert* del primo, indicato come $\hat{x}(t) = \mathcal{H}\{x(t)\}$, ed il cui andamento in frequenza ha espressione

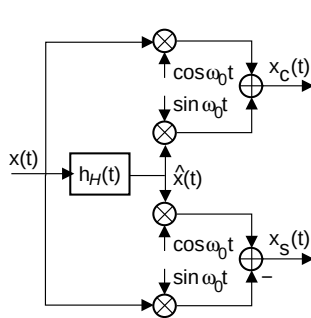
$$\hat{X}(f) = \mathcal{F}\{\hat{x}(t)\} = H_{\mathcal{H}}(f) X(f) = -j \cdot \text{sgn}(f) \cdot X(f)$$

ossia differisce da $X(f)$ per uno sfasamento di $\mp \frac{\pi}{2}$ per frequenze rispettivamente positive o negative.

Senza soffermarci ora sulle proprietà¹² della trasformata di Hilbert, forniamo direttamente un risultato¹³ necessario per proseguire:
$$\begin{cases} \mathcal{H}\{x_c(t) \cos \omega_0 t\} = x_c(t) \sin \omega_0 t \\ \mathcal{H}\{x_s(t) \sin \omega_0 t\} = -x_s(t) \cos \omega_0 t \end{cases}$$
 che ci permette di esprimere la trasformata di Hilbert di un segnale modulato. Quest'ultima, assieme all'espressione di $x(t)$ in funzione di $x_c(t)$ ed $x_s(t)$, costituisce un sistema di due equazione in due incognite:

$$\begin{cases} x(t) = x_c(t) \cos \omega_0 t - x_s(t) \sin \omega_0 t \\ \hat{x}(t) = x_c(t) \sin \omega_0 t + x_s(t) \cos \omega_0 t \end{cases}$$

Il sistema può essere risolto¹⁴, permettendo in definitiva di esprimere le componenti analogiche di bassa frequenza in termini di $x(t)$ e di $\hat{x}(t)$:



$$\begin{cases} x_c(t) = x(t) \cos \omega_0 t + \hat{x}(t) \sin \omega_0 t \\ x_s(t) = \hat{x}(t) \cos \omega_0 t - x(t) \sin \omega_0 t \end{cases}$$

Pertanto le componenti analogiche di bassa frequenza possono essere estratte direttamente da $x(t)$, utilizzando un filtro di Hilbert per ottenere $\hat{x}(t)$, e combinando i due segnali per mezzo di oscillatori in quadratura. Infine, $x_c(t)$ e $x_s(t)$ permettono di risalire alle modulazioni di ampiezza ed angolare:

$$\begin{cases} a(t) = |\underline{x}(t)| = \sqrt{x_c^2(t) + x_s^2(t)} \\ \varphi(t) = \angle \underline{x}(t) = \arctan \frac{x_s(t)}{x_c(t)} \end{cases}$$

Prima di procedere a calcolare $\mathcal{P}_{\underline{x}}(f)$, occorre introdurre l'ulteriore concetto di *segnale analitico*.

¹² Accenniamo brevemente alle principali proprietà della trasformata di Hilbert:

1) $\mathcal{H}\{x(t) = x_0\} = 0$: una costante ha trasformata di Hilbert nulla, e la trasformata di Hilbert è definita a meno di una costante. Il valore medio di $x(t)$ non si ripercuote su $\hat{x}(t)$;

2) $\mathcal{H}\{\mathcal{H}\{x(t)\}\} = \hat{\hat{x}}(t) = -x(t)$: infatti una rotazione di fase pari a π radianti corrisponde ad una inversione di segno;

3) $\int_{-\infty}^{\infty} x(t) \hat{x}(t) dt = 0$: ortogonalità tra un segnale e la sua trasformata di Hilbert;

4) $\mathcal{H}\{x(t) * h(t)\} = \hat{x}(t) * \hat{h}(t) = x(t) * \hat{\hat{h}}(t)$: la trasformata di Hilbert di una convoluzione (cioè dell'uscita di un filtro) è la convoluzione tra un operando trasformato e l'altro no.

¹³ Il risultato a cui ci riferiamo è valido solamente se $x_c(t)$ ed $x_s(t)$ sono limitate in banda $\pm W$ con $W < f_0$; infatti in tal caso le parti a frequenze positive e negative si combinano con le fasi del filtro di Hilbert e si ottiene il risultato.

¹⁴ Potremmo notare come la matrice dei coefficienti costituisca una rotazione di assi, rotazione che "ruota" letteralmente a velocità angolare ω_0 . Tale rotazione stabilisce che le coppie di segnali $(x_c(t), x_s(t))$ e $(x(t), \hat{x}(t))$ rappresentano comunque l'evoluzione dell'involuppo complesso $\underline{x}(t) = a(t) e^{j\varphi(t)}$, mentre $x_c(t)$ e $x_s(t)$ lo rappresentano su due assi ad esso solidali, $x(t)$ e $\hat{x}(t)$ sono definiti su assi ruotanti che tengono conto della frequenza portante.

8.2.4 Segnale Analitico

Il segnale analitico associato al segnale $x(t)$ corrisponde al suo contenuto a frequenze positive $x^+(t)$, che abbiamo già introdotto; si può mostrare che $x^+(t)$ è esprimibile in termini di $\hat{x}(t)$, secondo l'espressione¹⁵:

$$x^+(t) = \frac{1}{2} (x(t) + j\hat{x}(t)) \quad (8.1)$$

Molto utile è anche la relazione che lega il segnale analitico all'involuppo complesso:

$$x^+(t) = \frac{1}{2} \underline{x}(t) e^{j\omega_0 t} \quad (8.2)$$

che si ottiene tenendo conto dalla (8.1), come illustrato alla nota¹⁶. Effettivamente, l'ultima relazione rappresenta il contenuto a frequenze positive di $x(t)$, a patto che $\underline{x}(t)$ sia di banda base con frequenza massima $W < f_0$; in tal caso infatti si ottiene che, trasformando la (8.2), risulta

$$X^+(f) = \mathcal{F}\{x^+(t)\} = \frac{1}{2} \underline{X}(f - f_0)$$

che giace tutta nel semipiano $f > 0$.

Alternativamente alla (8.1), si può ottenere $x^+(t)$ senza utilizzare $\hat{x}(t)$, pensandolo come il risultato del passaggio di $x(t)$ attraverso un filtro $H_{fp}(f)$ ¹⁷ con funzione di trasferimento a gradino unitario:

$$x^+(t) = x(t) * h_{fp}(t)$$

Invertendo la (8.2), otteniamo ora $\underline{x}(t) = 2x^+(t) e^{-j\omega_0 t}$, che trasformata, ci consente di valutare l'espressione di $\underline{X}(f)$:

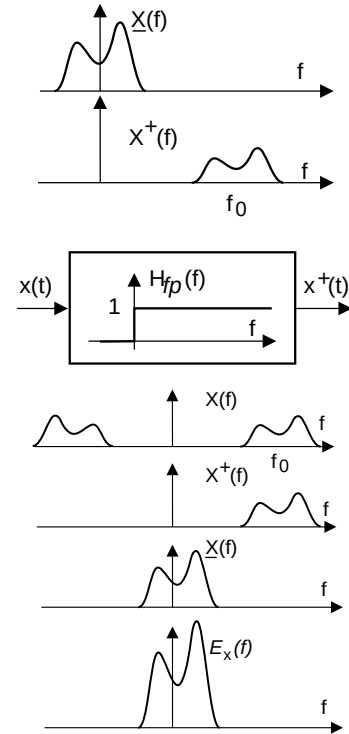
$$\underline{X}(f) = 2X^+(f + f_0)$$

Ricordando ora che $\mathcal{E}_x(f) = |X(f)|^2$, otteniamo

$$\mathcal{E}_{\underline{x}}(f) = 4|X^+(f + f_0)|^2 = 4\mathcal{E}_{x^+}(f + f_0)$$

Un risultato del tutto simile può essere ottenuto per segnali di potenza, ovvero

$$\mathcal{P}_{\underline{x}}(f) = 4\mathcal{P}_{x^+}(f + f_0)$$



¹⁵L'eguaglianza si dimostra valutandola nel dominio nella frequenza, ricordando la definizione di filtro di Hilbert, in quanto risulta:

$$X^+(f) = \frac{1}{2} (X(f) + j\hat{X}(f)) = \begin{cases} \frac{1}{2} \{X(f) + j[-jX(f)]\} = X(f) & \text{con } f > 0 \\ \frac{1}{2} \{X(f) + j[jX(f)]\} = 0 & \text{con } f < 0 \end{cases}$$

infatti, a frequenze negative il prodotto $j \cdot j = -1$ costituisce uno sfasamento di π radianti per tutte le frequenze, provocando l'elisione tra $X(f)$ e $-X(f)$ per tutti i valori $f < 0$.

¹⁶Sviluppando il secondo membro di (8.2) si ottiene:

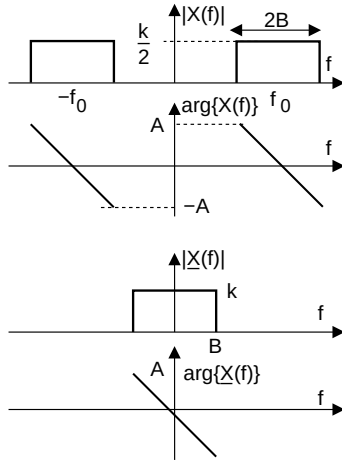
$$\begin{aligned} \frac{1}{2} \underline{x}(t) e^{j\omega_0 t} &= \frac{1}{2} (x_c(t) + jx_s(t)) (\cos \omega_0 t + j \sin \omega_0 t) = \\ &= \frac{1}{2} [(x_c(t) \cos \omega_0 t - x_s(t) \sin \omega_0 t) + j(x_c(t) \sin \omega_0 t + x_s(t) \cos \omega_0 t)] = \frac{1}{2} (x(t) + j\hat{x}(t)) \end{aligned}$$

che corrisponde al secondo membro di (8.1), e quindi a $x^+(t)$.

¹⁷Il pedice $_{fp}$ stà per frequenze positive.

Pertanto, la densità di potenza di $\underline{x}(t)$ si ottiene da quella a frequenze positive di $x(t)$, traslata nell'origine e moltiplicata per 4.

8.2.5 Esempi



• Sia dato il segnale $x(t)$ la cui trasformata è riportata in figura. Quali sono le sue componenti analogiche di bassa frequenza, espresse nel dominio della frequenza?

• Notiamo che $|X^+(f)| = \frac{k}{2} \text{rect}_{2B}(f - f_0)$, e dunque

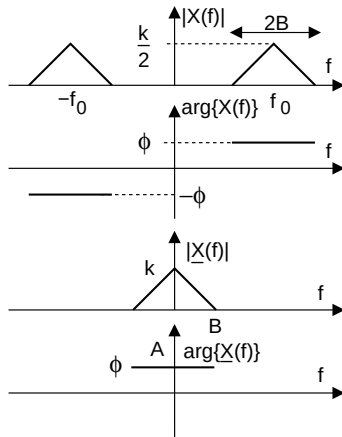
$$|\underline{X}(f)| = 2 |X^+(f + f_0)| = k \text{rect}_{2B}(f)$$

Per la fase si opera una traslazione analogica, ma senza moltiplicare per il fattore 2 che, in quanto fattore, incide solo sul modulo.

Osserviamo ora che $\underline{X}(f)$ ha modulo pari e fase dispari, e dunque la sua antitrasformata è un segnale reale: $\underline{x}(t) = x_c(t) + jx_s(t) = x_c(t)$, ovvero la componente in quadratura $x_s(t)$ è nulla. Pertanto, risulta¹⁸

$$\begin{cases} X_c(f) = k \text{rect}_{2B}(f) e^{-j2\pi \frac{A}{2\pi B} f} \\ X_s(f) = 0 \end{cases}, \text{ ed effettuando l'antitra-}$$

sformata di $X_c(f)$ si ottiene $x_c(t) = 2Bk \frac{\sin 2\pi B(t - \frac{A}{2\pi B})}{2\pi B(t - \frac{A}{2\pi B})}$ in cui la traslazione nel tempo è dovuta alla fase lineare presente in $\underline{X}(f)$.



• Lo stesso problema precedente, ma applicato al nuovo segnale in figura.

• Eseguendo di nuovo le operazioni di traslazione si ottiene l'involuppo complesso riportato in basso. Questa volta la fase di $\underline{X}(f)$ non è dispari, e dunque non si verificano le condizioni di simmetria coniugata, quindi $\underline{x}(t)$ è complesso.

Si ha: $\underline{x}(t) = kB \left(\frac{\sin \pi Bt}{\pi Bt} \right)^2 e^{j\phi}$ e dunque

$$\begin{cases} x_c(t) = kB \left(\frac{\sin \pi Bt}{\pi Bt} \right)^2 \cos \phi \\ x_s(t) = kB \left(\frac{\sin \pi Bt}{\pi Bt} \right)^2 \sin \phi \end{cases} \Rightarrow \begin{cases} X_c(f) = k \left(1 - \frac{|f|}{B} \right) \cos \phi \\ X_s(f) = k \left(1 - \frac{|f|}{B} \right) \sin \phi \end{cases}$$

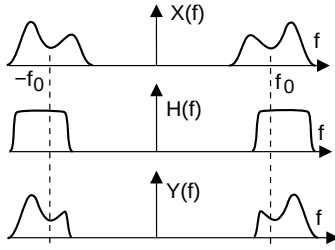
con $|f| < B$.

8.3 Transito dei segnali in banda traslata nei sistemi fisici

Trattiamo qui di come esprimere le trasformazioni di un segnale nel passaggio attraverso un sistema fisico, in funzione degli involuppi complessi del segnale $\underline{x}(t)$ e della risposta impulsiva $\underline{h}(t)$. L'ovvia utilità di questa forma di descrizione è che ci consente di svolgere i calcoli maneggiando segnali di banda base.

¹⁸Approfittiamo dell'occasione per notare che, pur potendo scrivere $\underline{X}(f) = X_c(f) + jX_s(f)$, non è assolutamente lecito dire che $\Re\{\underline{X}(f)\} = X_c(f)$ e $\Im\{\underline{X}(f)\} = X_s(f)$; infatti sia $X_c(f)$ che $X_s(f)$ possono a loro volta essere complessi (mentre $x_c(t)$ e $x_s(t)$ sono necessariamente reali).

8.3.1 Filtraggio



Poniamoci ora nelle condizioni generali di un segnale $x(t)$ che attraversa un filtro $h(t)$, che per entrambi valgano le condizioni di limitazione in banda, e che si adotti una stessa f_0 di riferimento. Anche il segnale in uscita $y(t)$ è dello stesso tipo, e può essere mostrato¹⁹ che il suo inviluppo complesso vale:

$$\underline{y}(t) = \frac{1}{2} \underline{x}(t) * \underline{h}(t)$$

Deriviamo inoltre le espressioni di $y_c(t)$ ed $y_s(t)$ in funzione delle c.a. di b.f. di $x(t)$:

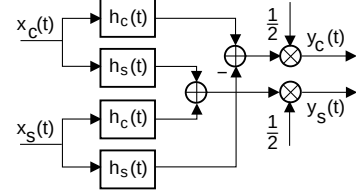
$$\underline{y} = \frac{1}{2} \underline{x} * \underline{h} = \frac{1}{2} [x_c + jx_s] * [h_c + jh_s] = \frac{1}{2} [x_c * h_c - x_s * h_s] + j \frac{1}{2} [x_s * h_c + x_c * h_s]$$

e dunque:

$$\begin{cases} y_c(t) = \frac{1}{2} [x_c(t) * h_c(t) - x_s(t) * h_s(t)] \\ y_s(t) = \frac{1}{2} [x_s(t) * h_c(t) + x_c(t) * h_s(t)] \end{cases}$$

Intermodulazione tra C.A. di B.F.

Osservando il risultato ottenuto, notiamo come sia $y_c(t)$ che $y_s(t)$ dipendano in generale da entrambe $x_c(t)$ e $x_s(t)$: questo fenomeno prende il nome di *intermodulazione* tra componenti analogiche di bassa frequenza, ed è fonte di una distorsione *ineliminabile* in banda base. Infatti, le informazioni contenute in $x_c(t)$ ed $x_s(t)$ sono ora mescolate in modo tale che, anche disponendo sia di $y_c(t)$ che di $y_s(t)$, non possono essere separate. Gli unici casi in cui ciò *non* si verifica sono relativi all'evenienza che $\underline{x}(t)$ oppure $\underline{h}(t)$ presentino *una sola* delle due C.A. di B.F., ossia almeno uno dei due sia solo reale o solo immaginario.



Equalizzazione Ad esempio, se $\underline{h}(t) = h_c(t)$, si ottiene $\begin{cases} y_c(t) = \frac{1}{2} x_c(t) * h_c(t) \\ y_s(t) = \frac{1}{2} x_s(t) * h_c(t) \end{cases}$ e

le componenti trasmesse $x_c(t)$ e $x_s(t)$ possono essere ri-ottenute a partire da $y_c(t)$ e $y_s(t)$ mediante un procedimento di *equalizzazione*, che consiste nell'utilizzo di un filtro $g_{eq}(t)$ la cui risposta in frequenza risulta $G_{eq}(f) = 1/H_c(f)$, permettendo di ottenere

$$\begin{cases} x_c(t) = 2y_c(t) * g_{eq}(t) \\ x_s(t) = 2y_s(t) * g_{eq}(t) \end{cases}.$$

¹⁹Per dimostrare il risultato, mostriamo innanzitutto che il segnale analitico in uscita vale $y^+(t) = x^+(t) * h^+(t)$. Infatti, omettendo di indicare nei passaggi la variabile (t) per compattezza di notazione, risulta

$$x^+(t) * h^+(t) = [x * h_{fp}] * [h * h_{fp}] = [x * h] * [h_{fp} * h_{fp}] = y * h_{fp} = y^+$$

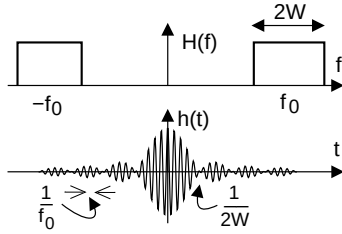
in cui $h_{fp}(t)$ è la risposta impulsiva del filtro necessario ad estrarre il segnale analitico. Non resta ora che mostrare lo sviluppo per il risultato anticipato:

$$\begin{aligned} \frac{1}{2} \underline{x}(t) * \underline{h}(t) &= \frac{1}{2} [2x^+(t) e^{-j\omega_0 t}] * [2h^+(t) e^{-j\omega_0 t}] = \\ &= 2 \int_{-\infty}^{\infty} x^+(\tau) e^{-j\omega_0 \tau} h^+(t - \tau) e^{-j\omega_0 (t - \tau)} d\tau = \\ &= 2e^{-j\omega_0 t} \int_{-\infty}^{\infty} x^+(\tau) h^+(t - \tau) d\tau = 2e^{-j\omega_0 t} y^+(t) = \underline{y}(t) \end{aligned}$$

8.3.2 Condizioni per $\underline{x}(t)$ reale

Data la rilevanza dei segnali con inviluppo complesso ad una sola componente, determiniamo quali condizioni si debbano verificare per dar luogo ad una simile circostanza, iniziando da un esempio.

Filtro passa banda ideale E' descritto da una risposta in frequenza $H(f)$ nulla ovunque, tranne che negli intervalli di frequenze $f_0 - W \leq |f| \leq f_0 + W$ dove ha valore unitario.



Pertanto risulta $H(f) = \text{rect}_{2W}(f - f_0) + \text{rect}_{2W}(f + f_0)$, da cui si ottiene facilmente

$$\begin{aligned} h(t) &= \mathcal{F}^{-1}\{H(f)\} = 2W \text{sinc}(2Wt) \left(e^{j2\pi f_0 t} + e^{-j2\pi f_0 t} \right) = \\ &= 4W \text{sinc}(2Wt) \cos 2\pi f_0 t \end{aligned}$$

D'altra parte, l'andamento di $H(f)$ è quello tipico dei segnali modulati, e quindi per $h(t)$ vale la sua rappresentazione in termini di inviluppo complesso $\underline{h}(t) = h_c(t) + jh_s(t)$, per

cui possiamo scrivere

$$h(t) = h_c(t) \cos 2\pi f_0 t - h_s(t) \sin 2\pi f_0 t$$

Confrontando questa espressione con quella trovata prima, si osserva che deve necessariamente risultare $h_s(t) = 0$ ed $h_c(t) = 4W \frac{\sin 2\pi Wt}{2\pi Wt}$, per cui $\underline{h}(t)$ è reale.

Simmetria coniugata attorno ad f_0 Il filtro passa banda ideale presenta $\underline{h}(t)$ reale, in quanto $\underline{H}(f) = 2H^+(f + f_0) = 2\text{rect}_{2W}(f)$ esibisce simmetria coniugata attorno all'origine. E' proprio questa la condizione cercata, che ci permette di enunciare

Un segnale modulato $x(t)$ possiede un inviluppo complesso $\underline{x}(t)$ reale, se lo spettro di quest'ultimo $\underline{X}(f)$ ha simmetria coniugata attorno all'origine $\underline{X}(f) = \underline{X}^*(-f)$, ovvero il segnale analitico $X^+(f)$ ha simmetria coniugata attorno ad f_0 : $X^+(f_0 + \phi) = [X^+(f_0 - \phi)]^*$ ($|\phi| < W$).

In altre parole, $\underline{x}(t) = x_c(t)$ se $X^+(f)$ ha modulo pari e fase dispari rispetto ad f_0 .

8.3.3 Estrazione delle componenti analogiche di bassa frequenza

Lo schema precedentemente adottato per ottenere le C.A. di B.F., basato sul filtro di Hilbert, non è l'unico. Un secondo metodo molto pratico prevede ancora l'uso di portanti *sincrone* (od *omodina*)²⁰, e di due filtri passa basso anzichè del filtro di Hilbert (vedi figura seguente). Il funzionamento è basato sul fatto che, considerando $x(t)$ espresso in termini delle sue C.A. di B.F., si ha:

$$\begin{aligned} x(t) \cos \omega_0 t &= [x_c(t) \cos \omega_0 t - x_s(t) \sin \omega_0 t] \cos \omega_0 t = \\ &= x_c(t) \cos^2 \omega_0 t - x_s(t) \sin \omega_0 t \cos \omega_0 t = \\ &= \frac{1}{2} x_c(t) + \frac{1}{2} x_c(t) \cos 2\omega_0 t - \frac{1}{2} x_s(t) \sin 2\omega_0 t \end{aligned}$$

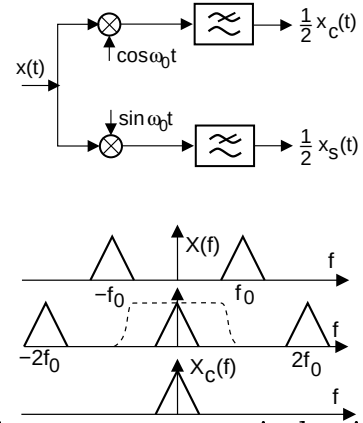
²⁰Con queste parole si indica l'uso in ricezione della stessa identica portante usata per la trasmissione, senza errori né di fase né di frequenza.

I termini in cui compaiono $\cos 2\omega_0 t$ e $\sin 2\omega_0 t$ rappresentano componenti di segnale centrate attorno a $2f_0$, ed il filtro passa basso²¹ provvede ad eliminarli: la banda del filtro deve quindi essere maggiore di W ma inferiore a $2f_0 - W$. Pertanto, il filtro non deve necessariamente essere ideale, e se $f_0 \gg W$ non pone particolari problemi realizzativi. Procedendo in maniera simile, per il ramo in quadratura si ottiene:

$$x(t) \sin \omega_0 t = \frac{1}{2} x_s(t) - \frac{1}{2} x_c(t) \cos 2\omega_0 t - \frac{1}{2} x_s(t) \sin 2\omega_0 t$$

e dunque, anche in questo caso, il filtro passa-basso rimuove le componenti a frequenza doppia.

Se i filtri non sono ideali, ma hanno ad esempio una fase lineare, saranno equivalenti ad un ritardo; se presentano distorsioni più severe (modulo non costante o fase non lineare), allora introducono distorsioni aggiuntive; per ridurre al minimo gli effetti di queste ultime, si tenta almeno di realizzare i due filtri quanto più identici tra loro.



8.4 Rappresentazione dei processi in banda traslata

Finora abbiamo trattato i casi di segnali di energia e di potenza; per ottenere una rappresentazione adeguata anche dei processi, occorre ancora un pò di teoria. Il lettore impaziente, o timoroso di perdersi tra i calcoli (che sono effettivamente intricati), può saltare direttamente alle conclusioni, che saranno le uniche che ci serviranno per il resto del testo. Altrimenti, armiamoci di pazienza e partiamo.

Siamo ora interessati ad ottenere, una volta noto uno spettro di densità di potenza $\mathcal{P}_x(f)$ limitato in banda attorno ad f_0 , delle rappresentazioni utili per gli spettri di densità di potenza delle componenti analogiche di bassa frequenza, ovvero espressioni per le loro funzioni di autocorrelazione. Infatti, come abbiamo visto, il passaggio di un segnale in banda traslata attraverso un filtro può essere scomposto in 4 filtraggi in banda base: pertanto la rappresentazione delle C.A. di B.F. è sufficiente per ottenere tutte le altre grandezze di interesse.

Osserviamo innanzitutto che, se un processo aleatorio presenta una $\mathcal{P}_x(f)$ limitata in banda attorno ad f_0 , allora la funzione di autocorrelazione $\mathcal{R}_x(\tau) = \mathcal{F}^{-1}\{\mathcal{P}_x(f)\}$ può essere espressa in termini delle componenti analogiche di bassa frequenza della funzione di autocorrelazione stessa:

$$\mathcal{R}_x(\tau) = \mathcal{R}_c(\tau) \cos \omega_0 \tau - \mathcal{R}_s(\tau) \sin \omega_0 \tau$$

Pertanto, non si ottengono direttamente le C.A. di B.F. del processo, come invece accade per segnali di energia di cui è noto $\underline{X}(f)$. D'altra parte, è innegabile che una realizzazione di $x(t)$ sia limitata in banda centrata a f_0 , e che quindi per essa debba esistere la rappresentazione $x(t) = x_c(t) \cos \omega_0 t - x_s(t) \sin \omega_0 t$; data la natura aleatoria di $x(t)$, gli stessi $x_c(t)$ ed $x_s(t)$ sono realizzazioni di processi. Questi ultimi in generale non sono indipendenti tra loro, in quanto la loro combinazione deve produrre un $x(t)$ che appartiene

²¹Il simbolo $\approx$ rappresenta un filtro passa-basso, poichè viene cancellata l'ondina superiore. Nello stesso stile, sono a volte indicati un passa-alto $\approx$ ed un passa-banda $\approx$.

al processo originario. Si pensi ad esempio al segnale $x(t) = x_c(t) \cos \omega_0 t$, in cui $x_c(t)$ è stazionario ed ergodico: come già osservato al § 7.6.1, $x(t)$ è solamente ciclostazionario.

Come prima cosa, proviamo a calcolare la funzione di autocorrelazione dell'involuppo complesso di una generica realizzazione:

$$\begin{aligned}\mathcal{R}_{\underline{x}}(\tau) &= E\{\underline{x}^*(\tau) \underline{x}(t+\tau)\} = E\{[x_c(\tau) - jx_s(\tau)][x_c(t+\tau) + jx_s(t+\tau)]\} = \\ &= E\{x_c(\tau)x_c(t+\tau) + x_s(\tau)x_s(t+\tau) + j[x_c(\tau)x_s(t+\tau) - x_s(\tau)x_c(t+\tau)]\} = \\ &= \mathcal{R}_{x_c}(\tau) + \mathcal{R}_{x_s}(\tau) + j[\mathcal{R}_{x_c x_s}(\tau) - \mathcal{R}_{x_s x_c}(\tau)]\end{aligned}$$

queste 4 quantità sono calcolate in appendice. Il risultato finale dei calcoli, nel caso in cui $x_c(t)$ e $x_s(t)$ siano stazionari ed ergodici, fornisce le espressioni:

$$\begin{cases} \mathcal{R}_{x_c}(\tau) &= \mathcal{R}_{x_s}(\tau) = \mathcal{R}_x(\tau) \cos \omega_0 \tau + \hat{\mathcal{R}}_x(\tau) \sin \omega_0 \tau \\ \mathcal{R}_{x_c x_s}(\tau) &= -\mathcal{R}_{x_s x_c}(\tau) = \hat{\mathcal{R}}_x(\tau) \cos \omega_0 \tau - \mathcal{R}_x(\tau) \sin \omega_0 \tau \end{cases}$$

in cui $\hat{\mathcal{R}}_x(\tau) = \mathcal{H}\{\mathcal{R}_x(\tau)\}$ è la trasformata di Hilbert di $\mathcal{R}_x(\tau)$. Osserviamo quindi come risulti $\mathcal{R}_{\underline{x}}(\tau) = 2[\mathcal{R}_{x_c}(\tau) + j\mathcal{R}_{x_c x_s}(\tau)]$, e pertanto $\mathcal{P}_{\underline{x}}(f) = 2[\mathcal{P}_{x_c}(f) + j\mathcal{P}_{x_c x_s}(f)]$.

- Da quest'ultima espressione, sembrerebbe che $\mathcal{P}_{\underline{x}}(f)$ possa assumere valori complessi, perdendo così il senso fisico di potenza: ma non è così. Osserviamo infatti che $\mathcal{R}_{x_c x_s}(\tau)$ è un segnale reale dispari²²: pertanto $\mathcal{P}_{x_c x_s}(f) = \mathcal{F}\{\mathcal{R}_{x_c x_s}(\tau)\}$ è completamente immaginario, e dunque $\mathcal{P}_{\underline{x}}(f)$ è reale.
- Se risultasse $\mathcal{R}_{x_c x_s}(\tau) = 0$ per ogni τ allora $\mathcal{P}_{x_c x_s}(f) = 0$ e $\mathcal{P}_{\underline{x}}(f) = 2\mathcal{P}_{x_c}(f)$ sarebbe reale pari; la presenza di $\mathcal{P}_{x_c x_s}(f)$ lo può invece rendere asimmetrico, permettendo di ottenere ancora $\mathcal{P}_{\underline{x}}(f) = 4\mathcal{P}_x^+(f + f_0)$ (non dimostrato),
- Corollario del punto precedente è che, se $\mathcal{P}_x(f)$ è simmetrico rispetto ad f_0 , allora $\mathcal{R}_{x_c x_s}(\tau) = 0$ e le due C.A. di B.F. sono incorrelate; se inoltre queste sono congiuntamente gaussiane, allora risultano anche statisticamente indipendenti.
- Dato che $\mathcal{R}_{x_c x_s}(\tau) = -\mathcal{R}_{x_s x_c}(\tau)$ sono dispari, deve risultare che $\mathcal{R}_{x_c x_s}(0) = -\mathcal{R}_{x_s x_c}(0) = 0$; se i processi sono a media nulla, allora la potenza è pari al valore dell'autocorrelazione per $\tau = 0$, e quindi

$$\mathcal{P}_{\underline{x}} = \sigma_{\underline{x}}^2 = \mathcal{R}_{\underline{x}}(0) = 2\mathcal{R}_{x_c}(0) = 2\mathcal{R}_{x_s}(0) = 2\sigma_{x_c}^2 = 2\sigma_{x_s}^2 = 2\mathcal{P}_{x_c} = 2\mathcal{P}_{x_s}$$

In definitiva, le componenti analogiche di bassa frequenza hanno entrambe potenza metà di quella dell'involuppo complesso.

- $x_c(t)$ e $x_s(t)$ hanno (ciascuna) potenza pari a quella di $x(t)$, ovvero $\mathcal{P}_{x_c} = \mathcal{P}_{x_s} = \mathcal{P}_x$; infatti, ricordando che $\mathcal{P}_{\underline{x}}(f) = 4\mathcal{P}_x^+(f + f_0)$, si ottiene $\mathcal{P}_{\underline{x}} = 4\mathcal{P}_x^+$. Dovendo chiaramente risultare $\mathcal{P}_x = \mathcal{P}_x^+ + \mathcal{P}_x^-$, si ottiene

$$\mathcal{P}_x = \frac{1}{4}[\mathcal{P}_{\underline{x}} + \mathcal{P}_{\underline{x}}] = \frac{1}{2}\mathcal{P}_{\underline{x}} = \mathcal{P}_{x_c} = \mathcal{P}_{x_s}$$

²²Infatti $\mathcal{R}_{x_c x_s}(\tau) = \hat{\mathcal{R}}_x(\tau) \cos \omega_0 \tau - \mathcal{R}_x(\tau) \sin \omega_0 \tau$, in cui $\mathcal{R}_x(\tau) = \mathcal{F}^{-1}\{\mathcal{P}_x(f)\}$ è pari e $\sin \omega_0 \tau$ è dispari, mentre $\hat{\mathcal{R}}_x(\tau)$ è dispari (non è stato dimostrato, ma vale per le trasformate di Hilbert di segnali pari) e $\cos \omega_0 \tau$ è pari. Inoltre, essendo $x_c(t)$ ed $x_s(t)$ reali, $\mathcal{R}_{x_c x_s}(\tau)$ è reale.

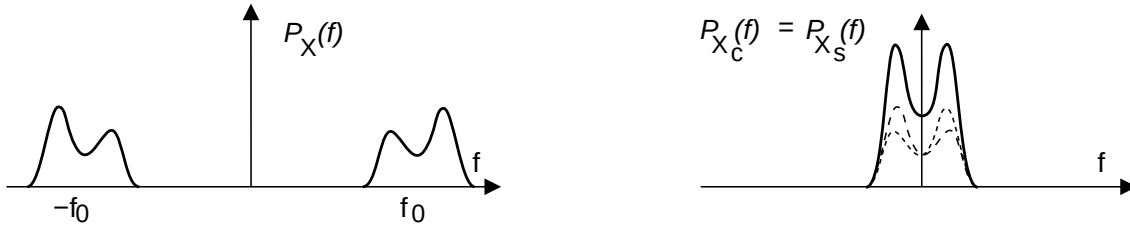


Figura 8.1: Segnale modulato e densità di potenza delle componenti analogiche di b.f.

- E' possibile mostrare che, volendo esprimere l'autocorrelazione di $x(t)$ in termini delle sue C.A. di B.F. $\mathcal{R}_x(\tau) = \mathcal{R}_c(\tau) \cos \omega_0 \tau - \mathcal{R}_s(\tau) \sin \omega_0 \tau$, risulta

$$\begin{cases} \mathcal{R}_c(\tau) = \mathcal{R}_{x_c}(\tau) \\ \mathcal{R}_s(\tau) = -\mathcal{R}_{x_s}(\tau) \end{cases}$$

da cui è possibile mostrare che $\mathcal{R}_x(\tau) = \mathcal{R}_{\hat{x}}(\tau)$.

- Volendo valutare $\mathcal{P}_{x_c}(f)$, questo risulta identico a $\mathcal{P}_{x_s}(f)$, in quanto (come già visto) $\mathcal{R}_{x_c}(\tau) = \mathcal{R}_{x_s}(\tau) = \mathcal{R}_x(\tau) \cos \omega_0 \tau + \hat{\mathcal{R}}_x(\tau) \sin \omega_0 \tau$; applicando ora la formula di Eulero per seno e coseno si ottiene

$$\begin{aligned} \mathcal{R}_{x_c}(\tau) = \mathcal{R}_{x_s}(\tau) &= \mathcal{R}_x(\tau) \frac{e^{j\omega_0 \tau} + e^{-j\omega_0 \tau}}{2} + \hat{\mathcal{R}}_x(\tau) \frac{e^{j\omega_0 \tau} - e^{-j\omega_0 \tau}}{2j} \\ &= \frac{1}{2} \left[\mathcal{R}_x(\tau) - j\hat{\mathcal{R}}_x(\tau) \right] e^{j\omega_0 \tau} + \frac{1}{2} \left[\mathcal{R}_x(\tau) + j\hat{\mathcal{R}}_x(\tau) \right] e^{-j\omega_0 \tau} \\ &= \mathcal{R}_x^-(\tau) e^{j\omega_0 \tau} + \mathcal{R}_x^+(\tau) e^{-j\omega_0 \tau} \end{aligned}$$

infatti i termini tra parentesi quadre corrispondono alla definizione di componenti a frequenze positive e negative ottenute tramite trasformata di Hilbert.

8.4.1 Conclusioni

Abbiamo mostrato che $\mathcal{R}_{x_c}(\tau) = \mathcal{R}_{x_s}(\tau) = \mathcal{R}_x^-(\tau) e^{j\omega_0 \tau} + \mathcal{R}_x^+(\tau) e^{-j\omega_0 \tau}$. Risulta quindi:

$$\mathcal{P}_{x_c}(f) = \mathcal{P}_{x_s}(f) = \mathcal{P}_x^-(f - f_0) + \mathcal{P}_x^+(f + f_0)$$

e dunque lo spettro di densità di potenza delle componenti analogiche di un processo si ottiene traslando e sovrapponendo (vedi fig. 8.1) le componenti a frequenze positive e negative del $\mathcal{P}_x(f)$ di partenza.

8.4.2 Processo gaussiano bianco limitato in banda

Se $x(t)$ è gaussiano stazionario ergodico e bianco, con $\mathcal{P}_x(f) = \frac{N_0}{2}$ limitato in banda $\pm W$ attorno ad f_0 e a media nulla, allora (vedi fig. 8.2):

- $x_c(t)$ e $x_s(t)$ sono congiuntamente gaussiane, stazionarie, ergodiche e indipendenti, con $\mathcal{P}_{x_c} = \mathcal{P}_{x_s} = \mathcal{P}_x = 2N_0W$, e pari alle varianze $\sigma_{x_c}^2$, $\sigma_{x_s}^2$ e σ_x^2 . Le rispettive densità di potenza valgono:

$$\mathcal{P}_{x_c}(f) = \mathcal{P}_{x_s}(f) = \mathcal{P}_x^+(f + f_0) + \mathcal{P}_x^-(f - f_0) = N_0 \text{rect}_{2W}(f)$$

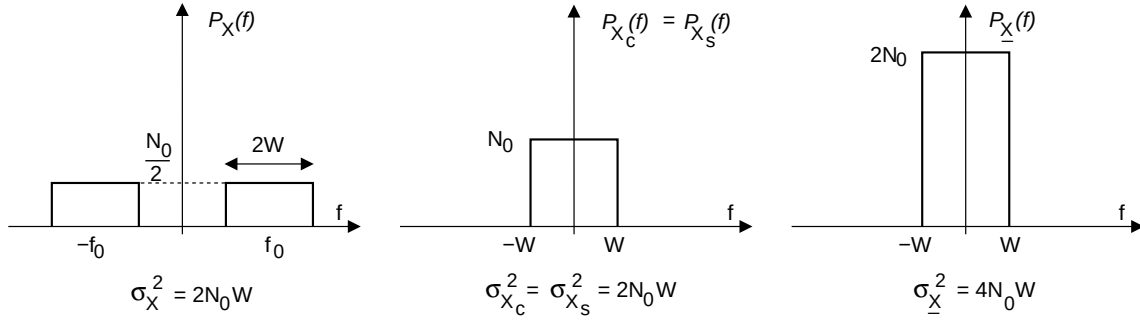


Figura 8.2: Densità di potenza per rumore passabanda

- L'inviluppo complesso ha potenza doppia:

$$\mathcal{P}_{\underline{x}} = 2\mathcal{P}_{x_c} = 4N_0W; \quad \mathcal{P}_{\underline{x}}(f) = 2N_0 \text{rect}_{2W}(f)$$

8.5 Appendice

8.5.1 Autocorrelazione di Processi Passa-Banda

Svolgiamo qui il calcolo relativo al valore di $\mathcal{R}_{x_c}(\tau)$, $\mathcal{R}_{x_s}(\tau)$, $\mathcal{R}_{x_c x_s}(\tau)$ e $\mathcal{R}_{x_s x_c}(\tau)$.

Ricordando che $x_c(t) = x(t) \cos \omega_0 t + \hat{x}(t) \sin \omega_0 t$, iniziamo a svolgere i calcoli per $\mathcal{R}_{x_c}(\tau)$:

$$\begin{aligned} \mathcal{R}_{x_c}(\tau) &= E\{x_c(\tau) x_c(t+\tau)\} = \\ &= E\{[x(t) \cos \omega_0 t + \hat{x}(t) \sin \omega_0 t][x(t+\tau) \cos \omega_0(t+\tau) + \hat{x}(t+\tau) \sin \omega_0(t+\tau)]\} = \\ &= E\{x(t)x(t+\tau)\} \cdot \cos \omega_0 t \cdot \cos \omega_0(t+\tau) + E\{x(t)\hat{x}(t+\tau)\} \cdot \cos \omega_0 t \cdot \sin \omega_0(t+\tau) + \\ &+ E\{\hat{x}(t)x(t+\tau)\} \cdot \sin \omega_0 t \cdot \cos \omega_0(t+\tau) + E\{\hat{x}(t)\hat{x}(t+\tau)\} \cdot \sin \omega_0 t \cdot \sin \omega_0(t+\tau) \end{aligned}$$

Valutiamo quindi i quattro valori attesi singolarmente, indicando con $\overline{x(t)}$ la media temporale di $x(t)$, ossia $\overline{x(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt$:

$$E\{x(t)x(t+\tau)\} = \overline{x(t)x(t+\tau)} = \mathcal{R}_x(\tau)$$

$$\begin{aligned} E\{x(t)\hat{x}(t+\tau)\} &= \overline{x(t)\hat{x}(t+\tau)} = \mathcal{R}_{x\hat{x}}(\tau) = x(-\tau) * \hat{x}(\tau) = x(-\tau) * x(\tau) * \frac{1}{\pi\tau} = \\ &= \mathcal{R}_x(\tau) * \frac{1}{\pi\tau} = \hat{\mathcal{R}}_x(\tau) \end{aligned}$$

$$\begin{aligned} E\{\hat{x}(t)x(t+\tau)\} &= \overline{\hat{x}(t)x(t+\tau)} = \mathcal{R}_{\hat{x}x}(\tau) = \hat{x}(-\tau) * x(\tau) = x(-\tau) * \left(-\frac{1}{\pi\tau}\right) * x(\tau) = \\ &= x(-\tau) * x(\tau) * \left(-\frac{1}{\pi\tau}\right) = \mathcal{R}_x(\tau) * \left(-\frac{1}{\pi\tau}\right) = -\hat{\mathcal{R}}_x(\tau) \end{aligned}$$

$$\begin{aligned} E\{\hat{x}(t)\hat{x}(t+\tau)\} &= \overline{\hat{x}(t)\hat{x}(t+\tau)} = \mathcal{R}_{\hat{x}\hat{x}}(\tau) = \hat{x}(-\tau) * \hat{x}(\tau) = x(-\tau) * \left(-\frac{1}{\pi\tau}\right) * x(\tau) * \frac{1}{\pi\tau} = \\ &= x(-\tau) * x(\tau) * \left(-\frac{1}{\pi\tau}\right) * \frac{1}{\pi\tau} = -\hat{\hat{\mathcal{R}}}_x(\tau) = \mathcal{R}_x(\tau) \end{aligned}$$

Sostituendo le relazioni ora trovate nella espressione di $\mathcal{R}_{x_c}(\tau)$, si ottiene

$$\begin{aligned}
\mathcal{R}_{x_c}(\tau) &= \mathcal{R}_x(\tau) \cdot \cos \omega_0 t \cdot \cos \omega_0(t + \tau) + \widehat{\mathcal{R}}_x(\tau) \cdot \cos \omega_0 t \cdot \sin \omega_0(t + \tau) + \\
&- \widehat{\mathcal{R}}_x(\tau) \cdot \sin \omega_0 t \cdot \cos \omega_0(t + \tau) + \mathcal{R}_x(\tau) \cdot \sin \omega_0 t \cdot \sin \omega_0(t + \tau) = \\
&= \frac{1}{2} \mathcal{R}_x(\tau) [\cos \omega_0(-\tau) + \cos \omega_0(2t + \tau)] + \frac{1}{2} \widehat{\mathcal{R}}_x(\tau) [\sin \omega_0(\tau) + \sin \omega_0(2t + \tau)] + \\
&- \frac{1}{2} \widehat{\mathcal{R}}_x(\tau) [\sin \omega_0(-\tau) + \sin \omega_0(2t + \tau)] + \frac{1}{2} \mathcal{R}_x(\tau) [\cos \omega_0(-\tau) - \cos \omega_0(2t + \tau)] = \\
&= \mathcal{R}_x(\tau) \cdot \cos \omega_0 \tau + \widehat{\mathcal{R}}_x(\tau) \cdot \sin \omega_0 \tau
\end{aligned}$$

che costituisce il risultato anticipato. Per l'espansione dei termini trigonometrici, si è fatto uso delle relazioni:

$$\begin{aligned}
\cos \alpha \cdot \cos \beta &= \frac{1}{2} [\cos(\alpha - \beta) + \cos(\alpha + \beta)] \\
\sin \alpha \cdot \sin \beta &= \frac{1}{2} [\cos(\alpha - \beta) - \cos(\alpha + \beta)] \\
\sin \alpha \cdot \cos \beta &= \frac{1}{2} [\sin(\alpha - \beta) + \sin(\alpha + \beta)]
\end{aligned}$$

I calcoli relativi al valore di $\mathcal{R}_{x_s}(\tau)$ sono del tutto simili, ed il loro svolgimento porta al risultato $\mathcal{R}_{x_c}(\tau) = \mathcal{R}_{x_s}(\tau)$.

Per quanto riguarda $\mathcal{R}_{x_c x_s}(\tau)$, applichiamo la relazione $x_s(t) = \widehat{x}(t) \cos \omega_0 t - x(t) \sin \omega_0 t$, per ottenere:

$$\begin{aligned}
\mathcal{R}_{x_c x_s}(\tau) &= E\{x_c(\tau) x_s(t + \tau)\} = \\
&= E\{[x(t) \cos \omega_0 t + \widehat{x}(t) \sin \omega_0 t] [\widehat{x}(t + \tau) \cos \omega_0(t + \tau) - x(t + \tau) \sin \omega_0(t + \tau)]\} = \\
&= E\{x(t) \widehat{x}(t + \tau)\} \cdot \cos \omega_0 t \cdot \cos \omega_0(t + \tau) - E\{x(t) x(t + \tau)\} \cdot \cos \omega_0 t \cdot \sin \omega_0(t + \tau) + \\
&+ E\{\widehat{x}(t) \widehat{x}(t + \tau)\} \cdot \sin \omega_0 t \cdot \cos \omega_0(t + \tau) - E\{\widehat{x}(t) x(t + \tau)\} \cdot \sin \omega_0 t \cdot \sin \omega_0(t + \tau)
\end{aligned}$$

I valori attesi che vediamo comparire sono stati già calcolati, e quindi possiamo scrivere direttamente lo sviluppo dei calcoli, in cui si applicano nuovamente le identità trigonometriche note:

$$\begin{aligned}
\mathcal{R}_{x_c x_s}(\tau) &= \widehat{\mathcal{R}}_x(\tau) \cdot \cos \omega_0 t \cdot \cos \omega_0(t + \tau) - \mathcal{R}_x(\tau) \cdot \cos \omega_0 t \cdot \sin \omega_0(t + \tau) + \\
&+ \mathcal{R}_x(\tau) \cdot \sin \omega_0 t \cdot \cos \omega_0(t + \tau) - \widehat{\mathcal{R}}_x(\tau) \cdot \sin \omega_0 t \cdot \sin \omega_0(t + \tau) = \\
&= \frac{1}{2} \widehat{\mathcal{R}}_x(\tau) [\cos \omega_0(-\tau) + \cos \omega_0(2t + \tau)] - \frac{1}{2} \mathcal{R}_x(\tau) [\sin \omega_0(\tau) + \sin \omega_0(2t + \tau)] + \\
&- \frac{1}{2} \mathcal{R}_x(\tau) [\sin \omega_0(-\tau) + \sin \omega_0(2t + \tau)] - \frac{1}{2} \widehat{\mathcal{R}}_x(\tau) [\cos \omega_0(-\tau) - \cos \omega_0(2t + \tau)] = \\
&= -\mathcal{R}_x(\tau) \cdot \sin \omega_0(2t + \tau) + \widehat{\mathcal{R}}_x(\tau) \cdot \cos \omega_0(2t + \tau)
\end{aligned}$$

Per quanto riguarda gli argomenti delle funzioni trigonometriche, il valore di t è lasciato non specificato. Pertanto, visto che il processo è stazionario per ipotesi, può sensatamente essere posto a zero, e dunque ottenere il risultato previsto.

I calcoli relativi al valore di $\mathcal{R}_{x_s x_c}(\tau)$ sono del tutto simili, ed il loro svolgimento porta al risultato $\mathcal{R}_{x_s x_c}(\tau) = -\mathcal{R}_{x_c x_s}(\tau)$.

Capitolo 9

Modulazione per Segnali Analogici

Trattiamo qui delle tecniche comunemente usate per imprimere su una portante l'informazione di un messaggio modulante di natura analogica¹, illustrando allo stesso tempo l'andamento in frequenza del segnale ottenuto. Sono quindi discussi i metodi più idonei a realizzare la funzione di demodulazione, e gli accorgimenti che influenzano il risultato finale.

9.1 Modulazione di Ampiezza - AM

Si è illustrato come un segnale modulato possa essere rappresentato nei termini delle sue componenti analogiche di bassa frequenza: $x(t) = x_c(t) \cos \omega_0 t - x_s(t) \sin \omega_0 t$. Nel caso in cui x_c e x_s siano segnali indipendenti, la trasmissione congiunta di entrambi sulla medesima portante costituisce un segnale *QAM* (QUADRATURE AMPLITUDE MODULATION). Nei casi più tipici invece, i segnali x_c ed x_s non sono qualsiasi, ed in base alla loro definizione sono distinte le seguenti classi di segnali modulati in ampiezza:

- *Banda Laterale Doppia*: la componente $x_s(t)$ è nulla, cosicchè $\mathcal{P}_x(f)$ è simmetrico rispetto ad f_0 . Si tratta del caso a noi già noto, ed è indicato dagli acronimi *BLD* o *DSB* (DOUBLE SIDE BAND).
- *Banda Laterale Unica*: sono presenti sia $x_c(t)$ che $x_s(t)$, e risulta $x_s(t) = \hat{x}_c(t)$. Questo fa sì che (come vedremo) la densità $\mathcal{P}_x(f)$ del segnale modulato giaccia tutta all'esterno (od all'interno) di $\pm f_0$ (*BLU* o *SSB* - SINGLE SIDE BAND).
- *Banda Laterale Ridotta*: è una via di mezzo tra i due casi precedenti, e cioè $\mathcal{P}_x(f)$ non è simmetrica rispetto ad f_0 , ma comunque giace da ambo i lati (*BLR* o *VSB* - VESTIGIAL SIDE BAND²).

Per completare la classificazione, per ognuna delle possibilità precedenti può verificarsi uno tra tre sottocasi, che si riferiscono alla presenza o meno, in $\mathcal{P}_x(f)$, di una concentrazione di potenza (ossia di un impulso) a frequenza f_0 , corrispondente alla trasmissione

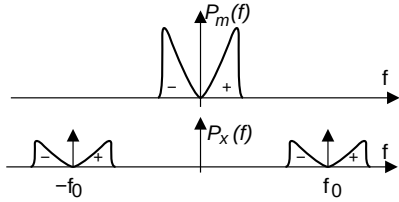
¹Per i segnali numerici si usano tecniche peculiari, esposte nel prossimo capitolo.

²Come sarà più chiaro nel seguito, l'acronimo *VRB* simboleggia che, anziché sopprimere completamente una delle due bande laterali, se ne mantengono delle vestigia.

di potenza non associata al messaggio $m(t)$, ma solamente alla portante, e quindi priva di contenuto informativo ai fini della trasmissione. I tre sottocasi citati sono indicati come:

- Portante Soppressa (*PS* o *SC* - SUPPRESSED CARRIER);
- Portante Intera (*PI* o *LC* - LARGE CARRIER);
- Portante Parzialmente Soppressa (*PPS*).

9.1.1 Banda Laterale Doppia - *BLD*



Questo è il caso a cui ci si riconduce in presenza di una sola componente analogica di bassa frequenza, che per convenzione è posta pari a $x_c(t)$ ³. La dipendenza di $x_c(t)$ da $m(t)$ è posta nella forma generale $x_c(t) = a_p + k_a m(t)$, e quindi

$$x_{BLD}(t) = (a_p + k_a m(t)) \cos \omega_0 t$$

L'involuppo complesso pertanto risulta $\underline{x}(t) = a_p + k_a m(t)$ da cui $\mathcal{P}_{\underline{x}}(f) = a_p^2 \delta(f) + k_a^2 \mathcal{P}_M(f)$ e quindi, dato che $\mathcal{P}_x(f) = \mathcal{P}_{x^+}(f) + \mathcal{P}_{x^-}(f)$ e che $\mathcal{P}_{x^+}(f) = \frac{1}{4} \mathcal{P}_{\underline{x}}(f - f_0)$ e $\mathcal{P}_{x^-}(f) = \frac{1}{4} \mathcal{P}_{\underline{x}}(f + f_0)$, si ottiene che il segnale modulato presenta una densità di potenza pari a

$$\mathcal{P}_x(f) = \frac{a_p^2}{4} [\delta(f - f_0) + \delta(f + f_0)] + \frac{k_a^2}{4} [\mathcal{P}_m(f - f_0) + \mathcal{P}_m(f + f_0)]$$

La potenza totale di $x(t)$ risulta pertanto $\mathcal{P}_x = \frac{a_p^2}{2} + \frac{k_a^2}{2} \mathcal{P}_m$, mentre il suo andamento in frequenza è quello riportato in figura, dove si è posto $k_a = 1$.

9.1.1.1 Portante soppressa - *PS*

Esaminando l'ultima espressione trovata per $\mathcal{P}_x$, è evidente come $\frac{a_p^2}{2}$ sia pari alla potenza della portante non modulata (concentrata per metà ad f_0 e per metà a $-f_0$), che quindi svanisce per $a_p = 0$, dando luogo in quest'ultima circostanza al sottocaso di *portante soppressa*.

La demodulazione di questo segnale si effettua in modo coerente (9.2.1), dopo aver ricostruito la portante per quadratura (9.4.2), oppure mediante demodulatore ad involuppo (9.2.2), dopo aver elaborato la portante ricostruita come spiegato in 9.1.1.3.

9.1.1.2 Portante Intera - *PI*

Nel caso in cui $a_p \neq 0$, si può scegliere che risulti sempre $x_c(t) \geq 0$, e quindi $a_p \geq k_a \cdot \max\{|m(t)|\}$, da cui risulta che deve essere

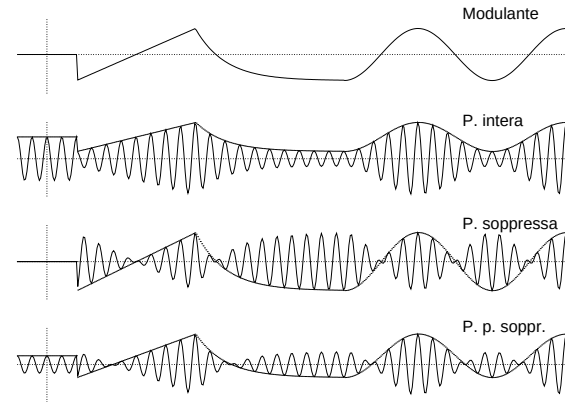
$$a_p^2 \geq k_a^2 m^2(t) \text{ per } \forall t$$

³Considerando che la portante di modulazione può avere una fase iniziale arbitraria, e che con una traslazione temporale ci si può sempre ricondurre ad usare una funzione $\cos \omega_0 t$, la convenzione posta tratta il caso di un segnale modulato $x(t) = a(t) \cos(\omega_0 t + \varphi)$ generico, con φ costante.

Queste ultime sono le condizioni che caratterizzano il caso di *portante intera*. Il rapporto $\left(\frac{a_p}{k_a}\right)^2$ rappresenta la massima *potenza istantanea*⁴ per $m(t)$, e consente di dimensionare i parametri a_p e k_a in modo da realizzare le condizioni richieste⁵.

Nel caso di portante intera, questa non inverte mai il segno, mentre nel caso di portante soppressa $x_c(t)$ ha media nulla (se $m(t)$ ha media nulla) e la portante presenta segni alterni con probabilità $\frac{1}{2}$, cosicchè per $f = f_0$ non compaiono impulsi in $\mathcal{P}_x(f)$.

La ragione principale dell'utilizzo della portante intera è che in tal caso il processo di decodifica non richiede la conoscenza di f_0 , e può svolgersi facendo uso di un semplice demodulatore di involuppo, descritto in 9.2.2.



9.1.1.3 Portante parzialmente soppressa - PPS

Se a_p è inferiore al valore necessario per avere la portante intera, si ottiene il caso della portante *parzialmente soppressa*, che ci permette di risparmiare potenza (vedi 9.1.1.4). Il residuo di portante presente, può essere usato per ri-generarla "al ricevitore" mediante un PLL (9.2.1.3), e sommarla al segnale ricevuto, ri-producendo così il termine $a_p \cos \omega_0 t$. In tal modo, ci si riconduce al caso *PI*, e si può effettuare la demodulazione di involuppo (9.2.2).

9.1.1.4 Efficienza di PI-PPS

Nell'espressione della potenza totale $\mathcal{P}_x = \frac{1}{2} (a_p^2 + k_a^2 \mathcal{P}_m)$ per un generico segnale AM, notiamo che solo $\mathcal{P}_u = \frac{k_a^2}{2} \mathcal{P}_m$ è relativa al segnale utile, mentre $\frac{a_p^2}{2}$ viene spesa sulla portante, che non trasporta informazione. Pertanto, si definisce una efficienza energetica

$$\eta = \frac{\mathcal{P}_u}{\mathcal{P}_x} = \frac{\frac{1}{2} k_a^2 \mathcal{P}_m}{\frac{1}{2} (a_p^2 + k_a^2 \mathcal{P}_m)} = \frac{1}{1 + \frac{a_p^2}{k_a^2 \mathcal{P}_m}}$$

che rappresenta la frazione di potenza trasmessa, utile ai fini della ricostruzione del messaggio⁶.

⁴Si definisce come potenza istantanea (o di picco) di $m(t)$, il segnale $\mathcal{P}_{M_I}(t) = m^2(t)$, per cui $\mathcal{P}_M = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \mathcal{P}_{M_I}(t) dt$.

⁵Ad esempio, nel caso in cui $m(t)$ sia un processo con densità di probabilità uniforme tra $\pm \frac{\Delta}{2}$, la potenza di picco risulta essere $\frac{\Delta^2}{4} = 3\sigma_M^2$ (ricordiamo che $\sigma_M^2 = \frac{\Delta^2}{12}$), mentre ad esempio se $m(t) = a \sin 2\pi f_M t$ si ha potenza di picco $a^2 = 2\sigma_M^2$ (dato che $\mathcal{P}_M = \sigma_M^2 = \frac{a^2}{2}$), oppure ancora se $m(t)$ è gaussiano, la potenza di picco (e dunque a_p^2/k_a^2 per ottenere la portante intera) risulta infinita (e cosa accade allora? Si avrà necessariamente una portante ridotta...).

⁶Ad esempio, se $m(t) = \sin 2\pi f_M t$ si ha $\mathcal{P}_M = 1/2$ e, nel caso di portante intera, deve risultare $a_p = k_a$ e dunque $\eta = \frac{1}{1+2} = 0.33$. Ovvero solo 1/3 della potenza trasmessa è utile al ricevitore!

9.1.2 Banda Laterale Unica - *BLU*

Come abbiamo visto, la modulazione *BLD* determina una occupazione di banda per $x(t)$ doppia di quella di $m(t)$. Per impegnare invece una banda pari a quella di $m(t)$, il segnale modulato deve dipendere da entrambe le componenti analogiche $x_c(t)$ ed $x_s(t)$, che devono risultare: $x_c(t) = m(t)$ e $x_s(t) = \hat{m}(t)$. Infatti in tal modo si ottiene:

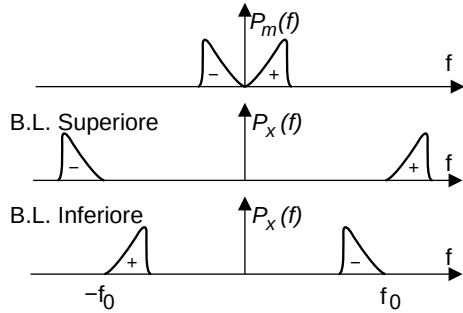
$$\begin{aligned} x_{BLU}(t) &= m(t) \cos \omega_0 t - \hat{m}(t) \sin \omega_0 t = m(t) \frac{e^{j\omega_0 t} + e^{-j\omega_0 t}}{2} - \hat{m}(t) \frac{e^{j\omega_0 t} - e^{-j\omega_0 t}}{2j} = \\ &= e^{j\omega_0 t} \frac{1}{2} [m(t) + j\hat{m}(t)] + e^{-j\omega_0 t} \frac{1}{2} [m(t) - j\hat{m}(t)] \end{aligned}$$

Ricordando ora che $\frac{1}{2} [m(t) \pm j\hat{m}(t)] = m^\pm(t)$ (vedi pag. 191) è proprio il contenuto a frequenze positive (negative), se $x(t)$ è di energia, effettuando la trasformata di Fourier di ambo i membri si ottiene

$$\begin{aligned} X_{BLU}(f) &= \delta(f - f_0) * M^+(f) + \delta(f + f_0) * M^-(f) = \\ &= M^+(f - f_0) + M^-(f + f_0) \end{aligned}$$

e quindi il segnale modulato *AM-BLU* è formato dai contenuti a frequenze positive e negative di $m(t)$, traslati ai lati della portante f_0 .

Qualora si consideri invece $m(t)$ un processo, si può dimostrare (passando dalla trasformata di $\mathcal{R}_x(\tau)$) un risultato del tutto analogo, ovvero



$$\mathcal{P}_x(f) = \mathcal{P}_{m+}(f - f_0) + \mathcal{P}_{m-}(f + f_0)$$

Nel caso descritto abbiamo considerato soppressa la portante, ed il segnale modulato (considerato nel dominio della frequenza) risulta “esterno” ad f_0 : questa circostanza è indicata con il termine di *banda laterale superiore*. Il caso opposto (*banda laterale inferiore*) si ottiene cambiando segno a $x_s(t)$. Scriviamo dunque

$$x_{BLU}(t) = \frac{k_a}{\sqrt{2}} m(t) \cos \omega_0 t \mp \frac{k_a}{\sqrt{2}} \hat{m}(t) \sin \omega_0 t$$

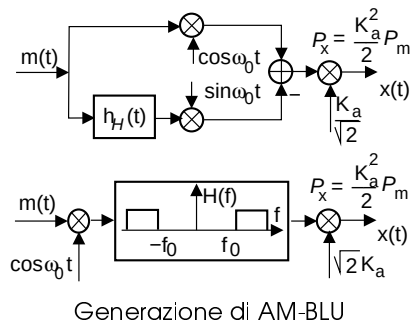
con $-$ e $+$ rispettivamente per ottenere un segnale *BLU* con banda superiore o inferiore. Con le costanti indicate, il segnale modulato *BLU* ha una potenza $\mathcal{P}_x = 2 \cdot (\frac{k_a^2}{2} \cdot \mathcal{P}_m \cdot \frac{1}{2}) = \frac{k_a^2}{2} \mathcal{P}_m$ (vedi 9.4.1), eguale a quella di un segnale *AM-BLD* in cui $x_c(t) = k_a m(t)$ e $x_s(t) = 0$.

I vantaggi di un tale metodo di modulazione sono subito evidenti: consente infatti di risparmiare banda, permettendo la trasmissione di più messaggi in divisione di frequenza (FDM).

9.1.2.1 Generazione di segnali *BLU*

Un segnale *BLU* può essere generato in due diversi modi. Il primo consiste nell’uso di un filtro di Hilbert per calcolare $\hat{m}(t)$, da usare assieme ad $m(t)$ in un modulatore in fase

ed in quadratura. E' subito evidente come si possano presentare problemi se $m(t)$ ha contenuti energetici prossimi a frequenza zero, che rendono assai stringenti le specifiche per approssimare il filtro di Hilbert.

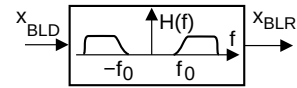


Un problema simile si presenta anche con il secondo metodo di generazione del segnale *BLU*, in cui si genera un segnale *BLD*, che viene filtrato in modo da eliminare una delle bande laterali; la necessità di trasmettere frequenze di $m(t)$ prossime allo zero, complica infatti la realizzazione dei filtri.

La trasmissione FDM di segnali *BLU* è stata lungamente usata per i ponti radio telefonici. Pertanto, la limitazione sulle minime frequenze telefoniche a non meno di 300 Hz sono motivate anche dalla necessità di effettuare modulazioni *BLU*.

9.1.3 Banda laterale ridotta - *BLR*

Si può verificare il caso in cui non si possa assolutamente fare a meno di componenti di segnale a frequenza molto bassa, come avviene, ad esempio, nel segnale televisivo⁷ (vedi appendice 9.4.4). Si ricorre allora alla modulazione a Banda Laterale Ridotta (*BLR*), che viene generata inviando il segnale modulato *BLU* attraverso uno specifico filtro, che presenta una transizione tra la banda passante e quella attenuata, più dolce di quella di un passa-banda ideale, e che si estende oltre f_0 .



9.1.4 Potenza di un segnale AM

Segue uno schema delle espressioni del segnale modulato per i diversi tipi di modulazione AM, ed i vincoli sui parametri legati all'ottenimento di una potenza totale P_x .

	Segnale modulato	Potenza	k_a per P_x dato
BLD-PS	$x(t) = k_a m(t) \cos(\omega_0 t)$	$P_x = \frac{k_a^2}{2} P_m$	$k_a = \sqrt{\frac{2P_x}{P_m}}$
BLU-PS	$x(t) = \frac{k_a}{\sqrt{2}} m(t) \cos(\omega_0 t) - \frac{k_a}{\sqrt{2}} \hat{m}(t) \sin(\omega_0 t)$	$P_x = \frac{k_a^2}{2} P_m$	$k_a = \sqrt{\frac{2P_x}{P_m}}$
BLD-PI	$x(t) = [a_p + k_a m(t)] \cos(\omega_0 t)$ con $a_p \geq k_a \cdot \max\{ m(t) \}$	$P_x = \frac{a_p^2}{2} + \frac{k_a^2}{2} P_m$	$k_a = \sqrt{\frac{2P_x - a_p^2}{P_m}}$

9.2 Demodulazione di ampiezza

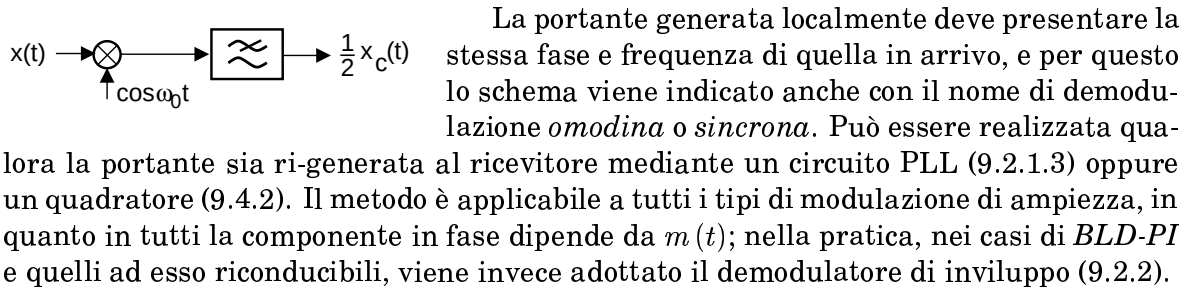
9.2.1 Demodulazione coerente (omodina)

Si tratta del circuito già noto (vedi 8.3.3) di estrazione della componente in fase $x_c(t)$ mediante moltiplicazione⁸ di $x(t)$ per una portante di demodulazione $\cos \omega_0 t$, e rimozione

⁷Nel caso ad esempio di ampie zone di immagine a luminosità costante, ed in lento movimento, il segnale è praticamente costante.

⁸In appendice 9.4.3 sono illustrate due tecniche di realizzazione del moltiplicatore.

delle componenti a frequenza $2f_0$ mediante un filtro passa-basso, e che viene riproposto nella figura che segue.



9.2.1.1 Errori di fase e frequenza

Esaminiamo ora cosa accade nel caso in cui la portante di demodulazione $\cos [2\pi (f_o + \Delta f) t + \theta]$ presenti un errore di fase θ e/o di frequenza Δf . Il risultato di una demodulazione coerente risulta⁹:

$$\begin{aligned} y(t) &= x_c(t) \cos \omega_0 t \cos [(\omega_o + \Delta \omega) t + \theta] \\ &= \frac{1}{2} x_c(t) [\cos (\Delta \omega t + \theta) + \cos ((2\omega_0 + \Delta \omega) t + \theta)] \end{aligned}$$

Pertanto, mentre il termine a frequenza (circa) doppia viene eliminato come al solito da un filtro apposito, il segnale utile è affetto da una distorsione pari a :

- per errori di fase: si ottiene $\frac{1}{2} x_c(t) \cos \theta$ che... può annullarsi per $\theta = \pm \frac{\pi}{2}$!
- per errori di frequenza: si ottiene $\frac{1}{2} x_c(t) \cos [\Delta \omega t]$ e dunque il segnale demodulato, oltre ad invertire periodicamente polarità, presenta una notevole oscillazione di ampiezza che, ad esempio, nel caso di segnale audio può rendere il risultato inintelligibile già con Δf pari a pochi Hertz.

9.2.1.2 Demodulazione in fase e quadratura

Se il demodulatore dispone anche del ramo in quadratura (quello con il seno) possiamo, in presenza di errore di fase, scegliere ad esempio quale dei due rami è meno attenuato, ed ovviare al problema. Lo stesso schema può essere utile in fase di ricerca della regione di frequenza in cui è presente un segnale¹⁰, oppure qualora si desideri solo verificare la presenza o meno di un segnale ad una determinata frequenza, come nel caso del radar¹¹.

⁹Si applichi $\cos \alpha \cos \beta = \frac{1}{2} [\cos (\alpha + \beta) + \cos (\alpha - \beta)]$.

¹⁰La ricerca dell'emittente, che può essere banalmente l'azione di sintonizzare la propria radio sul programma preferito, può richiedere interventi automatici, qualora si tratti ad esempio di dover compensare le variazioni di frequenza dovute al movimento reciproco di trasmettitore e ricevitore (*effetto doppler*), come per il caso delle comunicazioni con mezzi mobili.

In questi casi, prima della comunicazione vera e propria, è necessario prevedere una fase di *acquisizione della portante*, svolta ad esempio mediante un circuito del tipo di quello che stiamo per discutere, in cui vengono provate diverse portanti di demodulazione, finché non si produce un segnale in uscita.

¹¹Un radar trasmette ad elevata potenza per periodi molto brevi, e stima la presenza di oggetti basandosi sul ritardo con cui il segnale, riflesso da questi, torna indietro. Per questo, il ritardo di fase rappresenta proprio la grandezza che fornirà l'informazione relativa alla distanza, e può essere qualsiasi. Prima di iniziare a stimare tale informazione, è essenziale per il sistema accertarsi che *ci sia* un segnale da stimare.

Nel caso in cui il segnale ricevuto presenti una fase θ incognita

$$x(t) = m(t) \cos(\omega_0 t + \theta)$$

l'involuppo complesso risulta

$$\underline{x}(t) = m(t) e^{j\theta} = m(t) \cos \theta + jm(t) \sin \theta$$

e quindi $x_c(t) = m(t) \cos \theta$ e $x_s(t) = m(t) \sin \theta$.

Come noto, il demodulatore in fase e quadratura estrae proprio $x_c(t)$ ed $x_s(t)$, e dunque il segnale $y(t)$ risulta pari a:

$$y(t) = \sqrt{x_c^2(t) + x_s^2(t)} = |m(t)| \sqrt{\cos^2 \theta + \sin^2 \theta} = |m(t)|$$

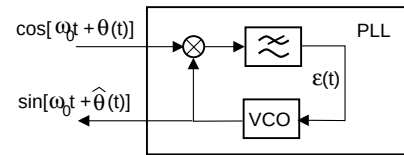
Pertanto, nonostante l'ignoranza della fase θ , siamo ancora in grado di individuare la *presenza* di un segnale modulante. L'operazione di modulo impedisce l'uso dello schema per demodulare generici segnali BLD-PS (mentre il caso PI sarebbe perfettamente demodulabile, ma in tal caso è più che sufficiente un demodulatore di involuppo (9.2.2)). Al § 5.6.7 sono esposti alcuni risultati relativi alla probabilità di detezione per questo demodulatore, nel caso in cui l'ingresso sia costituito da rumore, più una eventuale sinusoide.

9.2.1.3 Phase Locked Loop - PLL

Trattiamo qui del problema di generare una portante di demodulazione *coerente* (in fase) con quella della portante del segnale ricevuto. Una soluzione molto usata si basa su di un circuito controreazionato che prende il nome di *anello ad aggancio di fase*, e basa il suo funzionamento su di un dispositivo chiamato *oscillatore controllato in tensione* (VCO, VOLTAGE CONTROLLED OSCILLATOR).

Il VCO genera una sinusoide $y(t) = \sin\left(\omega_0 t + 2\pi k_f \int_{-\infty}^t x(\tau) d\tau\right)$, la cui fase varia nel tempo con l'integrale del segnale di ingresso¹².

Supponiamo ora di disporre di un segnale AM in cui sia presente un residuo di portante, come ad esempio nel caso di *portante parzialmente soppressa*: in tal caso la portante di demodulazione può essere ottenuta mediante il circuito in figura, che prende il nome di *PLL*. Indicando con $\hat{\theta}(t)$ la fase già integrata dal VCO fino all'istante t , all'uscita del moltiplicatore è presente un segnale¹³:



$$\frac{1}{2} \sin[2\omega_0 t + \theta(t) + \hat{\theta}(t)] + \frac{1}{2} \sin[\theta(t) - \hat{\theta}(t)]$$

¹²Se quest'ultimo ad esempio è costante ($x(t) = \Delta$), allora si avrà: $y(t) = \sin(2\pi f_0 t + 2\pi \Delta t) = \sin[2\pi(f_0 + \Delta)t]$, ovvero la frequenza si è alterata di una quantità pari a Δ . Il lettore più attento avrà riconosciuto che il VCO realizza il processo di *modulazione di frequenza*.

¹³Si utilizzi $\cos \alpha \sin \beta = \frac{1}{2} [\sin(\alpha + \beta) + \sin(\alpha - \beta)]$.

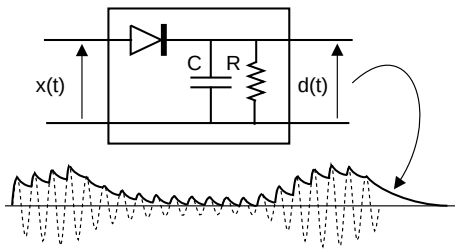
Il termine centrato a frequenza doppia ($2\omega_0$) viene eliminato dal filtro passa basso, e dunque rimane

$$\varepsilon(t) = \frac{1}{2} \sin [\theta(t) - \hat{\theta}(t)] = \frac{1}{2} \sin(\Delta\theta)$$

in cui $\Delta\theta$ rappresenta l'errore di fase, e $\varepsilon(t)$ è la grandezza in ingresso al VCO. Qualora risulti $\Delta\theta = 0$, si ha che anche $\varepsilon(t) = 0$, ed il VCO non altera la fase (esatta) della portante generata. Se $\Delta\theta > 0$ (o $\Delta\theta < 0$), allora $\varepsilon(t) > 0$ (o $\varepsilon(t) < 0$)¹⁴, e dunque il VCO è portato ad aumentare (diminuire) la fase della propria portante riducendo di conseguenza l'errore di fase¹⁵.

Nel caso in cui, invece, la fase entrante $\theta(t)$ vari nel tempo, allora il PLL insegue tali variazioni tanto più da vicino quanto più è elevato k_f (il coefficiente di proporzionalità tra $\hat{\theta}(t)$ e l'integrale di $\varepsilon(t)$)¹⁶.

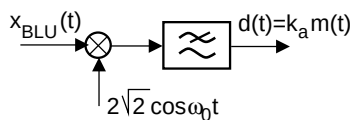
9.2.2 Demodulatore di inviluppo



Si tratta del semplice circuito non lineare riportato in figura¹⁷. Durante i periodi in cui il segnale in ingresso $x(t)$ è positivo rispetto alla tensione $d(t)$ accumulata dal condensatore, quest'ultimo si carica, inseguendo l'andamento dell'ingresso. Quando diviene $x(t) < d(t)$, il condensatore si scarica sulla resistenza con una costante di tempo $\tau = RC$, abbastanza grande rispetto ad $\frac{1}{f_0}$, da per-

mettere la ricostruzione dell'andamento di $x_c(t)$. Le oscillazioni a frequenza f_0 (e sue armoniche) infatti possono essere rimosse da un successivo filtro passa-basso, mentre la costante a_p è rimossa mediante un passa alto¹⁸. La semplicità del circuito è tale da farlo usare nel maggior numero di casi possibili.

9.2.2.1 Segnali Banda Laterale Unica e Ridotta



Nel caso di segnali *BLU*, il segnale modulante può essere riottenuto a partire da $x(t)$ utilizzando un demodulatore omodina, in quanto la componente in fase $x_c(t)$ dell'involuppo complesso è proprio pari al messaggio modulante $m(t)$.

¹⁴La grandezza di controllo $\varepsilon(t) \propto \sin(\Delta\theta)$ si mantiene direttamente proporzionale a $\Delta\theta$ nei limiti di $|\Delta\theta| < \frac{\pi}{2}$, denominato appunto *intervallo di aggancio*.

¹⁵Notiamo che un moltiplicatore, seguito da un filtro passabasso, esegue l'intercorrelazione tra gli ingressi del moltiplicatore, che nel nostro caso è una sinusoidale.

¹⁶In particolare, le realizzazioni pratiche del PLL dipendono fortemente dalla banda e dall'ordine del filtro di loop, in quanto è quest'ultimo che limita la velocità di variazione di $\varepsilon(t)$ e l'estensione dell'intervallo di aggancio. Lo studio teorico delle prestazioni si basa sull'uso della trasformata di Laplace e sulla linearizzazione di $\sin(\Delta\theta) \simeq \Delta\theta$, in quanto così il PLL può essere studiato come un sistema di controllo "linearizzato". Questa soluzione è brevemente illustrata in 9.3.1.1.

¹⁷Il simbolo ∇ rappresenta un *diode*, costituito da un bipolo di materiale semiconduttore drogato, che ha la particolarità di condurre in un solo verso (quello della freccia).

¹⁸Mentre la frequenza di taglio superiore del filtro complessivo può assumere un qualunque valore tra f_0 e la massima frequenza di $m(t)$, la frequenza inferiore dev'essere minore di quella minima del segnale. Perciò il metodo non è adatto, nel caso in cui $m(t)$ abbia componenti energetiche prossime a frequenza zero.

In questo caso, occorre prestare particolare attenzione ad eventuali errori di frequenza e di fase (Δf e θ) della portante di demodulazione perché, essendo presenti entrambe le componenti $x_c(t)$ ed $x_s(t)$, in uscita al demodulatore si ottiene (per il caso di banda laterale superiore):

$$d(t) = k_a m(t) \cos(\Delta\omega t + \theta) - k_a \hat{m}(t) \sin(\Delta\omega t + \theta)$$

Pertanto si nota come la modulazione *BLU* sia più sensibile di quella *BLD* agli errori della portante di demodulazione, in quanto ora un semplice errore di fase θ produce non solo un affievolimento, ma una vera intermodulazione tra $m(t)$ e $\hat{m}(t)$.

Per i segnali *BLU* è possibile anche usare un demodulatore di inviluppo: in tal caso è necessario che nel segnale trasmesso sia presente la portante, di ampiezza sufficiente perché $x(t)$ non si inverta neanche per i picchi di segnale più ampi. L'analisi di questa esigenza determina una efficienza inferiore a quella del caso *BLD*.

Anche nel caso *BLR*, è possibile ricorrere ad un demodulatore di tipo omodina, purché il filtro $H(f)$ usato in trasmissione per rimuovere parte di una banda laterale presenti alcune condizioni di simmetria attorno a f_0 ¹⁹.

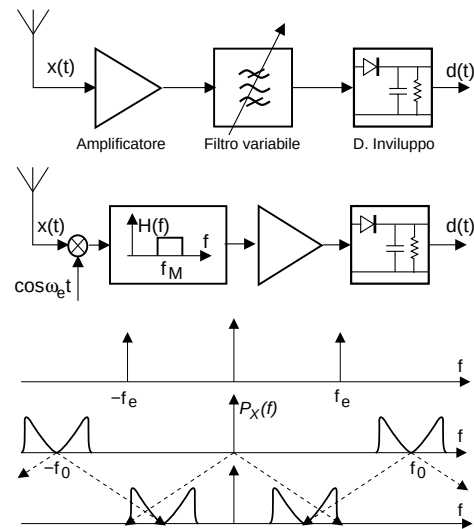
9.2.3 Demodulatore eterodina

Con questo nome si indica l'uso di una frequenza di demodulazione *differente* da quella della portante. Nella descrizione che segue illustriamo le motivazioni di tale scelta, assieme alle metodologie realizzative.

Per usare un demodulatore di inviluppo (purché le emittenti trasmettano a portante intera) occorrerebbe sintonizzare la stazione con un filtro passa banda variabile (rappresentato in figura da una freccia), la cui realizzazione a radio frequenza non è priva di difficoltà.

Attuare invece una demodulazione omodina per ricevere diverse emittenti (ad es. trasmissioni radio di tipo *broadcast*), richiede di variare la frequenza di demodulazione, ed estrarre il segnale modulante della emittente desiderata mediante il passa basso, che ora deve avere una banda pari alla distanza tra le portanti.

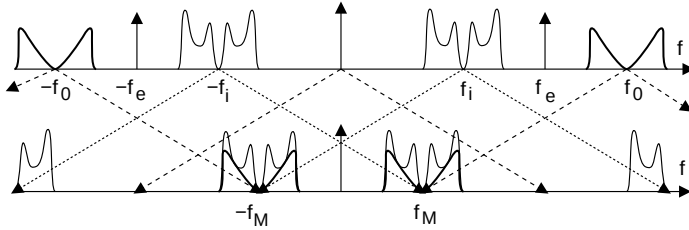
Si ricorre allora ad un diverso schema: il segnale ricevuto viene moltiplicato per una portante *eterodina* $f_e \neq f_0$, che ha l'effetto di centrare la frequenza zero dei segnali in ingresso alle frequenze $\pm f_e$. Allo stesso tempo, la componente a frequenze positive $\mathcal{P}_{x+}(f)$ centrata in f_0 si trasla in $f_0 \pm f_e$, così come la componente a frequenze negative $\mathcal{P}_{x-}(f)$ centrata in $-f_0$ va in $-f_0 \pm f_e$. Il risultato così ottenuto, attraversa quindi un filtro passa basso, centra-



¹⁹Si può dimostrare che per l'inviluppo complesso $\underline{H}(f)$ di $H(f)$ deve risultare: $\underline{H}(f) + \underline{H}^*(-f) = \text{cost}$ perché in tal modo il residuo di banda parzialmente soppressa si combina esattamente con ciò che manca alla banda laterale *non* soppressa.

to su di una *media frequenza* (MF) f_M ⁽²⁰⁾ particolare: all'uscita del filtro è presente allora solo l'emittente con portante $f_0 = f_M + f_e$, che potrà essere demodulata da un demodulatore fisso (ad esempio di involuppo). Per sintonizzare una diversa emittente, a frequenza f'_0 , è sufficiente porre $f'_e = f'_0 - f_M$, mentre il resto del ricevitore (con i suoi amplificatori e filtri) opera su di un segnale centrato sempre alla stessa media frequenza f_M , indipendentemente dall'emittente.

9.2.3.1 Frequenze immagine



L'uso di un ricevitore eterodina necessita della presenza di un filtro, che impedisca il verificarsi del problema seguente. Accade infatti che il filtro a MF riceverà nella propria banda, oltre alle emittenti centrate a $\pm f_0 = \pm (f_e + f_M)$, anche le emittenti al-

le portanti $\pm f_i = \pm (f_e - f_M)$, per le quali cioè $f_e - f_i = f_M$. La frequenza f_i prende il nome di *frequenza immagine*, in quanto è l'immagine speculare di f_0 rispetto ad f_e ; in altre parole, l'utilizzo di una f_e provoca la traslazione entro la MF sia della stazione desiderata, che della sua immagine a distanza $2f_M$. Pertanto, a monte del mixer va anteposto un filtro, che elimini dal segnale di ingresso le frequenze immagine, ovvero, una volta nota la gamma di frequenze che si vuol sintonizzare, tutte quelle a distanza $2f_M$ dalla banda di interesse.

Esempio: La maggior parte dei ricevitori di trasmissioni AM Broadcast (540-1600 KHz) utilizza un ricevitore supereterodina, con $f_M = 455 \text{ KHz}$, ed $f_e > f_0$. Volendo sintonizzare una stazione con $f_p = 600 \text{ KHz}$, occorre una $f_e = f_M + f_p = 1055 \text{ KHz}$. Ma allo stesso tempo, si porta a MF anche l'emittente a portante $f_i = f_e + f_M = 1510 \text{ KHz}$. Dunque, prima del mixer occorre un filtro che elimini le stazioni con $f_0 > f_e$ ⁽²¹⁾.

La scelta $f_M = 455 \text{ KHz}$, inferiore alla minima frequenza di 510 KHz, è motivata dalla necessità di utilizzare una regione di frequenze libera da altre trasmissioni, che altrimenti sarebbero state amplificate dallo stadio ad alto guadagno posto dopo il filtro MF.

²⁰Nel caso in cui $f_M < f_0$, ma maggiore della massima frequenza modulante, il ricevitore prende il nome di *supereterodina*.

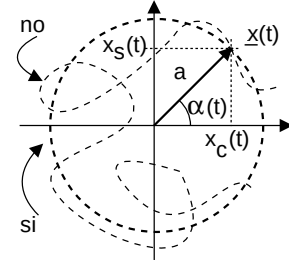
²¹Se l'esempio non sembra in accordo con la teoria precedente, è perché ora si è scelto $f_e > f_0$, mentre prima si poneva $f_e < f_0$, per motivi di leggibilità dei grafici. Nei ricevitori casalinghi invece si usa $f_e > f_0$ per motivi tecnologici, legati alla migliore regione di frequenze dell'oscillatore variabile di eterodina.

9.3 Modulazione angolare

In questo tipo di modulazione, l'informazione contenuta nel messaggio $x(t)$ è impressa sulla portante modificandone la fase: $x(t) = a \cos(\omega_0 t + \alpha(t))$.

Si ottiene allora un inviluppo complesso

$$\begin{aligned}\underline{x}(t) &= 2x^+(t) e^{-j\omega_0 t} = 2 \cdot a \frac{1}{2} e^{j\omega_0 t} e^{j\alpha(t)} \cdot e^{-j\omega_0 t} = a e^{j\alpha(t)} \\ &= a [\cos \alpha(t) + j \sin \alpha(t)] = x_c(t) + j x_s(t)\end{aligned}$$



Notiamo subito che, a differenza della AM, il modulo di $\underline{x}(t)$ è rigorosamente costante e pari ad a , mentre la fase $\alpha(t)$ varia continuamente.

Si è già mostrato come siano adottati 2 diversi tipi di legame tra messaggio $m(t)$ e fase dell'inviluppo complesso $\alpha(t)$, indicati con PM (modulazione di fase) ed FM (modulazione di frequenza) nella tabella a fianco. In questa compare la *frequenza istantanea*

	$\alpha(t)$	$f_i(t)$
PM	$k_\phi m(t)$	$f_0 + \frac{k_\phi}{2\pi} \frac{d}{dt} m(t)$
FM	$2\pi k_f \int_{-\infty}^t m(\tau) d\tau$	$f_0 + k_f m(t)$

$$f_i(t) = \frac{1}{2\pi} \frac{d}{dt} \psi(t) = f_0 + \frac{1}{2\pi} \frac{d}{dt} \alpha(t)$$

che è definita come la derivata della *fase istantanea* $\psi(t) = 2\pi f_0 t + \alpha(t)$. Le due alternative (PM e FM) sono analizzate assieme, in quanto reciprocamente intercambiabili qualora si effettui

- una PM con $m(t)$ pari all'integrale del messaggio informativo oppure
- una FM con $m(t)$ pari alla derivata del messaggio informativo.

Illustriamo subito alcune particolarità della modulazione angolare, prima di applicarci al problema della ricezione, ed alla determinazione della densità di potenza del segnale modulato.

Non linearità La caratteristica *fondamentale* della modulazione angolare è che il segnale modulato ha una dipendenza da $m(t)$ *fortemente non lineare*, e pertanto lo spettro di densità di potenza $\mathcal{P}_x(f)$ del segnale modulato non può essere calcolato con le tecniche tradizionali. Infatti, l'inviluppo complesso di un segnale modulato di angolo può essere espresso²² come:

$$\underline{x}(t) = a e^{j\alpha(t)} = a \left[1 + j\alpha(t) - \frac{\alpha^2(t)}{2} - j \frac{\alpha^3(t)}{3!} + \dots \right]$$

da cui risulta evidente che, anche se $\mathcal{P}_\alpha(f)$ è esprimibile a partire da $\mathcal{P}_M(f)$, nulla può essere detto in generale per $\mathcal{P}_x(f)$ (e dunque per $\mathcal{P}_x(f) = \frac{1}{4} \mathcal{P}_x(f - f_0) + \frac{1}{4} \mathcal{P}_x(f + f_0)$). La presenza delle potenze dell'angolo $\alpha(t)$ infatti, determina la non-applicabilità del principio di sovrapposizione degli effetti, ovvero, anche se sono noti i risultati della modulazione per due diversi messaggi $x_1(t) = FM\{m_1(t)\}$, $x_2(t) = FM\{m_2(t)\}$, il risultato ottenibile modulando la loro somma, non è quello della somma dei risultati individuali: $FM\{m_1(t) + m_2(t)\} \neq FM\{m_1(t)\} + FM\{m_2(t)\}$.

²²Si fa qui uso della espansione in serie di potenze dell'esponenziale: $e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots$

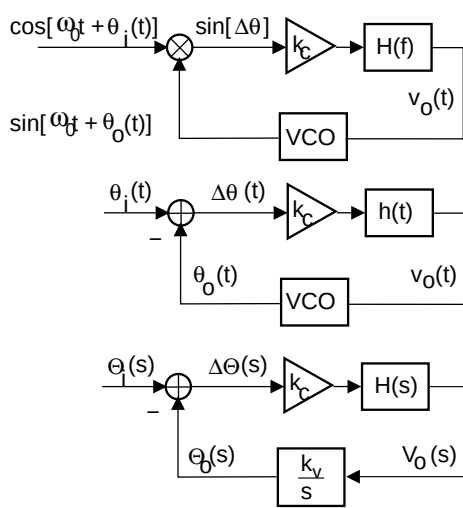
Ampiezza costante La circostanza che $\underline{x}(t) = ae^{j\alpha(t)}$ presenti un modulo costante pari ad a , indipendentemente dalle ampiezze del segnale modulante, è particolarmente utile, qualora per $m(t)$ siano da aspettarsi forti variazioni di dinamica. Questo è proprio il caso del segnale FDM, utilizzato per moltiplicare in frequenza più canali telefonici. In questo caso infatti, non essendo noto a priori il numero di canali effettivamente impegnati, la potenza del segnale $y(t) = \sum_{n=1}^N BLU\{m_n(t), f_n\}$ ottenuto sommando i diversi canali (ognuno a modulazione BLU con una diversa portante) può variare di molto: allora, il segnale complessivo $y(t)$ viene applicato all'ingresso di un modulatore FM e trasmesso come tale.

Generazione di un segnale a modulazione angolare Come anticipato, per effettuare una modulazione PM $x(t) = k_\phi m(t)$ si può usare un modulatore FM, in cui $\alpha(t) = 2\pi k_f \int_{-\infty}^t m'(\tau) d\tau$, ponendo $m'(t) = \frac{1}{2\pi} \frac{k_\phi}{k_f} \frac{d}{dt} m(t)$. Pertanto, consideriamo nel seguito solo le operazioni di modulazione/demodulazione FM. Un metodo *diretto* di generare un segnale FM è quello di utilizzare un VCO (già introdotto in (9.2.1.3)), ossia un oscillatore controllato in tensione, che produce il segnale $x(t) = a \sin(\omega_0 t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau)$, e dunque realizza proprio la funzione desiderata. Un secondo metodo verrà illustrato per un caso particolare in appendice 9.4.5. Infine, è sempre valido il modulatore in fase e quadratura, in cui si pone $x_c(t) = \cos \alpha(t)$ e $x_s(t) = \sin \alpha(t)$.

9.3.1 Ricezione di un segnale a modulazione angolare

In linea di principio, una volta ottenute $x_c(t)$ ed $x_s(t)$ del segnale modulato (ad esempio mediante un demodulatore in fase e quadratura) è sempre valida la relazione $\alpha(t) = \arctan \frac{x_s(t)}{x_c(t)}$. D'altra parte, tale soluzione si presta esclusivamente a realizzazioni digitali, in quanto è difficile realizzare un dispositivo che presenti esattamente la relazione non lineare di tipo arcotangente. Illustriamo allora i due metodi più comunemente usati:

9.3.1.1 Ricevitore a PLL



per $\Delta\theta$ piccolo.

In (9.2.1.3) si è già mostrato l'uso del circuito PLL per l'aggancio della fase della portante di modulazione. Lo stesso schema può essere usato per inseguire l'andamento temporale della fase di una portante modulata angularmente, ottenendo in tal modo l'informazione desiderata. La figura a lato riporta lo schema generale di un PLL, in cui il VCO genera un segnale pari a $\sin(\omega_0 t + \theta_o(t))$, con $\theta_o(t) = k_v \int_{-\infty}^t v_o(\tau) d\tau$, mentre il segnale ricevuto ha la forma $x(t) = \cos(\omega_0 t + \theta_i(t))$.

Lo schema può essere analizzato con i metodi dei controlli automatici, in quanto rappresenta un sistema che tenta di mantenere nullo l'errore $\sin \Delta\theta$, con $\Delta\theta(t) = \theta_i(t) - \theta_o(t)$; l'analisi si basa quindi sulla linearizzazione $\sin \Delta\theta \simeq \Delta\theta$, valida

L'analisi di Laplace fornisce allora il risultato

$$\Theta_o(s) = \frac{k_c k_v H(s)}{s + k_c k_v H(s)} \Theta_i(s)$$

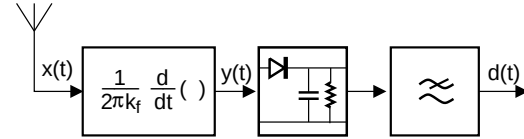
che, antitrasformato, permette di esprimere $\theta_o(t)$ (fase del VCO) come una versione filtrata della fase della portante modulata $\theta_i(t)$, da parte della funzione di trasferimento ad anello chiuso

$$\left. \frac{k_c k_v H(s)}{s + k_c k_v H(s)} \right|_{s=j2\pi f}$$

Inoltre, dato che il VCO produce $\theta_o(t) = k_v \int_{-\infty}^t v_o(\tau) d\tau$, si riconosce subito che l'uscita $v_o(t)$ del filtro di loop $H(s)$ corrisponde alla ricostruzione del messaggio modulante $m(t)$ nel caso di modulazione FM. Pertanto, l'uscita del filtro di loop del PLL realizza la demodulazione di frequenza.

9.3.1.2 Ricevitore a discriminatore

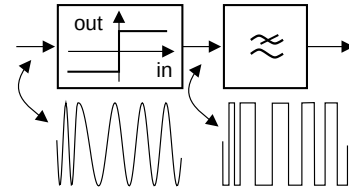
E' realizzato mediante il circuito di figura, in cui il derivatore effettua una conversione FM-AM, di cui viene demodulato l'involuppo di ampiezza. Infatti, la grandezza $y(t)$ risulta pari a



$$\begin{aligned} y(t) &= \frac{1}{2\pi k_f} \frac{d}{dt} a \cos \left(\omega_0 t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right) = \\ &= \frac{1}{2\pi k_f} (2\pi f_0 + 2\pi k_f m(t)) a \sin \left(\omega_0 t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right) \end{aligned}$$

e corrisponde dunque ad un segnale modulato sia angularmente che in ampiezza, di ampiezza $a(t) = a \left(\frac{f_0}{k_f} + m(t) \right)$. Siamo dunque in presenza di una modulazione di ampiezza *BLD-PI* e quindi, con una scelta opportuna²³ di $\frac{f_0}{k_f}$, il messaggio $m(t)$ può essere estratto mediante un demodulatore d'involuppo.

Il risultato ottenuto è valido purché $x(t)$ sia privo esso stesso di variazioni di ampiezza: per questo, spesso il derivatore è preceduto da un blocco *squadratore*, che produce una versione, appunto, "squadrata" del segnale ricevuto e quindi priva di modulazioni di ampiezza. Essendo lo squadratore fortemente non lineare, in uscita saranno presenti, oltre al segnale originario, anche componenti centrate a frequenze multiple di quella della portante, che vengono rimosse dal filtro passa basso a valle dello squadratore.



9.3.2 Densità spettrale di segnali modulati angularmente

Riprendiamo l'espressione dell'involuppo complesso di un segnale modulato angularmente

$$\underline{x}(t) = a e^{j\alpha(t)} = a \sum_{n=0}^{\infty} \frac{[j\alpha(t)]^n}{n!}$$

²³ Per utilizzare il demodulatore involuppo, deve risultare sempre $\frac{f_0}{k_f} + m(t) > 0$, e dunque occorre scegliere $\frac{f_0}{k_f} > \max_t \{|m(t)|\}$.

Osserviamo subito che la potenza totale di $\underline{x}(t)$ vale sempre $\mathcal{P}_{\underline{x}} = a^2$, indipendentemente da $\alpha(t)$, e dunque $\mathcal{P}_x = \frac{a^2}{2}$. Per ciò che riguarda $\mathcal{P}_{\underline{x}}(f)$, in linea di principio non si potrebbe neanche affermare che $\underline{x}(t)$ sia limitato in banda, vista la presenza delle potenze di qualunque ordine di $\alpha(t)$. D'altro canto, la presenza dei fattoriali fa sí che la serie di potenze possa essere troncata ad un certo ordine $\nu < \infty$. Se poniamo ora $\alpha(t) = k_\phi m(t)$, osserviamo che quanto piú $|k_\phi m(t)|$ è piccolo, tanto prima può essere troncata la serie con errori trascurabili. In particolare, se $\alpha(t)$ si mantiene sempre *molto piccolo*, la serie di potenze può essere troncata al primo termine ($n = 1$), dando luogo ad un comportamento praticamente lineare.

Se invece $\alpha(t)$ assume valori *molto elevati*, e quindi la serie di potenze comprende parecchi termini, subentra un secondo aspetto peculiare dell'FM, e cioè quello della conversione ampiezza $\rightarrow$ frequenza. Infatti, dato che $f_i(t) = f_0 + k_f m(t)$, la frequenza istantanea presenta scostamenti rispetto ad f_0 completamente dipendenti dalle ampiezze di $m(t)$, e quindi l'andamento della densità di potenza $\mathcal{P}_x(f)$ risulta strettamente dipendente da quello della densità di probabilità di $p_M(m)$ che descrive le ampiezze di $m(t)$.

Per valori intermedi della dinamica di $\alpha(t)$, invece, la $\mathcal{P}_x(f)$ risultante sarà una via di mezzo tra i due casi estremi discussi, che pertanto possono essere pensati come casi limite tra cui porre la densità di potenza effettiva.

Come anticipato, la natura non lineare della modulazione angolare rende necessario studiare ogni caso individualmente; pertanto la determinazione di $\mathcal{P}_x(f)$ viene svolta per due casi particolari, considerando per questi le due possibilità estreme di $\alpha(t)$ molto piccolo o molto grande, ed i risultati estrapolati per approssimare altre situazioni; i due casi esaminati sono:

- $m(t)$ sinusoidale e
- $m(t)$ membro di un processo stazionario ergodico.

9.3.2.1 Segnale modulante sinusoidale

Ponendo $m(t) = \cos(2\pi wt)$, si ottiene che la fase modulante $\alpha(t)$ e la frequenza istantanea $f_i(t)$ per i due casi PM ed FM, relativi al segnale $x(t) = a \cos(2\pi f_0 t + \alpha(t))$, risultano:

	$\alpha(t)$	$f_i(t)$	$\Delta\alpha$	Δf
PM	$k_\phi \cos(2\pi wt)$	$f_0 + w k_\phi \sin(2\pi wt)$	k_ϕ	$w k_\phi$
FM	$2\pi k_f \int_{-\infty}^t m(\tau) d\tau = \frac{k_f}{w} \sin(2\pi wt)$	$f_0 + k_f \cos(2\pi wt)$	$\frac{k_f}{w}$	k_f

in cui si è anche indicata la massima deviazione di fase $\Delta\alpha = \max\{|\alpha(t)|\}$ e di frequenza $\Delta f = \max\{|f_i(t) - f_0|\}$. Notiamo subito che, in entrambi i casi, sia la fase $\alpha(t)$ che la frequenza istantanea $f_i(t)$ variano sinusoidalmente con periodo $\frac{1}{w}$; nel caso PM l'entità di Δf aumenta con w , mentre nell'FM la $\Delta\alpha$ diminuisce con w . Nel seguito si farà riferimento all'*indice di modulazione angolare* β , corrispondente alla massima escursione della fase $\Delta\alpha$, che risulta:

$$\beta = \begin{cases} k_\phi & \text{(PM)} \\ \frac{k_f}{w} & \text{(FM)} \end{cases}$$

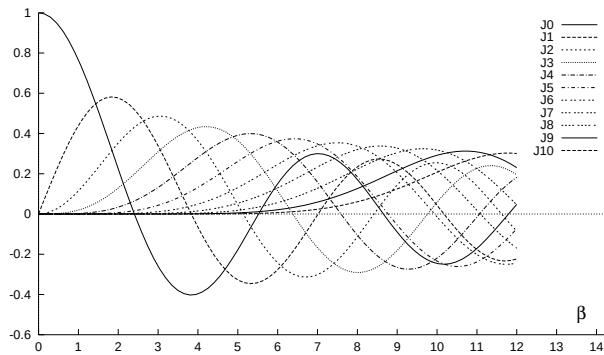
Con questa convenzione, possiamo trattare congiuntamente entrambi i casi PM ed FM riscrivendo l'involuppo complesso come ²⁴

$$\underline{x}(t) = a e^{j\beta \sin(2\pi w t)}$$

Notiamo ora che $\underline{x}(t)$ è periodico di periodo $\frac{1}{w}$, e dunque per esso vale lo sviluppo in serie di Fourier $\underline{x}(t) = a \sum_{n=-\infty}^{\infty} X_n e^{j2\pi n w t}$, i cui coefficienti risultano

$$X_n = w \int_{-\frac{1}{2w}}^{\frac{1}{2w}} e^{j\beta \sin(2\pi w t)} e^{-j2\pi n w t} dt = \mathcal{J}_n(\beta)$$

ovvero sono pari²⁵ alle *funzioni di Bessel* del primo tipo, ordine n ed argomento β . Queste hanno l'andamento mostrato in figura, assieme alle proprietà che le caratterizzano:



• $\mathcal{J}_n(\beta)$ è reale con $\forall n, \beta$

$$\bullet \begin{cases} \mathcal{J}_n(\beta) = \mathcal{J}_{-n}(\beta) & n \text{ pari} \\ \mathcal{J}_n(\beta) = -\mathcal{J}_{-n}(\beta) & n \text{ dispari} \end{cases}$$

$$\bullet \sum_{n=-\infty}^{+\infty} \mathcal{J}_n^2(\beta) = 1$$

$$\bullet \mathcal{J}_n(\beta) \simeq 0 \text{ con } n > \beta \text{ se } \beta \gg 1$$

e quindi i valori di X_n si ottengono tracciando una linea verticale nel diagramma di figura in corrispondenza del valore β adottato, e individuando il valore di ciascuna $\mathcal{J}_n$ per quel β . Osserviamo ora che l'ultima proprietà mostra come, in presenza di un valore di β elevato, le funzioni di Bessel di ordine $n > \beta$ siano praticamente nulle: è quindi lecito in tal caso limitare lo sviluppo in serie di Fourier di $\underline{x}(t)$ ai primi β termini (positivi e negativi), ovvero: $\underline{x}_{FM}(t) \simeq a \sum_{n=-\beta}^{\beta} \mathcal{J}_n(\beta) e^{j2\pi n w t}$. Pertanto, il segnale modulato $x(t) = \Re \{ \underline{x}(t) e^{j\omega_0 t} \}$ risulta:

$$x(t) \simeq a \sum_{n=-\beta}^{\beta} \mathcal{J}_n(\beta) \cos 2\pi(f_0 + n w)t$$

e quindi lo spettro di densità di potenza di $x(t)$ ha espressione

$$\mathcal{P}_x(f) \simeq \frac{a^2}{4} \sum_{n=-\beta}^{\beta} |\mathcal{J}_n(\beta)|^2 [\delta(f - f_0 - n w) + \delta(f + f_0 + n w)]$$

ed è formato da impulsi centrati a frequenze $f = \pm f_0 \pm n w$ (²⁶).

²⁴Si è sostituito *cos* con *sin* nel caso PM per omogeneità di formulazione, senza alterare la sostanza delle cose.

²⁵Le *funzioni di Bessel* del primo tipo, ordine n ed argomento β sono definite come $\mathcal{J}_n(\beta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(\beta \sin x - n x)} dx$.

²⁶Integrando l'espressione di $\mathcal{P}_x(f)$, e ricordando che $\sum_{n=-\infty}^{+\infty} \mathcal{J}_n^2(\beta) = 1$, si ottiene ancora un risultato già noto, e cioè che la potenza totale del segnale modulato risulta pari a quella della portante non modulata, e pari a $\mathcal{P}_x = \frac{a^2}{2}$, indipendentemente da β .

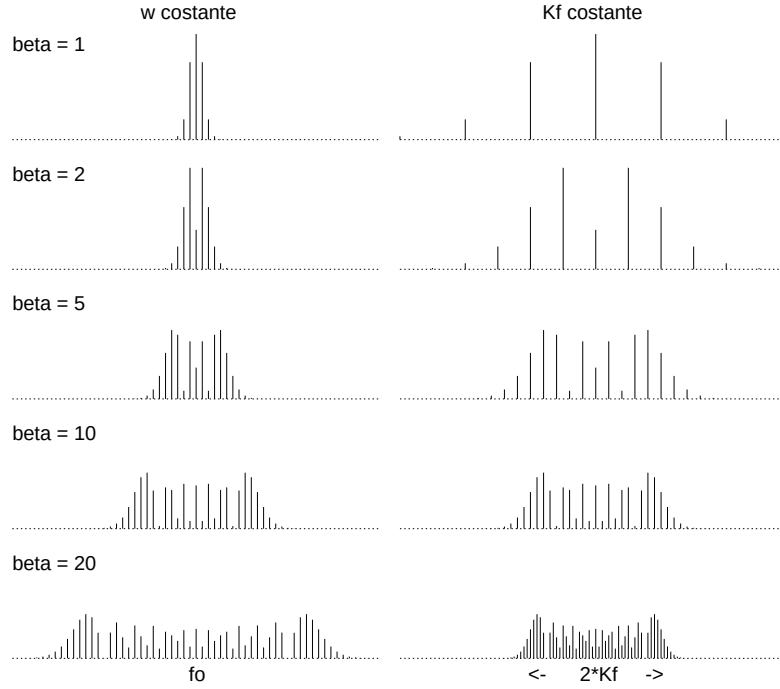


Figura 9.1: Spettri di ampiezza per segnale FM a modulazione sinusoidale

La fig. 9.1 mette a confronto $|X(f)|$ per f vicino ad f_0 , ovvero mostra $|X^+(f)| = \frac{a}{2} \sum_{n=-\beta}^{\beta} |\mathcal{J}_n(\beta)| \delta(f - f_0 - nw)$, calcolato per diversi valori di β , mantenendo fisso w oppure k_f (a sinistra e destra rispettivamente), e ci aiuta a comprendere i ragionamenti che seguono.

Modulazione a basso indice In questo caso $\beta \ll 1$, e tale da rendere trascurabili le funzioni di Bessel $\mathcal{J}_n(\beta)$ con $n > 1$. Allora, $x(t)$ occupa una banda pari a $2w$, in modo del tutto simile all'AM-BLD.

Modulazione ad alto indice In tal caso sono presenti più funzioni di Bessel, ed il comportamento non lineare tende a legare $\mathcal{P}_x(f)$ ai valori assunti da $f_i(t) = k_f m(t)$, realizzando una conversione ampiezza $\rightarrow$ frequenza.

In Fig. 9.1 sono evidenziati gli effetti dell'aumento di $\beta = \frac{k_f}{w}$ nelle due circostanze:

1. Si mantiene w fisso, aumentando k_f . Il numero di righe spettrali a frequenza $f_0 \pm nw$ aumenta, occupando una banda crescente, e per $\beta \rightarrow \infty$ si verifica che $\mathcal{J}_n(\beta) = 0$ per $n > \beta$. Pertanto, la banda occupata tende a $B = 2\beta w = 2\frac{k_f}{w} \cdot w = 2k_f$;
2. Si mantiene k_f fisso, diminuendo w . La banda occupata tende a ridursi, mentre le nuove righe spettrali a frequenza $f_0 \pm nw$ si infittiscono. Per $\beta \rightarrow \infty$, la spaziatura w tra le righe spettrali tende ad annullarsi, producendo una $\mathcal{P}_x(f)$ praticamente continua, e con una banda $B = 2k_f$, ossia pari alla massima deviazione di frequenza istantanea Δf .

Regola di Carson Come mostrato, nei due casi a basso ed alto indice, la banda occupata da $x(t)$ varia tra $2w$ e $2k_f$ rispettivamente. Nei casi intermedi, è pratica comune ricorrere all'espressione

$$B_C \simeq 2(k_f + w) = 2w(\beta + 1)$$

nota come *Regola di Carson*²⁷, che tiene conto di entrambi i fattori che concorrono alla determinazione della banda, e che fornisce i valori esatti sia per $\beta \ll 1$, che per $\beta \rightarrow \infty$, in entrambi i casi $k_f \rightarrow \infty$ o $w \rightarrow 0$.

Sebbene la determinazione approssimata della banda mediante la *Regola di Carson* sia stata ottenuta nel caso di $m(t) = \cos(2\pi wt)$, la stessa espressione è spesso adottata come una buona approssimazione anche per segnali non sinusoidali, ma limitati in banda tra $-W$ e W , e contraddistinti da una $\Delta f = k_f \cdot \max\{|m(t)|\}$. In tal caso, la regola di Carson si applica ponendo ora $B_C \simeq 2W(\beta + 1)$ con $\beta = \frac{\Delta f}{W}$. Per un approfondimento della questione, si veda l'appendice 9.4.6 e la sottosezione 9.3.3.1.

A prima vista, l'estensione del risultato per $m(t) = \cos(2\pi wt)$ al caso qualunque appare più che ragionevole; il comportamento non lineare della modulazione angolare impedisce però una sua verifica analitica. D'altra parte, i risultati sperimentali mostrano che l'approssimazione fornita dalla banda di Carson può effettivamente costituire una stima plausibile della banda occupata per segnali modulanti qualsiasi.

In base alla regola di Carson, notiamo ora che la banda occupata dal segnale modulato può risultare $\beta + 1$ volte più estesa di quella ottenibile mediante modulazione AM. Nonostante questo aumento di banda possa apparire un fatto negativo, vedremo nel capitolo 10 che ciò consente un migliore SNR dopo la demodulazione rispetto al caso AM. Al contrario, se $\beta \ll 1$, il comportamento si avvicina molto a quello lineare (vedi appendice 9.4.5).

9.3.3 Densità spettrale FM con processo aleatorio modulante

Riprendiamo il ragionamento iniziato in 9.3.2, relativo all'influenza di $p_M(m)$ su $\mathcal{P}_x(f)$. Considerando che la frequenza istantanea ha espressione $f_i = f_0 + k_f m(t)$, la frazione di potenza tra f_1 ed f_2 sarà pari alla frazione di tempo che il segnale $m(t)$ si trova tra $m_1 = \frac{f_1 - f_0}{k_f} \leq m(t) \leq m_2 = \frac{f_2 - f_0}{k_f}$. Nel caso in cui $m(t)$ sia sinusoidale, con fase iniziale aleatoria a distribuzione uniforme, $m(t)$ è una realizzazione di un processo armonico, e la frazione di tempo su indicata equivale alla $\text{Prob}\{m_1 \leq m(t) \leq m_2\}$. Pertanto le righe spettrali, addensandosi, tendono a disporsi in accordo all'andamento della densità $p_M(m)$ ²⁸.

Il risultato a cui siamo pervenuti nel caso di modulante sinusoidale è generale, e pertanto si può affermare che qualora si generi un segnale FM ad alto indice, a partire da un processo con densità di probabilità nota, lo spettro di densità di potenza del segnale modulato acquisisce l'andamento proprio della densità di probabilità del processo modulante, indipendentemente dal suo spettro di densità di potenza.

La conclusione riportata si mantiene valida purchè $\beta \gg 1$; nel caso contrario, sono validi i ragionamenti sviluppati alla sezione 9.3.3.2.

²⁷J.R. Carson fu uno dei primi ad investigare l'FM negli anni '20.

²⁸In particolare, per $\beta \rightarrow \infty$ risulterà $\mathcal{P}_x(f) = \frac{a^2}{1 - (f/k_f)^2}$, che è l'andamento a cui tendono (per $\beta \rightarrow \infty$) i grafici in basso di fig. 9.1.

Esempio: un processo uniforme $m(t)$ limitato in banda $\pm W$, con densità di probabilità $p_M(m) = \frac{1}{\Delta_M} \text{rect}_{\Delta_M}(m)$, modula ad alto indice la frequenza di una portante, con frequenza f_0 ed ampiezza a , con un coefficiente di modulazione k_f . Determinare $\mathcal{P}_x(f)$ del segnale modulato.

Notiamo subito che la frequenza istantanea f_i rimane limitata tra $f_0 - \frac{\Delta_M}{2} k_f$ e $f_0 + \frac{\Delta_M}{2} k_f$. Inoltre, la potenza totale deve risultare ancora pari a $\frac{a^2}{2}$. Pertanto si ottiene²⁹:

$$\mathcal{P}_x(f) = \frac{a^2}{4\Delta_M k_f} [\text{rect}_{\Delta_M k_f}(f - f_0) + \text{rect}_{\Delta_M k_f}(f + f_0)]$$

9.3.3.1 Indice di modulazione per processi

Ai fini dell'applicazione della regola di Carson, si è posto l'indice di modulazione $\beta = \frac{\Delta f}{W}$, con $\Delta f = k_f \max\{|m(t)|\}$. Nel caso di processi, può accadere che $m(t)$ non sia limitata in ampiezza, come ad esempio nel caso gaussiano, rendendo problematica la quantificazione di β . Per risolvere la questione, l'indice di modulazione β è ridefinito ancora una volta, e nel caso in cui $m(t)$ sia un generico processo si pone

$$\beta' = \begin{cases} \sigma_\alpha & \text{(PM)} \\ \frac{\sigma_f}{W} & \text{(FM)} \end{cases}$$

in cui W è la banda a frequenze positive del segnale modulante, $\sigma_f = k_f \sqrt{\mathcal{P}_m}$ rappresenta la deviazione standard della frequenza istantanea³⁰, e $\sigma_\alpha = k_\phi \sqrt{\mathcal{P}_m}$ è la deviazione standard della fase modulante³¹. L'applicazione della regola di Carson con il nuovo valore di β' , fornisce per la banda un risultato che non indica più la banda *totale* occupata, ma individua una *banda efficace* entro cui $\mathcal{P}_x(f)$ è in larga parte (ma non completamente) contenuta (vedi anche 9.4.6).

Nel caso in cui *non* risulti $\beta \gg 1$, lo spettro di potenza del segnale modulato FM torna a dipendere da quello del segnale modulante, e si ricade nella trattazione che segue.

9.3.3.2 Modulazione a basso indice

Ora l'indice di modulazione β si assume piccolo a sufficienza, da far sí che lo sviluppo in serie dell'involuppo complesso del segnale modulato possa essere arrestato ai primi termini.

Sotto opportune ipotesi, si può mostrare che vale il risultato

$$\mathcal{P}_x(f) \simeq a^2 e^{-\sigma_\alpha^2} \left[\delta(f) + \mathcal{P}_\alpha(f) + \frac{1}{2} \mathcal{P}_\alpha(f) * \mathcal{P}_\alpha(f) + \frac{1}{3!} \mathcal{P}_\alpha(f) * \mathcal{P}_\alpha(f) * \mathcal{P}_\alpha(f) + \dots \right]$$

avendo indicando con σ_α^2 la varianza della fase modulata e con $\mathcal{P}_\alpha(f)$ il relativo spettro di densità di potenza, pari rispettivamente a

²⁹Volendo applicare la regola di Carson per calcolare la banda, si avrebbe (considerando $\beta \gg 1$) $B_C = 2W(\beta + 1) \simeq 2\frac{\Delta f}{W}W = 2\Delta f$, in cui $\Delta f = k_f \frac{\Delta_M}{2}$. Pertanto risulta $B_C = 2k_f \frac{\Delta_M}{2} = k_f \Delta_M$, in accordo al risultato previsto nel caso di modulazione ad alto indice.

Qualora si fosse invece posto $\beta = \frac{\sigma_f}{W}$ (vedi 9.3.3.1) si sarebbe ottenuto $B_C = 2W(\beta + 1) \simeq 2\frac{\sigma_f}{W}W = 2\sigma_f = 2k_f \sqrt{\mathcal{P}_M} = 2k_f \sqrt{\frac{\Delta_M^2}{12}} = 2k_f \frac{\Delta_M}{2\sqrt{3}} = \frac{\Delta_M k_f}{\sqrt{3}}$, un risultato che è circa pari a 0.58 volte quello precedente. Data la particolarità di $p_M(m)$ uniforme, in questo caso è da preferire il primo risultato.

³⁰Infatti, dalla definizione $f_i(t) = f_0 + k_f m(t)$ si ottiene che $\sigma_f^2 = k_f^2 \sigma_M^2$, in cui $\sigma_M^2 = \mathcal{P}_M$ se $m(t)$ è un processo stazionario ergodico a media nulla.

³¹Come sopra, partendo dalla relazione $\alpha(t) = k_\phi m(t)$.

	$\mathcal{P}_\alpha(f)$	σ_α^2
PM	$k_\phi^2 \mathcal{P}_m(f)$	$k_\phi^2 \mathcal{P}_m$
FM	$k_f^2 \frac{\mathcal{P}_m(f)}{f^2}$	$k_f^2 \int_{-w}^w \frac{\mathcal{P}_m(f)}{f^2} df$

Osserviamo che se k_ϕ (o k_f) tendono a zero, $\mathcal{P}_\alpha(f)$ si riduce ad un impulso, corrispondente alla portante non modulata. All'aumentare di k_ϕ (o k_f), aumenta anche σ_α^2 e dunque il termine $e^{-\sigma_\alpha^2}$ diminuisce, riducendo la concentrazione di potenza a frequenza portante. Dato che risulta comunque $\mathcal{P}_\alpha = a^2$, la potenza residua si distribuisce sugli altri termini, rappresentati da $\mathcal{P}_\alpha(f)$ e le sue autoconvoluzioni. E' immediato notare come, al crescere di k_ϕ (o k_f), cresca la banda.

In appendice 9.4.5 è illustrata una tecnica di modulazione per segnali FM modulati a basso indice.

9.4 Appendici

9.4.1 Calcolo della potenza di un segnale AM BLU

Mostriamo che se $X_{BLU}(t) = \frac{k_a}{\sqrt{2}} (m(t) \cos \omega_0 t - \hat{m}(t) \sin \omega_0 t)$, allora $\mathcal{P}_x = \frac{k_a^2}{2} \mathcal{P}_m$. Possiamo innanzitutto scrivere che

$$\mathcal{P}_x = \mathcal{P}_{x^+} + \mathcal{P}_{x^-} = 2\mathcal{P}_{x^+}$$

in quanto le componenti a frequenza positiva e negativa di $x(t)$ sono ortogonali (infatti $\int_{-\infty}^{\infty} X^+(f) X^-(f) df = 0$), e lo spettro di densità di potenza è una funzione pari della frequenza: $\mathcal{P}_x(f) = \mathcal{P}_x(-f)$. Inoltre, invertendo la relazione $\mathcal{P}_x(f) = 4\mathcal{P}_{x^+}(f - f_0)$ valida per la densità di potenza dell'involuppo complesso, otteniamo $\mathcal{P}_{x^+}(f) = \frac{1}{4}\mathcal{P}_x(f + f_0)$, e quindi

$$\mathcal{P}_{x^+} = \frac{1}{4} \int_{-\infty}^{\infty} \mathcal{P}_x(f + f_0) df = \frac{1}{4} \mathcal{P}_x$$

che, sostituita nella prima relazione mostrata, fornisce $\mathcal{P}_x = 2\mathcal{P}_{x^+} = \frac{1}{2}\mathcal{P}_x$.

Nel caso AM-BLU si ha inoltre $\underline{x}(t) = \frac{k_a}{\sqrt{2}} [m(t) + j\hat{m}(t)]$, tenendo ora conto che $\mathcal{P}_x = \mathcal{R}_{\underline{x}}(0)$, si ottiene

$$\mathcal{P}_x = \frac{1}{2} \left(\frac{k_a}{\sqrt{2}} \right)^2 [\mathcal{R}_{MM}(0) + \mathcal{R}_{\hat{M}\hat{M}}(0) + 2j\mathcal{R}_{M\hat{M}}(0)]$$

Osserviamo ora che $\mathcal{R}_{M\hat{M}}(0) = \int_{-\infty}^{\infty} m(t) \hat{m}(t) dt = 0$ in quanto $m(t)$ ed $\hat{m}(t)$ sono ortogonali; inoltre, $\mathcal{R}_{MM}(0) = \mathcal{P}_M = \mathcal{R}_{\hat{M}\hat{M}}(0)$ (non dimostrato ma intuibile). Pertanto si ottiene

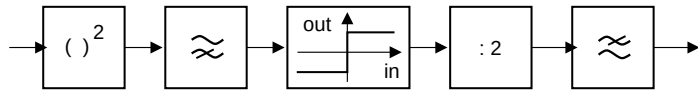
$$\mathcal{P}_x = \frac{1}{2} \frac{k_a^2}{2} [\mathcal{P}_m + \mathcal{P}_m] = \frac{1}{4} k_a^2 \cdot 2\mathcal{P}_m = \frac{k_a^2}{2} \mathcal{P}_m$$

9.4.1.1 Calcolo della potenza di segnali BLD-PI, PS, PPS

Vale lo stesso procedimento adottato sopra, in cui ora

$$\mathcal{P}_x = \mathcal{P}_{x_c} = \begin{cases} k_a^2 \mathcal{P}_m & \text{(BLD-PS)} \\ a_p^2 + k_a^2 \mathcal{P}_m & \text{(BLD-PI, PPS)} \end{cases}$$

9.4.2 Ricostruzione della portante mediante quadratura



Nel caso di una trasmissione BLD-PS, la portante di demodulazione può essere ottenuta mediante lo schema riportato. Il segnale

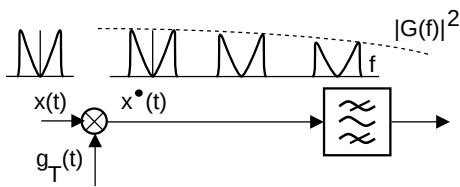
$x(t) = m(t) \cos(\omega_0 t + \varphi)$ viene elevato al quadrato producendo

$$\frac{1}{2}m^2(t) [1 + \cos(2\omega_0 t + 2\varphi)]$$

il cui termine di banda base $\frac{1}{2}m^2(t)$ viene rimosso dal filtro passa alto. Successivamente, lo squadratore produce un'onda quadra a frequenza $2f_0$ che viene divisa per 2 da un contatore binario. Infine, un passa basso provvede ad eliminare le armoniche dell'onda quadra a frequenza nf_0 , fornendo così la portante desiderata, a meno di una ambiguità di segno.

9.4.3 Il mixer

Il dispositivo moltiplicatore, presente negli schemi di mo-demodulazione, viene anche chiamato *mixer*, in quanto miscela tra loro due segnali.



Non è strettamente necessario disporre di un oscillatore sinusoidale per realizzare il prodotto di un segnale con una portante: ridotto ai minimi termini... è sufficiente un'onda quadra ed un filtro! Infatti, un qualunque segnale periodico

$$g_T(t) = g(t) * \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

di periodo $T = k/f_0$ (con k intero), possiede una densità di potenza

$$\mathcal{P}_{G_T}(f) = |G(f)|^2 \cdot \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - nf_0)$$

Il prodotto di tale segnale per $x(t)$, produce un $x^*(t)$ con densità di potenza

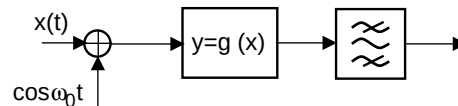
$$\mathcal{P}_{x^*}(f) = \mathcal{P}_x(f) * \mathcal{P}_{G_T}(f) = \frac{|G(f)|^2}{T} \sum_{n=-\infty}^{\infty} \left| X\left(f - \frac{n}{k}f_0\right) \right|^2$$

Pertanto, il desiderato spettro di potenza si ottiene inserendo dopo il moltiplicatore un filtro passa banda centrato su f_0 , ossia sulla k -esima replica spettrale di $\mathcal{P}_{x^*}(f)$. Lo stesso dispositivo può essere usato anche per i moltiplicatori di ricezione: in tal caso, il filtro da usare sarà un passa basso.

Un secondo metodo di realizzare il mixer è con un sommatore, un oscillatore, un dispositivo non lineare, e di nuovo un filtro passa-banda. Il dispositivo non lineare è del tipo

$$y = a_1x + a_2x^2 + a_3x^3 + \dots$$

e quando in ingresso viene applicata la somma di due segnali $x(t) + \cos \omega_0 t$, produce in uscita



$$y(t) = a_1(x(t) + \cos \omega_0 t) + a_2(x^2(t) + \cos^2 \omega_0 t + 2x(t) \cos \omega_0 t) + a_3(\dots) + \dots$$

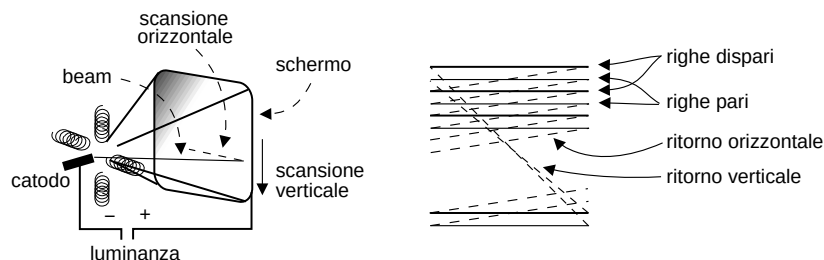
in cui, osservando che i termini $\cos^n \omega_0 t$ sono relativi a termini a frequenza $n f_0$, è possibile ancora una volta estrarre il termine che ci interessa.

9.4.4 Trasmissione televisiva

Illustriamo molto brevemente le principali caratteristiche delle trasmissioni televisive broadcast analogiche, con riferimento agli standard attuati in Italia.

Codifica dell'immagine Una trasmissione televisiva avviene riproducendo 25 diverse immagini (dette *quadri*) al secondo. Ogni immagine è scomposta in 625 linee orizzontali, che vengono trasmesse in due fasi: prima le linee dispari, poi quelle pari. In questo modo un singolo quadro è riprodotto due volte³² ogni $\frac{1}{25} = 0.04$ secondi (seppure in modo alternato) portando così a 50 semiquadri/secondo³³ la frequenza di rinfresco, in modo da impedire i fenomeni di *sfarfallamento* (FLICKER) ottico³⁴.

La riproduzione di un quadro avviene mediante un *tubo catodico*, il quale dispone posteriormente di un catodo che emette elettroni, che sono accelerati da un segnale di luminanza positivo applicato all'anodo, e che terminano la loro corsa contro lo strato di fosforo distribuito sulla parte anteriore (schermo) del tubo. Il *fascio* (BEAM) di elettroni è focalizzato elettronicamente, e viene deflesso ciclicamente sia in orizzontale alla frequenza di $625 \frac{\text{linee}}{\text{quadro}} \cdot 25 \frac{\text{quadri}}{\text{secondo}} = 15625 \text{ Hz}$ (*frequenza di riga*), sia verticalmente con velocità di $50 \frac{\text{semiquadri}}{\text{secondo}}$.



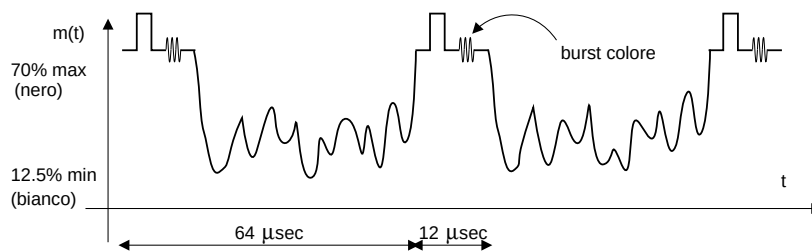
Modalità di scansione interlacciata dell'immagine televisiva

³²La riproduzione di metà quadro alla volta è chiamata *scansione interlacciata* dell'immagine. Nulla vieta al costruttore del ricevitore di prevedere una *memoria di quadro* e di riprodurre le immagini in modo non interlacciato; il segnale trasmesso invece presenta sempre le righe in formato interlacciato.

³³La frequenza di 50 semiquadri/secondo è stata scelta di proposito uguale alla frequenza di funzionamento della rete elettrica, in modo che eventuali disturbi elettrici avvengano sempre *nello stesso punto* dell'immagine, riducendo gli effetti fastidiosi.

³⁴Il *flicker* si manifesta nel caso in cui la frequenza di rinfresco è inferiore al tempo di persistenza delle immagini sulla retina, pari a circa $\frac{1}{40}$ di secondo.

Segnale televisivo in bianco e nero Il segnale televisivo contiene sia le informazioni di temporizzazione necessarie a sincronizzare la scansione dell'immagine, che l'informazione di luminanza che pilota la tensione anodica, e quindi la forza con cui l'elettrone urta lo schermo.



Forma d'onda del segnale televisivo

Durante la trasmissione di ogni semiquadro, ogni riga dispone di $\frac{1}{15625} = 64 \mu\text{secondi}$. Il segnale modulante è sempre positivo, ed associa ai valori più piccoli la maggiore luminanza³⁵, trasmettendo in logica negata, in modo che gli impulsi di sincronismo orizzontale siano di ampiezza superiore al livello del nero (il 70 % dell'ampiezza). Il tempo dedicato alla trasmissione della luminanza di una riga è di 52 μsec , mentre nei restanti 12 il segnale oltrepassa il livello del nero (in modo da rendere invisibile il beam) e quindi un impulso determina il ritorno orizzontale. In figura è anche mostrato un *burst colore* che è presente nelle trasmissioni a colori per sincronizzare la *portante di colore* (vedi di seguito).

Formato dell'immagine Ogni singolo quadro è realizzato con un rapporto di aspetto 4:3 (che rappresenta il rapporto tra le dimensioni orizzontale e verticale), e solo 575 delle 625 linee vengono mostrate (infatti 25 linee per ogni semiquadro cadono al di fuori dello schermo³⁶).

Occupazione spettrale Diverse considerazioni³⁷ hanno portato a stabilire che la banda del segnale televisivo sia di circa $\pm 5 \text{ Mhz}$, e nell'ultima versione del sistema PAL questa è stata portata a 6 MHz. In particolare, dato che le immagini presentano spesso ampie zone uniformi, corrispondenti ad un segnale di luminanza pressoché costante, la densità spettrale del segnale televisivo è piuttosto concentrata nella regione delle basse frequenze. Per questo motivo, si è deciso di trasmettere il segnale mediante modulazione di ampiezza a banda laterale ridotta, conseguendo un risparmio di banda e contemporaneamente preservando le componenti del messaggio a frequenze più basse.

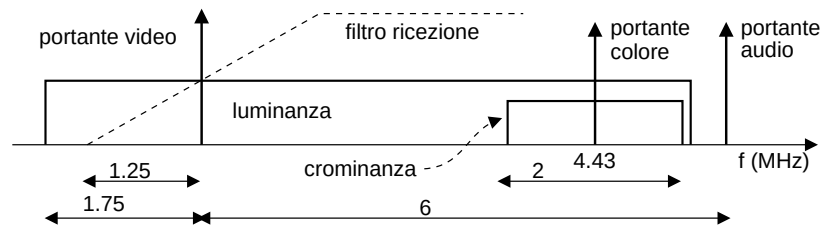
La figura seguente mostra la situazione, in cui solo parte (1.75 MHz) della banda inferiore del segnale di luminanza viene trasmessa, mentre il filtro di ricezione provvede

³⁵In questo modo si riduce mediamente la potenza trasmessa, dato che sono più frequenti scene chiare che scure.

³⁶Nel tempo destinato alle linee che non sono mostrate, vengono comunque trasmesse altre informazioni, come ad esempio i dati che compaiono nelle pagine del televideo.

³⁷Ad esempio, si può stabilire di realizzare la stessa risoluzione orizzontale e verticale. A fronte delle 625 linee, il rapporto di aspetto di $\frac{4}{3}$ determina l'esigenza di individuare $625 \cdot \frac{4}{3} = 833 \frac{\text{punti}}{\text{linea}}$, e quindi $833 \cdot 625 = 520625 \frac{\text{punti}}{\text{quadro}}$, ossia circa $13 \cdot 10^6 \frac{\text{punti}}{\text{secondo}}$. Per il teorema del campionamento, il segnale deve avere una banda minore od uguale di $\frac{f_c}{2} = 6.5 \text{ MHz}$.

a realizzare un filtraggio complessivo tale che $\underline{H}(f) + \underline{H}^*(-f) = \text{cost}$ (vedi nota 19 a pag. 209).



Occupazione spettrale di un segnale televisivo

Segnale audio Nella figura precedente è rappresentata anche una portante audio, che viene trasmessa oltre la banda del segnale video, mediante una modulazione FM con $\Delta f_{Max} = 25$ KHz.

Segnale di crominanza Il requisito che più di altri ha determinato quale soluzione adottare per effettuare trasmissioni a colori, è che queste dovessero essere correttamente visibili anche da parte di ricevitori in bianco e nero.

Un risultato di colorimetria è che ogni colore è scomponibile nella somma di tre colori fondamentali (verde, rosso e blu), effettivamente operata dagli apparati di acquisizione. La somma³⁸ della tre componenti fornisce il segnale di luminanza L , che viene utilizzato esattamente come per il bianco e nero. Il segnale di crominanza è invece costruito da una coppia di segnali differenza $\begin{cases} \Delta_R = R - L \\ \Delta_B = B - L \end{cases}$, che sono usati per modulare in ampiezza, portante soppressa, una portante di colore, usando Δ_R come componente in fase e Δ_B come componente in quadratura³⁹.

L'occupazione spettrale del segnale di crominanza è ridotta (± 1 MHz) rispetto a quello di luminanza, in quanto la *risoluzione spaziale* dell'occhio umano è ridotta per stimoli colorati, e quindi Δ_R e Δ_B possono variare più lentamente di L .

Sincronizzazione Per impedire fenomeni di interferenza tra C.A. di B.F. nella ricezione del segnale di crominanza, occorre effettuare una demodulazione omodina, e l'oscillatore del ricevitore si mantiene coerente con la portante di colore, grazie ai *burst di colore* presenti dopo l'impulso di sincronizzazione orizzontale, costituiti da 8 cicli di portante. Questo segnale ha inoltre lo scopo di segnalare la *presenza* della componente di crominanza: in caso contrario infatti (trasmissione B/N) il ricevitore deve disattivare il circuito del colore, per non produrre deterioramenti dell'immagine.

³⁸In realtà ogni componente è pesata mediante un opportuno coefficiente che tiene conto della diversa sensibilità dell'occhio ai tre colori fondamentali. Infatti, per ottenere il bianco, i tre colori non devono essere mescolati in parti uguali, bensì 59% di verde, 30% di rosso e 11% di blu.

³⁹Le ampiezze delle componenti in fase e quadratura del segnale di crominanza devono essere opportunamente scalate, per impedire al segnale complessivo (luminanza più crominanza) di assumere valori troppo elevati.

Interferenza La presenza di entrambi i segnali di luminanza e cromaticanza nella stessa banda sembrerebbe dare luogo a difficili problemi di interferenza. Innanzitutto osserviamo che, come anticipato, il segnale di luminanza è concentrato attorno alla portante video, e dunque arreca un disturbo ridotto⁴⁰ alla cromaticanza. Quest'ultima quindi, prima di essere demodulata, viene filtrata per rimuovere il segnale di luminanza fuori della banda di cromaticanza, ed il disturbo è generalmente trascurabile. Viene inoltre adottata una soluzione che riduce anche l'interferenza di cromaticanza su luminanza. Queste presentano infatti una spiccata periodicità, legata alla frequenza di riga f_r ed alla presenza degli impulsi di sincronismo ogni $64 \mu\text{sec}$, che determina uno spettro con energia concentrata alle armoniche di $f_r = 15625 \text{ Hz}$. Pertanto, si colloca la portante di colore a metà tra due armoniche del segnale di luminanza, in modo che le densità spettrali risultino, pur se sovrapposte, intercalate. L'uso di filtri a pettine nel ricevitore può quindi ridurre notevolmente l'interferenza.

9.4.5 Modulazione FM a basso indice

Riprendiamo qui il caso in cui $\beta \ll 1 \Rightarrow \Delta\alpha \ll 1$, e quindi l'espansione in serie di potenze di $\underline{x}(t)$ può arrestarsi al primo ordine; se il segnale modulante è cosinusoidale, il segnale FM risulta

$$x_{FM}(t) = a \cos \left(\omega_0 t + 2\pi k_f \int_{-\infty}^t \cos(2\pi w\tau) d\tau \right) = a \cos(\omega_0 t + \beta \sin(2\pi w t))$$

Ricordando che $\cos(\alpha + \beta) = \cos\alpha \cos\beta - \sin\alpha \sin\beta$, $x_{FM}(t)$ può essere riscritto come

$$x_{FM}(t) = a \cos \omega_0 t \cos(\beta \sin 2\pi w t) - a \sin \omega_0 t \sin(\beta \sin 2\pi w t)$$

che, se $\beta \ll 1$, diviene

$$x_{FM}(t) = a \cos \omega_0 t - \beta a \sin \omega_0 t \sin 2\pi w t$$

che confrontiamo con l'espressione

$$x_{AM}(t) = a_p \cos \omega_0 t + k_a \cos \omega_0 t \cos 2\pi w t$$

che si otterrebbe per modulazione a portante intera, o ridotta, dello stesso $m(t)$.

Il confronto rivela che il segnale modulante, mentre nell'AM opera *in fase* alla portante, nell'FM a basso indice opera *in quadratura*. Il risultato esposto costituisce ad ogni modo uno *schema di modulazione* per segnali FM a basso indice, realizzabile sommando alla portante un segnale modulato AM su di una portante in quadratura.

Resta il fatto che uno schema di modulazione del genere produce anche una modulazione AM parassita: quest'ultima è eliminata in ricezione dall'azione congiunta di uno squadratore e di un filtro passa basso.

9.4.6 FM BroadCast

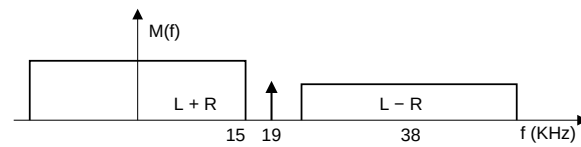
Illustriamo brevissimamente i parametri delle trasmissioni FM ricevibili mediante "la radio di casa". Nella banda 88-108 MHz operano le radio FM, con spaziatura di 200 KHz l'una dall'altra. Ad ogni emittente è concessa una deviazione massima della frequenza istantanea pari a $\Delta f = 75 \text{ KHz}$.

⁴⁰Possiamo riflettere su quali siano le circostanze che producono la massima interferenza della luminanza sulla cromaticanza: ciò avviene in corrispondenza di scene molto definite, relative ad immagini con elevato contenuto di frequenze spaziali elevate, ad esempio nel caso di righe fitte; il disturbo è più appariscente nel caso in cui la zona ad elevato contrasto sia povera di componenti cromatiche. Avete mai notato cravatte a righe bianche e nere, divenire cangianti?

Il trasmettitore viene tarato mediante un $m(t)$ sinusoidale a frequenza di 15 KHz, e k_f regolato in modo da ottenere $\Delta f = 75$ KHz. In queste condizioni, risulta $\beta = \frac{k_f}{w} = \frac{75}{15} = 5$, e la regola di Carson fornisce $B_C = 2(k_f + w) = 2(75 + 15) = 180$ KHz. Un esame degli andamenti riportati in Fig. 9.1 mostra che per $\beta = 5$, le $J_n(\beta) \neq 0$ sono le prime 8, e dunque la “vera” banda ha una estensione $B = 2 \cdot 8w = 16 \cdot 15 \cdot 10^3 = 240$ KHz, mostrando l'approssimazione della regola di Carson. D'altra parte, risulta che $2 \sum_{n=6}^8 |J_n(\beta)|^2 = 2[(.13)^2 + (.05)^2 + (.02)^2] = 2 \cdot 0.0198 = 0.0396$, e dunque l'errore commesso esclude circa il 4% della potenza totale.

Qualora il segnale sinusoidale venga sostituito da un messaggio limitato in banda con $\pm W = \pm 15$ KHz, con potenza eguale a quella del seno e cioè $P_m = \frac{1}{2}$, la Δf non è più definita con esattezza, e conviene ricorrere alla definizione di $\beta' = \frac{\sigma_f}{W} = \frac{k_f \sqrt{P_m}}{W} = \beta \sqrt{P_m} = \beta \frac{1}{\sqrt{2}} = 0.707 \cdot \beta$, a cui corrisponde una banda efficace $B = 2W(\beta' + 1) = 2 \cdot 15 \cdot 10^3 \cdot (0.707 \cdot 5 + 1) = 136$ KHz.

Nell'FM stereo, il segnale trasmesso deve essere compatibile con i ricevitori mono, ed allora il segnale modulante è un segnale multiplato FDM, e “composto” da tre “canali”:



- La somma di Left + Right (L+R) come segnale di banda base, che consente la compatibilità con gli apparati “mono”;
- Il segnale L-R è centrato a frequenza di 38 KHz mediante modulazione AM-BLD;
- A 19 KHz, è presente una portante a cui si concede il 10% di $\mathcal{P}_M$, mentre il restante 90% di $\mathcal{P}_M$ è condiviso tra L+R e L-R. Il tutto è poi modulato FM.

La portante a 19 KHz può essere impiegata per sincronizzare il ricevitore, e generare la portante (a frequenza doppia, di 38 KHz) necessaria a demodulare il canale L-R. Se assente, indica la ricezione di un canale mono.

A prima vista, sembrerebbe che la presenza del canale L-R possa aumentare la massima deviazione di frequenza. In realtà non è così, per due motivi:

- quando L+R è grande, vuol dire che i due canali sono simili, e dunque L-R è piccolo, e viceversa;
- Il canale L-R, trovandosi a frequenze più elevate, è caratterizzato da un indice di modulazione inferiore. Infatti, la massima deviazione di frequenza istantanea dipende dalle ampiezze di $m(t)$, e non dalla sua banda.

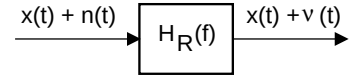
Capitolo 10

Prestazioni delle Trasmissioni Modulate

10.1 Il rumore nei segnali modulati

Consideriamo un segnale modulato, affetto da un rumore additivo gaussiano bianco $n(t)$, con densità di potenza

$$\mathcal{P}_n(f) = \frac{N_0}{2}$$



Prima di effettuare la demodulazione, il segnale ricevuto transita in un filtro di ricezione $H_R(f)$, che ha lo scopo di limitare la banda del rumore ricevuto, e quindi ridurre l'entità della potenza di rumore in ingresso al ricevitore. Il filtro $H_R(f)$ presenta un modulo costante nella banda del segnale, mentre tende a zero al di fuori di tale banda. In questo modo, il segnale utile $x(t)$ transita inalterato, ed il rumore $n(t)$ viene limitato in banda, producendo $v(t)$.

10.1.1 Rapporto Segnale-Rumore e Banda di Rumore

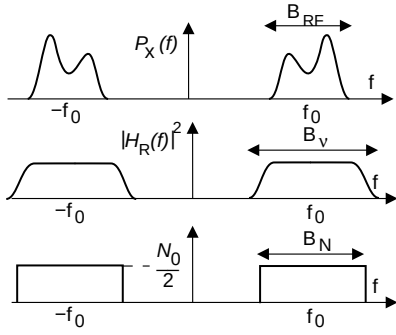
Per descrivere la qualità del segnale modulato ricevuto, si ricorre al rapporto

$$SNR_{RF}(f) = \frac{\mathcal{P}_x(f)}{\mathcal{P}_v(f)}$$

indicato come SNR LOCALE, a sottolineare la sua dipendenza dalla frequenza. Per ciò che riguarda $\mathcal{P}_x(f)$, questo è ottenibile come illustrato al capitolo precedente, una volta noto il segnale modulante e la tecnica di modulazione, mentre $\mathcal{P}_v(f)$ risulta pari a

$$\mathcal{P}_v(f) = \mathcal{P}_n(f) |H_R(f)|^2 = \frac{N_0}{2} |H_R(f)|^2$$

D'altra parte, ha senso valutare l'*SNR complessivo*, ovvero il rapporto tra le potenze di segnale e di rumore *totali*. Per ciò che riguarda il segnale, è ben noto che $\mathcal{P}_y = \int_{-\infty}^{\infty} \mathcal{P}_y(f) df$; in modo analogo, si ottiene $\mathcal{P}_v = \int_{-\infty}^{\infty} \frac{N_0}{2} |H_R(f)|^2 df$, valutandone cioè la potenza che attraversa il filtro di ricezione $H_R(f)$. Data l'impossibilità pratica di realizzare un filtro ideale (rettangolare), $H_R(f)$ è caratterizzato da una banda (a frequenze positive) B_v , più estesa di B_{RF} (che è la banda di segnale).



La potenza totale del rumore uscente da $H_R(f)$ risulta pertanto pari a

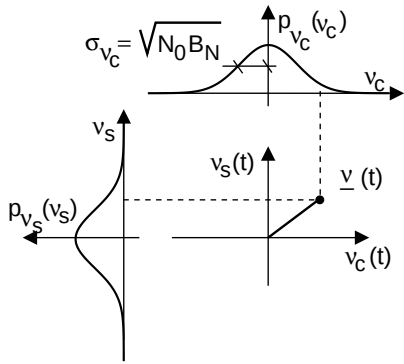
$$\begin{aligned} \mathcal{P}_\nu &= \frac{N_0}{2} \cdot \int_{-\infty}^{\infty} |H_R(f)|^2 df = \frac{N_0}{2} \cdot 2 \cdot \int_0^{\infty} |H_R(f)|^2 df = \\ &= N_0 B_N |H_R(f_0)|^2 \end{aligned}$$

Il termine B_N rappresenta la cosiddetta *banda di rumore* definita come

$$B_N = \frac{\int_0^{\infty} |H_R(f)|^2 df}{|H_R(f_0)|^2}$$

ossia come la banda di un filtro ideale (rettangolare) che lascia passare la stessa quantità di rumore.

10.1.2 Demodulazione di un processo di rumore



Il rumore $\nu(t)$ che esce dal filtro di ricezione $H_R(f)$ è di tipo passa-banda, e può quindi essere descritto in termini delle sue componenti analogiche di bassa frequenza:

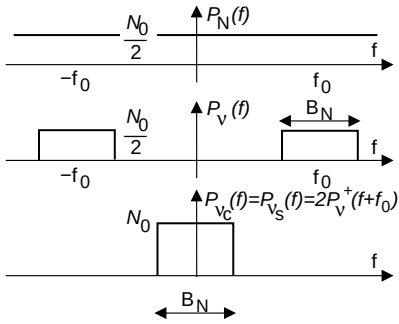
$$\nu(t) = \nu_c(t) \cos \omega_0 t - \nu_s(t) \sin \omega_0 t$$

Osserviamo che nel caso in cui la banda di $\nu(t)$ sia *stretta rispetto a f_0* , l'involuppo complesso $\underline{\nu}(t) = \nu_c(t) + j\nu_s(t)$ evolve *lentamente* rispetto alla velocità di rotazione di $\underline{\nu}(t) e^{j\omega_0 t}$.

Ricordando ora i risultati ottenuti al § 8.4.2, nel caso in cui il rumore $n(t)$ in ingresso a $H_R(f)$ sia un processo ergodico bianco con densità di potenza $\mathcal{P}_n(f) = \frac{N_0}{2}$, e se il filtro di ricezione ha una risposta in frequenza unitaria a frequenza portante, ossia $|H_R(f_0)|^2 = 1$, allora si ottiene che $\nu_c(t)$ e $\nu_s(t)$ risultano essere due processi congiuntamente gaussiani, ergodici, a media nulla ed uguale varianza (e potenza)

$$\sigma_{\nu_c}^2 = \sigma_{\nu_s}^2 = \mathcal{P}_\nu = N_0 B_N$$

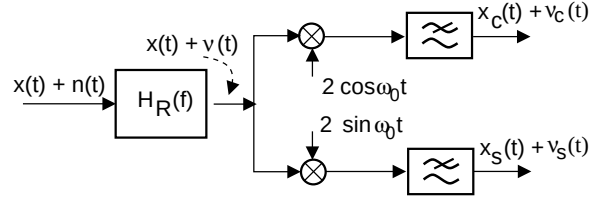
Pertanto, nel caso in cui si operi una demodulazione coerente in fase ed in quadratura del segnale ricevuto, nelle componenti analogiche risultanti saranno presenti i termini additivi $\nu_c(t)$ e $\nu_s(t)$, entrambi di potenza $\mathcal{P}_\nu = N_0 B_N$.



10.2 Prestazioni delle trasmissioni AM

Per valutare il rapporto SNR per le diverse tecniche di modulazione AM, esprimiamo il segnale modulato nei termini delle sue componenti analogiche

$$x_{AM}(t) = x_c(t) \cos \omega_0 t - x_s(t) \sin \omega_0 t$$



Operando una demodulazione coerente del segnale $x(t)$ affetto da rumore bianco $n(t)$ e filtrato da $H_R(f)$ come in figura, si ottengono le due componenti analogiche per il segnale demodulato: $\begin{cases} d_c(t) = x_c(t) + \nu_c(t) \\ d_s(t) = x_s(t) + \nu_s(t) \end{cases}$.

Tra la potenza del segnale ricevuto e quella delle sue C.A. di B.F. sussiste¹ la relazione: $\mathcal{P}_x = \frac{1}{2}\mathcal{P}_{x_c} + \frac{1}{2}\mathcal{P}_{x_s}$.

10.2.1 Potenza di segnale e di rumore dopo demodulazione. SNR

Nel caso di modulazione AM, siamo interessati alla sola componente in fase, che è sufficiente a fornire $m(t)$; il rapporto tra $\mathcal{P}_{x_c}$ e $\mathcal{P}_{\nu_c}$ fornirà dunque il valore di SNR che stiamo cercando. La tabella che segue mostra i valori delle componenti di segnale e di rumore, assieme alle rispettive potenze espresse in funzione di una medesima potenza di messaggio $\mathcal{P}_m$, per tre casi di modulazione AM, di cui discutiamo individualmente.

	$x_c(t)$	$x_s(t)$	$\mathcal{P}_{x_c}$	$\mathcal{P}_x$	B_N	$\mathcal{P}_{\nu_c}$
BLD-PS	$m(t)$	0	$\mathcal{P}_m$	$\frac{1}{2}\mathcal{P}_m$	$2W$	$2WN_0$
BLD-PI	$\sqrt{\eta}(a_p + m(t))$	0	$\eta(a_p^2 + \mathcal{P}_m)$	$\frac{1}{2}\eta(a_p^2 + \mathcal{P}_m) = \frac{1}{2}\mathcal{P}_m$	$2W$	$2WN_0$
BLU-PS	$\frac{1}{\sqrt{2}}m(t)$	$\mp \frac{1}{\sqrt{2}}\hat{m}(t)$	$\frac{1}{2}\mathcal{P}_m$	$\frac{1}{2}\mathcal{P}_m$	W	WN_0

Rispetto alla notazione adottata al Capitolo 9, si considera ora il termine k_a inglobato in $m(t)$, e quindi non evidenziato in tabella. Inoltre, la banda di rumore B_N presa in considerazione nella tabella è la minima possibile, pari a quella del segnale modulato B_{RF} , direttamente legata (nella modulazione AM) a quella del segnale modulante $\pm W$. Pertanto, i risultati che otterremo sono i migliori possibili: infatti, se $B_N > B_{RF}$, l' SNR risulterà peggiore. Precisiamo infine che nella valutazione dell' SNR che segue, ci si riferisce sempre ad una medesima potenza ricevuta $\mathcal{P}_x$ e ad una densità di rumore $\mathcal{P}_n(f) = \frac{N_0}{2}$, allo scopo di rendere confrontabili i risultati.

10.2.1.1 BLD-PS

Si ha $\mathcal{P}_{x_c} = \mathcal{P}_m$ e $\mathcal{P}_x = \frac{1}{2}\mathcal{P}_m$; dopo demodulazione il segnale $d_c(t) = x_c(t) + \nu_c(t)$ presenta dunque una potenza di segnale utile $\mathcal{P}_{x_c} = \mathcal{P}_m = 2\mathcal{P}_x$ (con $\mathcal{P}_x$ pari alla potenza del segnale ricevuto) ed una potenza di rumore $\mathcal{P}_{\nu_c} = 2WN_0$; dunque un

$$SNR = \frac{\mathcal{P}_{x_c}}{\mathcal{P}_{\nu_c}} = \frac{\mathcal{P}_m}{2WN_0} = \frac{\mathcal{P}_x}{WN_0} = SNR_0$$

in cui nell'ultima eguaglianza si definisce:

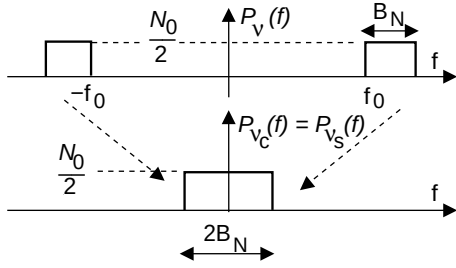
¹Infatti i segnali $x_c(t) \cos \omega_0 t$ e $x_s(t) \sin \omega_0 t$ risultano ortogonali, e le potenze si sommano.

SNR convenzionale La grandezza SNR_0 è definita come *rapporto segnale-rumore convenzionale*, ed è il rapporto tra la *potenza di segnale ricevuto* e la *potenza di rumore in una banda pari a quella del messaggio di banda base*.²

Osserviamo dunque che la modulazione BLD-PS non altera il rapporto SNR_0 (è come se il processo di modulazione fosse assente). Notiamo infine (e questo è valido anche per i casi che seguono) che SNR può riferirsi indifferentemente sia alle potenze di segnale che a quelle disponibili (vedi Cap. 12), in quanto $SNR_0 = \frac{\mathcal{P}_x}{WN_0} = \frac{\mathcal{P}_x}{WN_0} \frac{4R_g}{4R_g} = \frac{W_{d_x}}{W_{d_N}}$.

10.2.1.2 BLU-PS

In questo caso, per ottenere una $\mathcal{P}_x = \frac{1}{2}\mathcal{P}_m$ uguale al caso BLD-PS, le componenti $x_c(t)$ ed $x_s(t)$ devono essere poste pari a $\frac{1}{\sqrt{2}}m(t)$ e $\frac{1}{\sqrt{2}}\hat{m}(t)$, rispettivamente (vedi § 9.4.1).



A seguito del processo di demodulazione, si ottiene un rumore in banda base che occupa ancora una banda B_N , ma possiede una densità uguale a quella del rumore a RF, in quanto i contenuti a frequenze positive e negative non si sovrappongono, come mostrato in figura. Risulta:

$$SNR = \frac{\mathcal{P}_{x_c}}{\mathcal{P}_{\nu_c}} = \frac{\frac{1}{2}\mathcal{P}_m}{WN_0} = \frac{\mathcal{P}_x}{WN_0} = SNR_0$$

Dunque, si ottengono prestazioni identiche a quelle BLD. Si noti che il risultato è valido solo se $\nu(t)$ è effettivamente limitato alla sola banda B_{RF} . Se infatti si fosse adottato un filtro con banda più larga, come ad esempio un $H_R(f)$ con $B_N = 2W$, si sarebbe ottenuto $\mathcal{P}_{\nu_c}(f) = N_0$, ed SNR risulterebbe dimezzato.

10.2.1.3 BLD-PI

Per ottenere una potenza di segnale ricevuto $\mathcal{P}_x = \frac{1}{2}\mathcal{P}_m$ uguale ai due casi precedenti, si considera la ricezione di un segnale

$$x(t) = \sqrt{\eta}(a_p + m(t)) \cos \omega_0 t \quad \text{in cui} \quad \eta = \frac{\mathcal{P}_m}{a_p^2 + \mathcal{P}_m}$$

ovvero η è proprio pari all'efficienza della BLD-PI introdotta al § 9.1.1.4. Nel valutare l' SNR , faremo riferimento alla sola componente di messaggio $\sqrt{\eta}m(t)$ del segnale demodulato $\mathcal{P}_{x_c}$, che ha potenza $\eta\mathcal{P}_m = 2\eta\mathcal{P}_x$. La quantità a_p^2 si riferisce infatti alla portante non modulata, e non fornisce informazione. Si ha pertanto:

$$SNR = \frac{\mathcal{P}_{x_c}}{\mathcal{P}_{\nu_c}} = \frac{2\eta\mathcal{P}_x}{2WN_0} = \eta \frac{2\mathcal{P}_x}{2WN_0} = \eta SNR_0$$

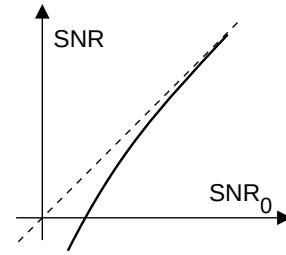
Dunque in questo caso constatiamo che la presenza della portante comporta una riduzione di prestazioni proprio pari all'efficienza $\eta = \frac{\mathcal{P}_m}{a_p^2 + \mathcal{P}_m}$.

²In virtù di questa definizione, SNR_0 è una grandezza che caratterizza le condizioni operative ($\mathcal{P}_x$, $\mathcal{P}_n(f) = \frac{N_0}{2}$ e W sono grandezze *indipendenti*) ma non è legata alla particolare tecnica di modulazione adottata. Pertanto, esprimere SNR in funzione di SNR_0 permette il confronto tra i diversi casi *a parità di condizioni operative*.

L'analisi fin qui esposta si riferisce però al caso di demodulazione coerente: invece per BLD-PI si usa il demodulatore di inviluppo! In tal caso, il segnale demodulato è il modulo dell'inviluppo complesso, ovvero

$$d(t) = |\underline{x}(t) + \underline{\nu}(t)| = \sqrt{[\sqrt{\alpha}(a_p + m(t)) + \nu_c(t)]^2 + \nu_s^2(t)}$$

Nel caso in cui $\underline{\nu}(t)$ sia piccolo, si può ottenere una approssimazione che ci riconduce al caso precedente. In caso contrario, sorgono termini *prodotto* tra $m(t)$ e $\nu_c(t)$, ed in definitiva l' SNR risulta peggiore (per bassi SNR_0) del caso di demodulazione sincrona omodina, come illustrato nella curva riportata a fianco.



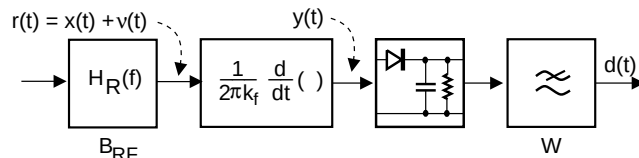
10.3 Prestazioni delle trasmissioni FM

Quando si è analizzata la tecnica di modulazione FM, si è fatto più volte notare la sua caratteristica non lineare. E' lecito aspettarsi che questa caratteristica determini dei risvolti "bizzarri" per quanto riguarda l' SNR del segnale demodulato: e difatti è proprio così. Anticipiamone due:

- La potenza del rumore *diminuisce* all'aumentare della potenza ricevuta
- L' SNR *migliora* all'aumentare della banda occupata.

10.3.1 Rumore dopo demodulazione FM

Analizziamo innanzitutto il comportamento di un demodulatore a discriminatore³, quando è presente in ingresso una portante non modulata di ampiezza⁴ $A = \sqrt{2P_x}$ sovrapposta ad un rumore gaussiano passa-banda:



$$r(t) = A \cos \omega_0 t + \nu_c(t) \cos \omega_0 t - \nu_s(t) \sin \omega_0 t$$

A differenza del caso AM, ora il filtro $H_R(f)$ lascia passare una banda di frequenze di estensione almeno pari alla banda occupata dal segnale FM, stimata applicando ad esempio la *regola di Carson*.

E' immediato verificare come le componenti analogiche di bassa frequenza di

$$\underline{r}(t) = r_c(t) + jr_s(t)$$

valgano $\begin{cases} r_c(t) = A + \nu_c(t) \\ r_s(t) = \nu_s(t) \end{cases}$, come mostrato nella

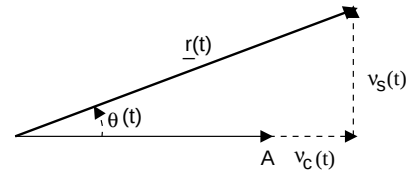


figura a fianco. Notiamo come, per piccoli valori (rispetto ad A) di $\nu_c(t)$ e $\nu_s(t)$, l'inviluppo complesso ricevuto $\underline{r}(t)$ rimanga "prossimo" al "segnale" $r_c(t) = A$. Come noto, $\nu_c(t)$ e $\nu_s(t)$ appartengono a due processi congiuntamente gaussiani, a media nulla e

³Descritto al § 9.3.2.2.

⁴Con questa posizione, la potenza della portante risulta $\frac{(\sqrt{2P_x})^2}{2} = \frac{2P_x}{2} = P_x$.

deviazione standard $\sigma_{\nu_c} = \sigma_{\nu_s} = \sqrt{N_0 B_N}$, in cui $B_N \geq B_{RF}$ è la banda di rumore del ricevitore, ed $N_0/2$ è la densità di potenza del rumore in ingresso.

Ricordiamo ora che nel caso FM, il *segnale informativo* è legato alla *derivata* della fase $\theta(t)$. Esprimiamo dunque $r(t)$ mettendo $\theta(t)$ in evidenza:

$$r(t) = \Re \{ \underline{r}(t) e^{j\omega_0 t} \} = \Re \{ |\underline{r}(t)| e^{j\theta(t)} e^{j\omega_0 t} \} = |\underline{r}(t)| \cos(\omega_0 t + \theta(t))$$

Osserviamo quindi che il termine $|\underline{r}(t)|$ viene rimosso dal limitatore che usualmente è anteposto al discriminatore. Il segnale $y(t)$ in uscita dal derivatore è quindi dato da

$$y(t) = \left(\frac{f_0}{k_f} + \frac{1}{2\pi k_f} \frac{d}{dt} \theta(t) \right) \sin(\omega_0 t + \theta(t))$$

che viene a sua volta elaborato da parte del demodulatore di inviluppo come fosse un segnale BLD-PI, fornendo in definitiva un segnale demodulato

$$d(t) = \frac{1}{2\pi k_f} \frac{d}{dt} \theta(t)$$

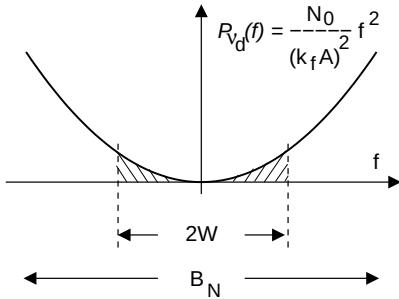
10.3.2 Caso di basso rumore

Se $\mathcal{P}_x = \frac{A^2}{2} \gg \sigma_{\nu_c}^2 = \sigma_{\nu_s}^2 = N_0 B_N$ allora, come osservato, l'involuppo complesso del rumore ha modulo "abbastanza" più piccolo di A . Pertanto si può scrivere

$$\theta(t) = \arctan \frac{\nu_s(t)}{A + \nu_c(t)} \simeq \arctan \frac{\nu_s(t)}{A} \simeq \frac{\nu_s(t)}{A}$$

e dunque

$$\mathcal{P}_\theta(f) = \frac{1}{A^2} \mathcal{P}_{\nu_s}(f) = \frac{N_0}{A^2}$$



Ricordando ora che l'operazione di derivata (svolta dal discriminatore) equivale a moltiplicare lo spettro di ampiezza del segnale che si deriva per $j2\pi f$, ovvero moltiplicare la sua densità di potenza per $(2\pi f)^2$, si ottiene che la densità di potenza del segnale demodulato (che in questo caso è tutta dovuta al rumore) risulta pari a

$$\mathcal{P}_{\nu_d}(f) = \frac{1}{(2\pi k_f)^2} (2\pi f)^2 \mathcal{P}_\theta(f) = \left(\frac{f}{k_f} \right)^2 \frac{N_0}{A^2}$$

Infine, troviamo che la potenza totale di rumore dopo demodulazione risulta pari a

$$\mathcal{P}_{\nu_d} = \sigma_{\nu_d}^2 = 2 \int_0^W \left(\frac{f}{k_f} \right)^2 \frac{N_0}{A^2} df = \frac{2}{3} \frac{N_0}{(k_f A)^2} W^3 \quad (10.1)$$

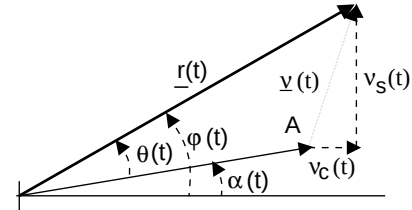
in cui W è la banda del segnale modulante, ed il rumore è limitato in tale banda in virtù del filtro di ricezione posto a valle del discriminatore. Si noti che invece le potenze $\sigma_{\nu_c}^2$ e $\sigma_{\nu_s}^2$ sono relative alla banda B_N , pari a quella del segnale modulato.

Notiamo subito la veridicità della prima affermazione fatta ad inizio capitolo: la potenza *complessiva* del rumore dopo demodulazione FM *diminuisce* all'aumentare della potenza del segnale ricevuto $\mathcal{P}_x = \frac{A^2}{2}$. Una seconda osservazione molto importante è che, per effetto della derivata, la densità di potenza del rumore demodulato ha un andamento *parabolico*.

Segnale presente

Continuando ad ipotizzare $\mathcal{P}_x = \frac{A^2}{2} \gg \sigma_{\nu_c}^2 = \sigma_{\nu_s}^2 = N_0 B_N$ possiamo osservare che, in presenza di una fase modulante $\alpha(t)$, la fase $\varphi(t)$ dell'involuppo complesso del segnale ricevuto $r(t)$ è costituita dalla somma di $\alpha(t)$ con l'angolo $\theta(t)$ dovuto al rumore sovrapposto alla portante di ampiezza A . Pertanto all'uscita del discriminatore si ottiene

$$d(t) = \frac{1}{2\pi k_f} \frac{d}{dt} (\alpha(t) + \theta(t))$$



Il rapporto SNR è ora definito come $SNR = \frac{\mathcal{P}_d}{\mathcal{P}_{\nu_d}}$, in cui $\mathcal{P}_{\nu_d}$ è la potenza del rumore demodulato, calcolata in (10.1), e $\mathcal{P}_d$ è la potenza di segnale utile demodulato, pari a $d(t) = \frac{1}{2\pi k_f} \frac{d}{dt} \alpha(t)$. Sappiamo che $\alpha(t) = 2\pi k_f \int_{-\infty}^t m(\tau) d\tau$, e pertanto la potenza di segnale utile demodulato risulta proprio pari a $\mathcal{P}_d = \mathcal{P}_m = \int_{-W}^W \mathcal{P}_m(f) df$. Quindi:

$$SNR = \frac{\mathcal{P}_d}{\mathcal{P}_{\nu_d}} = \frac{\mathcal{P}_m}{\frac{2}{3} \frac{N_0}{(k_f A)^2} W^3} = 3 \frac{\mathcal{P}_m k_f^2}{W^2 N_0 W} \frac{A^2}{2} = 3 \frac{\sigma_{f_d}^2}{W^2} \frac{\mathcal{P}_x}{N_0 W} = 3\beta^2 SNR_0$$

Il risultato ottenuto conferma la seconda affermazione di inizio sezione: si ha un *miglioramento* rispetto all' SNR convenzionale (e dunque rispetto all'AM) *tanto maggiore* quanto più grande è l'indice di modulazione β , ovvero *quanto maggiore è la banda occupata* dal segnale modulato.

Discussione dei passaggi Innanzi tutto, è ovvio che $\frac{A^2}{2} = \mathcal{P}_x$ (la potenza ricevuta non cambia ed è sempre uguale a quella della portante non modulata) e che $\frac{\mathcal{P}_x}{N_0 W} = SNR_0$, il rapporto tra potenza ricevuta e potenza di rumore *nella banda del messaggio*. Mostriamo ora che $\mathcal{P}_m k_f^2 = \sigma_{f_d}^2$. Ricordiamo che $f_d(t)$ rappresenta la deviazione istantanea di frequenza rispetto ad f_0 , ovvero la derivata della fase istantanea diviso 2π , e sottratta la frequenza portante:

$$f_d(t) = \frac{1}{2\pi} \frac{d}{dt} \left(2\pi f_0 t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right) - f_0 = (f_0 - k_f m(t)) - f_0 = -k_f m(t)$$

Pertanto si ha $\sigma_{f_d}^2 = k_f^2 \sigma_m^2 = k_f^2 \mathcal{P}_m$ se $m(t)$ è a media nulla: praticamente, σ_{f_d} rappresenta la deviazione standard della frequenza istantanea, e per questo è una grandezza rappresentativa della larghezza di banda del segnale modulato. Infine, il rapporto $\frac{\sigma_{f_d}^2}{W^2} = \beta^2$ è posto pari all'indice di modulazione perché appunto rappresenta una misura di quanto sia variata l'occupazione di banda del segnale modulato, rispetto alla massima frequenza W presente nel segnale modulante.

Discussione del risultato $SNR = 3\beta^2 SNR_0$.

Notiamo innanzitutto che se $\beta < \frac{1}{\sqrt{3}} = 0,57$ non si ha miglioramento, anzi si peggiora. Ma con bassi indici di modulazione abbiamo già visto che FM ha un comportamento che può avvicinarsi a quello lineare dell'AM, e dunque ci possiamo non-sorprendere. D'altra parte, SNR può migliorare (e di molto) con $\beta > \frac{1}{\sqrt{3}}$: ad esempio, se $\beta = 5 \Rightarrow 3\beta^2 = 75$ volte meglio, ovvero 17,75 dB di miglioramento! In compenso, la regola di Carson ci dice che la banda occupata aumenta di circa $2(\beta + 1) = 12$ volte quella di banda base... dunque il miglioramento di SNR avviene *a spese dell'occupazione di banda*. (La cosa non preoccupa per le trasmissioni via satellite, in quanto c'è riuso di frequenze in diversità di spazio).

Potrebbe ora sembrare che si possa aumentare indefinitamente β (nei limiti della banda disponibile) per migliorare a piacere l' SNR . In realtà non è così, dato che ad un certo punto l'analisi effettuata perde validità. Questo accade perché se β è troppo elevato, occorre che la banda di rumore del ricevitore sia più ampia (essendo aumentata la banda del segnale modulato) e perciò non si verifica più che $\mathcal{P}_x = \frac{A^2}{2} \gg \sigma_{\nu_c}^2 = \sigma_{\nu_s}^2 = N_0 B_N$. Le conseguenze di questo fatto verranno illustrate tra breve.

Esercizio

Sia dato un trasmettitore FM con potenza trasmessa 1 Watt e segnale modulante $m(t)$ con banda $\pm B = \pm 10$ MHz. Un collegamento con attenuazione disponibile $A_d = 100$ dB lo interfaccia ad un ricevitore con temperatura di sistema $T_{ei} = 2900$ °K. Volendo un $SNR = 40$ dB, calcolare:

- 1) Il fattore di rumore del ricevitore in dB;
- 2) Il minimo valore dell'indice di modulazione e la banda occupata a radiofrequenza B_{RF} ;
- 3) Se il valore di β non sia troppo grande, e quale sia il suo massimo valore;
- 4) Il nuovo valore β' , se si richiede di fornire al collegamento un margine di 25 dB.

Soluzione

- 1) Questa domanda va affrontata dopo lo studio del capitolo sul rumore termico, dove è mostrato che $T_{ei} = T_0(F - 1) + T_G = T_0 F$ se $T_G = T_0$; assumiamo quest'ipotesi per vera e dunque $F = \frac{T_{ei}}{T_0} = 10$; pertanto $F_{dB} = 10$ dB. Per proseguire l'esercizio con le nozioni fin qui acquisite, esplicitiamo che $\mathcal{P}_n(f) = \frac{N_0}{2} = \frac{1}{2} k T_{ei} = \frac{1}{2} \cdot 1.38 \cdot 10^{-23} \cdot 2900 \simeq \frac{1}{2} \cdot 4 \cdot 10^{-20}$ Watt/Hz.
- 2) $SNR = 3\beta^2 SNR_0 = 3\beta^2 \frac{W_R}{N_0 B} = 3\beta^2 \frac{W_T G_d}{N_0 B}$; il valore numerico di SNR risulta $10^{\frac{SNR_{dB}}{10}} = 10^4$, mentre quello di A_d è $10^{\frac{A_d(dB)}{10}} = 10^{10}$ e quindi $G_d = 1/A_d = 10^{-10}$. Sostituendo i valori, ed invertendo la relazione, si ottiene $\beta_{min} = \sqrt{\frac{SNR \cdot N_0 B}{3 \cdot W_T G_d}} = \sqrt{\frac{10^4 \cdot 4 \cdot 10^{-20} \cdot 10^7}{3 \cdot 10^{-10}}} = 3.65$. Applicando al regola di Carson per la banda: $B_{RF} \simeq 2B \cdot (\beta + 1) = 2 \cdot 10^7 \cdot 4.65 = 9.3 \cdot 10^7 = 93$ MHz.
- 3) Perché l'analisi svolta abbia valore, deve risultare $W_R \gg \sigma_{n_c}^2 = \sigma_{n_s}^2 = N_0 B_N = N_0 B_{RF} = 4 \cdot 10^{-20} \cdot 9.3 \cdot 10^7 = 3.72 \cdot 10^{-12}$ Watt, ma poichè $W_R = \frac{W_T}{A_d} = \frac{1}{10^{10}} = 10^{-10}$, si ha $\frac{W_R}{\sigma_{n_c}^2} = \frac{10^{-10}}{3.72 \cdot 10^{-12}} = 26$. Il valore di 26 soddisfa quindi pienamente l'esigenza di *grande segnale*. Per trovare β_{Max} , scriviamo allora $\beta_{Max} \Rightarrow W_R = 10 \cdot \sigma_{n_c}^2 = 10 \cdot N_0 \cdot B_{RF} = 10 \cdot N_0 \cdot 2B \cdot (\beta_{Max} + 1)$, e dunque $\beta_{Max} = \frac{W_R}{10 \cdot N_0 \cdot 2B} - 1 = \frac{10^{-10}}{8 \cdot 10^{-12}} - 1 = 12.5 - 1 = 11.5$, a cui corrisponde una banda $B_{RF} = 2B \cdot (\beta_{Max} + 1) = 2 \cdot 10^7 \cdot 12.5 = 250$ MHz, ed un guadagno di $SNR = 10 \lg_{10} 3\beta_{Max}^2 \simeq 26$ dB, mentre con β nominale si sarebbe ottenuto $10 \lg_{10} (3 \cdot 3.65^2) = 16$ dB.
- 4) Un margine di 25 dB equivale a far fronte ad una attenuazione supplementare $A'_d = 10^{2.5} = 316$ volte. Proviamo ad ottenere lo stesso SNR con un nuovo valore β' : $SNR = 10^4 = 3\beta'^2 \frac{W_T G_d G'_d}{N_0 B} = 3\beta^2 \frac{W_T G_d}{N_0 B} \frac{\beta'^2}{\beta^2} G'_d$; dunque deve risultare $\frac{\beta'^2}{\beta^2} G'_d = 1$ e quindi $\beta'^2 = \beta^2 \sqrt{\frac{1}{G'_d}} =$

$3.65\sqrt{316} = 3.65 \cdot 17.7 = 64.88 \dots$ non ce la facciamo. Infatti, al più (con $\beta = \beta_{Max} = 11.5$) si ha un margine di 10 dB.

10.3.3 Caso di elevato rumore

Qualora il valore efficace del rumore in ingresso al discriminatore sia confrontabile con quello del segnale utile ricevuto, si verifica un effetto soglia all'aumentare del rumore, e l' SNR degrada molto rapidamente.

Riprendendo lo schema che mostra l'involuppo complesso della portante non modulata A , del rumore in ingresso $\underline{v}(t)$, e del segnale ricevuto $\underline{r}(t)$, notiamo che se i valori efficaci dei primi due sono comparabili, può verificarsi il caso che $\underline{r}(t)$ ruoti attorno all'origine.

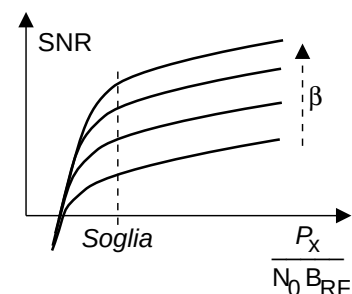
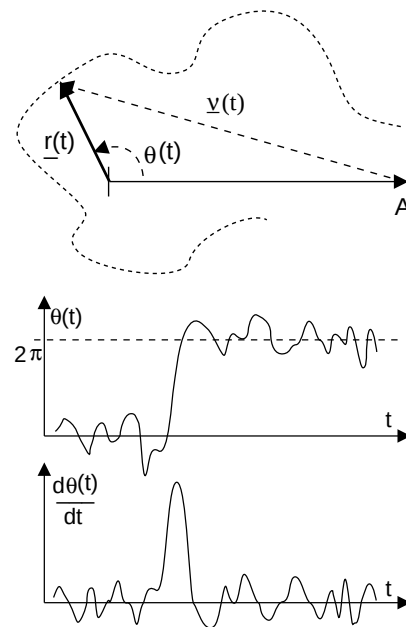
Quando ciò si verifica, a valle del derivatore che è presente nel discriminatore si determina un *click*, ovvero un segnale impulsivo di area pari a 2π , come illustrato nella figura seguente. Questo fatto è facilmente verificabile, ascoltando una radio FM broadcast, che in condizioni di cattiva ricezione manifesta la comparsa di un rumore, appunto, impulsivo.

All'aumentare della potenza di rumore, aumenta la frequenza con la quale $\underline{r}(t)$ "aggira" l'origine, e pertanto aumenta la frequenza dei *click*, che tendono a produrre un crepitio indistinto. Si è trovato che questo effetto si manifesta a partire da un SNR di ingresso pari a 10 dB, e per SNR peggiori di tale valore l'effetto aumenta molto rapidamente, cosicché si parla di effetto soglia.

Nel grafico che segue è riportato un tipico andamento dell' SNR dopo demodulazione, con l'indice di modulazione β che svolge il ruolo di parametro, e possiamo osservare come con un SNR di ingresso inferiore alla soglia, le prestazioni degradino rapidamente.

Si è trovato che demodulando con un PLL, anziché con un discriminatore, la soglia si riduce a circa 7 dB.

Nella pratica comune il segnale di rumore può essere costituito da una interferenza dovuta ad una emittente adiacente (ossia con una portante prossima alla nostra) che *sovramodula*, ovvero adotta un indice di modulazione troppo elevato, ed invade la banda delle emittenti contigue.



10.3.4 Enfasi e de-enfasi

Abbiamo osservato che, in presenza di rumore bianco in ingresso, il rumore dopo demodulazione ha un andamento parabolico. Questo comporta che, se il messaggio modulante $m(t)$ avesse un $\mathcal{P}_m(f)$ a sua volta bianco, l' $SNR(f)$ (alle frequenze più elevate) sarebbe molto peggiore, rispetto al suo valore per frequenze inferiori. Nella pratica, si possono verificare (ad esempio) i seguenti problemi:

- 1) Nelle trasmissioni FDM (in cui piú canali telefonici vengono modulati BLU, multiplati in frequenza e trasmessi come FM a basso indice) i canali agli estremi della banda sono piú rumorosi;
- 2) nell'FM commerciale, il segnale modulante è molto piú ricco di energia alle basse frequenze, dunque il problema del rumore elevato in alta frequenza è aggravato dal “basso segnale”.

Il rimedio a tutto ciò consiste nel modificare $m(t)$, in modo che anch'esso presenti uno spettro “parabolico”, e poi aggiungere una rete di de-enfasi in ricezione (praticamente un integratore, ovvero un passa-basso) tale da ripristinare l'originale sagoma spettrale del segnale e rendere la densità di potenza del rumore nuovamente uniforme.

Con un po di riflessione, ci si accorge che l'uso di una coppia enfasi-deenfasi equivale ad effettuare una trasmissione PM! In realtà, la rete di enfasi non è un derivatore perfetto (altrimenti annullerebbe le componenti del segnale a frequenza prossima allo zero), ed esalta le frequenze solo se queste sono maggiori di un valore minimo. Pertanto, si realizza un metodo di modulazione “misto”, FM in bassa frequenza e PM a frequenze (di messaggio) piú elevate.

Capitolo 11

Modulazione numerica

Sono qui discusse le tecniche adottate qualora il segnale da trasmettere in forma modulata sia numerico (ovvero binario). Il contesto applicativo può variare su un insieme di casi molto ampio, come ad esempio la trasmissione in ponte radio numerico di un flusso informativo ottenuto per moltiplicazione temporale di più sorgenti campionate, le forme di broadcast numerico (radiofonico o televisivo, terrestre o satellitare); in particolare può essere citato il caso dei modem, per mezzo dei quali un computer effettua trasmissione dati mediante un normale canale telefonico che, come illustrato in precedenza, è di tipo passabanda e quindi incapace di trasportare codici di linea di banda base.

11.1 Modulazione di Ampiezza e di Frequenza

11.1.1 BPSK

E' l'acronimo di *Bi-Phase Shift Keying*¹, e individua una tecnica per il trasporto dell'informazione basata sull'utilizzo di 2 possibili fasi per la portante:

$$x_{BPSK}(t) = a \sin(\omega_0 t + \varphi(t)) \quad \text{dove} \quad \varphi(t) = \sum_{k=-\infty}^{\infty} \varphi_k \text{rect}_{T_b}(t - kT_b)$$

con i valori φ_k pari a $\pm \frac{\pi}{2}$ per rappresentare le cifre binarie 0 ed 1. Sebbene, nei termini posti, questa soluzione appaia come una modulazione di fase, è facile mostrare come possa essere realizzata mediante una comune modulazione di ampiezza BLD.

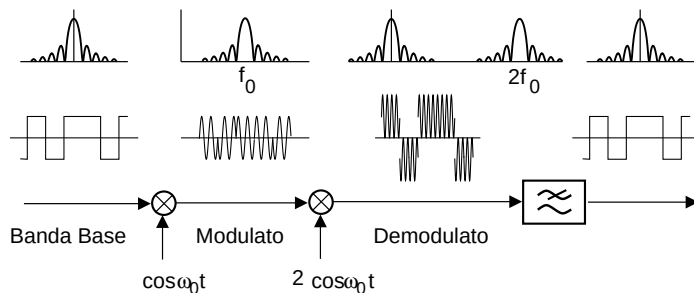
Se definiamo infatti un segnale $m(t)$ come un codice di linea NRZ bipolare, che assume valori ± 1 in corrispondenza delle cifre binarie 0 ed 1, il segnale

$$x_{BPSK}(t) = m(t) \cos \omega_0 t$$

è equivalente al precedente, e la sua mo-demodulazione coerente avviene come rappresentato in figura. Il segnale uscente dal moltiplicatore di demodulazione ha espressione

$$y(t) = x(t) \cdot 2 \cos \omega_0 t = 2m(t) \cdot \cos^2 \omega_0 t = m(t) + m(t) \cdot \cos 2\omega_0 t$$

¹Letteralmente, *slittamento di tasto a due fasi*.



e dunque il codice di linea $m(t)$ può essere riottenuto mediante filtraggio passa-basso.

Una buona caratteristica di questa tecnica è l'andamento costante dell'ampiezza della portante, che permette di utilizzare la massima potenza del trasmettitore, appena inferiore al valore che inizia a produrre fenomeni di distorsione. L'aspetto negativo è l'elevata occupazione di banda, dovuta all'uso di forme d'onda rettangolari in $m(t)$ che causano una riduzione sensibile della massima frequenza binaria, nel caso di trasmissione su canali con limitazioni di banda. Per questi motivi, il metodo è applicabile principalmente al caso di comunicazioni satellitari, in cui la potenza di trasmissione è limitata e la banda no, essendo quest'ultima condivisa a divisione di spazio, in quanto le antenne in ricezione sono fortemente direttive, e vedono un solo satellite.

Le caratteristiche ora illustrate possono essere "sovertite" qualora il segnale $m(t)$ sia generato utilizzando forme d'onda con occupazione di banda ridotta, come ad esempio la famiglia a coseno rialzato. In tal caso la banda occupata a frequenze positive risulta pari a $f_b(1 + \gamma)$, doppia rispetto al caso di banda base, a causa della modulazione BLD. In tal caso, l'ampiezza del segnale modulato non è più costante: mentre in corrispondenza degli istanti di Nyquist questa assume esattamente il valore del segnale dati corrispondente, nell'intervallo tra due istanti di Nyquist assume valori prodotti dall'effetto della somma di tutte le code delle funzioni $g(t)$ relative ai simboli trasmessi.

11.1.2 L-ASK

Ci riferiamo ora al caso in cui si operi ancora una AM-BLD (da cui il termine *Amplitude Shift Keying* - ASK) a partire da un segnale $m(t)$ caratterizzato dall'uso di L diversi livelli. L'ampiezza del segnale modulato subisce dunque variazioni, mentre la banda occupata è ridotta di un fattore pari a $\log_2 L$, pari al numero di bit necessari ad individuare uno degli L livelli², e quindi pari al rapporto tra frequenza binaria f_b e frequenza di segnalazione $f_s = \frac{f_b}{\log_2 L}$. Di nuovo, nel caso in cui $m(t)$ sia generato mediante forme d'onda con trasformata a coseno rialzato, la banda occupata a frequenze positive risulta pertanto pari a

$$B = f_s(1 + \gamma) = \frac{f_b}{\log_2 L}(1 + \gamma)$$

Se consideriamo $\gamma = 0$, possiamo definire:

Efficienza Spettrale ρ (o *densità di informazione*) come il rapporto tra la frequenza binaria e la banda occupata

$$\rho = \frac{f_b}{B} = \log_2 L$$

Questa si esprime in bit/sec/Hz e rappresenta appunto quanti bit/sec sono trasmessi per ogni Hz utilizzato.

Il valore trovato $\rho = \log_2 L$, è l'efficienza spettrale dell'L-ASK quando utilizza impulsi di Nyquist a banda minima; per altre forme di modulazione e/o di impulsi si ottengono altri valori, ed il loro confronto esprime la bontà del metodo rispetto all'utilizzo della banda a disposizione. Ad esempio, se confrontiamo il risultato ottenuto ora con quello

²Ad esempio: se $L = 32$ livelli, la banda si riduce di 5 volte, ed infatti con $M = 5$ bit si individuano $L = 2^M = 32$ configurazioni. Dato che il numero M di bit/simbolo deve risultare un intero, si ottiene che i valori validi di L sono le potenze di 2.

relativo ad una trasmissione numerica di banda base, notiamo un peggioramento di un fattore 2, dovuto all'uso di una AM-BLD. Sebbene a prima vista questo possa annullato adottando una AM-BLU, vedremo tra breve che si preferisce seguire approcci diversi, come ad esempio PSK e QAM.

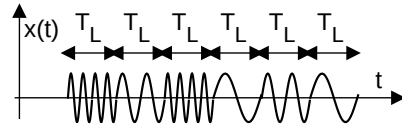
11.1.3 L-FSK

Nel caso in cui si richieda una ampiezza del segnale modulato rigidamente costante, si può adottare l'FSK (*Frequency Shift Keying*), che produce per il segnale modulato l'espressione

$$x_{FSK}(t) = \cos[2\pi(f_0 + m(t))t] \quad \text{dove} \quad m(t) = \Delta \cdot \sum_{k=-\infty}^{\infty} f_k \cdot \text{rect}_{T_L}(t - kT_L)$$

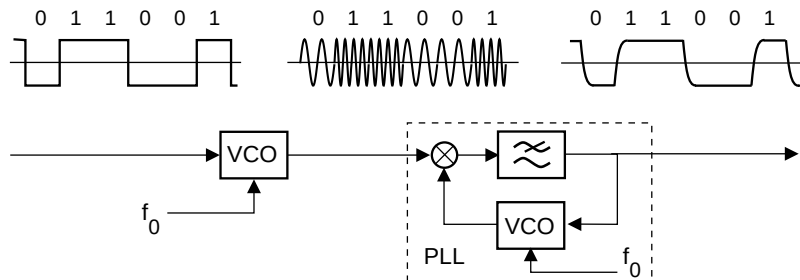
con $f_k \in \{0, 1, 2, \dots, L-1\}$.

Si tratta in pratica di una portante la cui frequenza nominale f_0 è alterata di una quantità $\Delta \cdot f_k$ Hz per un intervallo di T_L secondi, pari al periodo di simbolo. Δ rappresenta quindi la spaziatura (uniforme) tra le frequenze, ognuna rappresentativa di uno degli L livelli. Pertanto l'espressione può essere riscritta come



$$x_{FSK}(t) = \sum_{k=-\infty}^{\infty} \cos[2\pi(f_0 + \Delta f_k)t] \cdot \text{rect}_{T_s}(t - kT_s)$$

Il risultato è senza dubbio ad ampiezza costante; se $T_s \gg \frac{1}{f_0}$ si può adottare uno schema di mo-demodulazione basato su di un PLL (vedi pagg. 9.2.1.3 e 9.3.1.1) riportato (per $L = 2$) in figura, in cui all'uscita del passa basso ritroviamo il segnale modulante.



Questo schema è effettivamente utilizzato per modem a bassa velocità e basso costo, ed ha il pregio di funzionare anche in presenza di errori tra l' f_0 usata al trasmettitore e quella al ricevitore. Per raggiungere velocità più elevate occorre ridurre T_s (ovvero aumentare $f_s = \frac{f_b}{\log_2 L}$ che, a parità di L , consente un aumento di f_b).

FSK ortogonale Nel caso in cui si realizzi la condizione $\Delta = \frac{n}{2T_s}$ con n intero, le diverse frequenze $f_0 + \Delta f_k$ sono ortogonali³, e può essere adottato un *demodulatore a correlazione*, mostrato appresso e discusso alla nota⁴.

Nel caso di *modulazione coerente*, la banda occupata può essere approssimata⁵ come

$$B \simeq \frac{L}{2T_s}$$

(considerando L elevato e dunque $B \gg \Delta$), pertanto l'efficienza spettrale risulta

$$\rho = \frac{f_b}{B} = \frac{f_L \log_2 L}{L/2T_s} = \frac{2 \log_2 L}{L}$$

ossia $\frac{L}{2}$ volte peggiore dell' L-ASK. Nell'Appendice 11.5.2 è riportato un approfondimento dell'analisi relativa all'FSK ortogonale.

Ma: se l'efficienza spettrale è così bassa, che vantaggi ci sono ad usare l'FSK? ... a sua difesa, portiamo i seguenti argomenti:

Il caso semplice (con $T_s \gg \frac{1}{f_0}$) è di facile realizzazione e poco costoso; ad esempio, veniva usato per salvare su *compact cassette* audio i dati degli *home computer* degli anni '70⁶

Se $L=2$ l'efficienza spettrale è uguale all'ASK⁷.

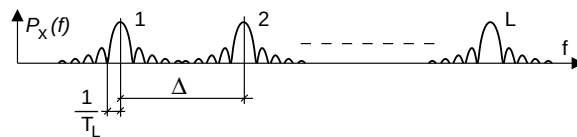
³L'ortogonalità tra le forme d'onda associate ai diversi simboli è sinonimo di intercorrelazione nulla, ovvero $\int_0^{T_s} \cos[2\pi(f_0 + m\Delta)t] \cos[2\pi(f_0 + n\Delta)t] dt = \begin{cases} \frac{1}{2} T_s & \text{se } n = m \\ 0 & \text{altrimenti} \end{cases}$. Si può dimostrare (vedi appendice 11.5.2) che una spaziatura $\Delta = \frac{1}{2T_s}$ garantisce l'ortogonalità solo nel caso in cui tra le forme d'onda non sussistano ritardi di fase, così come sopra espresso. Se invece i diversi simboli presentano una fase aleatoria ϕ_k , ossia hanno espressione $\cos[2\pi(f_0 + \Delta f_k)t + \phi_k]$ con ϕ_k casuale e diversa per $\forall k$, allora si ottengono segnali incorrelati solo adottando una spaziatura doppia, e cioè $\Delta = \frac{1}{T_s}$.

Questo secondo caso è detto di *modulazione incoerente*, per distinguerlo da quello in cui $\phi = 0$ detto *coerente*.

⁴Ognuno dei correlatori del banco esegue l'integrale indicato alla nota precedente, integrando su T_L il prodotto tra il segnale ricevuto e tutte le possibili frequenze $f_0 + m\Delta$ con $m \in \{0, 1, 2, \dots, L-1\}$. Se le frequenze sono ortogonali, al termine dell'intervallo di integrazione una sola delle uscite sarà diversa da zero. Il confronto tra i risultati indicato in figura è necessario, perché la presenza di rumore additivo *corrompe* l'ortogonalità tra simboli.

Nel caso di modulazione coerente, sia il trasmettitore che il ricevitore devono rispettare specifiche realizzative più stringenti, dovendo necessariamente realizzarsi un errore di fase nullo tra le frequenze di confronto ed il segnale ricevuto.

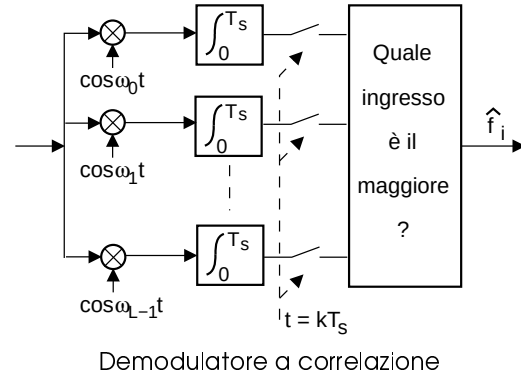
⁵In generale, se ogni diversa f_k è equiprobabile, l'FSK ha una densità spettrale del tipo:



Se $L \gg 1$, le L diverse frequenze occupano una banda (circa) uguale a $L \cdot \Delta$; qualora $\Delta = \frac{1}{2T_s} = \frac{f_s}{2}$, la banda risulta $L \cdot \frac{f_s}{2} = \frac{L f_s}{2}$.

⁶tipo: Sinclair Spectrum, Commodore Vic20 e 64 ... Come noto, le cassette audio soffrono di variazioni di velocità di trascinamento del nastro (*wow & flutter*), ma il PLL non ne risente.

⁷Tranne che, essendo ora presenti solo 2 frequenze, l'approssimazione $B \simeq \frac{L}{2T_s}$ non è più valida.

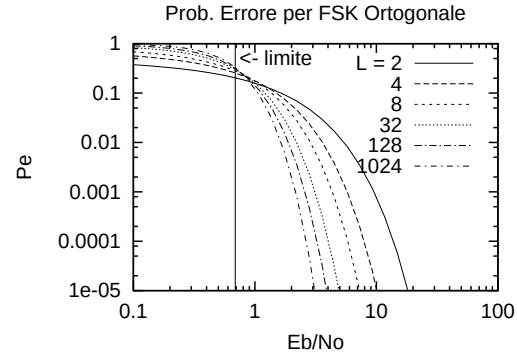


La probabilità di errore può essere resa piccola a piacere nei limiti della teoria dell'informazione⁸. Il grafico seguente mostra i valori di E_b/N_0 necessari per ottenere le varie P_e con diversi valori di L , e mostra come, all'aumentare di L , sia necessaria sempre meno potenza per ottenere la stessa P_e , a patto che risulti $E_b/N_0 > 1/\log_2 e = 0,69$. Questo valore è noto come *limite di Shannon-Hartley*⁹

Il risultato evidenziato merita qualche considerazione ulteriore: osserviamo infatti che la banda occupata

$$B \simeq \frac{L}{2T_L} = \frac{L}{2\frac{1}{f_L}} = \frac{L}{2\frac{\log_2 L}{f_b}} = \frac{f_b}{2} \frac{L}{\log_2 L}$$

aumenta (a parità di f_b) all'aumentare di L . Pertanto, per un E_b/N_0 assegnato (ovvero con f_b , potenza di segnale, e potenza di rumore preassegnate), l'FSK permette di ottenere P_e arbitrariamente piccole, a spese di una occupazione di banda sempre maggiore. L'aumento di L non può però essere qualunque, oltre che per le limitazioni del canale, anche a causa della complessità del ricevitore!



11.1.4 Prestazioni di L-ASK

Allo scopo di permettere un confronto, esaminiamo la probabilità di errore di un ASK al variare del numero di livelli, in funzione di E_b/N_0 . Abbiamo osservato come l'ASK possa essere ottenuto per modulazione AM BLD di un segnale dati di banda base, ed al Cap. 10 si è mostrato che in questo caso $SNR = SNR_0$; la probabilità di errore *per simbolo* è pertanto (nella nota¹⁰ una dimostrazione alternativa) identica¹¹ a quella ricavata in

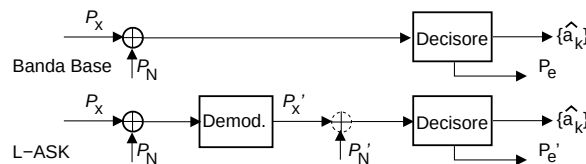
⁸Per l'esattezza, l'espressione teorica della probabilità di errore per simbolo risulta essere

$$P_e = 1 - \frac{1}{\sqrt{\pi L}} \int_{-\infty}^{\infty} e^{-z^2} \left(\int_{-\infty}^{z + \sqrt{\log_2 L \cdot E_b/N_0}} e^{-y^2} dy \right)^{L-1} dz$$

che deve essere valutata numericamente per ricavare le curve mostrate.

⁹In appendice è esposta una motivazione informale del comportamento descritto.

¹⁰Con riferimento alla figura seguente, il calcolo della P_e per l'L-ASK si imposta definendo valori di E_b ed N_0 equivalenti a quelli di banda base, ma ottenuti dopo demodulazione, e cioè $E'_b = P'_x T_b$ e $N'_0 = P'_N/W$ (infatti, $P'_N = \frac{N'_0}{2} 2W$, con $W = \frac{f_s}{2} = \frac{f_b}{2 \log_2 L}$).



L'equivalenza è presto fatta, una volta tarato il demodulatore in modo che produca in uscita la componente in fase $x_c(t)$ limitata in banda tra $\pm W$. Infatti in tal caso $P'_x = P_{x_c} = k_a^2 P_M = 2P_x$ e quindi $E'_b = P'_x T_s = 2P_x T_s = 2E_b$; per il rumore si ottiene $N'_0 = \frac{P'_N}{W}$ in cui $P'_N = P_{n_c} = \sigma_{n_c}^2 = \sigma_n^2 = \frac{N_0}{2} 4W$ e quindi $N'_0 = 2N_0$. Pertanto, il valore E'_b/N'_0 su cui si basa ora il decisore è lo stesso E_b/N_0 in ingresso al demodulatore.

¹¹Le curve di $P_e \left(\frac{E_b}{N_0} \right)$ sul bit sono quindi quelle di pag. 98, dove si tiene anche conto dell'uso di un codice di Gray per associare i livelli a configurazioni binarie.

§ 5.5.6 per il caso di banda base:

$$P_e = \left(1 - \frac{1}{L}\right) \operatorname{erfc} \left\{ \sqrt{3 \frac{E_b}{N_0} \frac{\log_2 L}{L^2 - 1}} \right\}$$

in cui ci si riferisce ad un segnale a coseno rialzato con $\gamma = 0$. Osserviamo inoltre che, se $L = 2$, si ottiene

$$P_e = \frac{1}{2} \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0}} \right\}$$

che rappresenta le prestazioni ottenibili per modulazione BPSK.

Per completare il confronto con l'FSK, osserviamo che ora, all'aumentare di L , la banda che (per $\gamma = 0$) risulta $B = f_s = \frac{f_b}{\log_2 L}$ si riduce, mentre la P_e aumenta: un comportamento diametralmente opposto all'FSK, e che può tornare utile in presenza di canali con limitazioni di banda ma non di potenza; quest'ultima può infatti essere aumentata per compensare il peggioramento di prestazioni legato all'uso di molti livelli e di una banda ridotta.

Natura di E_b/N_0 L'analisi delle prestazioni di questo capitolo sono tutte riferite alla grandezza $\frac{E_b}{N_0}$, che in questo caso rappresenta l'equivalente del rapporto segnale rumore convenzionale $SNR_0 = \frac{P_x}{N_0 W}$. Infatti, considerando che la potenza ricevuta può essere espressa come $\mathcal{P}_x = \frac{E_b}{T_b} = E_b \cdot f_b$, si ottiene

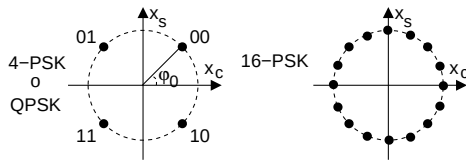
$$\frac{E_b}{N_0} = \frac{\mathcal{P}_x}{N_0 f_b}$$

che dipende unicamente da grandezze che identificano la qualità del collegamento ($\mathcal{P}_x$ ed N_0) e la natura del messaggio (f_b) senza alcun riferimento al metodo di modulazione.

11.2 Modulazione di fase

11.2.1 QPSK ed L-PSK

A completamento del quadro delle modulazioni “canoniche” consideriamo il caso in cui l'informazione modulata sia impressa, anziché sulle ampiezze o sulla frequenza, direttamente sulla fase del segnale modulato, in modo tale da ottenere



$$x_{L-PSK}(t) = a \cos(\omega_0 t + \varphi(t))$$

e quindi un inviluppo complesso

$$\underline{x}_{L-PSK}(t) = a [\cos \varphi(t) + j \sin \varphi(t)]$$

in cui

$$\varphi(t) = \sum_{k=-\infty}^{\infty} \varphi_k \operatorname{rect}_{T_s}(t - kT_s) \quad \text{e} \quad \varphi_k \in \{\varphi_0, \varphi_1, \dots, \varphi_{L-1}\}$$

La generica fase $\varphi_i = \frac{\pi}{4} + i \cdot \frac{2\pi}{L}$ individua quindi uno dei punti rappresentati alla figura precedente, e l'insieme di tutte le possibili fasi (punti) è chiamata *costellazione*. Nel caso in cui $\varphi(t)$ è costituita da rettangoli, $\varphi(t)$ stesso è un codice di linea NRZ; l'espressione di $\underline{x}_{L-PSK}(t)$ in termini di $\{x_c, x_s\}$ evidenzia inoltre come il risultato sia ottenibile mediante una modulazione AM in fase e quadratura (che *non* è una BLU dato che $x_s \neq \hat{x}_c$).

L'uso di un codice NRZ

per $\varphi(t)$, e quindi per x_c

ed x_s , produce una occupa-

zione di banda elevata per $\underline{x}_{L-PSK}(t)$, la

cui distribuzione di po-

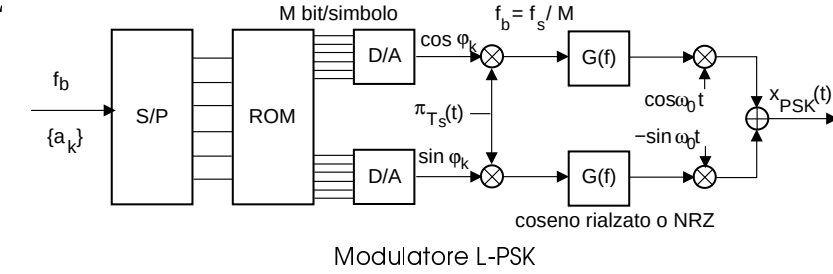
tenza assume una sago-

ma $\frac{\sin x}{x}$ centrata in f_0 e

con lobo principale di esten-

sione pari ad f_s . L'occupazione di banda può essere limitata a $B = f_s (1 + \gamma)$ se si realizza

$\underline{x}_{L-PSK}(t)$ mediante impulsi di Nyquist $g(t)$ a coseno rialzato, ottenendo



$$\underline{x}_{L-PSK}(t) = a \sum_{k=-\infty}^{\infty} e^{j\varphi_k} \cdot g(t - kT_s)$$

Mentre però, con $g(t)$ rettangolare, il modulo $|\underline{x}|$ giace su di un cerchio, con $g(t)$ a coseno rialzato $\underline{x}(t)$ passa dai punti della costellazione solo negli istanti significativi, mentre nell'intervallo tra due istanti segue traiettorie con ampiezza variabile. In questo caso la scelta tra NRZ e coseno rialzato dipende dalla necessità di limitare la dinamica delle ampiezze, oppure l'estensione della banda.

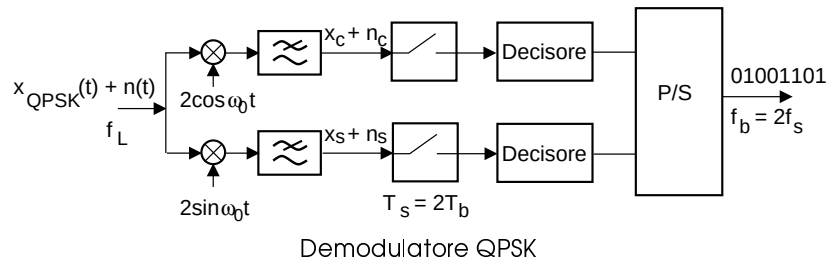
Da un punto di vista dell'efficienza spettrale, questa risulta identica a quanto ottenuto per l'ASK con ugual numero di livelli, in quanto si ha la medesima frequenza di simbolo $f_s = \frac{f_b}{\log_2 L}$ e dunque $\rho = \frac{f_b}{B} = \frac{f_s \log_2 L}{f_s} = \log_2 L$ (per coseno rialzato con $\gamma = 0$). In figura è rappresentata una possibile architettura per il modulatore.

Dal punto di vista delle prestazioni, occorre distinguere il caso in cui $L = 4$ (indicato come QPSK = *Quadrature Phase Shift Keying*) da quello con L generico, in quanto sussistono due diverse architetture per il demodulatore.

11.2.2 Prestazioni QPSK

In questo caso (PSK con 4 livelli) il demodulatore è costituito da due rami indipendenti in fase e quadratura, operanti a frequenza di simbolo f_s metà di quella binaria.

Ognuno dei due rami effettua una decisione per uno dei due bit che compongono il simbolo, e le due decisioni vengono re-serializzate, come mostrato nella figura a lato. Entrambi i



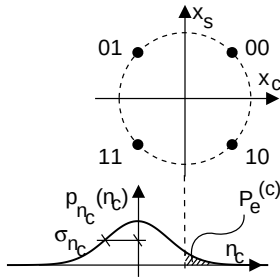
rami si comportano pertanto come un demodulatore L-ASK con $L=2$, ovvero (a parte una

rotazione di fase) un BPSK, e la probabilità di errore¹² di ogni singolo ramo risulta (per $\gamma = 0$):

$$P_e^{(c)} = P_e^{(s)} = \left(1 - \frac{1}{L}\right) \operatorname{erfc} \left\{ \sqrt{3 \frac{E_b}{N_0} \frac{\log_2 L}{(L^2 - 1)}} \right\} \Big|_{L=2} = \frac{1}{2} \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0}} \right\}$$

mentre la probabilità di errore in un bit della sequenza re-serializzata risulta¹³

$$P_{e_{QPSK}} = P_e^{(c)} \cdot P^{(c)} + P_e^{(s)} \cdot P^{(s)} = \frac{1}{2} (P_e^{(c)} + P_e^{(s)}) = P_e^{(c)} = P_e^{(s)} = \frac{1}{2} \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0}} \right\}$$

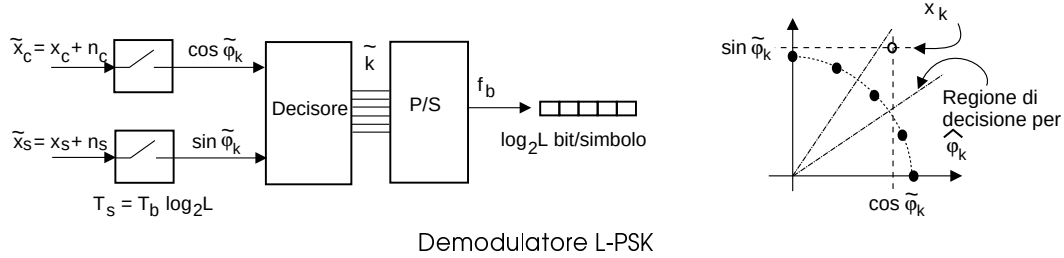


in cui $P^{(c)}$ e $P^{(s)}$ sono le probabilità che il bit ricevuto provenga dal ramo in fase o da quello in quadratura, e si assume che $P_e^{(c)} \cdot P_e^{(s)} \ll P_e^{(c)}$ e quindi trascurabile.

Osserviamo quindi come il QPSK consenta di ottenere le stesse prestazioni del BPSK, utilizzando solo metà banda: $B = f_s = \frac{f_b}{2}$ (con $\gamma = 0$). Il risultato ha una giustificazione intuitiva: osserviamo infatti che, dimezzando la banda, si dimezza anche la varianza del rumore gaussiano in ingresso al demodulatore, e questo fatto compensa la riduzione di ampiezza delle componenti analogiche di bassa frequenza ricevute nel caso QPSK.

11.2.3 Prestazioni L-PSK

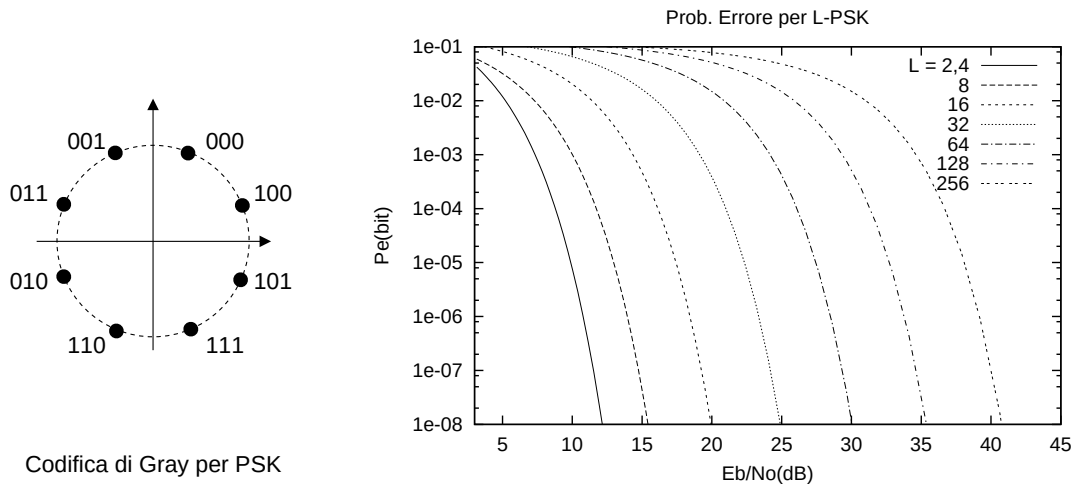
In questo caso il demodulatore ha una differente architettura, ed il decisore opera congiuntamente su entrambi i rami, per ottenere la stima del gruppo di $\log_2 L$ bit associati ad una delle possibili fasi φ_k .



La decisione avviene calcolando $\tilde{\varphi}_k = \arctan \frac{\sin \tilde{\varphi}_k}{\cos \tilde{\varphi}_k}$ e stabilendo all'interno di quale regione di decisione $\hat{\varphi}_k$ cada la fase ricevuta $\tilde{\varphi}_k$. All'aumentare del numero di livelli L , la potenza di rumore (che concorre alla probabilità di errore) diminuisce con la stessa legge di riduzione della banda, ovvero con il $\log_2 L$. Al contrario, la spaziatura tra le regioni di decisione diminuisce con legge lineare rispetto ad L ; pertanto, l'aumento del numero di livelli produce un peggioramento della P_e . Senza approfondire i relativi conti, forniamo direttamente il risultato (con $\gamma = 0$) della probabilità di errore sul simbolo,

¹²Per ottenere questo risultato occorre dimostrare che ogni ramo opera una demodulazione BPSK con $\frac{E'_b}{N'_0} = \frac{E_b}{N_0}$. Il calcolo di $N'_0 = 2N_0$ è identico a quello svolto nella nota 10; mostriamo che $E'_b = 2E_b$. Per ogni ramo infatti si ottiene $E'_b = P_{x_{c(s)}} T_s = \frac{1}{2} \cdot 2P_x \cdot 2T_s = 2P_x T_b = 2E_b$ in quanto $P_{x_{c(s)}} = \frac{1}{2} \cdot 2P_x$, dove $\frac{1}{2}$ tiene conto che $x_{c(s)}$ ha una ampiezza ridotta di $\frac{1}{\sqrt{2}}$ rispetto al BPSK, ed il fattore 2 tiene conto che ogni ramo opera una modulazione AM-BLD; infine, $T_s = 2T_b$ in quanto i bit di ognuno dei due rami transitano a metà della velocità di quelli in uscita.

¹³La probabilità di errore per simbolo risulta invece $P_e(\text{simbolo}) = P_e^{(c)} + P_e^{(s)} = \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0}} \right\}$, considerando di nuovo trascurabile la probabilità di un errore contemporaneo su entrambi i rami.



$$P_{eL-PSK} = \text{erfc} \left\{ \sin \left(\frac{\pi}{L} \right) \sqrt{\frac{E_b}{N_0} \log_2 L} \right\}$$

che rappresenta la probabilità di decidere di aver ricevuto un $\hat{\varphi}_k \neq \varphi_k$ (diverso da quello trasmesso) e che, se $P_e \ll 1$, è approssimata con la probabilità di invadere (a causa del rumore) una regione di decisione contigua.

Confrontando il risultato con quello per l'ASK, osserviamo che l'assenza del termine $(1 - \frac{1}{L})$ è dovuto alla "circolarità" della costellazione, il termine $\sin(\frac{\pi}{L})$ è un fattore che rappresenta il peggioramento all'aumentare di L , ed il $\log_2 L$ sotto radice è il miglioramento dovuto alla riduzione di banda. Il risultato esposto è una approssimazione valida se $P_e \ll 1$, e via via più accurata con L crescente.

In tabella è riportato il risultato del confronto, a parità di P_e , dei valori di $\frac{E_b}{N_0}$ necessari per L-PSK, contro quelli necessari per L-ASK: si è eseguito il rapporto tra gli argomenti degli $\text{erfc}\{\}$, si è elevato al quadrato, ed il risultato è stato espresso in dB.

L	$\Delta y^2 = \frac{1}{3} (L^2 - 1) \sin^2 \frac{\pi}{L}$	Δy_{dB}^2
4 (QPSK)	2.5	4
8	3.07	4.88
16	3.23	5.1
32	3.28	5.2
64	3.29	5.2

Esaminando la tabella per diversi valori di L , si trova (a parte il termine $(1 - \frac{1}{L})$ dell'L-ASK) il miglioramento di prestazioni in dB dell'L-PSK rispetto ad L-ASK, ovvero i dB di potenza risparmiata a parità di probabilità di errore. Il risultato (4-5 dB di miglioramento) ha portato a prediligere sempre il PSK rispetto all'ASK.

E' opportuno osservare che, qualora si desideri ottenere un valore di probabilità di errore *per bit*, questo è pari a

$$P_e(\text{bit}) = \frac{P_e(\text{simbolo})}{\log_2 L}$$

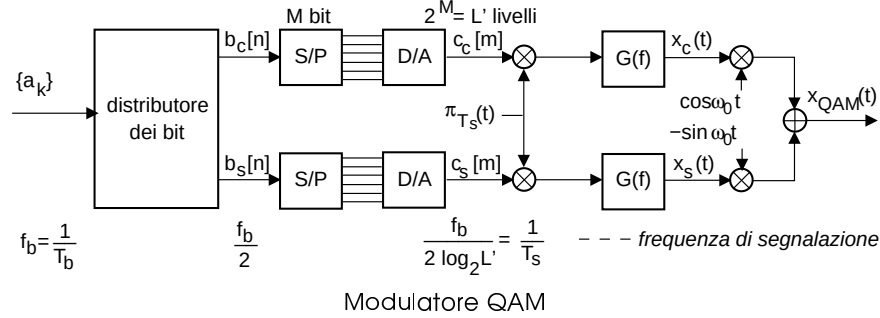
a patto di associare, a livelli contigui, gruppi di bit differenti in una sola posizione, come previsto dal codice di Gray (mostrato nella figura seguente), in modo che l'errore tra due livelli provochi l'errore di un solo bit nel gruppo di $\log_2 L$ bit associati a ciascun livello. Le curve di probabilità di errore per bit, riportate in figura, sono calcolate in questo modo.

11.3 QAM

Questo acronimo, che sta per "*Quadrature Amplitude Modulation*", individua una tecnica di modulazione che utilizza (come il PSK) due portanti in quadratura. La differenza rispetto al PSK è che ora le componenti di banda base x_c ed x_s , che danno luogo a

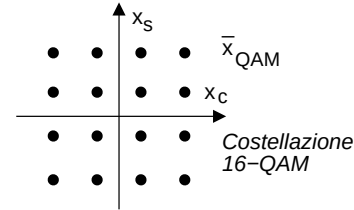
$$x_{QAM}(t) = x_c(t) \cos \omega_0 t - x_s(t) \sin \omega_0 t$$

non dipendono più da una stessa sequenza di fasi, ma sono originate da due flussi di dati distinti.



Nella pratica, questi due flussi di dati hanno origine da un medesimo messaggio numerico, i cui bit vengono distribuiti alternativamente sui due rami a frequenza metà. Questi vengono raggruppati per formare due segnali di banda base x_c ed x_s ad L' livelli, prodotti mediante un filtro formatore di impulsi $G(f)$ che determina l'occupazione spettrale del segnale x_{QAM} ottenuto a partire da x_c ed x_s mediante modulazione in fase e quadratura. Se $G(f)$ è a coseno rialzato con roll off γ , allora la banda a frequenze positive di x_{QAM} risulta pari a $\frac{f_b}{2 \log_2 L'} (1 + \gamma)$ Hz.

La sequenza di operazioni determina una costellazione quadrata, composta da $L = (L')^2$ punti, che rappresentano le coordinate, nel piano dell'involuppo complesso, in cui quest'ultimo è forzato a transitare in corrispondenza degli istanti di Nyquist multipli di T_s .



In definitiva, il periodo di simbolo risulta

$$T_s = \frac{2 \log_2 L'}{f_b} = \frac{2 \log_2 (L)^{1/2}}{f_b} = \frac{1}{f_b} \log_2 L$$

mentre la banda occupata è

$$B = f_s (1 + \gamma) = \frac{f_b}{\log_2 L} (1 + \gamma)$$

e quindi uguale al PSK con uguale numero di livelli (di cui condivide quindi anche l'efficienza spettrale). Notiamo che, per come si è impostato lo schema di distribuzione dei bit tra i rami, L deve risultare un quadrato perfetto. Nulla impedisce di elaborare schemi più complessi in cui L' è diverso per i due rami, oppure in cui la costellazione non sia quadrata.

11.3.1 Prestazioni di QAM

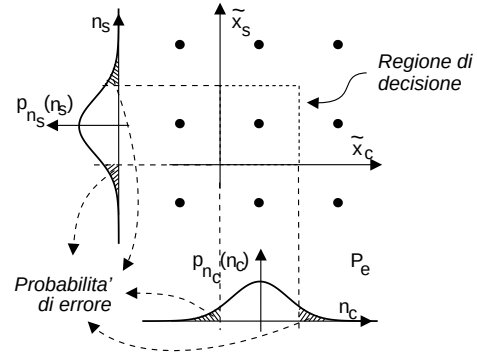
Notiamo subito che la “distanza” tra due punti della costellazione QAM è maggiore (a parità di L) del caso PSK; pertanto, c'è da aspettarsi un miglioramento delle prestazioni (a parità di E_b/N_0) in quanto l'area che individua la probabilità di errore è ridotta. Il valore della probabilità di errore si determina dopo aver osservato che ciascuno dei due rami in fase e quadratura costituisce un segnale ASK multilivello con $L' = \sqrt{L}$.

Abbiamo già calcolato che, per tale segnale, il rapporto E_b/N_0 dopo demodulazione è esattamente pari a quello del segnale modulato; pertanto otteniamo

$$\begin{aligned} P_\alpha &= P_e^c(\text{simbolo}) = P_e^s(\text{simbolo}) = \\ &= \left(1 - \frac{1}{L'}\right) \operatorname{erfc} \left\{ \sqrt{3 \frac{E_b}{N_0} \frac{\log_2 L'}{(L')^2 - 1}} \right\} \end{aligned}$$

Ricordando ora che $L' = \sqrt{L} = (L)^{1/2}$ e dunque $\log_2 L' = \frac{1}{2} \log_2 L$, si ottiene

$$P_\alpha = \left(1 - \frac{1}{\sqrt{L}}\right) \operatorname{erfc} \left\{ \sqrt{\frac{3}{2} \frac{E_b}{N_0} \frac{\log_2 L}{L - 1}} \right\}$$

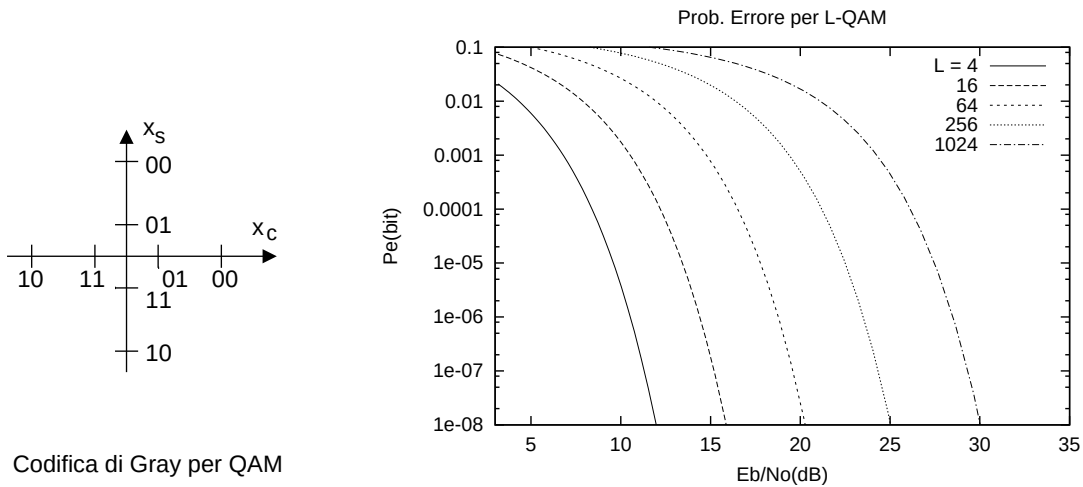


La probabilità di errore (a carattere) complessiva, cioè la probabilità che il segnale ricevuto $\tilde{x} = \underline{x} + \underline{n}$ cada fuori della regione di decisione relativa all' $\underline{x}$ trasmesso, risulta $P_e(\text{carattere}) = P_\alpha + P_\alpha + (P_\alpha)^2 \simeq 2P_\alpha$, in cui si assume trascurabile la probabilità di sbagliare entrambe x_c ed x_s .

Questa stessa ipotesi, assieme all'utilizzo di un codice di Gray per codificare i gruppi di bit associati a livelli dei due rami, consente di esprimere la probabilità di errore per bit come

$$P_e(\text{bit}) = \frac{P_e(\text{carattere})}{\log_2 L} = \frac{2}{\log_2 L} \left(1 - \frac{1}{\sqrt{L}}\right) \operatorname{erfc} \left\{ \sqrt{\frac{3}{2} \frac{E_b}{N_0} \frac{\log_2 L}{L - 1}} \right\}$$

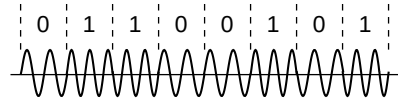
Di seguito, sono raffigurate le curve con i valori di $P_e(\text{bit})$ per diversi valori di L al variare di $\frac{E_b}{N_0}$ espresso in dB; il confronto con le curve relative al PSK permette di valutare l'entità del miglioramento di prestazioni. Come è evidente, la modulazione QAM offre prestazioni sensibilmente superiori rispetto alla PSK.



11.4 Altre possibilità

Sebbene il lettore possa pensare di averne già viste più di quante non si fosse mai aspettato, quelle illustrate sono solamente le tecniche di base. Senza ora entrare in dettagli troppo elaborati per essere riportati qui, menzioniamo alcune altre tecniche:

MSK (*Minimum Shift Keying*) simile all'FSK, ma con la variante di mantenere una continuità di fase tra simboli contigui. Questa caratteristica consente una riduzione della banda occupata, in virtù dell'assenza di brusche variazioni di fase.



Offset Keying Una variante del QAM e QPSK, in cui i periodi di simbolo per i 2 rami sono sfasati del 50%. La capacità di sincronizzazione del ricevitore risulta migliore.

Partial Response QAM Il segnale modulato è filtrato, e si introduce deliberatamente una ISI in modo controllato. Migliora l'efficienza spettrale.

Codifica e Detezione Differenziali Nel PSK sussiste una ambiguità rispetto a quale sia la fase di riferimento. Una soluzione è inviare un flusso binario costruito a partire dalle *differenze* di bit contigui nel messaggio originario, oppure demodulare la fase di ogni simbolo *relativamente* alla fase del simbolo precedente. L'operazione causa un peggioramento di prestazioni di qualche dB.

TCM (*Trellis Coded Modulation*). Trellis significa *traliccio*, e rappresenta un modo di realizzare una codifica di canale che impone vincoli alle possibili sequenze. Il numero di livelli è artificialmente aumentato, ma i punti della costellazione risultante non sono tutti possibili, anzi solo un ristretto numero lo è, in funzione dei simboli precedenti. Il risultato è un miglioramento della P_e (od una riduzione di E_b necessaria) a spese di una maggiore occupazione di banda.

OFDM (*Orthogonal Frequency Division Multiplex*). Simile sotto certi aspetti all'FDM, in quanto suddivide la banda in più portanti, che sono però ora attive contemporaneamente. Ogni portante effettua tipicamente una modulazione QAM (con più livelli), e la spaziatura tra portanti è scelta in modo da renderle ortogonali, annullando così le interferenze tra canali. La realizzazione si basa su *componenti hardware* che effettuano operazioni di FFT (Fast Fourier Transform) per sintetizzare il segnale e demodularlo. Il vantaggio principale è l'assenza di necessità di equalizzazione. Viene impiegato per ottenere velocità di trasmissione molto elevate su mezzi trasmissivi scarsamente condizionati, come nel caso dell'adsl su linea telefonica. Alla appendice 11.5.3 è esposta una analisi dettagliata della tecnica..

SPREAD SPECTRUM (*Modulazione ad espansione di spettro*). La stessa banda di frequenze è contemporaneamente utilizzata da più trasmissioni differenti, che non interferiscono tra loro perché ognuna utilizza forme d'onda ortogonali a quelle delle altre, e che sono caratterizzate da una occupazione spettrale *molto* superiore a quella minima. La tecnica di trasmissione risultante prende anche il nome di *Multiplicazione a Divisione di Codice*.

Esercizio Consideriamo un sistema di modulazione numerica PSK con 16 fasi, per il quale si riceve una potenza di segnale $\mathcal{P}_x = 10^{-3}$ (Volt)², in presenza di una densità di potenza di rumore $\mathcal{P}_N(f) = 2 \cdot 10^{-11}$ (Volt)²/Hz. Si desidera trasmettere un flusso numerico a velocità $f_b = 1$ Mbit/sec e si considerino impulsi a coseno rialzato con $\gamma = 0$.

- 1) Quale è la P_e per bit al ricevitore? E la banda occupata?
- 2) Quale nuovo valore di P_e si ottiene usando invece una modulazione QAM con lo stesso numero di punti di costellazione?
- 3) Nel caso 16-QAM, qualora si desideri ancora la P_e ottenuta al punto 1), quanta potenza è sufficiente ricevere?
- 4) Nel caso QAM con la P_e del punto 1), qualora si desideri dimezzare la banda occupata, quanta potenza è necessario ricevere?
- 5) Nel caso 16-QAM con la P_e del punto 1) e $\mathcal{P}_x = 10^{-3}$ (Volt)², quale nuova f_b è possibile raggiungere?

Soluzione

- 1) Osserviamo che $E_b = \mathcal{P}_x \cdot T_b = \frac{\mathcal{P}_x}{f_b} = \frac{10^{-3}}{10^6} = 10^{-9}$ (Volt)²/Hz, mentre $N_0 = 2\mathcal{P}_N(f) = 4 \cdot 10^{-11}$ (Volt)²/Hz, pertanto $\frac{E_b}{N_0} = 25$ e $\left(\frac{E_b}{N_0}\right)_{dB} = 10 \lg_{10} 25 \simeq 14$ dB.
 - Dalle curve delle prestazioni per il PSK si trova che con $E_b/N_0 = 14$ dB, si ottiene $P_e = 10^{-3}$ qualora si utilizzino 16 livelli.
 - La banda occupata risulta $B = \frac{f_b}{\log_2 L} = \frac{10^6}{4} = 250$ KHz.
- 2) Le curve delle prestazioni per il QAM mostrano che con $E_b/N_0 = 14$ dB e 16 livelli, si ottiene $P_e \simeq 3 \cdot 10^{-6}$.
- 3) le stesse curve mostrano che, con il 16-QAM, la $P_e = 10^{-3}$ si ottiene con $E_b/N_0 = 10.5$ dB, ovvero $14 - 10.5 = 3.5$ dB in meno, che corrispondono ad una potenza $\mathcal{P}'_x = \frac{\mathcal{P}_x}{10^{0.35}} = \frac{10^{-3}}{2.24} = 4.47 \cdot 10^{-5}$ (Volt)².
- 4) Dimezzare la banda equivale a raddoppiare $\log_2 L$, ovvero utilizzare un numero di livelli $L' = L^2 = 256$. Le curve delle prestazioni per il 256-QAM mostrano che per ottenere $P_e = 10^{-3}$ occorre $E_b/N_0 \simeq 18.3$ dB, pari ad un aumento di $18.3 - 14 = 4.3$ dB, che equivale ad una potenza $\mathcal{P}'_x = 10^{0.43} \mathcal{P}_x \simeq 2.7 \cdot 10^{-3}$ (Volt)².
- 5) Ci ritroviamo nelle stesse condizioni del punto 3), con un eccesso di 3.5 dB nel valore di E_b/N_0 , che può essere eliminato riducendo in ugual misura T_b , e quindi aumentando f_b . Risulta: $T'_b = \frac{T_b}{10^{0.35}}$ e quindi $f'_b = \frac{1}{T'_b} = \frac{10^{0.35}}{T_b} = 10^{0.35} \cdot f_b = 10^{0.35} \simeq 2.24$ Mb/sec.
 - E se $\gamma \neq 0$? La trattazione del caso di banda base, mostra che y^2 subisce un peggioramento di $(1 + \gamma) \left(1 - \frac{\gamma}{4}\right)$, che (per esempio) con $\gamma = 0.5$ fornisce 1.31, che deve essere compensato da una uguale diminuzione di E_b/N_0 . Nel caso 5), ad esempio, la f_b risulterà quindi limitata a $f''_b = f'_b / 1.31 = 1.71$ Mb/sec.

11.5 Appendici

11.5.1 Schema riassuntivo delle prestazioni

Metodo	M bit/simbolo	P_e (*) (codifica di Gray)	Banda RF	ρ (bit/sec/Hz)
BPSK	1	$\frac{1}{2} \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0}} \right\}$	$f_b (1 + \gamma)$	$\frac{1}{(1+\gamma)}$
L-ASK	$\log_2 L$	$\frac{1}{M} \left(1 - \frac{1}{L} \right) \operatorname{erfc} \left\{ \sqrt{3 \frac{E_b}{N_0} \frac{M}{(L^2-1)}} \right\}$	$f_b \frac{(1+\gamma)}{M}$	$\frac{M}{(1+\gamma)}$
L-FSK incoerente	$\log_2 L$		$f_b \frac{L}{M}$	$\frac{M}{L}$
L-FSK coerente	$\log_2 L$	nota 8 a pag. 241	$f_b \frac{L}{2M}$	$\frac{L}{2L}$
QPSK	2	$\frac{1}{2} \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0}} \right\}$	$f_b (1 + \gamma)$	$\frac{2}{(1+\gamma)}$
L-PSK	$\log_2 L$	$\frac{1}{M} \operatorname{erfc} \left\{ \sin \left(\frac{\pi}{L} \right) \sqrt{\frac{E_b}{N_0} M} \right\}$	$f_b \frac{(1+\gamma)}{M}$	$\frac{M}{(1+\gamma)}$
L-QAM	$\log_2 L$	$\frac{2}{M} \left(1 - \frac{1}{\sqrt{L}} \right) \operatorname{erfc} \left\{ \sqrt{\frac{3}{2} \frac{E_b}{N_0} \frac{M}{L-1}} \right\}$	$f_b \frac{(1+\gamma)}{M}$	$\frac{M}{(1+\gamma)}$

(*) Se $\gamma \neq 0$, si deve aggiungere un termine $(1 + \gamma) (1 - \frac{\gamma}{4})$ al denominatore sotto radice.

11.5.2 FSK Ortogonale

A pagina 239 è stata introdotta la modulazione FSK, e nelle note si è iniziata la discussione relativa alle condizioni di ortogonalità tra le frequenze di confronto ed il segnale ricevuto; prendiamo qui in considerazione segnali del tipo generale $\cos [2\pi (f_0 + \Delta f_k) t + \phi_k]$, in cui è incluso un errore di fase aleatorio tra simboli, in modo da esaminare le differenze tra il caso di modulazione coerente ed incoerente.

Iniziamo dunque con lo sviluppare l'espressione dell'integrale di intercorrelazione $\rho = \int_0^{T_s} \cos [2\pi (f_0 + m\Delta) t] \cos [2\pi (f_0 + n\Delta) t + \phi] dt$ facendo uso della relazione $\cos \alpha \cos \beta = \frac{1}{2} [\cos (\alpha + \beta) \cos (\alpha - \beta)]$ e riferendoci per semplicità al caso di due frequenze contigue (ponendo $m = 0$ ed $n = 1$):

$$\rho = \frac{1}{2} \int_0^{T_s} \{ \cos [2\pi (2f_0 + \Delta) t + \phi] + \cos [2\pi \Delta t - \phi] \} dt = \quad (11.1)$$

$$= \frac{1}{2} \int_0^{T_s} \cos [2\pi (2f_0 + \Delta) t + \phi] dt + \frac{1}{2} \int_0^{T_s} \cos [2\pi \Delta t - \phi] dt \quad (11.2)$$

Per quanto riguarda il primo integrale, esso assume un valore nullo se $2f_0 + \Delta = \frac{k}{T_s}$, perché in tal caso in un intervallo T_s entrano un numero intero di periodi, ed il coseno ha valor medio nullo. Concentriamoci allora sul valore di Δ che annulla anche il secondo integrale, che riscriviamo facendo uso della relazione $\cos (\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$:

$$\int_0^{T_s} \cos (2\pi \Delta t - \phi) dt = \int_0^{T_s} [\cos (2\pi \Delta t) \cos \phi + \sin (2\pi \Delta t) \sin \phi] dt = \quad (11.3)$$

$$= \frac{\sin (2\pi \Delta t)}{2\pi \Delta} \Big|_0^{T_s} \cdot \cos \phi - \frac{\cos (2\pi \Delta t)}{2\pi \Delta} \Big|_0^{T_s} \cdot \sin \phi = \quad (11.4)$$

$$= T_s \left[\frac{\sin (2\pi \Delta T_s)}{2\pi \Delta T_s} \cdot \cos \phi + \frac{1 - \cos (2\pi \Delta T_s)}{2\pi \Delta T_s} \cdot \sin \phi \right] \quad (11.5)$$

Osserviamo ora che, nel caso in cui $\phi = 0$, il secondo termine si annulla per qualunque Δ . Esaminiamo quindi ora solamente il primo termine, individuando così il risultato relativo al caso di

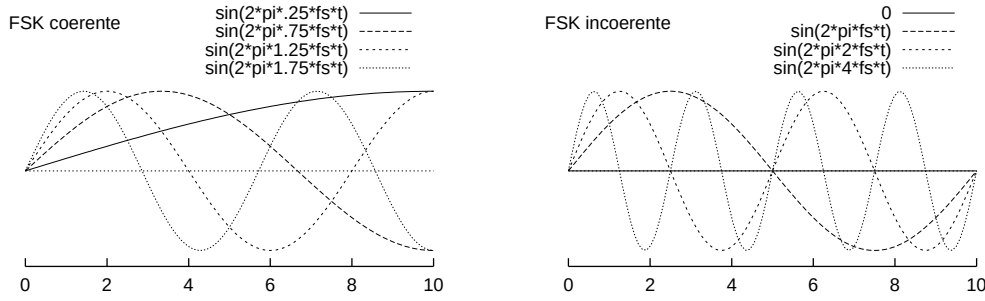


Figura 11.1: Forme d'onda ortogonali nei casi di modulazione *coerente* ed *incoerente*

Modulazione coerente Il termine $\frac{\sin(2\pi\Delta T_s)}{2\pi\Delta T_s}$ si annulla per $\Delta = \frac{k}{2T_s}$, e quindi la minima spaziatura tra portanti risulta $\Delta = \frac{1}{2T_s} = \frac{f_s}{2}$; pertanto, le frequenze utilizzate dovranno essere del tipo $f_0 + k\frac{f_s}{2}$.

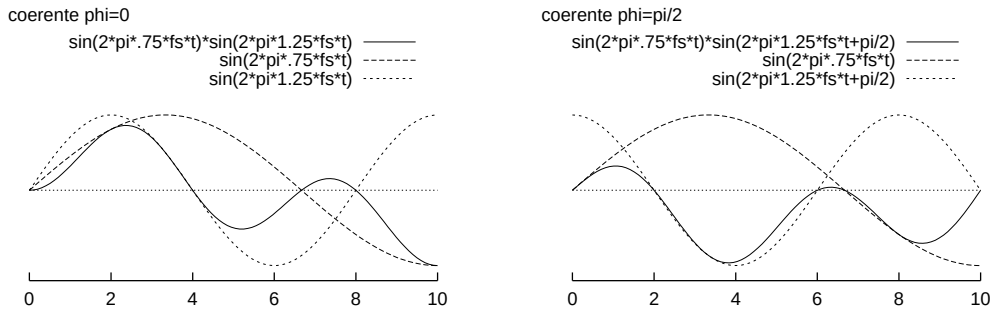
Per quanto riguarda il primo termine della 11.2, anche questo si annulla se sussiste la relazione $2f_0 + \Delta = 2f_0 + \frac{f_s}{2} = \frac{k}{T_s} = kf_s$, che fornisce la condizione $f_0 = f_s \frac{2k-1}{4}$, ossia f_0 può essere scelta come uno tra i valori $\frac{1}{4}f_s, \frac{3}{4}f_s, \frac{5}{4}f_s, \frac{7}{4}f_s, \dots$. Notiamo come la spaziatura $\frac{f_s}{2}$ tra i possibili valori per la portante, coincida con quella tra le frequenze di segnalazione. Pertanto la parte sinistra della figura 11.1 rappresenta, disegnate in un intervallo pari a T_s , sia le portanti che possono essere usate, sia le prime frequenze che è possibile adottare per un modulazione FSK *coerente* basata sul valore minimo di f_0 pari a $\frac{1}{4}f_s$.

Nel caso in cui f_0 non assuma uno dei valori individuati, il primo termine di 11.2 non si annulla, ma se $f_0 \gg \frac{1}{T_s}$, risulta trascurabile rispetto al secondo. Pertanto, nel caso di trasmissioni su canali di tipo passa-banda, la scelta di f_0 non è più determinante; d'altra parte, la spaziatura tra le frequenze di segnalazione pari a $\frac{f_s}{2}$ produce comunque il risultato che due frequenze di segnalazione contigue, accumulano in un intervallo T_s una differenza di fase di mezzo periodo.

Modulazione incoerente In questo caso si ha $\phi \neq 0$. In generale la 11.5 presenta entrambi i termini; mentre il primo (come ora esaminato) si annulla per $\Delta = \frac{k}{2T_s}$, il secondo invece è nullo solo se $\Delta = \frac{k}{T_s}$. Questa circostanza determina il risultato che occorre ora adottare una spaziatura tra portati doppia della precedente, e pari cioè a $\Delta = f_s$.

Tornando ad esaminare la 11.2, il suo primo termine si annulla ora se $2f_0 + \Delta = 2f_0 + f_s = kf_s$, che determina la condizione $f_0 = f_s \frac{k-1}{2}$, ossia $f_0 = 0, \frac{1}{2}f_s, f_s, \frac{3}{2}f_s, \dots$. Notiamo come la spaziatura $\frac{f_s}{2}$ tra i possibili valori per la portante sia identica al caso precedente, ma sia ora pari alla metà della spaziatura necessaria alle frequenze di segnalazione. La circostanza che sia adesso ammessa anche una portante a frequenza nulla consente di tracciare la parte destra della figura 11.1, che mostra le prime frequenze di segnalazione che è possibile adottare per una modulazione FSK *incoerente* basata sul valore minimo di $f_0 = 0$.

Verifica grafica La figura che segue mostra il risultato del prodotto di due frequenze ortogonali distanti $\frac{f_s}{2}$ e calcolate in assenza di errore di fase (a sinistra) e con un errore di fase pari a $\phi = \frac{\pi}{2}$. Si può notare come in questo secondo caso si perda l'ortogonalità tra i segnali, essendo il risultato prevalentemente negativo.



Discussione sull'ottimalità per $L \rightarrow \infty$ Osserviamo innanzitutto che il ricevitore a correlazione commette errore nel caso in cui il rumore sovrapposto al segnale di ingresso sia casualmente “simile” ad una delle cosinusoidi utilizzate per la trasmissione. In tal caso, l'uscita dell'integratore relativo alla frequenza “simile” può superare quella relativa alla frequenza trasmessa.

All'aumentare di L (per f_b fisso) aumenta il periodo di simbolo $T_s = \frac{\log_2 L}{f_b}$ e quindi diventa sempre più “difficile” per il rumore emulare “bene” una delle frequenze di segnalazione, e quindi si riduce la probabilità di errore.

Chiaramente, all'aumentare di L aumenta proporzionalmente la complessità del ricevitore, che deve disporre di un numero di correlatori crescente. Pertanto, le prestazioni ideali per L che tende ad infinito rivestono solamente un interesse teorico.

11.5.3 OFDM

La sigla sta per ORTHOGONAL FREQUENCY DIVISION MULTIPLEX, ossia *multiplazione a divisione di frequenza ortogonale*. Si tratta della tecnica di modulazione numerica adottata per le trasmissioni ADSL¹⁴, ovvero la possibilità di utilizzare la linea telefonica per realizzare un collegamento numerico ad alta velocità, occupando una regione di frequenze più elevata di quella in cui avviene la trasmissione del segnale vocale¹⁵. Il termine *Asymmetric* si riferisce alla circostanza che è si realizza una velocità elevata in ricezione e ridotta in trasmissione.

La trasmissione OFDM ha la particolarità di utilizzare in modo ottimo la banda del canale, e di ridurre l'operazione di equalizzazione ad un prodotto tra vettori.

11.5.3.1 Rappresentazione nel tempo ed in frequenza

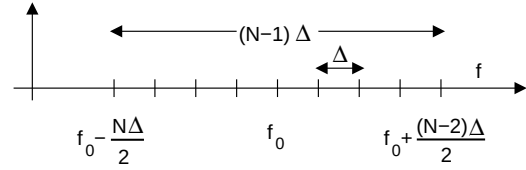
La sigla OFDM sintetizza che si tratta di una trasmissione a divisione di frequenza, in cui i diversi canali adottano forme d'onda ortogonali.

¹⁴ADSL = *Asymmetric Digital Subscriber Loop*, dove il *Subscriber Loop* rappresenta il circuito di utente che si realizza tra apparecchio e centrale quando si solleva il telefono.

¹⁵Per questo è a volte indicata come OVER VOICE.

Concettualmente possiamo pensare l'OFDM come una evoluzione¹⁶ della modulazione FSK, in cui però tutte le diverse frequenze $f_n = f_0 + \Delta \cdot \left(n - \frac{N}{2}\right)$, con $n = 0, 1, \dots, N-1$, sono utilizzate contemporaneamente, ed ognuna realizza una modulazione numerica anche a più

livelli (es. QPSK o QAM) con impulso NRZ rettangolare. Indicando ora con $\underline{a}_n^k = a_{n_c}^k + ja_{n_s}^k$ le coordinate nel piano dell'involuppo complesso di un generico punto della costellazione realizzata per la portante f_n all'istante $t = kT$, il segnale OFDM può essere scritto come



$$x_{OFDM}(t) = \sum_k \text{rect}_T(t - kT) \sum_{n=0}^{N-1} \left(a_{n_c}^k \cos \omega_n(t - kT) - a_{n_s}^k \sin \omega_n(t - kT) \right) \quad (11.6)$$

$$= \sum_{k=-\infty}^{\infty} \delta(t - kT) * \left(\text{rect}_T(t) \sum_{n=0}^{N-1} \left(a_{n_c}^k \cos \omega_n t - a_{n_s}^k \sin \omega_n t \right) \right) \quad (11.7)$$

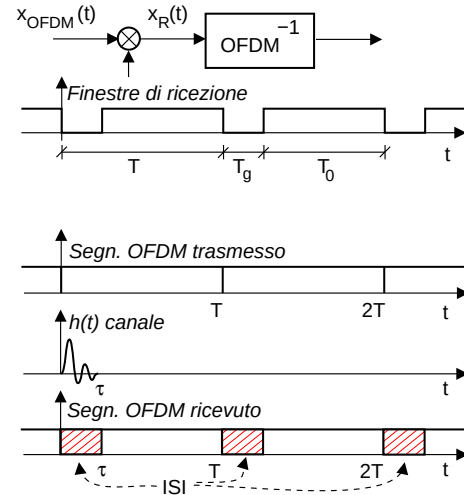
in cui la prima sommatoria (su k) identifica gli istanti di simbolo, e la seconda (su n) le diverse portanti.

E' facile osservare¹⁷ che tale segnale presenta un involuppo complesso pari a

$$\underline{x}_{OFDM}(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT) * \left(\text{rect}_T(t) \sum_{n=0}^{N-1} \underline{a}_n^k e^{j2\pi[\Delta(n - \frac{N}{2})]t} \right)$$

L'espressione (11.7) non vincola la durata T di un simbolo ad un valore particolare; deve però risultare $T \geq T_0 = \frac{1}{\Delta}$, in quanto il ricevitore opera sul segnale una finestrata temporale di estensione $T_0 = \frac{1}{\Delta}$ allo scopo di rendere ortogonali tra loro le frequenze $f_n = f_0 + \Delta \cdot \left(n - \frac{N}{2}\right)$, e mettere in grado il ricevitore di calcolare i valori $\underline{a}_n^k$ per tutti gli n presenti all'istante $t = kT$, mediante un ricevitore a correlazione.

L'intervallo T_0 è detto *periodo principale* del simbolo OFDM, mentre la differenza $T_g = T - T_0$ è indicata come *tempo di guardia*, od anche *preambolo*, ed il segnale ricevuto durante T_g non è usato in



¹⁶La trasmissione numerica contemporanea su più portanti prende il nome di *Multi Carrier Modulation* (MCM). La modulazione FSK utilizza invece una portante alla volta, in quanto la sua definizione prevede la presenza di un solo oscillatore.

¹⁷Se

$$x(t) = \cos \omega_1 t = \frac{1}{2} (e^{j\omega_1 t} + e^{-j\omega_1 t})$$

allora $x^+(t) = \frac{1}{2} e^{j\omega_1 t}$, e l'involuppo complesso $\underline{x}(t)$ calcolato rispetto ad f_0 risulta

$$\underline{x}(t) = 2x^+(t) e^{-j\omega_0 t} = 2 \frac{1}{2} e^{j\omega_1 t} e^{-j\omega_0 t} = e^{j(\omega_1 - \omega_0)t}$$

ricezione. Il motivo di tale “spreco”¹⁸ risiede nel fatto che, in presenza di un canale non perfetto, la parte iniziale di ogni simbolo risulta corrotta (vedi figura) da una interferenza intersimbolica (ISI) dovuta al risultato della convoluzione tra la coda del simbolo precedente e l' $h(t)$ del canale.

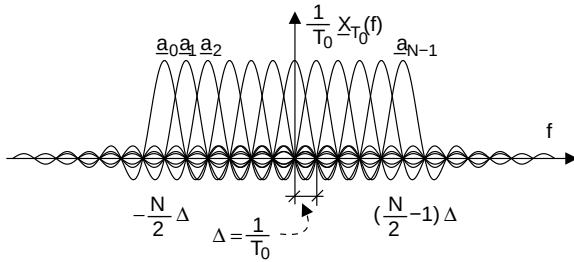
Consideriamo ora un solo simbolo (fissiamo $k = 0$ e consideriamo l'origine dei tempi ritardata di T_g) ricevuto nell'intervallo $T_0 = \frac{1}{\Delta} \leq T$, con inviluppo complesso

$$\underline{x}_{T_0}(t) = \text{rect}_{T_0}(t) \cdot \sum_{n=0}^{N-1} \underline{a}_n e^{j2\pi[\Delta(n - \frac{N}{2})]t} \quad (11.8)$$

e calcoliamone la trasformata per determinare l'occupazione di banda:

$$\underline{X}_{T_0}(f) = T_0 \text{sinc}(fT_0) * \sum_{n=0}^{N-1} \underline{a}_n \delta\left(f - \Delta\left(n - \frac{N}{2}\right)\right) = \quad (11.9)$$

$$= T_0 \sum_{n=0}^{N-1} \underline{a}_n \text{sinc}\left(\left(f - \Delta\left(n - \frac{N}{2}\right)\right)T_0\right) \quad (11.10)$$



Otteniamo quindi il risultato mostrato in figura, dove si evidenzia come ogni funzione *sinc* risulti moltiplicata per uno dei coefficienti $\underline{a}_n$, che potrebbero quindi essere ri-ottenuti in ricezione campionando (in modo complesso) $\underline{X}(f)$ a passo Δ .

Dalla (11.10) si ottiene la densità di potenza $\mathcal{P}_{\underline{x}_R}(f)$ dell'inviluppo complesso $\underline{x}_R(t)$ ricevuto e finestrato, di cui $\underline{X}_{T_0}(f)$ rappresenta la trasformata di un generico periodo principale, dopo aver specificato il numero di bit M_n e la potenza $\mathcal{P}_n$ assegnate alla portante n -esima, vincolate a fornire

$$\sum_{n=0}^{N-1} M_n = M \quad \text{e} \quad \sum_{n=0}^{N-1} \mathcal{P}_n = \mathcal{P}$$

Essendo le portanti ortogonali nel periodo T_0 , possiamo applicare la relazione $\mathcal{P}(f) = \sigma_a^2 \frac{\mathcal{E}(f)}{T}$ (vedi § 7.2.2) alle singole componenti e sommare i contributi. Nel caso in cui la sequenza $\{\underline{a}_n\}$ sia a valori indipendenti ed uniformemente distribuiti su di una costellazione QAM quadrata di lato $2\sqrt{3\mathcal{P}_n \frac{\sqrt{L_n}-1}{\sqrt{L_n}+1}}$ (vedi nota¹⁹) in cui $L_n = 2^{M_n}$, si ottiene che $\sigma_{\underline{a}_n}^2 = E\{\underline{a}_n^2\} = 2\mathcal{P}_n$, permettendo di scrivere la densità potenza di $\underline{x}_R(t)$ in ingresso al demodulatore come

$$\mathcal{P}_{\underline{x}_R}(f) = \frac{1}{T} \sum_{n=0}^{N-1} 2\mathcal{P}_n T_0^2 \text{sinc}^2\left(\left(f - \Delta\left(n - \frac{N}{2}\right)\right)T_0\right)$$

¹⁸Infatti la frequenza di simbolo $f_L = \frac{1}{T} = \frac{1}{T_0 + T_g}$ risulta ridotta rispetto al caso in cui T_g sia nullo.

¹⁹Al § 7.9.3 si è mostrato che se gli a_n sono v.a. indipendenti e distribuite uniformemente su L' livelli tra $\pm A$, si ottiene $\sigma_a^2 = \frac{A^2}{3} \frac{L'+1}{L'-1}$. Nel caso di una costellazione QAM quadrata ad L livelli si ha $L' = \sqrt{L}$, e se i rami in fase e quadratura sono indipendenti risulta $\sigma_{\underline{a}_n}^2 = E\{(a_{n_c} + ja_{n_s})^2\} = 2\sigma_a^2 = \frac{2A^2}{3} \frac{\sqrt{L}+1}{\sqrt{L}-1}$; volendo eguagliare tale valore a $2\mathcal{P}_n$, occorre quindi scegliere $A = \sqrt{3\mathcal{P}_n \frac{\sqrt{L}-1}{\sqrt{L}+1}}$.

a cui corrisponde una potenza complessiva²⁰ pari a

$$\mathcal{P}_{x_R} = \frac{1}{T} \sum_{n=0}^{N-1} 2\mathcal{P}_n T_0 = 2 \frac{T_0}{T} \sum_{n=0}^{N-1} \mathcal{P}_n$$

Infine, la potenza totale di $x_r(t)$ risulta

$$\mathcal{P}_{x_R} = \mathcal{P}_{x_R}^+ + \mathcal{P}_{x_R}^- = 2 \frac{1}{4} \mathcal{P}_{x_R} = \frac{T_0}{T} \sum_{n=0}^{N-1} \mathcal{P}_n$$

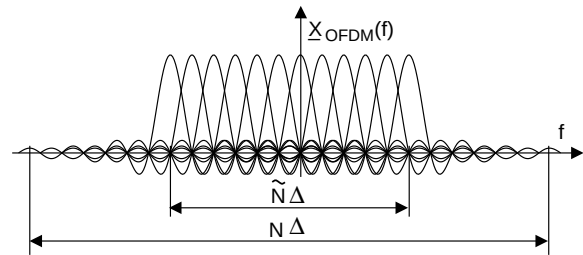
in cui è evidenziata la perdita di potenza legata alla presenza del preambolo.

11.5.3.2 Architettura di modulazione

Una caratteristica fondamentale della modulazione OFDM è quella di essere realizzata senza *oscillatori e integratori*, ma completamente tramite circuiti digitali.

Con riferimento alla figura 11.2, il flusso binario a frequenza f_b viene parallelizzato per formare simboli ad $L = 2^M$ livelli a frequenza $f_s = \frac{f_b}{M} = \frac{f_b}{\log_2 L}$. Questi M bit/simbolo sono suddivisi in $\tilde{N}$ gruppi di M_n ($n = 0, 1, \dots, \tilde{N} - 1$) bit ciascuno, con $M = \sum_{n=0}^{\tilde{N}-1} M_n$. Ogni gruppo di M_n bit produce un punto di costellazione $\underline{a}_n$ scelto tra $L_n = 2^{M_n}$ punti possibili.

La sequenza $\{\underline{a}_n\}$ viene “arricchita” con $N - \tilde{N}$ valori nulli (metà all’inizio e metà alla fine) in modo da ottenere una nuova sequenza $\{\underline{a}_n\}$ di N valori, in modo che la sommatoria di (11.10) dia luogo ad un inviluppo complesso praticamente limitato in banda (vedi figura) tra (circa) $\pm N \cdot \Delta$ Hz, che può essere pertanto rappresentato dai suoi campioni presi a frequenza $f_c = N \cdot \Delta \frac{\text{campioni}}{\text{secondo}}$. Il blocco indicato come FFT^{-1}



²⁰Si è fatto uso del risultato $\int_{-\infty}^{\infty} T_0^2 \text{sinc}^2(fT_0) df = T_0$.

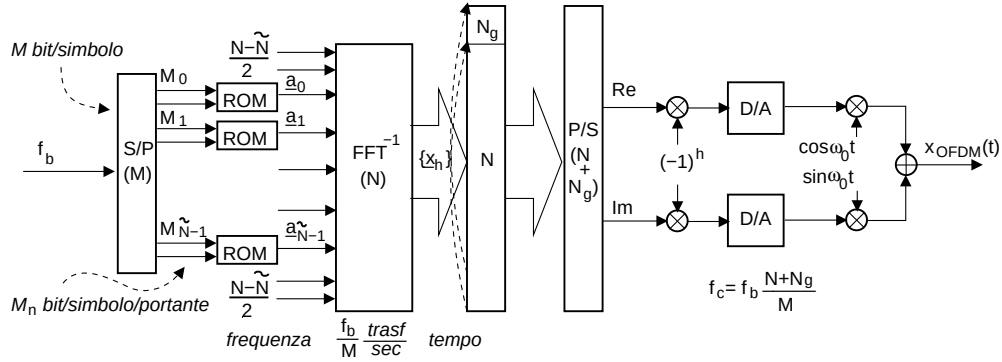


Figura 11.2: Architettura di un modulatore OFDM numerico

svolge proprio questa operazione: esso infatti esegue efficientemente²¹ il calcolo

$$\sum_{n=0}^{N-1} a_n e^{j2\pi \frac{n}{N} h} = \frac{1}{(-1)^h} x_{T_0}(hT_c) \quad (11.11)$$

Il risultato della FFT^{-1} è quindi una sequenza di coefficienti complessi $\{x_h\}$, che a meno di un segno alterno sono uguali ai campioni dell'involuppo complesso $x_{T_0}(t)$ fornito dalla (11.8) relativo ad un simbolo. Il preambolo da trasmettere durante il tempo di guardia T_g si ottiene “aggiungendo” in testa a $\{x_h\}$ un gruppo di campioni prelevati dalla coda²².

Infine, le parti reale ed immaginaria di $\{x_h\}$ sono inviate ad una coppia di convertitori

²¹La (11.11) è in qualche modo simile alla formula di ricostruzione per il segnale periodico limitato in banda $\pm \frac{N}{2}F$

$$x(t) = \sum_{m=-N/2}^{N/2} X_m e^{j2\pi m F t}$$

che calcolata per $t = hT_c = \frac{h}{NF}$ fornisce $x(hT_c) = \sum_{m=-N/2}^{N/2} X_m e^{j2\pi \frac{m}{N} h}$. Ponendo ora $n = m + \frac{N}{2}$ e $Y_n = X_{n-\frac{N}{2}}$ otteniamo

$$x(hT_c) = \sum_{n=0}^{N-1} Y_n e^{j2\pi \frac{n-\frac{N}{2}}{N} h} = e^{-j\pi h} \sum_{n=0}^{N-1} Y_n e^{j2\pi \frac{n}{N} h}$$

Osservando ora che $e^{-j\pi h} = (-1)^h$ e confrontando con la (11.8) si ottiene la (11.11). La coppia di relazioni

$$X_n = \frac{1}{N} \sum_{h=0}^{N-1} x_h e^{-j2\pi \frac{h}{N} n} \quad \text{e} \quad x_h = \sum_{n=0}^{N-1} X_n e^{j2\pi \frac{n}{N} h}$$

sono chiamate *Discrete Fourier Transform* (DFT) diretta e inversa, in quanto costituiscono la versione discreta della trasformata di Fourier (vedi § 3.9.4), e consentono il calcolo di una serie di campioni in frequenza a partire da campioni nel tempo e viceversa.

La FFT (*Fast Fourier Transform*) esegue le stesse operazioni, ma organizza i calcoli sfruttando le proprietà di periodicità degli esponenziali complessi, in modo da realizzare una mole di calcoli non superiori a $N \cdot \log_2 N$ per trasformate ad N punti. Questo risultato è possibile solamente se N è una potenza di 2, e quindi la modulazione OFDM opera necessariamente su $N = 2^H$ portanti, con H intero.

²²In effetti la (11.11) fornisce un risultato periodico rispetto ad h , con periodo N , ossia con periodo $N \cdot T_c = N \frac{1}{f_c} = N \frac{1}{\Delta N} = \frac{1}{\Delta} = T_0$ per la variabile temporale. Per questo motivo il *preambolo* dell'OFDM è detto anche *estensione ciclica*.

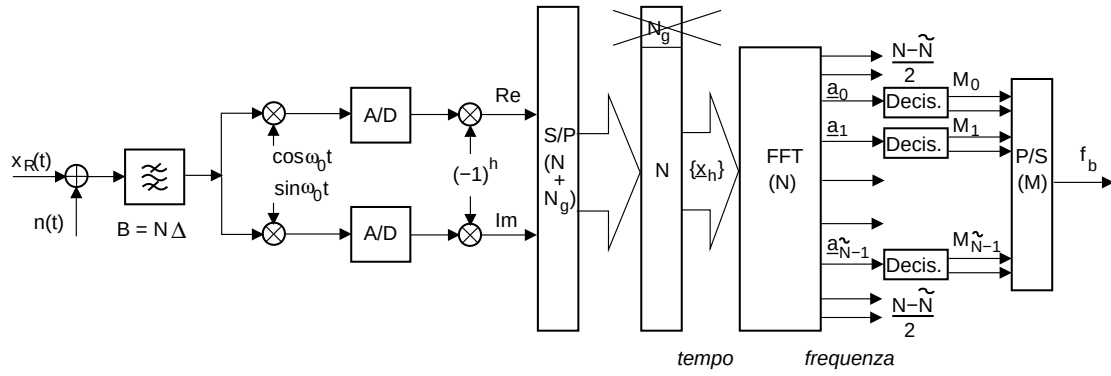


Figura 11.3: Architettura di un demodulatore OFDM numerico

D/A operanti a $f_c = \frac{N+N_g}{T}$ in modo da ottenere le C.A. di B.F., utilizzate per produrre il segnale $x_{OFDM}(t)$ mediante una coppia di modulatori in fase e quadratura.

11.5.3.3 Efficienza dell'OFDM

Come vedremo tra breve, questa è una tra le tecniche di modulazione che meglio approssima i risultati della teoria dell'informazione; tanto più quanto maggiore è la sua efficienza. Quest'ultima si ottiene considerando che solo $\tilde{N}$ portanti su N trasportano informazione, e che solo $f_c \cdot T_0$ campioni su $f_c \cdot T$ sono unici; combinando queste quantità si ottiene

$$\rho = \frac{\tilde{N} T_0}{N T} = \frac{\tilde{N}}{N} \frac{T - T_g}{T} = \frac{\tilde{N}}{N} \left(1 - \frac{T_g}{T} \right)$$

che misura la frazione di segnale utile rispetto all'occupazione di banda ed al numero di campioni/simbolo presenti in $x_{OFDM}(t)$. La ridondanza introdotta (le portanti vuote ed il preambolo) è della stessa natura di quella introdotta dal roll-off γ di un impulso a coseno rialzato, in quanto ha lo scopo di evitare che si verifichino fenomeni di interferenza tra simboli. Osserviamo che l'efficienza migliora all'aumentare di T e di N , dato che T_g ed $N - \tilde{N}$ sono fissi.

11.5.3.4 Architettura di demodulazione

Per ottenere gli elementi della sequenza $\{a_n\}$ e quindi il gruppo di M bit che hanno originato il simbolo, si adotta l'architettura mostrata in figura che svolge una azione del tutto inversa a quella del modulatore.

Innanzitutto il ricevitore deve acquisire il sincronismo di simbolo per determinare l'inizio della ricezione di un singolo blocco di campioni. Il segnale ricevuto viene quindi demodulato in fase e quadratura, e le C.A. di B.F. campionate a frequenza $f_c = \frac{N+N_g}{T}$. Dopo l'inversione di segno ad indici alterni, gli $f_c \cdot T_g = N_g$ campioni del preambolo sono rimossi, ed una FFT permette di ottenere i valori

$$\frac{1}{N} \sum_{h=0}^{N-1} x_h e^{-j2\pi \frac{h}{N} n} = \underline{x}_{T_0} \left(\left(n - \frac{N}{2} \right) \Delta \right) = a_n \quad (11.12)$$

Solo gli $\tilde{N}$ valori centrali sono avviati verso altrettanti decisori, che determinano il punto di costellazione più vicino all' $\underline{a}_n$ ricevuto per ogni portante, lo codificano con M_n bit, ed il risultato finale è nuovamente serializzato per produrre gli M bit che hanno dato origine al simbolo.

11.5.3.5 Prestazioni

Il calcolo della P_e si basa su quello relativo alle probabilità di errore P_{e_n} condizionato alle singole portanti. Dato che la portante n -esima trasporta M_n bit/simbolo, la probabilità che un bit generico provenga dalla portante n -esima risulta pari a $Pr(n) = \frac{M_n}{M}$ e quindi la probabilità che sia errato è pari a

$$P_e = \sum_{n=0}^{\tilde{N}-1} Pr(n) P_{e/n} = \frac{1}{M} \sum_{n=0}^{\tilde{N}-1} M_n P_{e_n} \quad (11.13)$$

Calcolo della P_e per portante La P_{e_n} dipende dal numero di livelli $L_n = 2^{M_n}$ scelto per la portante n -esima, e dal rapporto $\left(\frac{E_b}{N_0}\right)_n$ locale.

Per determinare il valore di P_{e_n} conviene applicare i risultati trovati al § 11.3.1 per la modulazione QAM, particolarizzati al caso attuale, in cui si adottano impulsi rettangolari di durata $T_0 = \frac{1}{\Delta}$. Attribuendo ai punti delle costellazioni gruppi di bit secondo la codifica di Gray, risulta

$$P_{e/n} = \frac{2}{\log_2 L_n} P_{\alpha_n} \quad \text{in cui} \quad P_{\alpha_n} = \left(1 - \frac{1}{\sqrt{L_n}}\right) \text{erfc} \left\{ \sqrt{\frac{3}{2} SN R_n \frac{1}{L_n - 1}} \right\} \quad (11.14)$$

è la probabilità di errore su di uno dei rami (in fase od in quadratura) della n -esima costellazione QAM, come ottenuta in § 5.5.3 per il caso di banda base²³.

Per il calcolo di

$$SN R_n = \frac{\mathcal{P}_{R_n}^c}{\mathcal{P}_{N_n}^c} = \frac{\mathcal{P}_{R_n}^s}{\mathcal{P}_{N_n}^s} = \frac{\frac{1}{2} \mathcal{P}_{R_n}}{\frac{1}{2} \mathcal{P}_{N_n}} = \frac{\mathcal{P}_{R_n}}{\mathcal{P}_{N_n}}$$

osserviamo che la potenza $\mathcal{P}_{R_n}$ dell'involuppo complesso del segnale ricevuto sulla portante n -esima, è pari a

$$\mathcal{P}_{R_n} = 2\mathcal{P}_{R_n} = 2 \frac{T_0}{T} \alpha_n \mathcal{P}$$

in cui $\mathcal{P}$ è la potenza totale ricevuta, e $\alpha_n = \frac{\mathcal{P}_n}{\mathcal{P}}$ è la frazione di potenza assegnata alla n -esima portante. Resta quindi da determinare $\mathcal{P}_{N_n}$.

Potenza di rumore per portante Per quanto riguarda $\mathcal{P}_{N_n}$, si tratta di applicare la (11.12) alla sequenza $\{(-1)^h \underline{n}(hT_c)\}$ dei campioni dell'involuppo complesso del rumore, e determinare il valore

$$\mathcal{P}_{N_n} = E \left\{ (\underline{N}_n)^2 \right\} = \sigma_{\underline{N}_n}^2 \quad \text{in cui} \quad \underline{N}_n = \frac{1}{N} \sum_{h=0}^{N-1} (-1)^h \underline{n}(hT_c) e^{-j2\pi \frac{h}{N} n}$$

²³Si consideri che il valore L presente al § 5.5.3 è pari alla radice di $L_n = 2^{M_n}$ della n -esima costellazione OFDM.

in virtù del fatto che i valori $\underline{n}(hT_c)$ sono a media nulla, che (con n fissato) la FFT ne effettua una combinazione lineare con coefficienti $e^{-j2\pi \frac{h}{N}n}$, e che essendo $\underline{n}(t)$ ergodico è possibile scambiare medie temporali e di insieme. Sviluppando

$$(\underline{N}_n)^2 = \underline{N}_n \underline{N}_n^* = \frac{1}{N^2} \sum_{h=0}^{N-1} \sum_{k=0}^{N-1} (-1)^{h-k} \underline{n}(hT_c) \underline{n}^*(kT_c) e^{-j2\pi \frac{h-k}{N}n}$$

e tenendo conto che $E \left\{ (-1)^{h-k} \underline{n}(hT_c) \underline{n}^*(kT_c) \right\} = e^{j\pi(h-k)} \mathcal{R}_{\underline{N}}((h-k)T_c)$ otteniamo²⁴

$$\begin{aligned} \mathcal{P}_{\underline{N}_n} &= \frac{1}{N^2} \sum_{h=0}^{N-1} \sum_{k=0}^{N-1} \mathcal{R}_{\underline{N}}((h-k)T_c) e^{j\pi(h-k)} e^{-j2\pi \frac{h-k}{N}n} = \\ &= \frac{1}{N} \sum_{m=-(N-1)}^{N-1} \frac{N-|m|}{N} \mathcal{R}_{\underline{N}}(mT_c) e^{j2\pi \frac{mT_c}{2T_c}} e^{-j2\pi \frac{m}{N}n} = \\ &= \frac{1}{N} \sum_{m=-(N-1)}^{N-1} z(m) e^{-j2\pi \frac{m}{N}n} \end{aligned} \quad (11.15)$$

in cui l'ultima riga semplifica l'espressione introducendo la sequenza $\{z(m)\}$ di lunghezza N , che si ottiene campionando

$$z(t) = \left(1 - \frac{|t|}{NT_c}\right) \mathcal{R}_{\underline{N}}(t) e^{j2\pi \frac{t}{2T_c}} \quad (11.16)$$

agli istanti $t = mT_c$ con $T_c = \frac{1}{N\Delta}$.

Mostriamo ora come, per N sufficientemente elevato, la (11.15) possa essere calcolata in funzione dei campioni di $Z(f) = \mathcal{F}\{z(t)\}$, ed in particolare di come risulti $\mathcal{P}_{\underline{N}_n} \simeq \Delta \cdot Z(f)|_{f=n\Delta} \simeq 4\Delta \cdot \mathcal{P}_N(f_n)$.

Analizzando i termini che compaiono in (11.16), osserviamo che il prodotto $\mathcal{R}_{\underline{N}}(t) e^{j2\pi \frac{t}{2T_c}}$ ha trasformata pari a $\mathcal{P}_{\underline{N}}(f)$, translata in frequenza di $-\frac{1}{2T_c} = -\frac{N\Delta}{2}$, ovvero

$$\mathcal{F}\left\{\mathcal{R}_{\underline{N}}(t) e^{j2\pi \frac{t}{2T_c}}\right\} = \mathcal{P}_{\underline{N}}\left(f - \frac{N\Delta}{2}\right)$$

mentre il termine $\left(1 - \frac{|t|}{NT_c}\right) = \text{tri}_{2NT_c}(t) = \text{tri}_{\frac{2}{\Delta}}(t)$ possiede come noto trasformata $\mathcal{F}\left\{\text{tri}_{\frac{2}{\Delta}}(t)\right\} = \frac{1}{\Delta} \text{sinc}^2\left(\frac{f}{\Delta}\right)$; pertanto per N elevato il prodotto $z(t) = \mathcal{R}_{\underline{N}}(t) e^{j2\pi \frac{t}{2T_c}} \cdot \text{tri}_{\frac{2}{\Delta}}(t)$ ha trasformata

$$Z(f) = \mathcal{P}_{\underline{N}}\left(f - \frac{N\Delta}{2}\right) * \frac{1}{\Delta} \text{sinc}^2\left(\frac{f}{\Delta}\right) \simeq \mathcal{P}_{\underline{N}}\left(f - \frac{N\Delta}{2}\right)$$

avendo approssimato $\frac{1}{\Delta} \text{sinc}^2\left(\frac{f}{\Delta}\right)$ come un impulso di area unitaria, per $N\Delta$ grande rispetto a Δ .

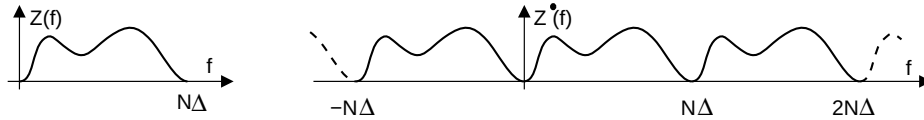
²⁴La riduzione da due ad una sommatoria, si ottiene scrivendo esplicitamente tutti i termini della doppia sommatoria, e notando che si ottiene per N volte lo stesso termine $\mathcal{R}_{\underline{N}}(0)$, $N-1$ volte i termini $\mathcal{R}_{\underline{N}}(T_c) e^{j\pi} e^{-j2\pi \frac{1}{N}n}$ e $\mathcal{R}_{\underline{N}}(-T_c) e^{-j\pi} e^{j2\pi \frac{1}{N}n}$, $N-2$ volte quelli $\mathcal{R}_{\underline{N}}(2T_c) e^{j2\pi} e^{-j2\pi \frac{2}{N}n}$ e $\mathcal{R}_{\underline{N}}(-2T_c) e^{-j2\pi} e^{j2\pi \frac{2}{N}n}$, e così via.

Dato che $\mathcal{P}_N(f)$ è limitato in banda tra $\pm \frac{N\Delta}{2}$, allora $Z(f)$ è limitato in una banda compresa tra $f = 0$ ed $f = N\Delta$, e $z(t)$ è perfettamente rappresentato dai suoi campioni $z(m) = z(mT_c)$ che compaiono nella (11.15); in particolare, per N sufficientemente elevato, si ottiene²⁵ che

$$\begin{aligned}\mathcal{P}_{N_n} &= \frac{1}{N} \sum_{m=-(N-1)}^{N-1} z(m) e^{-j2\pi \frac{m}{N} n} \simeq \Delta \cdot Z(f)|_{f=n\Delta} = \\ &= \Delta \cdot \mathcal{P}_N\left(n\Delta - \frac{N\Delta}{2}\right) = \Delta \cdot \mathcal{P}_N\left(\Delta\left(n - \frac{N}{2}\right)\right) = 4\Delta \cdot \mathcal{P}_N^+\left(f_0 + \Delta\left(n - \frac{N}{2}\right)\right) = \\ &= 4\Delta \cdot \mathcal{P}_N(f_n) = 2\Delta \cdot \mathcal{N}_0(f_n)\end{aligned}$$

in cui si è tenuto conto che $\mathcal{P}_N(f) = 4\mathcal{P}_N^+(f + f_0)$ e si è indicata la densità di potenza in ingresso come $\mathcal{P}_N(f) = \frac{\mathcal{N}_0(f)}{2}$.

²⁵Se campioniamo $z(t)$ con periodo $T_c = \frac{1}{N\Delta}$, il segnale $Z^\bullet(f) = \sum_{m=-\infty}^{\infty} Z(f - m \cdot N\Delta)$ non presenta aliasing (vedi figura), ed il passaggio di $z^\bullet(t) = \sum_{m=-\infty}^{\infty} z(mT_c) \delta(t - mT_c)$ attraverso un filtro di



ricostruzione $H(f) = \frac{1}{N\Delta} \text{rect}_{N\Delta}\left(f - \frac{N\Delta}{2}\right)$ restituisce il segnale originario. Scriviamo pertanto

$$z(t) = z^\bullet(t) * h(t) = \sum_{m=-\infty}^{\infty} z(mT_c) \delta(t - mT_c) * \text{sinc}(N\Delta t) e^{j\pi N\Delta t}$$

ed effettuiamone la trasformata:

$$\begin{aligned}Z(f) &= \mathcal{F}\left\{\sum_{m=-\infty}^{\infty} z(mT_c) \delta(t - mT_c)\right\} \cdot \frac{1}{N\Delta} \text{rect}_{N\Delta}\left(f - \frac{N\Delta}{2}\right) \\ &= \left[\sum_{m=-\infty}^{\infty} z(mT_c) e^{-j2\pi \frac{m}{N\Delta} f}\right] \cdot \frac{1}{N\Delta} \text{rect}_{N\Delta}\left(f - \frac{N\Delta}{2}\right)\end{aligned}$$

che, calcolata alle frequenze $f = n\Delta$ con $n = 0, 1, \dots, N-1$ fornisce

$$Z(f)|_{f=n\Delta} = \frac{1}{N\Delta} \sum_{m=-\infty}^{\infty} z(mT_c) e^{-j2\pi \frac{m}{N} n}$$

Se ora non disponiamo di tutti i campioni $z(mT_c)$, ma solo degli $2N-1$ valori con $m = -(N-1), \dots, 0, 1, \dots, N-1$, la relazione precedente si applica ad un nuovo segnale $z'(t) = z(t) \cdot \text{rect}_{2NT_c}(t)$, fornendo

$$Z'(f)|_{f=n\Delta} = \frac{1}{N\Delta} \sum_{m=-(N-1)}^{N-1} z(mT_c) e^{-j2\pi \frac{m}{N} n}$$

In virtù delle proprietà delle trasformate, risulta

$$Z'(f) = Z(f) * \mathcal{F}\{\text{rect}_{2NT_c}(t)\} \simeq Z(f) * \delta(f) = Z(f)$$

in cui l'approssimazione è lecita per N elevato.

Prestazioni per portante Siamo finalmente in grado di scrivere

$$SNR_n = \frac{\mathcal{P}_{R_n}}{\mathcal{P}_{N_n}} = \frac{2\frac{T_0}{T}\alpha_n\mathcal{P}}{2\Delta\mathcal{N}_0(f_n)} = \frac{T_0}{T}\alpha_n\frac{T_0\mathcal{P}}{\mathcal{N}_0(f_n)} = \frac{T_0}{T}\alpha_n\frac{E_s}{\mathcal{N}_0(f_n)} = \frac{T_0}{T}\alpha_n M_n \frac{E_{b_n}}{\mathcal{N}_0(f_n)}$$

avendo posto $E_s = T_0\mathcal{P}$ pari all'energia di un simbolo di durata $T_0 = \frac{1}{\Delta}$. L'energia per bit risulta dunque $E_b = \frac{E_s}{M}$, mentre per la portante n -esima si ha $E_{b_n} = \frac{E_s}{M_n}$. La P_e per portante risulta quindi

$$P_{e/n} = \frac{2}{M_n} \left(1 - \frac{1}{\sqrt{L_n}}\right) \operatorname{erfc} \left\{ \sqrt{\frac{3}{2} \frac{T_0}{T} \frac{E_{b_n}}{\mathcal{N}_0(f_n)} \frac{\alpha_n M_n}{L_n - 1}} \right\} \quad (11.17)$$

Caso di rumore bianco Se $\mathcal{P}_N(f)$ non dipende da f , possiamo scrivere

$$\mathcal{P}_N^+(f) = \frac{\mathcal{N}_0}{2} \operatorname{rect}_{N\Delta}(f - f_0)$$

e semplificare la (11.17), sostituendo ad $\mathcal{N}_0(f_n)$ la costante $\mathcal{N}_0$. In questo caso, il risultato $\mathcal{P}_{N_n} = 2\Delta \cdot \mathcal{N}_0$ può essere ottenuto direttamente dalla (11.15): infatti, risulta

$$\mathcal{R}_N(t) = \mathcal{F}^{-1} \{ \mathcal{P}_N(f) \} = \mathcal{F}^{-1} \{ 4\mathcal{P}_N^+(f + f_0) \} = 2\mathcal{N}_0 N \Delta \operatorname{sinc}(N\Delta t)$$

e dunque $\mathcal{R}_N(t) = 0$ con $t = mT_c = \frac{m}{N\Delta}$ per $m \neq 0$. Ciò permette di scrivere in definitiva

$$\mathcal{P}_{N_n} = \frac{1}{N} \mathcal{R}_N(0) = \frac{1}{N} 2\mathcal{N}_0 N \Delta = 2\Delta \cdot \mathcal{N}_0$$

Confronto con la portante singola Proviamo a verificare se la modulazione OFDM è vantaggiosa in termini di prestazioni, per una medesima occupazione di banda ed a parità di potenza. Nel caso in cui il tempo di guardia $T_g = T - T_0$ sia nullo, in presenza di rumore bianco, e scegliendo un intervallo di simbolo $T_0 = \frac{1}{\Delta}$ da cui derivare $M^{OFDM} = T_0 \cdot f_b$, $M_n = \frac{M^{OFDM}}{\tilde{N}}$ e $\alpha_n = \frac{1}{\tilde{N}}$, si ottengono valori $\frac{E_{b_n}}{\mathcal{N}_0}$ uguali per le diverse portanti, a cui corrisponde il miglior valore di

$$P_e^{OFDM} = P_{e/n} = \frac{2\tilde{N}}{M^{OFDM}} \left(1 - \frac{1}{\sqrt{L_n}}\right) \operatorname{erfc} \left\{ \sqrt{\frac{3}{2} \frac{E_b}{\mathcal{N}_0} \frac{1}{\tilde{N}} \frac{M^{OFDM}}{L_n - 1}} \right\}$$

ottenuta tenendo conto che $E_{b_n} M_n = E_s = E_b M^{OFDM} = E_b \log_2 L^{OFDM}$.

Nel caso in cui si adotti una modulazione a portante singola con impulso a coseno rialzato e roff-off $\gamma = \frac{N}{\tilde{N}} - 1$, si determina una occupazione di banda pari a $B = f_L(1 + \gamma)$ che, se eguagliata a quella del caso OFDM, fornisce $f_L = \tilde{N}\Delta = \frac{\tilde{N}}{T_0}$ e quindi $M^{QAM} = \frac{f_b}{f_L} = \frac{M^{OFDM}}{\tilde{N}}$. Pertanto in questo caso si ottiene

$$\begin{aligned} P_e^{QAM} &= \frac{2}{M^{QAM}} \left(1 - \frac{1}{\sqrt{L}}\right) \operatorname{erfc} \left\{ \sqrt{\frac{3}{2} \frac{E_b}{\mathcal{N}_0} \frac{M^{QAM}}{L - 1}} \right\} \\ &= \frac{2\tilde{N}}{M^{OFDM}} \left(1 - \frac{1}{\sqrt{L}}\right) \operatorname{erfc} \left\{ \sqrt{\frac{3}{2} \frac{E_b}{\mathcal{N}_0} \frac{1}{\tilde{N}} \frac{M^{OFDM}}{L - 1}} \right\} \end{aligned}$$

che risulta identico a P_e^{OFDM} qualora si noti che $L_n = 2^{M_n} = 2^{\frac{M^{OFDM}}{N}}$ e $L = 2^{M^{QAM}} = 2^{\frac{M^{OFDM}}{N}} = L_n$.

E allora dov'è la convenienza? È il tema delle prossime sottosezioni.

11.5.3.6 Equalizzazione

Consideriamo il caso in cui la trasmissione attraverso un canale descritto da un inviluppo complesso $\underline{H}(f)$ in cui il modulo non è costante e/o la fase non è lineare: in tal caso $\underline{X}_{T_0}(f)$ di (11.10) si altera, ed i valori $\underline{a}_n$ restituiti dalla (11.12) si modificano in $\underline{b}_n = \underline{a}_n \cdot \underline{H}\left(f - \Delta\left(n - \frac{N}{2}\right)\right)$. Come anticipato, l'equalizzazione è pertanto ridotta ad eseguire un semplice prodotto vettoriale tra i valori $\{\underline{b}_n\}$ e la sequenza di valori $\left\{\frac{1}{\underline{H}\left(f - \Delta\left(n - \frac{N}{2}\right)\right)}\right\}$.

Modulazione differenziale Nel caso in cui la distorsione non sia eccessiva, si può evitare del tutto lo stadio di equalizzazione, e ricorrere ad una modulazione differenziale. In presenza di distorsione di fase infatti, il piano dell'inviluppo complesso subisce, per ogni portante consecutiva, una rotazione pari alla differenza della fase di $\underline{H}(f)$ calcolata per le due frequenze contigue. Se questa quantità non è eccessiva, si può prendere come riferimento di fase il risultato della demodulazione della portante precedente.

11.5.3.7 Sensibilità alla temporizzazione

Nel caso in cui il ricevitore non acquisisca una perfetta sincronizzazione di simbolo, il calcolo della FFT su di un gruppo di campioni presi a partire dalla coda del preambolo, non altera per nulla il risultato²⁶ in virtù della caratteristica di periodicità dello stesso. Inoltre, non è indispensabile ottenere un perfetto sincronismo di bit: ritardi di temporizzazione sono come noto equivalenti ad incrementi lineari della risposta in fase, eliminabili mediante equalizzazione o modulazione differenziale.

11.5.3.8 Ottimalità

Come stiamo per mostrare, questa proprietà è intimamente legata alla possibilità dell'OFDM di assegnare valori di potenza differenti alle diverse portanti.

La trasmissione numerica con una f_b elevata, eseguita utilizzando una sola portante, deve necessariamente occupare una banda molto ampia; nel caso in cui $H(f)$ presenti una elevata distorsione di ampiezza, l'equalizzazione della stessa causa una colorazione del rumore in ingresso al demodulatore, ed un peggioramento delle prestazioni. Un problema analogo nasce nel caso in cui il rumore non sia bianco, ad esempio perchè derivante da un segnale interferente.

In entrambi i casi, la teoria di SHANNON che prevede²⁷ una *capacità di canale* pari a $C = W \log_2 \left(1 + \frac{P_r}{W N_0}\right)$ in presenza di un rumore bianco, con $\mathcal{P}_N(f) = \frac{N_0}{2}$ e con una

²⁶Infatti non siamo nelle condizioni di demodulazione coerente dell'FSK, e le portanti del simbolo OFDM ricevuto mantengono ortogonalità purchè finestrate su di un periodo $T_0 = \frac{1}{\Delta}$.

²⁷Il risultato della teoria di Shannon asserisce che è possibile conseguire una velocità di trasmissione $f_b = C$ con probabilità di errore nulla, ma non indica come fare. Una soluzione al problema è quella di adottare una codifica di canale a ridondanza elevata, capace di correggere un elevato numero di errori.

potenza ricevuta $\mathcal{P}_r$ in una banda (positiva) W , si modifica nel seguente modo, per tenere conto dell'andamento incostante di $\mathcal{P}_r(f)$ e $\mathcal{P}_N(f)$. Considerando il canale scomposto in infinite sottobande entro le quali le densità possono ritenersi costanti, l'espressione della capacità diviene ora:

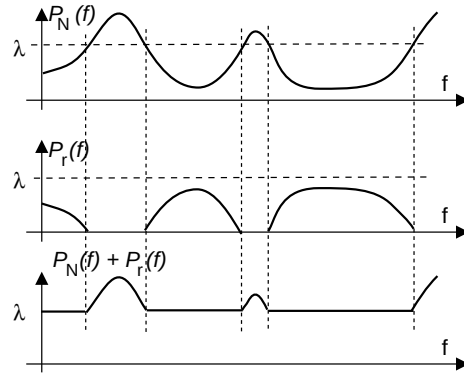
$$C = \sup_{\mathcal{P}_R(f)} \int_{f \in I_f} \log_2 \left(1 + \frac{\mathcal{P}_r(f)}{\mathcal{P}_N(f)} \right) df \quad (11.18)$$

in cui $\mathcal{P}_r(f)$ viene fatto variare in tutti i modi possibili e tali che $\int_{f \in I_f} \mathcal{P}_r(f) df = \mathcal{P}_r$, con $\mathcal{P}_r(f) \geq 0$, ed I_f rappresenta l'insieme delle frequenze in cui è presente il segnale: $I_f = \{f : \mathcal{P}_r(f) > 0\}$. La (11.18) asserisce quindi che, nel caso in cui $\mathcal{P}_N(f)$ in ingresso al canale non sia bianco, le migliori prestazioni (f_b che tende a C) si ottengono solamente sagomando la densità di potenza del segnale ricevuto in modo opportuno.

Per determinare l'andamento ottimo di $\mathcal{P}_r(f)$ si ricorre allora al *calcolo delle variazioni* basato sui *moltiplicatori di Lagrange*, che in questa sede non affrontiamo, e che fornisce la condizione

$$\mathcal{P}_r(f) + \mathcal{P}_N(f) = \begin{cases} \lambda & \text{se } \mathcal{P}_N(f) < \lambda \\ \mathcal{P}_N(f) & \text{se } \mathcal{P}_N(f) \geq \lambda \end{cases} \quad (11.19)$$

detta anche del *riempimento d'acqua* perché asserisce che (vedi figura) il segnale debba essere presente in misura maggiore nelle regioni di frequenza dove il rumore è sufficientemente ridotto. La costante λ è scelta in modo tale da ottenere $\int \mathcal{P}_r(f) df = \mathcal{P}_r$.



In un sistema di modulazione numerica a singola portante, $\mathcal{P}_r(f)$ non può essere modificato a piacere, in quanto il suo andamento deve essere quello legato alla particolare caratteristica di Nyquist $G(f)$ scelta per ottenere una ricezione priva di ISI. Nel caso dell'OFDM invece, la potenza assegnata a ciascuna portante può essere variata liberamente, ed in particolare può essere scelta in modo da realizzare le condizioni (11.19) e quindi rendere massima la velocità di trasmissione conseguibile.

In particolare, si ottiene che la massima velocità f_b è conseguibile attribuendo a tutte le portanti la medesima probabilità di errore, e quindi in definitiva determinando dei valori $\left(\frac{E_b}{N_0}\right)_n$ per ogni portante $n = 0, 1, \dots, \tilde{N} - 1$ tali da rendere le $P_{e/n} = P_e$. Questo risultato può essere ottenuto sia adottando per ogni portante lo stesso numero di livelli $L_n = 2^{\frac{M}{N}}$ e variando $\mathcal{P}_n$, sia variando congiuntamente entrambi $\mathcal{P}_n$ e M_n , sia tenendo fisso $\mathcal{P}_n = \frac{\mathcal{P}_r}{N}$ e variando unicamente M_n .

11.5.3.9 Codifica

Abbiamo appena mostrato come, conoscendo la $H(f)$ e la $\mathcal{P}_N(f)$ del canale, sia possibile determinare $\mathcal{P}_x(f) = \frac{\mathcal{P}_r(f)}{|H(f)|^2}$ in modo da soddisfare (11.19) e rendere massima la f_b . Nel caso di collegamenti tempo-varianti però, la $H(f)$ non è nota, ed anche se lo fosse non esiste garanzia che rimanga costante. In tal caso allora non ha senso determinare una distribuzione ottima della potenza e dei bit sulle portanti, mentre invece occorre

aggiungere della ridondanza al segnale trasmesso mediante un codice di canale, allo scopo di correggere i bit errati.

Osserviamo ora che, nel caso di una modulazione a portante singola, in presenza di una $H(f)$ tempo-variante, il processo di equalizzazione è particolarmente complesso in quanto deve “inseguire” le variazioni di $H(f)$. Se l’equalizzazione non è perfetta, insorge ISI e la trasmissione diviene rapidamente così piena di errori da renderne impossibile la correzione anche adottando codici di canale.

Nel caso dell’OFDM, al contrario, l’andamento di $H(f)$ determina un peggioramento di prestazioni solamente per quelle portanti per le quali $H(f)$ si è ridotto²⁸. Pertanto, l’applicazione di un codice di canale al blocco di M bit che costituisce un simbolo, seguito da una operazione di scrambling, consente al lato ricevente di recuperare l’informazione trasmessa anche nel caso in cui per alcune portanti si determini un elevato tasso di errore.

La trasmissione OFDM in cui è presente una codifica di canale prende il nome di trasmissione COFDM (*Coded* OFDM).

²⁸Si consideri ad esempio il caso in cui $H(f)$ ha origine da un fenomeno di cammini multipli, che determina un andamento di $H(f)$ oscillante in frequenza.

Capitolo 12

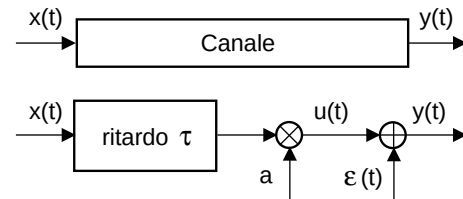
Transito dei segnali nei circuiti

Trattiamo qui della descrizione dei circuiti elettrici in termini di sistemi fisici, e delle modifiche che questi generano nei segnali in transito.

Elaborazione e Distorsione Quando un segnale viene “manipolato” di proposito, si dice che questo è *elaborato*. Se viceversa il segnale si altera per una causa indipendente dalla volontà, allora il segnale subisce una *distorsione*¹.

Canale perfetto Per valutare l'entità della distorsione, stabiliamo un criterio con cui distinguere la componente di segnale *utile* dal *disturbo*.

Come è noto, la ricezione di un segnale identico a quello trasmesso, tranne che per un fattore di scala ed un ritardo temporale, non altera la sostanza del messaggio: pertanto, un canale che presenti una risposta impulsiva



$$h(t) = a\delta(t - \tau)$$

¹L'elaborazione di un segnale è indicata anche come suo “processamento” (dall'inglese PROCESSED=trattato). In altri contesti non “comunicazionistici” la terminologia può essere ancora più varia, come ad esempio... le alterazioni prodotte sul suono di uno strumento musicale sono indicate come *effetti* ed il segnale risultante è “effettato” (!).

²Nella pratica, i valori a e τ non si conoscono, mentre invece possiamo disporre di coppie di segnali $(x(t), y(t))$. I valori vengono dunque definiti come quelli che *rendono SNR massimo* ovvero $\mathcal{P}_\varepsilon$ minimo. Considerando segnali di potenza, ossia processi stazionari ergodici, si ha

$$\begin{aligned}\mathcal{P}_\varepsilon(a, \tau) &= E\{(y(t) - ax(t - \tau))^2\} = E\{y^2(t)\} + a^2 E\{x^2(t)\} - 2a E\{y(t)x(t - \tau)\} = \\ &= \mathcal{P}_y + a^2 \mathcal{P}_x - 2a \mathcal{R}_{xy}(\tau)\end{aligned}$$

in cui si è operata la sostituzione $E\{y(t)x(t - \tau)\} = \mathcal{R}_{yx}(-\tau) = \mathcal{R}_{xy}^*(\tau) = \mathcal{R}_{xy}(\tau)$.

Il valore di a che rende minimo $\mathcal{P}_\varepsilon(a, \tau)$ si ottiene eguagliando a zero la derivata: $\frac{\partial}{\partial a} \mathcal{P}_\varepsilon(a, \tau) = 2a\mathcal{P}_x - 2\mathcal{R}_{xy}(\tau) = 0 \Rightarrow a_{opt} = \frac{\mathcal{R}_{xy}(\tau)}{\mathcal{P}_x}$, che sostituita nell'espressione di $\mathcal{P}_\varepsilon$ fornisce

$$\mathcal{P}_\varepsilon(\tau) = \mathcal{P}_y + \left(\frac{\mathcal{R}_{xy}(\tau)}{\mathcal{P}_x}\right)^2 \mathcal{P}_x - 2\frac{\mathcal{R}_{xy}(\tau)}{\mathcal{P}_x} \mathcal{R}_{xy}(\tau) = \mathcal{P}_y - \frac{(\mathcal{R}_{xy}(\tau))^2}{\mathcal{P}_x} = \mathcal{P}_y \left(1 - \frac{(\mathcal{R}_{xy}(\tau))^2}{\mathcal{P}_x \mathcal{P}_y}\right)$$

Il valore di $\mathcal{P}_\varepsilon$ evidentemente è minimo per quel valore di $\tau = \tau_{opt}$ che rende massima $(\mathcal{R}_{xy}(\tau))^2$, ovvero per quella traslazione temporale che rende “più simili” i segnali di ingresso ed uscita.

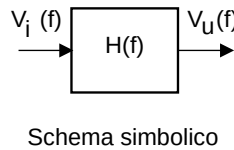
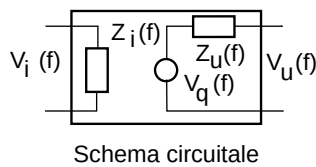
viene indicato come *canale perfetto*, ed il segnale ricevuto $y(t) = u(t) = ax(t - \tau)$ è *tutto utile*. Se invece non vale quest'uguaglianza, viene definito *disturbo additivo* la differenza $\varepsilon(t) = y(t) - ax(t - \tau) = y(t) - u(t)$ ⁽²⁾. Infine, il rapporto $\frac{P_u}{P_\varepsilon}$ tra la potenza del segnale utile e quella del disturbo, prende il nome di *rapporto segnale rumore* (SNR).

12.1 Caratterizzazione dei circuiti

12.1.1 Numero di porte

Si indicano come *porte* le coppie di morsetti a cui applicare o da cui prelevare un segnale. Pertanto, i *generatori* ed le *impedenze* costituiscono reti ad *una* porta, mentre un sistema fisico dotato di relazione ingresso-uscita è una rete *due porte*.

12.1.2 Modelli di rappresentazione



Un circuito può essere rappresentato mediante il suo modello circuitale, in cui sono evidenziati generatori, resistenze, impedenze, generatori controllati..., oppure il suo schema simbolico, in cui sono solo mostrate le relazioni funzionali tra i segnali in transito.

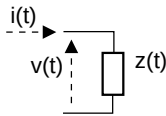
12.1.3 Proprietà delle reti due porte

Le seguenti proprietà, già definite a suo tempo per i sistemi fisici in § 7.5, possono essere presenti o meno nelle reti due porte: linearità, permanenza, realizzabilità ideale e fisica, stabilità.

12.2 Bipoli

Passivi Non contengono generatori, e sono caratterizzati dalle relazioni esistenti tra la tensione ai loro capi e la corrente che vi scorre (entrante).

Il legame tra le due grandezze è una *convoluzione*



$$v(t) = i(t) * z(t)$$

in cui si suppone $i(t)$ la causa e $v(t)$ l'effetto. La trasformata di Fourier fornisce $V(f) = I(f) \cdot Z(f)$ in cui $Z(f)$ prende il nome di *impedenza*, e può scriversi in termini di parte reale ed immaginaria:

$$Z(f) = R(f) + jX(f)$$

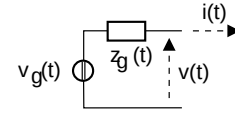
in cui $R(f)$ (*resistenza*) è una funzione pari di f e positiva, ed $X(f)$ (*reattanza*) è dispari. Pertanto, $Z(f) = Z^*(-f)$ e $z(t)$ è reale. Allo stesso tempo, è definita l'*ammettenza*

$$Y(f) = \frac{1}{Z(f)} = \frac{R(f) - jX(f)}{|Z(f)|^2}$$

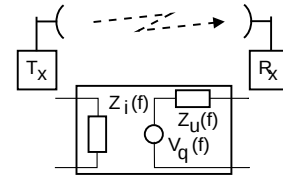
e la corrispondente $y(t) = \mathcal{F}^{-1}\{Y(f)\}$, che permette di scrivere $i(t) = v(t) * y(t)$.

Attivi

Sono bipoli al cui interno è presente un generatore. Per il teorema di THEVENIN, qualunque circuito può essere ridotto ad un generatore di tensione con in serie una impedenza (vedi figura), in cui $V_g(f)$ rappresenta la tensione a vuoto, ossia quando $I(f) = 0$ (considerata uscente nei bipoli attivi).

**Esempio**

Una antenna trasmittente è schematizzabile come un bipolo passivo, di impedenza pari all'impedenza di ingresso dell'antenna, che assorbe la potenza erogata dal trasmettitore. Una antenna ricevente è schematizzabile come un generatore di tensione, con in serie la propria impedenza di uscita, ed eroga la potenza ricevuta per via elettromagnetica.

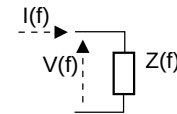
**12.3 Potenza di segnale e grandezze elettriche**

Tutta la caratterizzazione energetica svolta finora sui segnali è stata svolta *a prescindere* dalla natura fisica degli stessi: ovvero, non si è mai specificato se si trattasse di tensioni o correnti, né si sono indicate le impedenze in gioco. Trattando ora di grandezze elettriche, le potenze di segnale, di tensione o di corrente, saranno misurate in $(V\text{olt})^2$ o in $(A\text{mpere})^2$ rispettivamente.

Esempio Sia $x(t)$ un segnale di tensione. La sua potenza $\mathcal{P}_x$ ha unità di misura $[V^2]$, mentre la sua densità di potenza $\mathcal{P}_x(f)$ si esprime in $\left[\frac{V^2}{Hz}\right]$.

12.3.1 Potenza assorbita da un bipolo

Se ad un bipolo passivo di impedenza $Z(f)$ è applicato un segnale di tensione con spettro di densità di potenza $\mathcal{P}_v(f)$, la potenza *dissipata* sul bipolo, indicata come $\mathcal{W}_z(f)$ per distinguerla da quella di segnale, ha densità



$$\mathcal{W}_z(f) = \mathcal{P}_v(f) \cdot \Re\{Y(f)\} = \mathcal{P}_v(f) \frac{R(f)}{|Z(f)|^2} \quad \left[\frac{V^2}{\Omega \cdot Hz} \right] = \left[\frac{Watt}{Hz} \right] \quad (12.1)$$

La dimostrazione della relazione illustrata è fornita in appendice 12.8.1. La dipendenza di $Y(f)$ dalla frequenza svolge una azione filtrante, e la potenza totale assorbita (o dissipata) vale

$$\mathcal{W}_z = \int_{-\infty}^{\infty} \mathcal{P}_v(f) \frac{R(f)}{|Z(f)|^2} df \quad [Watt]$$

12.3.2 Misure di potenza

Le potenze assorbite, ed espresse in Watt, sono grandezze fisiche, e come tali percepibili dai nostri sensi³. La misura delle potenze è spesso espressa in *decibel*⁴, già introdotti al § 5.6.3, e che rappresentano la misura logaritmica del rapporto tra potenze:

$$R_{dB} = 10 \log_{10} \frac{\mathcal{W}_1}{\mathcal{W}_2} [dB]$$

Se una potenza $\mathcal{W}_1$ è R_{dB} dB maggiore di un'altra, il loro rapporto risulta $\frac{\mathcal{W}_1}{\mathcal{W}_2} = 10^{\frac{R_{dB}}{10}}$, e pertanto $\mathcal{W}_1$ (in Watt) è $10^{\frac{R_{dB}}{10}}$ volte maggiore di $\mathcal{W}_2$ (in Watt). Ad esempio, con $R_{dB} = 20$ dB si ha $\frac{\mathcal{W}_1}{\mathcal{W}_2} = 100$ volte, e $\mathcal{W}_1[Watt] = 100 \cdot \mathcal{W}_2[Watt]$. Quindi, il dB è una misura di potenza *relativa*.

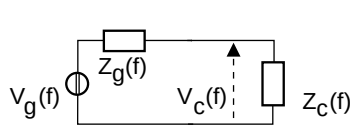
Per contro, il dB può essere usato come unità di misura *assoluta*, ponendo come potenza di riferimento $\mathcal{W}_2$ un valore noto, ad esempio 1 Watt, e riflettendo questo valore nel nome dell'unità di misura:

$$R_{dBW} = 10 \log_{10} \frac{\mathcal{W}_1}{1 Watt} [dBW]$$

In tal caso, se $\mathcal{W}_1$ misura 30 dBW, vuol dire che $\frac{\mathcal{W}_1}{1 Watt} = 10^{\frac{30}{10}}$, e quindi $\mathcal{W}_1 = 1000 \cdot 1 Watt = 1000 Watt$. Allo stesso modo, si usa il dBm, che vuol dire che il riferimento è 1 milliWatt. Si può passare dall'uno all'altro ricordando che se $1000 mW = 1 W$, allora $30 dBm = 0 dBW$ ⁵.

Usando i dB, i prodotti ed i rapporti divengono somme e differenze⁶. Se invece sommiamo due potenze fornite in dBW, si deve prima tornare a Watt, sommarle, e poi portarle di nuovo a dBW. Usando i dB può capitare di avere a che fare con... potenze negative: in realtà, ciò rappresenta solamente un rapporto minore di 1. Esempio: a quanto equivalgono $-18 dBW$ espressi in dBm ? A $+12 dBm$.

12.4 Connessione tra generatore e carico



La tensione ai capi del carico è valutabile applicando la *regola del partitore*:

$$V_c(f) = V_g(f) \frac{Z_c(f)}{Z_c(f) + Z_g(f)}$$

ossia $V_c(f) = V_g(f) H(f)$ con $H(f) = \frac{Z_c(f)}{Z_c(f) + Z_g(f)}$. La potenza di segnale ai capi del carico vale $\mathcal{P}_{v_c}(f) = \mathcal{P}_{v_g}(f) |H(f)|^2$, e la potenza dissipata su $Z_c(f)$ risulta

$$\mathcal{W}_{z_c}(f) = \mathcal{P}_{v_c}(f) \frac{R_c(f)}{|Z_c(f)|^2} = \mathcal{P}_{v_g}(f) \left| \frac{Z_c(f)}{Z_c(f) + Z_g(f)} \right|^2 \frac{R_c(f)}{|Z_c(f)|^2} = \mathcal{P}_{v_g}(f) \frac{R_c(f)}{|Z_c(f) + Z_g(f)|^2}$$

³Potenza acustica della musica, potenza termica della stufa, potenza luminosa della lampada, potenza meccanica di un motore... tutte le nostre sensazioni seguono una legge percettiva non lineare, e che somiglia piuttosto ad una progressione geometrica.

⁴Il decibel è la decima parte del Bel. Chissà, forse dopo che definirono il Bel, si accorsero che era troppo grande ? :-)

⁵Esempio: quanto fa 12 dBW in dBm ? 42 dBm.

⁶Esempio: Se $\mathcal{W}_1$ è 15 dB maggiore di $\mathcal{W}_2 = 7 dBW$, allora $\mathcal{W}_1 (dBW) = \mathcal{W}_2 (dBW) + 15 dB = 22 dBW$, pari a $10^{2.2} = 158.5 Watt$. Infatti, $\mathcal{W}_1 (Watt) = 10^{1.5} \mathcal{W}_2 (Watt) = 31.6 \cdot 10^{0.7} = 31.6 \cdot 5 = 158.5 Watt$.

Osserviamo dunque che la potenza dissipata dal carico dipende da $Z_c(f)$, che compare sia a denominatore, che a numeratore con $R_c(f)$. Ci chiediamo allora quale sia il valore di Z_c che realizza il *massimo trasferimento di potenza* tra generatore e carico, sfruttando così appieno la

12.4.1 Potenza disponibile

Questo avviene per quel valore di $Z_c(f)$ che rende massima la potenza ceduta al carico. Innanzi tutto osserviamo che, se si pone $X_c = -X_g$, la parte immaginaria del denominatore si annulla; se poi si pone anche $R_c = R_g$, allora si ottiene il risultato cercato:

$$\text{Se } Z_c(f) = Z_g^*(f) \quad \text{allora} \quad \mathcal{W}_{z_c}(f) = \max_{Z_c(f)} \{\mathcal{W}_{z_c}(f)\} = \frac{\mathcal{P}_{v_g}(f)}{4R_g(f)} = \mathcal{W}_{d_g}(f)$$

Il risultato appena trovato prende il nome di spettro di potenza *disponibile* del generatore, dipende solo dai suoi parametri ($\mathcal{P}_{v_g}(f)$ e $R_g(f)$) e rappresenta la *massima* potenza ceduta ad un carico che è *adattato* per il *massimo trasferimento di potenza*.

Pertanto, la potenza disponibile $\mathcal{W}_{d_g}(f) = \frac{\mathcal{P}_{v_g}(f)}{4R_g(f)}$ è una grandezza caratteristica del generatore; la potenza effettivamente ceduta ad un carico generico $Z_c(f) \neq Z_g^*(f)$, risulta inferiore a $\mathcal{W}_{d_g}(f)$ di una quantità

$$\alpha(f) = \frac{4R_g(f)R_c(f)}{|Z_g(f) + Z_c(f)|^2}$$

(vedi appendice 12.8.2) e quindi in generale si ha $\mathcal{W}_{z_c}(f) = \alpha(f)\mathcal{W}_{d_g}(f)$.

12.4.2 Assenza di distorsioni lineari

Abbiamo già osservato come la tensione ai capi del carico abbia valore $V_c(f) = V_g(f) \cdot \frac{Z_c(f)}{Z_c(f) + Z_g(f)} = V_g(f)H(f)$. Ci chiediamo ora quali condizioni debbano sussistere affinché $H(f)$ si comporti come un *canale perfetto*, ovvero risulti $|H(f)| = \text{cost}$ e $\arg\{H(f)\} = 2\pi f\tau$: tali condizioni sono anche conosciute come *assenza di distorsioni lineari*. Il risultato cercato si ottiene qualora si ponga

$$Z_c(f) = \alpha Z_g(f) \quad \text{con } \alpha \text{ reale}$$

infatti in tal caso risulta $H(f) = \frac{\alpha Z_g(f)}{(1+\alpha)Z_g(f)} = \frac{\alpha}{1+\alpha}$, ossia $H(f)$ costante. La condizione $Z_c(f) = \alpha Z_g(f)$ prende il nome di *adattamento di impedenza*, a volte ristretto al caso in cui $\alpha = 1$.

12.4.3 $Z_g(f)$ reale

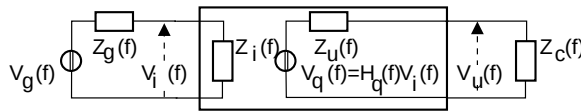
Notiamo che massimo trasferimento di potenza ed assenza di distorsioni lineari possono sussistere *congiuntamente*, a patto che $Z_g(f) = R_g$, ovvero che sia il generatore che il carico siano ad impedenza reale.

12.5 Reti due porte

Una rete elettrica accessibile a partire da due coppie di morsetti è detta rete due porte, e può essere rappresentata secondo almeno due diversi formalismi: il *modello circuitale* e lo *schema simbolico*.

12.5.1 Modello circuitale

In figura è mostrato un possibile modello circuitale⁷ per una rete due porte, caratterizzata in termini di impedenza di ingresso $Z_i(f)$, di uscita $Z_u(f)$, e di un generatore controllato con tensione a vuoto $V_q(f) = H_q(f) V_i(f)$; le condizioni di chiusura sono quelle di un generatore $V_g(f)$ con impedenza $Z_g(f)$ in ingresso, e di una impedenza di carico $Z_c(f)$ in uscita.



La tensione all'ingresso della rete

$$V_i(f) = V_g(f) H_i(f)$$

dipende da quella del generatore $V_g(f)$ mediante il rapporto di partizione $H_i(f) = \frac{Z_i(f)}{Z_g(f) + Z_i(f)}$, così come la tensione in uscita

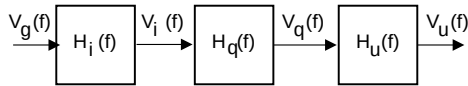
$$V_u(f) = V_q(f) H_u(f)$$

dipende da quella del generatore controllato $V_q(f)$ mediante il rapporto di partizione $H_u(f) = \frac{Z_c(f)}{Z_u(f) + Z_c(f)}$. Combinando queste relazioni, si ottiene che la funzione di trasferimento complessiva $H(f)$ risulta:

$$V_u(f) = V_g(f) H_i(f) H_q(f) H_u(f) = V_g(f) H(f) \quad (12.2)$$

La relazione mostra come $H(f)$ dipenda, oltre che dalla risposta in frequenza intrinseca della rete $H_q(f)$, anche dalle condizioni di adattamento che si realizzano in ingresso ed in uscita.

12.5.2 Schema simbolico



Lo stesso modello circuitale descritto può essere rappresentato equivalentemente mediante lo schema simbolico rappresentato a lato, in cui sono evidenziate le tre funzioni di trasferimento sopra

ricavate, e che operano sui segnali indicati. Lo schema simbolico ha il vantaggio di trascendere dal modello circuitale adottato, e di rendere del tutto evidente come la funzione di trasferimento complessiva abbia origine dal prodotto di tre termini di cui solo uno ($H_q(f)$) rappresenta strettamente la rete.

⁷Sono chiaramente possibili modelli diversi, basati su topologie e relazioni differenti. Esistono infatti circuiti a T, ad L, a scala, a traliccio, a pigreco; le relazioni tra le grandezze di ingresso ed uscita possono essere espresse mediante modelli definiti in termini di impedenze, ammettenze, e parametri ibridi.

Il caso qui trattato è quello di un modello ibrido, con la particolarità di non presentare influenze esplicite dell'uscita sull'ingresso. Qualora il circuito che si descrive presenti una dipendenza, ad esempio di Z_i da Z_c , o Z_u da Z_g , questo deve risultare nell'espressione della grandezza dipendente. Viceversa, qualora il circuito presenti in ingresso un generatore controllato da una grandezza di uscita, il modello non è più applicabile.

12.5.3 Trasferimento energetico

Applicando ora la (12.1) alla potenza ceduta al carico $Z_c(f)$ dal generatore controllato $V_q(f)$, e tenendo conto della (12.2), si ottiene:

$$\mathcal{W}_c(f) = \mathcal{P}_{v_u}(f) \frac{R_c(f)}{|Z_c(f)|^2} = \mathcal{P}_{v_g}(f) |H(f)|^2 \frac{R_c(f)}{|Z_c(f)|^2}$$

Proseguiamo ora l'analisi cercando di individuare una relazione di trasferimento energetico che possa rappresentare caratteristiche esclusive della rete.

Guadagno di tensione E' definito come il rapporto tra tensione di uscita e di ingresso:

$$G_v(f) = \frac{V_u(f)}{V_i(f)} = H_q(f) H_u(f)$$

Evidentemente, dipende dalle condizioni di chiusura della rete.

Guadagno di potenza E' il rapporto tra la potenza ceduta al carico e quella assorbita all'ingresso della rete:

$$\begin{aligned} G_{\mathcal{W}}(f) &= \frac{\mathcal{W}_c(f)}{\mathcal{W}_i(f)} = \mathcal{P}_{v_g}(f) |H(f)|^2 \frac{R_c(f)}{|Z_c(f)|^2} \cdot \frac{1}{\mathcal{P}_{v_g}(f)} \frac{|Z_g(f) + Z_i(f)|^2}{R_i(f)} = \\ &= |H(f)|^2 \frac{R_c(f)}{R_i(f)} \cdot \left| \frac{Z_g(f) + Z_i(f)}{Z_c(f)} \right|^2 = |H_q(f)|^2 \cdot \frac{R_c(f)}{R_i(f)} \cdot \left| \frac{Z_i(f)}{Z_u(f) + Z_c(f)} \right|^2 \end{aligned}$$

ed evidentemente è ancora funzione di $Z_c(f)$ ⁽⁸⁾. Notiamo ora che, qualora il carico sia adattato per il massimo trasferimento di potenza ($Z_c(f) = Z_u^*(f)$), la potenza ceduta a $Z_c(f)$ (e quindi $G_{\mathcal{W}}(f)$) è massima, e la dipendenza di $G_{\mathcal{W}}(f)$ da $Z_c(f)$ decade, risultando

$$G_{\mathcal{W}_{Max}}(f) = |H_q(f)|^2 \cdot \frac{|Z_i(f)|^2}{4R_i(f) R_u(f)} \quad (12.3)$$

Guadagno disponibile Il rapporto tra la potenza disponibile di uscita, e quella disponibile del generatore posto in ingresso della rete (indipendentemente dal fatto che il generatore presenti o meno le condizioni per il massimo trasferimento di potenza) è detto *guadagno disponibile*, e risulta:

$$\begin{aligned} G_d(f) &= \frac{\mathcal{W}_{d_u}(f)}{\mathcal{W}_{d_g}(f)} = \frac{\mathcal{P}_{v_q}(f)}{4R_u(f)} \cdot \frac{4R_g(f)}{\mathcal{P}_{v_g}(f)} = \mathcal{P}_{v_g}(f) |H_i(f)|^2 |H_q(f)|^2 \cdot \frac{R_g(f)}{R_u(f)} \cdot \frac{1}{\mathcal{P}_{v_g}(f)} = \\ &= |H_i(f)|^2 |H_q(f)|^2 \cdot \frac{R_g(f)}{R_u(f)} \end{aligned} \quad (12.4)$$

⁸L'ultimo passaggio tiene conto che (omettendo la dipendenza da f):

$$|H|^2 \cdot \left| \frac{Z_g + Z_i}{Z_c} \right|^2 = \left| \frac{Z_i}{Z_i + Z_g} H_q \frac{Z_c}{Z_c + Z_u} \right|^2 \left| \frac{Z_g + Z_i}{Z_c} \right|^2 = |H_q|^2 \cdot \left| \frac{Z_i}{Z_u + Z_c} \right|^2$$

La relazione trovata mostra la dipendenza di $G_d(f)$ dalle condizioni di chiusura in ingresso; se l'impedenza $Z_g(f)$ del generatore è scelta in modo da conseguire il massimo trasferimento di potenza $Z_g(f) = Z_i^*(f)$, la dipendenza decade ed $|H_i(f)|^2 = \left| \frac{Z_i(f)}{Z_i(f) + Z_i^*(f)} \right|^2 = \frac{|Z_i(f)|^2}{4R_i^2(f)}$; considerando inoltre che $R_g(f) = R_i(f)$, la (12.4) diviene:

$$G_{d_{Max}}(f) = |H_q(f)|^2 \cdot \frac{|Z_i(f)|^2}{4R_u(f)R_i(f)} \quad (12.5)$$

Quest'ultima quantità è chiamata *guadagno disponibile della rete due porte* ed è quella che appunto dipende solo dai parametri della rete stessa. Confrontando (12.5) con (12.3) notiamo che $G_{d_{Max}}(f)$ coincide con $G_{\mathcal{W}_{Max}}(f)$. Confrontando (12.5) con (12.4), troviamo che $G_d(f) = |H_i(f)|^2 G_{d_{Max}}(f) \frac{4R_g(f)R_i(f)}{|Z_i(f)|^2}$. Considerando ora che $|H_i(f)|^2 \frac{1}{|Z_i(f)|^2} = \left| \frac{Z_i(f)}{Z_i(f) + Z_g(f)} \right|^2 \frac{1}{|Z_i(f)|^2} = \frac{1}{|Z_i(f) + Z_g(f)|^2}$, otteniamo

$$G_d(f) = \frac{4R_g(f)R_i(f)}{|Z_g(f) + Z_i(f)|^2} \cdot G_{d_{Max}}(f)$$

che ci consente di valutare $G_d(f)$ nelle reali condizioni di chiusura in ingresso, a partire da $G_{d_{Max}}(f) = G_{\mathcal{W}_{Max}}(f)$ che dipende solo dalla rete.

Collegamento di un generatore ad un carico mediante una rete due porte

- Considerando generatore e rete adattati per il massimo trasferimento di potenza, la densità di potenza disponibile in uscita risulta

$$\mathcal{W}_{d_u}(f) = G_{d_{Max}}(f) \mathcal{W}_{d_g}(f)$$

e dunque l'uscita della rete due porte si comporta come un generatore equivalente, caratterizzato da una nuova $\mathcal{W}_{d_u}(f)$ ed una diversa impedenza interna $Z_u(f)$.

- Nel caso in cui in ingresso non si verifichi il massimo adattamento, $G_d(f)$ si riduce di un fattore $\beta(f) = \frac{4R_g(f)R_i(f)}{|Z_g(f) + Z_i(f)|^2}$, e dunque la nuova potenza disponibile di uscita risulta

$$\mathcal{W}_{d_u}(f) = \beta(f) \cdot G_{d_{Max}}(f) \mathcal{W}_{d_g}(f) = \frac{4R_g(f)R_i(f)}{|Z_g(f) + Z_i(f)|^2} \cdot G_{d_{Max}}(f) \mathcal{W}_{d_g}(f)$$

- Nel caso infine in cui il carico $Z_c(f)$ in uscita alla rete non sia adattato, quest'ultimo assorbe una potenza inferiore a $\mathcal{W}_{d_u}(f)$ e pari a (vedi Appendice 12.8.2)

$$\mathcal{W}_c(f) = \alpha(f) \cdot \mathcal{W}_{d_u}(f) = \frac{4R_u(f)R_c(f)}{|Z_u(f) + Z_c(f)|^2} \cdot \mathcal{W}_{d_u}(f)$$

Reti passive Se una rete non contiene elementi attivi, allora $G_{d_{Max}}(f) \leq 1$ per qualunque f . In questo caso si parla più propriamente di *attenuazione disponibile* $A_d(f) = \frac{1}{G_d(f)}$ ovvero $A_d(f) [\text{dB}] = -G_d(f) [\text{dB}]$.

Reti in cascata Se più reti sono connesse tra loro l'una di seguito all'altra, e si verificano le condizioni di massimo trasferimento di potenza, il guadagno disponibile complessivo è il prodotto dei singoli guadagni disponibili: $G_{dTot} = G_{d1} \cdot G_{d2} \cdot \dots \cdot G_{dN}$.

Collegamento radio Con riferimento al circuito equivalente per una coppia di antenne, puntualizziamo che la potenza trasmessa è quella assorbita dall'impedenza di ingresso dell'antenna trasmittente, mentre quella ricevuta è quella ceduta dal generatore equivalente dell'antenna ricevente, all'impedenza di ingresso del ricevitore.

12.6 Distorsioni lineari

Sono quelle derivanti dal passaggio del segnale attraverso un canale che non è perfetto, ossia per il quale $H(f)$ non ha modulo costante e/o fase lineare.

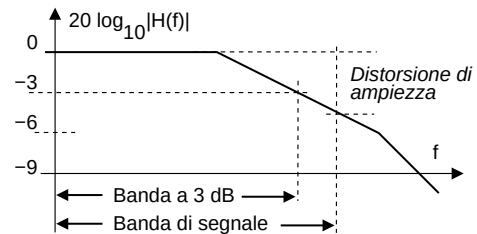
12.6.1 Rappresentazioni di $H(f) = |H(f)| e^{j\varphi(f)}$

Anzichè fornire una espressione analitica di modulo e fase, sono spesso utilizzate rappresentazioni grafiche di:

Guadagno in dB Il valore

$$G_{dB}(f) = 10 \log_{10} |H(f)|^2 = 20 \log_{10} |H(f)|$$

esprime il guadagno di potenza in deciBel. Nel caso in cui il segnale in transito occupi una banda entro la quale $G_{dB}(f)$ non sia costante, l'escursione di $G_{dB}(f)$ nella banda stessa è indicata come *distorsione lineare di ampiezza*, e quantificata appunto in *dB*. Al contrario, qualora si specifichi quale debba essere la massima distorsione lineare in *dB*, mediante il grafico si individua la banda entro la quale $G_{dB}(f)$ si mantiene all'interno della fascia consentita, ottenendo così il valore della banda per distorsione assegnata (ad esempio, la *banda a 3 dB*, corrispondente alla frequenza di taglio, è definita per questa via).



Tempo di ritardo di gruppo Il valore

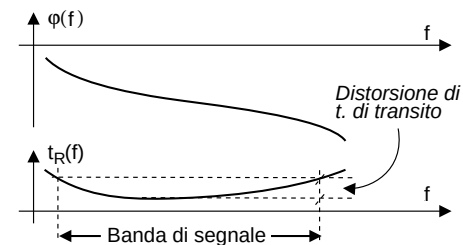
$$t_R(f) = -\frac{1}{2\pi} \frac{d}{df} \varphi(f)$$

rappresenta il ritardo di transito di un gruppo di frequenze molto stretto attorno ad f (al limite una sola frequenza). Ad esempio, un $\cos[2\pi ft]$ diviene

$$\cos[2\pi f(t - t_R(f))] = \cos[2\pi ft - 2\pi f t_R(f)]$$

ossia con uno sfasamento $\varphi(f) = 2\pi f t_R(f)$.

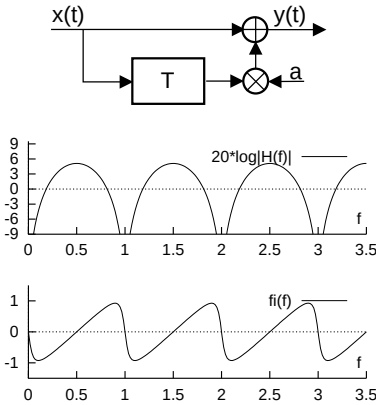
Se $t_R(f)$ è costante per qualunque f , allora tutte le componenti spettrali mantengono, all'uscita di $H(f)$, le stesse relazioni di fase che presentano in ingresso. In tal caso,



se anche $G_{dB}(f)$ si mantiene costante nella banda di segnale, siamo nelle condizioni di canale perfetto.

Se $t_R(f)$ non è costante, si ha una *distorsione lineare di fase* o di *tempo di transito*. Tale distorsione viene quantificata in termini di millisecondi (o μ , o nano) di massima differenza tra i tempi di transito nella banda di interesse.

Esempio



Il filtro trasversale in figura rappresenta un collegamento radio in cui sia presente una eco dovuta a riflessione. Risulta⁹:

$$|H(f)|^2 = 1 + a^2 + 2a \cos 2\pi fT$$

e dunque è presente sia distorsione lineare di ampiezza che di fase (mostrate in figura per $a = .8$ e $T = 1$).

In particolare, $|H(f)|^2$ è periodica di periodo $f = \frac{1}{T}$ e dunque può produrre una forte attenuazione (per $a \simeq 1$) anche a frequenze elevate. Se poi T cambia (perché si sposta il corpo riflettente, oppure si spostano trasmettitore o ricevitore) allora si ha a che fare con un canale tempo-variante, il cui studio non è per ora affrontato.

12.6.2 Segnali di banda base

E' noto che l'orecchio umano non è sensibile allo spettro di fase¹⁰ e dunque distorsioni di fase non modificano la qualità del segnale audio - mentre alterazioni di modulo sì.

Se il segnale è numerico, la risposta di fase è importante, perché altrimenti i diversi ritardi di fase alterano l'arrivo delle componenti spettrali dell'impulso elementare, che si deforma e perde la caratteristica di Nyquist. In conseguenza, insorge il fenomeno di interferenza intersimbolica (ISI) e aumenta la probabilità di errore.

12.6.3 Segnali modulati

Abbiamo mostrato a pag. 8.3.1 che l'involuppo complesso $\underline{y}(t) = \frac{1}{2}\underline{x}(t) * \underline{h}(t)$ presenta C.A di B.F. pari a

$$\begin{cases} y_c(t) = \frac{1}{2} [x_c(t) * h_c(t) - x_s(t) * h_s(t)] \\ y_s(t) = \frac{1}{2} [x_s(t) * h_c(t) + x_c(t) * h_s(t)] \end{cases}$$

Si è anche osservato come, nel caso in cui $H(f)$ risulti a simmetria coniugata rispetto ad f_0 , si ottiene $h_s(t) = 0$. In tal caso, l'effetto si riduce a quello che si avrebbe in banda base, filtrando il messaggio modulante mediante $h_c(t)$. Pertanto, l'effetto può essere rimosso *equalizzando* i segnali di banda base $y_c(t)$ e $y_s(t)$ con un filtro $H_e(f) = \frac{ae^{j2\pi f\tau}}{H_c(f)}$.

⁹L'espressione di $|H(f)|^2$ è stata ricavata al capitolo 7. Per la fase (mostrata in figura), osservando che $H(f) = 1 + ae^{-j2\pi fT}$ e che $\varphi(f) = \arctan \frac{\Im}{\Re}$, si ottiene $\varphi(f) = \arctan \frac{a \sin 2\pi fT}{1 + \cos 2\pi fT}$.

¹⁰Al contrario, è sensibile alle sue variazioni: queste ultime sono infatti elaborate dal cervello per estrarne informazioni relative al "movimento" dei suoni percepiti. Confrontando i ritardi differenti e variabili dei segnali pervenuti alle orecchie, si può comprendere se la sorgente degli stessi è in movimento.

12.6.3.1 Segnali a banda stretta

Nel caso in cui un segnale modulato $x(t)$ presenti una occupazione di frequenza molto piccola rispetto alla frequenza portante, si assume spesso che $H(f)$ nella banda di segnale non vari di molto, ossia presenti sia modulo che fase pressoché costanti e pari al valore assunto per $f = f_0$, ossia $H(f) \simeq H(f_0) = G_0 e^{j\phi_0}$. Questa approssimazione permette di trascurare l'effetto delle distorsioni lineari, che in questo caso equivalgono ad una semplice rotazione degli assi dell'involuppo complesso, reversibile scegliendo opportunamente la fase della portante di demodulazione.

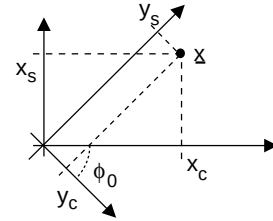
Nelle ipotesi poste, risulta infatti $\underline{H}(f) = 2H^+(f + f_0) = 2G_0 e^{j\phi_0}$ e quindi $\underline{h}(t) = \mathcal{F}^{-1}\{\underline{H}(f)\} = 2G_0 e^{j\phi_0} \delta(t)$; pertanto l'involuppo complesso di $h(t)$ è un impulso di area complessa $2G_0 e^{j\phi_0} = 2G_0 (\cos \phi_0 + j \sin \phi_0)$. All'uscita del canale $H(f)$ troviamo quindi

$$\begin{aligned} \underline{y}(t) &= \frac{1}{2} \cdot \underline{x}(t) * \underline{h}(t) = (x_c(t) + jx_s(t)) * G_0 (\cos \phi_0 + j \sin \phi_0) \delta(t) = \\ &= G_0 [(x_c(t) \cos \phi_0 - x_s(t) \sin \phi_0) + j(x_c(t) \sin \phi_0 + x_s(t) \cos \phi_0)] \end{aligned}$$

che identifica la trasformazione subita come una rotazione

$$\begin{bmatrix} y_c(t) \\ y_s(t) \end{bmatrix} = \begin{bmatrix} \cos \phi_0 & -\sin \phi_0 \\ \sin \phi_0 & \cos \phi_0 \end{bmatrix} \begin{bmatrix} x_c(t) \\ x_s(t) \end{bmatrix}$$

con una matrice dei coefficienti costante. Il risultato della rotazione è esemplificato in figura.



12.6.3.2 Modulazione di ampiezza

BLD-PS In questo caso $x_s(t) = 0$ ed allora $\begin{cases} y_c(t) = \frac{1}{2} [x_c(t) * h_c(t)] \\ y_s(t) = \frac{1}{2} [x_c(t) * h_s(t)] \end{cases}$. Adottando, ad esempio, una demodulazione omodina, si ottiene un segnale demodulato pari a $d(t) = x_c(t) * h_c(t)$, equivalente al caso di distorsione lineare di banda base.

BLD-PI Il problema maggiore con la portante intera può verificarsi se $H(f_0)$ è molto ridotto (ad esempio a causa di una attenuazione selettiva, esemplificata all'esempio precedente), perché in tal caso il demodulatore involuppo non funziona più.

BLU In questo caso il segnale modulato contiene ambedue le C.A. di B.F., e dunque la presenza di distorsioni lineari provoca il fenomeno noto come *intermodulazione tra componenti analogiche di bassa frequenza*, in quanto in entrambe ($y_c(t)$, $y_s(t)$), si trovano entrambe ($x_c(t)$, $x_s(t)$), mescolate tra loro tramite ($h_c(t)$, $h_s(t)$).

12.6.3.3 Modulazione angolare

Qualora un segnale modulato angolarmente attraversi un canale che presenta distorsioni lineari (di modulo, di fase, od entrambe), si verificano due fenomeni indicati come conversione PM-AM e PM-PM. Si manifesta infatti una modulazione di ampiezza sovrapposta, ed anche la modulazione di fase presenta delle alterazioni. Mentre la modulazione AM "parassita" può essere rimossa da un limitatore in ricezione, quella di fase no; inoltre quest'ultima presenta anche termini non-lineari e dunque non eliminabili mediante equalizzazione.

12.6.4 Calcolo dell'SNR

Se il segnale ricevuto ha subito distorsioni di ampiezza, lo spettro di densità di potenza si è deformato, ed il calcolo dell'SNR al ricevitore deve tenere conto dell'effetto filtrante introdotto dal canale attraversato, ossia:

$$SNR = \frac{\mathcal{P}_y}{\mathcal{P}_N} = \frac{\int_B \mathcal{P}_y(f) df}{\int_B \mathcal{P}_N(f) df} = \frac{\int_B \mathcal{P}_x(f) |H(f)|^2 df}{N_0 B}$$

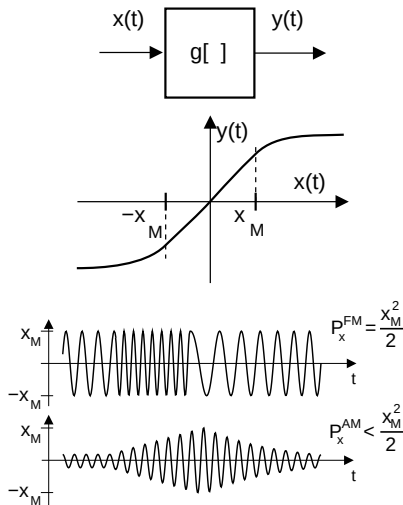
in cui si è indicata con B la banda a frequenze positive occupata dal segnale.

12.6.5 Equalizzazione

In una prossima edizione verrà svolta la trattazione relativa all'equalizzazione ottima.

12.7 Distorsioni di non linearità

Queste descrivono il deterioramento subito dal segnale nel transito attraverso un dispositivo dal comportamento non-lineare, per il quale cioè la relazione ingresso-uscita sia del tipo $y(t) = g[x(t)]$; espandendo la relazione ingresso-uscita $y = g[x]$, ed arrestando lo sviluppo al terzo ordine, otteniamo:



$$y(t) = G \left[x(t) + \alpha x^2(t) + \beta x^3(t) \right]$$

Un caso "tipico" di questo fenomeno si osserva allo stadio finale di un amplificatore di potenza che, per ampiezze del segnale di ingresso maggiori di $\pm x_M$, presenta fenomeni di *saturazione* dei valori di uscita. Già da questo esempio si nota l'importanza del coefficiente β , che tiene conto di distorsioni "cubiche" che meglio possono rappresentare il comportamento illustrato.

Preseguendo nell'analisi del caso dell'amplificatore che satura, osserviamo subito come, nel caso di segnali con modulazione AM, il fenomeno sia particolarmente grave: infatti questi segnali presentano valori di ampiezza direttamente dipendenti da quelli del segnale modulante. Per evitare di operare in regione non lineare, la potenza del segnale in ingresso all'elemento non lineare deve quindi essere ridotta (questa operazione è chiamata *back-off*), e conseguentemente la trasmissione avviene ad un livello di potenza $\mathcal{P}_x^{AM}$ inferiore a quello consentito ($\frac{x_M^2}{2}$) dall'amplificatore.

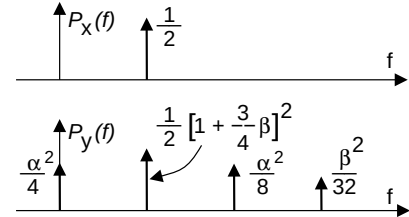
Nel caso di trasmissione FM, invece, abbiamo visto che il segnale modulato mantiene sempre la stessa ampiezza, e dunque si può effettuare la trasmissione a piena potenza; in altre parole, una volta fissato il livello di trasmissione, non occorre ricorrere ad un amplificatore sovradimensionato. Esaminiamo ora come quantificare l'effetto delle distorsioni non lineari.

12.7.1 Ingresso sinusoidale

Ponendo $x(t) = A \cos \omega_0 t$ si ottiene¹¹

$$\begin{aligned} y(t) &= G \left[A \cos \omega_0 t + \alpha A^2 \cos^2 \omega_0 t + \beta A^3 \cos^3 \omega_0 t \right] = \\ &= GA \left[\frac{\alpha A}{2} + \left(1 + \frac{3}{4} \beta A^2 \right) \cos \omega_0 t + \frac{\alpha A}{2} \cos 2\omega_0 t + \frac{\beta A^2}{4} \cos 3\omega_0 t \right] \end{aligned}$$

che corrisponde allo spettro di densità di potenza disegnato a lato (che è uno spettro *unilatero*, e calcolato per $A = G = 1$). Si osserva subito la comparsa di termini a frequenza multipla di quella di ingresso (oltre che di una componente continua). In particolare, non conoscendo a priori i valori di α e β , le relazioni ottenute sono usate per misurare l'entità delle distorsioni ponendo in ingresso una sinusoide, e misurando i valori di potenza in uscita alle diverse armoniche.



Fattori di intermodulazione Le caratteristiche tecniche che accompagnano gli amplificatori riportano, invece di α e β , i valori dei *fattori di intermodulazione* μ_2 e μ_3 (detti di seconda e di terza armonica), ottenuti utilizzando un ingresso sinusoidale e misurando le potenze $\mathcal{P}_I$, $\mathcal{P}_{II}$ e $\mathcal{P}_{III}$ alla frequenza in ingresso ed la sua seconda e terza armonica, e derivando da queste le quantità

$$\mu_2^2 = \frac{\mathcal{P}_{II}}{\mathcal{P}_I^2} \quad \text{e} \quad \mu_3^2 = \frac{\mathcal{P}_{III}}{\mathcal{P}_I^3}$$

e che sono legate ai coefficienti α e β dalle relazioni $\mu_2 = \frac{\alpha}{\sqrt{2}G}$; $\mu_3 = \frac{\beta}{2G^2}$ come mostrato alla nota¹².

Scrivendo $\mathcal{P}_{II} = \mu_2^2 \mathcal{P}_I^2$ e $\mathcal{P}_{III} = \mu_3^2 \mathcal{P}_I^3$, osserviamo che per piccoli valori di $\mathcal{P}_I$, la distorsione prodotta sia da $\mathcal{P}_{II}$ che da $\mathcal{P}_{III}$ è trascurabile; all'aumentare di $\mathcal{P}_I$, $\mathcal{P}_{II}$ cresce con il quadrato, mentre $\mathcal{P}_{III}$ con il cubo, e pertanto è quest'ultima componente che poi predomina.

12.7.2 Ingresso aleatorio

Nel caso in cui l'ingresso dell'elemento non lineare sia un processo gaussiano, la densità spettrale in uscita può ottenersi $\mathcal{F}$ -trasformando la funzione di autocorrelazione del-

¹¹Si fa uso delle relazioni $\cos^2 \alpha = \frac{1}{2} + \frac{1}{2} \cos 2\alpha$ e $\cos^3 \alpha = \frac{3}{4} \cos \alpha + \frac{1}{4} \cos 3\alpha$.

¹²Le relazioni mostrate si ottengono scrivendo

$$\begin{aligned} \mathcal{P}_I &= \frac{G^2 A^2}{2} \left(1 + \frac{3}{4} \beta A^2 \right)^2 \simeq \frac{G^2 A^2}{2} \quad \left(\text{se } \beta \ll \frac{4}{3A^2} \right) \\ \mathcal{P}_{II} &= \frac{G^2 A^4 \alpha^2}{8} = \frac{G^4 A^4}{4} \frac{1}{G^2} \frac{\alpha^2}{2} = \mathcal{P}_I^2 \mu_2^2 \\ \mathcal{P}_{III} &= \frac{G^2 A^6 \beta^2}{32} = \frac{G^6 A^6}{8} \frac{1}{G^4} \frac{\beta^2}{4} = \mathcal{P}_I^3 \mu_3^2 \end{aligned}$$

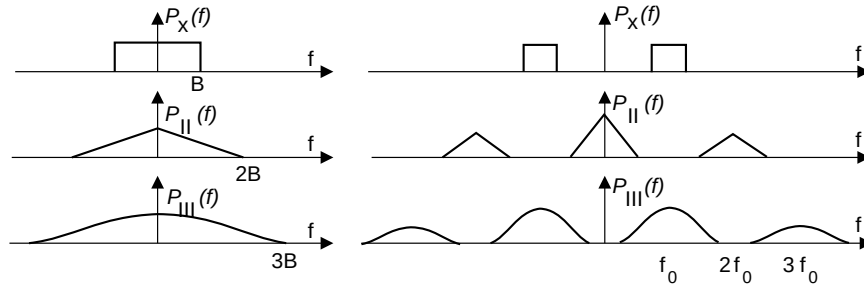


Figura 12.1: Densità spettrale per non linearità; a sin. per banda base, a ds per segnale modulato

l'uscita: in virtù di alcune proprietà¹³ dei momenti di variabili aleatorie gaussiane, si ottiene in questo caso:

$$\mathcal{P}_{II}(f) = G^2 2\alpha^2 \mathcal{P}_x(f) * \mathcal{P}_x(f); \quad \mathcal{P}_{III}(f) = G^2 6\beta^2 \mathcal{P}_x(f) * \mathcal{P}_x(f) * \mathcal{P}_x(f)$$

ovvero compaiono termini di distorsione di “2^a e 3^a armonica” che hanno origine dalla convoluzione della densità di potenza del segnale utile con se stesso.

Nel caso in cui il processo $x(t)$ sia limitato in banda contigua all'origine, i termini $\mathcal{P}_{II}(f)$ e $\mathcal{P}_{III}(f)$ hanno una banda rispettivamente doppia e tripla di quella di $\mathcal{P}_x(f)$ (vedi fig. 12.1). Nel caso di segnali modulati, oltre ad un allargamento di banda, avviene un fatto diverso e degno di commento: $\mathcal{P}_{II}(f)$ giace in bande diverse da quelle di $\mathcal{P}_x(f)$, pertanto può essere non considerato fonte di disturbo. $\mathcal{P}_{III}(f)$ invece ha una componente anch'essa centrata su f_0 , e dunque è solo questa la fonte disturbo.

La potenza totale delle due componenti di disturbo risulta inoltre pari a

$$\mathcal{P}_{II} = 4\mu_2^2 \mathcal{P}_I^2 \quad \text{e} \quad \mathcal{P}_{III} = 32\mu_3^2 \mathcal{P}_I^3$$

In definitiva, vi sono almeno tre buone ragioni per tenere d'occhio il valore di β , che è causa delle distorsioni di terza armonica:

1. è il coefficiente che tiene conto dei fenomeni di saturazione;
2. produce interferenza “in banda” per i segnali modulati;
3. produce interferenza “fuori banda” che danneggia le trasmissioni a frequenza tripla.

¹³Dato che l'uscita ha espressione $y(t) = G[x(t) + \alpha x^2(t) + \beta x^3(t)]$, il calcolo di $\mathcal{R}_y(\tau) = E\{y(t)y(t+\tau)\}$ si sviluppa calcolando i momenti misti $m_x^{(i,j)}(\tau) = E\{x^i(t)x^j(t+\tau)\}$. Se $x(t)$ è un processo gaussiano a media nulla, accade che $m_x^{(i,j)}(\tau) = 0$ se $i+j$ è dispari, mentre in caso contrario si applica il risultato per il valore atteso del prodotto di più v.a. estratte in tempi diversi:

$$E\{x_1 \cdot x_2 \cdot \dots \cdot x_n\} = \sum (E\{x_{p_1} \cdot x_{p_2}\} \cdot E\{x_{p_3} \cdot x_{p_4}\} \cdot \dots \cdot E\{x_{p_{n-1}} \cdot x_{p_n}\})$$

in cui la somma è estesa a tutte le possibili permutazioni non equivalenti di $(1, 2, \dots, n)$ (sono equivalenti se accoppiano con ordine diverso o in in posizione diversa le stesse v.a.). Ad esempio, per quattro v.a. si ha:

$$E\{x_1 \cdot x_2 \cdot \dots \cdot x_n\} = E\{x_1 \cdot x_2\} \cdot E\{x_3 \cdot x_4\} + E\{x_1 \cdot x_3\} \cdot E\{x_2 \cdot x_4\} + E\{x_1 \cdot x_4\} \cdot E\{x_2 \cdot x_3\}$$

12.7.3 Effetto sulla modulazione FM

Se consideriamo un segnale $x(t) = \cos[\omega_0 t + \varphi(t)]$, l'effetto della non linearità produce il segnale

$$y(t) = G \left[\frac{\alpha}{2} + \left(1 + \frac{3}{4}\beta \right) \cos[\omega_0 t + \varphi(t)] + \frac{\alpha}{2} \cos[2\omega_0 t + 2\varphi(t)] + \frac{\beta}{4} \cos[3\omega_0 t + 3\varphi(t)] \right]$$

Osserviamo che i termini a frequenza $2\omega_0$ e $3\omega_0$, nonché il livello in continua, possono essere eliminati mediante un filtro passa-banda centrato in $f = f_0 = \frac{\omega}{2\pi}$; dopo tale operazione, la modulazione di fase $\varphi(t)$ è proprio quella impressa dal modulatore, e pertanto i fenomeni non lineari *non hanno conseguenze sulla FM* (tranne che per le interferenze causate ai canali vicini !)

Il risultato appena illustrato è stato sfruttato nei ponti radio progettati per trasmettere un segnale FDM in FM. Si usa un basso indice di modulazione (risparmiando banda) e si trasmette a piena potenza (senza backoff). La potenza del segnale modulato *non dipende* dal numero di canali contemporaneamente attivi.

12.8 Appendici

12.8.1 Potenza assorbita da un bipolo

La dimostrazione inizia definendo una potenza istantanea assorbita dal bipolo come $w(t) = v(t)i(t) = v(t) \cdot (v(t) * y(t))$. La potenza media allora risulta

$$\begin{aligned} \mathcal{W}_z &= \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} w(t) dt = \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} v(t) \left[\int_{-\infty}^{\infty} v(t-\tau) y(\tau) d\tau \right] dt = \\ &= \int_{-\infty}^{\infty} y(\tau) \left[\lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} v(t) v(t-\tau) dt \right] d\tau = \int_{-\infty}^{\infty} y(\tau) \mathcal{R}_v(-\tau) d\tau = \\ &= \int_{-\infty}^{\infty} Y(f) \mathcal{P}_v(f) df = \int_{-\infty}^{\infty} [\Re\{Y(f)\} + j\Im\{Y(f)\}] \mathcal{P}_v(f) df = \\ &= \int_{-\infty}^{\infty} \mathcal{P}_v(f) \frac{R(f)}{|Z(f)|^2} df \end{aligned}$$

Nel terzultimo passaggio si è fatto uso del teorema di Parseval, e del fatto che $\mathcal{R}_v(\tau)$ è pari; nell'ultimo, si è tenuto conto che $\mathcal{P}_v(f)$, $R(f)$ e $|Z(f)|^2 = R^2(f) + X^2(f)$ sono funzioni pari di f , mentre $X(f)$ è dispari: pertanto il termine $\int_{-\infty}^{\infty} \Im\{Y(f)\} \mathcal{P}_v(f) df = \int_{-\infty}^{\infty} \mathcal{P}_v(f) \frac{X(f)}{|Z(f)|^2} df$ è nullo. Notiamo che quest'ultimo termine rappresenta la *potenza reattiva*, che non è trasformata in altre forme di energia, e viene accumulata e restituita dalla componente reattiva del carico. Al contrario, il termine relativo a $\Re\{Y(f)\}$ rappresenta la potenza assorbita dalla componente resistiva, nota come *potenza attiva*, che viene completamente dissipata.

Avendo espresso la potenza assorbita $\mathcal{W}_z$ nella forma di un integrale in f , la funzione integranda è intuitivamente associabile allo spettro di densità di potenza: $\mathcal{W}_z(f) = \mathcal{P}_v(f) \frac{R(f)}{|Z(f)|^2}$. Lo stesso risultato può essere confermato svolgendo il seguente calcolo più diretto, pensando al bipolo come ad un filtro la cui grandezza di ingresso è $v(t)$ e quella di uscita $i(t)$.

La definizione di potenza media $\mathcal{W}_z = \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} w(t) dt$, in cui $w(t) = v(t)i(t)$, mostra come questa sia equivalente alla funzione di intercorrelazione tra i e v calcolata in $\tau = 0$: $\mathcal{W}_z = \mathcal{R}_{vi}(0)$. Allora, è ragionevole assumere che $\mathcal{W}_z(f) = \mathcal{F}\{\mathcal{R}_{vi}(\tau)\}$. Indicando infatti con $\otimes$ l'integrale di intercorrelazione, e ricordando che gli operatori di convoluzione e correlazione godono della proprietà commutativa, possiamo scrivere

$$\mathcal{R}_{vi}(\tau) = v(t) \otimes i(t) = v(t) \otimes (v(t) * y(t)) = (v(t) \otimes v(t)) * y(t) = \mathcal{R}_v(\tau) * y(t)$$

quindi, risulta che

$$\mathcal{W}_z(f) = \mathcal{F}\{\mathcal{R}_{vi}(\tau)\} = \mathcal{P}_v(f) Y(f) = \mathcal{P}_v(f) \frac{R(f) - jX(f)}{|Z(f)|^2}$$

In base alle stesse considerazioni già svolte, si verifica come il termine immaginario non contribuisce alla potenza media assorbita, e quindi può essere omesso dalla definizione di *potenza attiva* $\mathcal{W}_z(f)$.

12.8.2 Potenza ceduta ad un carico $Z_c(f) \neq Z_g^*(f)$

Avendo a disposizione un generatore di segnale di potenza disponibile $\mathcal{W}_d(f)$ ed impedenza interna $Z_g(f)$ assegnate, la tensione a vuoto ai suoi capi ha densità di potenza (di segnale) pari a $\mathcal{P}_v(f) = \mathcal{W}_d(f) 4R_g(f)$. Collegando al generatore un carico generico $Z_c(f)$, la potenza dissipata da quest'ultimo risulta pari a $\mathcal{W}_{z_c}(f) = \mathcal{P}_v(f) \frac{R_c(f)}{|Z_g(f) + Z_c(f)|^2}$. Il rapporto tra la densità di potenza effettivamente ceduta a $Z_c(f)$, e quella che le sarebbe ceduta se questa fosse adattata per il massimo trasferimento di potenza, fornisce la perdita di potenza subita:

$$\frac{\mathcal{W}_{z_c}(f)}{\mathcal{W}_d(f)} = \mathcal{W}_d(f) 4R_g \frac{R_c(f)}{|Z_g(f) + Z_c(f)|^2} \cdot \frac{1}{\mathcal{W}_d(f)} = \frac{4R_g(f) R_c(f)}{|Z_g(f) + Z_c(f)|^2} = \alpha(f)$$

Pertanto, se $Z_c(f) \neq Z_g^*(f)$, su $Z_c(f)$ si dissipa una potenza pari a $\mathcal{W}_{z_c}(f) = \alpha(f) \mathcal{W}_d(f)$. Il medesimo risultato è valido anche per l'analisi dell'accoppiamento tra il generatore equivalente di uscita di una rete due porte ed un carico.

Esempio Consideriamo un generatore con $Z_g(f)$ resistiva e pari a 50Ω , e con densità di potenza disponibile (a frequenze positive)

$$\mathcal{W}_d^+(f) = \frac{\mathcal{W}_d}{4W} \text{rect}_{2W}(f - f_0)$$

in cui $\mathcal{W}_d = 1$ Watt è la potenza disponibile totale, distribuita uniformemente in una banda $2W = 10$ KHz centrata a frequenza $f_0 = 1$ MHz. Il generatore è collegato ad un carico

$$Z_c(f) = R_c(f) + jX_c(f)$$

con $R_c(f) = 50 \Omega$ ed $X_c(f) = 2\pi fL = 50 \Omega$ per $f = f_0$ (da cui $L = \frac{50}{2\pi 10^6} = 7.96 \mu H$).

Essendo la banda di segnale $2W \ll f_0$, approssimiamo la dipendenza da f di $X_c(f)$ come una costante. In queste ipotesi, la potenza effettivamente ceduta al carico risulta $\mathcal{W}_{z_c} = \alpha \mathcal{W}_d$, con

$$\alpha = \frac{4R_g R_c}{|Z_g + Z_c|^2} = \frac{4 \cdot 50 \cdot 50}{|50 + 50 + j50|^2} = \frac{10000}{12500} = 0.8$$

e quindi $\mathcal{W}_{z_c} = 0,8$ Watt.

Il valore $\alpha_{dB} = 10 \log_{10} \alpha = 0,97$ dB rappresenta il valore della perdita di potenza causata dal mancato verificarsi delle condizioni di massimo trasferimento di potenza, e può essere tenuto in conto come una attenuazione supplementare al collegamento, in fase di valutazione del *link budget* (vedi capitolo 13).

Capitolo 13

Mezzi Trasmissivi

In questo capitolo sono caratterizzati i mezzi trasmissivi cavo, radio e fibra ottica, applicando agli stessi le considerazioni fin qui svolte. Le caratteristiche dei diversi mezzi intervengono nella analisi del dimensionamento dei sistemi, e quando possibile sono discusse tecniche particolari di realizzazione.

13.1 Dimensionamento di un collegamento

Introduciamo subito alcune definizioni e concetti, che aiutano a meglio analizzare il problema del dimensionamento di un collegamento. I parametri fondamentali del collegamento sono espressi dalla potenza disponibile del trasmettitore W_{dT} , dalla minima potenza che occorre ricevere W_{RMin} (spesso indicata come *sensibilità*) e dall'attenuazione disponibile A_d del mezzo di trasmissione che si intende utilizzare.

Guadagno di Sistema

Il rapporto $G_s = \frac{W_{dT}}{W_{RMin}}$ ⁽¹⁾ è detto *Guadagno di sistema* e rappresenta il massimo valore di attenuazione A_d che è possibile superare. La differenza in decibel $G_{sdB} = W_{dT} [dBW] - W_{RMin} [dBW]$ rappresenta la stessa quantità, in una forma che rende più intuitivo il suo utilizzo nel determinare un limite alla massima attenuazione disponibile: deve infatti risultare

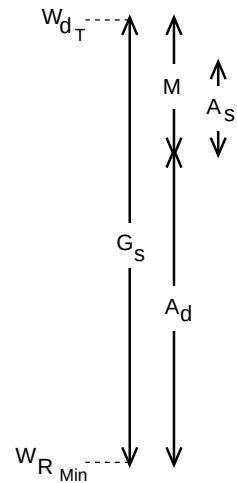
$$A_{dB} \leq G_{sdB}$$

Margine di sistema

La differenza tra G_{sdB} ed A_{dB} , che per quanto appena detto deve risultare ≥ 0 , prende il nome di *Margine di sistema*, e rappresenta l'eccesso di potenza (in dB) che viene trasmessa, rispetto alla minima indispensabile:

$$M_{dB} = G_{sdB} - A_{dB}$$

¹Notiamo che G_s è definito come ingresso/uscita, contrariamente agli altri guadagni. Infatti, non è una *grandezza* del collegamento, bensì una *potenzialità* dello stesso.



Attenuazione supplementare L'eccesso di potenza M_{dB} deve comunque risultare maggiore della somma (in dB) di tutte le possibili ulteriori cause di attenuazione del segnale, indicate collettivamente come *attenuazioni supplementari*:

$$\sum A_{s_{dB}} \leq M_{dB}$$

In questa categoria rientrano tutte le cause di attenuazione non previste nella situazione ideale e che possono, ad esempio, avere origine dal fallimento delle condizioni per il massimo trasferimento di potenza, oppure essere causate da un fenomeno piovoso in un collegamento radio, o dipendere dalla perdita di segnale dovuta alla giunzione tra tratte in fibra ottica....

Grado di servizio Nel capitolo 5 il concetto di grado di servizio è stato associato al valore di probabilità con cui può verificarsi un fenomeno di blocco in un elemento di commutazione. Un concetto del tutto analogo sussiste, qualora le attenuazioni supplementari siano grandezze aleatorie, e la loro somma possa superare il valore del margine a disposizione: in tal caso, la potenza ricevuta si riduce sotto la minima $W_{R_{Min}}$, ed il collegamento “*va fuori specifiche*”. Pertanto, in sede di dimensionamento di un collegamento, indichiamo con grado di servizio la percentuale di tempo per la quale si mantiene $W_R > W_{R_{Min}}$, ovvero la probabilità che le attenuazioni supplementari *non* superino il margine, ossia

$$\text{Grado di Servizio} = Pr \left\{ \sum A_{s_{dB}} < M_{dB} \right\}$$

Esempio: Un grado di servizio del 99.99 % equivale a poco meno di 1 ora l'anno di fuori servizio, e corrisponde a richiedere che $Pr \{ \sum A_{s_{dB}} > M_{dB} \} = 10^{-4}$.

13.1.1 Casi particolari

Segnale a banda stretta Nel caso in cui il segnale trasmesso sia un segnale modulato con banda $B \ll f_0$, le impedenze $Z_i(f)$ e $Z_u(f)$ non variano di molto per $|f - f_0| < \frac{B}{2}$, e pertanto le condizioni di adattamento di impedenza sono approssimate ponendo $Z_g(f) = Z_i(f_0)$ e $Z_c(f) = Z_u(f_0)$.

In alternativa, si può scegliere di realizzare le condizioni di massimo trasferimento di potenza, subendo una distorsione di fase praticamente costante, che causa una rotazione tra le C.A. di B.F. del segnale modulato.

Equalizzazione Qualora la funzione di trasferimento complessiva del mezzo trasmissivo e delle sue chiusure $H(f) = H_i(f) H_q(f) H_u(f)$ non soddisfi la condizione di canale perfetto, l'equalizzazione si ottiene inserendo elementi filtranti presso il trasmettitore ($H_T(f)$) ed il ricevitore ($H_R(f)$), in modo da realizzare $H_T(f) H(f) H_R(f) = ae^{-j2\pi f\tau}$.

Se non sono presenti entrambi i filtri, ma uno solo di essi, si può scegliere di utilizzare solo $H_T(f)$, riservando ad $H_R(f)$ il solo scopo di filtrare il rumore esterno alla banda di segnale.

La scelta opposta, permette invece di realizzare $H_R(f)$ in modo adattativo, modificandone cioè l'andamento senza conoscere l' $H(f)$ da correggere², sfruttando conoscenze

²L'operazione può rendersi necessaria qualora il canale cambi da un collegamento all'altro, come ad esempio nel transito in una rete commutata, od in una comunicazione radiomobile.

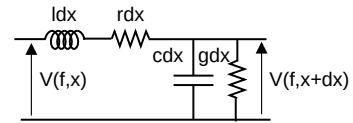
a riguardo della densità di potenza del segnale in arrivo, od in altri modi³. D'altra parte, attribuire il compito dell'equalizzazione ad $H_R(f)$, può causare una "colorazione" del rumore in ingresso al ricevitore, il cui effetto dovrebbe essere analizzato per verificarne le conseguenze sul segnale ricevuto.

13.2 Collegamenti in cavo

La descrizione completa delle caratteristiche e delle prestazioni dei cavi in rame è una materia molto vasta, di cui forniamo di seguito solo alcuni risultati, strettamente legati agli aspetti di telecomunicazione, il più rilevante dei quali è senz'altro la manifestazione dell'effetto pelle, che determina (per $f > 100$ KHz) una attenuazione in dB proporzionale a $\sqrt{f}$. La sezione è completata da una breve catalogazione dei cavi usati per telecomunicazioni.

13.2.1 Costanti distribuite, grandezze derivate, e condizioni generali

Un conduttore elettrico uniforme e di lunghezza infinita, è descritto in base ad un modello a costanti distribuite, espresso in termini delle *costanti primarie* costituite dalla resistenza r , la conduttanza g , la capacità c e l'induttanza l per unità di lunghezza. La teoria delle linee uniformi definisce quindi due grandezze derivate dalle costanti primarie: l'*impedenza caratteristica* $Z_0(f)$ e la *costante di propagazione* $\gamma(f)$.



Impedenza caratteristica E' definita come

$$Z_0(f) = R_0(f) + jX_0(f) = \sqrt{\frac{r + j2\pi fl}{g + j2\pi fc}}$$

e rappresenta il rapporto tra $V(f)$ ed $I(f)$ in un generico punto del cavo, permettendo di scrivere $I(f) = \frac{V(f)}{Z_0(f)}$.

Costante di propagazione E' definita come

$$\gamma(f) = \alpha(f) + j\beta(f) = \sqrt{(r + j2\pi fl)(g + j2\pi fc)}$$

mentre la grandezza $e^{-\gamma(f)d}$ rappresenta il rapporto dei valori di tensione presenti tra due punti di un cavo di lunghezza infinita, distanti d , permettendo di scrivere: $V(f, x + d) = e^{-\gamma(f)d} V(f, x)$.

³Conoscendo la densità di potenza del segnale in arrivo, è possibile generare un segnale "differenza" tra quello che ci si aspetta e quel che invece si osserva, e controllare in modo automatico un filtro in modo da ridurre la differenza tra le densità di potenza al minimo.

Nel caso in cui occorra correggere una distorsione di fase (ad es. nelle trasmissioni numeriche), non ci si può basare sul solo spettro di potenza, e possono essere previste fasi di "apprendimento" dell'equalizzatore, durante le quali il segnale trasmesso è noto al ricevitore, e si può costruire un segnale differenza basato direttamente sulla forma d'onda.

Condizioni di chiusura Qualora il cavo di lunghezza d sia chiuso ai suoi estremi su di un generatore con impedenza $Z_g(f)$ e su di un carico $Z_c(f)$, risultano definiti i *coefficienti di riflessione* del generatore e del carico:

$$r_g(f) = \frac{Z_g(f) - Z_0(f)}{Z_g(f) + Z_0(f)} \quad \text{e} \quad r_c(f) = \frac{Z_c(f) - Z_0(f)}{Z_c(f) + Z_0(f)} \quad (13.1)$$

Osserviamo subito che nel caso in cui $Z_g(f) = Z_c(f) = Z_0(f)$, risulta $r_g(f) = r_c(f) = 0$.

Quadripolo equivalente L'impedenza vista dai morsetti di *ingresso* e di *uscita* di un cavo, interposto tra generatore e carico, vale rispettivamente

$$Z_i(f) = Z_0(f) \frac{1 + r_c(f) \cdot e^{-2d\gamma(f)}}{1 - r_c(f) \cdot e^{-2d\gamma(f)}} \quad \text{e} \quad Z_u(f) = Z_0(f) \frac{1 + r_g(f) \cdot e^{-2d\gamma(f)}}{1 - r_g(f) \cdot e^{-2d\gamma(f)}} \quad (13.2)$$

Allo stesso tempo, la funzione di trasferimento *intrinseca* risulta

$$H_q(f) = 2 \frac{e^{-d\gamma(f)}}{1 - r_g(f) \cdot r_c(f) \cdot e^{-2d\gamma(f)}} \quad (13.3)$$

Condizioni di adattamento Nel caso in cui $Z_g(f) = Z_c(f) = Z_0(f)$, come sappiamo, il quadripolo si comporta in modo perfetto. In tal caso, risultando $r_g(f) = r_c(f) = 0$, si ottiene che $Z_i(f) = Z_u(f) = Z_0(f)$ e $H_q(f) = \frac{V_q(f)}{V_i(f)} = 2e^{-d\gamma(f)}$: il cavo si comporta allora come se avesse lunghezza infinita. In tal caso, inoltre, risulta che $H_i(f) = \frac{1}{2}$ ed $R_g(f) = R_u(f)$; pertanto il guadagno disponibile risulta

$$G_d(f) = |H_i(f)|^2 |H_q(f)|^2 \frac{R_g(f)}{R_u(f)} = \frac{1}{4} |2e^{-d[\alpha(f) + j\beta(f)]}|^2 = e^{-2d\alpha(f)}$$

Condizione di Heaviside Nel caso in cui i valori delle costanti primarie sia tale da risultare $r \cdot c = l \cdot g$, note come *condizioni di Heaviside*, allora per la costante di propagazione si ottiene $\gamma(f) = \alpha(f) + j\beta(f) = \sqrt{rg} + j2\pi f\sqrt{lc}$: pertanto, $\alpha(f)$ è costante e $\beta(f)$ cresce linearmente con la frequenza, realizzando così le condizioni di un canale perfetto, in quanto $H_q(f) = 2e^{-d\alpha(f)}e^{-jd\beta(f)} = 2e^{-d\sqrt{rg}}e^{-jd2\pi f\sqrt{lc}}$.

Allo stesso tempo, l'impedenza caratteristica $Z_0(f) = R_0(f) + jX_0(f) = \sqrt{\frac{r}{g}} = \sqrt{\frac{l}{c}} = R_0$ è solo resistiva ed indipendente dalla frequenza, rendendo semplice realizzare la condizione di adattamento $Z_g(f) = Z_c(f) = R_0$, che determina al contempo anche il massimo trasferimento di potenza.

In definitiva, la funzione di trasferimento complessiva $H(f) = H_i(f) H_q(f) H_u(f)$ vale in questo caso $H(f) = \frac{1}{2}2e^{-d\sqrt{rg}}e^{-jd2\pi f\sqrt{lc}}\frac{1}{2} = \frac{1}{2}e^{-d\sqrt{rg}}e^{-jd2\pi f\sqrt{lc}}$, equivalente quindi ad un canale perfetto con guadagno $G = \frac{1}{2}e^{-d\sqrt{rg}}$ e ritardo $t_R = d\sqrt{lc}$; al contempo, l'attenuazione disponibile risulta indipendente da f , e pari a

$$A_d(f) = 1/G_d(f) = e^{2d\sqrt{rg}}$$

13.2.2 Trasmissione in cavo

In generale, le costanti primarie del cavo non soddisfano le condizioni di Heaviside, e le impedenze di chiusura non sono adattate. In tal caso si ha $r_g(f) \neq 0$ e/o $r_c(f) \neq 0$, e devono essere applicate le (13.2) e (13.3).

Cavo molto lungo Se il cavo è sufficientemente lungo da poter porre $e^{-2d\gamma(f)} \ll 1$, ossia $|e^{-2d\gamma(f)}| = e^{-2d\alpha(f)} \ll 1$, le (13.2) divengono $Z_i(f) = Z_u(f) \simeq Z_0(f)$, mentre la (13.3) si semplifica in $H_q(f) = 2e^{-d\gamma(f)}$; nel caso generale risulta pertanto

$$G_d(f) = |H_q(f)|^2 \cdot |H_i(f)|^2 \cdot \frac{R_g(f)}{R_u(f)} = 4 \cdot e^{-2d\alpha(f)} \cdot |H_i(f)|^2 \cdot \frac{R_g(f)}{R_u(f)}$$

che mostra due cause di distorsione lineare.

La prima dipende dal disadattamento di impedenze in ingresso ed uscita. Qualora si realizzi invece la condizione $Z_g(f) = Z_c(f) = Z_0(f)$, si ottiene

$$A_d(f) = \frac{1}{G_d(f)} = e^{2d\alpha(f)}$$

La seconda causa di distorsione dipende dal comportamento non perfetto di $H_q(f) = 2e^{-d\gamma(f)}$, che in teoria può essere completamente corretto, solo nel caso in cui le costanti primarie soddisfino le condizioni di Heaviside. In pratica, però, il risultato è diverso, perchè.... le “costanti primarie” *non sono* costanti !!!

Effetto pelle Si tratta di un fenomeno legato all’addensamento del moto degli elettroni verso la superficie del cavo, al crescere della frequenza. Per questo motivo, si riduce la superficie del conduttore realmente attraversata da corrente elettrica, a cui corrisponde un aumento della resistenza per unità di lunghezza r . Si può mostrare che, per frequenze maggiori di 50-100 KHz, la resistenza per unità di lunghezza r aumenta proporzionalmente a $\sqrt{f}$, e quindi si può scrivere $\alpha(f) = \alpha_0\sqrt{f}$, in cui la costante α_0 dipende dal tipo di cavo.

In tali condizioni, l’attenuazione disponibile risulta $A_d(f) = e^{2d\alpha(f)} = e^{2d\alpha_0\sqrt{f}}$, a cui corrisponde un valore in dB pari a

$$A_d(f)|_{dB} = 10 \log_{10} e^{2d\alpha_0\sqrt{f}} = d\sqrt{f} \cdot \alpha_0 10 \log_{10} e^2 = A_0 \cdot d \cdot \sqrt{f}$$

Il valore A_0 riassume in sé tutte le costanti coinvolte, prende il nome di *attenuazione chilometrica*, ed è espresso in dB/Km, ad una determinata frequenza (ad es. 1 MHz). Pertanto, poiché nell’applicare la formula occorre mantenere congruenza dimensionale, si ottiene in definitiva:

$$A_d(f)|_{dB} = A_0(f_R) \cdot d_{Km} \cdot \sqrt{\frac{f}{f_R}}$$

in cui f_R rappresenta la frequenza di riferimento per la quale è disponibile il valore A_0 . Questo risultato può essere usato come formula di progetto, e mette in evidenza come l’attenuazione in dB dei cavi sia linearmente proporzionale alla lunghezza⁴.

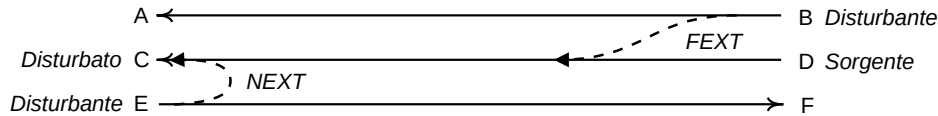
⁴Questa circostanza è comune con le trasmissioni in fibra ottica, ed è legato alla presenza nel mezzo di una componente dissipativa, in questo caso la resistenza.

Equalizzazione In presenza di effetto pelle, la funzione di trasferimento intrinseca $H_q(f) = 2e^{-d\gamma(f)}$ presenta una dipendenza da f tutt'altro che perfetta, causando potenzialmente distorsioni lineari sui segnali in transito. Un problema analogo sorge, anche in assenza di effetto pelle, qualora si manifesti un disadattamento di impedenze ed il cavo non sia sufficientemente lungo (vedi appresso).

Se la banda di segnale è sufficientemente estesa da causare una distorsione lineare non trascurabile, o se la particolare natura del segnale (ad es. numerico) richiede la presenza di un ritardo strettamente costante con f , è necessario prevedere uno stadio di equalizzazione.

Diafonia La diafonia, indicata in inglese con il termine di *crosstalk*, consiste nei fenomeni di interferenza tra i messaggi trasportati su cavi disposti in prossimità reciproca, e dovuti a fenomeni di induzione elettromagnetica ed accoppiamenti elettrostatici. Il fenomeno è particolarmente rilevante in tutti i casi in cui molti cavi giacciono affasciati in una medesima canalizzazione, condividendo un lunghezza significativa di percorso. Nel caso di comunicazioni foniche, la diafonia può causare l'ascolto indesiderato di altre comunicazioni⁵; nel caso di trasmissioni numeriche o di segnali modulati, la diafonia produce un disturbo additivo supplementare, che peggiora le prestazioni espresse in termini di probabilità di errore o di SNR.

Con riferimento allo schema di figura, consideriamo un collegamento D-C su cui gravano due cause di interferenza di diafonia: il collegamento da E ad F produce il fenomeno di *paradiafonia* (in inglese NEXT, *near end crosstalk*), mentre il collegamento da B ad A produce il fenomeno di *telediafonia* (FEXT, *far end crosstalk*). Nel primo caso, il segnale disturbante ha origine in prossimità del punto di prelievo del segnale disturbato, mentre nel secondo ha origine in prossimità del punto di immissione.



L'entità del disturbo è quantificata mediante un valore di attenuazione di diafonia tra le sorgenti disturbanti e l'estremo disturbato. La circostanza che, nei rispettivi punti di immissione, i segnali disturbanti hanno la stessa potenza della sorgente che emette il segnale disturbato, permette di definire lo *scarto di paradiafonia*

$$\Delta A_{EC}|_{dB} = A_{EC}|_{dB} - A_{DC}|_{dB}$$

come la differenza in dB tra l'attenuazione di *paradiafonia* $A_{EC}|_{dB}$ e l'attenuazione del collegamento $A_{DC}|_{dB}$. Il livello di potenza del segnale disturbante proveniente da E ed osservato al punto C risulta quindi pari a⁶ $W_E^{next} = W_E - A_{EC} = W_D - A_{EC} = W_C + A_{DC} - A_{EC} = W_C - \Delta A_{EC}$, ossia di ΔA_{EC} dB inferiore al segnale utile. Una definizione del tutto analoga risulta per la *telediafonia* (FEXT), per la quale il livello di potenza del segnale disturbante proveniente da B ed osservato al punto C risulta $W_B^{fext} = W_C - \Delta A_{BC}$ in cui lo *scarto di telediafonia* ha il valore

$$\Delta A_{BC}|_{dB} = A_{BC}|_{dB} - A_{DC}|_{dB}$$

⁵.. le famose interferenze telefoniche, di molto ridotte dall'avvento della telefonia numerica (PCM).

⁶Omettiamo di indicare di operare in dB per chiarezza di notazione.

13.2.2.1 Casi limite

Cavo a basse perdite E' un modello applicabile per tutte quelle frequenze per cui risulti $r \ll 2\pi fl$ e $g \ll 2\pi fc$. In tal caso si ha

$$Z_0(f) = R_0 = \sqrt{\frac{l}{c}} \text{ reale} \quad \text{e} \quad \gamma(f) = j2\pi f\sqrt{lc}$$

Di conseguenza, è facile realizzare $Z_g = Z_c = R_0$, che determina

$$H_q(f) = 2e^{-jd2\pi f\sqrt{lc}}$$

quindi il cavo non presenta distorsioni di ampiezza, ha una attenuazione trascurabile, e manifesta una distorsione di fase lineare in f , realizzando quindi le condizioni di canale perfetto.

Cavo corto E' il caso di collegamenti interni agli apparati, o tra un trasmettitore-ricevitore e la relativa antenna. La ridotta lunghezza del cavo permette di scrivere

$$e^{-d\gamma(f)} = e^{-d\alpha(f)}e^{-jd\beta(f)} \simeq e^{-jd\beta(f)}$$

in quanto $e^{-d\alpha(f)} \simeq 1$.

Qualora si manifesti un disadattamento di impedenze, i coefficienti di riflessione $r_g(f)$ e $r_c(f)$ risultano diversi da zero, rendendo

$$H_q(f) = 2 \frac{e^{-jd\beta(f)}}{1 - r_g(f) \cdot r_c(f) \cdot e^{-j2d\beta(f)}}$$

periodica con d e con f (quest'ultimo in assenza di effetto pelle). In particolare, se il carico viene sconnesso, o l'uscita del cavo posta in corto circuito, l'eq. (13.1) mostra come risulti $r_c(f) = \pm 1$ rispettivamente, e la prima delle (13.2) diviene

$$Z_i(f) = Z_0(f) \frac{1 \pm e^{-j2d\beta(f)}}{1 \mp e^{-j2d\beta(f)}}$$

e si vede che per quei valori (ricorrenti) di frequenza f che rendono $e^{-jd\beta(f)} = \pm 1$ ⁽⁷⁾, l'impedenza di ingresso del cavo può risultare infinita o nulla.

Evidentemente, le distorsioni lineari prodotte in questo caso hanno un andamento del tutto dipendente dalle particolari condizioni operative, e dunque la loro equalizzazione deve prevedere componenti in grado di adattarsi alla $H_q(f)$ del caso⁸. D'altra parte, una volta equalizzato il cavo, non sono necessari ulteriori aggiustamenti, a parte problemi di deriva termica. Diverso è il caso dal punto di vista di un terminale di rete, per il quale il cavo effettivamente utilizzato può essere diverso da collegamento a collegamento, e pertanto i dispositivi modem a velocità più elevate devono disporre di un componente di equalizzazione adattiva, da regolare ogni volta ad inizio del collegamento⁹.

⁷Ovvero, tali che $\cos 2d\beta(f) - j \sin 2d\beta(f) = \pm 1$, e quindi $d\beta(f) = k\frac{\pi}{2}$ con $k = 0, 1, 2, \dots$

⁸Può ad esempio rendersi necessario "tarare" un trasmettitore radio, la prima volta che lo si collega all'antenna.

⁹E' questa la fase in cui il modem che usiamo per collegarci al provider internet emette una serie di orribili suoni...

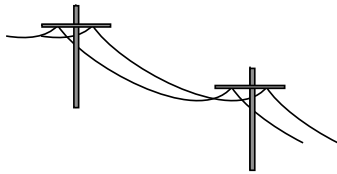
13.2.3 Tipologie di cavi per le telecomunicazioni

Descriviamo i principali tipi di cavo utilizzati. per i quali forniamo in tabella i valori tipici delle grandezze essenziali, nelle condizioni illustrate nel testo che segue.

Tipo di cavo	A_0 [dB/Km]	Z_0 [Ω]	r, g, l, c per 1 Km
Linee aeree	0.036 ad 1 KHz 0.14 a 100 KHz	600	$5 \cdot 10^{-6}, 2 \cdot 10^{-3}, 5 \cdot 10^{-9}$ ad 1 KHz
Coppie ritorte	1.2 ad 1 KHz 6 a 100 KHz 20 a 1 MHz	$600e^{-j\frac{\pi}{4}}$	$100, 5 \cdot 10^{-5}, 10^{-3}, 5 \cdot 10^{-8}$ ad 1 KHz
Coassiale 1.2/4.4 mm	5.3 ad 1 MHz	75 con politene	$89, 1.88 \cdot 10^{-7}, .26 \cdot 10^{-6}, 10^{-10}$
" 2.6/9.5 mm	2.3 ad 1 MHz	50 con aria	41, ", ", "
" 8.4/38 mm	.88 ad 1 MHz	$\frac{138}{\sqrt{\epsilon_r}} \log_{10} \frac{D}{d}$	1.45, ", ", " ad 1 MHz

13.2.3.1 Coppie simmetriche

Linee aeree



Sono costituite da una coppia di conduttori nudi, di bronzo od acciaio rivestito in rame, con diametro ϕ da 2 a 4 mm, sostenuti da una palificazione che li mantiene a distanza di 15 - 30 cm. L'uso delle linee aeree è andato estinguendosi con il tempo, ma rimane largamente diffuso nei paesi più poveri.

I valori riportati in tabella sono riferiti a conduttori con $\phi = 3$ mm, a frequenza di 1 KHz; la r già a 100 KHz cresce al valore di 20 Ω /Km, mentre la conduttanza g a 100 KHz e con tempo molto umido, può crescere fino a decine di volte il suo valore nominale ad 1 KHz. I valori riportati mostrano come le condizioni di Heaviside non siano rispettate, in quanto $rc \gg lg$, anche se lo scarto è inferiore rispetto al caso delle coppie ritorte.

L'impedenza caratteristica riportata in tabella, di circa 600 Ω , è ottenuta applicando il modello a basse perdite, con le costanti primarie indicate.

Coppie ritorte



Sono costituite da una coppia di conduttori in rame, con ϕ da 0.4 ad 1.3 mm, rivestiti di materiale isolante, ed avvolti tra loro secondo eliche con passo grande rispetto al diametro. Un numero variabile di tali coppie (tra qualche decina e qualche centinaio) sono poi raggruppate assieme, e rivestite con guaine protettive isolanti o metalliche; il risultato dell'operazione è interrato o sospeso mediante una fune in acciaio. L'uso delle coppie ritorte, nato allo scopo di realizzare il collegamento tra utente e centrale telefonica, si è recentemente esteso al cablaggio di reti locali (LAN) con topologia a stella; in tale contesto, i cavi sono indicati come UTP (*unshielded twisted pair*).

I valori riportati in tabella sono riferiti a conduttori con $\phi = .7$ mm, a frequenza di 1 KHz; la r a 100 KHz è circa doppia. La g dipende sostanzialmente dall'isolante utilizzato, mentre l'aumento di c è evidentemente legato alla vicinanza dei conduttori.

Anche in questo caso, risulta $rc \gg lg$, e dunque le condizioni di Heaviside non sono verificate. Nel passato, si è fatto largo uso dell'espedito di innalzare artificialmente l , collocando ad intervalli regolari una induttanza "concentrata" (le cosiddette bobine *Pupin*), realizzando così nella banda del canale telefonico un comportamento approssimativamente perfetto. Al crescere della frequenza, però, le bobine Pupin producono un effetto passabasso, aumentando di molto il valore di attenuazione; attualmente, le stesse coppie ritorte sono utilizzate per la trasmissione di segnali numerici PCM, e dunque le bobine Pupin sono state rimosse, ed al loro posto inseriti ripetitori rigenerativi.

L'impedenza caratteristica riportata in tabella, di circa 600Ω , è valida a frequenze audio, con cavi $\phi = .7$ mm. Prevalendo l'aspetto capacitivo, al crescere della frequenza Z_0 si riduce a $100\text{-}200 \Omega$, con fase di -10 gradi.

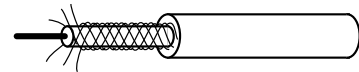
L'attenuazione chilometrica riportata, è sempre relativa al caso $\phi = .7$ mm; per diametri di 1.3 mm si ottengono valori circa dimezzati, mentre con $\phi = .4$ mm il valore di A_0 risulta maggiore.

Una ultima osservazione deve essere fatta, per spiegare che l'avvolgimento della coppia su se stessa ha lo scopo di ridurre i disturbi di diafonia. Infatti, se il passo dell'elica è diverso tra le coppie affasciate in unico cavo, le tensioni e correnti indotte da una coppia su di un'altra non interessano sempre lo stesso conduttore, ma entrambi in modo alternato. L'avvolgimento della coppia disturbante, inoltre, produce una alternanza dei conduttori in vicinanza della coppia disturbata, aggiungendo una ulteriore alternanza del verso del fenomeno di disturbo. Con questi accorgimenti, si trovano attenuazioni di diafonia a frequenze vocali, dell'ordine di $80\text{-}90$ dB su 6 Km. All'aumentare della frequenza, e della lunghezza del percorso comune, l'attenuazione di diafonia diminuisce (e quindi l'interferenza aumenta), fino a mostrare valori di $60\text{-}70$ dB a 750 KHz su 1.6 Km.

13.2.3.2 Cavo Coassiale

Un conduttore centrale è ricoperto di dielettrico, su cui è avvolto il secondo conduttore, intrecciato a formare una sorta di calza, e racchiuso a sua volta in una guaina isolante. La particolare conformazione del cavo lo rende molto più resistente ai fenomeni di interferenza; indicando con ϕ il diametro del conduttore interno e con D quello esterno, la teoria mostra che si determina un minimo di attenuazione se $D/\phi = 3.6$; per questo sono stati normalizzati i diametri mostrati nella tabella a pag. 290. Il tipo con $\phi/D = 8.4/38$ mm è sottomarino, e presenta la minima attenuazione chilometrica; A_0 aumenta al diminuire della sezione del cavo.

Finchè $D/\phi = 3.6$, l'impedenza caratteristica dipende solo dal dielettrico, con l'espressione generale fornita in tabella, ottenendo i valori di 50 e 75Ω con dielettrico aria e polietilene rispettivamente.



I valori delle costanti primarie riportati in tabella sono ottenuti facendo uso delle seguenti relazioni: $r = 8.4 \cdot 10^{-8} \sqrt{f} \left(\frac{1}{D} + \frac{1}{\phi} \right) \Omega/m$; $l = 0.46 \log_{10} \frac{D}{\phi} \mu H/m$; $g = 152 \cdot 10^{-12} \frac{f \varepsilon_r \tan \delta}{\log_{10} \frac{D}{\phi}} S/m$; $c = \frac{24.2 \cdot \varepsilon_r}{\log_{10} \frac{D}{\phi}} pF/m$; in cui si è posto f (in Hz nelle formule) pari a 1 MHz, D e d sono espressi in metri, ε_r è la costante dielettrica, e $\tan \delta$ è l'angolo di perdita del dielettrico; nel caso del polietilene, risulta $\varepsilon_r = 2.3$, $\tan \delta = 3 \cdot 10^{-4}$.

Esercizio Si desidera effettuare una trasmissione FDM di 120 canali telefonici, ognuno modulato AM-BLU, su di un cavo coassiale, nella banda di frequenze $1 \div 1.48$ MHz. Desiderando una potenza ricevuta per ogni canale di almeno 1 mW, e disponendo di un trasmettitore in grado di erogare 10 W, determinare la massima lunghezza del collegamento, supponendo verificate le condizioni di adattamento agli estremi del cavo, con impedenza caratteristica resistiva, ed attenuazione chilometrica $A_0 = 5.3$ dB/Km ad 1MHz. Di quanto dovrebbe aumentare la potenza trasmessa W_{dT} per raddoppiare la lunghezza?

Soluzione Supponendo tutti i canali contemporaneamente attivi, la potenza trasmessa per ciascuno di essi risulta pari a

$$W_{dT}^{(n)} = \frac{10}{120} = 83.3 \text{ mW, con } n = 1, 2, \dots, 120.$$

Tra tutti i canali, quello che subisce la massima attenuazione chilometrica è quello con portante più elevata, per il quale

$$A_d^{(120)} \text{ (dB/Km)} = A_0 \sqrt{1.48} = 5.3 \cdot 1.22 = 6.46 \text{ dB/Km.}$$

Per questo canale, il *guadagno di sistema* risulta pari a

$$G_s^{(120)} \Big|_{dB} = 10 \log_{10} \frac{W_{dT}^{(120)}}{W_{RMin}} = 10 \log_{10} \frac{83.3}{1} = 19.2 \text{ dB,}$$

essendo $W_{RMin} = 1$ mW come indicato nel testo. Come noto, G_s corrisponde alla massima attenuazione A_{dTot} che può essere accettata, e pertanto

$$A_{dTot}^{(120)} \Big|_{dB} = A_d^{(120)} \text{ (dB/Km)} \cdot L_{Km} = 19.2 \text{ dB,}$$

da cui si ricava per la massima lunghezza

$$L_{Km} = \frac{A_{dTot}^{(120)} \Big|_{dB}}{A_d^{(120)} \text{ (dB/Km)}} = \frac{19.2}{6.46} = 2.97 \text{ Km,}$$

che come vediamo è imposta dal canale più attenuato.

Per il primo canale, invece, si ha $A_d^{(1)} \text{ (dB/Km)} = A_0$, e dunque

$$A_{dTot}^{(1)} \Big|_{dB} = A_0 \text{ (dB/Km)} \cdot L_{Km} = 5.3 \cdot 2.97 = 15.74 \text{ dB.}$$

La differenza tra $G_s \Big|_{dB}$ (uguale per tutti i canali) e $A_{dTot}^{(1)} \Big|_{dB}$ rappresenta il margine di sistema per il primo canale, pari a

$$M = G_s - A_d = 19.2 - 15.74 = 3.46 \text{ dB.}$$

La stessa quantità, è anche uguale alla *distorsione lineare di ampiezza* del cavo nella banda del segnale.

Nel caso in cui si voglia superare una lunghezza doppia, anche $A_{dTot}^{(120)} \Big|_{dB}$ raddoppia, e per mantenere $W_{RMin} = 1$ mW, deve raddoppiare anche $G_s^{(120)} \Big|_{dB}$. Pertanto la nuova potenza/canale risulta

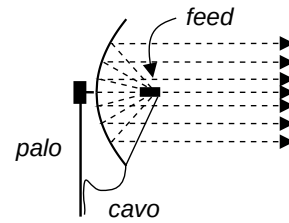
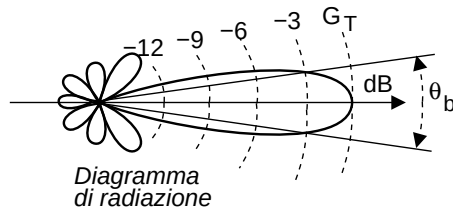
$$W'_{dT} \text{ (dBm)} = W_{Rmin} \text{ (dBm)} + G'_s \text{ (dB)} = 0 + 2G_s \text{ (dB)}; \text{ quindi}$$

$$W'_{dT} \text{ (mW)} = 10^{\frac{W'_{dT} \text{ (dBm)}}{10}} = 10^{\frac{2G_s \text{ (dB)}}{10}} = 10^{3.84} = 6918.3 \text{ mW,}$$

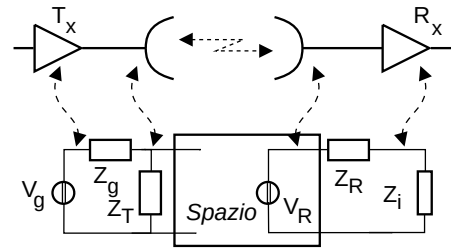
cioè 6.91 Watt/canale, e quindi $6.91 \cdot 120 = 830$ Watt complessivi !

13.3 Collegamenti radio

I segnali modulati occupano in genere una banda molto stretta attorno alla portante, tanto da poterli assimilare ad una singola sinusoide.



Pertanto, le condizioni di massimo trasferimento di potenza tra amplificatore finale e antenna trasmittente ($Z_g = Z_T^*$) e tra antenna ricevente e stadio di ingresso al ricevitore ($Z_R = Z_i^*$) danno luogo, nella banda di segnale, ad una componente di distorsione lineare che non dipende dalla frequenza (modulo e fase costante), e questo corrisponde (a parte una rotazione di fase) all'assenza di distorsioni lineari. Tutta la potenza disponibile fornita dall'amplificatore finale $W_{dT} = \frac{V_{Teff}^2}{4R_g}$ viene ceduta all'antenna, e da questa allo spazio. In realtà Z_T dipende dalla geometria dello spazio circostante; perciò l'amplificatore Tx va *accordato* dopo aver posizionato l'antenna.

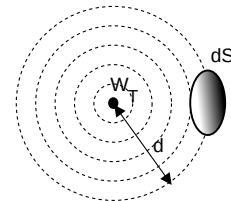


13.3.1 Trasduzione Elettromagnetica

Antenna Isotropa

Se l'antenna trasmittente irradia allo stesso modo in tutte le direzioni, W_{dT} si distribuisce su di una sfera; dunque una superficie dS , posta a distanza d , è attraversata da una potenza pari a

$$dW = W_{dT} \frac{dS}{4\pi d^2} \quad [\text{Watt}]$$



Si noti che il denominatore rappresenta la superficie di una sfera di raggio d .

Antenna direttiva Sono antenne che hanno direzioni privilegiate di emissione. Ad esempio, le antenne paraboliche dispongono di un illuminatore (*feed*¹⁰), e la superficie della parabola riflette il segnale del feed, focalizzandolo¹¹. Perciò la stessa W_{dT} si distribuisce *non* sfericamente, e la direzione di propagazione massima esibisce un guadagno G_T , il cui valore dipende dal rapporto tra le dimensioni dell'antenna e quelle della lunghezza d'onda λ secondo la relazione

$$G_T = 4\pi \frac{A}{\lambda^2}$$

¹⁰Dall'inglese *to feed* = alimentare.

¹¹Il processo di focalizzazione parabolica, comunemente usato ad esempio nei *fanali* degli autoveicoli, era ben noto ad un certo siracusano...

avendo indicato con A l'area dell'antenna.

Può essere definita una *larghezza del fascio* (BEAM WIDTH), che misura l'angolo θ_b entro cui la potenza irradiata è superiore alla metà della massima potenza presente nella direzione privilegiata¹². Ovviamente minore è θ_b , e maggiore è G_T .

Antenna ricevente Se la stessa antenna trasmittente viene usata (dall'altro lato del collegamento) per ricevere, questa mantiene lo stesso guadagno $G_R = G_T$ e lo stesso θ_b . Si definisce allora la sua *area efficace* come il valore

$$A_e = G_R \frac{\lambda^2}{4\pi}$$

legato alla forma e dimensione dell'antenna, a meno di un fattore di efficienza ρ (¹³). Perciò una stessa antenna (A_e fisso) aumenta il suo guadagno (e stringe il *beam*) all'aumentare della frequenza, ovvero al diminuire di $\lambda = \frac{c}{f}$ (¹⁴).

13.3.2 Bilancio energetico

Potenza ricevuta Usando l'area efficace dell'antenna ricevente per intercettare parte della potenza irradiata, si ottiene

$$W_R = W_{dT} G_T \frac{A_e}{4\pi d^2} = W_{dT} G_T G_R \left(\frac{\lambda}{4\pi d} \right)^2 [\text{Watt}]$$

Ovviamente, anche il ricevitore ha la propria $Z_i = Z_R^*$ accordata per il massimo trasferimento di potenza, e la banda di segnale è sempre stretta a sufficienza da garantire l'assenza di distorsioni lineari. Quindi la $W_R = W_{dR}$ è proprio la potenza ricevuta.

Attenuazione di spazio libero Il termine

$$\left(\frac{4\pi d}{\lambda} \right)^2 = \left(\frac{4\pi df}{c} \right)^2$$

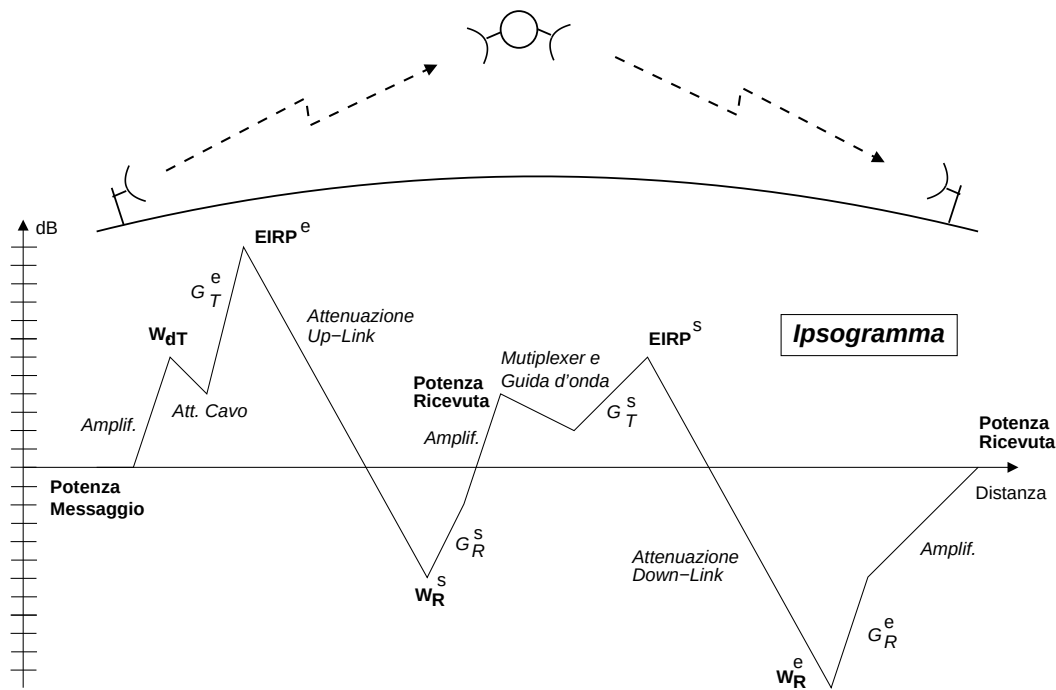
è chiamato *attenuazione di spazio libero*. A prima vista dipende da f^2 . In realtà tale dipendenza dalla frequenza si elide con quella relativa al guadagno delle antenne: $G_T = A_e \frac{4\pi}{\lambda^2} = A_e \frac{4\pi f^2}{c^2}$ (¹⁵).

¹²Si tratta di un concetto del tutto analogo alla "frequenza di taglio a 3 dB", ma applicata ad un dominio spaziale con geometria radiale.

¹³Indicando con A_r l'area *reale* (fisica) dell'antenna, risulta $A_e = \rho A_r$, con $\rho < 1$. La disuguaglianza tiene conto delle perdite dell'antenna, come ad esempio le irregolarità nella superficie della parabola, o l'ombra prodotta dalle strutture di sostegno.

¹⁴La costante $c = 3 \cdot 10^8$ metri/secondo rappresenta la velocità della luce nel vuoto, ossia la velocità di propagazione dell'onda elettromagnetica nello spazio.

¹⁵Mantenendo fissa la dimensione delle antenne, si ottiene il risultato che trasmissioni operanti a frequenze più elevate permettono di risparmiare potenza. Purtroppo però, guadagni superiori a 30-40 dB (corrispondenti a piccoli valori di θ_b) sono controproducenti, per i motivi esposti più avanti.



Attenuazione disponibile Il rapporto

$$A_d = \frac{W_{dT}}{W_{dR}} = \left(\frac{4\pi df}{c} \right)^2 \frac{1}{G_T G_R} \quad (13.4)$$

è chiamato *attenuazione disponibile*, ed indica di quanto si riduce la potenza trasmessa. Può essere espresso in decibel, tenendo conto delle costanti che vi compaiono, ed usando le unità di misura più idonee:

$$A_d \text{ (dB)} = 32.4 + 20 \log_{10} f \text{ (MHz)} + 20 \log_{10} d \text{ (Km)} - G_T \text{ (dB)} - G_R \text{ (dB)} \quad (13.5)$$

Osserviamo che, a differenza della trasmissione in cavo, l'attenuazione cresce con il quadrato della distanza, e quindi con il suo logaritmo quando espressa in decibel. Infatti ora l'attenuazione è dovuta esclusivamente all'aumentare della superficie su cui si distribuisce la potenza irradiata, e non a fenomeni dissipativi, come accade invece per cavo e fibra ottica.

Ipsogramma Questo è il nome di un diagramma che mostra l'andamento schematico dei livelli di potenza in dB lungo un collegamento.

In figura, è mostrata una ipotetica trasmissione terra-satellite-terra, assieme all'ipsogramma relativo. Il messaggio da trasmettere è amplificato a potenza W_{dT} , parte della quale si perde nel cavo che collega l'antenna trasmittente di guadagno G_T^e . L'EIRP^e (*effective irradiated power*) rappresenta la potenza effettivamente irradiata, che si riduce notevolmente nella trasmissione da terra a satellite (*Up-Link*). Nel caso, ad esempio, in cui la portante sia di 2 GHz e l'orbita sia geostazionaria (36.000 Km da terra), l'attenuazione di Up-Link (eguale a quella del *Down-Link* da satellite a terra) è di circa 190 dB.

Il segnale ricevuto, di potenza W_R^s , attraversa l'antenna ricevente del satellite di guadagno G_R^s , e l'amplificatore seguente eleva ulteriormente il livello del segnale trasmesso, che subisce alcune perdite nel collegamento con l'antenna trasmittente del satellite di guadagno G_T^s , determinando così il valore della EIRP^s all'uscita del *trasponder*¹⁶ satellitare.

Dal lato della stazione ricevente terrestre, si riceve il segnale di potenza W_R^e che ha subito l'attenuazione del down-link; tale segnale è quindi riportato ad un livello di potenza appropriato sia grazie al guadagno di antenna, che per mezzo di uno stadio di amplificazione.

13.3.3 Condizioni di propagazione e Attenuazioni supplementari

Come mostrato nella discussione del collegamento satellitare, il calcolo della potenza ricevuta non è così banale come sembrava a prima vista. In particolare, al valore A_d (dB) devono essere sommate (in decibel) tutte le attenuazioni *in più*:

Perdite di accoppiamento Dovute al mancato verificarsi delle condizioni di massimo trasferimento di potenza: ammontano a qualche dB.

Assorbimento terrestre Quando l'antenna è distante dal suolo meno di qualche lunghezza d'onda, l'energia si propaga per onda superficiale, ovvero la crosta terrestre fa da conduttore. Questa forma di propagazione provoca una attenuazione supplementare che aumenta con la frequenza, tanto che già a 3 MHz raggiunge i 25 dB ogni 10 Km (equivalente ad una riduzione di potenza di $10^{2.5} = 316$ volte). Le *onde medie* (0,3-3 MHz) sono meno attenuate, ed ancora meno le *onde lunghe* (10-300 KHz) che viaggiano appunto via terra.

13.3.3.1 Condizioni di visibilità

Come ricavabile anche dall'espressione dell'area efficace, all'aumentare della frequenza si possono ottenere antenne di dimensioni ridotte e contemporaneamente di elevato guadagno. Allo stesso tempo, per evitare l'assorbimento terrestre, occorre posizionare l'antenna in alto (in cima ad una torre), e trasmettere per onda diretta.

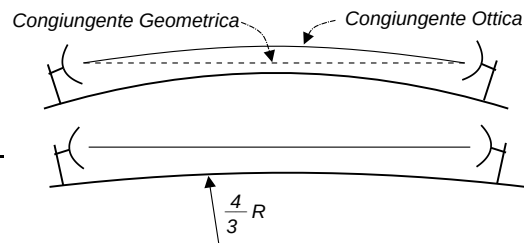
A causa della curvatura terrestre, esiste una altezza minima da rispettare: ad esempio con torri da 60 metri si raggiungono distanze (in visibilità) di 50 Km. Ovviamente, il problema si presenta in pianura. Tratte più lunghe richiedono torri più alte, ma anche guadagni di antenna maggiori (e quindi antenne più grandi e più direttive). Questa non è però una soluzione molto praticabile, in quanto in presenza di vento forte le antenne "grandi" possono spostarsi e perdere il puntamento; inoltre, il costo delle torri aumenta esponenzialmente con l'altezza.

¹⁶Il termine *trasponder* è legato al fatto che il satellite non si comporta come un semplice ripetitore, bensì si occupa anche di *trasporre* la banda di frequenze occupata dalla trasmissione. Questa esigenza deriva dal fatto che, essendo molto elevata la differenza tra EIRP^s e W_R^s , se la frequenza portante utilizzata nell'uplink fosse uguale a quella del down-link, il segnale trasmesso costituirebbe un rilevante termine di interferenza per il lato ricevente del satellite, nonostante l'elevata direttività delle antenne, dando così luogo ad un fenomeno di *diafonia*.

Le considerazioni svolte in questa nota si posso applicare altrettanto bene anche al caso di ripetitori terrestri.

Orizzonte Radio

Nel calcolare l'altezza delle torri (ed il puntamento delle antenne) si deve considerare anche il fenomeno legato al fatto che l'onda elettromagnetica, propagandosi, *si piega* verso gli strati dell'atmosfera con indici di rifrazione maggiori (ossia verso terra). Pertanto, i calcoli vengono effettuati supponendo che il raggio terrestre sia $4/3$ volte quello reale. Inoltre, l'indice di rifrazione (che aumenta verso il basso) può variare con l'ora e con le condizioni climatiche: pertanto, anche in questo caso, le antenne con guadagni elevati (e molto direttive) possono andare fuori puntamento.

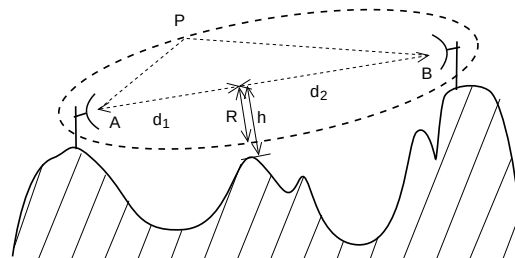


Ellissoidi di Fresnel Nella propagazione elettromagnetica occorre tenere conto dei fenomeni di diffrazione, che *deviano* nella zona in *ombra*¹⁷ le onde radio che transitano in prossimità di ostacoli. Pertanto, la determinazione dell'orizzonte radio deve prevedere un *margin* di distanza h tra la congiungente tra le antenne ed il suolo, od un eventuale ostacolo. La distanza h deve essere almeno pari al raggio del primo ellissoide di Fresnel, che è un solido di rotazione definito come il luogo dei punti P per i quali la somma delle distanze $d(A, P) + d(P, B)$ è pari a $d(A, B) + \frac{\lambda}{2}$, in cui $\lambda = \frac{c}{f}$ è la lunghezza d'onda della trasmissione a frequenza f .

Suddividendo la distanza $d(A, B)$ tra i due fuochi A e B in due segmenti d_1 e d_2 , individuati dalla posizione dell'ostacolo, si trova che il raggio dell'ellissoide è pari a

$$R = \sqrt{\frac{\lambda}{\frac{1}{d_1} + \frac{1}{d_2}}}$$

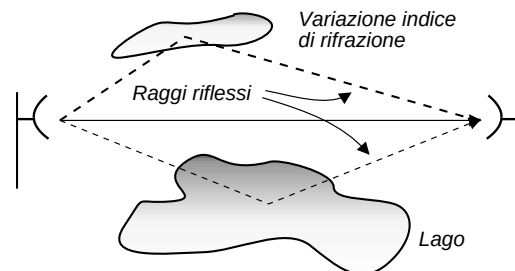
che, nel caso $d_1 = d_2 = \frac{d(A, B)}{2}$, assume il valore massimo $R_M = \frac{1}{2}\sqrt{\lambda d}$. Qualora si determini la condizione $h < R$, il collegamento subisce una attenuazione supplementare che aumenta al diminuire di $\frac{h}{R}$, ed è maggiore per gli "spigoli vivi", fino ad arrivare ad una decina di dB.



13.3.3.2 Cammini Multipli e Diffusione

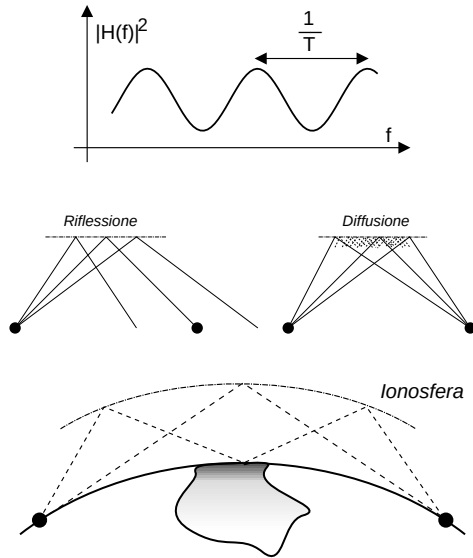
Oltre i 30 MHz, nonostante la direttività delle antenne, alcuni raggi obliqui possono incontrare superfici riflettenti (laghi o masse d'acqua), oppure brusche variazioni dell'indice di rifrazione, che causano la riflessione totale del raggio, e la ricezione di una eco ripetuta dello stesso segnale.

In questi casi il collegamento può essere caratterizzato mediante una risposta impulsiva del tipo $h(t) = \sum_{k=1}^N a_n \delta(t - T_k)$ in cui i valori T_k sono i ritardi con cui si presentano



¹⁷Lo stesso fenomeno di diffrazione è egualmente valido per l'energia luminosa, e può essere sperimentato illuminando una fessura, ed osservando le variazioni di luminosità dall'altro lato.

le diverse eco, ognuna caratterizzata da una ampiezza a_k . La risposta in frequenza corrispondente $H(f) = \sum_{k=1}^N a_k e^{-j2\pi f T_k}$ può produrre delle distorsioni lineari; ricordiamo infatti che il modello di cammino multiplo per una sola eco con ritardo T , che presenta una $|H(f)|^2 = 1 + a^2 + 2a \cos 2\pi f T$ periodica in frequenza con periodo $f = \frac{1}{T}$, osserviamo che all'aumentare di T , le oscillazioni di $|H(f)|^2$ si infittiscono¹⁸, e dunque aumenta la possibilità che $|H(f)|^2$ vari di molto nella banda del segnale, causando distorsioni lineari che devono essere equalizzate.



Qualora la banda del segnale sia invece sufficientemente piccola rispetto a $\frac{1}{T}$, e si possa considerare $|H(f)|^2$ costante, la presenza di cammini multipli può comunque dar luogo ad attenuazione, che prende il nome di *flat fading*¹⁹.

Tra 0,1 e 10 GHz, si può verificare il fenomeno della *diffusione troposferica* (lo strato dell'atmosfera fino a 20 Km di altezza), causata da turbolenze e particelle sospese, e che comportano un numero *infinito* di cammini multipli.

Tra qualche MHz e 30 MHz, intervengono fenomeni di radiodiffusione *ionosferica* (la fascia oltre gli 80 Km), dove strati ionizzati causano riflessioni del segnale, e consentono la trasmissione anche tra luoghi non in visibilità, ma con il rischio di cammini multipli. E' questo il caso tipico delle *onde corte*.

Per frequenze sotto il MHz la propagazione è per *onda di terra*, e l'assorbimento terrestre impedisce di coprire grandi distanze (tranne che per le *onde lunghe*, meno attenuate). Anche in questo caso può verificarsi la diffusione troposferica, specie di *notte*.

Per frequenze sotto il MHz la propagazione è per *onda di terra*, e l'assorbimento terrestre impedisce di coprire grandi distanze (tranne che per le *onde lunghe*, meno attenuate). Anche in questo caso può verificarsi la diffusione troposferica, specie di *notte*.

Collegamenti radiomobili Le condizioni di propagazione per terminali radiomobili, come nel caso della telefonia cellulare, presentano diversi aspetti particolari. Innanzitutto, l'antenna del terminale mobile è molto vicina al suolo, e ciò comporta la presenza di una eco fissa da terra, e quasi sempre il mancato rispetto delle condizioni di Fresnel²⁰. Inoltre, specialmente in ambito urbano, si verifica un elevato numero di cammini multipli e diffrazioni, che perdipiù variano nel tempo in conseguenza dello spostamento del terminale. Infine, l'uso condiviso di una stessa banda di frequenze radio da parte di una moltitudine di terminali, determina la necessità di riusare le stesse frequenze in regioni differenti, e l'attuazione di meccanismi di codifica di canale per ridurre gli effetti del-

¹⁸Ad esempio, desiderando $\frac{1}{T} > 1$ MHz, si ottiene $T_{Max} = 1 \mu\text{sec}$; se l'onda radio si propaga alla velocità $c = 3 \cdot 10^8$ m/sec, la massima differenza di percorso vale $\Delta_{max} = c \cdot T_{Max} = 3 \cdot 10^8 \cdot 10^{-6} = 300$ metri.

¹⁹Il termine *fading* si traduce come *affievolimento*, ma è spesso usato in inglese, cosicché le distorsioni lineari per segnali a banda stretta sono dette *condizioni di fading piatto*.

²⁰Alla frequenza di 1 GHz si ha $\lambda = 30$ cm e per una distanza di 100 metri dal trasmettitore si ottiene un raggio massimo dell'ellissoide pari a $\frac{1}{2}\sqrt{3 \cdot 100} = \frac{1}{2}\sqrt{300} \simeq 2.7$ metri.

le interferenze e del fading variabile²¹. Analizziamo brevemente i primi due fenomeni, rimandando la discussione sulle tecniche di accesso multiplo ad una prossima edizione.

Path loss La dipendenza della attenuazione dal quadrato della distanza presente in (13.4), si riferisce al caso ideale di spazio libero; l'esponente di d invece aumenta fino al valore 4, a seconda del tipo di ambiente (urbano, rurale) e dell'altezza dell'antenna ricevente. Pertanto, il termine $20 \log_{10} d (Km)$ che compare in (13.5) deve essere sostituito con

$$n \cdot 10 \log_{10} d (Km) + \alpha \quad [\text{dB}]$$

in cui n ed α sono determinati in base a misure, e tengono conto delle condizioni operative. Il valore di n varia da 4 a 3 con $d < 100$ metri, all'aumentare dell'altezza dell'antenna fissa, mentre il termine α può variare da 7 a 15 dB con antenna fissa alta 30 e 10 metri rispettivamente, e subire un incremento di quasi 30 dB passando da un ambiente aperto ad un ambito urbano.

Slow fading La stima delle grandezze n ed α ora introdotte è svolta *mediando* i risultati di diverse misure condotte nel territorio che si intende caratterizzare: infatti per territori diversi, si riscontrano valori di attenuazione complessiva diversi, anche per uguali valori di d . Questo fenomeno è indicato come *slow fading*, poiché non si presenta muovendosi di poco in una stessa zona, in quanto dipende dalla orografia del territorio e dalla natura degli oggetti limitrofi. Non conoscendo a priori in che zona ci si trovi, l'effetto dello *slow fading* si manifesta come una attenuazione supplementare a_s aleatoria, che risulta avere un andamento gaussiano in dB (che per questo è detto *lognormale*) e cioè del tipo²²

$$p_{A_s}(a_s(\text{dB})) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{a_s^2(\text{dB})}{2\sigma^2_{SF}}}$$

dove σ_{SF}^2 varia tra 6 ed 8 dB per elevazioni dell'antenna tra 5 e 15 metri²³. Per velocità del mobile non superiori ai 15 Km/h, si può assumere a_s costante in frequenza per qualche MHz, e nel tempo per poche centinaia di millisecondi.

²¹Mentre il fading produce una attenuazione variabile al segnale, la stessa variabilità delle condizioni di propagazione può portare a livelli interferenza variabili, causati da altre trasmissioni nella stessa banda. La variabilità temporale della qualità del segnale ricevuto, in particolare quella *veloce* (vedi appresso), produce errori a *burst*, che possono essere corretti mediante codifica di canale ed scrambling.

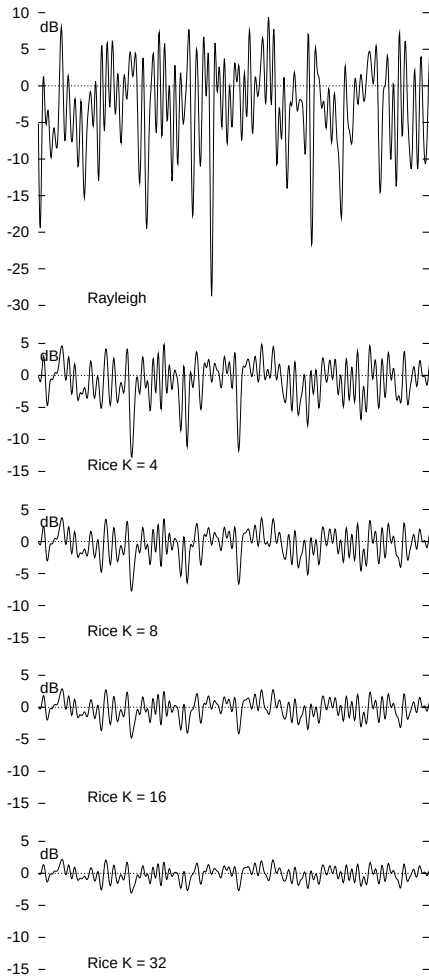
²²L'involuppo complesso $\underline{x}$ del segnale ricevuto è adeguatamente rappresentato da una v.a. di Rice (vedi pag. 111) $\underline{x} = a + \underline{r}$, in cui $\underline{r}$ ha d.d.p. di Rayleigh e rappresenta l'effetto di molte cause indipendenti, relative ai cammini multipli, ed a è l'ampiezza della eco di segnale ricevuta con la maggiore ampiezza. Se $a \gg |\underline{r}|$ possiamo scrivere

$$\begin{aligned} a_s(\text{dB}) &= 10 \log_{10} \frac{1}{|a + \underline{r}|^2} = -10 \log_{10} ((a + r_c)^2 + r_s^2) = -10 \log_{10} \frac{a^2}{a^2} (a^2 + 2ar_c + r_c^2 + r_s^2) = \\ &= 10 \left(\log_{10} a^2 + \log_{10} \left(1 + \frac{2r_c}{a} + \frac{r_c^2}{a^2} \right) \right) \simeq 10 \left(\log_{10} a^2 + \frac{2r_c}{a} \right) = 10 \log_{10} a^2 + 20 \frac{r_c}{a} \end{aligned}$$

in quanto $\log(1 + \alpha) \simeq \alpha$ con $\alpha \ll 1$, e quindi $a_s(\text{dB})$ ha media $10 \log_{10} a^2$ (compresa nel *path loss*) ed esibisce una d.d.p. gaussiana, la stessa di r_c .

²³All'aumentare dell'altezza dell'antenna, si estende l'area di copertura della stessa, ma in ambito urbano questo corrisponde ad una maggiore variabilità delle effettive condizioni operative.

Fast fading Rappresenta le rapide e profonde fluttuazioni nel livello del segnale radio, osservate durante il movimento. Queste fluttuazioni sono causate dalle variazioni dei ritardi di fase con cui giungono le varie onde riflesse: spostandosi infatti di $\frac{\lambda}{2}$ ⁽²⁴⁾ si può passare da una situazione di somma coerente ad una completa opposizione di fase.



Qualora la ricezione avvenga principalmente in assenza di visibilità, ed in presenza di un numero elevato di cammini multipli, i valori del modulo dell'involuppo complesso del segnale ricevuto $\rho(t) = |\underline{x}(t)|$ sono adeguatamente rappresentati da una v.a. di Rayleigh (vedi pag. 111), la cui d.d.p. ha espressione $p_P(\rho) = \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right)$ con $\rho \geq 0$, avendo indicato con σ^2 la potenza delle componenti in fase e quadratura²⁵. Sotto tali ipotesi la potenza istantanea s ricevuta, legata a ρ^2 , ha d.d.p. esponenziale negativa²⁶:

$$p_S(s = \rho^2) = \frac{1}{2\sigma^2} \exp\left(-\frac{s}{2\sigma^2}\right)$$

e quindi presenta frequentemente valori molto bassi e prossimi a zero, che in dB danno luogo a *profonde* attenuazioni (*deep fades*).

In figura è mostrato l'andamento in dB del livello di segnale ricevuto, relativo alle condizioni di ricezione nominale, nei casi di *fading di Rayleigh* e di *Rice* per diversi valori del *fattore di Rice K* definito come il rapporto $K = \frac{a^2}{2\sigma^2}$ tra la potenza $\frac{a^2}{2}$ dell'onda diretta e quella σ^2 della componente dovuta al multipath. Come si può notare, il fading di Rayleigh produce attenuazioni più profonde, mentre in presenza di una forte componente diretta, l'ampiezza del fading si riduce sensibilmente.

Dimensione di cella e velocità di trasmissione L'analisi ora svolta, prescinde dall'entità effettiva dei ritardi sui diversi percorsi, e l'applicazione diretta dei modelli di *Rayleigh* e di *Rice* sottintende che i raggi provenienti da diverse direzioni non presentino differenze di ritardo reciproche; in questo caso, non si determinano distorsioni lineari,

²⁴A frequenza di 1 Ghz, si ha $\lambda \simeq 30$ cm. Questo fenomeno può essere facilmente sperimentato, qualora si tenti di sintonizzare un televisore dotato di una antenna "interna", ed osservando come la qualità del segnale vari sensibilmente anche per piccoli spostamenti del ricevitore.

²⁵Le componenti in fase e quadratura sono costituite dalla somma di tutti i cammini multipli, e per questo motivo possono essere assunte di tipo gaussiano, a media nulla e varianza σ^2 .

²⁶Impostando il cambiamento di variabile $s = \rho^2$, si possono applicare le regole viste al § 5.6.6, individuando la funzione inversa come $\rho = \sqrt{s}$, la cui $\frac{d}{ds}\rho(s)$ fornisce $\frac{1}{2\sqrt{s}}$. Pertanto, la d.d.p. della nuova v.a. s vale:

$$p_S(s) = p_P(\sqrt{s}) \cdot \frac{d}{ds}\rho(s) = \frac{\sqrt{s}}{\sigma^2} \exp\left(-\frac{(\sqrt{s})^2}{2\sigma^2}\right) \cdot \frac{1}{2\sqrt{s}} = \frac{1}{2\sigma^2} \exp\left(-\frac{s}{2\sigma^2}\right)$$

ed il fading si dice *non selettivo* in frequenza, ovvero *piatto*. La banda occupata dalla trasmissione può allora cedere sensibilmente, e per trasmissioni numeriche è possibile realizzare un baud rate elevato, ed un periodo di simbolo ridotto. L'occasionale presenza di un *deep fade* può essere compensata mediante codifica di canale e scrambling dei dati.

Se al contrario la distorsione di tempo di transito prodotta dal multipath diviene confrontabile con la durata di un simbolo, sopraggiunge ISI, che per essere rimossa richiede l'attuazione di una procedura di equalizzazione.

Per celle molto grandi, la differenza di percorso delle diverse riflessioni può essere notevole, imponendo una bassa velocità di trasmissione, pena l'insorgenza di fading *selettivo* in frequenza. Riducendo la dimensione di cella, è possibile invece aumentare la velocità di trasmissione, dato che le differenze di ritardo si riducono. Per questo motivo, se celle con raggio di chilometri e dispersioni di ritardo di oltre 10 μsec necessitano di equalizzazione anche per trasmissioni a 64 kbps, comunicazioni *indoor* con dispersioni di ritardo inferiori ad 1 μsec possono presentare *flat fading* per velocità superiori al Mbps. Celle di dimensione minima, dette anche *picocelle*, presentano dispersioni temporali di solo qualche decina di picosecondi, permettendo di operare a molti Mbps anche senza equalizzazione.

13.3.3.3 Assorbimento Atmosferico

Per lunghezze d'onda di dimensione comparabile a quella delle molecole di ossigeno, si produce un fenomeno dissipativo di *assorbimento*; le frequenze interessate sono quelle superiori a 30 GHz, con un massimo di 20 dB/Km a 60 GHz⁽²⁷⁾. Inoltre, il vapor d'acqua (con molecole di dimensioni maggiori) produce una attenuazione supplementare di 1-2 dB/Km (al massimo) a 22 GHz. Sotto i 10 GHz non si verifica né assorbimento da ossigeno, né da vapore.

13.3.3.4 Dimensionamento di un collegamento soggetto a pioggia

In caso di pioggia, si manifesta una ulteriore causa di assorbimento atmosferico, detto appunto *da pioggia*, che costituisce la principale fonte di attenuazione supplementare per frequenze superiori a 10 GHz. L'attenuazione supplementare da pioggia aumenta con la frequenza portante, con l'intensità di precipitazione e con l'estensione della zona piovosa lungo il tragitto radio; questi ultimi due fattori sono evidentemente elementi aleatori, e per questo il dimensionamento mira a stabilire quale sia il margine necessario a garantire un grado di servizio prefissato. Il margine necessario, è pertanto pari al valore di attenuazione supplementare che viene superata con una probabilità minore o uguale al grado di servizio.

Una formula sperimentale che consente di determinare il valore in dB dell'attenuazione supplementare che viene superata con probabilità p è:

$$A_s(r_0, d, p) = K \cdot r_0^\alpha \cdot d \cdot \beta(d) \cdot \gamma(p) \quad [\text{dB}]$$

in cui r_0 è l'intensità di precipitazione (in mm/h) che viene superata per lo 0.01 % del tempo, d è la lunghezza del collegamento, e K ed α sono costanti che caratterizzano l'entità dell'interazione dell'onda radio con la pioggia, in funzione della frequenza portante e

²⁷L'elevata attenuazione chilometrica presente a 60 GHz può essere sfruttata nei sistemi di comunicazione cellulare, allo scopo di riusare una stessa banda di frequenze anche a breve distanza.

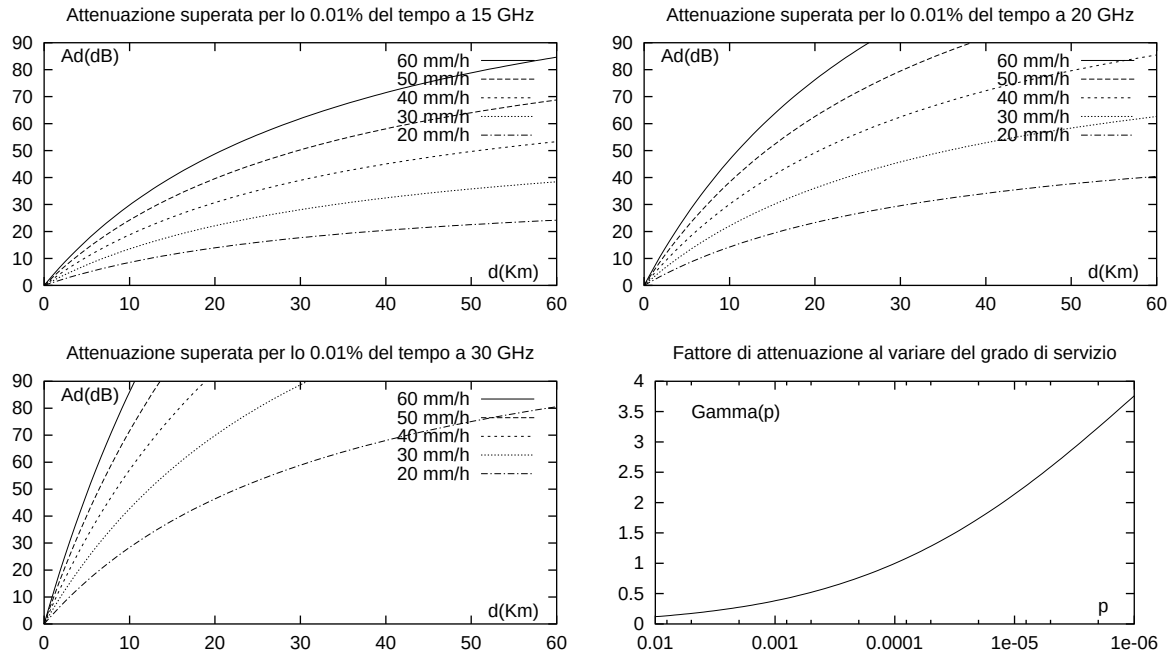


Figura 13.1: Curve di attenuazione supplementare per pioggia

di altre condizioni climatiche ed ambientali, i cui valori medi sono riportati nella tabella che segue.

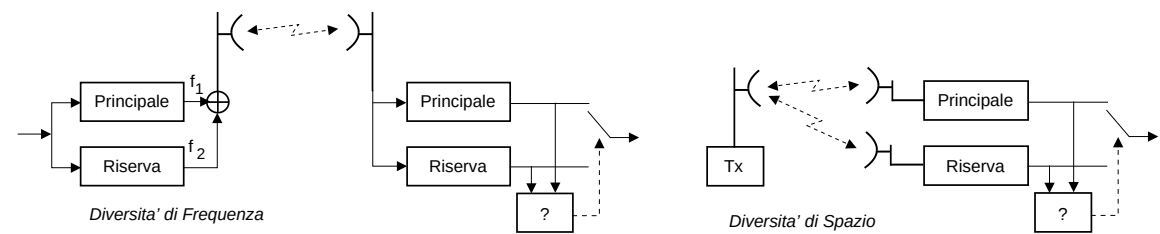
f_0 (GHz)	10	15	20	25	30	35
α	1.27	1.14	1.08	1.05	1.01	.97
K	.01	.036	.072	.12	.177	.248

Il valore di r_0 per l'Italia è compreso tra 20 e 60 mm/h, mentre il termine $\gamma(p) = 6.534 \cdot 10^{-3} \cdot p^{-(.718 + .043 \cdot \log_{10} p)}$, che vale 1 per $p = 10^{-4}$, permette di tener conto del grado di servizio che si vuole ottenere. Infine, $\beta(d) = 1 / (1 + .0286 \cdot d)$ è un fattore correttivo che tiene conto del fatto che *non piove lungo tutto* il collegamento. I grafici in fig. 13.1, mostrano l'andamento del termine $K \cdot r_0^\alpha \cdot d \cdot \beta(d)$ per diversi valori di f_0 ed r_0 , in funzione dell'estensione del collegamento; infine, è riportato il grafico della funzione $\gamma(p)$ per diversi valori di p .

Dimensionare un collegamento imponendo un margine elevato può dar luogo a problemi dal lato del ricevitore, che potrebbe trovarsi ad operare in regione non lineare a causa dell'eccesso di potenza ricevuta, qualora non siano presenti le attenuazioni supplementari: può essere allora utilizzato un canale di ritorno nell'altra direzione, in modo da regolare la potenza del trasmettitore.

13.3.3.5 Collegamenti in diversità

Il *fading* prodotto da cammini multipli può essere rilevante, specialmente se due repliche del segnale giungono al ricevitore con ampiezze molto simili. Il problema può essere risolto prevedendo una ridondanza degli apparati.



Diversità di frequenza Lo stesso collegamento è operato su due diverse portanti; nel caso in cui la trasmissione che avviene con una delle portanti subisca una attenuazione selettiva in frequenza, quella che adotta l'altra portante ne è esente (o viceversa). Se la banda è particolarmente affollata, la stessa configurazione può essere adottata per fornire una ridondanza $N : 1$. Ad esempio, in una trasmissione multiplata FDM, la portante di riserva viene assegnata al canale del banco FDM che presenta la maggiore attenuazione.

Diversità di spazio Adottando due diverse antenne riceventi in posizioni differenti, la differenza di percorso T tra cammini multipli è differente per le due antenne, e dunque la risposta in frequenza $|H(f)|^2 = 1 + a^2 + 2a \cos 2\pi fT$ ha una diversa periodicità nei due casi. Pertanto, anche se un ricevitore subisce una attenuazione selettiva, l'altro ricevitore ne è esente.

13.4 Collegamenti in fibra ottica

13.4.1 Generalità

Natura fisica della fibra Una fibra ottica è realizzata in vetro o silicio fuso, e purchè il materiale sia un dielettrico trasparente alla luce, può essere realizzata anche in plastica. Il suo utilizzo è quello di trasportare energia luminosa in modo guidato. Una caratteristica che deriva direttamente dalla sua natura, è l'immunità della fibra ottica ai disturbi di natura elettromagnetica; tale proprietà impedisce fenomeni di interferenza (diafonia), così come non permette di prelevare segnale dall'esterno (intercettazione).

Il segnale luminoso Le lunghezze d'onda delle radiazioni elettromagnetiche nel campo del visibile sono comprese tra 50 nm ($1 \text{ nm} = 10^{-9}$ metri) dell'ultravioletto fino a circa 100 μm dell'infrarosso, che corrispondono a frequenze (ricordando ancora che $f = \frac{c}{\lambda}$) che vanno da $6 \cdot 10^{15}$ Hz a $3 \cdot 10^{12}$ Hz.

Questi valori individuano una banda passante veramente notevole se comparata ad altri mezzi trasmissivi: supponiamo infatti di effettuare una modulazione che occupi una banda pari allo 0.1% della frequenza portante. Se $f_0 = 1$ GHz, si ha 1 MHz di banda; ma se $f_0 = 10^{14}$, si ha una banda di 100 GHz!

	Ultravioletto	→	Infrarosso	
λ	$50 \cdot 10^{-9}$	→	$9 \cdot 10^{-6}$	metri
f	$6 \cdot 10^{15}$	←	$3 \cdot 10^{12}$	Hz

Trasmissione ottica Anche se sono teoricamente possibili schemi di modulazione analogici, le fibre ottiche sono usate prevalentemente per trasportare informazione di

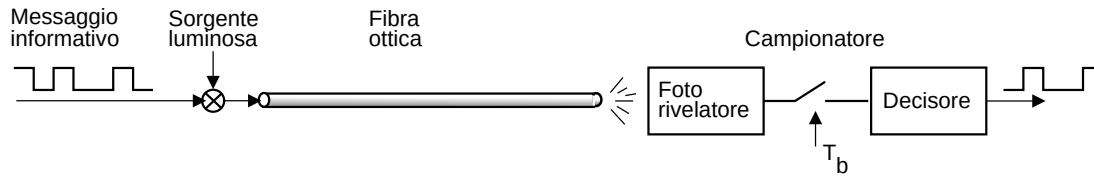
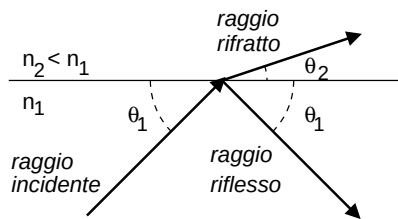


Figura 13.2: Schema di trasmissione in fibra ottica

natura *numerica* secondo lo schema di fig. 13.2, in cui la luce emessa da una sorgente è accesa o spenta (ovvero modulata in ampiezza con uno schema ON/OFF). All'altro estremo della fibra, un fotorivelatore effettua una rivelazione incoerente dell'energia luminosa, che viene nuovamente convertita in un segnale elettrico. Le prime fibre ottiche risalgono al 1970, e fornivano attenuazioni dell'ordine di 20 dB/Km. Attualmente si sono raggiunti valori di attenuazione di 0.2 dB/Km, pari ad un quarto di quella dei migliori cavi coassiali. D'altra parte, a differenza del rame, il materiale utilizzato per le fibre (vetro o silicio) è largamente disponibile in natura. Inoltre, a parità di diametro, una fibra ottica trasporta un numero anche 1000 volte maggiore di comunicazioni rispetto ad un cavo coassiale, fornendo quindi anche un risparmio di spazio.

Propagazione luminosa e indice di rifrazione Lo spazio libero è il mezzo di propagazione in cui la luce viaggia più velocemente. Il rapporto tra $c = 3 \cdot 10^8$ m/sec, e la velocità di propagazione v di un mezzo trasparente, è l'*indice di rifrazione* n del mezzo stesso.



Quando un raggio luminoso incontra un mezzo con diverso indice n (ad esempio, da n_1 ad $n_2 < n_1$) una parte di energia si riflette con angolo uguale quello incidente, e la restante parte continua nell'altro mezzo, ma con diverso angolo. Risulta $\frac{n_2}{n_1} = \frac{\cos \theta_1}{\cos \theta_2}$, e dunque il raggio rifratto è più inclinato nel mezzo con n inferiore (dove viaggia più veloce). Esiste un valore $\theta_c = \arccos \frac{n_2}{n_1}$

sotto il quale non si ha rifrazione, ma tutto il raggio viene riflesso.

E' proprio su questo fenomeno che si basa l'attitudine delle fibre ottiche di trasportare energia luminosa. La fibra ottica è infatti costituita da un nucleo (*core*) centrale con indice di rifrazione n_1 , circondato da un rivestimento (*cladding*) con indice $n_2 < n_1$; entrambi racchiusi in una guaina (*jacket*) di materiale opaco (vedi fig. 13.3).

Quando una sorgente luminosa è posta davanti alla fibra, l'energia si propaga mediante diversi *modi di propagazione*, definiti nel contesto della meccanica quantistica, e identificabili in chiave di ottica geometrica come i diversi angoli di incidenza. Il *modo principale* è quello che si propaga lungo l'asse rettilineo, mentre i *modi secondari* sono quelli con angolo $< \theta_c$, che si riflettono completamente al confine tra core e cladding. I modi associati ad angoli più elevati di θ_c vengono progressivamente assorbiti dalla guaina, e dunque non si propagano.

Il valore $\Delta = \sqrt{n_1^2 - n_2^2} = n_1 \sin \theta_c$ prende il nome di *apertura numerica*, e permette di risalire al massimo angolo di incidenza mediante la relazione $\theta_c = \arcsin \frac{\Delta}{n_1}$. Come si vede, Δ è tanto più piccolo quanto più n_1 ed n_2 sono simili; al diminuire di Δ , si riduce

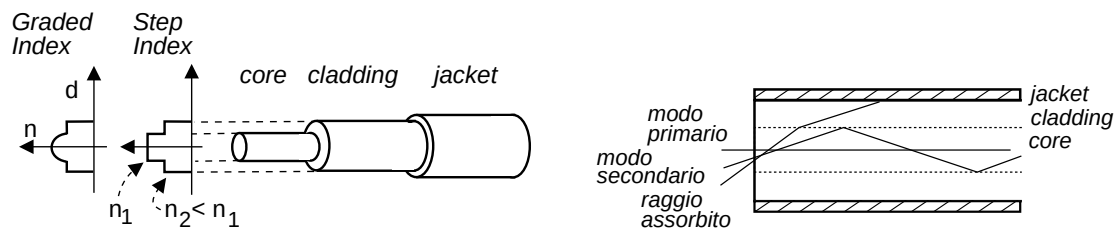


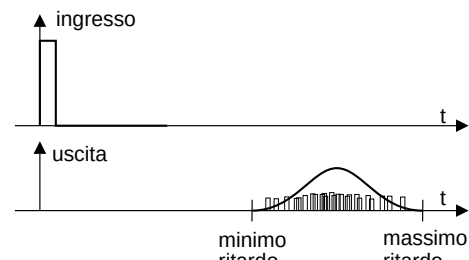
Figura 13.3: Tecnologia della fibra ottica e modi di propagazione

anche la potenza luminosa che viene immessa nella fibra ottica, ma si ottiene il beneficio illustrato di seguito.

13.4.2 Propagazione luminosa

Dispersione modale Questo fenomeno è dovuto al fatto che i modi propagazione relativi agli angoli di incidenza più elevati, percorrono di fatto *più strada*, e dunque impiegano più tempo per giungere a destinazione.

Pertanto, ogni singolo impulso luminoso presente in ingresso produce in uscita più impulsi distanziati nel tempo, uno per ogni modo di propagazione. Dato che inoltre avviene un continuo scambio di energia tra i diversi modi, si ottiene che l'uscita sarà un segnale con una sagomatura allargata (vedi figura). L'entità della *dispersione temporale* (differenza tra ritardo max e min) sarà tanto maggiore quanto più il collegamento è lungo, e quanti più modi partecipano alla propagazione. L'effetto più appariscente del fenomeno descritto consiste nella limitazione della massima frequenza con cui gli impulsi luminosi possono essere posti in ingresso alla fibra; impulsi troppo vicini risulterebbero infatti indistinguibili in uscita. Pertanto la massima frequenza di segnalazione in una fibra ottica, dipende dalla lunghezza della fibra stessa.

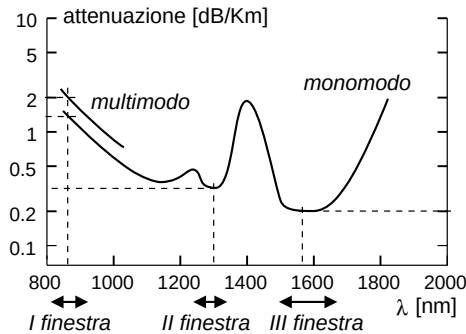


Si chiamano *fibre multimodo* le fibre ottiche in cui sono presenti più modi di propagazione. Queste sono del tipo STEP INDEX se n cambia in modo brusco, o GRADED INDEX se il core ha un indice graduato. Nel secondo caso la dispersione temporale è ridotta; infatti quando i modi secondari attraversano la sezione periferica del core, incontrano un indice di rifrazione n ridotto, e quindi viaggiano più veloci.

Una diversa (e drastica) soluzione al problema della dispersione temporale, è fornita dalle fibre ottiche *monomodo*: queste sono realizzate con un core di diametro così piccolo, da permettere la propagazione del solo modo primario.

Ovviamente le ultime due soluzioni (graded index e fibra monomodo) si sono rese possibili grazie ai progressi nei processi di fabbricazione. Per concretizzare il discorso, è sufficiente citare che il diametro del core passa dai 50 μm per le fibre multimodo, a circa 8 μm nel caso monomodo.

Attenuazione



dependente da λ , e sono stati individuati 3 intervalli di lunghezze d'onda (detti *finestre*) per i quali l'assorbimento è ridotto, ed in cui sono effettuate le trasmissioni ottiche.

Finestra	I	II	III
λ	$.8 \div .9 \mu\text{m}$	$1.2 \div 1.3 \mu\text{m}$	$1.5 \div 1.7 \mu\text{m}$
A_d (dB/Km)	1.2 (monom.) 2 (multim.)	.35	.2

La prima finestra (con attenuazione maggiore) è stata l'unica disponibile agli inizi, a causa dell'assenza di trasduttori affidabili a frequenze inferiori, ed è tuttora usata per collega-

menti economici e scarsamente critici. La seconda finestra ha iniziato ad essere usata assieme alle fibre monomodo, grazie all'evoluzione tecnologica dei trasduttori, mentre l'uso della III finestra si è reso possibile dopo essere riusciti a limitare la *dispersione cromatica* delle fibre (vedi appresso).

Un'altra fonte di attenuazione può avere origine dalle *giunzioni* tra tratte di fibre ottiche: l'uso di connettori produce una perdita di $0.4 \div 1$ dB, ed i giunti meccanici $\simeq 0.2$ dB oppure anche 0,05 dB se ottimizzati per via strumentale. Si possono infine *fondere* tra loro le fibre, con perdite tra 0,01 e 0,1 dB.

Dispersione cromatica e trasduttori elettro-ottici Dopo aver ridotto od eliminato il fenomeno di dispersione modale, si è individuata una ulteriore causa di dispersione temporale dell'energia immessa nella fibra ottica: il problema si verifica se il segnale di ingresso non è perfettamente monocromatico, ovvero sono presenti diverse lunghezze d'onda. Dato che il valore dell'indice di rifrazione dipende dalla lunghezza d'onda, λ diverse si propagano con velocità differenti e raggiungono l'altro estremo della fibra in tempi successivi. La dispersione cromatica della fibra si misura in [psec/Km·nm], e dà luogo ad una dispersione temporale tanto maggiore quanto più è lunga la fibra, e quanto più è estesa la gamma cromatica $\Delta\lambda$ della sorgente²⁸. Per ridurre il fenomeno è possibile:

- Utilizzare una lunghezza d'onda λ per la quale la dispersione cromatica è ridotta. Ad esempio, una fibra di silicio produce una dispersione cromatica 15 volte inferiore a $1.3 \mu\text{m}$ che non a $1.5 \mu\text{m}$.
- Scegliere una sorgente con la minima estensione cromatica $\Delta\lambda$ possibile.

Per ciò che riguarda il secondo punto, i trasduttori usati per primi sono stati gli economici LED (*Light Emitting Diode*), che richiedono una circuiteria di interfaccia semplice,

²⁸Il fenomeno della dispersione cromatica è l'equivalente ottico della distorsione di fase (o distorsione di ritardo) studiata per i segnali elettrici.

sono poco sensibili alle condizioni ambientali, e quindi risultano affidabili. D'altra parte, i LED emettono luce su più lunghezze d'onda, mentre per limitare la dispersione cromatica (e quindi raggiungere frequenze di segnalazione più elevate) occorre ricorrere ai *Diodi Laser* (LD).

I LASER forniscono anche una maggiore potenza, e quindi divengono indispensabili per coprire distanze maggiori²⁹. D'altra parte sono più costosi, hanno vita media ridotta rispetto ai LED, e richiedono condizioni di lavoro più controllate. Notiamo inoltre che una fibra ottica posta inizialmente in opera mediante sorgenti

LED, può essere potenziata (in termini di banda) sostituendo il LED con il LASER.

L'uso di sorgenti che operano in III finestra, che (presentando una attenuazione ridotta) permette di operare con tratte più lunghe, obbliga in generale a ridurre la frequenza di segnalazione, a causa della maggiore dispersione cromatica. Quest'ultima limitazione è stata rimossa da un particolare tipo di fibra, detta *dispersion shifted*, che presenta un minimo di dispersione cromatica D in III finestra anziché in II, e che raggiunge valori migliori di 3.5 psec/Km·nm.

Sorgente	λ (nm)	W_{dT} (dBm)	$\Delta\lambda$ (nm)
Si LED	850	-16	50
Ge LED	1300	-19	70
InGaAsP LED	1300	-10	120
DFB LASER	1300	-5	1
DFB LASER	1550	-5	0.4
IL/DFB LASER	1550	+2	0.8

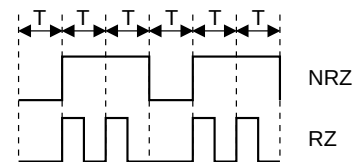
Prodotto Banda-Lunghezza e Codici di linea Come anticipato, la dispersione cromatica D risulta proporzionale alla lunghezza del collegamento L ed all'estensione cromatica $\Delta\lambda$ della sorgente. Se pensiamo di effettuare una trasmissione con codici NRZ e periodo $T = \frac{1}{f_L}$, ed imponiamo che la dispersione temporale sia non maggiore di $\frac{1}{4}T_L$, otteniamo $D \cdot L \cdot \Delta\lambda \leq 0.25 \cdot T$, in cui D è la dispersione cromatica [psec/Km·nm], L è la lunghezza [Km], $\Delta\lambda$ è l'estensione cromatica della sorgente [nm], e T è la durata di un bit [psec]. Associando ora il concetto di *banda* B alla frequenza di segnalazione $f_L = \frac{1}{T_L}$, si può affermare che il *prodotto della banda per la lunghezza* è pari al valore

$$PBL_{NRZ} = B \cdot L = \frac{.25}{D \cdot \Delta\lambda} \quad [\text{Tbps} \cdot \text{Km}]$$

che è una grandezza dipendente dalla coppia fibra-sorgente³⁰. Inserendo i valori di $\Delta\lambda$ (della sorgente) e D (della fibra), si ottiene una costante da usare per calcolare la banda (frequenza) massima trasmissibile per una data lunghezza (o viceversa).

Qualora si usi un codice RZ, i cui simboli hanno durata metà del periodo di bit T , la dispersione temporale tollerabile può essere elevata al 50% di T , e quindi in questo caso il prodotto banda-lunghezza risulta doppio rispetto al caso precedente:

$$PBL_{RZ} = \frac{.5}{D \cdot \Delta\lambda} = 2 \cdot PBL_{NRZ}$$



La tabella seguente riporta i valori di PBL_{NRZ} per alcune coppie fibra-sorgente.

²⁹La potenza emessa da un LASER non può aumentare a piacimento: oltre un certo valore intervengono infatti fenomeni non lineari, e la luce non è più monocromatica, causando pertanto un aumento della dispersione cromatica.

³⁰In questo senso, il prodotto *banda-lunghezza* costituisce un parametro di sistema che tiene conto di un concorso di cause. Un pò come il concetto di *tenuta di strada* di una autovettura, che dipende da svariati fattori, come il peso, i pneumatici, la trazione, il fondo stradale....

Fibra	Sorgente	λ (nm)	PBL (Gbps·Km)
Multimodo	LED	820	0.013
Graded Index	LED	820	2
Monomodo	LASER	1300	250
Monomodo	LASER	1550	36.8
Monomodo DS	LASER	1550	179

Esercizio determinare la lunghezza massima di un collegamento in fibra ottica monomodo, operante con $\lambda=1.3 \mu\text{m}$, e che garantisca una velocità $f_L=417$ Mbps, assumendo un guadagno di sistema di 42 dB (ovvero disponendo di una potenza di trasmissione 42 dB maggiore della minima potenza necessaria in ricezione).

Soluzione Si trova che per $\lambda=1300$ nm, l'attenuazione chilometrica è di 0,35 dB/Km, e questo valore determina una lunghezza non superiore a $\frac{42}{0.35}=120$ Km, che identifica il *Limite di Attenuazione* del collegamento. Verifichiamo quindi che non intervenga un limite più stringente per causa della dispersione cromatica. Supponendo di utilizzare la sorgente laser in grado di conseguire un PBL di 250 Gbps·Km, si ottiene una lunghezza massima pari a $\frac{250.000}{417}=600$ Km, che costituisce il *Limite di Dispersione*.

L'esercizio svolto ha lo scopo di mostrare la metodologia di progetto per un collegamento in fibra ottica, in cui vengono calcolati entrambi i limiti di *Attenuazione* e di *Dispersione*, e la massima lunghezza del collegamento è determinata dal vincolo più stringente. Nel caso dell'esercizio, la lunghezza è determinata dal limite di attenuazione, ed il progetto può essere rivisto utilizzando una sorgente *meno pura* per risparmiare, oppure una sorgente *più potente* per aumentare il guadagno di sistema e conseguentemente migliorare il limite di attenuazione. In questo secondo caso, può essere opportuno prestare attenzione al fatto che, aumentando la potenza di emissione, la purezza cromatica della sorgente può degradare (in quanto si verifica un aumento di $\Delta\lambda$ dovuto a fenomeni non lineari) con un conseguente peggioramento del limite di dispersione; è pertanto possibile ricercare la soluzione di migliore compromesso tra potenza di emissione e purezza spettrale.

Qualora non si riesca a rientrare nelle specifiche di progetto con una unica tratta, occorrerà suddividere il collegamento in più tratte, collegate da ripetitori rigenerativi, oppure ripartire la banda su più fibre poste in parallelo, ovvero bilanciarsi tra queste due soluzioni.

Trasduttori elettro-ottici La conversione del segnale uscente dalla fibra ottica meriterebbe una ampia trattazione approfondita, ma qui ci limitiamo a riferire esclusivamente poche cose fondamentali.

Il trasduttore utilizzato fin dall'inizio, economico ed affidabile, è il diodo P-I-N, che però non è adatto all'impiego con λ più elevate. Un secondo tipo di trasduttore molto usato è il diodo APD (*Avalanche Photo Detector*), caratterizzato da un "effetto valanga" che lo rende più sensibile di 10-15 dB rispetto ai P-I-N; d'altra parte però gli APD sono più delicati, e più sensibili alla temperatura.

La tabella che segue riporta i valori di sensibilità W_R (ossia la minima potenza che è necessario ricevere) di diversi fotorivelatori, per una probabilità di errore per bit $P_e = 10^{-11}$.

Fotorivelatore	λ (nm)	Sensibilità (dBm)	f_b (Mbps)
Si P-I-N	850	-48	50
Si APD	850	-58	50
InGaAs P-I-N	1310	-35	420
InGaAs APD	1310	-43	420
InGaAs APD	1550	-37.5	678

Nella tabella è riportato anche il valore della frequenza di segnalazione f_b a cui si riferisce la sensibilità, in quanto le prestazioni conseguite dal decisore che si trova a valle del trasduttore dipendono, come noto, da $\frac{E_b}{N_0}$, in cui E_b è l'energia per bit che vale $E_b = W_R \cdot T_b = \frac{W_R}{f_b}$. Pertanto, i trasduttori dimezzano la sensibilità (che aumenta di 3 dB) se la velocità f_b raddoppia, in quanto si dimezza l'energia per bit E_b . La sensibilità a frequenze diverse da quelle in tabella può quindi essere calcolata come³¹

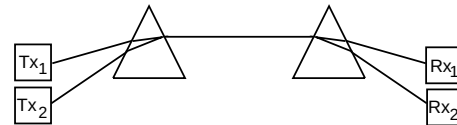
$$W_R(f'_b) \text{ [dBm]} = W_R(f_b) \text{ [dBm]} + 10 \log \frac{f'_b}{f_b}$$

13.4.3 Multiplazione a divisione di lunghezza d'onda - WDM

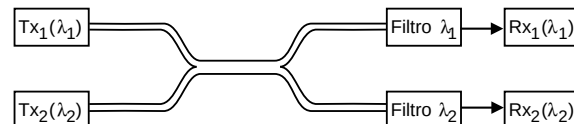
Abbiamo visto come nelle fibre ottiche siano presenti tre diverse finestre di valori di lunghezza d'onda per cui si ha una bassa attenuazione; è possibile effettuare allora tre diverse comunicazioni nella stessa fibra, ognuna nella propria finestra ! L'acronimo WDM (*Wavelength Division Multiplex*) identifica proprio questa tecnica di multiplazione, estesa al caso in cui le diverse comunicazioni avvengano anche nella medesima finestra.

Il modo più semplice ed intuitivo di realizzare la multiplazione di lunghezza d'onda è di adottare dei *rifrattori prismatici*, realizzando un circuito ottico del tipo illustrato in figura³².

I dispositivi di multiplazione di forma d'onda del tipo descritto vengono detti *passivi e reversibili*, in quanto non necessitano di alimentazione, ed uno stesso apparato può indifferentemente svolgere una funzione e la sua inversa. La passività del WDM rende questa tecnica attraente, qualora si pensi di distribuire fibre ottiche di casa in casa: ognuno avrebbe una sua lunghezza d'onda λ_i , e la fibra sarebbe una per tutto il condominio.



Se le λ_i sono troppo vicine tra loro (con una spaziatura dell'ordine di 0.015 nm), allora i prismi non riescono più nel compito di separazione geometrica, e conviene ricorrere ad una separazione della potenza (si fa uscire parte di segnale luminoso da *jacket*) ed un filtraggio (realizzato otticamente mediante gelatine) di ognuna delle λ_i . Così facendo però si perde molta potenza.



³¹Questo metodo di calcolo della sensibilità ad una f_L diversa da quella delle tabelle è approssimato, in quanto nei trasduttori avvengono fenomeni non-lineari che legano il livello di potenza del rumore alla potenza di segnale ricevuta. Trascurando questo effetto, si può applicare l'espressione sopra riportata.

³²Si sfrutta il principio "dell'arcobaleno" (o dei *Pink Floyd* ?), in quanto uno stesso materiale (il prisma) presenta indici di rifrazione differenti per lunghezze d'onda diverse, e quindi è in grado di focalizzare più sorgenti di diverso colore in un unico raggio.

13.4.4 Ridondanza e pericoli naturali

Le fibre vengono normalmente interrate, e per questo sono esposte ai pericoli di essere mangiate da talpe e topi, o di essere interrotte a causa di lavori stradali od agricoli. Quelle sottomarine sono a rischio per via di squali e reti a strascico. E' più che opportuno prevedere una adeguata ridondanza, in modo che in caso di interruzione di un collegamento sia possibile deviare tutto il traffico su di un altro collegamento.

13.4.5 Sonet

<i>Data Rate (Mbps)</i>	<i>Sigla CCITT</i>
51.84	
155.52	STM-1
466.56	STM-3
622.08	STM-4
933.12	STM-6
1244.16	STM-8
1866.24	STM-12
2488.32	STM-16

Questo è l'acronimo di *Synchronous Optical Network* ed è uno standard *mondiale* definito allo scopo di permettere l'interconnessione diretta tra reti in fibra ottica. Lo stesso standard è noto anche come *Synchronous Digital Hierarchy* (SDH). L'unità di moltiplicazione fondamentale è un flusso numerico da 51.84 Mbps, ovvero 3 volte tanto (155,52 Mbps) per l'Europa.

13.5 Appendici

13.5.1 Allocazione delle frequenze radio

La tabella che segue non pretende di essere completa né tantomeno esatta. Con le prossime edizioni del testo sarà progressivamente compilata.

<i>Intervallo</i>	λ	<i>Sigla</i>		<i>Uso</i>
30 - 300 Hz	$10^4 - 10^3$ Km	ELF	<i>Extremely Low</i>	
.3 - 3 KHz	$10^3 - 10^2$ Km	VF	<i>Voice Frequency</i>	
3 - 30 KHz	100 - 10 Km	VLF	<i>Very Low</i>	Radionavigazione a largo raggio. Attività nucleare.
30 - 300 KHz	10 - 1 Km	LF	<i>Low Frequency</i>	Radiolocalizzazione marittima ed aeronautica
.3 - 3 MHz	.1 - 1 Km	MF	<i>Medium Frequency</i>	Comunicazioni aeree e marittime. Radionavigazione. Broadcast AM
3 - 30 MHz	10 - 100 metri	HF	<i>High Frequency</i>	Collegamenti a lunga distanza fissi e mobili. Radioamatori.
30 - 300 MHz	1 - 10 metri	VHF	<i>Very High</i>	Broadcast FM e TV. Collegamenti in visibilità. Radiomobili civili e militari.
.3 - 3 GHz	.1 - 1 metro	UHF	<i>Ultra High</i>	Ponti radio e radiomobili terrestri. Broadcast TV. Satelliti meteo e TV.
3 - 30 GHz	10 - 100 mm	SHF	<i>Super High</i>	Ponti radio terrestri. Satelliti. Radar.
30 - 300 GHz	1 - 10 mm	EHF	<i>Extremely High</i>	Radar

13.5.1.1 Bande di frequenza Radar

Oltre alle bande HF, VHF ed UHF, le trasmissioni radar che operano in SHF ed EHF distinguono tra i seguenti intervalli di frequenze:

Intervallo (GHz)	1-2	2-4	4-8	8-12	12-18	18-27	27-40	40-75	75-110	110-300
Banda	L	S	C	X	K _u	K	K _a	V	W	millimetriche

13.5.1.2 Canali televisivi

VHF: Numerati da 1 a 6 a partire da 55.25 MHz, spaziatì di 6 MHz, fino a 83.25 MHz; numerati da 7 a 13 a partire da 175.25 MHz, fino a 211.25 MHz, ancora spaziatì di 6 Mhz. Nell'intervallo 88-108 Mhz e' presente il broadcast FM.

UHF: Numerati da 14 a 69 a partire dalla portante video di 471.25 MHz, fino a 801.25 MHZ, spaziatì di 6 MHz.

Capitolo 14

Rumore Termico

Descriviamo ora la natura e le fonti del processo di rumore sempre presente negli apparati di telecomunicazione, e di come questo sia tenuto in considerazione nel progetto degli stessi.

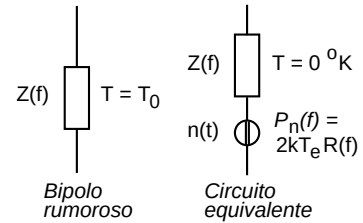
14.1 Rumore nei bipoli passivi

Ai capi di un resistore R a temperatura T è presente una *tensione a vuoto* $n(t)$, realizzazione di un processo gaussiano a media nulla, che è l'effetto del moto caotico degli elettroni all'interno della resistenza¹. Lo spettro di densità di potenza della tensione a vuoto ha espressione

$$\mathcal{P}_n(f) = 2R \frac{\hbar f}{e^{\frac{\hbar f}{kT}} - 1} \simeq 2kTR \quad [\text{Volt}^2]$$

in cui $k = 1.38 \cdot 10^{-23}$ joule/°K è la costante di Boltzman ed $\hbar = 6.62 \cdot 10^{-34}$ Joule·sec è la costante di Plank: questi valori fanno sì che l'approssimazione $\mathcal{P}_n(f) \simeq 2kTR$ sia valida ad ogni frequenza di interesse.

In un bipolo passivo di impedenza $Z(f) = R(f) + jX(f)$, solamente la parte reale (componente resistiva) concorre a generare il processo di rumore termico, che pertanto possiede una densità di potenza $\mathcal{P}_n(f) \simeq 2kTR(f)$. Nel caso in cui il bipolo contenga *più* resistori a temperature diverse, si può definire una temperatura equivalente $T_e(f)$; un bipolo passivo equivale pertanto allo stesso bipolo non rumoroso (a temperatura zero), con in serie un generatore di rumore con densità di potenza $\mathcal{P}_n(f) \simeq 2kT_e(f)R(f)$. Questo generatore equivalente, è quindi descritto da una *potenza disponibile di rumore*



$$\mathcal{W}_{dn}(f) = \frac{\mathcal{P}_n(f)}{4R(f)} = \frac{1}{2}kT_e(f) \quad \left[\frac{\text{Watt}}{\text{Hz}} \right]$$

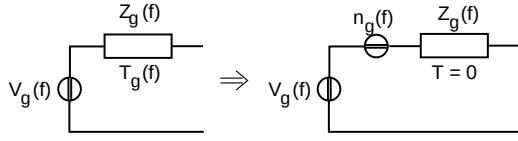
Nel caso in cui $T_e(f) = T_0 = 290$ °K (temperatura ambiente), il termine $kT_0 = 2\mathcal{W}_{dn}(f)$ assume i valori riportati di seguito per diverse unità di misura, da adottare in alternativa, allo scopo di rendere la grandezza omogenea con le altre che compaiono nelle formule di progetto:

¹Possiamo pensare che gli elettroni, qualora si trovino in maggior misura in una metà della resistenza, producano una differenza di potenziale negativa in quella direzione. Allo zero assoluto (- 273 °C) il moto caotico degli elettroni cessa, e si annulla così la tensione di rumore. Di qui l'aggettivo *termico* per descrivere il fenomeno.

$$kT_0 = -204 \text{ [dBW/Hz]} = -174 \text{ [dBm/Hz]} = -114 \text{ [dBm/MHz]}$$

Ad esempio, all'uscita di un filtro passa-banda ideale non rumoroso² di estensione 1 MHz, si ha una potenza disponibile di rumore pari a $10^{-11.4} \text{ mW}$.

14.2 Rapporto segnale rumore dei generatori

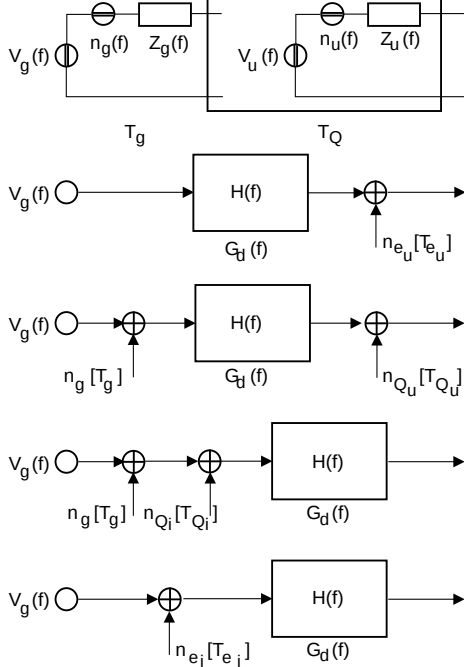


Un generatore di tensione $V_g(f)$, che possiede una propria impedenza interna $Z_g(f)$ a temperatura equivalente $T_g(f)$, produce anch'esso un processo di rumore in virtù della componente reale $R_g(f) = \Re\{Z_g(f)\}$ di $Z_g(f)$, e $Z_g(f)$ può quindi schematizzarsi con il circuito equivalente mostrato in figura. Pertanto, oltre alla potenza disponibile di segnale $\mathcal{W}_{dg}(f) = \frac{P_g(f)}{4R_g(f)}$, troviamo anche una potenza disponibile di rumore $\mathcal{W}_{dn}(f) = \frac{1}{2}kT_g(f)$, e dunque un rapporto segnale rumore disponibile

$$SNR_g(f) = \frac{\mathcal{W}_{dg}(f)}{\frac{1}{2}kT_g(f)}$$

che come osserviamo dipende da f , sia a causa di $\mathcal{W}_{dg}(f)$ che di $T_g(f)$.

14.3 Rumore nelle reti due porte



Se colleghiamo un generatore rumoroso all'ingresso di una rete due porte, è lecito aspettarsi all'uscita della rete un processo di rumore dipendente sia dal generatore che dalla rete, e la cui potenza disponibile $\mathcal{W}_{dnu}(f)$ può essere espressa in funzione di una temperatura equivalente di uscita $T_{eu}(f)$, tale che

$$\mathcal{W}_{dnu}(f) = \frac{1}{2}kT_{eu}(f)$$

D'altra parte a $T_{eu}(f)$ concorrono sia la temperatura del generatore $T_g(f)$, che la rete con una propria $T_{Qu}(f)$ "equivalente di uscita"; scriviamo dunque

$$\mathcal{W}_{dnu}(f) = \frac{1}{2}k \cdot [T_g(f) G_d(f) + T_{Qu}(f)]$$

in cui la potenza disponibile in ingresso alla rete con guadagno disponibile $G_d(f)$ è riportata in uscita, moltiplicandola per $G_d(f)$. Se effettuiamo l'operazione inversa per il contributo di rumore dovuto a T_{Qu} , otteniamo $\mathcal{W}_{dnu}(f) = \frac{1}{2}kG_d(f) \cdot [T_g(f) + T_{Qi}(f)]$ (in cui $T_{Qi}(f) = \frac{T_{Qu}(f)}{G_d(f)}$), ovvero

$$\mathcal{W}_{dnu}(f) = \frac{1}{2}kG_d(f) T_{ei}(f)$$

²Si intende dire che il filtro non introduce altro rumore oltre quello di natura termica.

in cui

$$T_{e_i}(f) = T_g(f) + T_{Q_i}(f) = T_g(f) + \frac{T_{Q_u}(f)}{G_d(f)}$$

è detta anche *temperatura di sistema* $T_s = T_{e_i}$, poiché riporta in ingresso alla rete tutti i contributi al rumore di uscita, dovuti sia al generatore che alla rete.

Siamo però rimasti con un problema irrisolto: che dire a riguardo di T_{Q_i} e T_{Q_u} ?

14.3.1 Reti passive

Supponiamo ora tutti i componenti alla stessa temperatura T_Q . In questo caso si può mostrare che risulta

$$\begin{cases} T_{Q_u}(f) &= [1 - G_d(f)] T_Q \\ T_{Q_i}(f) &= \frac{T_{Q_u}(f)}{G_d(f)} = [A_d(f) - 1] T_Q \end{cases}$$

in modo a poter scrivere:

$$\begin{cases} T_{e_u}(f) &= G_d(f) T_g(f) + T_{Q_u}(f) = G_d(f) T_g(f) + [1 - G_d(f)] T_Q \\ T_{e_i}(f) &= \frac{T_{e_u}(f)}{G_d(f)} = T_g(f) + [A_d(f) - 1] T_Q \end{cases}$$

Questo risultato evidenzia come per una rete passiva (con $0 \leq G_d \leq 1$), la temperatura di rumore equivalente in uscita sia una media pesata delle temperature del generatore e della rete. Nei casi limite in cui $G_d = 0$ oppure 1, la $T_{e_u}(f)$ è pari rispettivamente a T_Q e $T_g(f)$; infatti i due casi corrispondono ad una “assenza” della rete oppure ad una rete che non attenua.

14.3.2 Rapporto SNR in uscita

Se si valuta il rapporto segnale rumore in uscita alla rete, otteniamo

$$SNR_u(f) = \frac{\mathcal{W}_{dg}(f) G_d(f)}{\frac{1}{2}k T_{e_i}(f) G_d(f)} = \frac{\mathcal{W}_{dg}(f)}{\frac{1}{2}k \cdot [T_g(f) + [A_d(f) - 1] T_Q]}$$

Ricordando che il generatore in ingresso presenta un $SNR_i(f) = \frac{\mathcal{W}_{dg}(f)}{\frac{1}{2}k T_g(f)}$, possiamo valutare il peggioramento prodotto dalla presenza della rete:

$$\frac{SNR_i(f)}{SNR_u(f)} = \frac{\mathcal{W}_{dg}(f)}{\frac{1}{2}k T_g(f)} \cdot \frac{\frac{1}{2}k \cdot [T_g(f) + [A_d(f) - 1] T_Q]}{\mathcal{W}_{dg}(f)} = 1 + \frac{T_Q}{T_g(f)} \cdot [A_d(f) - 1]$$

14.3.3 Fattore di rumore

Il coefficiente $F(f) = 1 + \frac{T_Q}{T_g(f)} \cdot [A_d(f) - 1]$ è chiamato *fattore di rumore*³ della rete passiva, e rappresenta il peggioramento dell' SNR dovuto alla sua presenza. Notiamo subito che se $T_g(f) = T_Q$, allora $F = A_d$: una rete passiva che si trova alla stessa temperatura del generatore, esibisce quindi un fattore di rumore pari all'attenuazione. Infatti, mentre la potenza disponibile di rumore è la stessa (essendo generatore e rete alla stessa temperatura), il segnale si attenua di un fattore A_d .

³Il termine *fattore* trae origine dal fatto che l' SNR_i è ottenuto da quello di uscita mediante moltiplicazione per il fattore F : $SNR_i = F \cdot SNR_u$.

14.3.4 Reti attive

In questo caso il rumore introdotto dalla rete ha origine *non solo* dai resistori, e dunque *non è più vero* che $T_{Q_u}(f) = [1 - G_d(f)] T_Q$. Inoltre, il guadagno disponibile può assumere valori $G_d > 1$. In questo caso, si può esprimere l' SNR in uscita dalla rete come

$$SNR_u(f) = \frac{\mathcal{W}_{dg}(f) G_d(f)}{\frac{1}{2}k [G_d(f) T_g(f) + T_{Q_u}(f)]} = \frac{\mathcal{W}_{dg}(f)}{\frac{1}{2}k \cdot [T_g(f) + T_{Q_i}(f)]}$$

ed il peggioramento come

$$\frac{SNR_i(f)}{SNR_u(f)} = 1 + \frac{T_{Q_i}(f)}{T_g(f)} = F(f, T_g)$$

Quest'ultima espressione dipende ancora da T_g . Allo scopo di ottenere una grandezza che dipenda solamente dalla rete due porte, di definisce ora *fattore di rumore*⁴ il valore

$$F(f) = 1 + \frac{T_{Q_i}(f)}{T_0}$$

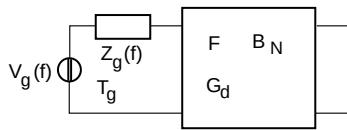
che rappresenta il peggioramento di SNR causato dalla rete quando il generatore è a temperatura ambiente $T_0 = 290$ °K = 17 °C. E' proprio $F(f)$ che *viene fornito* dal costruttore della rete due porte attiva, in modo da permettere il calcolo di $T_{Q_i}(f) = T_0 [F(f) - 1]$, e dunque

$$\frac{SNR_i(f)}{SNR_u(f)} = 1 + \frac{T_0}{T_g(f)} [F(f) - 1]$$

Riassunto

- Il fattore di rumore è definito come il peggioramento di SNR dovuto alla presenza della rete tra generatore e carico, quando il generatore è a temperatura $T_0 = 290$ °K = 17 °C.
- Se $T_g = T_0$ allora $T_{e_i}(f) = F(f) T_0$, e dunque la temperatura di sistema T_{e_i} è $F(f)$ volte quella del generatore;
- Se la rete non è rumorosa si ottiene $F = 1$ (pari a 0 dB);
- Se la rete è passiva allora $F(f) = [A_d(f) - 1] \frac{T_Q}{T_0} + 1$, e se $T_Q = T_0$ allora $F = A_d$;

Esempio



Sia data una rete due porte con assegnati guadagno disponibile G_d , banda di rumore B_N e fattore di rumore F . Valutare il rapporto segnale rumore disponibile in uscita nei due casi in cui il generatore si trovi ad una generica temperatura T_g oppure a T_0 .

Soluzione Sappiamo che la densità di potenza disponibile di rumore in uscita vale

$$\mathcal{W}_{dn_u}(f) = \frac{1}{2}k T_{e_i} G_d = \frac{1}{2}k \cdot [T_g + T_{Q_i}] \cdot G_d;$$

in generale $F = 1 + \frac{T_{Q_i}}{T_0}$ e quindi $T_{Q_i} = T_0 (F - 1)$, dunque

$$\mathcal{W}_{dn_u}(f) = \frac{1}{2}k \cdot [T_g + T_0 (F - 1)] \cdot G_d.$$

⁴A volte si incontra anche il termine *figura di rumore*, derivato dall'inglese NOISE FIGURE (che in realtà si traduce *cifra di rumore*), e che si riferisce alla misura di F in decibel.

Pertanto, la potenza disponibile di rumore risulta

$$\mathcal{W}_{dn_u} = k \cdot [T_g + T_0 (F - 1)] \cdot G_d B_N$$

che, nel caso in cui $T_g = T_0$, si riduce a $\mathcal{W}_{dn_u}(f) = kT_0 F G_d B_N$.

Per la potenza di segnale, si ha invece $\mathcal{W}_{ds_u} = \mathcal{W}_{dg} G_d$, e pertanto se $T_g = T_0$, risulta

$$SNR_u = \frac{SNR_i}{F} = \frac{\mathcal{W}_{dg}}{kT_0 F B_N}$$

ottenendo quindi lo stesso SNR in ingresso, ma con un rumore F volte più potente.

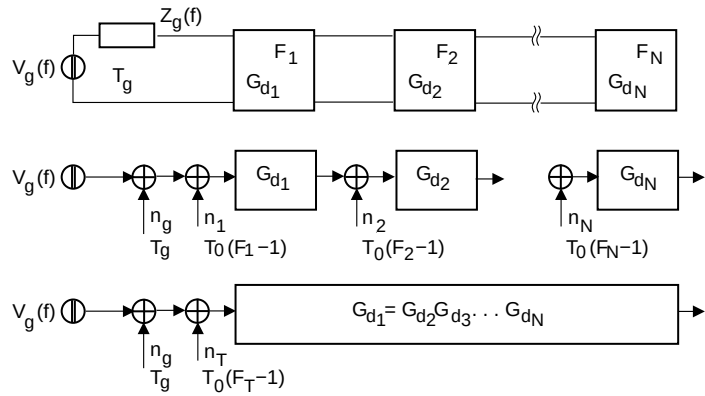
Nel caso in cui T_g sia generico, avendo posto il fattore di rumore indipendente dalla frequenza nella banda di rumore B_N , otteniamo:

$$SNR_u = \frac{SNR_i}{F(T_g)} = \frac{\mathcal{W}_{dg}}{kT_g B_N} \cdot \frac{1}{1 + \frac{T_{Q_i}}{T_g}} = \frac{\mathcal{W}_{dg}}{kT_g B_N} \cdot \frac{1}{1 + \frac{T_0(F-1)}{T_g}} = \frac{\mathcal{W}_{dg}}{k[T_g + T_0(F-1)]B_N}$$

14.3.5 Fattore di rumore per reti in cascata

Sappiamo che il guadagno disponibile dell'unica rete due porte equivalente alle N reti poste in cascata, è pari al prodotto dei singoli guadagni, ovvero $G_d = \prod_{n=1}^N G_{d_n}$. Come determinare invece il fattore di rumore equivalente complessivo?

Con riferimento alla figura riportata a lato, il singolo contributo di rumore dovuto a ciascuna rete può essere riportato all'ingresso della rete stessa, individuando così una temperatura $T_{Q_i}^{(n)} = T_0 (F^{(n)} - 1)$. I singoli contributi possono quindi essere riportati a monte delle reti che li precedono, dividendo la potenza (ovvero la temperatura) per il guadagno disponibile delle reti "scalvalcate". Dato che i contributi di rumore sono indipendenti, le loro potenze si sommano, e dunque è lecito sommare le singole temperature $T_{Q_i}^{(n)}$ riportate all'ingresso, in modo da ottenere un unico contributo complessivo di valore



$$T_{Q_i}^{(T)} = T_{Q_i}^{(1)} + T_{Q_i}^{(2)} \frac{1}{G_{d_1}} + T_{Q_i}^{(3)} \frac{1}{G_{d_1} G_{d_2}} + \dots + T_{Q_i}^{(N)} \frac{1}{\prod_{n=1}^{N-1} G_{d_n}}$$

in cui, sostituendo le espressioni per i $T_{Q_i}^{(n)}$ si ottiene

$$T_{Q_i}^{(T)} = T_0 \cdot \left[F_1 - 1 + \frac{F_2 - 1}{G_{d_1}} + \frac{F_3 - 1}{G_{d_1} G_{d_2}} + \dots + \frac{F_N - 1}{\prod_{n=1}^{N-1} G_{d_n}} \right]$$

Applicando la definizione $F^{(T)} = 1 + \frac{T_{Q_i}^{(T)}}{T_0}$, si ottiene

$$F^{(T)} = F_1 + \frac{F_2 - 1}{G_{d_1}} + \frac{F_3 - 1}{G_{d_1} G_{d_2}} + \dots + \frac{F_N - 1}{\prod_{n=1}^{N-1} G_{d_n}}$$

che costituisce proprio l'espressione cercata:

$$F^{(T)} = F_1 + \sum_{i=2}^N \frac{F_i - 1}{\prod_{j=1}^{i-1} G_{d_j}}$$

Il risultato si presta alle seguenti considerazioni:

- La *prima rete* due porte deve avere un F *più piccolo possibile*, in quanto quest'ultimo non può essere ridotto in alcun modo e contribuisce per intero ad $F^{(T)}$;
- la *prima rete* deve avere G_d *più elevato possibile*, in quanto quest'ultimo divide tutti i contributi di rumore delle reti seguenti.

Pertanto l'elemento che determina in modo preponderante il rumore prodotto da una cascata di reti due porte è la *prima* rete della serie, ed il suo progetto deve essere eseguito con cura particolare, anche tenendo conto del fatto che le due esigenze sopra riportate sono spesso in contrasto tra loro. E' inoltre appena il caso di ricordare che l'espressione ottenuta *non* è in dB, mentre spesso F è fornito appunto in dB; pertanto per il calcolo di $F^{(T)}$ occorre prima esprimere tutti gli F_i in unità lineari.

Esercizio Un trasmettitore con potenza di 50 mW e portante 30 MHz, modula AM un segnale con banda $\pm W = \pm 10$ KHz. Qualora si desideri mantenere un SNR in ricezione di almeno 25 dB, determinare la distanza che è possibile coprire adottando antenne isotrope, ed un ricevitore caratterizzato da un fattore di rumore $F = 10$ dB.

Svolgimento Assumendo che la trasmissione sia BLD-PS e che si verifichino le condizioni di massimo trasferimento di potenza, il valore desiderato $SNR = \frac{W_R}{W_N}$ può essere ottenuto se $W_R = W_N \cdot SNR = W \cdot N_0 \cdot SNR = B \cdot kT_0 F \cdot SNR$, e quindi

$$W_R(\text{dBm}) = 10 \log_{10} 10^4 (\text{Hz}) - 174 (\text{dBm/Hz}) + F_{dB} + SNR_{dB} = 40 - 174 + 10 + 25 = -99 \text{ dBm}.$$

Il guadagno di sistema risulta allora pari a

$$G_s(\text{dB}) = W_T(\text{dBm}) - W_R(\text{dBm}) = 10 \log_{10} 50 + 99 = 17 + 99 = 116 \text{ dB}.$$

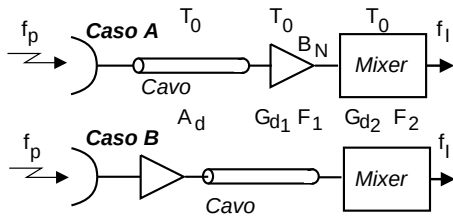
Non prevedendo nessun margine, l'attenuazione dovuta alla distanza è numericamente pari al guadagno di sistema, e pertanto scriviamo

$$A_d = 116 = 32.4 + 20 \log_{10} f(\text{MHz}) + 20 \log_{10} d(\text{Km}) = 32.4 + 29.5 + 20 \log_{10} d(\text{Km})$$

e quindi $2.7 = \log_{10} d(\text{Km})$, da cui $d = 10^{2.7} = 501 \text{ Km}$.

Svolgendo nuovamente i calcoli nel caso in cui il fattore di rumore del ricevitore sia pari a 20 dB e 100 dB, si ottiene che la nuova massima distanza risulta rispettivamente di 158 Km e di 15 metri.

Esercizio



Una trasmissione video modulata AM-BLU con portante $f_p = 2$ GHz viene ricevuta secondo uno dei due schemi in figura, indicati come caso **A** e **B**. E' presente una discesa in cavo coassiale con $\phi = 1.2/4.4$ mm lunga 50 metri, un filtro-amplificatore con guadagno disponibile $G_{d1} = 20$ dB, fattore di rumore $F_1 = .4$ dB e banda di rumore $B_N = 7$ MHz, ed un mixer che converte il segnale a frequenza intermedia f_I , e che esibisce $G_{d2} = 0$ dB e $F_2 = 10$ dB. Tutti i

componenti a valle dell'antenna si trovano alla stessa temperatura $T_0 = 290$ °K. Calcolare:

- 1) La minima potenza disponibile W_{dR} che occorre ricevere per ottenere $SNR_0 = 50$ dB nei due casi. Ripetere il calcolo supponendo l'antenna ricevente a temperatura $T_a = 10$ °K anziché T_0 .
- 2) La minima potenza che è necessario trasmettere per superare un collegamento terrestre lungo 50 Km, con antenne di guadagno $G_T = G_R = 30$ dB. Ripetere il calcolo per un down link satellitare in orbita geostazionaria, con $G_T = G_R = 40$ dB.

- 3) Il valore efficace della tensione ai capi del generatore equivalente di uscita dell'amplificatore di potenza del trasmettitore, per il caso migliore (tra **A** e **B**) del collegamento terrestre, nel caso di massimo trasferimento di potenza con $Z_u = Z_a = 50 \Omega$, oppure con $Z_u = 50 \Omega$ e $Z_a = 50 - j 50 \Omega$.

Svolgimento Determiniamo innanzitutto l'attenuazione del cavo coassiale, che risulta $A_d(f) = A_0 \sqrt{f(MHz)}$ dB/Km. Per il diametro indicato risulta $A_0 = 5.3$ dB/Km, ed alla frequenza di 2 GHz si ottiene $A_d(f)_{dB} = 5.3 \sqrt{2 \cdot 10^3} = 237$ dB/Km; e quindi in 50 metri si hanno $11.85 \approx 12$ dB. Dato che il cavo è a temperatura T_0 , risulta anche $F_{cavo} = A_d = 12$ dB. Riassumendo:

	$A_d = F_{cavo}$	F_1	G_{d1}	F_2	G_{d2}
dB	11.85	.4	20	10	0
lineare	15.3	1.1	100	10	1

1)

- A)** Il fattore di rumore complessivo risulta

$$F^A = F_{cavo} + A_d (F_1 - 1) + \frac{A_d}{G_{d1}} (F_2 - 1) = 15.3 + 15.3 \cdot (.6) + \frac{15.3}{100} (9) = 26 \rightarrow 14.1 \text{ dB.}$$

Dato che per la trasmissione televisiva AM-BLU risulta $SNR = SNR_0$, scriviamo

$$W_{dR} = SNR_i \cdot W_{dN} = SNR_0 \cdot F^A \cdot B_N \cdot kT_0 \text{ e quindi}$$

$$W_{dR}(\text{dBm}) = SNR_0(\text{dB}) + F^A(\text{dB}) + B_N(\text{dBMHz}) + KT_0(\text{dBm/MHz}) =$$

$$= 50 + 14.1 + 8.45 - 114 = -41.05 \text{ dBm}$$

- B)** Il fattore di rumore complessivo risulta ora

$$F^B = F_1 + \frac{(F_{cavo}-1)}{G_{d1}} + \frac{A_d}{G_{d1}} (F_2 - 1) = 1.1 + \frac{14.3}{100} + \frac{15.3}{100} (9) = 2.26 \rightarrow 3.5 \text{ dB.}$$

La differenza con il caso **A** è di 10.56 dB, e la potenza disponibile che occorre ricevere diminuisce pertanto della stessa quantità, e quindi ora risulta $W_{dR} = -51.6$ dBm.

Nel caso in cui $T_a = 10^\circ\text{K} \neq T_0$, non si ottiene più $T_{ei} = FT_0$, ma occorre introdurre la T_{Qi} della rete riportata al suo ingresso, e considerare la rete non rumorosa in modo da scrivere $T_{ei} = T_g + T_{Qi} = T_a + T_0 (F - 1)$. Ripetiamo i calcoli per i due casi **A** e **B**:

- A)** $W_{dR} = SNR \cdot W_{dN} = SNR \cdot B_N \cdot k \cdot (T_a + T_{Qi}) = SNR \cdot B_N \cdot k \cdot (T_a + T_0 (F^A - 1))$;

$$W_{dR}(\text{dBW}) = SNR(\text{dB}) + 10 \log_{10} 7 \cdot 10^6 + 10 \log_{10} (1.38 \cdot 10^{-23} (10 + 290 \cdot 25)) =$$

$$= 50 + 68.5 - 190 = -71.5 \text{ dBW} = -41.5 \text{ dBm}$$

- B)** $W_{dR}(\text{dBW}) = 50 + 68.5 + 10 \log_{10} (1.38 \cdot 10^{-23} (10 + 290 \cdot 1.26)) = -84.3 \text{ dBW} = -54.3 \text{ dBm}$

Notiamo che se la T_a è ridotta, le prestazioni per la configurazione **A** migliorano di soli 0.45 dB, mentre nel caso **B** il miglioramento è di circa 2.7 dB. Questo risultato trova spiegazione con il fatto che in **A** predomina comunque il T_{Qi} prodotto dal cavo.

- 2) In un collegamento radio terrestre si assume $T_a = 290^\circ\text{K}$. Inoltre, per il caso in esame si trova una attenuazione disponibile pari a

$$A_d = 32.4 + 20 \log_{10} f(\text{MHz}) + 20 \log_{10} d(\text{Km}) - G_T - G_R = 32.4 + 66 + 34 - 60 =$$

$$= 72.4 \text{ dB}$$

- A)** $W_{dT} = W_{dR} + A_d = -41.5 + 72.4 = 30.9 \text{ dBm} = .9 \text{ dBW} \rightarrow 1.23 \text{ Watt}$

- B)** $W_{dT} = W_{dR} + A_d = -54.3 + 72.4 = 18.1 \text{ dBm} = -11.9 \text{ dBW} \rightarrow 66 \text{ mWatt}$

Per il downlink si ha $d = 36.000$ Km, mentre $T_a = 10^\circ\text{K}$. Pertanto:

$$A_d = 32.4 + 20 \log_{10} f(\text{MHz}) + 20 \log_{10} d(\text{Km}) - G_T - G_R = 32.4 + 66 + 91.12 - 80 \\ = 109.5 \text{ dB}$$

e quindi, utilizzando il valore W_{d_R} ottenuto per il caso **B**, otteniamo

$$W_{d_T} = W_{d_R} + A_d = -54.3 + 109.5 = 55.2 \text{ dBm} = 25.2 \text{ dBW} \rightarrow 331 \text{ Watt}$$

- 3) Nel caso di adattamento, la potenza ceduta all'antenna T_x è proprio quella disponibile del generatore, e quindi si ha $W_{d_T} = \frac{\sigma_g^2}{4R}$, da cui

$$\sigma_g = \sqrt{W_{d_T} 4R} = \sqrt{66 \cdot 10^{-3} \cdot 4 \cdot 50} = 3.63 \text{ Volt}.$$

In caso di disadattamento, desiderando che la potenza ceduta all'antenna trasmittente rimanga la stessa, e supponendo le impedenze indipendenti dalla frequenza, scriviamo (in accordo alla relazione mostrata in (12.1))

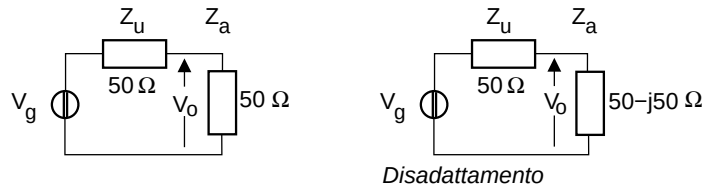
$$W_T = P_{v_o} \frac{R_a}{|Z_a|^2} = P_{v_o} \frac{50}{50^2 + 50^2} = P_{v_o} \cdot 10^{-2}$$

e quindi $P_{v_o} \simeq 6.6$ (Volt²). Applicando ora la regola del partitore, si ottiene

$$P_{v_o} = P_{v_g} \left| \frac{Z_a}{Z_a + Z_u} \right|^2 = P_{v_g} \left| \frac{50 - j50}{50 + 50 - j50} \right|^2 = P_{v_g} \frac{50^2 + 50^2}{100^2 + 50^2} = P_{v_g} \cdot 0.4.$$

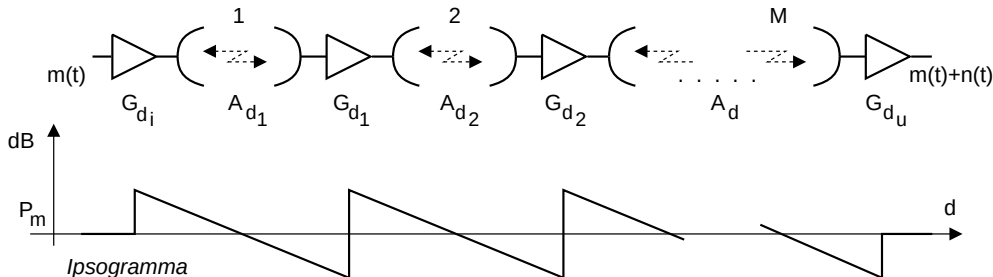
Dunque, $P_{v_g} = \frac{P_{v_o}}{0.4} = \frac{6.6}{0.4} = 16.5 \text{ Volt}^2$, ovvero $V_{g_{eff}} = \sqrt{16.5} \simeq 4 \text{ Volt}$.

Evidentemente, il disadattamento produce un innalzamento del valore efficace, se si vuol mantenere la stessa potenza di uscita.



14.3.6 Rumore nei ripetitori

Analizziamo il problema con riferimento ad un collegamento radio, anche se la trattazione può essere estesa ad altre tecniche trasmissive. Esaminiamo il caso in cui il collegamento da effettuare sia molto lungo, tanto da impedirne la realizzazione mediante un'unica *tratta*, od a causa dell'eccessiva attenuazione disponibile, oppure per la mancanza di condizioni di visibilità. In tal caso, occorre suddividere il collegamento in più *tratte* (in numero di M) dimensionate in modo tale che ognuna ripristini il livello di segnale ad un valore pari a quello in ingresso alla tratta stessa (tranne ovviamente l'ultima).



Tra ogni coppia di tratte si trova un *ripetitore*, che amplifica il segnale in misura pari al proprio guadagno disponibile $G_{d_i} = \frac{1}{A_{d_i}}$, e pari cioè all'inverso dell'attenuazione

disponibile della tratta precedente. Il rumore termico accumulato alla fine del collegamento può calcolarsi con i metodi tradizionali, ma considerando che il livello di segnale è lo stesso per tutti i ripetitori, si ottiene un risultato interessante, che ora illustreremo. Faremo quindi notare l'influenza della distorsioni *di non linearità* nel progetto.

14.3.6.1 Rumore termico accumulato

Osservando solamente il segnale in uscita dall'ultimo ripetitore, si può definire un SNR complessivo come $SNR_T = \frac{P_m}{P_n}$. D'altra parte, il rumore complessivo è dovuto ai contributi di rumore introdotti dai singoli ripetitori: essendo questi ultimi indipendenti tra loro, la potenza di rumore accumulata è la somma delle singole potenze di rumore:

$$P_n = \sigma_n^2 = E \{ n^2(t) \} = E \left\{ \left(\sum_i n_i(t) \right)^2 \right\} = E \left\{ \sum_i n_i^2(t) \right\} = \sum_i \sigma_{n_i}^2 = \sum_{i=1}^M P_{n_i}$$

Osserviamo ora che per ogni singolo ripetitore può essere definito un proprio $SNR_i = \frac{P_{m_i}}{P_{n_i}}$, e quindi $P_{n_i} = \frac{P_{m_i}}{SNR_i}$. Pertanto l' SNR complessivo risulta: $SNR_T = \frac{P_m}{\sum_i \frac{P_{m_i}}{SNR_i}}$. Notiamo ora che, essendo il rumore complessivo riferito ad un livello di segnale di riferimento, lo stesso deve avvenire per i singoli contributi P_{n_i} , cosicché nell'ultima espressione occorre considerare $P_{m_i} = P_m$ con $\forall i$, fornendo in definitiva

$$SNR_T = \frac{P_m}{P_m \sum_i \frac{1}{SNR_i}} = \frac{1}{\sum_i \frac{1}{SNR_i}}$$

Questo risultato può essere espresso con la frase

l' SNR prodotto da più cause indipendenti è il parallelo degli SNR dovuti alle diverse cause di rumore

per via della analogia formale con la resistenza equivalente di un parallelo di resistenze; l'analogia evidenzia, tra l'altro, che se una tratta è considerevolmente peggiore delle altre, SNR_T dipenderà essenzialmente da questa.

Il risultato a cui siamo giunti ha validità più generale del caso illustrato, e può essere invocato ogni volta che un sistema di comunicazione è affetto da più cause di disturbo additivo indipendenti tra loro, per ognuna delle quali si sia separatamente in grado di giungere ad una espressione di SNR .

Proseguiamo l'analisi ipotizzando ora che tutte le tratte siano uguali tra loro, ovvero con eguali A_d e G_d , uguali temperature di rumore, ed uguali SNR_i . In tal caso si ottiene

$$SNR_T = \frac{1}{\frac{M}{SNR_i}} = \frac{SNR_i}{M}$$

con $SNR_i = \alpha SNR_0 = \alpha \frac{P_R}{P_n}$, dove $P_n = kT_{e_i}W$ è la potenza di rumore nella banda di messaggio W , P_R è la potenza ricevuta da un ripetitore (uguale per tutti se le tratte sono uguali), e α è un fattore che dipende dal tipo di modulazione. Sembrerebbe dunque che per migliorare l' SNR complessivo sia sufficiente elevare il livello di trasmissione di tutti gli stadi, in modo da elevare la potenza ricevuta. In realtà la potenza trasmessa non può aumentare a piacere, in quanto intervengono fenomeni di non-linearità.

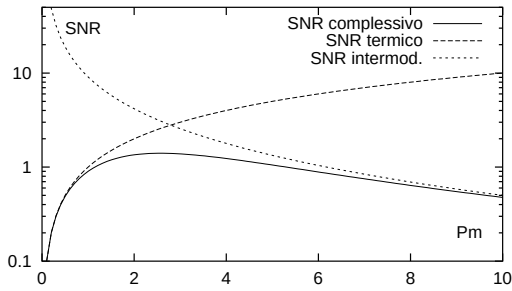
14.3.6.2 Compromesso tra rumore termico e di intermodulazione

A suo tempo si è osservato come per un segnale modulato, la presenza di un elemento a comportamento non lineare (tipicamente l'amplificatore di potenza del trasmettitore) produce interferenza in banda, la cui potenza dipende con legge cubica dalla potenza del segnale trasmesso. Indicando quindi con $SNR_I = \frac{P_m}{P_I}$ il rapporto SNR complessivo del collegamento dovuto a cause di non linearità, *indipendenti dal rumore termico*, osserviamo che questo *diminuisce* all'aumentare della potenza trasmessa da ogni ripetitore. Pertanto, l' SNR complessivo che tiene conto sia del rumore termico che di quello di intermodulazione, e che possiamo ottenere come "il parallelo" di entrambi, ossia

$$SNR = \frac{1}{\frac{1}{SNR_T} + \frac{1}{SNR_I}}$$

presenta un massimo per un certo valore di potenza trasmessa, ovvero esiste un dimensionamento ottimo in grado di fornire il miglior SNR complessivo.

Esempio



La figura a lato mostra l'andamento di

$$SNR = \frac{1}{\frac{1}{SNR_T} + \frac{1}{SNR_I}}$$

dovuto ai due termini

$$SNR_T = P_m \text{ e } SNR_I = \frac{P_m}{.1 \cdot P_m^2 + .01 \cdot P_m^3}$$

Come si vede, SNR presenta un massimo per $P_m \simeq 2.5$.

Bibliografia

- S. BARBAROSSA, T. BUCCIARELLI**, *Teoria dei Segnali*
- S. BENEDETTO, E. BIGLIERI**, *Teoria della Probabilità e Variabili casuali*, Quaderni di Elettronica, 1980 Boringhieri
- JOHN BELLAMY**, *Digital Telephony*, 1991 John Wiley and Sons, New York
- G. CANCELLIERI**, *Telecomunicazioni*, 2000 Pitagora editrice - Bologna
- A. B. CARLSON**, *Communication Systems*, 3rd Edition, 1986 Mc Graw Hill
- F. CUOMO**, *Telematica*, 2001 <http://net.infocom.uniroma1.it/tlem/lucidi/lucidi.php3>
- M. DECINA, A. ROVERI**, *Code e Traffico nelle Reti di Comunicazione*, 1991 Editrice La Goliardica - Roma
- M. DECINA, A. ROVERI**, *Introduzione alle Reti Telefoniche Analogiche e Numeriche*, 1989 Editrice La Goliardica - Roma
- M. G. DI BENEDETTO, P. MANDARINI**, *Comunicazioni Elettriche*, 2000 Editrice La Goliardica - Roma
- R. L. FREEMAN**, *Telecommunication System Enginnering*, 2nd Edition, John Wiley & Sons
- C. W. HELSTROM**, *Probability and Stochastic Processes for Engineers*, 2nd, 1991 Macmillan Publishing Company
- N. S. JAYANT, P. NOLL**, *Digital Coding of Waveforms*, 1984 Prentice-Hall, NJ
- M. LISTANTI, A. ROVERI**, *Comunicazioni Dati*, Appunti
- P. MANDARINI**, *Teoria dei Segnali*, 1979 Editrice La Goliardica - Roma
- A. V. OPPENHEIM, R. W. SHAFER**, *Digital Signal processing*, 1975 Prentice Hall, NJ
- A. PAPOULIS**, *Probability, Random variables, and Stochastic Processes*, 1991 McGraw-Hill Int.Eds.
- B. PERONI**, *Comunicazioni Elettriche*, Ed. Scientifiche Siderea, 1973 Roma
- A. ROVERI**, *Reti di telecomunicazione*, Appunti

- M. SCHWARTZ**, *Information Transmission, Modulation, and Noise*, 4th Edition, 1990, Mc Graw Hill
- R. STEELE**, *Mobile Radio Communications* 1992 Pentech Press London, 1994 IEEE press NJ
- F. G. STREMLER**, *Communication Systems*, 1990 Addison-Wesley
- A. S. TANENBAUM**, *Reti di Computer*, 1989 Gruppo Editoriale Jackson
- H. TAUB, D. L. SCHILLING**, *Principles of Comunication Systems*, 1986 Mc Graw Hill

Indice

1	Generalità	1
1.1	Trasmissione dell'informazione	1
1.2	Trasmissioni numeriche	2
1.3	Segnali analogici, certi ed aleatori	5
1.3.1	Rappresentazione di segnali analogici	6
1.3.2	Rappresentazione di processi aleatori	6
1.3.3	Transito dei segnali attraverso sistemi fisici	6
1.4	Segnali numerici	7
1.5	Teoria delle probabilità	8
1.6	Sistemi di Telecomunicazione	8
1.7	Segnali e sistemi	11
1.7.1	Caratteristiche dei sistemi	11
1.7.2	Caratteristiche dei segnali	12
	Segnale di potenza	12
	Segnale periodico	12
	Segnale di energia	12
	Segnale impulsivo	12
	Riassumendo	12
	Esempi di segnali di energia	13
	Esempi di segnale a durata limitata	13
1.7.3	Aspetti fisici delle grandezze energetiche	13
	Potenza istantanea	13
	Energia	13
	Potenza dissipata	14
	Valore efficace	14
2	Serie di Fourier	15
2.1	Prerequisiti trigonometrici	15
2.1.1	Numeri complessi	15
2.1.2	Formule di Eulero	16
2.1.3	Fasori	16
2.2	Serie di Fourier	17
2.2.1	Segnali reali	18
	2.2.1.1 Simmetria Coniugata	18
	2.2.1.2 Interpretazione degli X_n come fasori	19
	2.2.1.3 Serie Trigonometrica	19

	Esempio: Serie di Fourier di un'onda rettangolare	20
2.2.2	Serie di Fourier a banda limitata	21
2.3	Teorema di Parseval	22
	Ortogonalità degli esponenziali complessi	23
	Spettro di Potenza per segnali periodici	23
	Problema	23
	Soluzione	23
2.4	Appendici	24
2.4.1	Algebra Vettoriale	24
	Spazio normato	24
	Spazio dei segnali periodici	25
	Ri-definizione dei coefficienti di Fourier	25
	Disuguaglianza di Schwartz	25
2.4.2	Esempi di Sviluppo in serie	26
	Onda quadra simmetrica	26
	Treno di impulsi rettangolari	26
	Onda triangolare simmetrica	26
	Dente di sega simmetrico	27
	Rettificata a singola semionda	27
	Rettificata a onda intera	27
3	Trasformata di Fourier	29
3.1	Definizione	29
	Trasformata di un rettangolo	30
3.2	Densità di Energia	31
3.3	Prime proprietà della trasformata di Fourier	31
	Simmetria coniugata	31
	Dualità	31
	Linearità	32
	Valore medio e valore iniziale	32
	Traslazione nel tempo	32
	Conseguenze della linearità di fase	33
	Traslazione in frequenza (<i>Modulazione</i>)	34
	Coniugato	34
	Cambiamento di scala	34
3.4	Impulso matematico e trasformate notevoli	34
3.4.1	Trasformata di una costante	35
3.4.2	Trasformata per segnali periodici	35
	Trasformata di un coseno	36
3.4.3	Proprietà di <i>setacciamento</i>	36
3.5	Risposta Impulsiva e Convoluzione	36
3.5.1	Risposta Impulsiva	36
3.5.2	Integrale di convoluzione	37
3.5.3	$h(t)$ come funzione memoria	38
3.5.4	Convoluzione con l'impulso traslato	38
3.6	Moltiplicazione in Frequenza e nel Tempo	39
	Moltiplicazione in Frequenza	39

	Moltiplicazione nel Tempo (<i>Modulazione</i>)	40
3.7	Derivazione ed Integrazione nel Tempo	40
	Derivazione nel Tempo	41
	Integrazione nel Tempo	41
	Esempio: Trasformata di un triangolo	41
3.8	Trasformata di segnali periodici	42
	Treno di impulsi	43
	Segnale periodico	43
	Trasformata del treno di impulsi	43
	Trasformata di segnale periodico	44
3.9	Appendici	44
3.9.1	Esercizio	44
3.9.2	Sulla trasformata di una costante	44
3.9.3	Finestratura e stima spettrale	45
3.9.4	Trasformata discreta di Fourier	46
3.9.5	Trasformata di un gradino	48
3.9.6	Sintesi delle proprietà della trasformata di Fourier	48
4	Campionamento e Trasmissione dei Dati	49
4.1	Teorema del Campionamento	49
4.1.1	Aliasing	51
4.1.2	Energia di un segnale campionato	51
4.1.3	Uso pratico	52
4.2	Trasmissioni numeriche	53
4.2.1	Trasmissione numerica su canale analogico di Banda Base	53
	4.2.1.1 Segnale dati	53
	4.2.1.2 Requisiti per l'impulso di trasmissione	53
	Limitazione di banda	53
	Limitazione nel tempo	54
	Limitazione di precisione	54
	4.2.1.3 Condizioni di Nyquist	55
	Coseno Rialzato	56
4.2.2	Codifica di linea	58
	4.2.2.1 Codici di Linea a banda infinita	58
	4.2.2.2 Trasmissione multilivello	59
	4.2.2.3 Codice di Gray	60
4.2.3	Codifica di Canale	61
	4.2.3.1 Codici a blocchi	61
	Controllo di parità	61
	4.2.3.2 Codici a correzione	61
	Codice a ripetizione	62
	Distanza di Hamming.	62
	FEC e ARQ	62
4.2.4	Scrambling	62
4.3	Reti di trasmissione	63
4.3.1	Moltiplicazione di tempo numerica	63
	Trama PCM	64

	CAS e CCS	65
	Sincronizzazione di centrale	65
65		
4.3.3	Sincronizzazione di Rete	66
	Elastic Store	67
4.3.4	Messaggi di segnalazione	67
4.3.5	Topologia di rete	69
4.3.6	Instradamento	69
4.3.7	Commutazione	70
	4.3.7.1 Reti a divisione di spazio	70
	4.3.7.2 Reti multistadio	70
	4.3.7.3 Commutazione numerica a divisione di tempo	71
	Time Slot Interchanger	71
	Commutazione bidimensionale	72
4.4	Appendici	73
4.4.1	Multiplicazione a divisione di tempo mediante modulazione di am- piezza degli impulsi	73
4.4.2	Trasmissione Asincrona	73
4.4.3	Codici convoluzionali	74
	4.4.3.1 Diagramma di transizione	74
	Esempio: CC(2,1,3)	75
	4.4.3.2 Diagramma a traliccio	75
	4.4.3.3 Decodifica di Viterbi	76
	Riflessioni	77
5	Probabilità, Processi ed Errori	79
5.1	Teoria delle probabilità	79
5.1.1	Assiomi delle probabilità	80
5.1.2	Teoremi di base	80
5.1.3	Probabilità condizionali	80
5.1.4	Teorema di Bayes	81
5.1.5	Indipendenza statistica	82
5.1.6	Esempi riassuntivi	82
5.2	Variabili aleatorie	83
5.2.1	Funzioni di densità e di distribuzione di probabilità	83
	Istogramma	84
5.2.2	Medie, momenti e momenti centrati	84
5.2.3	Variabile aleatoria a distribuzione uniforme	85
5.3	Processi Stazionari ed Ergodici	86
5.3.1	Media di insieme	86
5.3.2	Medie temporali	87
5.3.3	Medie temporali calcolate come medie di insieme	87
5.3.4	Processi stazionari	88
5.3.5	Processi stazionari ed ergodici	88
	Esempio: la potenza di segnale	88
5.3.6	Riassumendo	89
5.3.7	Processo ad aleatorietà parametrica	89

5.4	SNR di Quantizzazione	90
5.5	Probabilità di Errore nelle Trasmissioni Numeriche di Banda base	91
5.5.1	Gaussiana	92
5.5.2	Probabilità di errore	93
5.5.3	Dipendenza di P_e da SNR	95
5.5.4	Dipendenza di P_e da E_b/N_0	96
5.5.5	Diagramma ad occhio	97
5.5.6	Uso del codice di Gray	98
5.5.7	Ricevitore ottimo	99
5.5.8	Esercizio	100
5.6	Appendici	102
5.6.1	Canale simmetrico binario	102
	Rapporto di verosimiglianza	102
	Verifica di Ipotesi di ML e Bayesiana	102
	Riflessioni	103
	Applicazione	103
5.6.2	Entropia e Codifica di Sorgente	103
	Codifica di Huffman e codifica entropica a lunghezza variabile	105
5.6.3	Misura delle grandezze in Decibel	106
5.6.4	Quantizzazione logaritmica	106
5.6.5	Funzione caratteristica	107
5.6.6	Trasformazioni di v.a. e cambio di variabili	108
	5.6.6.1 Caso unidimensionale	108
	5.6.6.2 Caso multidimensionale	109
	v.a. di Raileigh	110
	v.a. di Rice	111
5.6.7	Detezione di sinusoidi nel rumore	112
6	Traffico, Code e Reti a Pacchetto	115
6.1	Distribuzione Binomiale per popolazione finita	115
6.2	Distribuzione di Poisson ed Esponenziale negativa	117
6.3	Sistema di servizio orientato alla perdita	118
	6.3.1 Probabilità di rifiuto	119
	6.3.2 Efficienza di giunzione	120
	6.3.3 Validità del modello	121
6.4	Sistemi di servizio orientati al ritardo	122
	6.4.1 Risultato di Little	122
	6.4.2 Sistemi a coda infinita ed a servente unico	123
	6.4.3 Sistemi a coda finita e con più serventi	125
6.5	Reti per trasmissione dati	126
	6.5.1 Il pacchetto dati	127
	6.5.2 Modo di trasferimento delle informazioni	128
	6.5.2.1 Schema di moltiplicazione	128
	6.5.2.2 Principio di commutazione	128
	Commutazione di circuito:	129
	Commutazione di pacchetto a Circuito Virtuale:	129
	Commutazione di pacchetto a Datagramma	130

6.5.2.3	Architettura protocollare	131
	Stratificazione ISO-OSI	132
	Incapsulamento	133
	Indipendenza dei servizi tra pari rispetto al servizio di collegamento	133
6.6	Appendici	134
6.6.1	La rete Internet	134
6.6.1.1	Storia	134
6.6.1.2	Le caratteristiche	135
6.6.1.3	Gli indirizzi	136
	IP ed Ethernet	136
	Sottoreti	136
	Intranet	136
	DNS	137
	Indirizzi TCP	138
6.6.1.4	TCP	138
	Il pacchetto TCP	138
	Apertura e chiusura della connessione	139
	Protocollo a finestra	140
	UDP	141
6.6.1.5	IP	141
	L'intestazione IP	141
	Indirizzamento e Routing	142
	Subnetting e Supernetting	143
	Classless Interdomain Routing - CDIR	143
	Longest Match	143
	Sistemi Autonomi e Border Gateway	144
	Multicast	144
6.6.1.6	Ethernet	144
	Address Resolution Protocol - ARP	145
	Formato di pacchetto	146
	Collisione	146
	Trasmissione	146
6.6.1.7	Fast e Gigabit Ethernet	147
	Fast Ethernet	147
	LAN Switch	147
	Gigabit Ethernet	147
	Packet bursting	148
	Architettura	148
	Ripetitore full-duplex e controllo di flusso	148
	10 Gigabit Ethernet	149
6.6.2	Rete ATM	149
	Architettura	149
	Strato fisico	149
	Strato ATM	150
	Classi di traffico e Qualità del Servizio (QoS)	151
	Indirizzamento	153

	Strato di adattamento	153
	Ip su ATM classico	154
	LANE, NHRP e MPOA	155
	MPLS	156
7	Densità Spettrale e Filtraggio	159
7.1	Correlazione e Covarianza	159
7.1.1	Correlazione	160
7.1.2	Covarianza e Indipendenza Statistica	161
7.1.3	Statistiche dei Processi	162
7.1.4	Autocorrelazione	162
7.1.4.1	Proprietà dell'autocorrelazione	163
	Traslazioni temporali.	163
	Durata Limitata.	163
	Segnali Periodici.	163
	Massimo nell'origine.	163
	Simmetria coniugata:	163
7.2	Densità Spettrale	164
7.2.1	Teorema di Wiener	164
7.2.2	Esempi	164
	Processo armonico.	164
	Processo gaussiano bianco limitato in banda.	165
	Onda PAM.	165
7.3	Stima spettrale	166
7.3.1	Periodogramma	166
7.4	Filtraggio di segnali e processi	167
7.4.1	Segnali di energia	167
7.4.2	Segnali periodici	168
7.4.3	Processi ergodici	168
7.4.4	Esercizio	169
	Risposte	169
7.4.5	Filtro Adattato	170
	Segnalazione antipodale	171
	Segnalazione ortogonale	171
7.5	Caratteristiche dei sistemi fisici	172
7.6	Unità di elaborazione	173
7.6.1	Prodotto	173
7.6.2	Somma	174
7.7	Filtri Digitali	175
7.7.1	Filtro trasversale del 1° ordine	176
7.7.2	Stima della autocorrelazione di un processo ergodico	177
7.7.3	Filtro digitale a risposta impulsiva <i>infinita</i> del 1° ordine	177
7.8	Filtri Analogici	178
7.8.1	Filtro analogico ad un polo	178
7.8.2	Frequenza di taglio	179
7.8.3	Assenza di distorsioni lineari	179
7.9	Appendici	179

7.9.1	Coefficiente di Correlazione	179
7.9.2	Onda PAM	180
7.9.3	Potenza di un segnale dati	181
7.9.4	Autocorrelazione dell'uscita di un filtro	182
7.9.5	Grafici di esempio	182
8	Segnali modulati	183
8.1	Generalità	183
8.1.1	Multiplicazione a divisione di frequenza - FDM	183
	Collegamenti Punto-Multipunto.	184
	Accesso Multiplo.	184
	Collegamenti Punto-Punto.	185
8.1.2	Canale telefonico	185
8.1.3	Banda di segnale	186
8.1.4	Antenne e lunghezza d'onda	186
8.1.5	Trasmissione a banda laterale unica	186
8.2	Rappresentazione dei segnali modulati	187
8.2.1	Involuppo complesso e modulazione di ampiezza e/o angolare	187
8.2.2	Componenti analogiche di bassa frequenza	188
8.2.3	Filtro di Hilbert	189
8.2.4	Segnale Analitico	191
8.2.5	Esempi	192
8.3	Transito dei segnali in banda traslata nei sistemi fisici	192
8.3.1	Filtraggio	193
	Intermodulazione tra C.A. di B.F.	193
	Equalizzazione	193
8.3.2	Condizioni per $\underline{x}(t)$ reale	194
	Filtro passa banda ideale	194
	Simmetria coniugata attorno ad f_0	194
8.3.3	Estrazione delle componenti analogiche di bassa frequenza	194
8.4	Rappresentazione dei processi in banda traslata	195
8.4.1	Conclusioni	197
8.4.2	Processo gaussiano bianco limitato in banda	197
8.5	Appendice	198
8.5.1	Autocorrelazione di Processi Passa-Banda	198
9	Modulazione per Segnali Analogici	201
9.1	Modulazione di Ampiezza - AM	201
9.1.1	Banda Laterale Doppia - BLD	202
	9.1.1.1 Portante soppressa - PS	202
	9.1.1.2 Portante Intera - PI	202
	9.1.1.3 Portante parzialmente soppressa - PPS	203
	9.1.1.4 Efficienza di PI-PPS	203
9.1.2	Banda Laterale Unica - BLU	204
	9.1.2.1 Generazione di segnali BLU	204
9.1.3	Banda laterale ridotta - BLR	205
9.1.4	Potenza di un segnale AM	205

9.2	Demodulazione di ampiezza	205
9.2.1	Demodulazione coerente (omodina)	205
9.2.1.1	Errori di fase e frequenza	206
9.2.1.2	Demodulazione in fase e quadratura	206
9.2.1.3	Phase Locked Loop - <i>PLL</i>	207
9.2.2	Demodulatore di inviluppo	208
9.2.2.1	Segnali <i>Banda Laterale Unica e Ridotta</i>	208
9.2.3	Demodulatore eterodina	209
9.2.3.1	Frequenze immagine	210
9.3	Modulazione angolare	211
	Non linearità	211
	Ampiezza costante	212
	Generazione di un segnale a modulazione angolare	212
9.3.1	Ricezione di un segnale a modulazione angolare	212
9.3.1.1	Ricevitore a <i>PLL</i>	212
9.3.1.2	Ricevitore a discriminatore	213
9.3.2	Densità spettrale di segnali modulati angolarmente	213
9.3.2.1	Segnale modulante sinusoidale	214
	Modulazione a basso indice	216
	Modulazione ad alto indice	216
	Regola di Carson	217
9.3.3	Densità spettrale FM con processo aleatorio modulante	217
9.3.3.1	Indice di modulazione per processi	218
9.3.3.2	Modulazione a basso indice	218
9.4	Appendici	219
9.4.1	Calcolo della potenza di un segnale AM BLU	219
9.4.1.1	Calcolo della potenza di segnali BLD-PI, PS, PPS	219
9.4.2	Ricostruzione della portante mediante quadratura	220
9.4.3	Il mixer	220
9.4.4	Trasmissione televisiva	221
	Codifica dell'immagine	221
	Segnale televisivo in bianco e nero	222
	Formato dell'immagine	222
	Occupazione spettrale	222
	Segnale audio	223
	Segnale di cromaticità	223
	Sincronizzazione	223
	Interferenza	224
9.4.5	Modulazione FM a basso indice	224
9.4.6	FM BroadCast	224
10	Prestazioni delle Trasmissioni Modulate	227
10.1	Il rumore nei segnali modulati	227
10.1.1	Rapporto Segnale-Rumore e Banda di Rumore	227
10.1.2	Demodulazione di un processo di rumore	228
10.2	Prestazioni delle trasmissioni AM	229
10.2.1	Potenza di segnale e di rumore dopo demodulazione. SNR	229

10.2.1.1	BLD-PS	229
10.2.1.2	BLU-PS	230
10.2.1.3	BLD-PI	230
10.3	Prestazioni delle trasmissioni FM	231
10.3.1	Rumore dopo demodulazione FM	231
10.3.2	Caso di basso rumore	232
	Segnale presente	233
	Discussione dei passaggi	233
	Discussione del risultato	234
10.3.3	Caso di elevato rumore	235
10.3.4	Enfasi e de-enfasi	235
11	Modulazione numerica	237
11.1	Modulazione di Ampiezza e di Frequenza	237
11.1.1	BPSK	237
11.1.2	L-ASK	238
	Efficienza Spettrale ρ	238
11.1.3	L-FSK	239
	FSK ortogonale	240
11.1.4	Prestazioni di L-ASK	241
	Natura di E_b/N_0	242
11.2	Modulazione di fase	242
11.2.1	QPSK ed L-PSK	242
11.2.2	Prestazioni QPSK	243
11.2.3	Prestazioni L-PSK	244
11.3	QAM	246
11.3.1	Prestazioni di QAM	247
11.4	Altre possibilità	248
11.5	Appendici	250
11.5.1	Schema riassuntivo delle prestazioni	250
11.5.2	FSK Ortogonale	250
	Modulazione coerente	251
	Modulazione incoerente	251
	Verifica grafica	252
	Discussione sull'ottimalità per $L \rightarrow \infty$	252
11.5.3	OFDM	252
11.5.3.1	Rappresentazione nel tempo ed in frequenza	252
11.5.3.2	Architettura di modulazione	255
11.5.3.3	Efficienza dell'OFDM	257
11.5.3.4	Architettura di demodulazione	257
11.5.3.5	Prestazioni	258
	Calcolo della P_e per portante	258
11.5.3.6	Equalizzazione	262
	Modulazione differenziale	262
11.5.3.7	Sensibilità alla temporizzazione	262
11.5.3.8	Ottimalità	262
11.5.3.9	Codifica	263

12 Transito dei segnali nei circuiti	265
Elaborazione e Distorsione	265
Canale perfetto	265
12.1 Caratterizzazione dei circuiti	266
12.1.1 Numero di porte	266
12.1.2 Modelli di rappresentazione	266
12.1.3 Proprietà delle reti due porte	266
12.2 Bipoli	266
Passivi	266
Attivi	267
Esempio	267
12.3 Potenza di segnale e grandezze elettriche	267
12.3.1 Potenza assorbita da un bipolo	267
12.3.2 Misure di potenza	268
12.4 Connessione tra generatore e carico	268
12.4.1 Potenza disponibile	269
12.4.2 Assenza di distorsioni lineari	269
12.4.3 $Z_g(f)$ reale	269
12.5 Reti due porte	270
12.5.1 Modello circuitale	270
12.5.2 Schema simbolico	270
12.5.3 Trasferimento energetico	271
Guadagno di tensione	271
Guadagno di potenza	271
Guadagno disponibile	271
Collegamento di un generatore ad un carico mediante una rete due porte	272
Reti passive	272
Reti in cascata	273
Collegamento radio	273
12.6 Distorsioni lineari	273
12.6.1 Rappresentazioni di $H(f) = H(f) e^{j\varphi(f)}$	273
Guadagno in dB	273
Tempo di ritardo di gruppo	273
Esempio	274
12.6.2 Segnali di banda base	274
12.6.3 Segnali modulati	274
12.6.3.1 Segnali a banda stretta	275
12.6.3.2 Modulazione di ampiezza	275
BLD-PS	275
BLD-PI	275
BLU	275
12.6.3.3 Modulazione angolare	275
12.6.4 Calcolo dell'SNR	276
12.6.5 Equalizzazione	276
12.7 Distorsioni di non linearità	276
12.7.1 Ingresso sinusoidale	277

Fattori di intermodulazione	277
12.7.2 Ingresso aleatorio	277
12.7.3 Effetto sulla modulazione FM	279
12.8 Appendici	279
12.8.1 Potenza assorbita da un bipolo	279
12.8.2 Potenza ceduta ad un carico $Z_c(f) \neq Z_g^*(f)$	280
13 Mezzi Trasmissivi	283
13.1 Dimensionamento di un collegamento	283
Guadagno di Sistema	283
Margine di sistema	283
Attenuazione supplementare	284
Grado di servizio	284
13.1.1 Casi particolari	284
Segnale a banda stretta	284
Equalizzazione	284
13.2 Collegamenti in cavo	285
13.2.1 Costanti distribuite, grandezze derivate, e condizioni generali	285
Impedenza caratteristica	285
Costante di propagazione	285
Condizioni di chiusura	286
Quadripolo equivalente	286
Condizioni di adattamento	286
Condizione di Heaviside	286
13.2.2 Trasmissione in cavo	287
Effetto pelle	287
Equalizzazione	288
Diafonia	288
13.2.2.1 Casi limite	289
Cavo a basse perdite	289
Cavo corto	289
13.2.3 Tipologie di cavi per le telecomunicazioni	290
13.2.3.1 Coppie simmetriche	290
Linee aeree	290
Coppie ritorte	290
13.2.3.2 Cavo Coassiale	291
13.3 Collegamenti radio	292
13.3.1 Trasduzione Elettromagnetica	293
Antenna Isotropa	293
Antenna direttiva	293
Antenna ricevente	294
13.3.2 Bilancio energetico	294
Potenza ricevuta	294
Attenuazione di spazio libero	294
Attenuazione disponibile	295
Ipsogramma	295
13.3.3 Condizioni di propagazione e Attenuazioni supplementari	296

Perdite di accoppiamento	296
Assorbimento terrestre	296
13.3.3.1 Condizioni di visibilità	296
Orizzonte Radio	297
Ellissoidi di Fresnel	297
13.3.3.2 Cammini Multipli e Diffusione	297
Collegamenti radiomobili	298
Path loss	299
Slow fading	299
Fast fading	300
Dimensione di cella e velocità di trasmissione	300
13.3.3.3 Assorbimento Atmosferico	301
13.3.3.4 Dimensionamento di un collegamento soggetto a pioggia	301
13.3.3.5 Collegamenti in diversità	302
Diversità di frequenza	303
Diversità di spazio	303
13.4 Collegamenti in fibra ottica	303
13.4.1 Generalità	303
Natura fisica della fibra	303
Il segnale luminoso	303
Trasmissione ottica	303
Propagazione luminosa e indice di rifrazione	304
13.4.2 Propagazione luminosa	305
Dispersione modale	305
Attenuazione	306
Dispersione cromatica e trasduttori elettro-ottici	306
Prodotto Banda-Lunghezza e Codici di linea	307
Trasduttori elettro-ottici	308
13.4.3 Multiplazione a divisione di lunghezza d'onda - WDM	309
13.4.4 Ridondanza e pericoli naturali	310
13.4.5 Sonet	310
13.5 Appendici	310
13.5.1 Allocazione delle frequenze radio	310
13.5.1.1 Bande di frequenza Radar	311
13.5.1.2 Canali televisivi	311
14 Rumore Termico	313
14.1 Rumore nei bipoli passivi	313
14.2 Rapporto segnale rumore dei generatori	314
14.3 Rumore nelle reti due porte	314
14.3.1 Reti passive	315
14.3.2 Rapporto SNR in uscita	315
14.3.3 Fattore di rumore	315
14.3.4 Reti attive	316
Riassunto	316
Esempio	316
14.3.5 Fattore di rumore per reti in cascata	317

Esercizio	318
14.3.6 Rumore nei ripetitori	320
14.3.6.1 Rumore termico accumulato	321
14.3.6.2 Compromesso tra rumore termico e di intermodulazione	322
Esempio	322